A Room Acoustics Measurement System using Non-Invasive Microphone Arrays

by

Simon Edward Roper

A thesis presented to
The University of Birmingham
for the degree of
Doctor of Philosophy

Electrical, Electronic and Computer Engineering
School of Engineering
University of Birmingham
December 2009

Abstract

This thesis summarises research into adaptive room correction for small rooms and pre-recorded material, for example music of films. A measurement system to predict the sound at a remote location within a room, without a microphone at that location was investigated. This would allow the sound within a room to be adaptively manipulated to ensure that all listeners received optimum sound, therefore increasing their enjoyment.

The solution presented used small microphone arrays, mounted on the room's walls. A unique geometry and processing system was designed, incorporating three processing stages, temporal, spatial and spectral. The temporal processing identifies individual reflection arrival times from the recorded data. Spatial processing estimates the angles of arrival of the reflections so that the three-dimensional coordinates of the reflections' origin can be calculated. The spectral processing then estimates the frequency response of the reflection. These estimates allow a mathematical model of the room to be calculated, based on the acoustic measurements made in the actual room. The model can then be used to predict the sound at different locations within the room.

A simulated model of a room was produced to allow fast development of algorithms. Measurements in real rooms were then conducted and analysed to verify the theoretical models developed and to aid further development of the system. Results from these measurements and simulations, for each processing stage are presented.

Acknowledgements

I begin by thanking my mother, sisters Emma and Nicola, and wife Sonia for their support and belief throughout this project, and during my life to get me to this point. Without their gentle pressure, encouragement and help through the tough times I would not have got to University let alone completed this thesis.

I would like to thank my supervisor, Dr. Tim Collins for his guidance and advice throughout the last five years. Without his technical insight and patience I doubt I would have completed this work. Thanks to Gavin Philips who was also a significant help, spending many hours discussing algorithm details and helping my understanding.

I also thank all my friends who have listened and encouraged me after frustrating days, even if they did not understand what I was working on, particularly Ian and Heather Armitage. Michael Knowles has been a great motivator and kindly given up many hours to proof read this thesis.

Many thanks to all the team at M Squared Lasers where I now work, especially Bill Miller, for their flexibility in allowing me time to get this project completed between many strict deadlines for getting products finished.

Table of Contents

Table of figures

Table of tables

Chapter 1 Introduction

Historically the acoustics of a room could only be altered using architectural techniques. The study of room acoustics originally was focused on large rooms, often concert halls, lecture theatres and places of worship. In more recent times electronic products have been employed to alter the perceived acoustics of a room, for both live and playback situations. The number of such products has increased greatly in recent years as digital signal processing hardware has developed to allow more complex digital signal processing to be applied.

In 1996 the DVD format was released in Japan [CEA 2008]. This format supports discrete surround sound from Dolby and Digital Theater Systems (DTS). This technology proved very popular with consumers throughout the developed world, more so than its predecessor, the laser disc, and there was significant growth in consumer adoption of home cinema. A typical home cinema consists of a large screen, a surround sound processor and a surround sound speaker system. Surround sound is now available from a number of different sources, including downloadable 5.1 mixes [iTrax 2008], DVDA, SACD, Blueray disc, and digital television.

To enjoy surround sound as the producer intended, it is important for the loudspeakers and listeners to be in the correct locations within the room in accordance with the ITU BS 775 standard that was used for Dolby Digital and DTS 5.1 surround sound formats [ITU 1994]. The room's geometry and furnishings also have a large effect on the sound which the listeners will hear.

1.1 Research question

There are a number of existing theories, techniques and products which the designers believe increase the audio performance of home audio systems by manipulating the sound coming from the loudspeakers to better match the room's acoustics. These are detailed in Chapter 2. With all these systems, the acoustic measurements of the room and system are taken and then the microphones are removed from the room and the previously designed processing is applied to the signals before they are amplified and sent to the loudspeakers. The inherent problem with this approach is that the acoustics of the room change over time. For instance the temperature of the room may change and the furniture might get moved. The doors and windows may be opened, whereas the calibration measurements were done with them closed additionally the listener might not always sit in the same location.

This research is aimed at developing a measurement tool which could be used to improve the audio experience in home listening environments by alleviating the problem of outdated calibration data. The proposed solution is to predict the sound at points within the room, without putting a microphone at those locations. Knowing the listeners' locations, and an estimate of what they will hear would enable an adaptive system to be developed which could change with the listening environment and listeners' locations. There is also potential for dynamic compensation, depending on whether there are multiple listeners in the room, or a single listener. The listeners can be located in the room using a number of existing techniques; some are explained in Chapter 3.

The advantage to the user of a prediction system rather than a traditional one-off calibration procedure is that it is more convenient. The user would not have to worry about recalibrating the system as the room changes. Leading on from this, a requirement of the project was that the proposed solution must have the potential of being sold to consumers. This requirement guided some of the decisions which had to be made regarding the equipment to be placed in the room; where possible the solution was chosen which would be most suitable in a typical home living room.

The focus of the research was on the prediction of the sound rather than modifying the audio signals before they drive the loudspeakers, because this area has already been researched by others, as discussed in Chapter 2. It concentrated on small rooms, of the sizes typically found in homes.

1.2 Proposed system

The chosen solution to the problem of predicting the sound at a remote location was to place the minimum number of small arrays of microphones on the walls of the room. The number of arrays required was dependent on the room geometry: for a shoebox rectangular room, for example, two arrays were sufficient. These arrays were used to record the reflections produced when test tones were transmitted from the loudspeakers. Analysis of the recordings produced a model of the room's reflections, which could then be used to predict the sound at any location within the room. The analysis was performed in three stages. First the reflections were identified in time, allowing the distance to each reflection to be calculated. Next spatial processing was

applied, which resulted in the azimuth and elevation angles of the reflection, relative to the array, being estimated. A second, low frequency signal was then used to approximate the frequency response of the reflections.

The results from the temporal, spatial and frequency estimations were combined to produce a mathematical model of the room's reflections, produced by the output of the loudspeaker reflecting off objects within the room. This model could be used to predict the sound at different locations within the space, in a similar way as is done when simulating a room using the Image Source Method, described in Chapter 3, but using measured, acoustic data.

1.3 Thesis overview

Following this introductory chapter, chapters are arranged as follows:

Chapter 2 provides the background to the research project, explaining where the topic originated. A review of architectural acoustics leads into comments of what makes good room acoustics. The implications of having surround sound in a domestic environment and a synopsis of existing solutions to modify a room's acoustics, from theory through to commercial products are presented. Lastly the limitations of these solutions are identified.

Chapter 3 outlines possible approaches to predicting the sound at remote locations, along with a justification for the decision to use small microphone arrays. An overview of the proposed system is presented.

Chapter 4 looks at different methods of simulating the acoustics of a room. A model was required to provide a test bed in which solutions could be developed, tested and verified. The image source method was selected. The details of the implementation produced are summarised along with verification experiment results.

The transmit signal is critical; it must be designed so that the correct information is recorded for the post processing to extract the room's parameters. Also before, the recorded sound can be processed to locate each reflection spatially the reverberation needs to be separated into individual reflections. This work is shown in Chapter 5, whilst Chapter 6 presents the spatial processing of the signals. Here complexities of array design and direction of arrival algorithms for reverberant spaces are discussed. A bespoke array geometry was designed and implemented for practical experiments. Different direction of arrival processing techniques are detailed and the selected method developed to work with the room measurements. Results are presented in Chapters 5 and 6 for both a simulated and real measured room.

In Chapter 7 work undertaken to estimate the low frequency response of the reflections is discussed. Adaptive filters were found to be unsuitable so a maximum likelihood solution was used. Unfortunately, due to the high computational load of this algorithm, its development proved difficult and it was not possible to process all

of the experimental data. Results to demonstrate the concept are therefore presented. Discussions of how the estimated parameters can be combined into the room model for predicting the sound at the desired remote location are presented in Chapter 8.

Chapter 9 draws conclusions on the work undertaken; the solution's successes and shortcomings are discussed along with suggestions for future work.

1.4 Results presented

As each subsection of the processing was developed a large number of simulations and practical experiments were conducted to verify the decisions made, before the next section was investigated. Results are therefore presented throughout the thesis. Where possible a real room and a corresponding simulation of the real room were used to ensure that the algorithms developed would work within a real situation.

1.5 Original contributions

The prediction of what multiple listeners will hear, based on measured data, without a microphone at the listeners' locations has not previously been investigated. During this research project a unique room acoustics measurement system has been developed. The original contributions of this work are as follows:

- The time response of a reverberant room has been captured and separated into individual reflections using the matching pursuit algorithm.

- One of the design aims was that the solution would be small and discreet enough that it could comfortably fit into a domestic living room; this objective has successfully been met by a bespoke array geometry.
- The unique array geometry has allowed the direction of arrival algorithm to perform a very efficient two dimensional search.
- The direction of arrival algorithm has been developed to extract a large amount of information from the data available, successfully identifying first and second order images from real and simulated data.
- The microphone array has been implemented allowing real data to be captured. Often research into direction of arrival algorithms have only presented simulated results; this thesis throughout presents simulated results which are backed up with real experimental data.
- The problem of estimating the frequency responses of individual reflections has been investigated.

Chapter 2 Background

The aim of this thesis is to present a measurement tool which could be used to increase the fidelity and enjoyment of music and film in a typical home environment. Before this could be done, however, it was important to understand how the sound is affected by the room in which it is played back. There are five sections in this chapter:

- The origins of architectural acoustics
- What are good room acoustics?
- An introduction to surround sound.
- Existing solutions to modifying a room's acoustics
- Proposed improvements on existing solutions.

2.1 The origin of room acoustics - architectural acoustics

The environment in which music is played has influenced both architecture and the music being composed. It is thought that the tonal scale and melodic line in early European music developed, in part, due to the performances being conducted in caves, and later highly reverberant cathedrals. African music, however, which was performed outdoors, developed complex rhymes instead [Long et al 2005; Toole 2006].

The history of architectural acoustics can be traced back to the Greeks. As the need to communicate with large numbers of people for political and military reasons grew, amphitheatres were built. Their aim was to increase the intelligibility of speech for the audience. The best preserved example is the theatre built in Epidaurus, Greece in around 300 B.C. It has now been restored and it is used for performances. The audience frequently reports that it is astonishing how loud and clearly each listener, even in the rear rows, can hear the sounds from the stage without any electroacoustic amplification [Long et al 2005]. The seats were steeply tiered, much steeper than is needed solely to give a good line of sight to the stage, which reduced grazing attenuation.

With the birth of Christianity, churches spread throughout Europe. All these churches had very reverberant acoustics, which suited the slow chanting used in the worship; something done for the participants rather than the outside listener [Long et al 2005].

As commerce and towns grew, secular plays and music developed. The renaissance period was rich in musical composition, both sacred and secular. During this time plays were performed to smaller groups than the cathedrals of the time catered for. Often the performances took place in open courtyards which reduced reverberation with the high walls providing shielding from outside noise. High galleries gave good sightlines. The good intelligibility of speech allowed complex dialogue to be performed, as highlighted by the plays of William Shakespeare.

During the Baroque period (1600 to 1750) there were many developments in musical instruments: the violin family and harpsichord, used for ensemble music, were developed. So too were early wind instruments which later became the French horn, oboe and bassoon used today. Much of the music of this period was composed to be performed in small rooms, by about 25 musicians, where the room did not affect the acoustics significantly.

The Protestant Christian services developed differently to those of Catholics, with the Protestants relying more on the spoken word. Church architecture was adapted to better suit this style of service by reducing the size of churches, moving the pulpit into the centre and providing seating nearer the front. All these changes were aimed at increasing the intelligibility of speech by lessening the effects of reverberation. Some churches used drapes to absorb reflections [Forsyth 1985] and also popular was a Baldachino, a canopy above the pulpit. Baldachinos were designed specifically with acoustics in mind, to reflect the words of the priest into the congregation. An upward tilt and the smooth underside of the Baldachino help to guide the sound to the more distant members of the congregation [Cremer et al 1990].

Up to the Classical period (1720 to 1800), music had been performed in rooms usually built for other purposes. During the classical period, though, purpose-built concert halls were designed and music composed specifically for them. During this time mathematics had also developed and scientists started to study sound and vibration [Lindsay 1966].

Later, scale models of concert halls were built to analyze the acoustics of the room. Light or water waves were used to simulate sound waves. Schlieren photography was a technique first used by Sabine in 1913 [Cremer et al 1990]. An electric spark was generated and photographed, the photographs showing the reflected wavefronts and intensity. Because of the similarity of sound and light waves these models gave great insight into the early reflections and were used to guide the architecture of refurbishments or newly built auditoria. With the development of microphones, scale models were used for auralization, the wavelength of the transmitted signal being scaled by the same factor of the physical model, by adjusting the record and playback speeds [Rindel 2002]. These scale models were used until the 1970s, when computer models started to replace the physical models [Cremer et al 1990].

Many modern concert halls must be designed for all types of music and performing arts. As already stated, different music genres were developed to be performed in the rooms that were common at the time. Therefore a modern music venue will ideally have adjustable acoustics so the room can be set to best complement the music being performed. The Sage in Gateshead, UK, is a prime example. In the main auditorium there are panels in the roof which can be individually adjusted to alter the early reflections. Acoustic drapes are hidden in the walls so that the reverberation time can be altered [IOA 2007].

2.2 What are good room acoustics?

This is a very difficult question since people have different opinions as to what properties rooms must have to qualify as a 'good' room for music playback. Section 2.1 gives a brief summary of some of the important developments in architectural acoustics. A large proportion of the work undertaken on this subject has been concerned with concert halls. Only in recent times, with recorded media, have smaller rooms been studied. The goals of concert hall acoustic design are different from those of small rooms used for playback of recorded music. However, before either can be considered it is important to understand the impulse response of a room.

The impulse response of a room contains all the information about the effect the room has on the sound, at the specific location and time it was recorded. The impulse response of a room can be obtained in a number of ways by performing a recording using a microphone and processing the received signal. For example a starter pistol will produce an impulse which is recorded, or a swept sinusoidal wave can be transmitted from a loudspeaker which is later matched filtered to obtain the impulse response. Figure 1 shows an example of an impulse response.

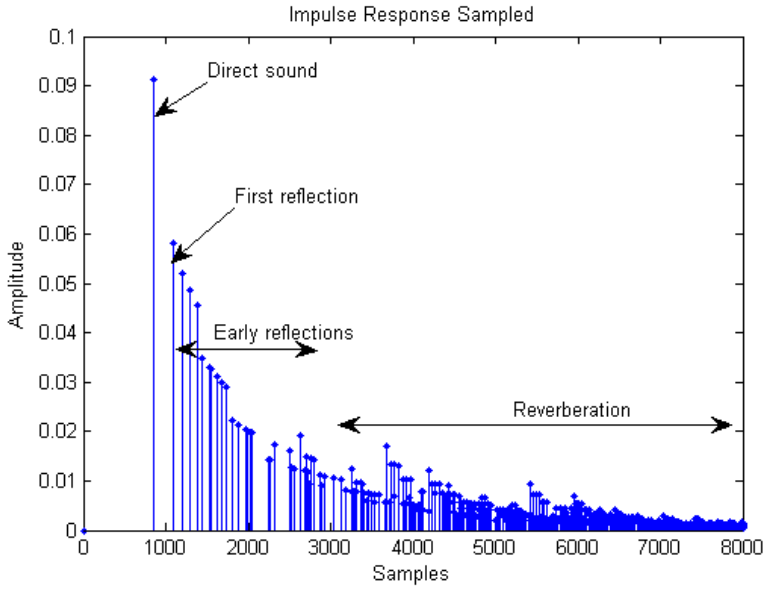


Figure 1: An impulse response.

The first signal arriving at the microphone is the direct sound, from the source. The distance to the source can be calculated from the difference between the transmitted and arrival time of the sound and the speed of sound, typically $344ms^{-1}$ at room temperature. The sound is reflected off surfaces and objects within the room and these reflections arrive at the microphone later, with reduced amplitude. The early reflections are caused by the sound reflecting off one, or two, surfaces. These early reflections are important for localization of the sound source and give information about the size of the listening space [Cremer et al 1990]. Ando [1998] quoted Bekesy's observation of the importance of sidewalls; Bekesy reported that a courtyard sound field was superior to that of any concert hall he had experienced. Early reflections were classified by Beranek [1996] in relation to concert halls, as those arriving before 80msec: he believes that these early reflections increase the apparent source width, which lends quality to the music heard. In comparison, Everest [2002] claims that for listening rooms (similar sized to living rooms) the early reflections

should be greatly reduced, using absorbers, so that the ratio of direct to reflected sound is increased, reducing the effect of comb filtering.

Angus's [2001] analysis of listening rooms found that the early reflections must arrive before 30ms to avoid the reflections being perceived as echoes. He believes that a short delay between the direct sound and early reflections can add 'intimacy' to the space, but that a longer delay improves the clarity of the sound; the ideal delay being within 20ms. The lateral reflections are important to add spaciousness to the sound; these ideally should have a flat frequency response. This agrees with Linkwitz's [2007] research on loudspeakers. Angus goes on to say that these rules are most applicable to concert halls and for recording studio control rooms where it is desirable to have a reflection free zone. This is where the early reflections are either absorbed or diffused to increase the delay between the direct sound and the reflections on the control room, thus allowing the engineer to listen to the space where the recording took place rather than the control room. Absorbing the early reflections will contradict the requirement of a diffuse reverberation field, so a balance is required. Toole [2006] points out that the reverberation is never diffuse in a living room due to the absorbing materials not being distributed evenly, for example carpets and sofas.

Linkwitz [2007] when considering small rooms, believes that there is great value in having loudspeakers with a uniform polar response over the audible frequency range. In this case the reflections caused by room reflections will be similar to the direct sound. He suggests that the human brain's cognitive facility is then better able to separate the static room acoustics from the acoustics embedded in the recording. He

hypothesizes that frequency dependent absorbers might, therefore, be undesirable, because the similarity between the direct and reflected sound is reduced.

As the sound reflects off the room's surfaces it becomes more diffuse and the reflections at the listener become more concentrated. The listener can no longer perceive where the reflections are coming from; something called the reverberant sound. This adds richness to music and helps integrate instruments together [Howard and Angus 2001]. Earlier, Knudsen [1929] suggested that the optimal reverberation time was longer for music than it was for speech, while MacNair [1930] urged that a longer reverberation time at low frequencies was desirable to supplement the loudness of music.

Over time, sound is absorbed by the surfaces, furniture and people in the room. The time it takes for sound to be absorbed is called the reverberation time, a term pioneered by W. C. Sabine [1912]. The RT60 reverberation time is specified as the time it takes for the steady state sound to reduce by 60dB of its original level, 60dB being the limit where the sound is perceived as inaudible after being cut off. This time is governed by the absorption of the room. Different materials reflect and absorb differently at different frequencies, so the reverberant time also varies with frequency. Raichel [2000] states that the reverberation time is still the most important parameter for gauging the acoustic performance of a room. It is used extensively in the design and quantitative analysis of concert and opera halls. Beranek [1996] has studied 76 such halls around the world. The average RT60 of the ten best halls, with regard to overall perceived sound quality, was 2 seconds.

The design goals of a concert hall can be summarised as follows: [Long et al 2005; Ando 1998; Beranek 1996]

1. Ensure a strong direct sound to all seats where possible.
2. Early reflections must arrive soon enough so that they are not perceived as echoes.
3. The audience should feel enveloped or surrounded by the sound, which requires strong lateral reflections.
4. Reflective surfaces close to the source or listener will give a short initial delay time gap, adding clarity.
5. The reverberant sound must be as diffuse as possible.
6. The reverberation time will depend on the music type, but should support the instruments.
7. The reverberation time should rise with frequencies below 500Hz to add warmth to the sound.
8. A wide bandwidth of sound should be supported, 30Hz to 12kHz.
9. Room modes should be evenly distributed and controlled.
10. The sound must have adequate loudness and background noise should be minimised.

Jackson and Leventhall [1972] measured the acoustics of 50 British living rooms and found that the average RT60 was 0.69 second at 125Hz and 0.4 second at 8kHz. These figures are slightly higher than research by Diaz and Pedrero [2005] who measured the reverberation time of over eleven thousand rooms in Madrid, Spain. They conducted measurements in bedrooms and living rooms, all with masonry walls

and ceilings together with heavy floor coverings. For rooms with a volume between 30 and 40m^3 they found that the average reverberation time at 125Hz was 0.55 seconds and at 4kHz it was 0.32 second.

At low frequencies sound does not behave in the same way as at high frequencies. At high frequencies (above about 200Hz for a typical room) the sound reflects diffusely off all the surfaces it contacts. At low frequencies this is not the case as sound reflects in cyclic paths, such that if the path length is a precise number of half wavelengths then they will form standing waves [Howard and Angus 2001]. These standing waves are different from the reverberant field because they are not chaotic, they do not strike every surface in the room with equal probability, and, in fact, they have a discrete path back to the sound source. This means that they only occur at discrete frequencies, determined by the room geometry.

These standing waves are called the modes of a room. The most basic modes are axial modes, which are caused by reflections off two surfaces (two opposite walls or floor and ceiling). Tangential modes are caused by reflections off four surfaces in a diamond shape, whilst oblique modes come from reflections off six surfaces. It is these standing waves which take longer to die away than other frequencies.

Wilson identified that frequencies where the reverberation time is significantly longer than the norm caused undesirable effects [Wilson et al 2003]. Research by Avis et al [2006] determined that the room modes as the biggest problem in reproducing music correctly; they are the main cause of acoustic variation between different rooms. He

notes that it is the Q factor, the decay time, of the modes which affects their subjective perception; reducing the Q and therefore decay time reduces the perception of the modal resonance. Waterfall plots can be used to compare the reverberation time against frequency, time and amplitude. Figure 2 shows an example of a waterfall plot, the dominant room modes can be identified at 32, 63, 70 and 92Hz.

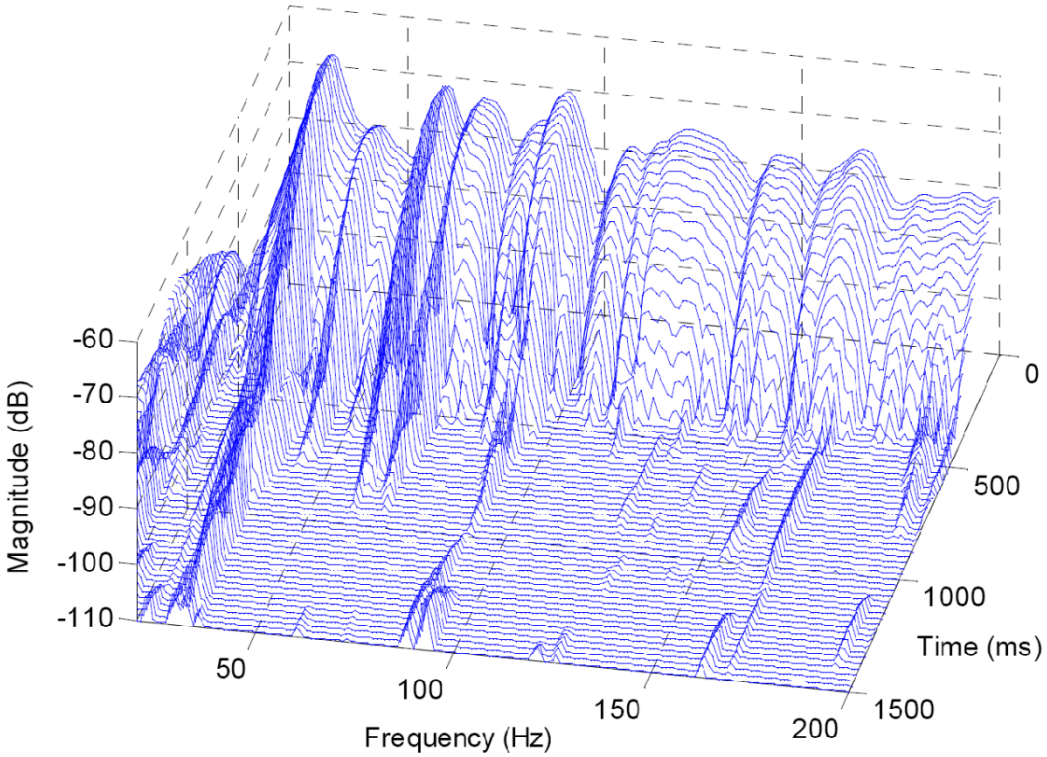


Figure 2: A waterfall plot example [Wilson et al 2003]

Toole [2006] claims that it is the behaviour of the room's boundaries at the locations of strong early reflections which dominates how a small room will sound; not the reverberation time. He also states (citing Genereux 1992, Craven 1992 and Rubak 2000) that at low frequencies where the room's effect is modal, the room responses are essentially minimum phase. What is heard can therefore be predicted by the

steady-state frequency-response measurements of the room. Welti and Devantier [2006] suggests that in his experience the axial modes dominate the low frequency performance of a room and maybe the first tangential mode has influence, but that the other tangential and oblique modes are rarely significant.

Gerzon [1990] commented on work conducted in the late 1950s by H D Harwood, which said that a delayed resonance, 40dB below the main loudspeaker response, severely coloured the reproduced sound. This work concluded that an amplitude variation of ±0.1dB and a variation in the phase response of 1˚ produced audible colouration. The human ear is therefore very sensitive to resonances. Standing waves within a listening environment are therefore undesirable and will be perceived as audible colouration.

2.3 Surround sound

There have been several speaker configurations aimed at delivering immersive audio to the listener, the original being quadraphonic, where four loudspeakers were arranged in a square at 90 degrees to each other. The most commercially successful solution has been the ITU BS 775 standard. This consists of five main loudspeakers and a subwoofer, commonly known as a 5.1 system. The speakers should be arranged as shown in Figure 3; there is a centre channel at 0˚, predominantly for dialogue in films. A left and right main channel at 30˚, and a pair of independent surround channels to add ambient effects between 100˚ and 120˚. The .1 channel refers to the subwoofer, or low frequency effects channel. It is used along with bass redirection

algorithms to reproduce the bass, allowing smaller, limited bandwidth loudspeakers to be used for the main channels. A 7.1 standard is becoming popular in home theatre equipment. It uses the same placement for the front three channels and subwoofer as in a 5.1 setup, but has two side and two rear channels for effects. The aim is to increase the listening area where surround sound can be experienced correctly. It has had limited market penetration though, due to the extra cost and cabling required. It is thought that a large number of people who have invested in a surround sound systems have not got them set-up correctly. This has led to a number of manufacturers adding automated set-up wizards to their surround sound processors.

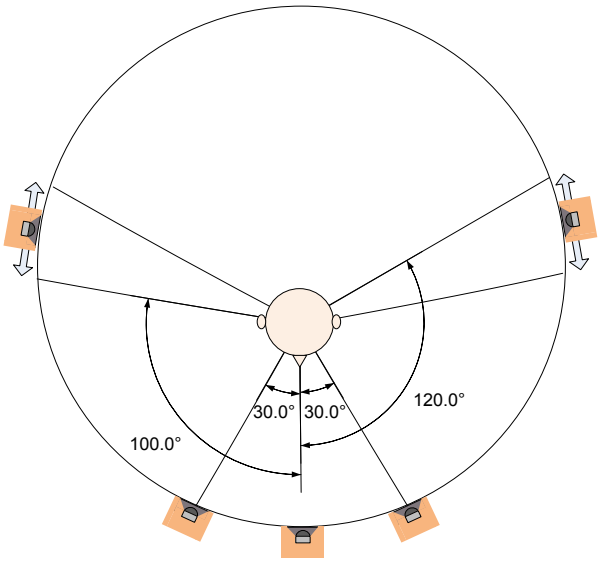


Figure 3: ITU 5.1 speaker configuration

Ambisonics is an alternative surround sound solution developed by a number of academics, including Michael Gerzon, Professor P B Fellgett and Duane Cooper in the 1970s. It was designed to overcome problems in Quadraphonic systems of a small

'sweet-spot' listening area and poor imaging. Ambisonics can be transmitted over one to four channels, depending on the bandwidth, encoding and decoding available, using either B-Format or UHJ, which require a decoder, or G-Format which can work with existing 5.1 decoders. It claims to offer a wider listening area and more freedom in loudspeaker placement than with current 5.1 standards by decomposing the soundfield into spherical harmonics. Height or periphony information can be derived if four channels and a decoder are present; this requires four additional loudspeakers, two on the floor and two above the listeners. Ambisonics has not, however, been a commercial success [Ambisonics 2009].

Wave field synthesis is a holophonic technique for reproducing a sound field. It works using an array of loudspeakers which are all driven independently. Using different time delays for different loudspeakers a wavefront can be synthesised; a superposition of signals allows a complete sound field to be created. The sound field can only truthfully be recreated if the loudspeaker array is a continuous line. In practice a finite array of loudspeakers is used. Any differences between the loudspeakers' responses will cause distortion to the wavefront and the result is colouration [Corteel 2006]. Equalization of each loudspeaker is therefore required to ensure that the errors are minimised. Such systems can potentially provide much greater localization of sound than a 5.1 or 7.1 home theatre system. However the large amount of equipment required for wave field synthesis currently makes it unsuitable for home use.

2.4 Modifying a room's acoustics

The approach that is taken to modifying the acoustics of a large room is quite different to that used in small rooms. Concert halls for example often use architectural structures and modifications to the room to improve the acoustics; such techniques are not suitable for domestic rooms. For example a study by Fuchs et al [2000] suggests that to absorb low frequencies absorbers need to cover 20% of the room's surfaces with panels measuring 1m by 1.5m by 0.1m. An alternative architectural solution is to construct a room where the dimensions spread the modal frequencies as evenly as possible. There have been many attempts at finding the optimal room dimensions using the equation of modal frequencies; Bolt [1946] suggests several ratios, one being where the width is 1.26 times the height and the length was 1.59 times the height. This is claimed to evenly spread the modal responses. Indeed Toole [2006] states that although designing a room to have the perfect dimension ratios will mathematically produce the optimum modal distribution, it is only valid if a single speaker and listener are in opposite three boundary corners. As soon as the listener moves, or the speaker is moved, or another speaker is introduced, the benefit of optimal dimension ratios is lost.

Cox et al [2004] have developed a more comprehensive model, which includes the absorption of the surfaces, to predict optimal room dimensions, loudspeaker positions and listener positions. Whilst these techniques are technically possible, large absorbing panels and custom built rooms with optimised dimensions and set listening positions are not feasible or desirable for most consumers.

A large amount of research has therefore been undertaken to find solutions that could be implemented electronically. Most of this research can be found in the Audio Engineering Society literature, both in their journals and conference proceedings. The design goals of these projects are varied. Some have attempted to improve the loudspeaker in isolation, whilst others have taken the loudspeaker and room as one, and tried to solve the deficiencies of both together. Another approach is to correct for the room and leave the loudspeaker's response largely untouched.

The first device widely used to adjust the sound electronically to better match a room's acoustics was the graphic equalizer. A 31 band graphic equalizer divides the frequency range from 20Hz to 20kHz into 1/3 octave bands. Such devices have been used for many years in live public address (PA) systems. They are used to compensate for the frequency response of the loudspeakers, and to increase the gain before feedback level of the system. Feilder [2003] states that the advantage of this type of equalization is that the filter's impulse response is short and so it is less likely to create time domain problems than much more complex dereverberation filters. However its simplicity does not permit all undesirable room characteristics to be removed. Toole [2006] believes that the minimum phase response of a room at low frequencies can be corrected for in the time and frequency domain with minimum-phase parametric filters.

A number of systems have been developed to enhance the reverberation of a room. These are aimed at halls where the acoustics are 'too dry' or in order to alter the early to late energy ratios. Examples are presented by Lissek and Meynail [2003],

Berkhout et al [1988] and Meynail and Vain [1998]. This thesis is concerned with the domestic environment where recorded music is played back. Information about desired goals of acoustic manipulation can still be drawn from such systems; however their solutions are not relevant to the focus of this research.

Since 1982, when the Compact Disc (CD) player was conceived an increasing amount of digital signal processing has been employed in consumer audio products with the aim of adding extra features or improving the quality of the audio [Greenfield and Hawksford 1991]. The signal processing duties have varied from decompressing stereo and surround sound formats to equalization of a loudspeaker or room through digital filtering.

Digital equalization filters are either Finite Impulse Response (FIR) or Infinite Impulse Response (IIR). The type of filter used depends on the developer and their design goals. The advantages of FIR filters are that they are always stable, are easy to design and can compensate for the phase of the system to produce a linear phase result. The disadvantage of linear phase filters is that they can produce pre-echo in the impulse response, so the sound can be heard before it should be. Several subjective listening tests have shown that the ear is sensitive to this pre-echo, so some developers choose to use IIR filters instead. Another disadvantage of FIR filters is that to equalize low frequencies with high resolution, a high order filter is required, which is costly computationally and its delay could be too great. This delay and the filter order can be reduced, however, by using a warped filter; here the filter is designed to have a nonlinear frequency resolution to increase the resolution of the

filter at low frequencies. Karjalainen et al [1999] demonstrated that warped filters could be designed to reduce the filter orders by a factor of five, but at the cost of increased computational complexity of 2.5 to 3 times. Downsampling and multirate filtering have also been used to reduce the delay of an FIR filter.

IIR filters generally need much lower orders than FIR filters, but they are not guaranteed to be stable and are more complex to design. Another consideration is that IIR filters are prone to noise problems, generated by quantatization effects, particularly at low frequencies. This effect can be minimised, however, as shown by Zolzer [1997], by using noise shaping and different filter topologies.

Considering the loudspeaker in isolation, it is known that equalizing the loudspeaker's axial response only can increase off axis distortion. A solution proposed by Greenfield and Hawksford [1991] was to average a weighted set of impulse responses measured in different places over the listening area, employing a two-stage equalization process. Firstly this used an IIR filter to equalize the magnitude response and then an optional second stage where an all-pass FIR filter equalized the phase response. Wilson [1991] proposed a least mean square loudspeaker equalization approach where the energy error of both on and off axis responses was minimised. Ser et al [2003] meanwhile provide another approach where binaural perception is considered. They conclude that it is nearly impossible to correctly equalize a loudspeaker without knowing where the listener is. Ramos and Lopez [2006] propose a scalable solution based on parametric IIR filters, with an automated design process

which used psychoacoustics to determine which parts of the response need to be corrected, thus simplifying the end implementation.

Several attempts have been made to completely undo the effects of the loudspeaker and room using inverse filtering. This involves recording the impulse response of the system and calculating an inverse filter which will remove all reflections and produce a Dirac function, i.e. a single impulse. This is like listening in a perfect anechoic chamber. To recreate the original recording space a further filter can then introduce reverberation back, either as a simulation, or recording of an optimised listening room. Inverse filters are implemented in a digital signal processor (DSP) as a digital FIR filter, allowing arbitrary amplitude and phase manipulation [Fielder 2003].

There are, however, several problems with this approach; firstly the performance of the dereverberation. Large gains are required to correct for dips in the response of the system. These could overdrive the system, increasing the nonlinear distortion produced by the speaker, particularly in the low frequencies where diaphragm displacement is greatest [Greenfield and Hawksford 1991]. To limit the gain a technique called regularization can be employed, though this introduces more problems because the regularization converts a minimum-phase filter into a non-minimum-phase filter, which is acausal, introducing pre-ringing in the time domain [Fielder 2003]. Fielder used a filter length of 10.9 second, significantly longer than might be practically achievable in consumer electronics, and three regularization levels, -80, -40, -20dB. His findings were that even this filter length might be too short, depending on the regularization setting. Regularization of -80 and -40dB both

caused extremely audible errors at ±5.46 second, the filter's boundaries. A regularization figure of -20dB can avoid these errors, but add errors in the time interval within 0.5 second of the direct sound, both pre and post direct sound which would be continuously audible.

More critically, correction is only valid at the location where the original measurement of the system's impulse response was taken. Fielder [2003] goes on to show that a displacement of 1.3cm is enough to severely corrupt the dereverberation. With this very small displacement the transfer function mismatch is significant enough to corrupt the dereverberation process above 500Hz. He points out that even if the receiver does not move, variations in room temperature would be enough to corrupt the results in a similar way. He concludes that dereverberation is therefore impractical in real world situations.

Hatziantoniou and Mourjopoulos [2003] recognise the problems with inverse filtering and propose pre-processing the impulse response with a complex smoothing operation. The inverse filter does not attempt to compensate for deep spectral notches, but still produces measurable and audible improvements. Norcross et al [2004] confirms that there are audible problems with inverse filtering, but that complex smoothing the impulse response can produce an overall improvement in subjective tests.

Stefanakis et al [2008] believes that the high frequency response of a room can be changed using passive absorbers, but it is the low frequencies where the room modes

dominate where active equalization is required. He refers to work undertaken by Mourjopoulos in 1985 who found that the radius of the equalization zone must be less than half a wavelength of the frequency being equalized so that the amplitude variations are within ±3dB. Stefanakis also referred to Santillan's 1997 research which found that compensating for peaks and dips at one position did not improve the magnitude response at other locations, but frequently made them worse.

Commercial products, based on the ideas presented above have been developed and are on sale by the following companies: Audyssey, Bang and Olufsen, Lyngdorf, Meridian and Trinnov. The techniques chosen for each solution will be reviewed in the following section.

Audyssey [2008] have developed a multi-listener room equalization scheme which has been widely adopted by a number of consumer electronics manufacturers. In Audyssey's room equalization scheme, a number of impulse response measurements are taken at different positions within the room. These responses are clustered, or grouped, according to their similarity, which can be as simple as the distance between the measurements [Bharitkar and Kyriakakis 2001]. The specific parameters used in Audyssey's commercial products are unknown and the information presented here is taken from published research by Bharitkar and Kyriakakis undertaken prior to them starting the company, Audyssey. The clustering uses a fuzzy c-means clustering algorithm. This means that a measurement may belong to more than one cluster by different degrees. A single equalization filter is then designed for each speaker using a non uniform weighted average of the clusters' responses, which is inverted with a

non linear, warped, frequency axis, to produce an FIR equalization filter. This is implemented with multirate filtering techniques to ease the computational load required to equalize at low frequencies [Bharitkar and Kyriakakis 2002; 2004].

Lyngdorf's solution meanwhile attempts to correct for the room, whilst leaving the loudspeaker's response untouched, thus keeping the timbre of the loudspeaker which the customer had chosen. Their design process involves measuring the loudspeaker-room transfer function at the primary listener's position, and a number of random room positions, with the room positions being averaged to give a global power response. The number of room positions is dependent on the room. A minimum of three positions are needed; the number of measurements taken is increased until a new room position measurement has little effect on the global response. These room positions are used to estimate the loudspeaker's low frequency slope and limit, the speaker's sensitivity, directivity index, and the high frequency roll off characteristics. These parameters are used to automatically specify the target frequency response of the system. The equalization filter is implemented with a two stage multirate FIR filter, calculated from the listener's position measurement and the target frequency response, upper and lower gain limited are calculated from the global and target responses to avoid non linear distortions or overloading the amplifier or loudspeakers [Pedersen 2006; Pedersen and Thomsen 2007].

A different approach, by Meridian Audio involves identifying the frequency of the room modes and decay time from a single spot measurement. They believe that the results would not vary significantly at different locations, thus the solution is spatially

robust. The modal frequencies are identified automatically from the loudspeaker-room impulse response. The RT60 decay time of the modes are compared to the RT60 decay time in the 500-2000Hz band. A notch filter is then designed which will reduce the mode's decay time to that of the 500-2000Hz band, for example a notch of - 6dB would halve the decay time [Wilson et al 2003]. A very similar technique was proposed by Makivirta et al [2001]. They identified the modes by their decay time and equalized the signal using either a notch filter or a secondary radiator to interact with the sound field produced by the primary source. An updated solution presented by the same research group in 2003 was to use a high order FIR filter with an exponential decay function at the modal frequencies to reduce their decay time to that of the mid frequencies [Karjalainen et al 2003].

Trinnov Audio have developed an innovative solution to overcome the practical placement of loudspeakers within a home environment. They first measure the loudspeakers' three dimensional positions, using a small three microphone array at the listener's location and then use acoustic field techniques to remap the sound from the music recording's speaker layout to the actual speaker layout in the listener's room. The optimizer then corrects for the frequency and phase response of a loudspeaker-room transfer function using an FIR filter along with the loudspeaker placement [Laborie, et al 2005; Trinnov 2008].

A quite different approach is taken by Bang and Olufsen, as presented by Pedersen [2003]. Rather than placing a microphone at the listening position and creating a correction filter based on an impulse response, they place the microphone at two

different positions just in front of the bass driver's diaphragm. These measurements are used to calculate the radiating resistance of the speaker. This measurement is performed in the reference listening room, at the reference location, and then in the customer's room at their desired location. A 16^{th} order IIR filter is designed to modify the bass signal so that the bass output power in the customer's room becomes the same as it was in the reference room. It is claimed that this method preserves the timbre of the loudspeaker, as the designer intended, regardless of the room and speaker position. However this seems to disagree with the majority of research on room modes which agrees that the effect of the room modes should be minimised so that the delayed resonance does not introduce colouration. In the Bang and Olufsen system any room modes which affect the sound in the reference room will be introduced in the customer's room. A better approach would possibly be to employ a theoretical radiation resistance in free field as the target response.

It is highlighted by Welti and Devantier [2006] that two subwoofers on opposite walls driven with the same signals can cancel out odd order modes, through destructive interference, leaving the even order modes untouched. In Welti's previous paper [2002], analysis of how many subwoofers were needed to optimise the sound at more than one position concluded that four subwoofers were optimal, with the subwoofers placed at the wall's mid points. Two units on opposite walls also performed well. The performance criterion was measured as seat-to-seat variation, rather than flat magnitude response, assuming that the overall response could then be equalized afterwards once a large listening area was achieved. In the later, 2006, paper a sound field management algorithm was proposed. This consisted of a search algorithm

which chose the subwoofer configuration with the least seat-to-seat variation. This time, along with the number of subwoofers and their locations, each subwoofer's gain, delay and the parameters of a simple bandstop filter per subwoofer were adjusted. The signal processing requirements of the implementation are, therefore, simpler than complex FIR filter solutions. The results showed that these simple signal manipulations reduced the spatial variation compared to optimising the number of subwoofers and their locations alone.

An alternative system, called CABS (Controlled Acoustically Bass System) has been developed by Neilsen and Celestinos [2007]. Rather than apply equalization filters to the existing speakers they propose adding four subwoofers into the room. These are placed in specific positions, two mounted on the front wall at half the room's height and a quarter of the width in from either side wall. These are used to reproduce low frequencies, below 100Hz, as a plane wave. Another pair of subwoofers are placed on the rear wall, in the same positions as those on the front wall. These rear subwoofers are driven with a delayed, phase inverted version of the signals sent to the front subwoofers. The result is that the rear subwoofers cancel out any axial standing waves, thus creating a much smoother frequency response within the room, with minimal spatial variation and a much reduced decay time which varies less with frequency. The drawback of such a system is the added subwoofers and their rigid placement requirements.

2.5 Room for improvement?

The drawback of all the systems evaluated above is that they are static. They rely on one or more microphone measurements being taken within the room, usually at the listener's possible locations and some other positions. The recordings are performed with test signals transmitted from each loudspeaker in turn. Using the developer's chosen design method a set of filters is then developed which are used to manipulate the signals in some way before they are amplified and drive the loudspeakers. If the room changes in any way, or the user changes their listening position the system needs to be recalibrated with new microphone recordings to give optimal correction and therefore listening experience. Over time the acoustics of the room will change, for example the furniture or loudspeaker might get moved, or a new item might be added to the room. Whilst some customers will recalibrate their playback system after such events, not all will. There is also variation introduced on a daily basis. For instance friends or family may visit so there are more people in the room, changing the sound's reflective path and the absorption of the room. The doors and windows might be opened or closed and be in different states than when the calibration was performed, and temperature variations will change the speed at which sound travels, which will change the response of the room.

This author believes that an adaptive system which alters with the changing acoustics of the room is, therefore, desirable. Such a system could use one, or a combination of the solutions reviewed above to design the correction filters, but be recalibrated on a continuous basis, based on current measurements. It would be technically possible to

do this with several microphones placed permanently in the room at the expected listening positions or in free space; however, most people would not accept this as a solution in their living room. It was from these considerations that this project was conceived. The aim of this research was to investigate whether it was possible to monitor the room's acoustics in a non-intrusive way and predict the sound at multiple different positions within the room, thus providing a powerful measurement tool for multi-listener room correction.

To provide focussed correction at the listeners' ears it is required to know the listeners' locations within the room. There are a variety of techniques, including video cameras, and acoustic measurements which have already been developed for person location within a room [Krumm et al 2000; Zhang et al 2006]. This project therefore concentrates on the impulse response prediction, assuming the listeners' locations were already known.

A literature survey has not found a solution which is able to predict the sound at remote locations within a reverberant enclosure. This is therefore the problem under investigation in the remaining sections of the thesis.

2.6 Chapter conclusions

This chapter has presented a brief review of some important acoustic developments and demonstrated that the room in which we listen to music has a large effect on what we hear. Techniques and solutions to modifying a room's acoustics have been

presented. Their application to domestic living rooms has been reviewed and their limitations identified. A solution has been suggested that could provide improvements, which will be the subject of the remaining chapters.

According to Beranek [1996], “the experience of music can never be divorced from the acoustics of the space in which it is performed”. Beranek goes on to say that, with regard to concert halls, “a superior space enhances the music in some way, without detracting from it in others” [Beranek 1996]. The aim of this research is to develop such a superior space in the context of the domestic environment through adaptive equalization.

Chapter 3 Possible Solutions and System Topology

Chapter 2 reviewed existing solutions and showed how the majority of room equalization systems calculate a correction filter as a result of microphone measurements at the listeners' positions and possibly other locations within the room. Currently there are no solutions which are able to predict the sound in the room using measured data, however, there are several possible solutions. Two methods were considered during this project, firstly placing microphones on the loudspeakers, and secondly mounting microphones on the room's walls.

This chapter compares these methods and presents the system topology proposed. The decisions taken which directed the development of the system topology were largely based on an investigation of array processing. While Chapter 6 provides an in depth analysis of this topic, a summary of the relevant points needed here follows: In a reflection free environment two sensors are sufficient to estimate the arrival angle of a single wavefront, using time difference of arrival and trigonometry [Carr 2001]. Several sensors are needed to estimate the angle of arrival of multiple signals arriving at an array of sensors at the same time. The number of sensors required and their placement depends on the angular resolution required, the algorithm used and the number of signals impinging on the array at the same time. A higher ratio of sensors to signals will give greater accuracy. The minimum number of sensors required to resolve the wavefronts is the number of wavefronts arriving at the array, during the sample duration, plus one [Van Trees 2002]. The accuracy of angular estimations is greatest when the direction of arrival of the wavefront is perpendicular to the array,

with the least accuracy when a wavefront arrives from a direction parallel to the array, known as the endifre condition [Johnson and Dudgeon 1993].

The feasibility of both methods was analysed considering the accuracy of a predicted impulse response and the viability of a final system. One of the research goals outlined in Chapter 1 was that any potential system should be practically viable in a real living room. This guided the selection process to a solution which could work with a potential user's stereo or home cinema system and have minimal installation requirements and impact on the user's living room's aesthetics. Four case studies of different room and loudspeaker layouts, shown in Figure 4, are used in this chapter to explain and justify the chosen direction.

Case A is a stereo setup, whilst case C is a five channel surround setup, both cases with a single listener and just loudspeakers in a rectangular room. However, since it would be very unusual for a potential user to have a lounge without furniture, cases B and D are examples of a more typical consumer's room layouts. Here it can be seen that the room is an 'L' shape with furnishings and loudspeaker positions which suit the room's layout, but are not acoustically ideal. These room layouts are considered to be a good compromise between acoustics and aesthetics, though many consumers will have layouts which deviate much further from the acoustic ideal.

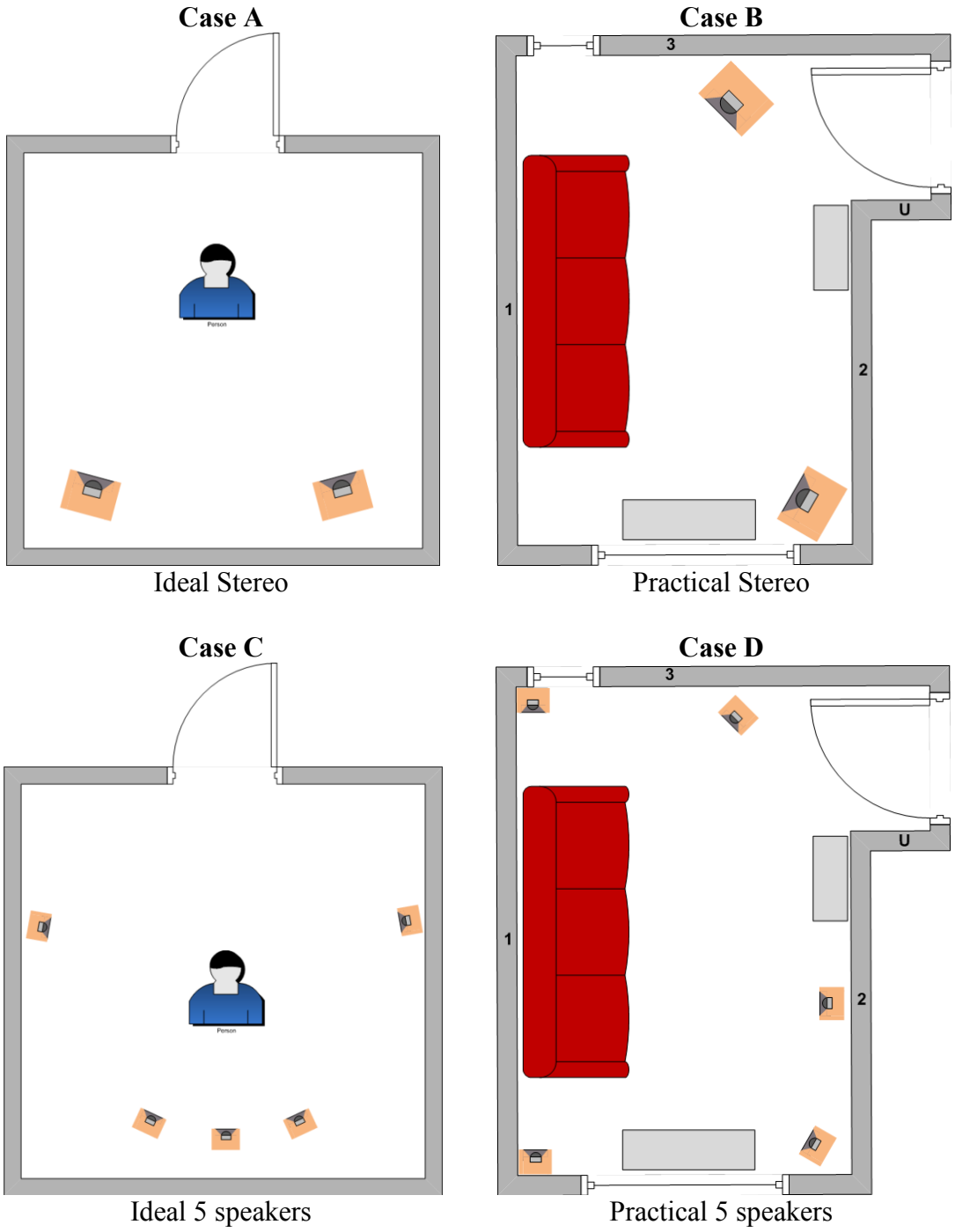


Figure 4: Solution case studies.

3.1 Why can a room simulator not be used to predict the sound?

Software packages exist which are able to calculate an impulse response at a location within a room using an acoustic simulation of the room. These are not designed for a typical customer of a stereo or home cinema system but are mainly aimed at academics and architects. Possible room simulation methods are described in Chapter 4. Briefly, simulators require a user to enter, at minimum, the room's dimensions, the surface absorption characteristics as well as the speaker and receiver positions. Most simulators do not consider furniture due to the complexity of modelling the acoustic impact of a particular object's shape and absorption. They require powerful processing to produce an accurate impulse response at one location. Without modelling the details of the room and the objects within, the impulse response calculated will be limited and may not match a real scenario, such as the room in case D. Another limitation of a simulator is that it would not provide an adaptive solution; a user updating a computer model periodically would be more labour intensive than the current solution of the user periodically re-measuring the room's response with a microphone at the listener's location.

3.2 Room parameter identification

The impulse response of a room contains all the acoustic information about the room, at that location. It will consist of the direct sound and reflections of the direct sound by the room's boundaries and objects within the room. Chapter 4 has details of

techniques for modelling reflections. In summary, when a sound wave hits a hard surface it reflects; this can be modelled as an image of the original source on the opposite side of the boundary, attenuated by the acoustic impedance of the boundary (see Figure 5) [Allen and Berkley 1979]. Thus if all the reflections within the room are identified and their parameters known, the impulse response can be calculated at different locations within the room.

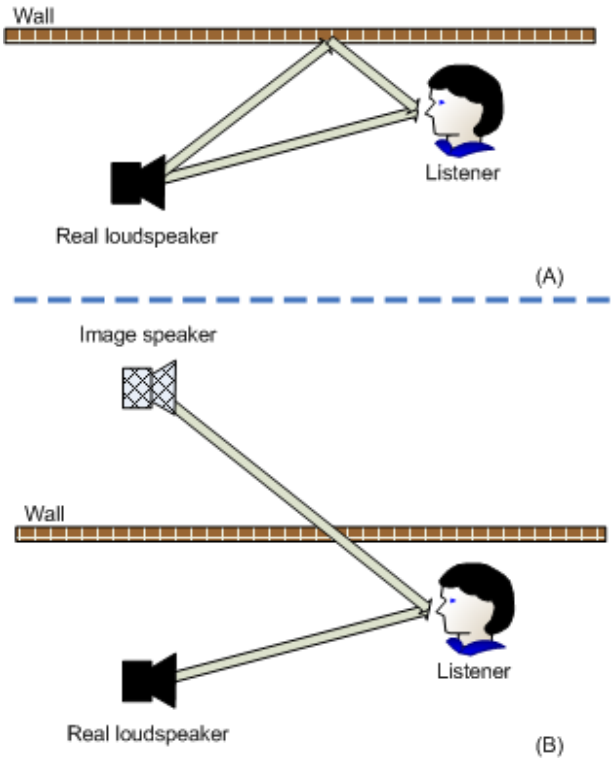


Figure 5: A) Reflected sound. B) Reflection modelled as an image source.

Chapter 2 showed that it was most worthwhile to improve the low frequency response of a room with electronic filtering to reduce the effect of standing waves. Therefore a measurement system need only be accurate at low frequencies if only low frequencies are to be manipulated by the filter. Because of the cyclic nature of standing waves it

is vital that all 1st order reflections are identified. A first order reflection from a far away object will arrive much later in time than higher order reflections from close objects, therefore part of the room's late, decaying reverberation will need to be analysed as well as the early reflection part. This is a significant challenge. The high number of signals present in reverberation means that many of the published methods of signal analysis were found to be unsuitable.

The distance from the receiver to the source, image or original, can be estimated by the difference in transmitted and received time and the speed of sound. This method assumes that the transmit and record equipment are synchronised and the latency of the test system is known.

Two methods, multidimensional scaling and direction of arrival techniques were considered as candidates for identifying the room's parameters. Multidimensional scaling is a statistical method which identifies similarities in data by comparing the Euclidean distance between sets of data [Young 2008]. In this application it could be used to identify parts of the impulse responses measured at different locations in the room, which are similar [Beard and Atkins 1998]. In the case of multiple loudspeakers with microphones mounted on them the relative positions of the loudspeakers to each other is easy to estimate, and repeated patterns of the loudspeaker layout caused by wall reflections might be identifiable in space, from the impulse responses, thus providing information about the room's boundary parameters.

A direction of arrival method requires several microphones. It could identify the azimuth and elevation angle of an image source relative to its location. The distance could be calculated from the arrival time of the reflection, providing a means of calculating the three-dimensional coordinates of the image.

Along with position data, the frequency response of the reflections will also need to be estimated. If successful, both methods would allow an acoustic picture of the room to be generated, from the measured data, which could be used to predict the impulse response at different locations within the room.

3.2.1 Microphones on speakers

If a single microphone was placed on each of the loudspeakers in the system and test signals were transmitted from each loudspeaker in turn it would be possible to calculate the distance from each speaker to every other speaker using the time delay, as described above. To estimate the exact position of a loudspeaker relative to the other loudspeakers, in three dimensions, a minimum of three microphones per loudspeaker would be required, arranged in a triangle with known positions. The time difference of the direct sound arriving at different microphones together with the known displacement between them could be used to calculate the angle of arrival of the direct sound. Thus the transmitting loudspeaker's position relative to the receiving loudspeaker could be estimated.

Multidimensional smoothing techniques could possibly identify an early reflection of the transmitting loudspeaker by identifying the repeated pattern later in the impulse

responses. Consider case A which only has two loudspeakers and a listener present, with the microphones only mounted on the front baffle of the loudspeakers. It might be possible, using a microphone placed on one of the loudspeakers to identify the reflections caused by the side and rear walls, along with floor and ceiling reflections. But a microphone mounted on the front baffle of a loudspeaker will not have an omnidirectional response, so identifying the wall behind the loudspeaker will be difficult or impossible. To overcome this microphones would need to be mounted on multiple sides of the loudspeaker, and combined to sample all the reflections.

This technique has limited information when only two loudspeakers are used because there would only be measurements at two locations. Results are therefore expected to be poor, or impossible for case A. In case C where a 5-channel speaker system is used the results are expected to be better. There are more points in space being sampled and the rear loudspeakers can aid identifying the reflections from the wall behind the main three speakers, meaning the microphones might only need to be mounted on the front baffle of each loudspeaker.

In cases B and D there is furniture present, the room is an 'L' shape and the loudspeakers are in non ideal positions. The latter will limit the sound field presentation, but will not limit the identification of each loudspeaker's position using time difference of arrival methods. The furniture in the room, however, will change the reflections and make identifying similarities between recordings much harder, which is predicted to considerably compromise the accuracy of the estimation of the room's parameters.

Another limitation of mounting the microphones on the loudspeaker is that it might not be possible to integrate with a customer's existing loudspeakers. It is more suitable for new loudspeaker designs where the microphones and required electronics can be incorporated into the loudspeaker cabinets.

3.2.2 Microphones on room's boundaries

The second solution considered was to place microphones on the walls of the room. As discussed in section 3.2.1, a microphone placed on a baffle will not have an omnidirectional response, thus placing a microphone on a wall of a room will prevent it from sampling sound from behind the wall it is mounted on. A number of microphones, forming an array, will be required so the direction of arrival of the sound can be estimated. The angles, along with the time delay of each image source could be used to calculate the three-dimensional coordinates of the image sources.

Two configurations of microphone arrays mounted on the wall were considered: either many arrays consisting of a minimum of three microphones each placed at intervals around the room, or fewer arrays consisting of more microphones. Both configurations would allow the sound field to be sampled. The results from many smaller arrays could be combined to reduce the need to analyse as much of the reverberation field, thus limiting the number of microphones needed per array. After consideration the idea of using many small arrays was rejected, however, due to the installation issues involved. It was decided that many arrays would require a large amount of cabling for signals and power and that they would disturb a living room's decor significantly so would not be a viable option to most consumers and therefore

would not fullfil one of the goals of this research. The smallest number of arrays possible was therefore opted for.

Cases A and C show a rectangular room: in this instance a minimum of two microphone arrays, mounted on different walls would be required to 'see' all the reflection origins in the room. The choice of which two walls the arrays would be mounted on is not constrained, though it would be advisable to place them on the longest walls; placing the array in free space, rather than in a corner would improve the angular resolution of the results. (Chapter 6 covers array processing where the reasoning for these decisions is presented in more detail.) For the room layouts shown in cases B and D, an array placed on the wall behind the sofa, in the position marked 1, would be able to sample reflections coming from all walls except the unseen wall marked 'U' and the wall it is mounted on (see Figure 4). A second array could be placed on the wall opposite the sofa, in the position marked 2 in Figure 4, which would be able to record the reflections behind the first array. Alternatively the second array could be placed on the top wall, in the position marked 3, which would be able to record reflections from behind the first array, and also reflections caused by the unseen wall, marked 'U', meaning that the entire room was covered. A more complex shaped room might require more or different array placements to ensure that all reflection origins could be sampled. If each microphone array included a single transmitter, then after installation, a test procedure could be used to estimate the positions of the arrays relative to each other. This is beneficial when combining the position estimates of multiple arrays.

An advantage of this solution is that it could lead to aftermarket products which would work with the customer's existing loudspeakers in both a stereo and multichannel layout, as well as with new loudspeaker designs.

3.3 Chosen solution

Based on the considerations presented in sections 3.2.1 and 3.2.2 it was decided that a minimum of two arrays, mounted on the walls of the room, was the strongest solution. It was believed to be the most versatile option with regard to room shapes and could be installed in a typical consumer's living room with minimal disruption, working equally well with a stereo or multichannel loudspeaker system.

3.3.1 Spatial processing

To limit the impact of the arrays on a room's decor it was decided to make the arrays as physically small as possible. The number of microphones and the geometry of the array are discussed in detail in Chapter 6, along with an explanation of the Incremental Multi-Parameter, (IMP) algorithm selected to estimate the angles [Clarke 1989]. The narrowband direction of arrival algorithm selected uses the phase differences at the microphones to estimate the angles of arrival. To maximise the phase difference and prevent spatial aliasing the signal should have a wavelength just less than twice the distance between the sensors. Because the wavelength at 100Hz is approximately 3.4m it would not be possible to have the microphones spaced to allow the direction of arrival estimate to take place at low frequencies within domestic

rooms. Therefore for a small array a high frequency signal is required for the angular estimations.

A large smooth surface will reflect narrowband high frequency sound the same as it will low frequency sound, but with a different attenuation. Therefore an estimate of the reflection's origin can be performed at high frequencies. If, however, the surface is not smooth but has some small variations, the reflection will be modified if the wavelength is small in comparison with the surface variations. In comparison, a surface which looks rough or uneven might well behave like a smooth surface at low frequencies. When the wavelength is small in comparison with the surface variations scattering will take place, resulting in multiple reflection origins close together [Vorlander 1989]. This is not a problem in the proposed system because the spatial processing will identify multiple targets close together. When the images' frequency response is being estimated the response of the room is measured with a low frequency signal and it is matched against a prediction of what the sound is estimated to be at the array, using the spatial estimates. Thus multiple images can be consolidated, whilst images which do not match can be removed.

It might be possible in a complete system to select relevant sections from the music being played for analysis, but for development a high frequency test signal was used for the identification of the reflections' location.

3.3.2 Reflection separation

To calculate the coordinates of an image its distance must be known, calculated from the delay between transmission and arrival of the individual reflection. Also, as stated, for a direction of arrival algorithm to be able to estimate the arrival angles of multiple wavefronts, the number of signals at the array must be less than the number of sensors. To fulfil both of these conditions the received signal must first be deconvolved into separate reflections, the matching pursuit algorithm being used to achieve this. Time sections can then be selected for direction of arrival processing. To reduce overlap of reflections a short burst of signal is required. This work is detailed in Chapter 5.

3.3.3 Reflection's frequency response

Once the reflection's coordinates are estimated the frequency response of the reflection needs to be predicted. This prediction needs to cover the frequency range over which the predicted impulse responses are required to be accurate. The same high frequency signal used for the spatial processing therefore is not suitable, therefore a low frequency logarithmic chirp was transmitted and recorded at the centre microphone of the array. A least mean squared maximum likelihood algorithm was then used to match an individual filter response to each image, which minimised the difference between the recording at the array and a prediction generated from the spatial estimates. This work is detailed further in Chapter 7.

The results from the arrays were combined. Using an estimate of the location of each reflection and their frequency response it was possible to predict the sound at different locations within the room.

3.4 Chapter conclusions

This chapter has investigated different solutions that have the potential to predict the sound at a remote location within a room. Two solutions were proposed, firstly microphones mounted on the loudspeakers and secondly microphones mounted on the room's walls. The solution chosen for investigation in this research was to place a minimum of two small arrays on the room's walls because this allows the location and frequency response of each reflection to be estimated using test signals and thus allows a model of the room to be generated from measured data. This model can be used to predict the sound at remote locations within the room.

Chapter 4 Room Modelling

To aid the development of the system a computer simulation of a room was developed. This was done for two main reasons, firstly to allow the results from real physical rooms to be compared to theoretical results, aiding identification of measurement errors and real world effects. Secondly, it allowed faster development by providing a test-bed for different configurations; the room geometry, source and receiver positions needed to be easily adjustable so that the algorithms developed could be validated for a range of setups. Comparing the correct image locations and the estimated locations, using simulated impulse responses and different spatial processing, a more effective solution was developed.

As stated in the introduction, to aid development of the system an empty shoebox, rectangular room was used for this research. There are four methods of simulating a room which were considered: Modal Calculations, Ray Tracing, Image Source Method (ISM) and Finite/Boundary Element analysis. These methods will now be explained, along with an analysis of their suitability for this application.

4.1 Modal frequency modelling

If a sound in a room reflects with equal probability off all surfaces then the sound energy will decay exponentially, due to the intensity decreasing according to the inverse square law and surface absorption. However at some frequencies the sound

energy will not reflect randomly, it will reflect in a cyclic pattern. This occurs when the path length of the sound equals an integer number of half wavelength. A standing wave is formed, having pressure and velocity distributions which are spatially static. Standing waves give large variation in sound pressure at different locations, occurring only at discrete frequencies which are determined by the room's geometry. An alternative name for a room's standing waves is room modes. There are three types of room mode, axial modes which reflect off two opposite surfaces, tangential modes which occur between four walls, and oblique modes which involve six walls. The frequencies of the modes can be determined using the formula below, if the room is empty with a cuboid shape.

$$f_{xyz} = \frac{c}{2}\sqrt{\left(\frac{x}{L}\right)^2 + \left(\frac{y}{W}\right)^2 + \left(\frac{z}{H}\right)^2}$$

where x, y, z are the number of half wavelengths between the surfaces and L, W, H are the Length, Width and Height, of the room, in metres and c is the speed of sound. The formula is valid only for integer numbers of half wavelengths x, y, z. The number of standing waves increases proportionally to the square of the frequency; this technique is therefore most suited to the modelling of small rooms at low frequencies [Howard and Angus 2001; Everest 2002]. Table 2 shows the modal frequencies for the room presented later in this thesis, calculated using the above formula.

Mode Number	x	y	z	Modal frequency, Hz
1	1.0	0.0	0.0	20.1
2	0.0	1.0	0.0	48.8
3	1.0	1.0	0.0	52.8
4	0.0	0.0	1.0	63.8
5	1.0	0.0	1.0	66.9
6	0.0	1.0	1.0	80.3

7	2.0	0.0	0.0	40.2
8	1.0	1.0	1.0	82.8
9	0.0	2.0	0.0	97.6
10	2.0	1.0	0.0	63.2
11	1.0	2.0	0.0	99.7
12	2.0	0.0	1.0	75.4
13	0.0	2.0	1.0	116.6
14	2.0	1.0	1.0	89.8
15	1.0	2.0	1.0	118.3
16	2.0	2.0	0.0	105.6
17	3.0	0.0	0.0	60.3
18	0.0	0.0	2.0	127.6
19	3.0	1.0	0.0	77.6
20	0.0	3.0	0.0	146.4
21	1.0	0.0	2.0	129.2
22	0.0	1.0	2.0	136.6
23	2.0	2.0	1.0	123.3
24	3.0	0.0	1.0	87.8
25	1.0	3.0	0.0	147.8
26	1.0	1.0	2.0	138.1
27	3.0	1.0	1.0	100.4
28	0.0	3.0	1.0	159.7
29	3.0	2.0	0.0	114.7
30	2.0	0.0	2.0	133.8
31	1.0	3.0	1.0	161.0
32	0.0	2.0	2.0	160.7
33	2.0	3.0	0.0	151.8
34	2.0	1.0	2.0	142.4
35	1.0	2.0	2.0	161.9
36	3.0	2.0	1.0	131.3
37	2.0	3.0	1.0	164.7
38	2.0	2.0	2.0	165.6
39	3.0	0.0	2.0	141.1
40	3.0	3.0	0.0	158.3
41	3.0	1.0	2.0	149.3
42	0.0	3.0	2.0	194.2
43	1.0	3.0	2.0	195.2
44	3.0	3.0	1.0	170.7
45	0.0	0.0	3.0	191.4
46	1.0	0.0	3.0	192.5
47	0.0	1.0	3.0	197.5
48	3.0	2.0	2.0	171.6
49	1.0	1.0	3.0	198.6
50	2.0	3.0	2.0	198.3
51	2.0	0.0	3.0	195.6

Table 1: Modal frequencies calculated up to 200Hz for the room used in this research project.

Length = 8.57 Width = 3.53 Height = 2.70m.

4.2 Ray tracing

Ray tracing is a geometrical approach which attempts to model the acoustics of a room numerically. The sound source is represented by a large number of rays, with energy distributed according to the source's directionality. These rays spread out over time, in a spherical pattern from the source. When a ray hits a hard surface it is reflected with the angle of incidence equalling the angle of reflection, as in optics. The reflected sound is attenuated to account for the sound absorption of the surface. The ray is removed from the simulation once its energy has fallen below a set threshold, or its path length has exceeded a specified maximum [Gerganova and Christov 1990; MacLaverty and Furlong 1992]. An estimate of the sound within the room is calculated by representing the microphone as a sphere. Every ray passing through the sphere has its energy registered, building a histogram of the energy over time [Gerganova and Christov 1990]. The accuracy of the result is determined by the angular separation between the transmitted rays, and thus the number of rays used to model the source.

The advantage of ray tracing is that it is computationally efficient. It can include specular and scattered reflections caused by rough surfaces and can easily model irregularly shaped rooms. The problem with the algorithm is that it cannot guarantee accuracy. The discrete angles at which the rays are transmitted means that some valid paths to the listener might be missed, meaning the impulse response is not a true representation of the sound. Increasing the spatial volume of the receiver reduces the

risk of missing paths, but increases the chance of false paths being included which would have passed a smaller received by.

4.3 Image source method

The Image source method (ISM) also uses geometric rules and is based on the same principles of optical reflections as ray tracing. When a sound wave hits a hard surface it reflects; this can be modelled as an image of the original source on the opposite side of the boundary, attenuated by the acoustic impedance of the boundary (see Figure 6); this is the basis of the ISM. It is assumed that the reflection is angularly independent, which is true if the surface is smooth in comparison to the wavelength being modelled.

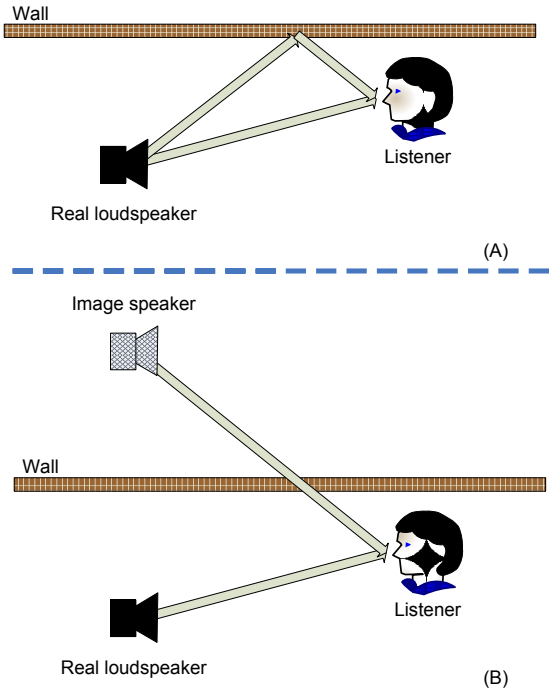


Figure 6: Image source equivalent of a reflection.

Allen and Berkley [1979] argued that 'the image solution of a rectangular enclosure rapidly approaches an exact solution of the wave equation as the walls of the room become rigid'. Under the condition of smooth surfaces, MacLaverty and Furlong [1992] state it has total accuracy. The ISM method is not restricted to rectangular rooms, but is simplest to calculate for such rooms.

The impulse response is calculated as a histogram using information about the delay to each image and their attenuation. The advantage of the image source method is that it is very accurate, once all the image sources which contribute to the impulse response are correctly identified. The disadvantage is that if the room is not rectangular considerable computation is required to locate image sources which contribute to the impulse response.

4.4 Finite and boundary element analysis

Finite element method and boundary element method analysis are numerical approximation techniques for solving complex problems. The finite element method is based around the theorem that a continuous physical problem can be transformed into a discretised finite element problem with unknown nodal values [Nikishkov 2007]. The nodal values can be found by solving linear equations, which allow the values inside the elements to be recovered. An important feature is that a piece-wise approximation of physical fields on finite elements provides good precision; by increasing the number of elements any precision can be achieved. The boundary

element method is similar; however it is the boundaries which are divided into surface elements, rather than volume elements as used for finite element method.

The drawbacks of both methods is that to achieve good accuracy at high frequencies, as required for the direction of arrival algorithm, the elements would need to be very small, thus there would be a high computational load and high memory requirements. Keeping the element size unchanged as the room size is increased, the computational load will increase rapidly due to an increased number of elements being used. Svensson and Kristiansen [2002] suggest that typically 6-10, or even more elements per wavelength are required, depending on the method used. Therefore they only recommend the finite element method for low frequencies. The direction of arrival algorithm works with signals in the 10kHz range, as explained in Chapter 6. Taking 8 elements per wavelength as an example, a one meter cubic volume would require nearly 13 million elements. The boundary element method is a numerical solution to the Kirchoff-Helmholtz integral equation: it requires that the elements are at least $1/8^{th}$ of the wavelength [Svensson and Kristiansen 2002]. Therefore 53824 elements would be required for a one meter square surface.

Kopuz and Lalor [1994] have successfully applied both the finite element method and the boundary element method for modelling the acoustics of a car's interior and boot, modelled as two connected chambers. They found that the results were almost identical for the two methods.

The obvious advantage of these methods is that they can work well with unusual room geometries, or furnishings inside the room. A specific problem for volume element methods is modelling a non enclosed space. To keep the modelling simple enough it would therefore have to be assumed that all doors and windows were closed. The boundary element method does not suffer this limitation.

4.5 Selected method

It was decided that the modal frequency method was not suitable for this project. This is because the direction of arrival processing needs to occur at high frequencies, where this method is not recommended. Also it is not possible to easily extend the model to work with the frequency response of the boundaries, or irregularly shaped rooms.

The ray tracing method was also rejected because it cannot guarantee accuracy, and the accuracy of the impulses is very important in direction of arrival processing, since any errors in the impulse response could reduce the performance of the direction of arrival algorithm and lead to mistaken decisions in the development and tuning of this algorithm.

Finite or boundary element methods are the most versatile, being able to include irregular shaped rooms, but at the expense of complexity and high computational load to obtain the required accuracy. The ISM is able to give guaranteed accuracy under

reasonable conditions and is simple to implement with a low computational load for rectangular rooms.

The proposed solution to predicting the sound within the room is to identify the image sources which cause reflections. Using the ISM for simulating the impulse response of the room, the theoretical locations of these images are calculated, and can easily be plotted. These can be compared to the estimated locations found using the direction of arrival algorithm. This fast and easy validation of the estimates is valuable in the development of algorithms, so the image source model was selected for simulating different rooms.

4.6 Model produced

4.6.1 Delay modelling

The method proposed by Allen and Berkley [1979] calculated the distance to each image; when creating a sampled impulse response the delay, relating to this distance, was rounded to the nearest sample. Although this is valid for low frequency modelling with a high sample rate it does not precisely simulate the arrival time, thus the phase of the signal, which is required for direction of arrival processing with multiple microphones [Peterson 1986]. In the sampling of a real impulse response the sound is measured by a microphone in the receiver's position; this signal would be low pass filtered by an antialiasing filter prior to sampling, limiting the signal to half the sample rate. The effect of this low pass filter is to spread the received energy over multiple samples. The filter needs to be designed to accurately preserve the signal,

having a flat frequency response below half the sampling frequency, and high attenuation above. The filter should perform well with impulses arriving at arbitrary times, thus a short time-domain response is required. There is a compromise between frequency-domain and time-domain requirements.

An alternative approach is to implement a fractional sample delay; this assumes that the signal is bandlimited. The theory is to implement a linear phase, allpass filter with unity amplitude and constant group delay. There are several techniques for designing fractional sample delay filters, as described by Laakos et al [1996] in their review paper. The chosen method was the Lagrange interpolator, due to its simplicity and fast coefficient calculation. An even order Lagrange filter has a constant phase delay with frequency, which is desirable. Low order Lagrange filters have a lowpass characteristic: a 6^{th} order filter was therefore selected because it has minimum attenuation below half the upper frequency limit.

The measurement equipment available had a maximum sample rate of 48kHz and the direction of arrival processing was designed to work with 10kHz signals. To minimise the errors introduced by the Lagrange filter, the simulations were performed at 96kHz and the experimental, measured signals were upsampled by a factor of two to make them equivalent to the simulated case. The antialiasing filter in the recording equipment was at a high enough frequency to not affect the transmitted signal, so it did not need to be included in the simulation.

4.6.2 Image positions

The order of an image refers to the number of reflections it represents. For example a 1st order image has reflected off one surface, whilst a 3rd order image has undergone 3 reflections. The impulse length for the simulation is set in advance, usually from a real measurement of the RT60 decay time, or by simulating a high order model and measuring the RT60 decay time. The model order must then be high enough to include all images that contribute within the desired impulse duration. Rindel [2002] shows that the number of image sources increases exponentially with model order; if n is the number of surface and i is the image order, N_{source} is the total number of possible image sources.

$$N_{Source} = 1 + \frac{n}{(n-2)}\left((n-1)^i - 1\right) \approx (n-1)^i$$

For example, if there are 6 surfaces and the reflection order, $i = 10$, $N_{Source} = 14648437$, the majority will not contribute to the impulse response. For an image to contribute to the impulse response it must satisfy three conditions:

- Distance: the image must be within a radius of the desired impulse length times the speed of sound, for example one 10th order image might fall within the desired radius, whereas an image created from a far away wall might be 4th order and be beyond the radius.
- The image must be valid: it must be caused by a source (real or image) reflecting from an inside wall out, not an outside a wall reflecting in.
- The image must be visible: there must be a direct line of sight from the listener to the image.

Borish [1984] provides details of how the locations of all images can be ascertained and how the above conditions can effectively be tested. However, if the room has a rectangular geometry the visible and viability conditions are always true and the image form a lattice structure. This greatly simplifies the search for images which contribute to the impulse response. For speed and ease of development the room's geometry was restricted to a rectangle: this was decided to be valid, considering that the majority of living rooms are either rectangular, or can be approximated as rectangular.

Figure 7 shows an example room with a single source in the top left corner of the room, the room being shown by the bold continuous lines.

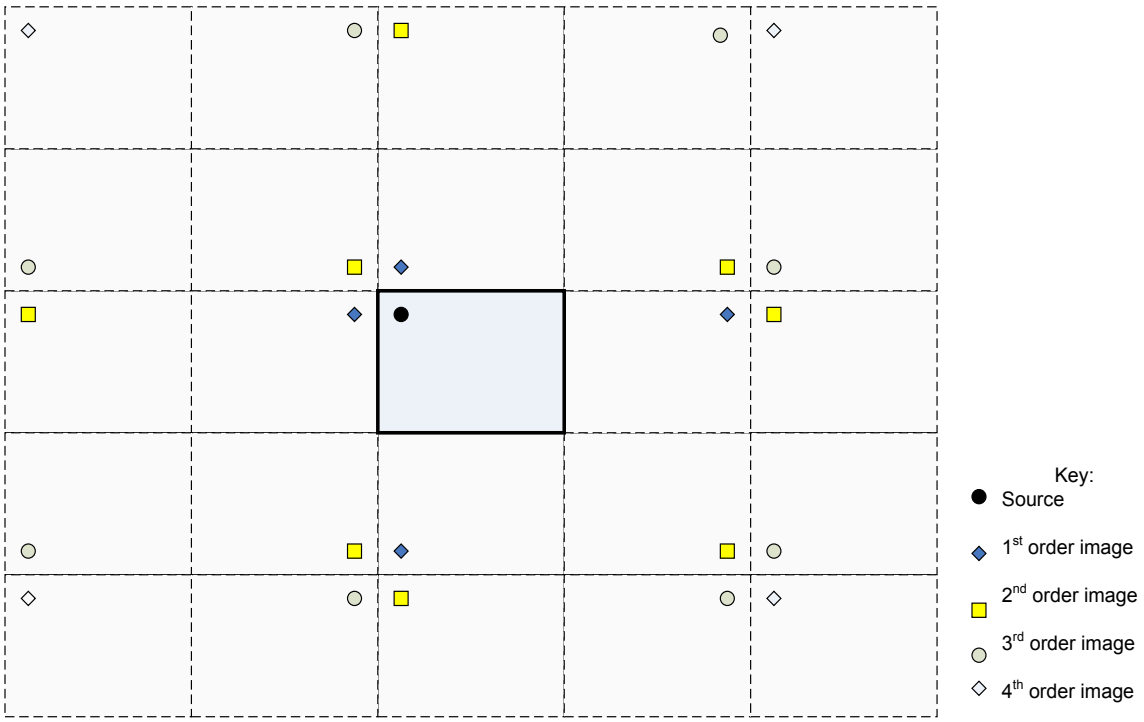


Figure 7: Image source method lattice example.

Several authors have published hybrid ray tracing and image source methods which are designed to ease the image identification and testing when a non rectangular room is being modelled. Ray tracing is used to efficiently locate the images, while back tracing is then used to test if they are visible at the listener's location, thus reducing the computational load of the pure image source method. The impulse response is then formed using image positions, as with the image source method. Rindel [2000] provides a detailed description and references to the original work by Vorlander [1989].

4.6.3 Boundary responses

The frequency response of each boundary in a real room will not be flat. Different materials will absorb sound differently and some will absorb more at high frequencies whilst others will have a greater absorption at low frequencies. Everest [2002] describes the reverberation chamber method of measuring a material's absorption, the method commonly used to specify architectural materials. A room specifically designed to be reverberant is used: the reverberation of the room is recorded with and without the material sample placed on the floor and the difference is used to calculate the absorption of the material at different frequencies. Research has located several tables containing the response of different materials typically found in rooms with respect to frequency [Beranek 1996; Raichel 2000; Long et al 2005]. The absorption is usually given in octave bands from 125Hz to 4kHz. Table 2 shows an average of absorption, taken from Douglas Campbell's Roomsim program [Campbell 2004].

The reverberant room method is a more realistic measure of absorption than the tube burst method, as described by [Everest 2002], because it is not constrained to sound hitting the test material at right angles, as in the tube burst method. However, a number of researchers have highlighted problems of inconsistency with the reverberant room method [Cops et al 1995, Takahashi et al 2005]. It is pointed out that it is very difficult to produce a truly diffuse sound field, and even if it is possible, when the test material is added into the room the sound field is unlikely to still be diffuse due to its absorption and edge diffraction.

Material	Frequency, Hz					
	125	250	500	1000	2000	4000
Anechoic	1.00	1.00	1.00	1.00	1.00	1.00
Quiet room	0.90	0.90	0.90	0.90	0.90	0.90
Percent50	0.50	0.50	0.50	0.50	0.50	0.50
Echoic	0.01	0.01	0.01	0.01	0.01	0.01
Acoustic tile average	0.10	0.30	0.80	0.85	0.75	0.65
Acoustic tile rigid	0.20	0.40	0.70	0.80	0.60	0.40
Acoustic tile suspended	0.50	0.70	0.60	0.70	0.70	0.50
Acoustic plaster	0.10	0.20	0.50	0.60	0.70	0.70
Plaster on lath	0.20	0.15	0.10	0.05	0.04	0.05
Gypsum wallboard	0.30	0.10	0.05	0.04	0.07	0.10
Plywood	0.60	0.30	0.10	0.10	0.10	0.10
Brick	0.03	0.03	0.03	0.04	0.05	0.07
Concrete unpainted	0.40	0.40	0.30	0.30	0.40	0.30
Concrete painted	0.10	0.05	0.06	0.07	0.10	0.10
Concrete poured	0.01	0.01	0.02	0.02	0.02	0.03
Vinyl tile on concrete	0.02	0.03	0.03	0.03	0.03	0.02
Heavy carpet on concrete	0.02	0.06	0.15	0.40	0.60	0.60
Heavy carpet felt backing	0.10	0.30	0.40	0.50	0.60	0.70
Platform floor wooden	0.40	0.30	0.20	0.20	0.15	0.10
Ordinary window glass	0.30	0.20	0.20	0.10	0.07	0.04
Heavy plate glass	0.20	0.06	0.04	0.03	0.02	0.02
Glazed wall	0.01	0.01	0.01	0.01	0.01	0.01
Draperies medium velour	0.07	0.30	0.50	0.70	0.70	0.60
Water	0.01	0.01	0.01	0.01	0.01	0.01

Table 2: Surface absorption data [Campbell 2004].

The values in these tables are purely for the material, since in a real room some sound will pass through the material to the next room, or outside. Additionally sound will escape from any non airtight joins in a room, for example the gaps around a door. Northwood [1968] performed experiments on plasterboard walls and found that the transmission loss through a plasterboard wall could be 12dB at 125Hz, meaning a

significant amount of energy could be heard outside the room. The amount of energy transmitted outside a room will depend greatly on the room's construction, therefore a frequency independent attenuation, per wall was added to the simulation to make the measured and simulated responses match.

To model the boundary's frequency response Campbell [2004] chose to take the frequency response and interpolate it and his programme then creates an FIR filter based on the magnitude response. The drawback of the approach is that the filter is acausal, the pre-ringing before the image's delay is not an accurate model of the effect of the boundary. The approach taken for this research was therefore to create an IIR filter using shelf and parametric filters. The parameters of these filters were identified using a LMS search, to best match the filter response to the corresponding values in Table 2 for the particular material being modelled, similar to that described by Ramos and Lopez [2006]. The response of each image is then a cascade of boundary responses, depending on the combinations of boundaries the image has reflected off. For efficient processing the cascading was implemented by multiplying the frequency responses of boundaries. Figure 8 to Figure 11 show frequency responses of the tabulated frequency response compared to the model filter response for different materials.

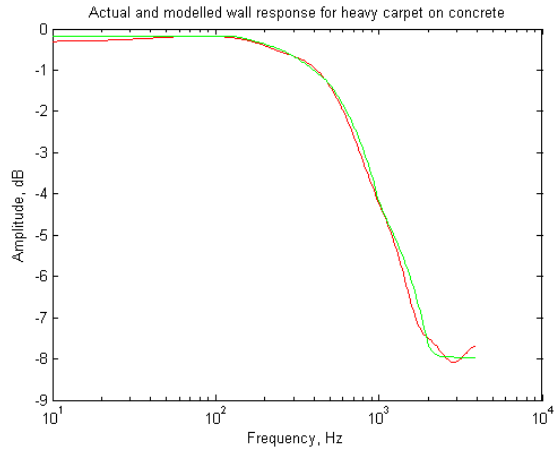


Figure 8: Absorption: Carpet on Concrete.

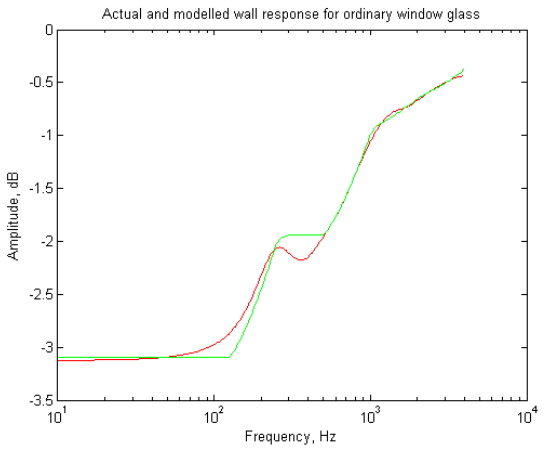


Figure 9: Absorption: Ordinary window glass.

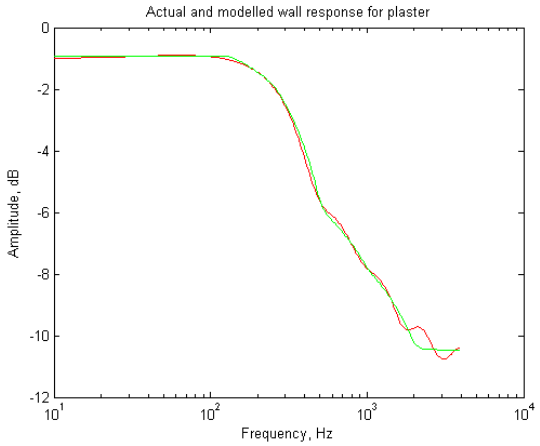


Figure 10: Absorption: Plaster

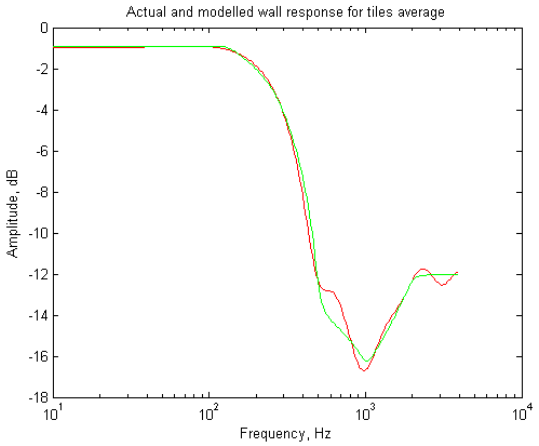


Figure 11: Absorption: Tiles average

Figure 8 to Figure 11:Room surface absorptions. Green actual, from Table 2. Red modelled

The proposed system uses a high frequency signal to locate the image locations, and a low frequency signal to estimate the frequency response of these images. Because data was not available at the high frequency required it was decided to model the boundaries with the filters described above at low frequencies and a constant attenuation at high frequencies. A high frequency attenuation coefficient of 1.9dB was selected to make the simulated responses similar to the measured responses recorded in real rooms with the high frequency LFM signal.

4.6.4 Comparisons

In comparison to a typical living room the model produced uses a good estimate for a given room's geometry. The model however does not include openings in the boundaries for objects like doors and windows. The model is also not able to model surfaces which are not smooth. Objects within the room, like furniture, are not currently considered but the model could be extended at a later date to include these. These compromises have been made for ease of development, both in terms of modelling computational load and clarity in interpreting the results.

4.7 Results

The accuracy of the room model was verified by comparing the measured time domain responses, recorded in a real room at different positions to a simulation of this room at the same locations. Presented here is one example position; the measured response after matched filtering is shown in Figure 13, and the corresponding simulation in Figure 12. Both were generated with a Linear Frequency Modulated chirp of centre frequency 10474Hz and duration of 1.67msec after matched filtering. These results were taken in the same room as the results presented later in this thesis. Its length was 8.57m by width 3.53m and height 2.70m. It can be seen that the shape is similar, but although there are some differences between the amplitude of some reflections, the arrival times are very accurate. Figure 14 and Figure 15 are zoomed in versions for added clarity. The real loudspeaker's response is also visible, particularly in Figure 15 where the ringing after the source and each reflection can be seen.

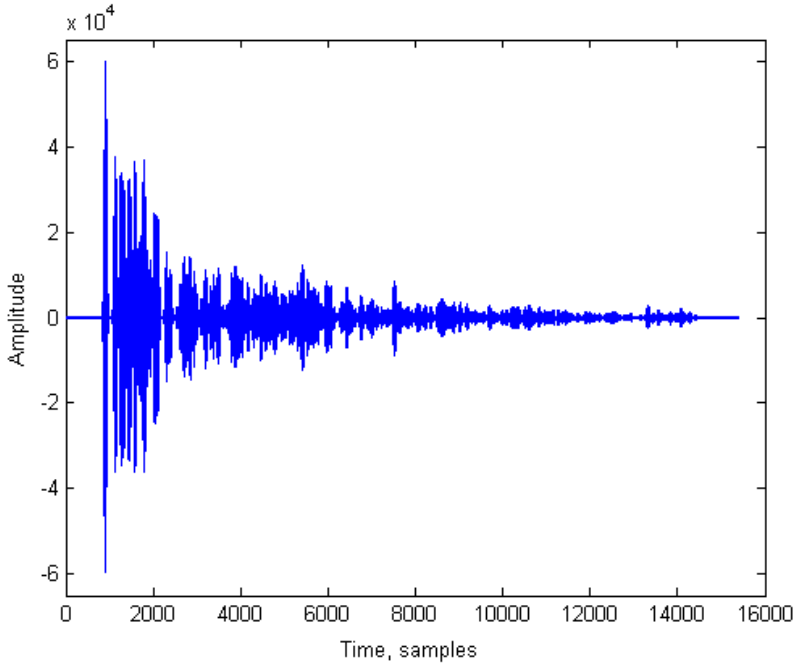


Figure 12: Simulated response, time domain, sample rate = 96kHz.

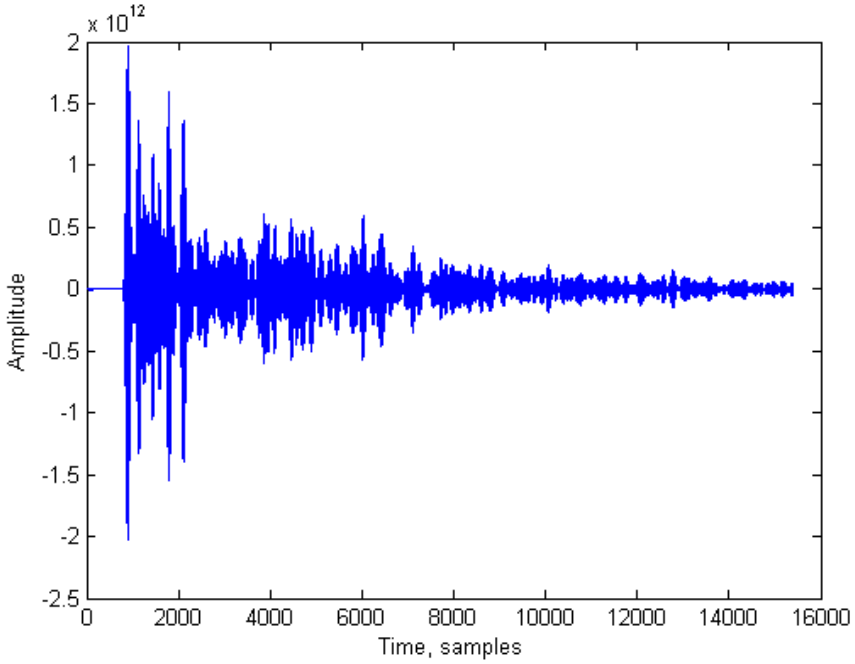


Figure 13: Measured response, time domain, sample rate = 96kHz.

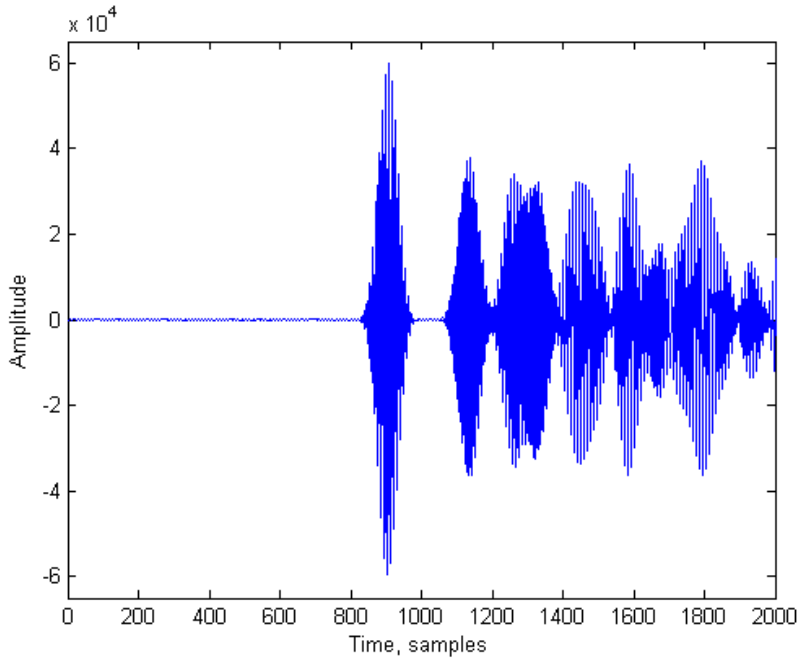


Figure 14: Simulated response, time domain, zoomed, sample rate = 96kHz.

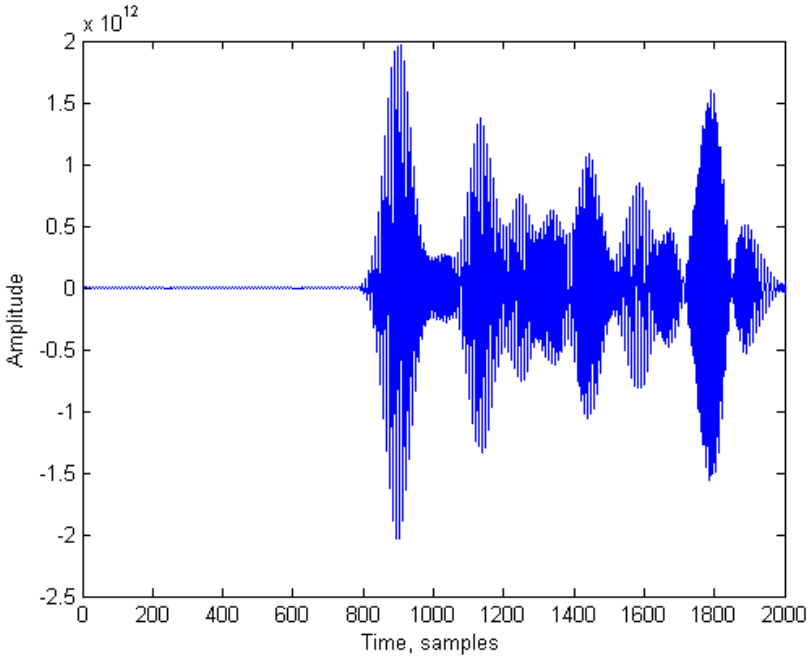


Figure 15: Measured response, time domain, zoomed, sample rate = 96kHz.

4.8 Chapter conclusions

Producing an efficient model of a room was identified as an important step in the development of the system. It was produced to aid the development of different algorithms which would be used to analyse the reflections in a room, caused by a loudspeaker. Different methods for simulating a room's acoustics have been presented and the most suitable was selected for accuracy and to aid interpretation of results. Details of the implementation developed are presented. The model has been extended to increase the accuracy of an image arrival time and by improving the way in which boundary frequency responses are incorporated. The model had been verified by comparing the arrival times of reflections in a real and simulated room.

Chapter 5 Temporal Processing

The first stage of the proposed system is to calculate the time of arrival of each reflection. This is required for two reasons; firstly it allows the distance to each reflection to be calculated, which is needed to calculate the reflection's coordinates. Knowing the time difference between the sound being transmitted and an individual reflection arriving at the microphone, the distance to the reflection can be calculated by multiplying the time by the speed of sound. The second reason is that the direction of arrival algorithm improves when there are only a few signals arriving at the array in a time segment. If the number of signals arriving at the array in a given time segment is larger than, or equal to, the number of microphones being used, the angle of arrival of all signals cannot be accurately resolved [Van Trees 2002]. Therefore for the direction of arrival algorithm to be successful small sections of the room's response need to be processed at a time. The arrival time of each reflection is used when identifying the correct time segments.

5.1 Transmitted signal

Methods exist that are able to estimate the parameters of a system without prior knowledge of the transmitted signal; however these blind algorithms have a higher computational load and are less accurate than when knowledge of the transmit signal is used to identify the response [Aichner et al 2005, Bell and Sejnowski 1995, Cardoso 1998]. Because a domestic loudspeaker system is used to play back pre-

recorded media there is no problem with knowing the signals transmitted from the loudspeaker.

In a final, complete system it would be desirable for the prediction system to work with the music or film soundtrack being played. It is proposed that the audio could be analysed and sections which have the appropriate characteristics could be used for the estimation of the image locations. Alternatively test signals could be added to the loudspeaker signal, at very low amplitude and transmitted for a long time which is undetectable to the listener, but after matched filtering the room's response could be recovered. However, for ease of development test signals were transmitted at amplitudes which produced a good signal to noise ratio.

Initially the distance to each reflection was estimated with a logarithmically swept sinewave as the transmit signal. (The choice of a logarithmic sweep for impulse measurements is explained in Chapter 7, along with its design.) The advantages of a logarithmic sweep are that linear response can be completely separated from the nonlinear response, and a good signal to noise ratio is easy to achieve. A separate, narrowband signal was then transmitted for the direction of arrival processing, followed by a third signal, a low frequency logarithmic sweep, used for estimating the frequency responses of the reflections. The direction of arrival algorithm requires a short burst of narrowband signals. During the development of the spatial processing the transmit signal was changed from a short burst of windowed sinewave to a linear frequency modulated chirp, (LFM), which after matched filtering looked like a short burst of windowed sine wave. The advantage of this was that the LFM signal and

matched filtering process dramatically increased the processing gain, and thus the signal to noise ratio was no longer a problem. The reduced noise level made it possible to identify angles of arrival late within the reverberation where the reflected amplitude was low.

The selected deconvolution method required an expected signal of an individual reflection, as described in section 5.2. The deconvolution method was proved to perform well with simulated data, where the loudspeaker's polar response was constant, although performance decreased with measured data. This was concluded to be because the response of the loudspeaker was varying with frequency and angle. At high frequencies it had greater off axis attenuation than the low frequencies. The expected signal used in the deconvolution was an on axis recording. Refer to section 6.7.3 for example figures of a real loudspeakers' polar response.

To reduce the effect of a real drive unit's polar response a narrowband signal was desirable. Although the signals attenuate with angle the spectral distortion will be less for a narrowband signal, resulting in a better deconvolution of the measured data. The complication, however, was that the deconvolution performs best when few signals overlap in time, thus a short pulse was desirable, which cannot be achieved for a truly narrowband signal. It was decided that a good compromise would be to use the same test signal for the distance estimation and the direction of arrival estimation. The first advantage of this was that the reflection's distances were used to select time sections for spatial processing; thus using the same signal reduced the scope for errors due to time variations affecting the results. Secondly because only two test signals were

required, the LFM chirp and the low frequency chirp, the measurement process was quicker and simplified.

5.1.1 Linear frequency modulated chirp design

The LFM signal can be designed so that after matched filtering it resembles a short burst of CW wave. The optimal pulse duration is calculated from twice the maximum dimension of the array, as will be explained in section 5.4. Because different array sizes were experimented with, the LFM signal was optimised for each array configuration. Taking, for example, a line array of 23 microphones with a spacing of 15mm and a speed of sound of $345ms^{-1}$, the desired pulse duration would be 2ms.

$$duration = \frac{spacing * microphones * 2}{speed\ of\ sound}$$

To prevent spatial aliasing the transmit signal must have a half wavelength greater than the microphone spacing. With a spacing of 15mm the maximum transmit frequency is 11.5kHz. This example will use a frequency of 10969Hz so that the frequency is 5% less than the maximum. An example of a sinusoid pulse meeting these requirements is shown in Figure 16A. The LFM signal therefore needs to be designed to have a pulse duration of 2ms after matched filtering and an upper frequency less than 11.5kHz. The bandwidth of the LFM signal is largest for the shortest duration pulse. This occurs when there are 9 line array microphones, the minimum number for a 24 microphone circle array, as described later in Chapter 6. With 9 line array microphones the upper -3dB point is 11.2kHz, meaning that spatial aliasing will not occur, regardless of the number of line microphones used.

Pulse compression is a technique which has been commonly used in radar and sonar applications to achieve a short pulse when the transmit power is limited [Collins and Atkins 1999; Varshney and Thomas 2003 and Blunt and Gerlach 2004]. These applications firstly require long range operation, which needs a long transmit signal, processed to minimise noise, and secondly fine range resolution, implying a large bandwidth. These prerequisites are the same for the proposed room acoustic measurements; to accurately identify individual reflections a high signal to noise ratio is required and to estimate their angles of arrival a short pulse is essential. These requirements can be accomplished with a Linear Frequency Modulated (LFM) chirp signal and matched filtering [Collins and Atkins 1999, Varshney and Thomas 2003]. The LFM signal, y(t) is given by the function

$$y(t) = \text{rect}\left(\frac{\text{t}}{\text{T}}\right) \cos(2\pi f_{start} t + \pi \alpha t^2)$$

where time, *t*, is from 0 to T, the transmit duration, f_{start} is the start frequency of the chirp and α is the LFM slope, $\alpha = \frac{B}{T}$ for a low to high frequency sweep and $\alpha = \frac{-B}{T}$ for a high to low frequency sweep, where *B* is the bandwidth, [Skolnik 2008]. The rect function is defined as

$$rect(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x \leq T \\ 0 & x > T \end{cases}$$

The received SNR can be maximised with use of a matched filter, this is the optimal processing for a signal buried in additive Gaussian noise [Stewart and Westfielf 1959]. The matched filter is implemented by cross-correlating the transmit signal with the received signal.

The equation for the LFM signal can be modified so that the bandwidth is set to meet the desired matched pulse duration, centred around the carrier frequency, *fc*, as follows:

$$tx(t) = \text{rect}\left(\frac{\text{t}}{\text{T}}\right) \sin\left(\begin{matrix} 2\pi\left(f_c - \frac{1}{matched\_duration} \times \frac{0.6}{2}\right) \times t \\ +\pi \times \frac{1}{T} \times matched\_duration \times 0.6t^2 \end{matrix}\right)$$

To prevent an initial transient that the amplifier and speaker could not reproduce, a short ramp at the start and end of *tx* was applied. The matching signal was the time reversed complex equivalent to extract the I and Q channels, improving the envelope estimation by reducing the ambiguity of locating the main lobe peaks, as expressed below:

$$match(t) = \text{rect}\left(\frac{-\text{t}}{\text{T}}\right) \exp\left(\begin{matrix} 2\pi j\left(f_c - \frac{1}{matched\_duration} \times \frac{0.6}{2}\right) \times -t \\ +\pi j \times \frac{1}{T} \times matched\_duration \times 0.6t^2 \end{matrix}\right)$$

The main lobe of the autocorrelation function of a LFM chirp looks like a short burst of sinusoid, with the same frequency as the LFM's centre frequency, as shown in Figure 16B. For high BT products the output approximates a sinc function, thus there are sidelobes at -13dB and below, as in Figure 16C where a logarithmic amplitude scale is used. The sidelobes can be suppressed by windowing the transmit signal, with the consequence of broadening the main lobe, therefore increasing the effective sinusoidal pulse duration and reducing the peak amplitude. Different window functions will attenuate the sidelobes differently, as detailed by Harris [1978]. The peak attenuation can be compensated for by increasing the transmit duration so that the processing gain cancels out the attenuation of the window. Experiments were

performed and the window selected was a Blackman window because it gave good sidelobe suppression of -87dB when the window was applied to both the transmit and replica signals, without broadening the main lobe excessively. Figure 16D shows the matched signal, whilst Figure 16E shows the logarithmic amplitude of the windowed signals. It can clearly be seen by comparing Figure 16C and Figure 16E that the window has greatly suppressed the sidelobes.

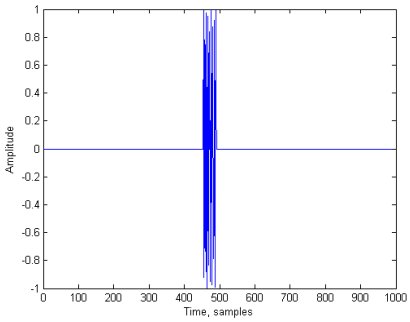
The effect of windowing the transmit signal caused a broadening of the received, matched pulse. To compensate for this and return the pulse to the required duration a compensation value of 0.2917 was experimentally found. A different compensation value would be needed for a different window function. The final transmit and matching functions are therefore:

$$tx(t) = \text{blackman(t)} \sin\left(\begin{matrix} 2\pi\left(f_c - \dfrac{1}{matched\_duration} \times \dfrac{0.6 \times 0.2917}{2}\right) \times t \\ +\pi \times \dfrac{1}{T} \times matched\_duration \times 0.6 \times 0.2917t^2 \end{matrix}\right)$$

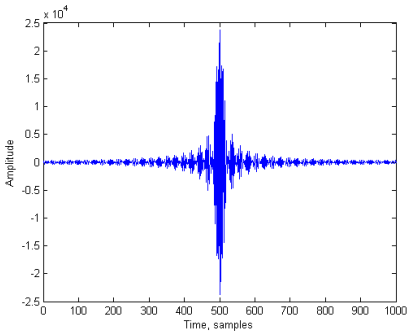
$$match(t)$$

$$= \text{blackman}(-\text{t}) \exp\left(\begin{matrix} 2\pi j\left(f_c - \dfrac{1}{matched\_duration} \times \dfrac{0.6 \times 0.2917}{2}\right) \times -t \\ +\pi j \times \dfrac{1}{T} \times matched\_duration \times 0.6 \times 0.2917t^2 \end{matrix}\right)$$

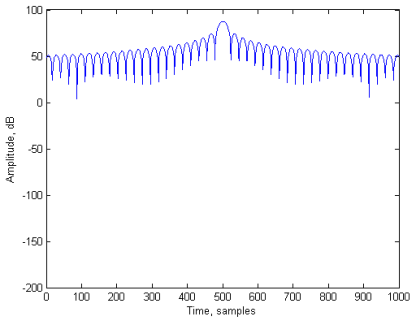
Figure 16F and Figure 16G show the matched output with linear and logarithmic amplitudes respectively.



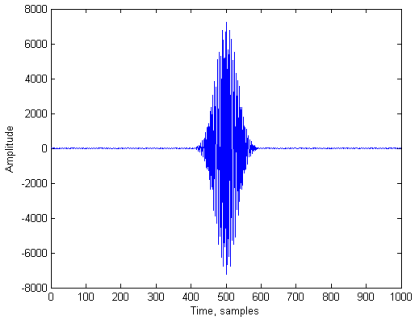
A: Pulse of sinewave



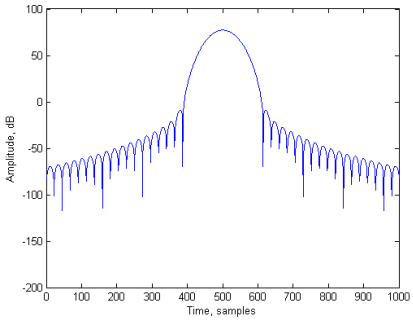
B: LFM matched



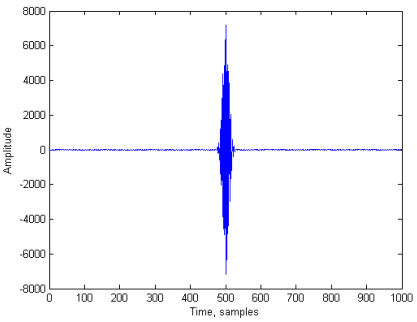
C: LFM matched output logarithmic amplitude



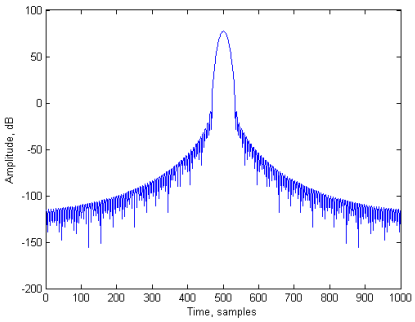
D: LFM windowed matched



E: LFM windowed matched logarithmic amplitude



F: LFM windowed, compensated matched



G: LFM windowed, compensated matched logorithmic amplitude

Figure 16: LFM design.

The processing gain of the solution indicates the reduction in noise level which has been achieved. When a burst of sinusoid was transmitted its duration was adjusted to match the array length being investigated, the duration of the LFM signal was a constant 1 second. Therefore the processing gain achieved varied between 27dB for an array length of 21 microphones to 32dB for an array length of 7 microphones.

5.2 Temporal separation

Figure 17 shows a simulated response, matched filtered, at the array's centre microphone, with a LFM signal designed for a 19 sensor line array (1.7ms) and centred at 10474Hz. The room model was the same as was used for results presented throughout this thesis. Figure 18 shows the room's response with each reflection modelled as a perfect impulse; no fractional sample delay was used, to add clarity to the graph.

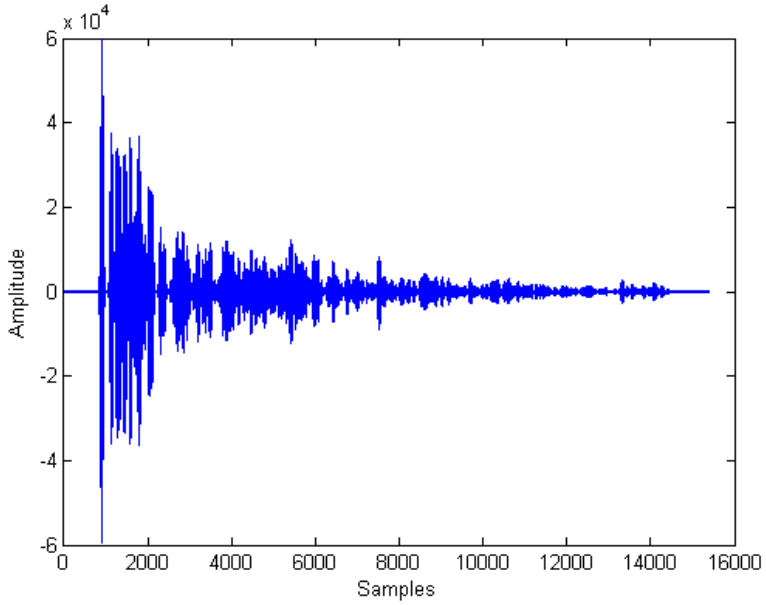


Figure 17: Simulated response, from LFM test signal

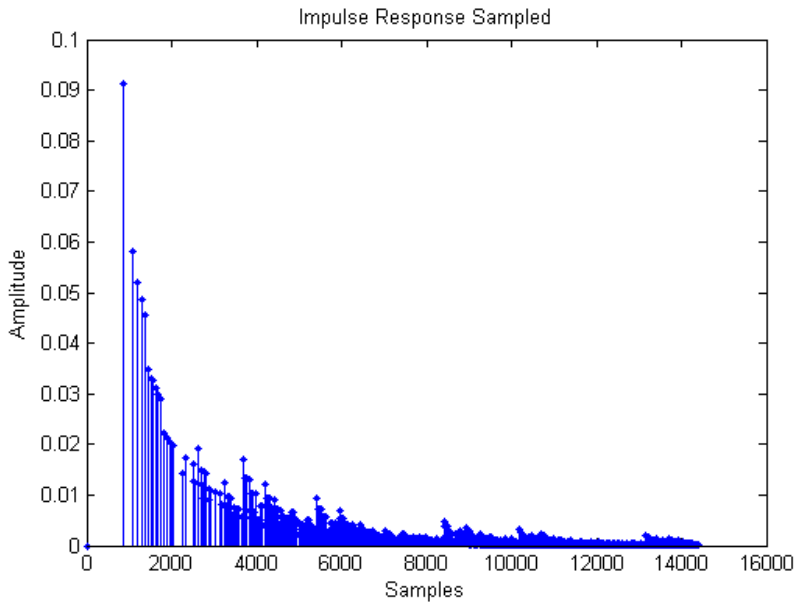


Figure 18: Simulated response, infinite bandwidth

It can clearly be seen that the sinusoidal pulses are overlapping with each other. After matched filtering the received signal must be separated into individual reflections. Possible methods of doing this are discussed below.

5.2.1 Peak detection

A simple algorithm to detect peaks was considered. Such an algorithm could work well when the reflections were independent; however when the reflected sound overlaps in time, with other reflected signals, it would not be possible to identify the multiple signals as separate signals. This would create errors in the distance estimation, but more crucially the correct time section might not be selected for the direction processing, and images would be missed altogether. To overcome this, a deconvolution algorithm was required to separate individual reflections.

5.2.2 Linear deconvolution

The disadvantage of an inverse filter is that the frequencies where the signal's amplitude are low will be boosted, resulting in an amplifying effect of any noise present. A linear deconvolution filter is therefore implemented by filtering the received signal with a band limited inverse filter, to limit the noise amplification. There is a trade-off in the design so that the noise is minimised and the bandwidth of the signal is preserved [Collins 2004]. In the ideal case (i.e. infinite bandwidth) linear deconvolution results in a series of Dirac functions with different delays and amplitudes. A further process would be needed to measure the time delays.

The problem with linear deconvolution is that nulls in the received signal will be emphasised in the deconvolved result [Zhou et al 1990]. For example if multiple signals overlap in time, with different phases some cancellation will occur. A linear deconvolution will not estimate the amplitudes of the signals correctly because the phase relationship will not be unwrapped.

5.2.3 Non linear deconvolution

Non linear deconvolution can be implemented using a number of techniques, but the idea is to minimise the difference, E, between the received, matched filtered signal, $y(t)$ and the convolution of, $h_{EST}(t)$, the channel estimate, and the transmission signal, $x(t)$ [Collins 2004].

$$E = \overline{[x(t) \otimes h_{EST}(t) - y(t)]^2}$$

Although the deconvolution uses nonlinear processing, the system being deconvolved is still linear. Non linear deconvolution algorithms fall into two catagories, statistical algorithms and maximum likelihood algorithms. For computational efficiency a statistical algorithm was selected, Matching Pursuit specifically, because of its proven performance with room response deconvolutions [Collins 2004].

5.2.3.1 Matching pursuit

The selected Matching Pursuit (MP) algorithm is an iterative algorithm which selects a waveform, from a dictionary of waveforms that is best adapted to approximate part of the received response; gradually building up a better estimate of the received signal each iteration. The waveforms which make up the dictionary are called atoms, with the atom having the largest correlation with the received signal, *y(t)* being selected. The weighted, delayed contribution of this atom is subtracted and the iteration proceeds on the residue [Goodwin 1997]. Because it maintains energy conservation, convergence is guaranteed [Mallat 1993]. It is described by the following pseudo-code [Collins 2004]:

Initialization: $h_{EST}(t) = 0$

Loop:

Cross-correlate *x(t)* signal with *y(t)*, find the delay τ where $x(t - \tau)$ best matches *y(t)*. Then find the amplitude A which minimises the mean-square-difference:

$\int [y(t) - Ax(t - \tau)]^2 dt$

Subtract $A\ x(t - \tau)$ from *y(t)*.

Add $A\,\delta(t-\tau)$ to $h_{EST}(t)$.

This is repeated until the residue is below a preset threshold; h_{EST} then represents a sampled estimate of the room's impulse response. The time delay applied to each atom represents the time delay to each reflection in the original signal.

5.3 Results

The deconvolution was performed on the central microphone of the array because the time section selection required the time delay result from the centre microphone. The matching pursuit algorithm relies on a dictionary of possible signals to compare the signal against. It had been identified that in the case of a real loudspeaker the response would change with angle. The angles of the reflections which make up the received signal are not yet known, or estimated and varied greatly. The options were therefore to use a measurement of the loudspeaker on axis, or to use measurements of the speaker at different angles. Because the surfaces causing the reflections will modify the signal and some reflections were caused by off axis loudspeaker wavefronts, it was not expected that an individual reflection would be represented by a single deconvolved atom. Therefore it was decided that to make the system as simple and computationally efficient as possible, a single on axis recording would be used. This was recorded with the array 1m from the speaker, in the centre of the room so reflections could be removed by time gating. In the simulated experiments a simulation of the loudspeaker was used as the expected signal.

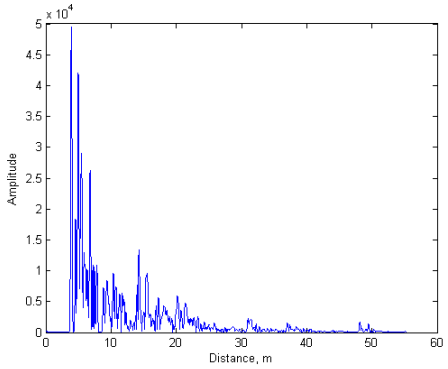
As expected some reflections were identified as a group of closely spaced atoms, which is correct, as a super position of these atoms will best match the reflection's response. This is not a problem because the consequence of multiple atoms is multiple time sections will be processed, and if multiple time sections represent one reflection this will result in several images being estimated very close together. The processing performed in the wall frequency response estimation stage, detailed in Chapter 7, will remove these multiple images and the best match will result.

Up to 200 iterations were used for all of the deconvolutions, with the stop criteria being the maximum of the residue reducing below 0.2% of its maximum amplitude. These numbers were chosen after careful tests. The values ensured that the residue was very small in all cases, meaning all reflections had been identified. A larger number of iterations or a lower threshold could have been used, but more time sections would have been processed, increasing the overall computational time with no benefit. To test the performance of the deconvolution a measure of the difference between the estimate of the received signal and the original signal, i.e. the residue left at the end of the deconvolution, can be used. Additionally the reflection times, known exactly in the simulated case, and known within a small margin in the measured case by simulating the room, were compared to the deconvolved results and found to match accurately.

The results presented here are for a transmit signal with a centre frequency of 10969Hz and pulse duration of 1.7ms, equivalent to twice the distance of 19 microphones with 15mm spacing. The received signal is shown in Figure 19A for the

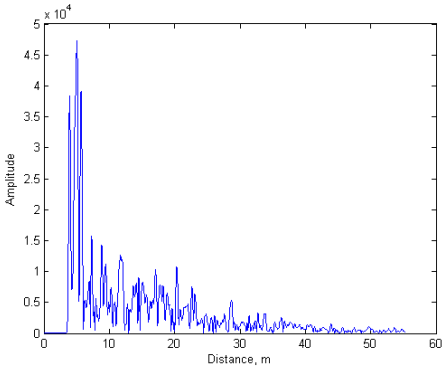
simulated case and Figure 19B for the measured, front drive unit's case. The corresponding residues, after matching pursuit, are shown in figures Figure 19C and D. The very low level of signal left after the matching pursuit deconvolution demonstrates that the deconvolution has been successful. Figure 19E and F show a zoomed in time section of the received signal and the atoms which were identified from it. Good peak identification can be seen, along with some reflections estimated by multiple atoms. It can be seen that the longer impulse length of the real drive unit is causing greater overlap of the reflections than in the simulated case, resulting in a more smeared response. The simulated case where the drive unit's polar response is not modelled works best; however all the reflections are still identified in the measured and simulated cases.

Simulated case:

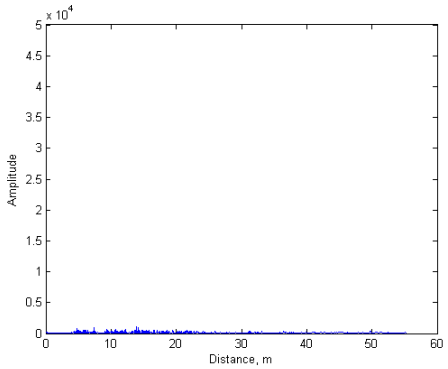


A: Received signal

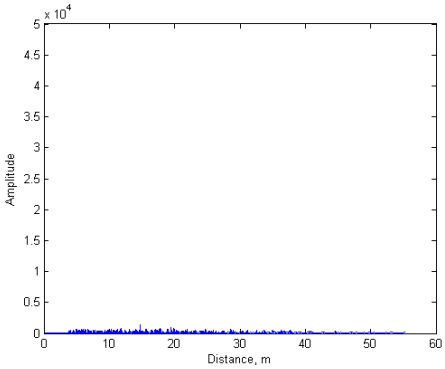
Measured case, front drive unit:



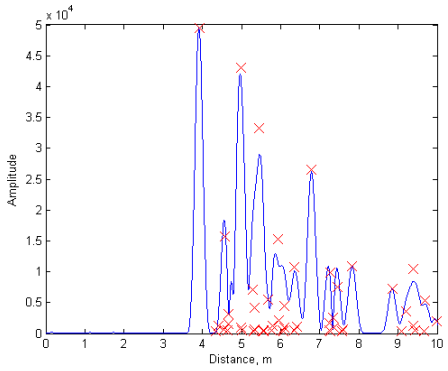
B: Received signal



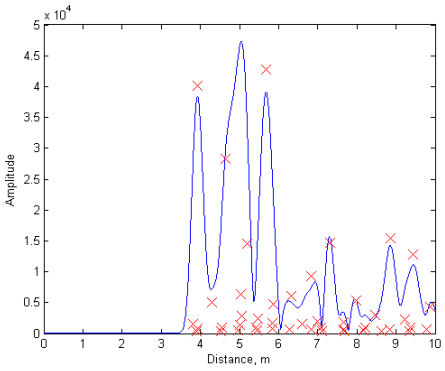
C: The residue left after MP deconvolution.



D: The residue left after MP deconvolution.



E: Zoomed in section of the deconvolution result. Blue is received signal, red crosses are estimated reflections.



F: Zoomed in section of the deconvolution result. Blue is received signal, red crosses are estimated reflections.

Figure 19: Matching pursuit results

5.4 Time section selection

In this section the pulse duration refers to the main lobe of the LFM signal after matched filtering. In a reverberant room there are many reflected signals, some of which will arrive at the microphone array at very similar times, as shown in Figure 17 and Figure 18. The optimal sinusoidal pulse duration for the direction finding is one which guarantees all microphones in the array are receiving the signal from the target under investigation, i.e. the array is fully sonified, as shown in Figure 21D. This occurs when the pulse duration is greater than the time it takes for the pulse to travel across the array. A long pulse will increase the overlap of reflections, which would reduce the ratio of microphones to signals impinging on the array; if there are more signals than microphones the signals' arrival angles cannot be resolved. Therefore the pulse duration should be as short as possible whilst still ensuring the array is fully sonified. The start of the time subsection for angular processing is not until the signal has arrived at all microphones (see Figure 21C), and must stop once this condition is not met (Figure 21E). This condition can be tested by measuring the arrival time of the signal at the centre microphone of the array and knowing the pulse duration.

Because the direction of arrival is not known in advance the worst case of a signal arriving at endfire, from either direction must be considered: Figure 20 illustrates both.

Figure 20: Signal arriving at endfire from the left and right. Signal in blue, array illustrated as a black box.

For the array to be fully sonified the wavefront must travel the effective duration of the array. A transmit pulse equivalent to one length of the array would provide one sample from each microphone for the direction processing, before the fully sonified condition fails. This assumes exact identification of the arrival time: any variation, as might be caused by the deconvolution algorithm, will cause the fully sonified condition to fail. Even if precise time measurements were achieved any phase noise would limit the angular resolution. These effects can be reduced by increasing the time subsection, and thus transmit pulse duration. Experiments indicated that the optimal pulse duration was twice the array length, resulting in time sections of one effective duration of the array being selected. The matched pulse has a double Blackman amplitude window applied, thus taking the middle section of the pulse maximises the signal to noise ratio, therefore minimising the chance of any phase errors. Because the size of the array was varied, the pulse duration and time section width were also varied for each array configuration investigated.

The following example is for a signal arriving from the left of the array, from endfire, parallel to the array and shows the start and stop conditions for the time section.

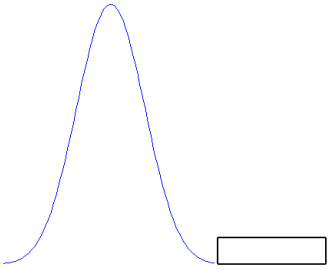


Figure 21A: The signal arriving at the array (illustrated as a black box), but not detected.

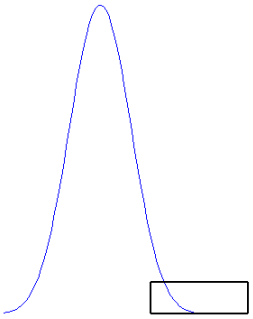


Figure 21B: The signal later in time when the start of the wavefront has reached the centre microphone. This time can be calculated from the deconvolution of the received centre microphone's signal.

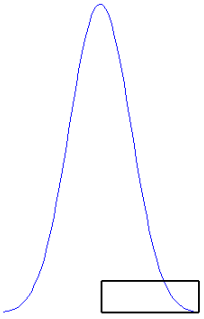


Figure 21C: Start of time section. Because the effective pulse duration is twice the array's length the start of the time section is the arrival time at the centre microphone plus a quarter of the pulse's duration.

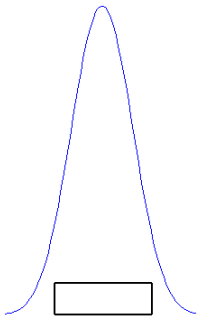


Figure 21D: Array fully sonified. As the wavefront move across the array sensors all the sensors are able to sample the wavefront.

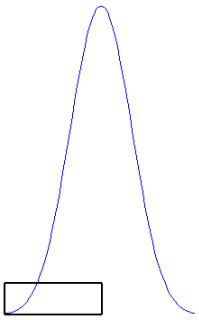


Figure 21E: End of time section. The time section must stop when the array is no longer fully sonified. This time is calculated by the arrival time at the centre microphone plus three quarters of the pulse's duration.

Figure 21: Time section selection

This can be expressed algebraically as follows:

m = Array sensors, number of microphones.

d = array sensor spacing

c = speed of sound in air = 345ms^{-1}

e = effective duration of the array

$$e = \frac{m \times d}{c}$$

p = pulse duration (matched LFM signal's duration)

$$p = 2 \times e$$

$t_{detected}$ = time deconvolution detects the signal arriving at the array's centre microphone.

t_{start} = *time section start*

t_{end} = *time section end*

$$t_{start} = t_{detected} + \frac{p}{4}$$

$$t_{stop} = t_{detected} + \frac{3 \times p}{4}$$

Therefore the start and stop times of each time section can be calculated from the time an individual reflection arrives at the array, identified by the matching pursuit algorithm.

5.5 Chapter conclusions

This chapter has shown the design requirements of the transmit signal used for both the detection of the reflection's arrival time and its angle of arrival. A linear frequency modulated chirp was selected and its design detailed. This section covered matched filtering for optimal noise processing, sidelobe suppression and design for a desired pulse of sinusoid duration.

Deconvolution methods were discussed so that the arrival time of each reflection could be estimated. Justification for a non linear algorithm, called matching pursuit, was presented along with details of the algorithm's implementation. An example set of results for simulated and measured experiments was shown and discussed. The deconvolution was proved to work well, with all major reflections identified. Some

reflections were represented by multiple atoms, this was expected and the consequence will be discussed in Chapter 7.

Lastly the results from the deconvolution were used to select the start and end times of the subsection which were used in the spatial processing to determine the arrival angles.

Chapter 6 Spatial Processing

In Chapter 5 the sound recorded at a single microphone was processed to identify the time delay to individual reflections using matching pursuit deconvolution. It was not able to identify how many signals were arriving at a particular delay: it is likely that in most time segments several reflections from different objects within the room will arrive at the microphone array, thus the spatial processing must be able to resolve these overlapping signals. The aim of the spatial processing was to estimate how many reflections were present and the angle of azimuth and elevation of each reflection relative to the array. Combining this information with the distance, estimated from the temporal processing, the coordinates of the reflections could be estimated. Spatial processing in a general case involves the processing of multiple sensors to estimate the signals' spatial parameters; it is defined by the following quotes:

"*Distributed sensor networks are used to monitor an area, with the aim of detection, identification, localization and tracking of one or more objects*" [Chen et al 2002a].

"*The quintessential goal of sensor array signal processing is the estimation of parameters by fusing temporal and spatial information, captured via sampling a wavefield with a set of judiciously placed antenna sensors*" [Krim and Viberg 1996].

There are many and varied applications of these techniques, from military surveillance or combat enemy identification, to hearing aid design and video conferencing systems to enhance the audio signal, or to seismic event localization.

Mobile communication systems use beamforming to reduce the effect of multipath from reflections. In the case of a hearing aid or conferencing system, beamforming can be used to suppress background noise so the user can hear a person talking more clearly. In sonar applications a submarine must identify and track a source, deciding if it is an enemy submarine or not; beamforming is used to estimate the angle of arrival of the source and track its position.

In each of these applications the commonality is the localization of a source, performed using a measure of the arrival time, in a wideband system, or phase in a narrowband system, of an impinging signal, and comparing this with the arrival time or phase at adjacent sensors. There are many branches of spatial processing due to the wide and diverse applications where it has been applied. Therefore, decisions had to be taken to narrow down the options, leading to a proposed solution. These different techniques and branches will now be briefly described along with decisions taken and their justifications. As the scope becomes more focused more details are covered. The requirements of the solution for this application are summarised, followed by the system's development, including the array parameters and algorithm development. Simplified simulated experiments were used to validate the design before simulated and real room data was analysed.

6.1 Background

The first consideration is what classification of wavefronts are expected to be received by the sensor array, farfield or nearfield. An ideal point source will be one which

transmits uniformly in all directions, called an isotropic source. A conventional, single cone loudspeaker is a point source, but it does not transmit sound uniformly in all directions. It will have a polar response where the off-axis high frequency response is attenuated [Colloms 1999]. (For further details, refer to section 6.7.2.) A point source will produce a spherical wave, where the amplitude is inversely proportional to the distance from the source [Colloms 1999, Krim and Viberg 1996]. Considering the surface of the sphere at a distance of radius R, all points on the surface will have the same phase; this is referred to as the wavefront. A nearfield source is one where a radiated wavefront, as samples by a line array will have different phases, whereas the wavefront of a farfield source will have the same phase, resulting from a plane wave. The classification as a nearfield or farfield source is therefore relative to the sensor array size; a small section of the wavefront, far from the source, will approximate to a straight line, therefore considered farfield, whereas for a source close to the array the curvature of the wavefront will be measureable [Krim and Viberg 1996]. Some nearfield solutions use information about the curvature of the wavefront to determine the range and the angle of the source. The proposed solution is to have a small array, the largest experimented with being 35cm in length, thus for a typical room the vast majority of sources are farfield. Therefore source localization techniques relying on the wavefront curvature were discounted.

There are two categories of source localization methods considered: model based and blind. Model based beamforming relies on prior knowledge of the sensors' location, their responses and received signals, whereas blind methods do not [Chen et al 2002a, Feng and Kammeyer 1998, Saruwatari et al 2002, van der Veen 1998]. An example

where blind methods are well suited is talker tracking, used in high quality hands free systems for in-car phone calls, or teleconferencing [Saruwatari et al 2002]. The received signals parameters are only loosely known, so that by locating the talker the speech can be enhanced and background noise suppressed; an example of the alternative, model based technique to solve the same problem is presented by Kobayashi et al [2002], using a conventional beamformer. Blind methods were discounted because in this application the sensors' responses and locations can be specified and known. This author believes that not using this information would have increased the complexity and computational requirements and would therefore probably have produced poorer results of the reverberation analysis, a view shared with Spence and Clarke [1999].

Another major consideration is the parameterisation or specification of the signals the array is expected to receive. They can be divided into two categories, wideband and narrowband; the array can also work in passive or active mode. Wideband solutions are often passive systems, particularly seismic event or vehicle detection where the system designer has no control over the received signal [Pham and Sadler 2004]. Chen et al [2002a] claims that it is best to transform the received data into the frequency domain and then perform narrowband DOA estimation on the frequency bins. This is the technique used by Valaee and Kabal, [1995] also by Wang and Kaveh [1984] and Bourennane and Bendjama [2000]. The results from the angle estimates from each frequency bin are then averaged together to give a final result. To prevent spatial aliasing the sensor array must be designed so that the spacing of the sensors is no greater than half the wavelength of the highest frequency signal

expected; the Nyquist rate. If the signal's wavelength is significantly larger than the sensors' spacing the phase difference between the sensors will be very small. Because narrowband direction of arrival algorithms utilize the phase difference of the signals, a significantly longer wavelength will reduce the resolution of the array. The optimal frequency bin to perform the DOA estimate is therefore just less than the spatial aliasing frequency. The wideband experiments presented by Bourennane and Bendjama [2000] had a centre frequency of 110Hz with a 40Hz bandwidth, demonstrating that wideband spatial processing cannot operate over the entire audio band (20Hz to 20kHz). Other wideband direction finding techniques estimate the time difference of arrival of the signals, but this only works for a single signal or uncorrelated signals [Feng et al 2001], which is not the case in a reverberant room. The conclusion drawn is that in this application there is no benefit from wideband processing: narrowband processing has a lower computational load and is likely to have superior resolution.

A direction finding system can operate in passive (receiving only) or active mode (transmits known signals and receives their reflections). This system will use an active mode of operation, through test signals being transmitted from each loudspeaker in the room in turn, because parameters of the received signal are then known, greatly reducing the computational load and reducing sources of error, thus increasing the resolution of the image positions. An active system reduces any benefits of a wideband transmit signal, reinforcing the decision to use narrowband signals.

6.2 Similar research

A literature search has found no solutions which directly address the issue of locating reflections within a reverberant room. A large amount of research has been conducted on array geometry and direction of arrival algorithm processing, but they have mainly focused on either freefield conditions or where only a few coherent sources are present. The case of reverberation spatial analysis has not previously been tackled.

The most closely related research found, however, was presented by Merimaa [2002], who used a small 3-D microphone array for room acoustics measurements. The 12 microphones were recorded and processed to form wideband differential directivity patterns, using 2^{nd} order spherical harmonics. These beams were used for surround sound recordings and sound intensity measurements in different directions. Whilst the topic of source localization was introduced and discussed, the author did not develop the ideas, identifying the project as future work.

Research presented over three papers from the Delft University of Technology [Berkhout et al 1997, Vries et al 1996 and 1997] cover an investigation into wave field analysis. Both simulated and measured experiments were presented for a line array of microphones spanning across the room, in free space and parallel to the source. They processed the data with the linear Radon transform, a technique borrowed from seismic signal processing allowing an image of the wavefronts to be calculated over time. Individual wavefronts can be seen, but because a line array was used, it is not possible to discriminate between different vertical angles. The recent

paper by Vries et al [1997] attempts to address this by using two microphones to form a cardoid response which is rotated, so the maximum power can then indicate the elevation angle. However they conclude that this is only valid for individual, non-coinciding plane wave components, thus it was not successful in 3 dimensional rooms. Additionally the array was significantly larger than would be acceptable in a domestic environment and the angle of arrival estimate would be of lower resolution than the solution proposed in this thesis.

6.3 Solutions requirements

The conclusions from the above review of localization categories led to the decision that the system proposed would operate in active, narrowband mode. It would be model based, using prior information about the transmit signal and array geometry to reduce the number of parameters being estimated and therefore increasing the chances of success. The received signals would be expected to be farfield, plane waves, highly correlated, or coherent, with many overlapping signals, therefore multiple small time sections would be processed to estimate each reflection's direction of arrival.

The remaining parameters affecting the accuracy and performance of spatial processing are the array geometry and calibration and the direction of arrival algorithm employed. The algorithms and array geometry are affected by the number of signals arriving at the array at the same time. In the case of a reverberant room the number of signals is unknown in advance; this posed a significant problem and meant

that several solutions which are reported to work well in free field situations with few sensors were unsuitable for the reverberant, complex situation faced in this research.

6.4 Array geometry

The performance of an array can be affected by the number and type of sensors used, their mounting and placement as well as the weighting applied to their signals. According to Van Trees [2002], there are three common measures used to assess an array's performance:

- Array gain versus white noise
- Directivity
- Sensitivity and tolerance factor

The first consideration in specifying an array design is the element number. The more sensors in the array, the higher the processing gain will be and the narrower the beamwidth. Therefore the signal to noise ratio can be improved by adding more sensors. Additionally the theoretical limit on the number of signals that can be resolved by a direction of arrival algorithm is one less than the number of sensors. Therefore in many applications, where there are no hardware or processing constraints, the more sensors the greater accuracy can be achieved. However all sensors must be receiving the signal at the same time. Increasing the sensor count means the density of sensors must increase or the array's size will increase, causing an increased sample time or time segment. In many applications this is not a problem. But in this application a longer segment time increases the number of signals at the

array, thus at some point there will be more signals than the array can resolve. There is therefore a tradeoff in the size of the array and its resolution. A closed form mathematical answer is not possible because the number of signals arriving at the array is affected by the room geometry, source and array location, all unknown when the array is being designed.

Mechanical arrays reduce the number of sensors required by mechanically moving one or more sensors, usually in a circle to sample a large space with fewer sensors than would otherwise be required. To prevent motor noise disturbing a listener in a final system, mechanical arrays were excluded.

The directivity of a receiving array is a measure of the rejection of isotropic uncorrelated noise [Acoustics and Sonar Group 2008], that can be manipulated by the sensors' placement, mounting and weighting. Different placement arrangements will be covered in the following sub-sections. Their spacing can be generalized as follows: if the sensor spacing is greater than a half wavelength of the received signal then diffraction secondaries will result. It is the same as sampling in the time domain where the Nyquist rate limits the upper frequency limit, without aliasing. The diffraction secondaries produce aliases in the array's response. In some instances this is acceptable because the array is only designed to work over a limited angular range, thus spatial aliasing can be ignored. In this application, however, it was important that the array should perform over all angles, thus to avoid ambiguity the sensors should be less than half wavelength spaced.

A design criterion was that the arrays should be as small as possible to fit in with a domestic environment, thus the sensors should be as close together as possible and this spacing will set the centre frequency of the transmitted signal. Due to the microphones available the closest spacing achievable was 15mm, which was the spacing chosen. The upper frequency limit of the array was therefore 11.5kHz. This was an appropriate frequency because a higher frequency would suffer more from the directivity of the loudspeaker and potentially the microphones. The frequency was also high enough above common crossovers between drive units in typical 2 or 3 way loudspeakers, thus avoiding crossover distortions or multiple displaced driven units transmitting at the same time, causing multiple reflections and complex directivity patterns. The loudspeaker's time response is also short at high frequencies, which helps reduce overlap of reflections, thus increasing performance.

Because the signals may arrive at any angle omni-directional sensors must be used, although the mounting of an individual sensor can affect its directivity. Mounting a sensor on a baffle, in comparison to in free space, can reduce the amount of energy received from behind the sensor. It was therefore decided that the arrays would be baffle mounted and placed very close to the wall with the aim of preventing image sources from behind the array from being detected. Because omni-directional sensors were selected the field of view of the array was a hemisphere, baffled by the mount and wall it was mounted against.

The sensors of an array can be weighted differently using a tapering function, and usually these are symmetrical about the centre of the array. The benefit of this

technique is that the directivity and sidelobes can be controlled. For a uniformly spaced linear array a taper function is equivalent to a window function for time samples. The result is therefore the same; the sidelobe level decreases at the cost of broadening the main lobe [Acoustics and Sonar Group 2008]. There was no benefit in taper functions with the angle of arrival algorithm selected; firstly because a narrow main lobe was desirable and because the output of the array was not going to be used directly sidelobes were not a constraint, as long as they were smaller than the main lobe. The second disadvantage of non uniform sensor weighting is that the sensitivity of the array to sensor imperfections, (placement, gain, phase) is increased when the sensors are not uniformly weighted because the array gain decreases with non uniform weighting [Van Trees 2002]. Therefore it was decided that all sensors would be uniformly weighted.

6.4.1 Line array

The simplest array geometry is the uniform linear array, as shown in Figure 22.

A line array is able to estimate the angle of arrival in one dimension. However, because it is symmetrical it is not possible to discriminate between signals arriving from other dimensions, thus there is a cone of ambiguity in the direction estimate, as shown in Figure 23. By mounting the sensors in a baffle the signals from one direction can be greatly attenuated, thus the array can 'see' signals in a hemisphere, rather than a sphere. A uniform line array is the optimal geometry for estimating the angle of arrival in one dimension. The resolution of a line array is greatest at broadside ($\theta=0$) where the element to element phase difference is zero. The resolution decreases as the signal's angle approaches endfire (along the axis of the

array) because of a broadening of the main beam; note that the directivity of the array remains constant with angle if the element spacing is half wavelength [Van Trees 2002].

Non-linear spaced line arrays do not have simple design formulae for the beamwidth and sidelobe levels: the latter increase in level compared to linear spaced arrays [Acoustics and Sonar Group 2008]. There is usually not a sufficient improvement over uniform spacing, so they were ruled out for this piece of research.

Figure 22: An example line array

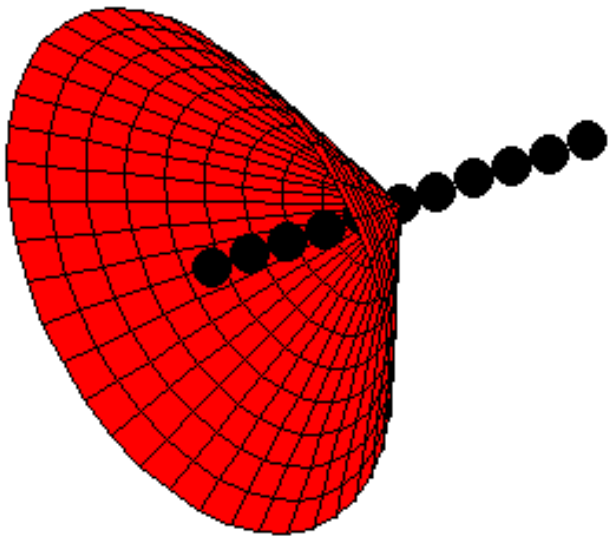


Figure 23: Line array's cone of ambiguity

6.4.2 Planar array

A planar array consists of sensors mounted in two dimensions, usually in a uniform grid. An example is shown in Figure 24. The sensors can be processed all together, or as separate, or sub, arrays. This has a good directivity, pattern and high processing gain but is costly to implement due to the high number of sensors required.

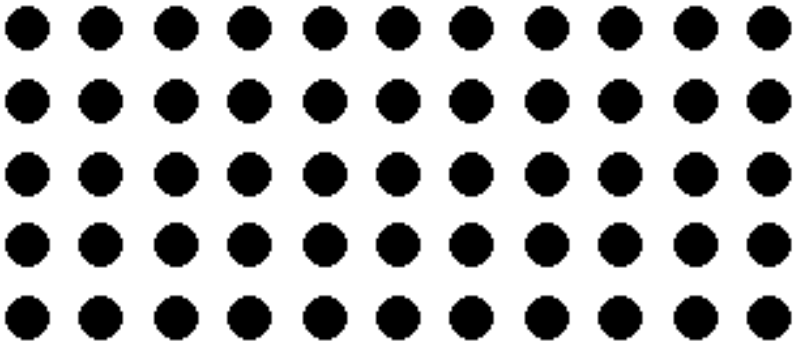


Figure 24: An example planar array

6.4.3 Circular array

A circular array is also a two dimensional array (see Figure 25), and has the advantage of using far fewer sensors than a uniform planar array, but it has a poor directivity pattern. When a signal arrives close to 90 degrees, say from the side, relative to broadside, the directivity is unevenly biased. The sensors at the side of the array have very similar phase; therefore summing their received signals gives a large amplitude. Alternatively the sensors at the top or bottom or the array have very different phases: when summed cancelation occurs so the result is significantly less than the sum from the side. The sensors are processed all together. This results in a directivity pattern where the amplitude varies with arrival angle.

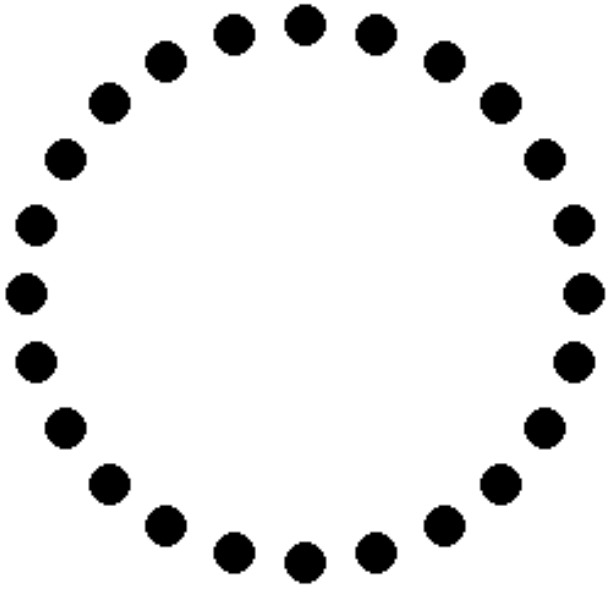


Figure 25: An example circular array

6.4.4 Cross array

A cross array is another example of a two dimensional array requiring only a few sensors. As shown in Figure 26, it composes of two mutually orthogonal line arrays, known as the Mills cross, after its designer Bernie Mills. It can either be processed as two separate line arrays, or as a two dimensional array. Processing as a single array, the directivity pattern is very uneven. Using the same example as given for a circular array, of a signal arriving from the side, all the sensors in the y axis will sum in phase whereas all the sensors from the x axis will have different phases. This means that its directivity is less even than a circular array, making it undesirable. If it is processed as two separate lines a single signal's azimuth angle can be estimated by the array in the x axis and the elevation by the array in the y axis. Working out the intersection of the two cones of ambiguity the two dimensional angle of arrival can be calculated. This approach is very computationally efficient because the entire two dimensional

space does not have to be searched by the direction of arrival algorithm. However, if two or more coherent signals arrive at the array at the same or very similar times it cannot be guaranteed to work because the azimuth and elevation angles cannot always be paired correctly.

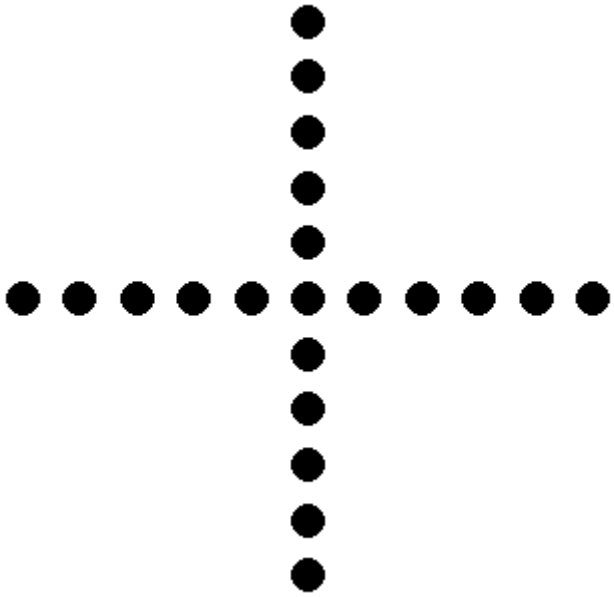


Figure 26: An example cross array

6.4.5 Sparse array

Sparse arrays can be designed to optimise the placements and weighting of a reduced number of microphones to best approximate a response. The techniques can be applied to a line, planar or even volumetric (3 dimensional) array. An ideal array response is chosen, which might not be implementable directly in practice due to sensor number or data acquisition limitations. The placement and weights of a lesser number of sensors are adjusted to approximate a desired response. Haubrich [1968] proposed a least mean squares algorithm to solve the non-linear problem. Due to their

sparse arrangement it is not suitable to process the data in sub arrays unless it has been designed accordingly.

6.4.6 Comparison and proposed solution

It was not possible to create a planar array for the practical experiments due to the high number of microphones and data acquisition required; the cost was prohibitive. A bespoke geometry was therefore designed to have a very high resolution to sensor number ratio. Out of the array geometries investigated a line array had the most even directionality in 1D and was thus used for determining the azimuth angle: a typical living room has a greater length/width than the height, so the azimuth requires greater resolution than the elevation. The elevation angle was then determined by a circular array, but with the azimuth angle fixed, as found by the line array. This minimises the directionality problems of a circle array. Also by fixing the azimuth whilst estimating the elevation in this way, the entire 2 dimensional space does not need to be searched by the direction finding algorithm. This two stage processing greatly reduces the computational load, thus providing an efficient, fast solution. This design is unique in that it takes the advantages of the Mills' cross's two stage processing but solves the problem of multiple signals. The resolution achieved by this arrangement, as tested by simulations, was no worse than the sparse array with the same number of microphones, and was better than using just a circular array on its own. The microphone placement is shown in Figure 27.

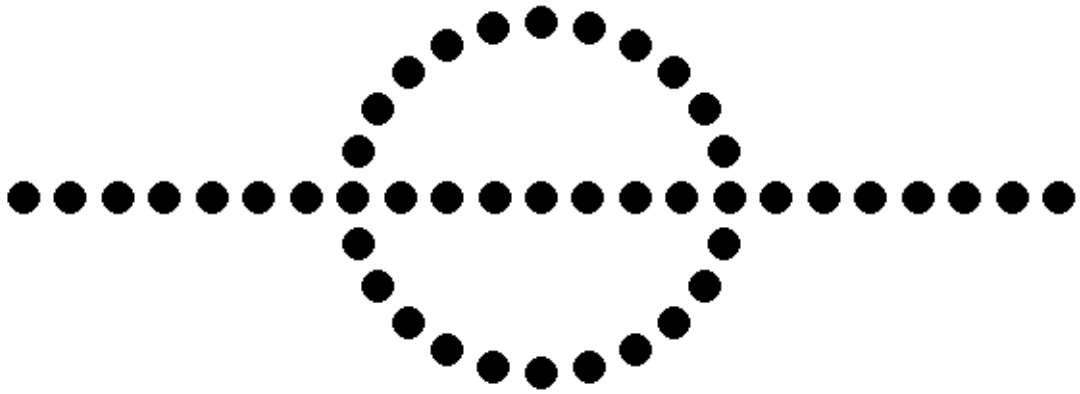


Figure 27: Proposed array

6.4.7 Practical array construction

The optimal number of microphones cannot be determined with closed form mathematics because it depends on the room size and array placement. The larger the array the longer the time section needs to be, which increases the number of signals to detect. The recording equipment available could record up to 24 channels simultaneously. Because of the physical size of the microphones used the line array needed to have an odd number of microphones to allow two of the microphones to be used for both the line and the circle array; this also provided a sensor at the centre of the array, used for the temporal processing. Therefore a line array of 23 sensors was implemented, the maximum length possible with the above constraints. The number of sensors could then be selected and different configurations experimented with. Initial tests were with three circular array sizes, consisting of 12, 16 and 24 microphones each. Only one circular array was used at a time. The simulated and practical results for the 12 microphone circle were poor, however, so this was discounted. With the 16 microphone circle a minimum of 7 line microphones could

be used because this was the diameter of the circle. For the 24 microphone circle the minimum number of line microphones was 9. With each circle the number of line array sensors selected was incremented by two to keep the array symmetrical, from its minimum to the maximum of 23 microphones. The effective pulse duration of the matched LFM transmit signal was optimised to match the array configuration being used. A total of 17 array configurations were therefore investigated.

Initially small round capsule microphones were used. A piece of wood formed the baffle, with holes drilled in it for the placement of the microphones. Because only 24 microphones and data acquisition were available the microphones needed to be moved from the line array to the circle arrays during the experiments. The performance of this arrangement was poor. The estimated DOAs were not consistent with a simulation of the room. The front of each microphone was supposed to be flush with the surface of the baffle, but there was undoubtedly a small variation each time in practice. With a centre frequency of 10474Hz (one of the test frequencies used) a placement error of 1mm would result in an 11 degree phase change. It was therefore decided that a microphone array with permanent microphone placement was necessary.

To ensure the microphones could not move and to give high precision in their location it was decided to construct the array using printed circuit boards (PCBs). A total of 59 microphones were used, 23 in the line, and a 24 and 16 microphone circle, but with two microphones from each circle also being used in the line array. Small, cheap electret microphones were chosen, due to their size, frequency response and

affordability. These were soldered onto one circuit board in the configuration shown in Figure 28. (See appendix A for schematic diagrams and PCB layout.) A second circuit board routed the microphone outputs to an ADAT recording device. Analogue switches were used to select the signals from different microphones, multiplexing the 59 microphones down to eight signals, prior to the microphone preamplifiers and balanced drivers. A PIC microcontroller was used to control the analogue switches through an RS232 interface. (See appendix B for schematic diagrams and PCB layout.) This design allowed the measurement process to be fully automated from a PC with a Matlab script performing the automation and recording; reducing human intervention and thus minimising the sources of errors. To reduce mutual coupling and unwanted reflections the microphones were baffled with a soft foam sheet. Figure 29 show the completed array.

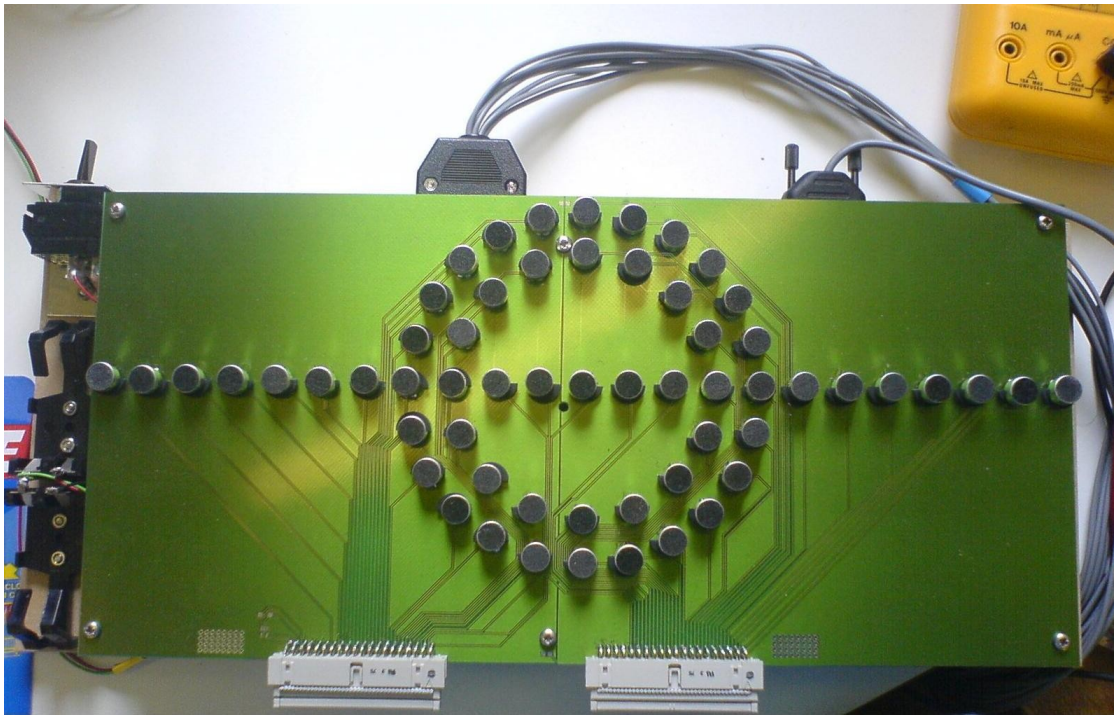


Figure 28: Photograph of the array before baffle added

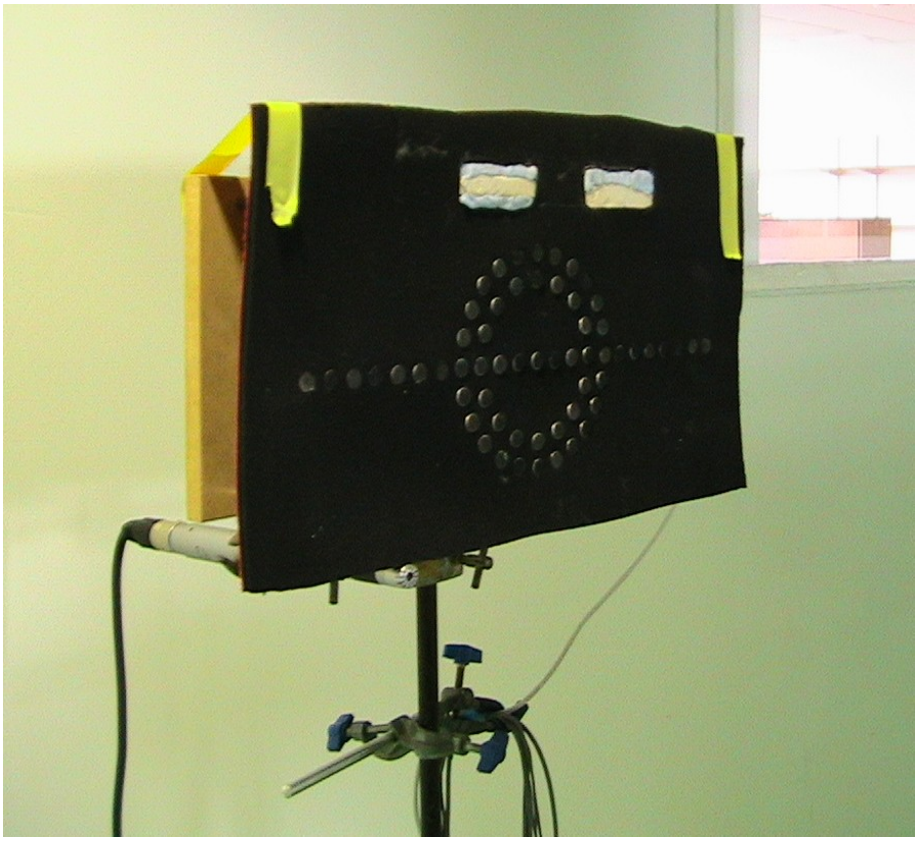


Figure 29: Photograph of the finished array complete with foam baffle

6.5 Direction estimation algorithms

Many different algorithms have been published which are claimed to estimate the angle of arrival of one or more signals using an array of sensors. These direction finding algorithms can be divided into two categories, Closed Form and Parameter based, with both needing to know the speed of sound through the medium. Closed form methods estimate the relative time delay of the signal arriving at the sensors and then localize the source, using these time delays. A parametric method estimates the source location using the received data directly. ESPRIT, developed by Roy et al [1986], is an example of a closed form solution which requires two spaced sub arrays.

A closed form method, even using advanced time delay estimation techniques, cannot guarantee a solution when multiple sources are present; a parametric solution, however, can [Chen et al 2002a], therefore closed form methods were discounted.

Because spatial processing has its origins in military applications the source of the signal being identified is conventionally referred to as 'the target'. In this project the target is the location of the original loudspeaker or the image source positions representing the individual reflection origins.

As outlined in section 6.3 the direction of arrival algorithm for this application must work with an unknown number of correlated or coherent signals arriving at the array in the same time section. Many of the published parametric methods were not suitable, though before they are discussed the conventional beamformer will be presented. This is because although it is known not to perform well in this application it has been the base algorithm used of many more advanced algorithms and is often used in performance comparison making it a suitable starting point.

6.5.1 Conventional beamformer

The conventional (Bartlett) beamformer can be considered as a spatial extension to the matched filter (or Wiener filter). In time domain filtering the time delayed signals are weighted and summed to form the output. Beamforming sums the weighted output of spatially distributed sensors to form a spatial filter. The beamformer enhances the signal from the desired direction and attenuates the signal from other directions [Chen et al 2002a, Krim and Viberg 1996]. This is because signals will

arrive at different sensors at different times; when the signals are summed, depending on the signals' arrival angle to the array, they may sum in phase, creating the main beam, whilst some signals, arriving at different angles will sum out of phase, causing cancellation, and resulting in signals from this direction being attenuated. The beam can be steered to different angles by adjusting the delays (in the wideband case), or phase shifts (in the narrowband case) of the sensor signals prior to summing. This is referred to as the delay and sum beamformer, the simplest of the conventional beamformers.

A system model can be formulated as follows: consider an array of M sensors, the output at each sensor, m, can be represented by $y_m(t)$

$$y_m(t) = a(\theta)x(t) + n_m(t)$$

where $a(\theta)$ is the sensor's response, with respect to angle. $x(t)$ is the received signal and $n_m(t)$ is additive noise. The superposition theorem applies, thus $x(t)$ may represent one or multiple signals arriving at the array. The output of the delay and sum beamformer is then

$$z(t) \equiv \sum_{m=0}^{M-1} w_m y_m(t - \Delta_m)$$

where w_m is the sensor weighting and Δ_m is the delay applied to the signal from sensor m. The time delay needed to steer the beam to an angle θ is calculated from the extra distance the signal has travelled, s_n as shown in Figure 30 for a line array.

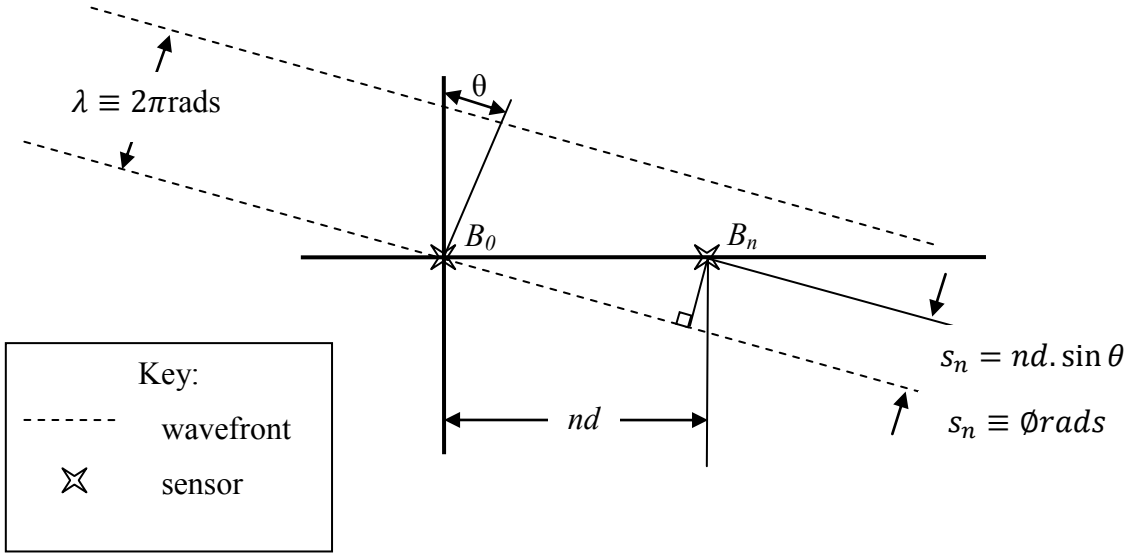


Figure 30: Phase delay required to steer a bean in the direction of a plane wave arriving at a line array.

B_0 is the reference point and B_n is the next sensor. The distance s_n travelled by the wavefront to sensor n is given by

$$s_n = nd.\sin\theta$$

If c is the speed of the signal the time delay is given by:

$$\tau = \frac{s_n}{c}$$

For a narrowband signal with a centre frequency of f the time delay is equal to the phase shift $\emptyset_s$

$$\emptyset_s \equiv \omega\tau_s = \omega\frac{s_n}{c} = \frac{2\pi f}{\lambda f}\sin\theta_s.nd = nK_s d$$

The phase shifts required for each sensor to steer the beam in a given direction are referred to as a steering vector, $m(\theta)$.

Direction finding can be performed with the conventional, delay and sum beamformer by scanning the main beam through all expected angles and recording the power received. Peaks in this power response represent signals detected and their angles of arrival.

The spatial covariance matrix of the received signals is a powerful representation of the received signals because it contains all the information required for direction of arrival processing. It represents the relative phase between pairs of sensors meaning that the overall propagation delay does not affect the beamforming [Johnson and Dudgeon 1993]. Because only the phase difference is used by the direction of arrival algorithm the covariance or the correlation can be used [Philips 2006]. The theoretical spatial covariance matrix can be defined as

$$\boldsymbol{R_{yy}} = \in \{\boldsymbol{y(t)}.\boldsymbol{y(t)^H}\}$$

where *H* represents the Hermitian transpose. A practical estimation of the covariance matrix over L samples can be expressed as

$$\boldsymbol{R_{est}} = \frac{1}{L}\sum_{l=1}^{L} y(t_l)\, y(t_l)^H$$

The power output of the delay and sum beamformer, $P(\theta)$, is therefore:

$$P(\theta) = \frac{m(\theta)^H\, R_{est}\, m(\theta)}{m(\theta)^H\, m(\theta)}$$

The steering vectors for all angles of arrival, as expected at the array can be combined into a matrix **M,** called the array manifold. By evaluating the power output over all expected angles of arrival the spatial power spectrum can be calculated. If only one

signal is present the maximum of this spectrum indicates the angle of arrival of this signal.

Krim and Viberg [1996] state that the drawback of the conventional beamformer is that the resolution can only be increased by increasing the aperture of the array; regardless of the data collection time or SNR. Manikas et al [1997] interprets their limitation differently, stating that the biggest problem of the conventional beamformer is its inability to resolve two closely spaced targets. Both concur than modern high and super resolution techniques do not suffer this limitation.

6.5.2 Maximum likelihood algorithms

Early attempts to improve the resolution of the conventional beamformer involved maximum likelihood (ML) solutions. A maximum likelihood solution is one which tests the data set against all conceivable solutions and uses a cost function to select the most likely match. By testing the data against all solutions it is guaranteed to identify the global maxima of the cost function (or minima for an error cost function), rather than a locally optimal solution as can be obtain from a gradient based solution. The disadvantage of the ML solution is that it has a very high computational cost, due to the nonlinear multidimensional search performed. There are many different ML direction of arrival algorithms proposed, varying by cost functions and efficiency of search methods [see, for example, Bresler and Macovski 1986, Cedervall and Moses 1997, Chen et al 2002a, Stoica et al 1996]. Because the maximum likelihood algorithm is not limited in its requirements it is able to estimate other signal parameters as well as the signals' direction of arrival, at the cost of increasing the

dimensionality of the search and thus the computational time. Examples of additional parameters estimated are the signals' frequencies or temporal responses [Featherstone and Strangeways 1999, Zatman and Strangeways 1995]. It is claimed that estimating the signal's parameters jointly rather than independently is advantageous, for example a DOA can be unambiguously linked with a particular frequency. A maximum likelihood algorithm can process wideband or narrowband signals [Chen et al 2002b], this is an advantage over many other direction estimation algorithms which only work with narrowband signals.

6.5.3 MUSIC

Because many of the applications of direction finding were military based, a real time, or near real time solution was necessary, which could not be achieved with a ML solution. This led to the development of suboptimal, but low computation algorithms. The most common of these is the Multiple Signal Classification (MUSIC) algorithm developed by Schmidt [1979]. It performs super resolution DOA estimation by utilizing the orthogonality between signal and noise subspace. Eigendecomposition is performed on the spatial covariance matrix, resulting in one eigenvalue and vector per detected signal. A search of the array manifold is performed to estimate which vector from the manifold, the steering vector, corresponds to which eigenvector, and therefore estimating the signal's angle of arrival.

A significant drawback of the MUSIC algorithm is that it is not possible to resolve correlated signals, as occurs when smart jammers are introduced in a military scenario or when multipath reflections arrive within the same time section as the signal under

investigation [Shan et al 1985]. The sensor to sensor noise must also be uncorrelated or treated as another source to resolve [Philips 2006]. This has led to many variations and extensions to the MUSIC algorithm to improve its performance with correlated signals [Chen et al 2002a]. The aim of these modifications is to decorrelate the signals prior to applying the MUSIC algorithm.

Decorrelation of the sensor signals can be performed with spatial smoothing, as proposed by Shan et al [1985], an extension to work presented by Evans et al [1981]. The technique works with a uniform linear array and employs sub arrays which overlap to form the full array. The outputs of the sub arrays' covariance matrices are averaged, thus the coherence averages to zero, and leaves a smoothed covariance matrix which can be processed by MUSIC to resolve the signals' directions of arrival. It will work if all signals are coherent, but at the cost of doubling the number of sensors required because there must be as many sub arrays as there are signals. An advancement was proposed by Pillai and Kwon [1989] where the sub array is processed twice, forward (as done by Shan et al) and backwards (making use of the complex conjugated backward subarrays), with the results averaged to decorrelate the signals. This version can work when all the received signals are coherent and reduces the number of sensors by 25% in comparison to Shan's original proposal, but again is limited to a uniform linear array. If K correlated signals arrive at the array, 3K/2 sensors are needed using forward/backward spatial smoothing. Chen [1997] proposed a generalized two dimensional spatial smoothing arrangement for a uniform planar array; however in order to decorrelate 8 coherent signals 94 sensors were required.

6.5.4 IMP

The incremental multi-parameter (IMP) algorithm uses the conventional beamformer as its base. The idea was first proposed by Clarke [1989], with further developments of the IMP algorithm, published by Mather [1990] and Clarke [1992]. It has the advantage of not requiring a prior knowledge of the number of signals, or their inter-signal correlation. The algorithm starts with a model containing one signal; its DOA is estimated with a conventional beamformer. The identified DOA is then nulled out, leaving a residue, which is tested. If the residue is greater than a pre-determined level it is assumed that another signal is present. Keeping the first null in place the conventional beamformer estimates the DOA of the second signal. These DOAs are iteratively refined until the estimated angle stops changing, meaning than the optimal combination has been detected. Both signals are then nulled out and the residue tested; if the residue is below the threshold all signals must have been detected, if not the algorithm proceeds with a third signal and repeats until all signals have been identified.

The solution therefore rapidly converges on the maximum likelihood solution, but is significantly more computationally efficient than a full ML search because the number of signals are iteratively increased and do not need to be known in advance.

6.5.5 ASPECT

The adaptive Signal Parameter Estimation and Classification Technique (ASPECT) was developed by Manikas and Turner [1991]. This is a signal subspace solution which is similar to MUSIC, but is claimed to work with coherent signals. In MUSIC

if two signals are uncorrelated the eigendecomposition of the covariance matrix will produce two eigenvectors, one per signal. If the signals are instead coherent, one eigenvector will span the signal subspace, representing a combination of the two signals' steering vectors. ASPECT works on the principle that if the dimensionality of the array manifold is large enough then there is a unique linear combination of steering vectors which will equal the eigenvector identified from the covariance matrix. ASPECT takes a guess at how many signals are present and searches for linear combinations of steering vectors which match the eigenvectors. It uses one of three different cost functions to select the best combination of steering vectors.

6.5.6 Comparison and selection

The most researched and published algorithm for high or super resolution direction finding is MUSIC. However because of its inability to work with correlated and coherent signals it needs to be combined with spatial smoothing to work with the correlated signals arriving at the array. Spatial smoothing increases the size of the array, relative to the number of signals which can be resolved. The response of the room to be analysed has many close and overlapping signals (as described in section 5.1), but to minimise overlapping, the duration of the signals needs to be short. To excite all sensors in a spatially smoothed array the transmit signal's duration and time section length must be increased, increasing the number of signals arriving at the array. It is likely that the number of signals would then become greater than the number of sensors in the sub array, meaning that the directions could not be resolved. Spatial smoothing and MUSIC were therefore judged to be unsuitable in this application.

A maximum likelihood solution could be selected that would succeed with the coherent signals, but at the cost of high computational load, making it unattractive. ASPECT is claimed to have a very good ability to resolve close coherent signals, with Manikas et al [1997] claiming it to be superior to IMP. However the number of signals has to be known or at least estimated to be larger than the number of signals present. This author believes that selecting the maximum number of signals that can be detected (number of sensors -1, for example 23 sensors and therefore 22 signals) is a safe guess, because it would guarantee the solution, but greatly increase the computational load. This is because the linear combinations of 22 steering vectors over the entire array's manifold would have to be tested, even if only one signal was present. Prior estimation of the number of signals is therefore advisable. Another disadvantage of the ASPECT algorithm is its limited published performance. One of the algorithm's original authors, Manikas, was the author of the 1997 comparison paper. The only other reference to this method is a paper by Qi et al [2004] which only presents results for the uncorrelated signal case.

IMP has the advantage that its performance approaches that of a maximum likelihood solution, but because the number of signals is iteratively increased the solution is computationally efficient. Its ability to work with an unknown number of signals, uncorrelated, correlated, or coherent makes it ideal for this solution. An added advantage is that IMP is able to estimate the signal's amplitude at the same time as the DOA. This is useful in this application when identifying false targets. There have been a number of DOA algorithms which rely on iterative null steering that are very

similar to IMP which are named differently by their authors, for example DOSE developed by Zatman et al [1993] and IDOA by Morrison et al [2000].

Warrington has undertaken research into high frequency sky wave propagation which has resulted in a number of papers in which he estimated the direction of arrival of the sky waves using both the MUSIC algorithm and "an iterative null steering algorithm based on IMP and DOSE algorithms". [Warrington 1995a and 1995b and Warrington and Moyle 1995c]. Subsequent papers were published progressing the topic, stating that an iterative null steering algorithm was used, but without details of the implementation. Warrington found that iterative null steering was superior to MUSIC, claiming it was more robust, required no prior knowledge of the number of signals and no pre-processing for coherent signals. It was also observed that the null steering algorithm was far better suited to short time snapshots of data than MUSIC, with a possible cause being that the correlation of signals appeared greater for small sample numbers. With the null steering algorithm a second target was often correctly identified which was not identified by MUSIC. An important aspect of this work is that it involved practical measurements instead of relying on simulations alone, which is a rarity in papers discussing super resolution direction of arrival algorithms. (A unique feature of this PhD thesis is the reverberation analysis to estimate direction of arrival).

Because of its proven application by several authors, its lower complexity and reduced computational load the IMP algorithm was selected over the ASPECT algorithm.

6.5.7 IMP implementation details

The IMP algorithm, as described by Mather and Clarke is shown in Figure 31. The conventional beamformer, as described in section 6.5.1 above, is a basic component of IMP. The IMP algorithm starts with one signal assumed, and estimates the DOA of the strongest signal with the conventional beamformer by identifying the maxima of $P(\theta)$, the spatial power spectrum

$$P(\Theta_i) = \frac{m^H(\Theta_i)\, R_{est}^H\, m(\Theta_i)}{m^H(\Theta_i)\, m(\Theta_i)}$$

where i=1 to the length of the array manifold, $m^H(\theta_i)$ is steering vector I, taken from the array manifold, R_{est} is an estimate of the covariance matrix from the sampled sensor output signals and H is the Hermitian transpose. The derivation is shown in section 6.5.1.

Once a signal is identified it needs to be removed so that other signals can be detected. The null matrix Q performs this task by projecting the identified signal into a subspace orthogonal to that spanned by the array matrix [Shannon 1998].

$$Q = I - M_s(M_s^H M_s)^{-1} M_s^H$$

where I in an identity matrix and M_s is a matrix containing the steering vectors of the identified signals to be nulled. For the first null this is the steering vector that produced a maximum in the conventional beamformer's power spectrum. The new search is performed with the modified conventional beamformer, below.

$$P(\Theta_i) = \frac{m^H(\Theta_i)\, Q\, R_{est}^H\, Q\, m(\Theta_i)}{m^H(\Theta_i)\, Q\, m(\Theta_i)}$$

By nulling out all but one signal the DOA of the remaining signal can be refined; this refinement is required because the presence of sidelobe leakage from the second

signal (as yet uncancelled) will have a bias on the first signal's direction estimate. The error then affects the null, resulting in a residual leakage of the other signal and a bias on the adapted spectrum. This bias is iteratively reduced by the re-estimation of the DOAs, improving the accuracy of the nulls and thus reducing the bias from sidelobes and leakage. When all the DOAs stop moving the bias is minimised and all signals are nulled out so the residue level can be checked. If there is still significant power left then another target is added and the search continues; otherwise it is assumed that all signals are now identified and the search stops. An example of the IMP algorithm, where the effect of the null can be seen is shown in section 6.5.7.6.1.

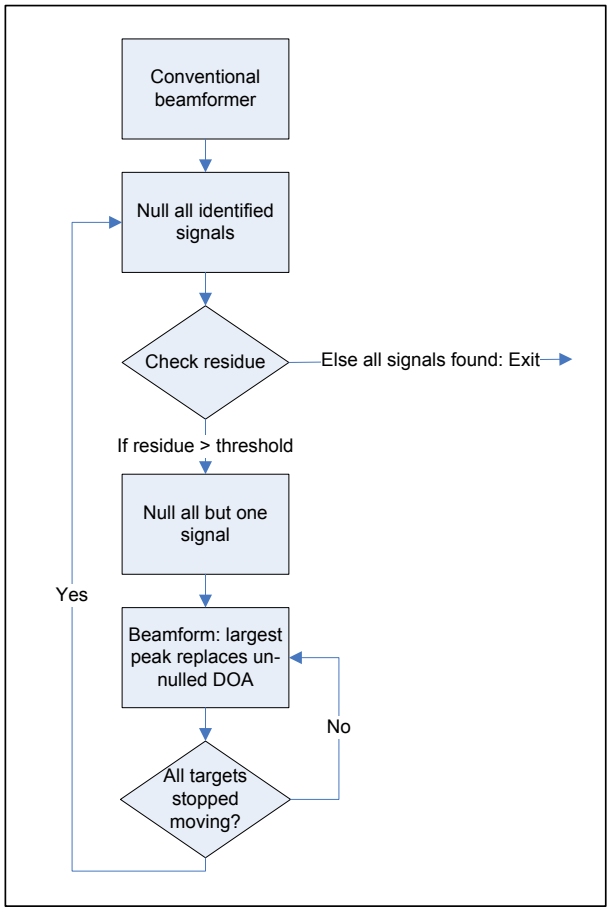


Figure 31: IMP algorithm flowchart.

A number of developments and refinements to the IMP algorithm were made. Some of these were a result of research conducted by this author, working in collaboration with Gavin Philips [2006] from within the same research group who was also using the IMP algorithm in his research.

6.5.7.1 Manifold calculations

A simulated manifold was used for both simulated room experiments and real room experiments; it was not possible to use a measured manifold, as detailed in section 6.6. The simulated manifold was calculated by simulating a source at a set distance from the simulated array and calculating the received signal's phase: if the source was a long way from the array the simulated wavefront would be a planewave, however if the source was close enough to the array the wavefront would be curved. (Refer to section 6.1 for definitions of nearfield and farfield sources). To select the correct distance, simulations were performed of a room with a source and array which were processed with the temporal processing described in Chapter 5 and two variations of the spatial processing described here. The first variation involved a manifold source to array distance at a constant 10m, while the second involved calculating the manifold each time, using the results of the temporal processing to set the manifold's source to array distance to be the same as the distance associated with that time section. Thus the second method eliminated any mismatch between the manifold and the received signals at the array caused by wavefront curvature. The conclusions were that the manifold, as calculated at a set 10m, performed only slightly worse for sources very close to the array and the same for far away sources, therefore the

additional computation involved was not considered worthwhile and the constant 10m manifold was used.

The manifold was calculated over an azimuth span of -90˚ to +90˚ for the line array and -90˚ to +90˚ for the two dimensional scan used for the circular array, both azimuth and elevation. Warrington [1995b] proposed using an interpolation of the manifold, between the estimated stepped values, to calculate a more accurate nulling vector. For a uniform linear array the interpolation can be linear [Philips 2006], however interpolation has the potential to introduce errors; therefore, the manifold was calculated at a resolution of 0.05˚. This direct approach is made possible by the affordability of computer RAM and the efficient processing arrangement, as described in section 6.5.7.3.

Because of the way the array was being processed, as two sub arrays, the azimuth angle was estimated using the line array. For the circular array, the complete two dimensional manifold was calculated, but strips containing only these azimuth angles matching the line array's results were selected, resulting in a two dimensional manifold consisting of x separate elevation strips, where x is the number of signals detected by the line array.

6.5.7.2 Sub array considerations

Tests found that if, for example, the line array identified two angles, A and B, meaning at least two targets were detected, sometimes the circular array would estimate two elevation angles for A degrees and nothing for B degrees, and IMP

would finish. The source at azimuth angle B would disappear. This was clearly not correct. As described in section 6.4, above where array geometry and array directionality were considered, the line array is optimal in one dimension and the circle array can suffer from a biased directionality pattern. Therefore the line array can be trusted more than the circle, and if it identifies two azimuth angles both must be kept.

To solve this problem the two dimensional IMP was modified. In the initial conventional beamformer stage of IMP, where no nulls are present, IMP selects the peak of the spectrum as the first DOA. For 2D IMP this stage was modified so that for each of the azimuth angles identified an elevation angle was estimated. To guarantee this happened the conventional beamformer was calculated for a temporary manifold, initially with all elevation strips at the defined azimuth angles. An elevation was identified for a particular azimuth from the maxima of the spectrum, as normal, but once identified that azimuth angle's strip was removed from the temporary manifold and the conventional beamformer repeated on the remaining 2D manifold strips, resulting in the same number of targets as there were azimuth angles, each with an elevation angle.

During the IMP iterations where the DOAs are refined, the spectrum was calculated with all but one signal nulled, the manifold used for this calculation only containing the azimuth angle of the target being refined, i.e. not nulled. Again this ensured that the azimuth angle was definitely preserved.

The test for more targets used the spectrum with all nulls present, calculated with the entire manifold. If a new target was added, i.e. one azimuth angle had multiple targets at different elevations, the maxima of this full search was used.

6.5.7.3 IMP computational efficiencies:

As noted by Warrington [1995b], imperfections in the steering vector used to calculate the null not only have the effect of introducing direction errors, but more significantly will reduce the dynamic range. To ensure that the null is correctly calculated the correct steering vector for the null must be selected. One solution is to have a manifold with very small steps, increasing the probability that the arriving signal's spatial response will exactly match one of the manifold vectors, or at least minimise the difference between them. While this will greatly increase the processing time it is, however, only the peak of the power spectrum which is of importance within the IMP algorithm when selecting the DOA associated with that signal and therefore it is unnecessary (and computationally wasteful) to calculate the entire spectrum at fine resolution.

The approach taken was therefore to perform the beamforming algorithm in two stages, firstly at a course resolution and then around the area of interest at a finer angular resolution. The course scan had to be performed at sufficient resolution to guarantee that the peak was not missed. The manifold angles either side of the maximum from the coarse scan were used as the bounds of the fine scan. The ratio of coarse to fine manifolds clearly affects the computational load, such that with a 1D manifold at 0.25˚ resolution, a full search would involve 721 vectors. If a coarse

manifold is sampled at 2˚ intervals there are 91 course vectors. If a source is detected at say -30˚ the fine manifold must cover the angles from -32˚ to -28˚, at full resolution of 0.25˚, so 17 fine vectors must be used to compute the fine spectrum around the peak, resulting in a total of 108 vectors to be used which, in comparison with the original figure of 721 vectors, results in the spectrum being calculated at over 6 time faster.

Algebraically, if the full resolution manifold contains L steering vectors and only every r vectors are used in the coarse manifold there are $\frac{L}{r}$ coarse vectors. The number of vectors used in the fine manifold section is $2r$. Therefore the total number of vectors, T is:

$$T = 2r + \frac{L}{r}$$

The most efficient ratio is when the total number of vectors is minimised, which can be calculated by finding the zero of the differential of T:

$$\frac{dT}{dr} = 2 - \frac{L}{r^2}$$

$$0 = 2 - \frac{L}{r^2}$$

Solving for r gives:

$$r^2 = \frac{L}{2}$$

$$r = \sqrt{\frac{L}{2}}$$

For the example above, r = 19, giving T = 76, which is over 9 times faster than the full manifold scan. The prerequisite that the coarse manifold must be performed at a

resolution high enough to guarantee not missing a peak is not included in this calculation. The width of the peak will vary with array configuration and the number of sensors; however if L is sufficiently large then the coarse manifold will be at a high enough resolution to not miss a peak and therefore, a target. This needed to be checked for the particular array and manifold under investigation, but as a minimum, it was understood that there needed to be more coarse vectors than the number of sensors.

In this research the manifold was from -90˚ to 90˚ at a resolution of 0.05˚, so L = 3601. Therefore the optimal ratio was 42.432, rounding down to a ratio of 42 results in a step size of 2.1˚ and T = 171, which is 21 times faster than the full manifold scan.

6.5.7.4 IMP: When to add another signal and sources of error.

The IMP algorithm starts with one target and then tests to see if another signal is present and if so increases the model order by one. When the DOAs for that order stop changing this test is repeated and the model order increased until all targets are identified. The decision on whether to add another order to the model is taken by nulling out all identified signals and computing the spectrum, if there are significant peaks another target is present. A prerequisite is that the model order, the number of targets, must be one less than the number of sensors for the DOAs to be resolvable. The criteria of 'significant peaks' can be tested using two methods; firstly a flatness test and secondly a threshold test. Referring to Figure 32, a spectrum with 2 signals successfully nulled out, it can be seen that the residue is not flat, but has a low amplitude peak. Using the flatness method this could be interpreted as another signal,

i.e. a target, still to be identified, but in reality it is simply a result of imperfect nulls. An amplitude method, based on a threshold test would not see this as significant, so would not continue searching and was therefore considered more reliable and less complex to implement.

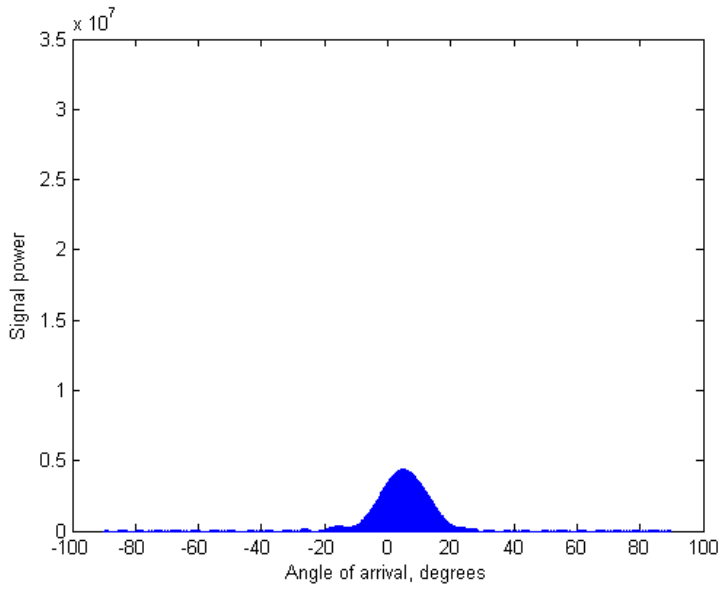


Figure 32: IMP residue; from, section 6.5.7.6.1 with amplitude zoomed in by a factor of 10

The setting of the threshold is crucial to correctly estimating the model order, and therefore the number of targets to estimate; if the threshold is too high a real target might be missed, if too low a noise spike might be falsely detected as a target, therefore a combined criterion was used where all of the following conditions had to be met before another target was added.

1. The number of targets had to be less than the number of sensors.
2. The maximum of the nulled spectrum had to be greater than the noise level. The recordings started at the time the sound was transmitted from the loudspeaker (real or simulated). In the time before this signal arrives at the

array (identified by the temporal processing) the noise of the room and system are being recorded. This noise recording was analysed, for each experiment, using the conventional beamformer, in the same way as the DOA time sections, to measure the noise level, including processing gain added by the matched filtering and array processing. The peak of these levels was used for the noise threshold.

3. The maximum of the nulled spectrum had to be less than the last maximum, i.e. the last iteration must be an improvement: this checks that the algorithm has not got lost and the DOAs accuracy decreased.
4. The new DOA, selected from the maximum of the fully nulled spectrum, had to be more than 1° away from an existing DOA's azimuth or elevation. This prevents another signal being added where a null has not successfully nulled out all of the energy from an identified signal. The threshold of 1° was chosen empirically from experiments.
5. The maximum of the nulled spectrum had to be greater than 5% of the non nulled, conventional spectrum maximum for this time section, i.e. there had been a significant reduction in the residue level [Mather 1990]. The 5% level was set after experimentation and fine tuning of the algorithm's parameters.

The reason that the final criterion was needed, rather than just considering the noise, is that in this application noise was introduced by the short time sections involving overlapping signals. For example, consider a time section which contained three different signals, which were shorter than the time section and all fully arriving early enough in the time section that they had finished before the end of the time section.

Assuming the angles of arrival were within the manifold and exactly matched it, there would be three steering vectors which ideally matched the DOAs of the three signals. Now consider one of the signals arrived later in time, so within the time section it only excited half of the sensors in the array. There would be no steering vector which matched the phase of the signal. The ideal vector would be one where sensors which were excited by the signal would have an element to element phase difference, (corresponding to the signal's DOA), and where those sensors which were not excited would have no phase difference corresponding to that signal. Because the manifold did not contain a matching steering vector the effective noise level would increase. The situation of overlapping signals is unavoidable in analysing the room's reflections, particularly later in time when the signals arrive closer together. Time was not available to further investigate this problem, though it would be interesting to experiment with manifolds containing partly excited sensors arrays to see if improvements could outweigh the dramatic increase in computational load caused by an increased manifold size.

Another source of errors affecting the spectrum is that *R_est*, the covariance matrix used is only an estimate. When the number of samples is small, as in this thesis, inaccuracies in *R_est* increase, resulting in a null which is not perfectly narrow. This theoretically limits the accuracy in resolving very close signal DOAs but cannot be avoided because small time sections are required to minimise overlap of signals, a greater source of inaccuracies in the system.

6.5.7.5 IMP post-processing

One of the limitations of IMP is that when there are multiple signals present a DOA is not accurately estimated until all other DOAs are accurately estimated. Over several iterations the DOAs will be refined as the nulls become more accurate, but it is only the last iteration when the nulls are the best they can be that the DOAs are therefore as accurate as they can be. Brandwood [1989] described an algorithm he called Iterative Scanned Nulling Beam (ISNB), a variant of IMP. He added a stage within the main loop in which the selected DOAs were checked and corrected after increasing the model order. It is considered that this checking and refinement is only required once the DOAs have settled to their final values.

There were two stages to the DOA verification implemented, firstly to test if the DOA contributed significantly to the spectrum by measuring the effect of the null on the residue spectrum and secondly to re-evaluate the DOAs to test for any movement. To test if a signal contributes significantly to the spectrum all signals are nulled out and the maximum measured. This is compared to the spectrum with all signals nulled except the one being verified. If the signal does significantly contribute to the spectrum a large peak will be present, but if the peak is small in amplitude the signal is likely to be noise, falsely detected, and is removed. Like the threshold for introducing a new target, the level at which a target is considered significant has a big impact on the results, but is too complex to compute quantitatively. It was empirically chosen through simulated experiments where the before and after DOAs to known signals were compared; the signal was considered significant if the maximum was 40% greater than the fully nulled maximum.

The DOAs were validated as per the main IMP loop, with all but one signal nulled out at a time and the new maximum taken as the best DOA. This was a repetition of the last iteration of IMP. It was rare that a signal would be removed as a result of the above test or that any of the DOAs would change by more than a fraction of a degree, but occasionally it did improve the results by removing a false signal and refining the DOAs more accurately.

6.5.7.6 IMP examples

The algorithm was developed in a test bed program where a number of known signals could be generated and the IMP algorithm used to estimate their angles of arrival. This allowed for quick verification and development. Once working well, the algorithm was tested with a simulated room and real room. The signals to be analysed for a room response were considerably more complex with varying numbers of signals and overlaps per time section. The algorithm was further refined to work well with the room analysis, particularly when adjusting the threshold levels.

Common parameters used in all experiments were as follows:

- Sample frequency = 96kHz
- Tx signal was a LFM chirp of matched filtered length equal to twice the effective distance of the line array.
- Speed of sound = $345ms^{-1}$

6.5.7.6.1 Example A

The first example, A, detailed in Table 3, is for two fully coherent signals arriving at the array at exactly the same time.

Array:	**Line sensors=15** **Therefore section length = 65 samples**		**Circle sensors= 24**	**Manifold step size = 0.03˚**
Signal 1:	Distance from array = 5m	Azimuth angle = -5˚	Elevation angle = -15˚	Centre frequency = 10474Hz
Signal 2:	Distance from array = 5m	Azimuth angle = 10˚	Elevation angle = 20˚	Centre frequency = 10474Hz

Table 3: IMP example A setup parameters

First considering only the line array: Figure 33 shows the spectrum resulting from the conventional beamformer, from which it can be seen that there are two distinct peaks, one for each source. Figure 34 shows the response with the first null in place, as selected by IMP, whilst Figure 35 shows the response once IMP has finished and both signals have been nulled. The DOAs estimated were -4.98˚ and 9.99˚

Now considering the circle array: Figure 36 shows the conventional beamformer spectrum; two peaks can again be seen, but there are much higher sidelobe levels. Figure 37 shows the residue with both signals identified. The final DOA estimates were azimuth -4.98˚, elevation -14.91 and azimuth 9.99, elevation 19.92. It can be seen that there is a slight error in the DOA estimates, the line array being more accurate, as expected.

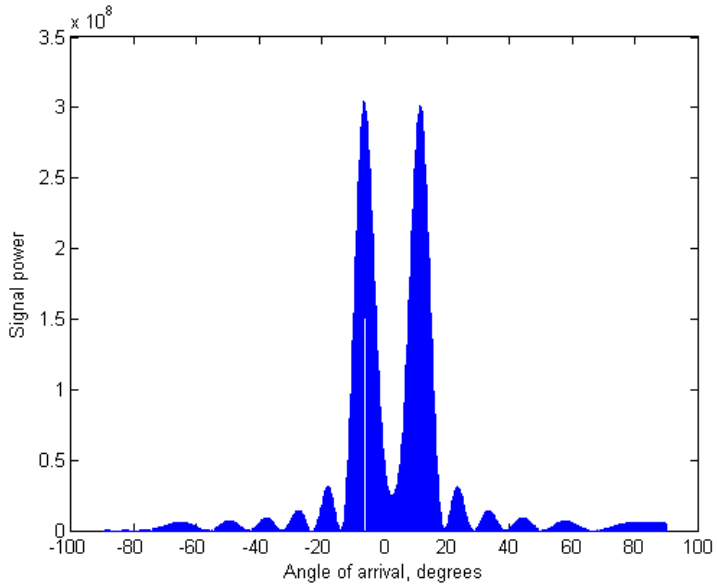


Figure 33: Example A conventional beamformer for line array, azimuth angle against amplitude.

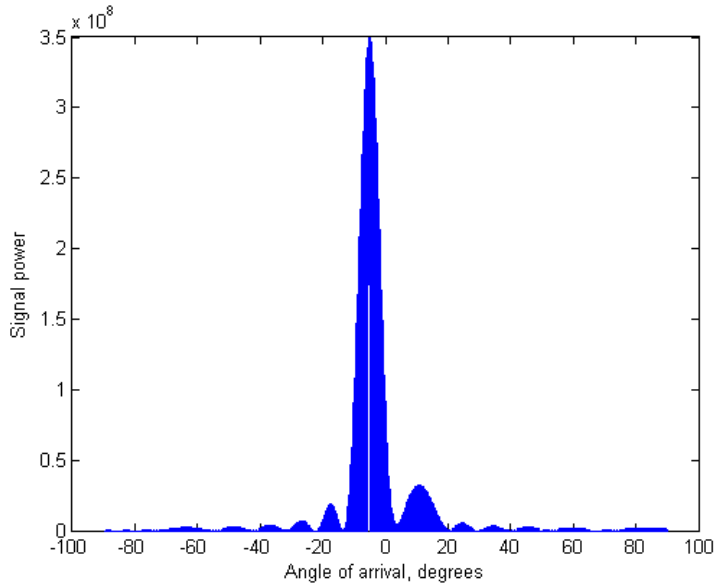


Figure 34: Example A: Line array, first null in place, azimuth angle against amplitude.

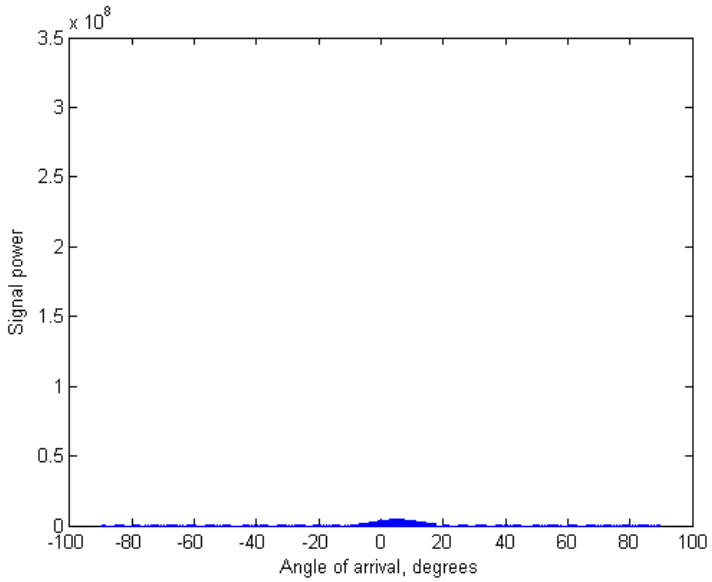


Figure 35: Example A: Line array, residue, both signals nulled, azimuth angle against amplitude.

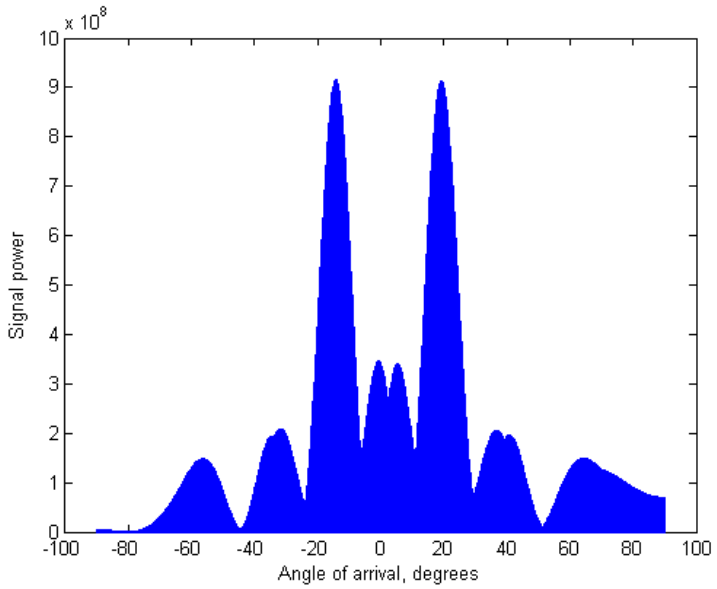


Figure 36: Example A conventional beamformer for circle array, elevation angle against amplitude.

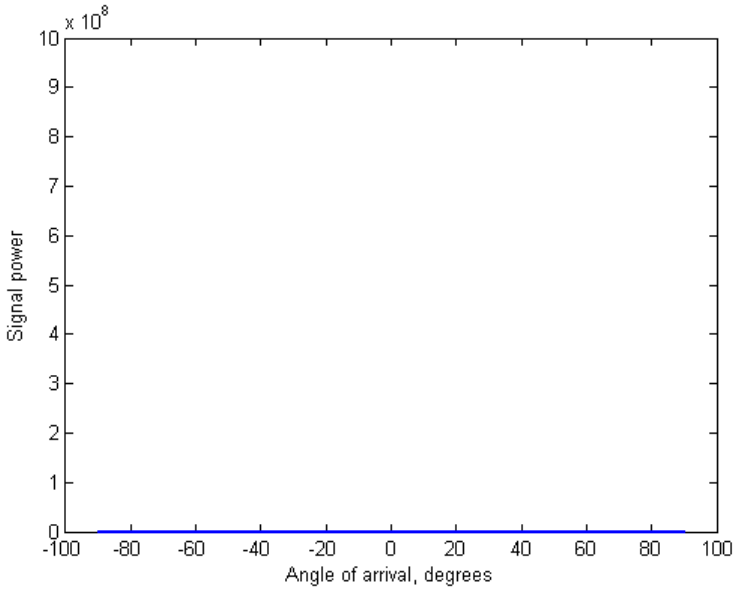


Figure 37: Example A: Circle array, residue, both signals nulled, elevation angle against amplitude.

6.5.7.6.2 Example B

This example is the same as example A, but with the sources closer together. The conventional beamformer output for the line array is shown in Figure 38.

Array:	**Line sensors=15** **Therefore section length = 65 samples**		**Circle sensors= 24**	**Manifold step size = 0.03˚**
Signal 1:	Distance from array = 5m	Azimuth angle = 18˚	Elevation angle = 15˚	Centre frequency = 10474Hz
Signal 2:	Distance from array = 5m	Azimuth angle = 10˚	Elevation angle = 20˚	Centre frequency = 10474Hz

Table 4: IMP example B setup parameters

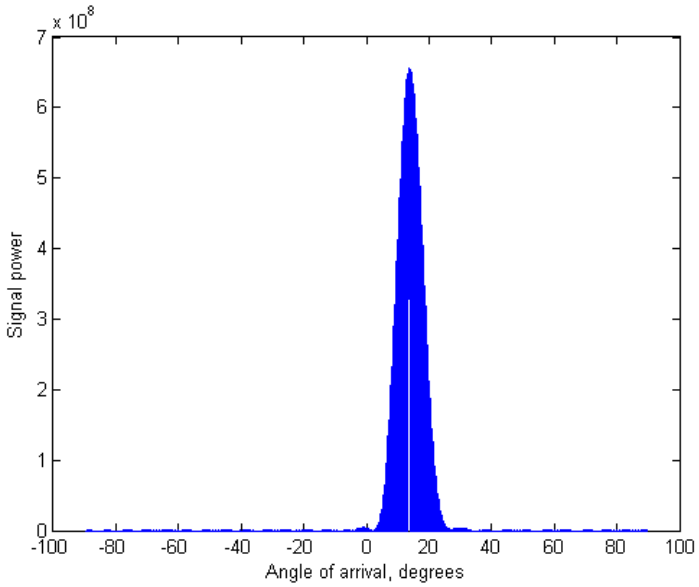


Figure 38: Example B conventional beamformer for line array, azimuth angle against amplitude.

The final DOA estimates were azimuth 10.32˚, elevation 19.74 and azimuth 18.42, elevation 14.97. It can be seen that IMP is able to resolve two closely spaced sources which would not be recoverable with a conventional beamformer. However resolving closer spaced signals was not possible with 15 line array sensors; this was because IMP was unaware that a second source was present when processing the line array; the logic for adding another signal had failed. It would have been easy to adjust the thresholds so that it was successful, but using too low a threshold with room response analysis caused many instances of a target to be detected as multiple targets, close together.

6.5.7.6.3 Example C

The more signals that arrive in a time section, or the closer they are together the greater the number of sensors required to accurately resolve them. This is demonstrated in the next example where four signals arrive the array at exactly the same time. The circle consisted of 24 sensors and the manifold step size was 0.03˚ for each experiment.

	Distance from array	**Azimuth angle**	**Elevation angle**	**Centre frequency**
Signal 1:	5m	10˚	-35˚	10474Hz
Signal 2:	5m	10˚	20˚	10474Hz
Signal 3:	5m	-66˚	-26˚	10474Hz
Signal 4:	5m	-45˚	-45˚	10474Hz

Table 5: IMP example C setup parameters

Line sensors : section samples	**Signal 1**		**Signal 2**		**Signal 3**		**Signal 4**		**Signal 5**	
	Az	**El**	**Az**	**El**	**El**	**El**	**Az**	**El**	**Az**	**El**
9:37	9.18	-35.43	9.18	19.86	-53.16	-90.00	9.18	-7.62		
11:47	10.65	-35.43	10.65	19.26	-52.98	-85.89	10.65	-6.93		
13:56	10.05	-35.25	10.05	20.40	-66.96	-26.16	-45.42	-46.11	10.05	-7.20
15:65	9.99	-35.07	9.99	19.77	-47.01	-57.24				
17:73	10.02	-35.61	10.02	19.80	-44.40	-0.72	-44.40	-54.06		
19:82	9.99	-34.92	9.99	20.04	-65.25	-24.87	-45.54	-45.72		
21:89	9.99	-34.86	9.99	19.98	-65.04	-24.30	-45.57	-45.72		
23:97	10.02	-34.86	10.02	20.04	-65.86	-24.51	-45.63	-45.84		

Table 6: IMP example C results

The results in Table 6 demonstrate that IMP is able to resolve coherent signals, if there are enough sensors. However if there were not enough sensors an extra signal might be falsely detected, as occurred in Table 6 with 13 line array sensors and Table 7 with 9 line array sensors. There are two reasons why the accuracy improves with more sensors, firstly because the mainlobe of the beam becomes narrower, increasing resolution and secondly because the time section length increases, making the covariance matrix estimate, *R_est,* more accurate. Inaccuracies in the covariance matrix make the null less accurate which can theoretically degrade the performance of IMP.

Another observation from these results is that the resolution is greatest at broadside to the array and degrades as the angle of arrival approaches endfire to the array. This is because of the mainlobe the beam broadens towards endfire.

To investigate whether the short time section length used (to reduce signal overlap) was limiting the resolution the experiments were conducted again, but this time with a section duration of 1000 samples for all array lengths. These results show that the section duration was having only a small effect on the DOA estimates; the accuracy was slightly increased, but the four signals were only resolved with nineteen line sensors, the same as with the short time sections. One observation is that for eleven and fifteen line sensors, with the longer time sections, only two signals were detected. This was due to the threshold levels being optimised for the shorter time sections and the compromise present in the covariance matrix estimate. It was therefore concluded

that the time section duration was not limiting the DOA processing, despite its theoretical ability to do so.

Line sensors : section samples	Signal 1		Signal 2		Signal 3		Signal 4		Signal 5	
	Az	El	Az	El	El	El	Az	El	Az	El
9:1000	9.24	-34.23	9.24	18.93	-54.09	-60.06	-54.09	-17.16	9.24	-72.63
11:1000	10.68	-35.64	10.68	18.93						
13:1000	9.99	-34.98	9.99	19.95	-65.82	-25.82	-45.94	-44.91		
15:1000	10.35	-35.82	1035	18.63						
17:1000	9.96	-35.70	9.99	19.74	-43.47	-1.98	-43.47	-53.13		
19:1000	9.99	-34.95	9.99	19.95	-65.97	-25.56	-45.00	-45.00		
21:1000	9.99	-34.95	9.99	19.95	-65.94	-25.53	-45.00	-45.00		
23:1000	9.99	-34.95	9.99	19.95	-65.94	-25.53	-45.03	-45.00		

Table 7: IMP example C results

6.6 Array calibration

A large amount of the research presented on the topic of direction finding solutions has been simulation based. This is often due to the large amount of equipment needed to create and record an array of many sensors, but is also due to the complications of real world variations and the degradation in the results which these produce. These variations can be corrected for if the array is accurately calibrated. Fistas and Manika [1994] describe array calibration as a prerequisite in signal subspace techniques, for example MUSIC, whilst Warrington [1995a] acknowledges that with null steering algorithms the results are markedly degraded if errors in the steering vectors result in an inability to fully null out the incoming energy.

6.6.1 Sources of error

Clearly whilst in a simulation it is possible to simulate a perfect array, in reality a real world array will deviate from ideal. The sensors will not be identical, and in the case of affordable electret microphones used in the experiments for this thesis, there will be tolerances in the sensors' sensitivity due to manufacturing compromises. Dobbins [1984] conducted some experiments to measure the tolerances of electret microphones and found that above 10kHz amplitude the maximum rms variation was 3dB, but the phase variations were less than 10 degrees. (The phase variation between 50Hz and 20kHz was quoted as 10 degrees: from the graph he presented it decreases significantly at high frequencies, probably to less than 5 degrees, but no figure is given.) However these experiments only used a sample of three microphones and were only on-axis measurements. The microphones used in this thesis were supposed to be omnidirectional, but no data for expected angular variations could be found. Dobbins also summarises work conducted by Hruska et al [1977] who discovered that for electret microphones the short term sensitivity varied by ±1dB, but long term (16 weeks) variations increased to ±1.5dB.

Additionally when these sensors are placed in an array the placement will not be exactly as per the design. In this project placement errors were minimised by using a PCB to mount the sensors where the tolerances in hole placement for the pins would be very high. These errors would be constant and would manifest themselves as gain and phase errors.

A real array will also suffer from mutual coupling of signals, caused by a number of interactions, including the vibrations which are conducted through one sensor, to the mount and up to the next sensor. Mutual coupling also occurs when a wavefront hits the sensor, and reflects off its body, some of which will then be received by other sensors slightly later in time. These multipaths will vary with the wavefront arriving at the array and therefore will not be constant and calibration experiments may not accurately produce a set of correction values which apply to a given setup. It is also worth noting that any calibration experiments will exhibit mutual coupling making the error measurements subject to the experimental signals themselves. Mutual coupling was minimised with the array produced by baffling the array in soft foam and adding rubber pads between each microphone and the PCB. The height was adjusted to make the surface as smooth as possible, and its snug fit around the sensors would absorb some of the vibrations induced into the sensor's body whilst the smooth surface and absorbent nature of the foam would reduce multipath coupling. For reference, Figure 29 is a photograph of the array with the baffle.

6.6.2 Calibration methods

There are two methods of array calibration, parameter based error source estimation and direct measurement. Parameter methods involve taking a relatively small number of measurements, with the source position nominally known, but not exactly. Different methods can be used to estimate the error parameters. Early methods, such as the one described by Levi and Messer [1990], identified sensor location errors only. Later solutions, as proposed by See [1995] for example, increased the model order to include mutual coupling and channel mismatch errors. Both methods resulted in a

multidimensional nonlinear problem. The advantage of the parametric methods are that they reduce the number of measurements required, in comparison to a direct manifold measurement, this comes, however, at a high computational cost and the calibration results are poorer than a direct search [Liu et al 1996]. This could be due to errors in the parameter estimation or variations in the individual sensor's response with respect to angle of arrival. They are assumed to be omnidirectional, but this might not be the case, even if claimed by the manufacturer. At high frequencies, as under investigation here, the wavelength of the received signal is small compared to the microphone's diameter, therefore the pressure on the diaphragm surface will not be equal at all areas of the surface, as it would at low frequencies. This will cause phase cancellations to occur, resulting in a polar response where the sensitivity of the microphone decreases at high angles and the output signal will change phase [Eargle 2001]. This cannot be quantified in the parametric method; only an average sensor phase and gain error are estimated, therefore it was decided to try and directly measure the array manifold at regular intervals and interpolate between them to create the high resolution manifold required.

6.6.3 Experimental setup and results

Array calibration experiments were undertaken in an anechoic chamber, with the array placed on a rotating table and a loudspeaker placed on an independently rotating arm. The rotating table and arm were driven by PC controlled stepper motors so an automated recording process could be used to record the array's response over a full 180 degree hemisphere. The distance from the loudspeaker to the array was 1.8m, the maximum possible in the anechoic chamber with the required equipment. Due to the

long time it took to perform the recordings over the whole hemisphere it was decided to increment the rotation at 10˚ intervals, with interpolation used to estimate finer resolution. The arm started at 90˚ and the table was rotated from 0˚ through to 350˚, with recordings made at every 10˚ step, then the arm was moved to 80˚ and the process repeated until all angles were covered. Figure 39 shows an experiment in progress.

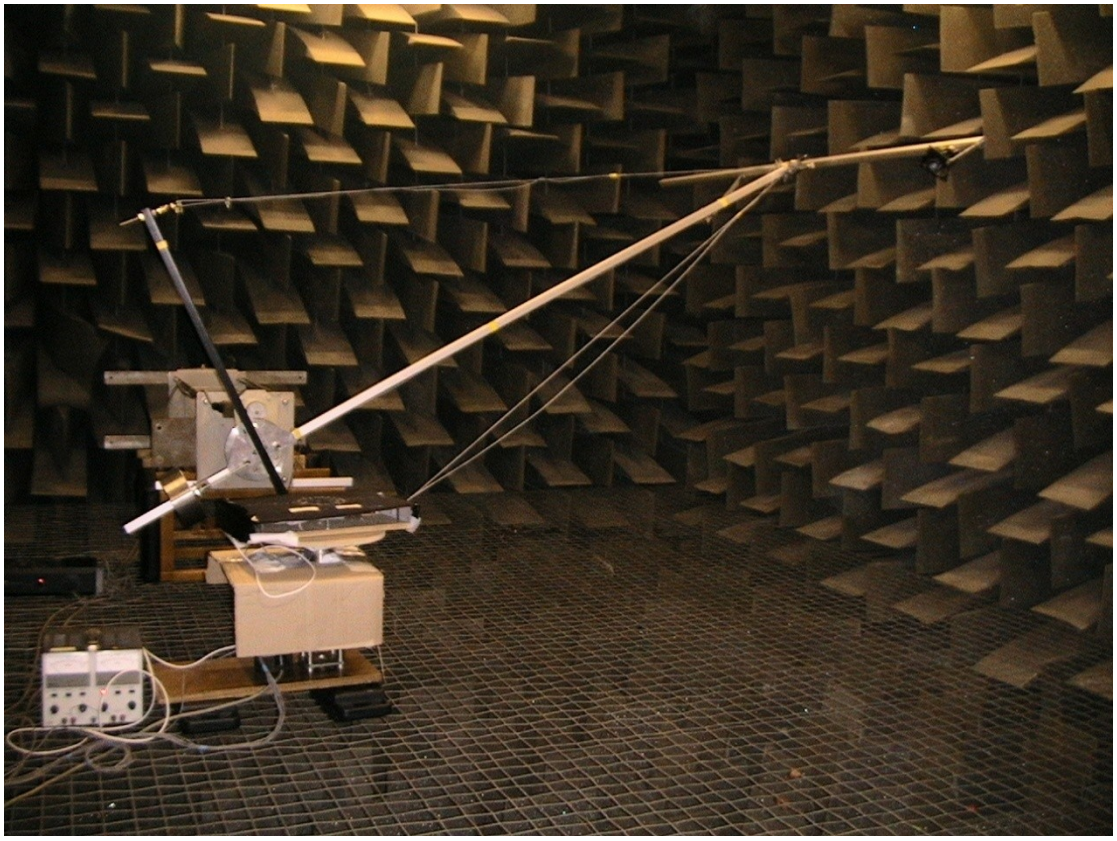


Figure 39: A photo of the array calibration experiments in progress.

These experiments were conducted before the LFM chirp was designed; therefore a short burst of 10940Hz sine wave was used with duration of 100ms and sample rate of 48kHz. A temporal window prevented a sudden turn on/off peak in the response. The noise was a significant problem (the reason why a LFM chirp was later adopted).

After matched filtering, the amplitude and phase were calculated using a DFT. The phase results for each microphone were plotted and compared to the expected phase. These showed that although great care had been taken in the setup the array centre was not exactly below the loudspeaker at 0 degrees azimuth and elevation and therefore the source positions were not exactly known for each recording. Due to the position and noise errors it was concluded that the calibration experiments had as many or greater phase errors than was expected in the array, and so there was no benefit to the measured manifold over a simulated manifold.

It is possible that repeating the experiments later with the LFM chirp would have reduced the noise to a level where the results would have been worthwhile; however the anechoic chamber was not available at the required time, due to building works in the department. The work conducted by Hruska et al [1977] demonstrated the variability of electret microphones over time, suggesting that calibration and room experiments needed to be performed close in time to make calibration valid. The above experiment had highlighted the sensitivity of calibration to other error sources that could not have been controlled, such as speed of sound variations caused by temperature and humidity, affecting the measured phase, or gain adjustment knobs on the recording equipment at different positions.

With the information gained it is expected that parametric calibration, with a few source positions taken in the room being measured might be worth additional investigation. However it was decided that time was limited and further work on array calibration had to be stopped so that the rest of the project could progress. An

alternative approach could have been to incorporate the hill climbing algorithm, as suggested by Moyle and Warrington [1997]. This algorithm does not try and calibrate the array, but in each iteration of the direction estimation algorithm the steering vector selected as the maximum of the spectrum is adjusted. The gain and phase of each sensor are adjusted by ±0.1dB and ±0.1˚ and the spectrum recalculated. The modified steering vector producing the maximum spectrum is used for the null in later iterations. The drawback of this approach is, however, that the steering vector selected might not actually be possible or it may not provide an accurate representation of the array's response.

Because it was not possible to calibrate the array to a high enough accuracy a simulated manifold had to be used, with the consequence of degrading the accuracy of the spatial estimates with measured data, compared to the ideal array used in the simulated room. As the direction of arrival estimation only requires phase information from the received signals, the simulated manifold was calculated from the phase of the signal from a simulated source, not the phase and amplitude. This meant that the amplitude variations of the microphones did not effect the direction of arrival estimates, reducing the affects of real world variations and therefore minimising the degradation that the measured estimates would suffer. Refer to section 6.5.7.1 for further details on the manifold calculations.

6.7 Results:

Throughout the development of the spatial processing the array designs and algorithms were tested first in a room simulator, its design being described in Chapter 4. Once the concept proved to be an improvement over previous iterations experiments were conducted in a real room. The real room used was varied to test for different scenarios and depended on room availability; they were seminar rooms within the department selected to be as close a match to a domestic room as possible. Simulations with the same setup as the real experiments were conducted to help understand and compensate for the differences between the real and simulated cases.

6.7.1 Real experimental setup

All experiments were automated by a Matlab script running on a PC with a RME audio interface card connected to three Behringer ADAT interfaces with analogue I/O. The test signals were created in Matlab and transmitted through the Behringer's analogue output to a powered KRK loudspeaker or a hi-fi amplifier and Linn drive unit. The received sound was synchronously recorded to hard disk for later processing. The simulations had used a sample rate of 96kHz to minimise the effect of the Lagrange fractional sample delay filter, used in the room simulation. Because the ADAT equipment available was limited to a maximum sample rate of 48kHz, the data was recorded at 48kHz and upsampled to 96kHz for processing. The eight balanced outputs from the microphone array were connected to the same Behringer interface. Analogue loopback tests had confirmed synchronisation and revealed that

the latency of the interface was constant at 68 samples for all channels in the first unit, and 69 and 70 samples for the second and third units' respectively. (The third unit was not used with the PCB array). Each test signal was transmitted eight times, with the microphone multiplex switches changed to select different microphones in the array, through an RS232 link between the PC, under Matlab control, and the microphone array's PIC microprocessor. Spot microphones at potential listener locations within the room were used to record the soundfield at those locations. These were to be used in the final system verification, so that predictions of the room's impulse response could be compared to the actual sound at different locations, (see Chapter 8). A combination of AKG and Behringer measurement microphones were used at these spot locations.

6.7.2 Performance measures

As described in section 6.4.7 there were 17 different array configurations investigated, comprising a different number of line sensors and either 16 or 24 sensors in the circle, along with different transmit frequencies. The position of the array and speaker were also varied to change the reflections under investigation. Therefore there was a large set of combinations to analyse and for each set many image positions were identified by the spatial processing. Full system verification required the three processing stages to be developed and joined together. Until this was ready it was decided to 'score' the estimations. The score was calculated using logic rules and measuring the difference between the actual source (original loudspeaker or image source) location and the estimated location, resulting from the temporal and spatial processing. This helped decide if a particular algorithm adjustment or setup parameter was beneficial because

results could quickly be compared; the more promising results were then investigated in greater detail. In the case of the simulations this was simple because the exact locations were known with certainty. For measured data the actual original source location was a direct measurement from the room boundaries, taken with a laser range finder, while the image locations were calculated from the geometry of the room. Although every effort was made in the measured experiments to ensure accuracy, there were many potential sources of error, as described in section 6.7.3.

For a configuration to be successful it was decided that the source and first order images must be identified with a small difference between the actual and estimated reflection origins, or images locations. This was a result of the research presented in Chapter 2 which highlighted that the room modes had a significant effect on what the listener heard, and their audible effect could be reduced with electronic filters. If the source and first order reflections were not identified then the room modes could not be modelled, so a correction filter could not accurately be designed to reduce their effects. It was decided that for the source loudspeaker to be successfully identified it must have been estimated within a radius of 10cm from the actual location: if so a score of 1000 was given. From the actual location of the first and second order images the distance to the nearest image was calculated: if this was less than 57.5cm for a first order image 100 was added to the score, while 115cm was used for the second order images threshold, which if successful each contributed 1 to the score. Therefore a maximum score of 1513 could be achieved, remembering that the array can only 'see' images in front of it. The distance limit for first order images was calculated as a quarter wavelength of a 150Hz signal. A quarter wavelength was

suggested as a minimum accuracy to be able to predict to that frequency and 150Hz was chosen as the minimum upper frequency that could be successful for a room correction system; double this was allowed for second order images. It was decided that for a setup to be suitable for predicting the sound a score of at least 1500 must be reached. Because the second order images are usually weaker than the source or first order images, due to them undergoing two wall absorptions and the sound having travelled further, they were considered less significant, hence a greater tolerance to deviate was allowed. To measure the quality of the estimates the maximum distance between an identified first order image and the actual position was also recorded.

It is worth noting that another measure of performance is the number of images estimated for a given score. Although it first seemed good if a high number of images had been detected, investigation showed that this was often because a single source had been estimated as multiple sources, close together. This was usually the case for reflections arriving late in time where the temporal processing had detected multiple signals or the spatial processing had falsely added another target into its search because of a failing of the threshold criteria. More sources were detected with measured, rather than simulated data. This could have been due to scattering of sound when it hit a surface rather than a clean reflection, or the impulse response of the drive unit increasing the duration of the reflected sound, thereby increasing signal overlap and temporal detections. Therefore the scoring was used to filter potentially successful setups which were later investigated further, using a 3 dimensional plot of the room and the actual and estimated source locations, allowing a visual inspection of the positions.

6.7.3 Differences between simulations and real experiments.

The situation faced in the real rooms was considerably more complex than in the simulated room. The value of the simulations is that they provide a controlled environment where problems can be quickly investigated. Building a simulation that was complex enough to incorporate all of the possible variables would have been an entire project and could not cover all eventualities. It was therefore important to conduct experiments in real rooms to assess the proposed system's actual potential and accuracy. The most significant differences between the simulated and real rooms and their effects will now be discussed.

The simulation assumed perfectly smooth, rigid walls, floor and ceiling which had the same frequency response for that entire surface. In the real rooms the walls and floor were concrete, considered rigid, but they were not smooth, or even, with the same frequency response over the entire surface. Often there were pillars near the walls, electrical trunking and sockets, doors and windows with radiators below. The ceilings were suspended tiles, so not rigid, with lights mounted in or on the ceiling. All these differences increased the scattering of the sound reflections or moved the origin of the reflection compared to the simplified room simulation.

Because it had been decided, for the sake for simplicity, to model empty rooms, before the real experiments could be started all furniture and equipment had to be emptied out of the room. However in some rooms lecture tables, for example, could not be removed, creating a reflective object which was not modelled in the simulations. This meant that when comparing the estimates to the calculated image

positions these additional objects had to be taken into consideration and more visual checks on the room plots were required.

The imperfections in the microphone array have been discussed in section 6.6. However, it was discovered that the loudspeaker's behaviour deviated more significantly from ideal than that of the microphones. The output of the loudspeaker decreased, and changed significantly with angle. The consequence of this was that there was very little energy radiating backwards, so the reflection from the wall behind the loudspeaker was often not identified. The polar response of the KRK loudspeaker used is shown in Figure 40.

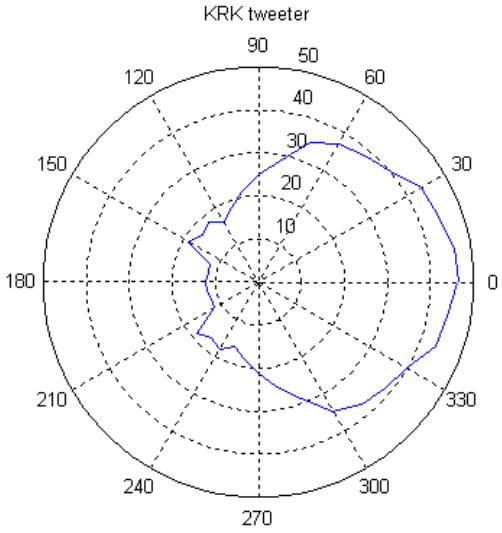


Figure 40: Horizontal polar response of the KRK loudspeaker at 10969Hz. Angle, degrees against magnitude, dB. Taken in the anechoic chamber at a distance of 1.6m.

To overcome this, two additional Linn tweeters were attached to the top of the loudspeaker, as shown in Figure 41. They were selected because they had no

faceplate to baffle them and their small size meant that they could be placed very close together. The tweeters were driven independently in case of interference patterns causing cancellations, therefore the recordings were conducted twice, once for the front tweeter and again for the rear. The polar response of the front and rear tweeters, mounted on the KRK loudspeaker are shown in Figure 42 and Figure 43 respectively. The data was processed separately, the rear unit was time aligned with the front tweeter before processing. Once the positions had been estimated both sets were concatenated.

Figure 41: A photo of the loudspeaker arrangement used.

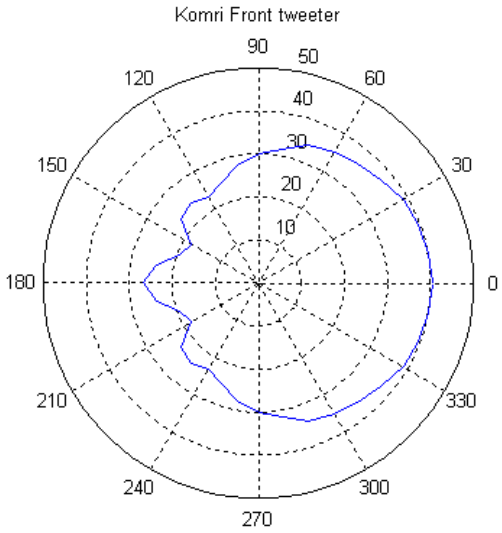


Figure 42: Horizontal polar response of the front tweeter at 10969Hz. Angle, degrees against magnitude, dB. Taken in the anechoic chamber at a distance of 1.6m.

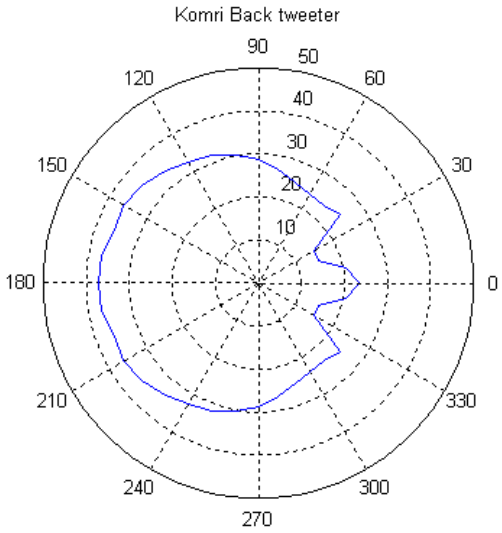


Figure 43: Horizontal polar response of the rear tweeter at 10969Hz. Angle, degrees against magnitude, dB. Taken in the anechoic chamber at a distance of 1.6m.

To minimise the effect of the real loudspeaker's response, attempts were made to equalize the measured data using an inverse filter designed with the on axis response

of the tweeter. The aim was to reduce the time domain ringing which occurred after the transmitted response as the drive unit's response decayed. These were unsuccessful and reduced the accuracy of the direction estimates. The problem was that the equalization needed to be correct for that time section under investigation, which meant it had to be designed with the signal's angle of arrival known, but it was the angle that was being estimated. Additionally multiple signals were often present within a time section so it would not be possible to correctly equalize for multiple, different angles of arrival. As equalization of the measured data was not possible, the angular dependant, time-spread response of the real drive units had to be accepted. The consequence of the drive units' non-ideal impulse responses was to degrade the performance of the measured data estimates. The effects of this degradation will be seen in the analysis of the results in section 6.7.5.

The performance measures employed relied heavily on an accurate calculation of the actual image positions, simple in the simulated situation, but prone to errors in the measured data case. All room dimensions, array and speaker positions were taken with a laser range finder, with 1mm accuracy. Although great care was taken in these measurements it is likely that the range finder was not always level, or the walls referenced to perfectly straight, meaning the exact location of the array and loudspeaker might be slightly different to the measured positions. Additionally if the line array was not perpendicular to the floor or wall all the estimated positions could be shifted slightly. When the array was moved its position was carefully rechecked with the range finder or a ruler to measure the distance from the back of the array to the wall; it was then set level with a spirit level, but some degree of error was still

expected. The combination of these position errors will have coloured the performance measures presented in this chapter.

6.7.4 Presented results parameters

As discussed above, there were 17 different array configurations investigated, both with measured data being recorded and analysed and simulations performed. Earlier analysis of simulated and measured data showed that the 16 sensor circle did not perform as well as the 24 sensor circle; it is therefore not presented here, leaving 8 array configurations with the 24 sensor circle and different numbers of line sensors, from 9 to 23 with increments of 2.

The speed of sound in air, *c,* varies with temperature and humidity, although because the effect of humidity is very small and relatively constant in the UK, this discussion only considers the temperature affect. The speed of sound can be calculated using the following formula, where T is the temperature in degrees Celsius.

$$c = 331.3 + 0.606T$$

Table 8 shows the resulting speed for the temperature range expected in a typical room in the UK.

Temperature , degrees c	**10**	**12**	**14**	**16**	**18**	**20**	**22**	**24**	**26**	**28**
c	337.4	338.6	339.8	341.0	342.2	343.4	344.6	345.8	347.1	348.3

Table 8: Temperature versus speed of sound in air.

Because the wavelength, λ, of a signal is related to the speed of sound by $\lambda = \frac{c}{f}$

temperature must be taken into consideration when selecting a centre frequency that prevents spatial aliasing in array processing; the sensors must be placed less than half a wavelength apart. The line array sensors are spaced 15mm apart, whilst the 24 sensor circle has sensors placed 15.7mm apart. A suitable frequency for the line array is therefore 10969Hz, for the 24 sensor circle a suitable frequency was 10474Hz. Note that the circle array will have slight spatial aliasing at 10969Hz with temperatures below 22˚C. Another frequency, 10722Hz, was therefore also transmitted because it was halfway between the two optimal frequencies and it prevented spatial aliasing down to 10˚C, the coolest temperature expected in a domestic room. To test the sensitivity of the processing to transmit frequencies, signals centred around 10940Hz were also transmitted. All four frequencies were transmitted and the signals recorded by both the line and circle array were processed using the same frequency, whilst different frequencies were transmitted to investigate the optimal frequency for the complete array.

As described in section 6.7.3 above, two small tweeters were used to transmit the LFM signals, with the full range KRK speaker transmitting the low frequency chirp, used in the reflection's frequency response estimation, described in Chapter 7.

The scale plan view of the room used for the tests presented here is shown in Figure 44. The room was 3.53m wide by 8.57m long with a height of 2.70m. The actual room is shown on the left, complete with the locations of the spot microphones. On the right is a plan view of the simulated room. The end wall with the window had a 40cm step behind which the radiator was mounted: Figure 45 shows its cross section.

A photograph of the room is shown in Figure 46 - notice the trunking. (This photograph was taken for an earlier setup when the loudspeaker and PC were at the opposite end to the results presented here).

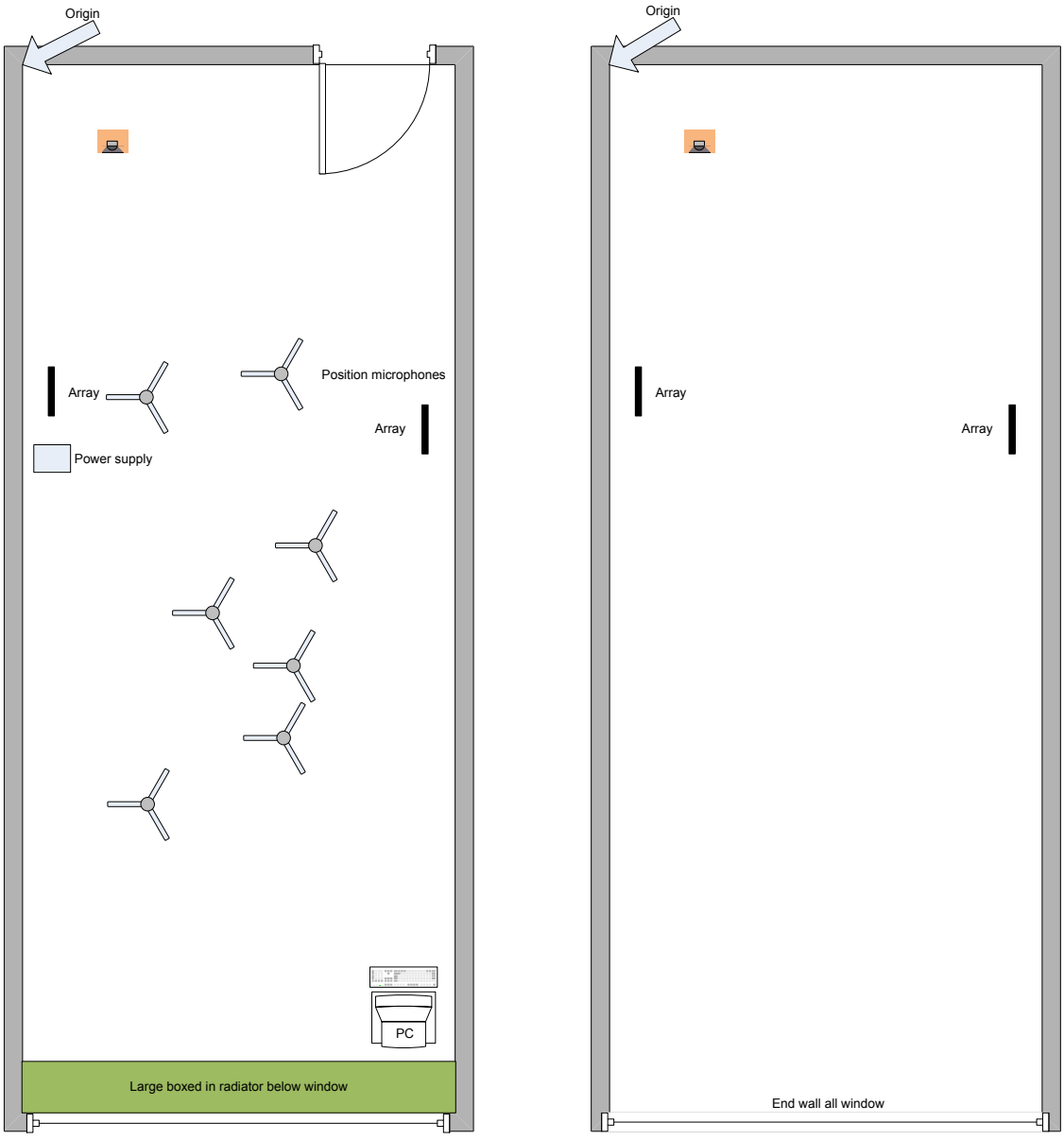


Figure 44: A scale plan view of the presented room, actual on the left and simulated on the right.

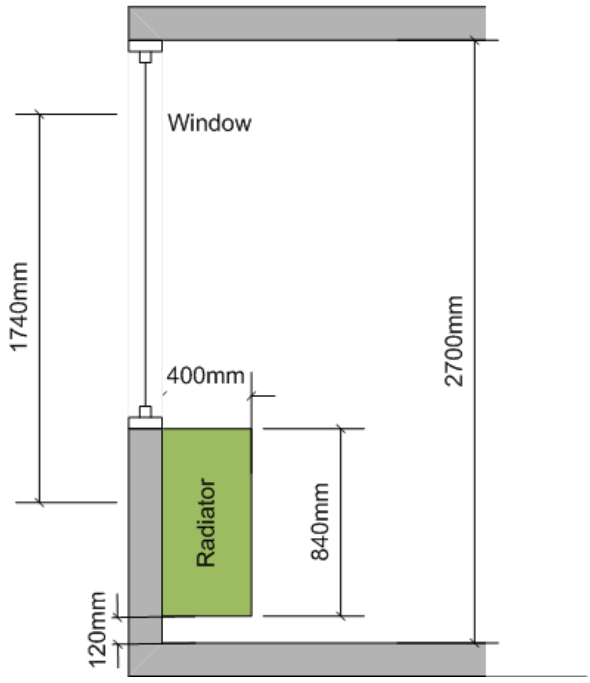


Figure 45: A scale cross section view of the end wall in the presented room.

Figure 46: A photograph of the experiments in the actual room

Different array and loudspeaker positions were tested; the combinations presented in this thesis are shown in Table 9. For each loudspeaker position, a pair of array positions, on opposite walls were required so that all the reflections could be captured. Experiments 3 and 4 are a pair as are 5 and 6, with the loudspeaker left in the same location to test the sensitivity of the solution to array placement. Figure 47 shows these locations within the plan view of the room.

Experiment	Array Position	Array Reference	Speaker Position
1	x = 4.10 y = 3.29 x = 1.17	A	x = 1.35 y = 1.33 x = 1.19
2	x = 4.56 y = 0.28 x = 1.17	B	x = 1.35 y = 1.33 x = 1.19
3	x = 4.10 y = 3.29 x = 1.17	A	x = 0.95 y = 0.75 x = 1.19
4	x = 4.56 y = 0.28 x = 1.17	B	x = 0.95 y = 0.75 x = 1.19
5	x = 2.71 y = 0.28 x = 1.17	D	x = 0.95 y = 0.75 x = 1.19
6	x = 3.02 y = 3.27 x = 1.17	C	x = 0.95 y = 0.75 x = 1.19
7	x = 3.02 y = 3.27 x = 1.17	C	x = 0.61 y = 1.46 x = 1.19
8	x = 2.71 y = 0.28 x = 1.17	D	x = 0.61 y = 1.46 x = 1.19
9	x = 2.71 y = 0.28 x = 1.17	D	x = 0.94 y = 1.14 x = 1.19
10	x = 3.02 y = 3.27 x = 1.17	C	x = 0.94 y = 1.14 x = 1.19
11	x = 2.71 y = 0.28 x = 1.17	D	x = 0.67 y = 0.50 x = 1.19
12	x = 3.02 y = 3.27 x = 1.17	C	x = 0.94 y = 0.50 x = 1.19

Table 9: Experimental setup - array and speaker location presented

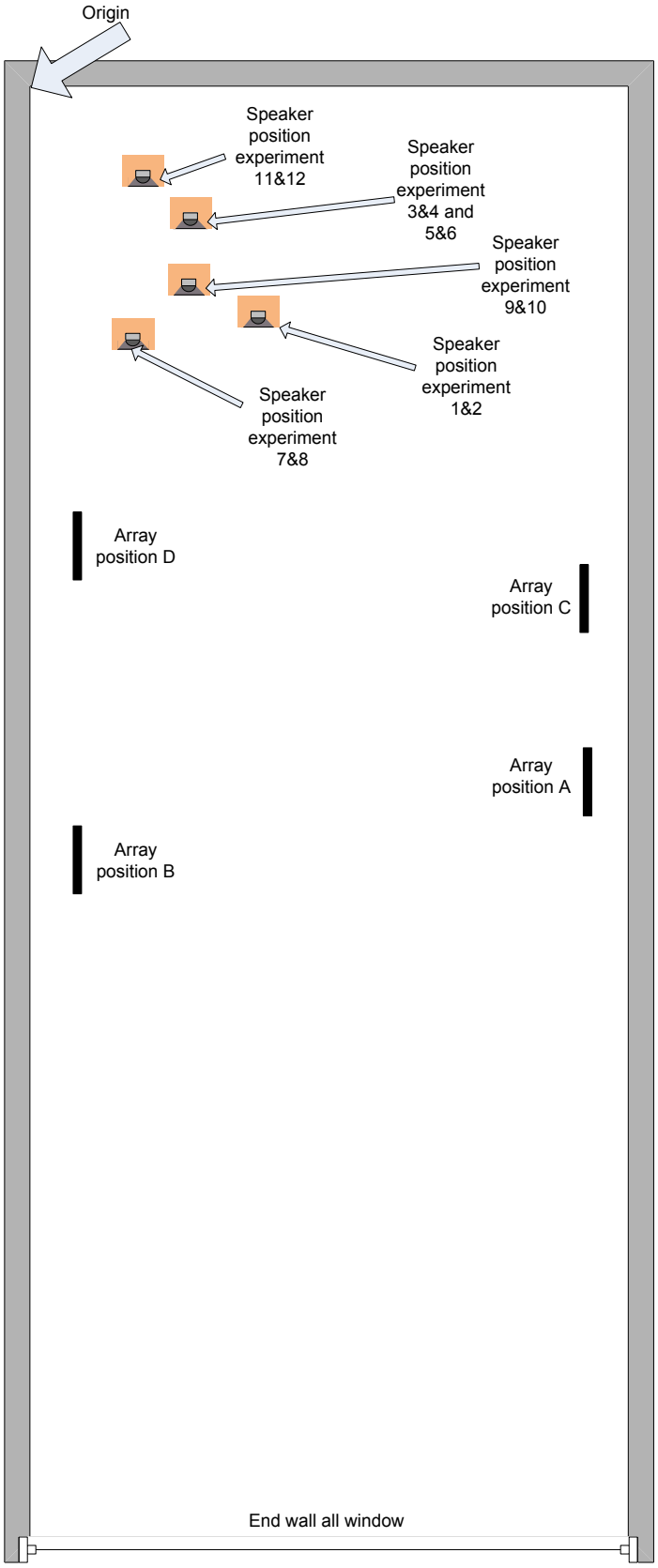


Figure 47: A scale plan view of the presented room with array and loudspeaker positions

6.7.5 Selection of best array configuration

The 12 experiments resulted in 384 different configurations, varying with the number of line array microphones used, the transmit frequencies, speaker and array positions. The solution should ideally work equally well, independently of the loudspeaker or array position, therefore to select the best array configuration the scores from the 12 experiments were plotted as a histogram of line microphones to number of occurrences where the score was above 1500, i.e. the source and all first order images were successfully identified. The simulated results can be seen in Figure 48, whilst the measured results are shown in Figure 49. A total of 158 simulated and 136 measured configurations had a score greater than 1500. The different distributions for line array microphones were unsurprising. As discussed in section 6.7.3, in the simulated room the reflections were clean, such that a single source intercepting with a single surface would produce a single reflecting sound wave, because the surface was modelled as a rigid, perfectly smooth surface. In the measured room where scattering occurred many more image sources were expected and identified. Additionally there was time domain ringing produced by the real loudspeaker, causing greater overlap of signals, and therefore more signals within one time section than in the simulated case. It is known that more microphones would be required to estimate the angle of arrival when more signals were arriving at the array in each time section.

Another reason why the measured results required more microphones than the simulated data to accurately estimate image locations is array phase errors. As discussed in section 6.6 the array parameters were ideal during the simulations, but there are many sources of variation present in the real array implementation which

could not be avoided. These variations will cause phase differences between microphones, which degrades the angle of arrival estimates. When a greater number of microphones are used the effect of these errors will average out and thus the direction of arrival estimates are less sensitive to phase errors when there are a large number of microphones.

Looking at the simulated and measured results together, the best number of line array microphones was 21.

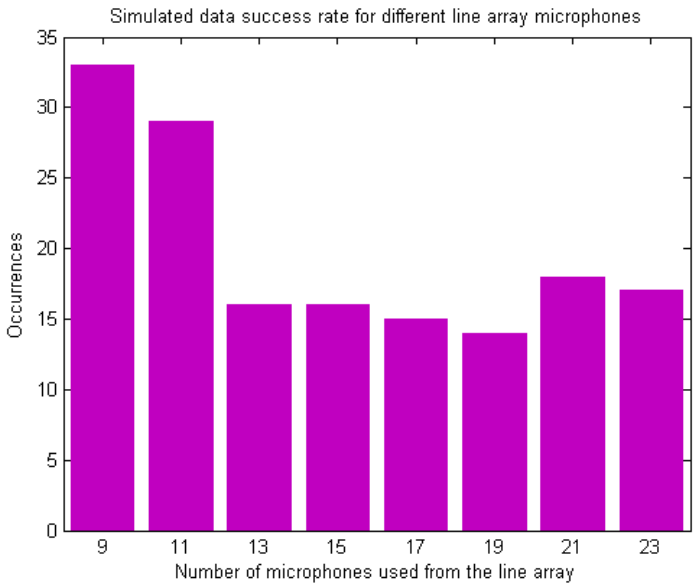


Figure 48: Histogram of successful line array configurations for all simulated experiments and frequencies

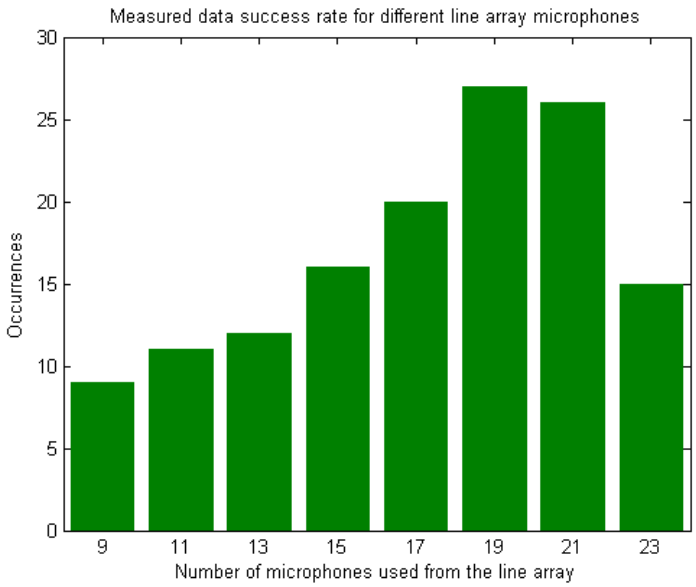


Figure 49: Histogram of successful line array configurations for all measured experiments and frequencies

I shall now consider the transmit frequency for all experiments and line array configurations which scored above 1500. Figure 50 shows the successful occurrences for simulated rooms, whilst Figure 51 shows the results for measured data. Theoretically the lower frequencies would perform less well than the higher frequencies because the phase differences at the microphones would be maximised at high frequencies, although it was possible that spatial aliasing would occur with the circular array at high frequencies, therefore reducing their performance. With both a simulated and real room the most successful transmit frequency was 10940Hz. The results show that the experiment is quite sensitive to transmit frequency. It is hypothesised that the results are more strongly affected by the performance of the line array than the circular array, because the direction of arrival is performed in two stages, with the azimuth angle set by the line array. In the line array the microphones were closer together, making a high frequency more appropriate. If the data was processed in a single stage it is expected that a lower frequency would be more successful.

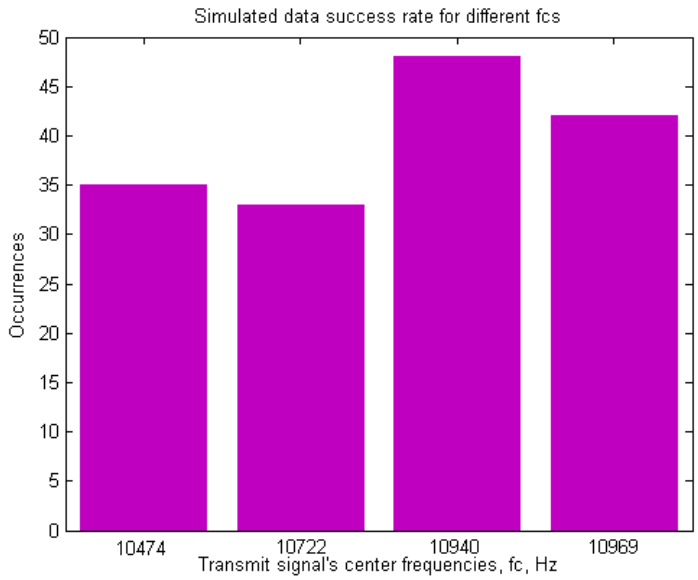


Figure 50: Histogram of successful transmit frequencies for all simulated experiments and array configurations

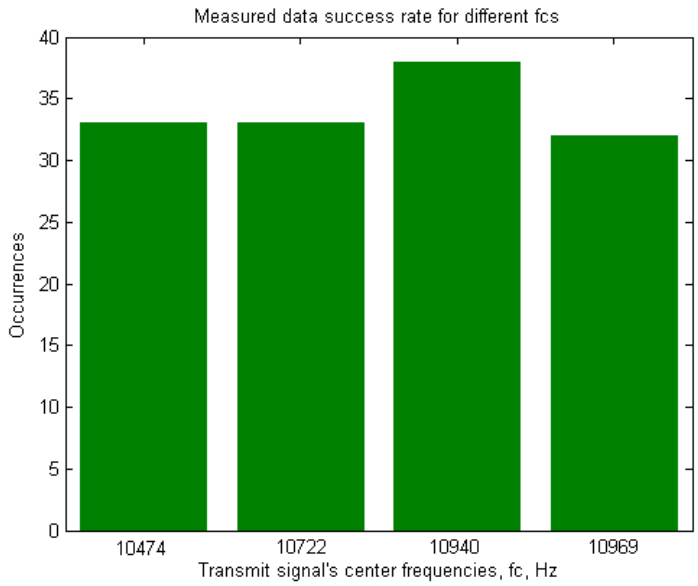


Figure 51: Histogram of successful transmit frequencies for all measured experiments and array configurations

Looking at the results for different numbers of line array microphones with a fixed transmit centre frequency of 10940Hz, as shown in Figure 52 for the simulated case and Figure 53 for the measured case, it can be seen that 21 line array microphones was indeed a sensible choice. For this configuration, over the 12 experiments 6 were successful in the simulated room, compared to 9 for the measured room.

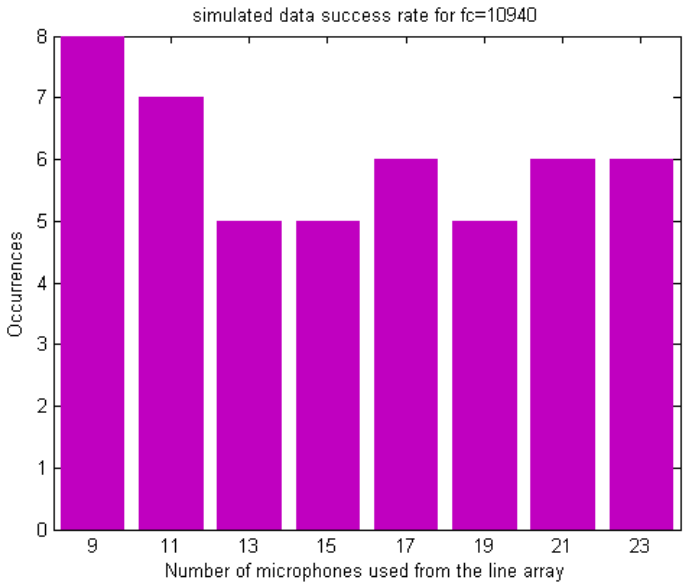


Figure 52: Histogram of successful line array microphones for all simulated experiments with a transmit frequency of 10940Hz

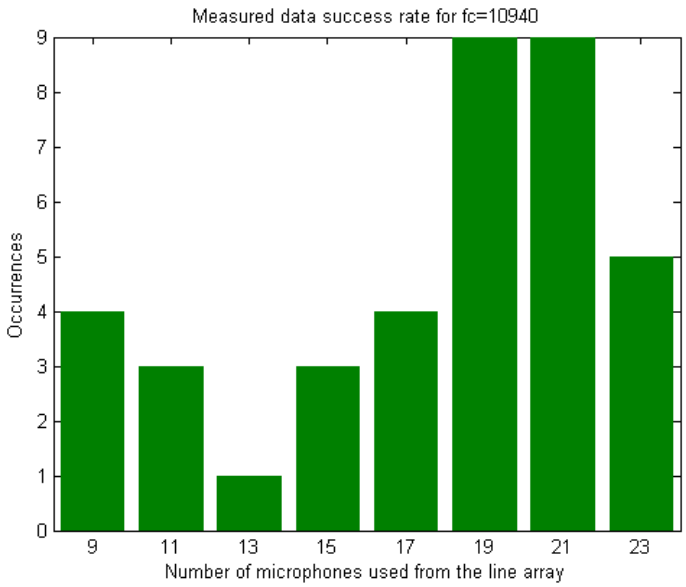


Figure 53: Histogram of successful line array microphones for all measured experiments with a transmit frequency of 10940Hz

6.7.6 Detailed results analysis

With the array configuration of 21 line array microphones, 24 circle array microphones and transmit frequency centred around 10940Hz, the precise differences between the actual and estimated positions were investigated, rather than using the scoring method. Table 10 shows that over the 12 experiments, the difference between the actual source position and the estimated position was very low for the simulated room, with a maximum difference of 0.9cm and an rms difference of 0.76cm. The measured results were slightly worse, with a maximum difference of 9.3cm and an rms difference of 7.12cm. Part of this difference might have been mistakes in recording the speaker and array positions in the real room, as described in section 6.7.3. It is worth noting that with this configuration the source was successfully identified in all cases.

Experiment	Simulated source estimation difference, m	Measured source estimation difference, m
1	0.007	0.058
2	0.008	0.093
3	0.008	0.081
4	0.008	0.044
5	0.008	0.036
6	0.006	0.067
7	0.007	0.091
8	0.008	0.046
9	0.008	0.093
10	0.009	0.056
11	0.008	0.090
12	0.007	0.065

Table 10: Source position differences, actual to estimated.

Table 11 shows the performance of the system when identifying the first order images. The very low minimum differences between the actual and estimated positions show that the solution can be successful, although the maximum errors show that in some experiments an image was not successfully identified. In all cases except the simulated room (experiments 7 and 8), the image which was not successfully identified by one array was successfully identified by the other array, on the opposite side of the room. When the results from both arrays are combined, as described in Chapter 8, logical rules could be used to identify the best image in the area where both arrays can see the same images. This means that in the majority of cases the maximum error will be less than 0.5m.

Experiment	Simulated room in metres			Measured room in metres		
	Minimum difference	**RMS difference**	**Maximum difference**	**Minimum difference**	**RMS difference**	**Maximum difference**
1	0.005	0.14	0.30	0.09	0.18	0.26
2	0.016	0.53	0.84	0.07	0.28	0.44
3	0.021	0.19	0.26	0.04	0.16	0.26
4	0.006	0.29	0.46	0.18	0.39	0.48
5	0.008	0.23	0.49	0.05	0.21	0.42
6	0.008	0.80	1.79	0.06	0.23	0.33
7	0.018	0.81	1.81	0.09	0.43	0.92
8	0.018	0.70	1.56	0.03	0.41	0.83
9	0.011	0.36	0.78	0.05	0.80	1.78
10	0.001	0.19	0.41	0.12	0.28	0.49
11	0.088	0.51	1.09	0.07	0.23	0.36
12	0.064	0.20	0.33	0.06	0.18	0.25

Table 11: First order image position differences, actual and estimated.

Disappointingly it can be seen by comparing the results from experiments 3 and 4 to experiments 5 and 6, where the loudspeaker was kept in the same location and the array positions changed, that in the simulated room the system was able to identify all the first order images with good accuracy in experiment 3, but in experiment 6 it was not. This shows that the solution is sensitive to array placement. The probable cause of this is that the number of reflections arriving in one time section is dependent on the array's location and in some locations there may be too many signals arriving for the array to resolve.

An alternative way of viewing the differences between the actual and estimated image positions is to look at the angle of arrival differences. This gives an insight into the

estimation errors attributable to the spatial processing, with greater isolation from the temporal processing. Table 12 shows the angular differences for the source, which confirm the results shown in Table 10, that there was little error in identifying the source with high accuracy. Table 13, however, shows that whilst the line array is consistent in its estimation, with only small differences between the estimated and actual angles, the circular array is not. The minimum elevation difference is consistently small, showing that the process can work, but there are some large maximum errors, as highlighted in pink. This reveals that the weakest part of the temporal and spatial processing is in the elevation estimations, which can sometimes fail completely. It is hypothesized that the results may improve if a larger number of microphones were used for the circular array, although it may not be possible to overcome the limitation of a circular array's directionality pattern and a planar array may in fact be the only way of successfully resolving the two dimensional angle of arrival when there are many signals arriving at the array in the same time section. The investigation into a possible solution has been left for future work.

Experiment	Simulated source estimation difference, degrees		Measured source estimation difference, degrees	
	Azimuth	Elevation	Azimuth	Elevation
1	0.04	0.01	0.46	0.04
2	0.00	0.01	0.60	3.56
3	0.02	0.00	0.43	0.55
4	0.00	0.01	0.30	1.09
5	0.03	0.01	0.13	2.74
6	0.05	0.00	0.95	0.85
7	0.04	0.02	1.39	0.77
8	0.06	0.02	0.36	0.03
9	0.05	0.02	2.65	0.73
10	0.02	0.01	0.52	0.86
11	0.03	0.03	2.13	0.32
12	0.01	0.01	0.56	1.09

Table 12: Source angular differences, actual and estimated

Experiment	Simulated room difference, in degrees						Measured room difference, in degrees					
	Minimum		RMS		Maximum		Minimum		RMS		Maximum	
	A	E	A	E	A	E	A	E	A	E	A	E
1	0.04	0.00	0.44	0.68	0.76	1.51	0.52	0.69	1.34	2.09	2.51	3.79
2	0.08	0.14	3.55	18.2	7.93	40.7	0.00	0.59	0.67	8.57	0.88	16.6
3	0.04	0.08	0.78	2.83	1.34	5.75	0.22	0.47	0.64	1.80	1.16	2.75
4	0.04	0.01	1.30	32.1	2.25	71.2	0.54	0.66	2.04	6.72	3.77	10.4
5	0.02	0.04	1.07	39.2	1.85	87.6	0.12	0.27	1.07	17.8	1.52	39.7
6	0.00	0.02	0.31	0.88	0.59	1.94	0.36	0.66	1.42	1.61	1.99	2.85
7	0.05	0.16	3.2	40.0	7.13	89.4	0.52	1.02	1.29	10.3	2.35	22.8
8	0.01	0.03	0.54	1.95	1.05	4.32	0.03	0.24	1.38	11.5	2.4	25
9	0.05	0.12	1.57	2.58	3.28	4.53	0.39	0.20	2.81	20.3	5.73	45.4
10	0.00	0.02	0.43	2.15	0.72	4.76	0.08	1.24	0.56	6.06	0.82	12.3
11	0.03	0.08	2.51	2.90	4.51	4.23	0.27	0.13	1.62	53.8	3.32	84.9
12	0.54	0.04	1.34	0.26	1.95	0.52	0.64	0.29	0.98	1.54	1.30	2.75

Table 13: First order image angular differences, actual and estimated: A=Azimuth, E=Elevation. Large errors are highlighted in pink.

During the development of the spatial processing three dimensional plots of the rooms were often used so that the location of the actual and estimated reflection origins could be visualized, giving a clear insight into the performance of a specific setup. Unfortunately these are not clear when printed on paper. Therefore two dimensional projections of the results must be used here, whist still illustrate the scattering of image estimates. Figure 54 shows the results for the simulated room whilst Figure 55 shows the results for the measured room, both for experiment 12, viewed from above. The room is represented by the black rectangle, the arrays by the red crosses. The actual image locations are shown as black squares and circles, depending on which array can see them. The green crosses represent the estimated images from the bottom array and the purple crosses are the estimated images from the top array.

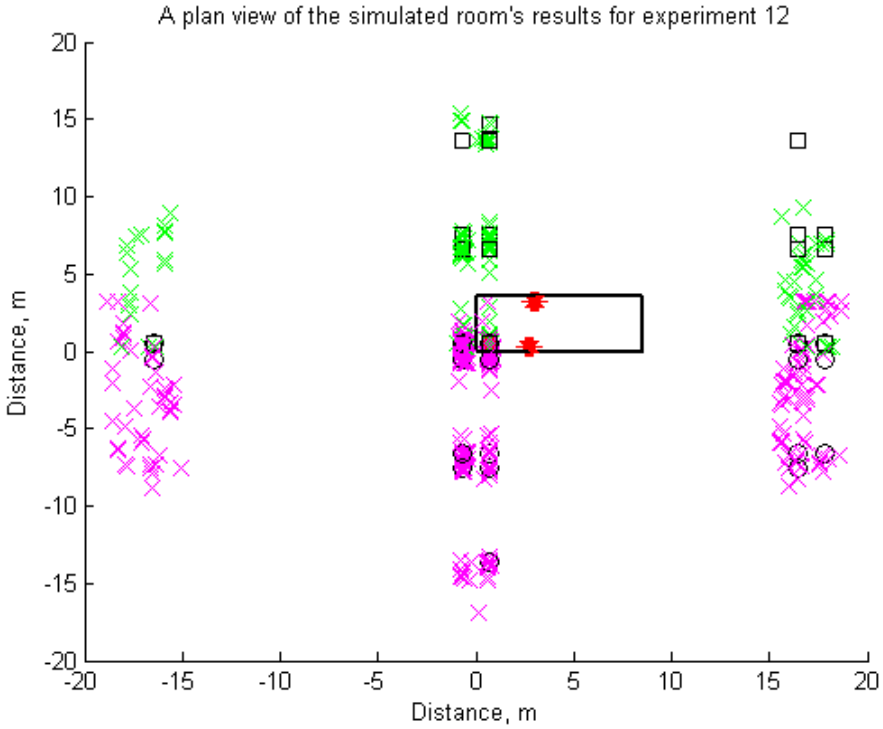


Figure 54: Plan view of the simulated room's results for experiment 12

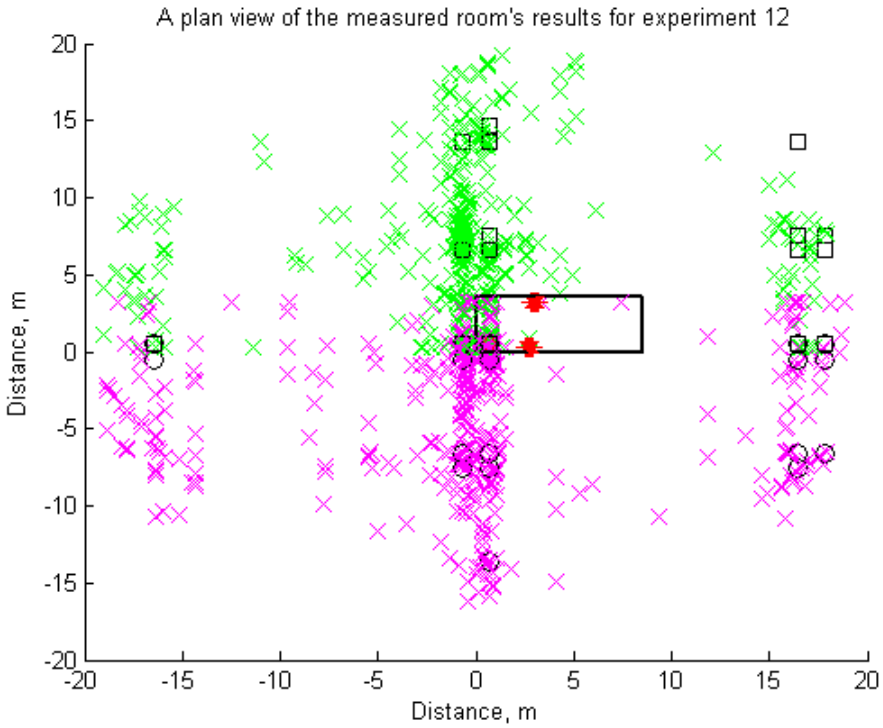


Figure 55: Plan view of the measured room's results for experiment 12

The plan views clearly show that in the simulated case the estimated images are all centred around the actual image locations, but in the measured room some estimated image locations were identified that were not close to an expected image. This is thought to be a result of the sound scattering when it hits an object, rather than it reflecting cleanly as it theoretically would from a rigid, smooth surface. One of the advantages of the IMP algorithm is that it can estimate the signal's amplitude at the same time as its direction of arrival. This information has been utilized to filter out some of the estimated images which were expected to be falsely detected by the processing, or a result of scattering, and therefore have low energy. Additionally if a number of images were estimated very close together a weighted average of their location was taken to further reduce the effects of scattering. The results, after filtering, are presented in Figure 56 and Figure 57. The filtering has reduced the number of images identified from 716 images to 184 images for the simulated room and from 2103 images to 540 images for the measured room. This filtering process is, however, not guaranteed to leave the correctly estimated images unaffected. Out of the 12 experiments presented here the source estimate was not degraded by the filtering in all experiments, measured and simulated. However a previously identified first order image was falsely removed in simulated experiments numbered 9, 10 and 11, with the measured data experiments 3, 4, 8, 9 and 11 also having a first order image falsely removed. Therefore the filtering algorithm was too severe and a refinement which does not remove important images is thus required. However, time constraints necessitated that these refinements be left for further work.

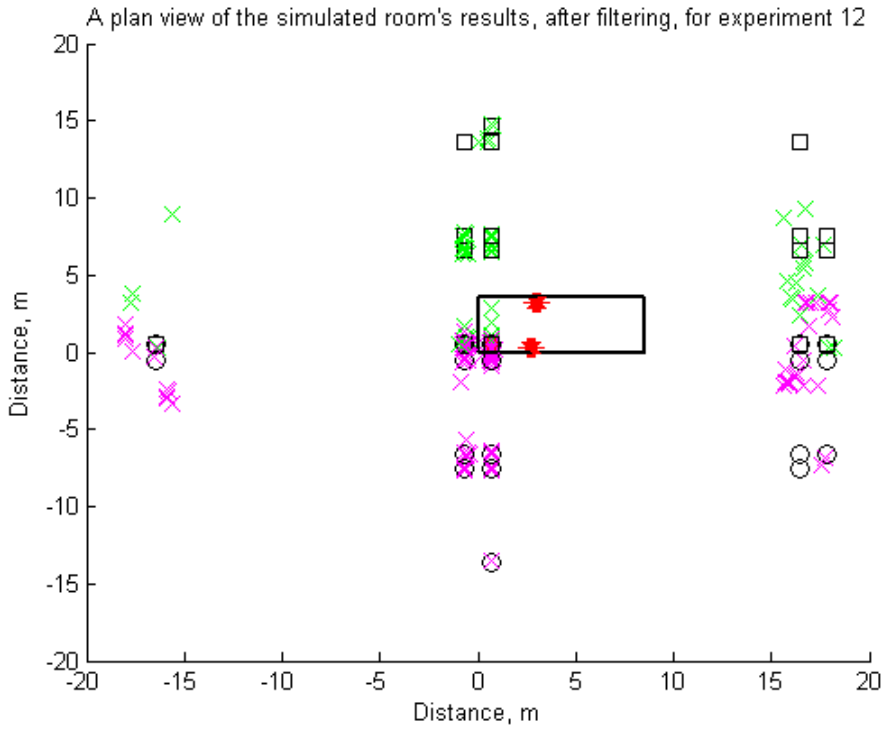


Figure 56: Plan view of the simulated room's results after filtering for experiment 12

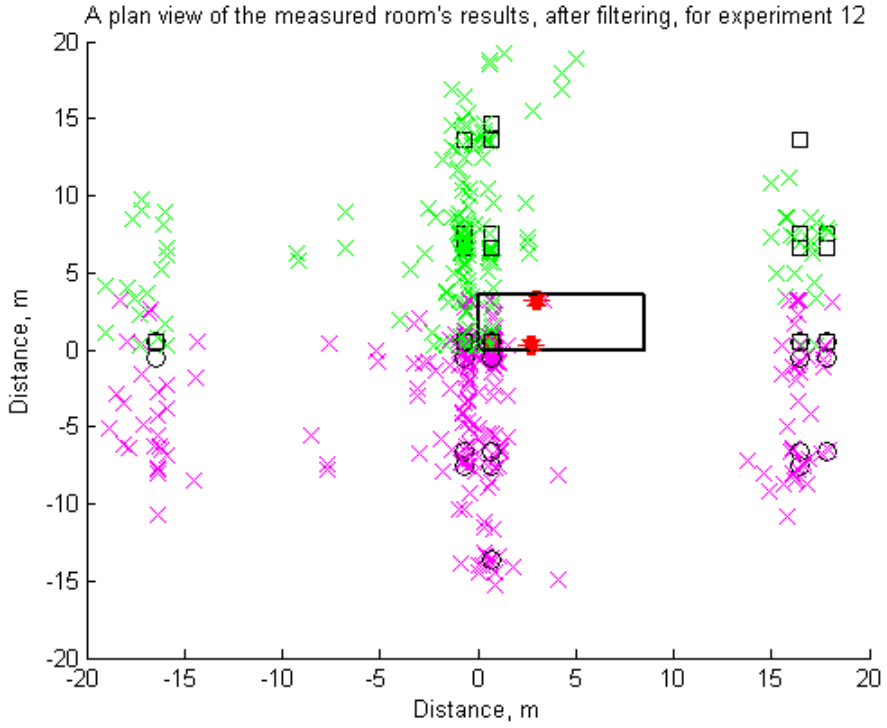


Figure 57: Plan view of the measured room's results after filtering for experiment 12

6.7.7 Measured room repeatability tests

Another measure of the system's performance was whether the results from a particular setup could be repeated. To ascertain this, experiment 9 was repeated 100 times in quick succession, with a repetition interval of 79 seconds. The data for this experiment had to be recorded before all the data was analysed, therefore the best array parameters were not known; an educated guess was made to select 19 line microphones and a transmit centre frequency of 10969Hz. The actual array configuration and transmit frequency were not crucial as it was the repeatability of the estimates that was under investigation, not their absolute accuracy. The results show that there was very little variation in either the source or first order image angles of arrival. The angular variation of the source is shown as a histogram in Figure 58 for the azimuth angle and in Figure 59 for elevation. The source loudspeaker's rms azimuth variation is 0.082 degrees, compared to an elevation rms variation of 0.000 degrees. The variation increases slightly for the first order images where 4 out of the 5 first order images were repeatedly identified by the scoring method: the variation of these 4 image estimates is shown in Figure 60 and Figure 61 for azimuth and elevation angles. The rms azimuth variation for all first order images detected was 0.152 degrees, compared to 0.274 degrees for elevation.

These results are encouraging, since although the repeatability experiment was recorded over 2 hours 12 minutes, the variation was very small. The test room was situated off a busy corridor, and so the main source of the variation is likely to be the changes in background noise from people walking by and talking or from adjacent doors banging. The small variation in angular estimations is evidence that the LFM

chirp signal was providing sufficient processing gain to minimise the effect of measurement noise to a negligible level.

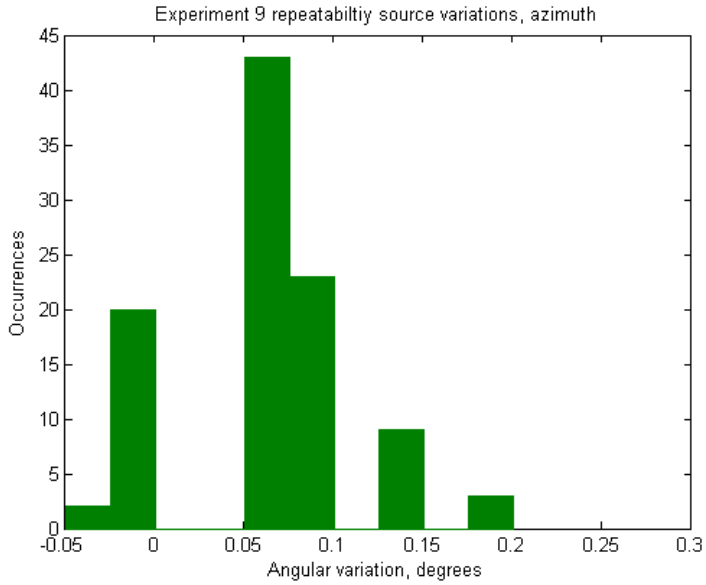


Figure 58: Experiment 9, source variation, azimuth angle

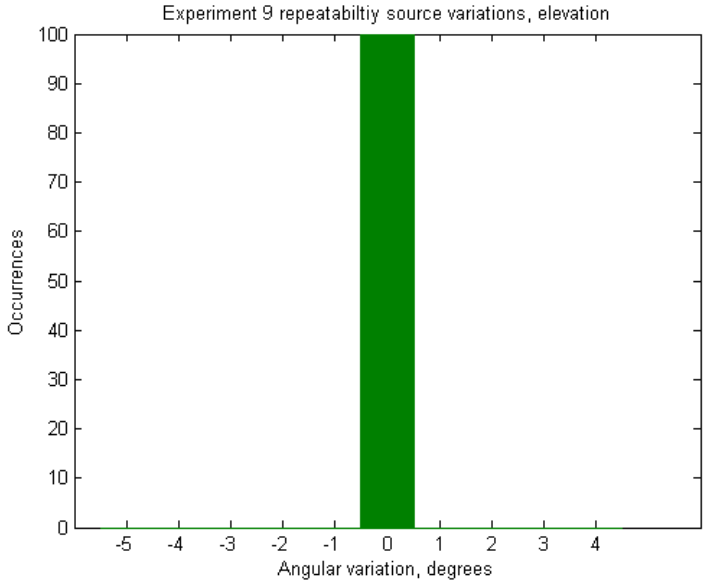


Figure 59: Experiment 9, source variation, elevation angle

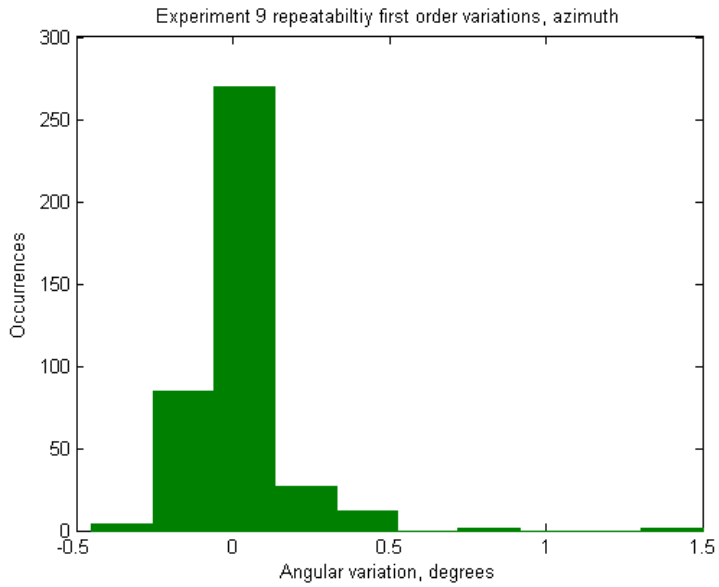


Figure 60: Experiment 9, first order images variation, azimuth angle

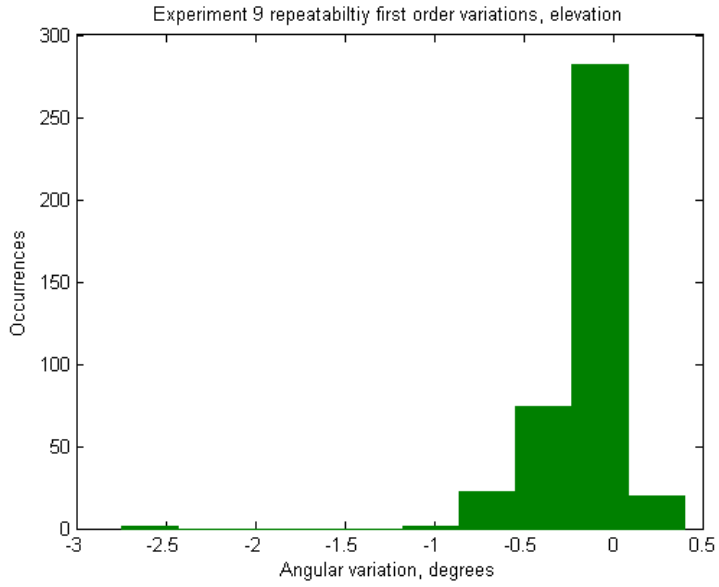


Figure 61: Experiment 9, first order images variation, elevation angle

6.8 Chapter conclusions

The topic of spatial processing is large and diverse. There have been a considerable number of publications on the topic; however, most of the existing research has focussed on a theoretical study of the subject, particularly on direction of arrival algorithms. An original feature of this research project has been the use of measured, experimental data to backup the simulations, verify the decisions made and guide the development of the solution.

The main application of spatial processing historically has been sonar, where reverberation is low and the solutions are usually concerned with uncorrelated signals. More recently spatial processing techniques have been employed in communications systems, where the reverberation and correlation of signals is also much less of an obstacle than it is within a small enclosed room as has been dealt with in this project. The limited pre-existing research into the spatial analysis of measured sound within a small room had not progressed to the level required for modelling a room using only measured acoustic data. Correlated signals within reverberation have, however, been analysed in this project to produce such a model with moderate success.

This chapter has discussed the many choices which had to be made throughout the development concerning array geometry and the processing algorithm. For each major decision a comparison of different options has been presented and the selected option has been justified. The limitations of available recording hardware and processing power has led to a bespoke solution of a line and circle array working

together to efficiently capture and process a three dimensional soundfield. The efficiency of the solution would make a commercial implementation of the system feasible for it's intended domestic application. This research demonstrates that the number of arrays required to capture the soundfield of an entire room depends on the room's geometry, but for the majority of domestic rooms which are rectangular, two arrays are sufficient. The line array estimates the azimuth angle of arrival of the signal, with this fixed the circle array then estimates the elevation angle. When the azimuth and elevation angles are combined with the distance estimated from the temporal processing, the three dimensional coordinates of each reflection have been estimated.

To allow practical evaluation of the ideas an implementation of the array has been produced, using a PCB to minimise the placement errors of the microphones. Associated electronics and software were designed to multiplex and record the microphone signals, allowing a fully automated recording process. Experiments were performed in an attempt to calibrate the array, however, the errors present in the setup were of a similar order to those expected in the microphones, so it was not successful.

A number of different direction finding algorithms have been investigated and are summarised in this chapter, with the IMP algorithm selected because of its proven ability to work with correlated signals. It is an iterative algorithm which searches a manifold, or library, to estimate the parameters and number of multiple, correlated signals. Work to adapt the algorithm to this application has resulted in a multi parameter threshold criterion for introducing a new target, or terminating the search.

This, combined with a validity test for each estimate, has set a good balance between reducing the number of false targets being identified and extracting the angles of arrival of the intended signals. The processing efficiency of the search has been increased by first performing a coarse search to identify the rough location of the spectral peak, and then by a finer search to accurately select the correct steering vector and angle of arrival of the signal. Small scale tests have demonstrated the performance limitations of the solutions prior to simulated and real room data results being presented.

A variety of different array configurations and transmit frequencies have been presented, and the analysis of the results has been enhanced by a logical scoring system which allowed quick evaluation of the results to be performed. From the experimental results presented, an array with 21 line array microphones, 24 circle array microphones and a transmit centre frequency of 10940Hz proved to be most successful. Detailed analysis of the results showed that the solution consistently identifies the source loudspeaker, but the identification of the all the first order images is not always possible. The analysis identified the elevation estimate to be the weakest part of the solution. This could be a limitation of the circle array, caused by its uneven directivity pattern, meaning that a planar array is, in fact, required. However, in hindsight it might have been better to have had an array with the same geometry and number of microphones but greater spacing and a lower transmit frequency, say 4kHz. Although this would have been less aesthetically desirable it would have had the advantage that the real microphones' responses would be closer to omnidirectional as the wavelength would be much larger than their diameter, making

array calibration easier or unnecessary. Additionally there would be less scattering from the walls because of the longer wavelength. It is believed that with these refinements the robustness of the solution could be improved.

Chapter 7 Spectral Processing

There were three stages to the proposed system; the temporal processing and spatial processing described in Chapters 5 and 6 estimated the coordinates of the reflections. The third stage of the process was to estimate the frequency responses of each of the reflections. The information available was the estimated coordinates of the image locations and any data which could be measured at the array locations. It was decided to transmit a second test signal which could be recorded at the array to capture the low frequency response of the room. Still processing each array separately, a comparison of this recorded impulse response and a prediction of the impulse response at the array could then be performed. The frequency response of each reflection was then adjusted to minimise the difference between the measurement and the prediction. The aim was that there would be a convergence between the two impulse responses and a best fit filter response for each reflection.

7.1 Methods of measuring the low frequency response of the room

For the system to work there needed to be a reference signal, recorded at the array, measuring the low frequency behaviour of the room. A review of the literature discovered two commonly used test methods for measuring the impulse response of an acoustic space, Maximum Length Sequence (MLS) and swept sinewaves. These will now be compared.

7.1.1 Maximum length sequence

The MLS technique is based on the assumption of perfect linearity and tine-invariance of the system [Farina 2000]. If the system is time invariant multiple measurements can be made and averaged together to increase the signal to noise performance of the system. When using the MLS technique the acoustic space is excited with a pseudorandom signal which has noise-like statistical properties. The length of the MLS signal must be greater than the impulse response of the system to be measured, to avoid time aliasing. The length, L, in number of samples, is defined as:

$$L = 2^m - 1$$

where m is the order of the signal. The impulse response of the system is obtained by circular crosscorrelation between the measured output and the input signal [Stan et al 2002].

The main problem with the MLS method is that, because of the circular deconvolution employed, any non linear distortions in the system appears as distortion peaks arriving after the initial, linear part of the system's impulse response, but possibly overlapping the reverberant tail of the response. This limits the overall signal to noise ratio of the measurement system such that careful optimisation of the transmitted amplitude is required to reduce the effects of background noise, but not increase the nonlinear distortion of the loudspeaker [Stan et al 2002, Muller and Massarani 2001].

7.1.2 Swept sine waves

An alternative method is to transmit a sinusoidal sweep in which the instantaneous frequency is either made to vary linearly with time, or exponentially with time. The

received signal is then matched filtered to linearly deconvolve the signal, resulting in the linear impulse response of the system [Stan et al 2002]. When a linear sweep is used any distortion products in the system appear as a 'sort of noise' everywhere in the deconvolved response [Farina 2000]. Because the distortion is correlated with the transmit signal it does not decrease with averaging. The smearing of the non linear response is more welcome than the overlapping distortion peaks produced by the MLS method, but it is still possible for the non linear and linear responses to overlap. When the sweep frequency is exponentially varied the distortion peaks are present, but it is possible to guarantee that these are separated in time from the linear part of the response. Harmonic distortions appear prior to the linear impulse as separated peaks representing different harmonics. Thus the linear impulse response can simultaneously be measured along with the harmonic distortion at various orders [Farina 2000, Stan et al 2002].

To obtain a flat frequency response with a logarithmic sweep the matched filter needs to be designed to compensate for the amplitude modulation caused by the different energy being generated at low and high frequencies [Farina 2000]. Additionally, to minimise the influence of transients at the start and end of the transmit signal, which show up as ringing at the frequency extremes of the resulting frequency response, a time window needs to be applied to the transmit signal [Stan et al 2002]. The compromise is that in order to control sidelobe levels, with the swept sinusoid, the temporal resolution must be slightly compromised.

7.1.3 Selected method

The nonlinearities of a loudspeaker increase with the cone excursion, which increases when the wavelength is large compared to the cone's diameter. Thus the nonlinearities of a loudspeaker are large at low frequencies when the wavelength is many times the cone diameter. The modal behaviour of a room is linear, thus a correction filter designed to modify the modal response needs to be linear, and therefore only the linear response of the room is required and any nonlinearities should be minimised from the measurement and isolated from the correction filter design.

The logarithmic sweep was, therefore, optimal for measuring the low frequency response of the room. For an MLS transmit signal there needs to be careful optimisation of output amplitude to ensure the nonlinearities of the loudspeaker are minimised whilst achieving a high signal to noise ratio. With a logarithmic sweep, because the linear and non linear responses are separated in time there is not the limit on output amplitude so, the signal to noise ratio can be increased by either extending the transit time, thus increasing the processing gain achieved by the matched filter or by increasing the amplitude [Farina 2000].

The response needed to be captured over the frequency range of 20Hz upwards. The usable upper frequency was limited by the reflections' location estimation errors. If the comparison of responses at the array was performed over a wider frequency range than was realistic to accurately model with the estimated image position, the optimal

filter response of each image might not be selected. Additionally a larger set of filters would have to be tested, increasing the computational load.

Research into room response correction, presented in Chapter 2, revealed that the modal response of the room was most in need of electronic filter equalization to improve the perceived sound quality [Stefanakis et al 2008]. By filtering the loudspeaker signal so that the Q factor of the room's modal frequencies are reduced the decay time of the room modes are reduced, which gives significant audible benefits. [Wilson et al 2003, Makivirta et al 2001, Karjalainen et al 2003]. The upper frequency limit of such systems is between 300 and 500Hz, which was therefore the upper frequency limit to which an image's response needed to be estimated to.

Because the test signal was recorded before the image positions were calculated the sweep was performed up to 2kHz, deliberately higher than needed. The bandwidth was later reduced by limiting the bandwidth of the matched filter. The mean error to all first order images in the experiments conducted, for the most successful array configuration, was 0.24m. This suggested that the image positions were not estimated with sufficient accuracy to reliable predict the sound at a frequency above 300Hz.

The transmit signal, *tx*, is calculated as:

$$tx(t) = sin\left(\frac{2T_{tx}\pi f_{start}}{log\left(\frac{f_{stop}}{f_{start}}\right)} \cdot \left|e^{\left(\frac{t}{T_{tx}} \cdot log\left(\frac{f_{stop}}{f_{start}}\right)\right)} - 1\right|\right) . r(t)$$

where

$$t = 0{:}\frac{1}{FS}{:}T_{tx}$$

and T_{tx} is the duration of the transmitted chirp, 10 second, f_{start} is the start frequency of the chirp, 20Hz here, f_{stop} is the stop frequency of the chirp, 2kHz and t is time. It was decided to transmit the chirp for 10 seconds to ensure that there was sufficient energy at the low frequencies to provide a good SNR. The transmit signal had the following ramp window applied to reduce the ringing caused by sudden transients.

$$r(t) = \begin{cases} r(t-1) + \dfrac{1}{T_{ramp\_up}} * FS & n = 0\ to\ T_{ramp\_up} \\ r(t-1) - \dfrac{1}{T_{ramp\_down}} * FS & n = T_{ramp\_down}\ to\ T_{tx} \\ 1 & else \end{cases}$$

where T_{ramp_up} is the ramp up time, a value of 1second was used. T_{ramp_down} is the ramp down time, a value of 1second was used and FS = sample rate of the recording (48kHz).

The matched filter is a windowed, time reversed version of *tx*, but only over a limited duration, $T_{matched}$, to limit the upper frequency limit. $T_{matched}$ = 7 second to give the desired upper frequency of 300Hz. Window *w1* equalizes the response to have a flat frequency response, whilst *w2*, a Tukey window, further reduces the ripple at the frequency extremes.

$$w1(t) = e^{\left(\log\left(\frac{f_{stop}}{f_{start}}\right).\frac{t}{T_{tx}}\right)}$$

$$w2(t) = \begin{cases} 1.0 & 0 \le |t| < \alpha\frac{T_{matched}}{2} \\ 0.5\left(1 + \cos\left(\pi\frac{t - \alpha\frac{T_{matched}}{2}}{2(1-\alpha)\frac{T_{matched}}{2}}\right)\right) & \alpha\frac{T_{matched}}{2} \le |t| < \frac{T_{matched}}{2} \end{cases}$$

$$tx_{match}(t) = sin\left(\frac{2T_{tx}\pi f_{start}}{log\left(\frac{f_{stop}}{f_{start}}\right)} . \left| e^{\left(\frac{-t}{T_{tx}} .log\left(\frac{f_{stop}}{f_{start}}\right)\right)} - 1 \right| \right) . w1(-t) . w2(-t) . r(-t)$$

where

$$t = 0{:}\frac{1}{FS}{:}T_{matched}$$

7.2 Adaptive filters:

The first method investigated to estimate the room's surfaces responses was to use adaptive filters. It was thought that one adaptive filter per image could be trained to create an estimated signal that matched the recorded signal at the array, the target response. Small scale tests were carried out where three signals with known responses were created and three adaptive filters tried to match their responses. Initially FIR adaptive filters were tested, because this topology is guaranteed to be stable, with the test signal's response being generated with an IIR filter. The adaptive filters were proven to work well if there was no overlap in time of their responses, but as soon there was an overlap the system failed. Although the combined estimated response matched the target, looking at the individual adaptive responses in the time domain it was clear that there was no discrimination of which signal was being

modelled by which filter. For example, if adaptive filter 1 had one peak for the first signal which then decayed to zero and a second signal arrived within its duration, the response of the second signal would be included partly within adaptive filter 1's response and partly within filter 2's response.

Attempts were made to reduce this effect by applying an exponential window to the training weights, so the filter's response could not have a significant second peak towards the end of its duration. This did improve the situation, but not solve it. Another attempt involved switching to IIR adaptive filters, with the risk of instability, but greatly reducing the duration of the filter and thus the overlapping signal problems. Again this led to improvements, but not solutions. It was therefore decided that this approach was not going to be reliable enough when room responses were analysed.

7.3 Maximum likelihood methods

Because adaptive filters had failed to perform adequately, the decision was taken to have a library of predefined filters and search for the filter which best matched the actual response of an image from the library.

Two different implementation methods were considered, first order estimation and expansion or individual image estimation. The idea for image expansion came from the image source method of room simulation, described in Chapter 4. If the parameters of the source and first order images are known, i.e. their positions and

frequency responses, it is possible to calculate the room's surfaces' absorption characteristics. All the information is then present to accurately calculate the positions and frequency responses of all higher order images. The frequency responses of the first order images can be calculated using a maximum likelihood search of the best filter responses for the source and first order images. Using image expansion to calculate the higher order image's parameters, an impulse response at the array can be calculated, which can be compared to the actual measured response at the array, and the difference used to select the best matching filter responses. The key assumption to this method working is that the first order images can be correctly identified from the spatial processing results. Theoretically this is possible if the room is an empty rectangle, but gets more difficult for complex room geometries. The disadvantage of this approach is that it is not easily scalable to rooms with objects in them, identifying an image's order from a large number images, without knowledge of the room layout would be very difficult. Looking at the results of the spatial processing, although some array configurations had successfully identified all first order images the error in their estimated positions would further complicate the identification of an image's order.

It was therefore decided that only estimating the responses of just the first order images and using image expansion was not practical. The alternative approach is to estimate an individual filter response for each image, using a Maximum likelihood search and comparing the estimated and actual responses at the array locations. The advantage here is that no assumptions about the room have to be made; it would be possible to work with odd shaped rooms, or objects within the rooms. However, there

is clearly a very high computational cost to this approach. Although this is not ideal it was considered to be the only possible solution, so was selected for investigation.

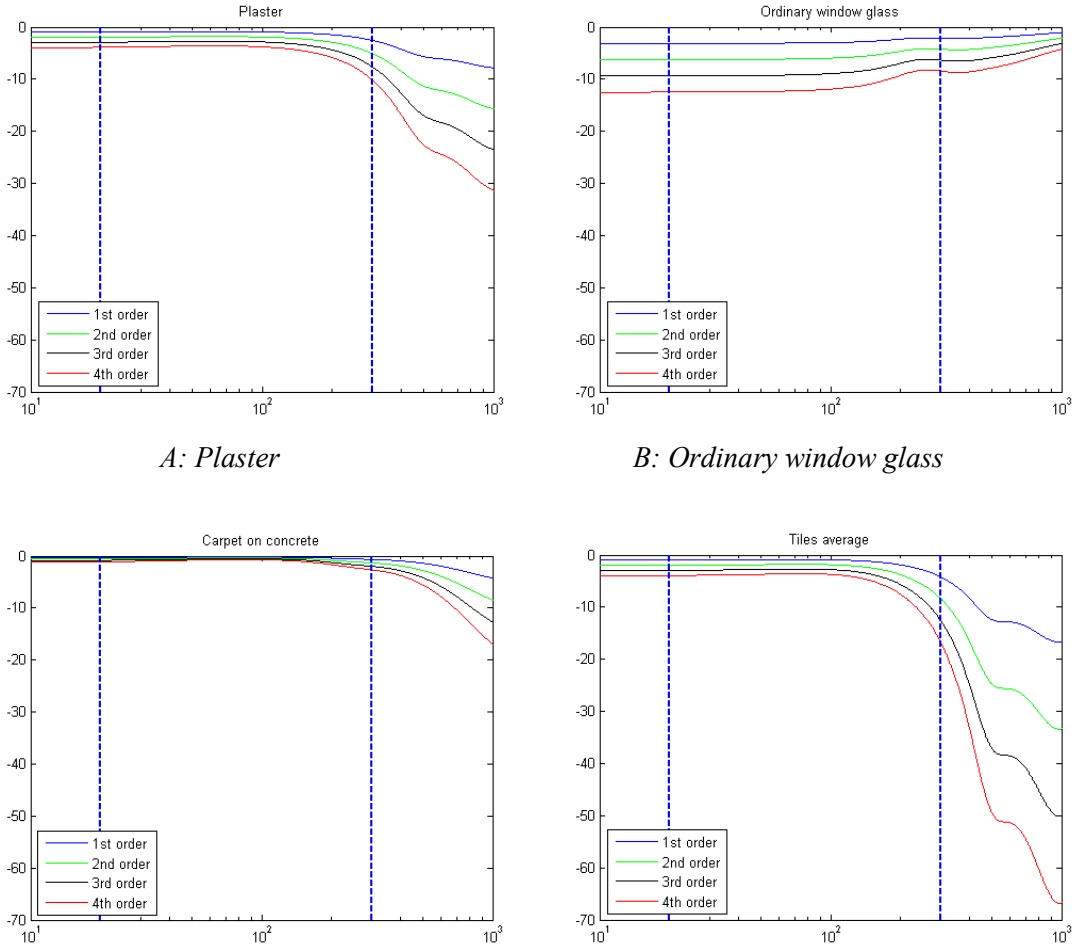
Both of the proposed methods use a maximum likelihood search. This was chosen because the problem was unknown. It was considered that there may be many local minima in the error function so it was prudent to conduct full searches to begin with, and later investigate more efficient search methods.

7.3.1 Filter library

During a maximum likelihood search, the filter response which best matches the response of a particular image is selected. Therefore, ideally, the library of filters would contain all possible filter responses; however, there is a trade-off between the number of the library options and the computational load of the search. The situation is more complex with the search for a filter for every image because high order images will have responses which are cascades of multiple surfaces responses, something that could be avoided if only first order images were being estimated.

Table 2 in Chapter 4 shows the frequency responses of various materials, taken from published material. The materials in the real room experiments have been classified as ‘plaster’ for three walls, ‘ordinary window glass’ for the other wall, (the end wall was made from multiple materials (see Figure 45), but was predominantly glass) ‘heavy carpet on concrete’ for the floor and ‘average tiles’ for the ceiling.

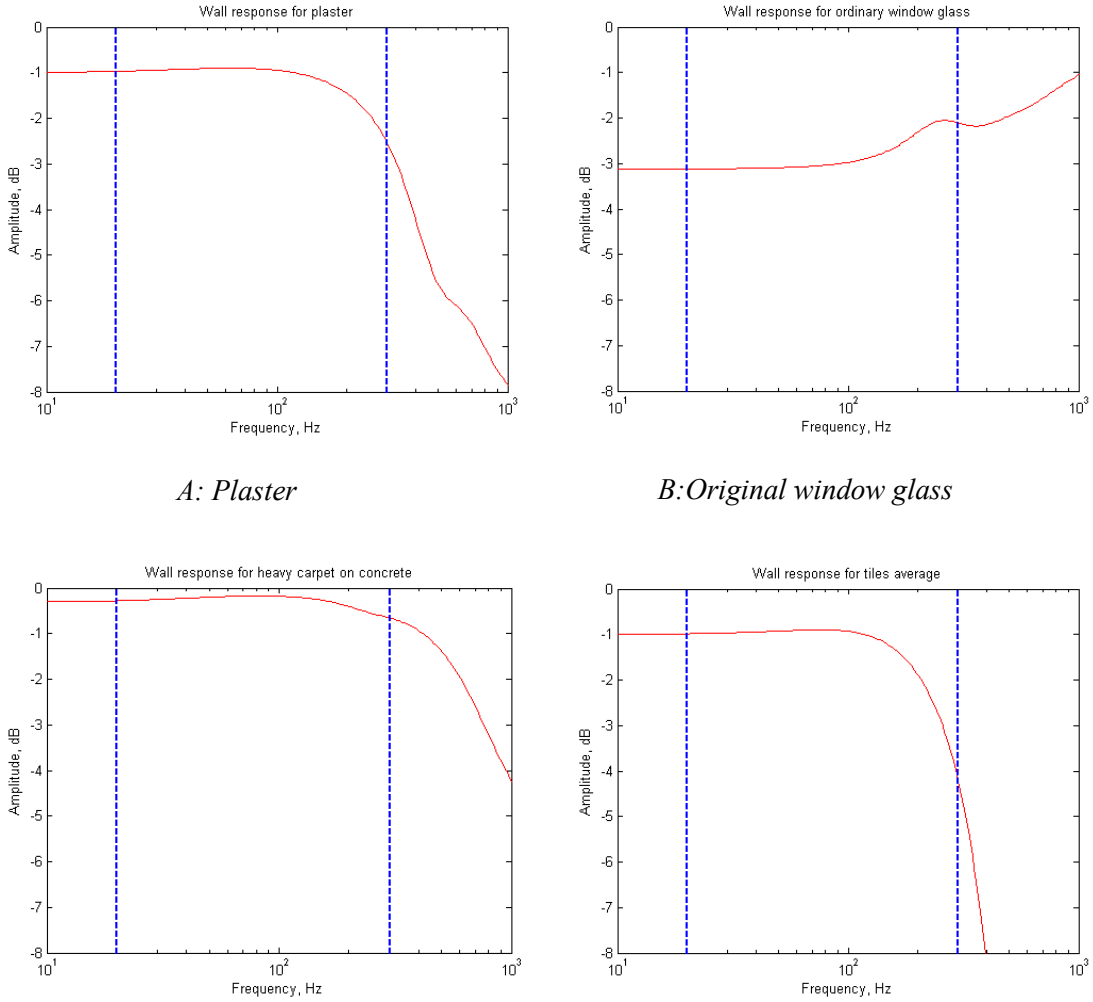
Figure 62 shows the absorption frequency responses of the four materials present in the room investigated. Each plot shows how the absorption response changes with the filter order for that material, e.g. the 3rd order plaster responses shows the frequency response of an ideal source reflected off three identical plaster surface and assumes no sound energy passing through the material. Clearly a large dynamic range in the filter options is required if there is to be a suitable filter available to match a high order image. Figure 63 shows only the 1st order response for clarity. The vertical lines in both figures indicate the bandwidth over which the surface responses were estimated, i.e. 20Hz to 300Hz.



A: Plaster *B: Ordinary window glass*

C: Carpet on concrete *D: Average tiles*

Figure 62: Absorption frequency responses for 1st to 4th order reflections. Vertical blue dashed lines represent the lower and upper bandwidth of interest (20-300Hz)



A: Plaster *B:Original window glass*

C: Carpet on concrete *D: Average tile*

Figure 63: Absorption frequency responses of the four materials. Vertical blue dashed lines represent the lower and upper bandwidth of interest (20-300Hz)

These responses did not exactly match the materials in the room and the highest order image within the estimated reflections origins could not be deduced from the measured data or estimations. It was therefore decided to implement a filter library which was a simplified response over a number of frequency and gain options rather than calculate a library from cascaded filters based on the responses in Table 2. This also had the advantage that the solution was generalised and could work with many

different room materials or objects within the room without the algorithm needing to be told what materials were present in the room. Different general filters were investigated, but in an attempt to keep the search computation time manageable a filter with three options was selected: low frequency gain, cut off frequency and high frequency gain. Two examples are shown in Figure 64.

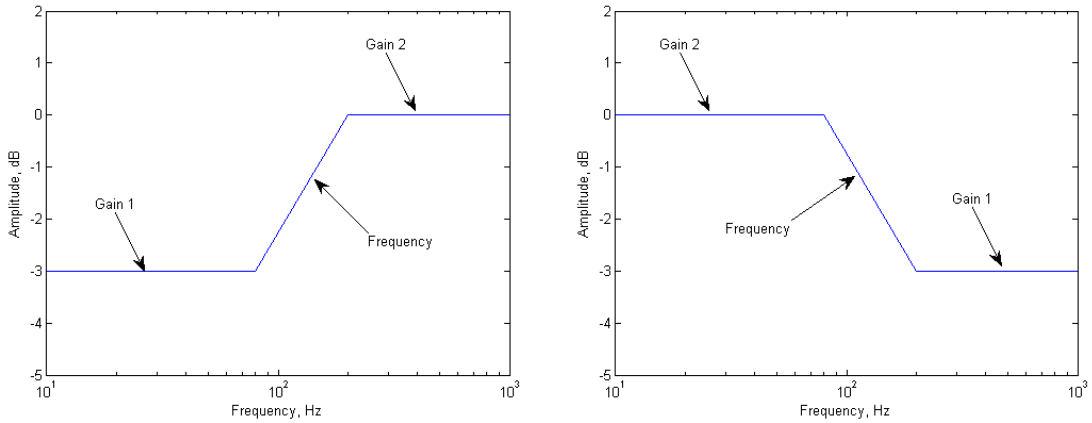


Figure 64: Filter model used for ML filter library

Before the maximum likelihood search can begin the actual response and the estimated response must be normalised. This removes the DC gain and ensures that the gain options of the library filters will cover the actual gains of the actual data. However, the estimated response is calculated only from the estimated image position, with no gain information at low frequencies at this point. The first image, assumed to be the source, was given a gain of 1 whilst the remaining images were given a gain of 0.5 (-3dB). An example of the actual and normalized estimated time domain responses, from the simulation of experiment 6, is shown in Figure 65. Because there are more close together images later in time the maximum of the estimated response was often not the source, meaning that a positive gain was required for some earlier

images. The gain options at both the high and low frequency were therefore selected to be from 10dB to -40 dB, in 2dB steps. Sometimes an image was falsely estimated by the direction of arrival processing, where an image was not really there. To remove this image a gain option of -80dB was also included so if an image did not contribute to reducing the difference between the estimated and actual response it could be attenuated to a level where it did not affect the estimated response significantly. The frequency options were chosen to be 80Hz to 400Hz in 1/3rd octave steps. This resulted in 4941 different filter options in the library; note that this excludes filters which have an identical response to an existing entry. For example, a filter with a low and high frequency gain of -4dB, is independent of frequency. The filter options selected were considered to be a sensible trade-off for accuracy and computational load. To reduce the computational load further the data was downsampled to 8kHz.

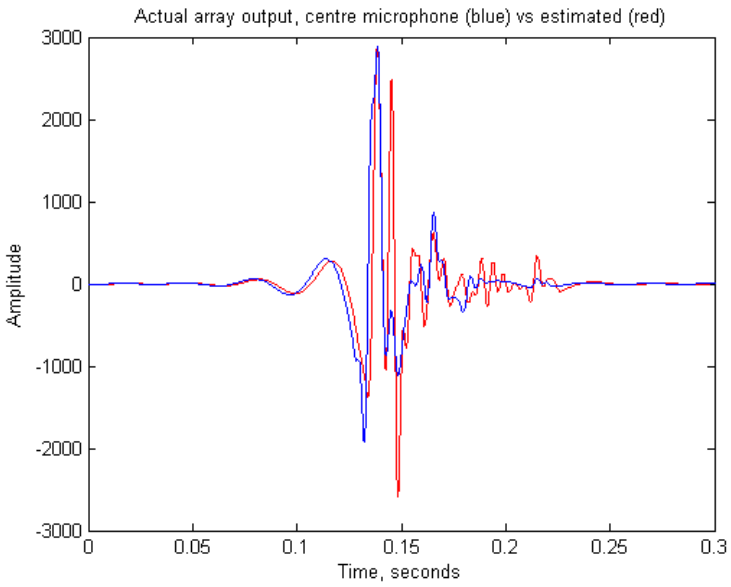


Figure 65: Time domain responses at the array's centre microphone, experiment 6, array position D, simulated data.. Actual (blue) and estimated (red)

7.3.2 ML search

Clearly with 4239 filter options to search and typically 350 image positions estimated by the spatial processing it would take a very long time to search all filter combinations for all images. It was therefore decided to limit the search to estimate the response of one image at a time and repeat this search until an individual response for each filter had been estimated. The image selected to have its' response optimized in a particular iteration was chosen by identifying the closest image to the maximum in the difference between the actual and estimated response. Another filter was applied to the rest of the images whose responses had not been estimated, thus minimising the effect of these images which overlap in time with the filter being optimized. There are, therefore, two different filters being tested in each iteration. Once an image had been selected and the search had selected the best matching filter, its response was used in the estimated response at the array. Therefore the difference between the estimated response at the array and the measured response was gradually reduced as the responses for individual images were being estimated. A block diagram of the search is shown in Figure 66.

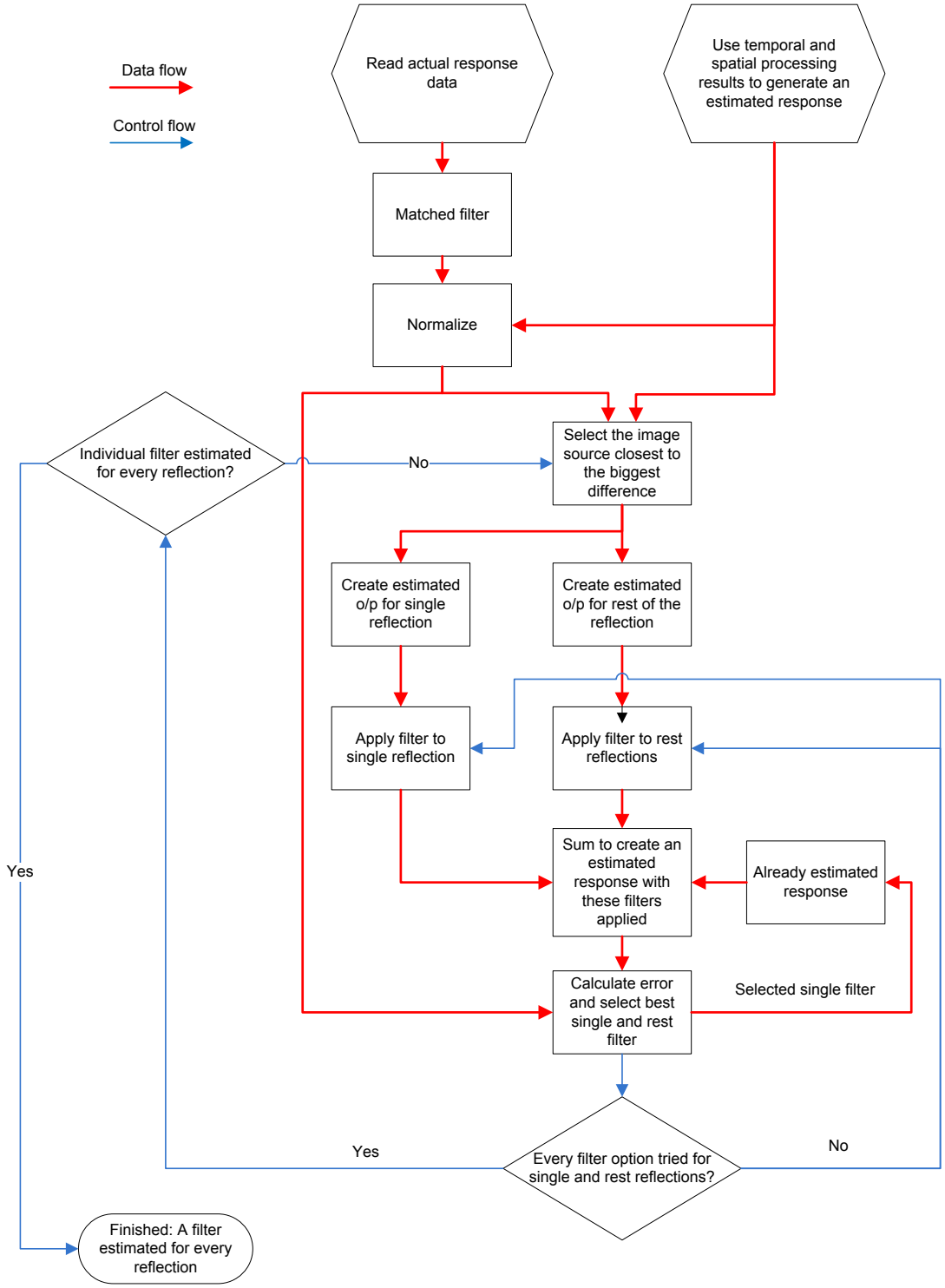


Figure 66: Maximum Likelihood search flow chart

All 4941 filter options were tested for each image response and the one which minimised the difference between the actual response at the array and the estimate was selected. Two error criteria were investigated:

a) The sum of the difference squared of the real parts of the measured and estimated response. (error = sum((real(measured) - real(estimated))^2))

a) The sum of the absolute difference between the measured and estimated responses. (error = sum(abs(measured - estimated)))

Tests indicated that more accurate results were produced when criteria a) was used.

7.4 Results:

The above processing was applied to measured and simulated room experiments 5 and 6, for both arrays (the setups described in Chapter 6). Figure 65 shows the time domain responses prior to the spectral processing. The time domain and frequency domain results from the ML processing for the simulation of experiment 6, array position D, are shown in Figure 67 and Figure 68. Figure 69 shows the actual and estimated time response for measured data prior to the spectral processing, whilst Figure 70 and Figure 71 show the equivalent results from the measured data. It is clear that in the measured case the reverberation lasted longer in time than for the simulation. Simulations were performed up to 100ms because this included all first, second and third order images which were initially considered sufficient. Increasing the duration of the simulations and refinement of the surface attenuations would be necessary to increase the accuracy of the simulations. It is unclear if the temporal and spatial processing would be able to resolve reflection locations if a longer duration was modelled. However, because of problems with the maximum likelihood search,

as described in the following section, it was decided that increasing the accuracy of the simulations was a lower priority and has therefore been left as further work due to time constraints. It can be seen that with the simulated data the time domain estimation was very close and the overall frequency response was similar up to 100Hz. With the limited time duration processed, the measured data also indicated that the process was indeed creating a more accurate estimate of the response at the array.

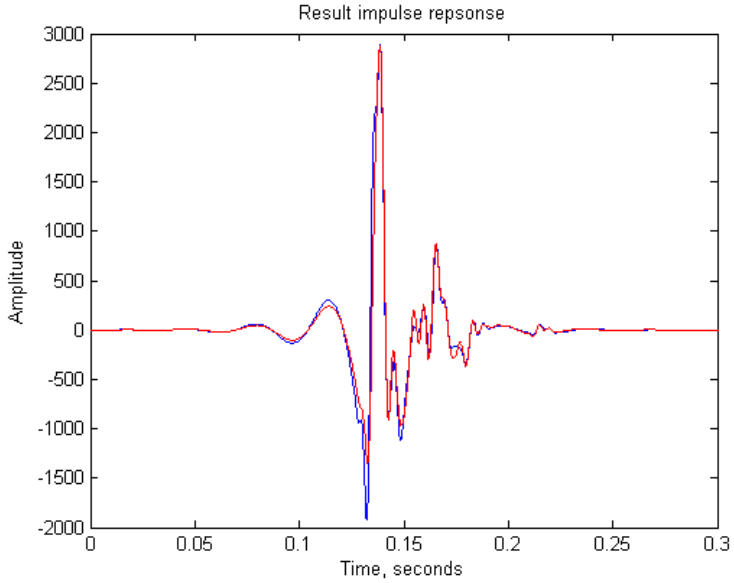


Figure 67: Time domain results from Maximum Likelihood search. Experiment 6, array position D, simulated data. Actual (blue) and estimated (red)

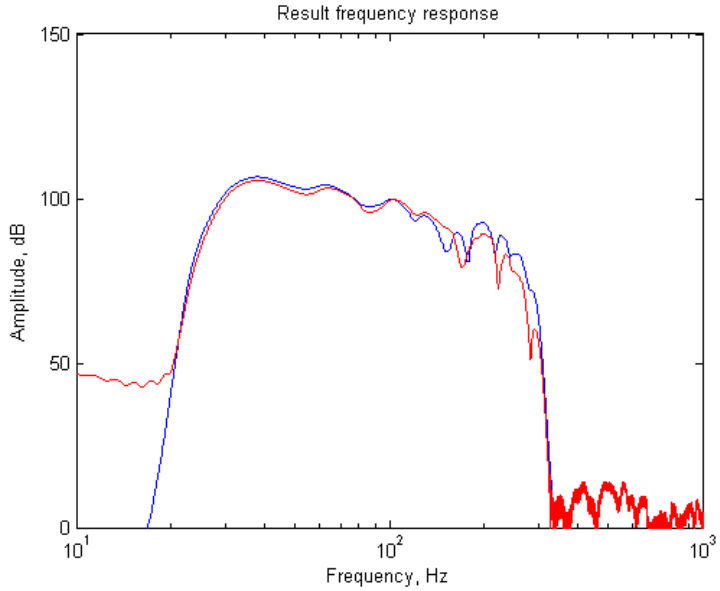


Figure 68: Frequency response results from ML search. Experiment 6, array position D, simulated data. Actual (blue) and estimated (red)

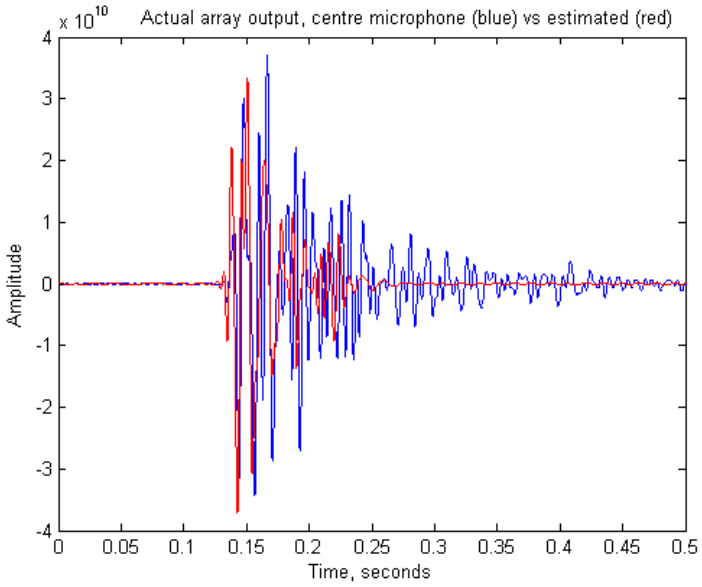


Figure 69: Time domain responses at the array's centre microphone, experiment 6, array position D, measured data. Actual (blue) and estimated (red)

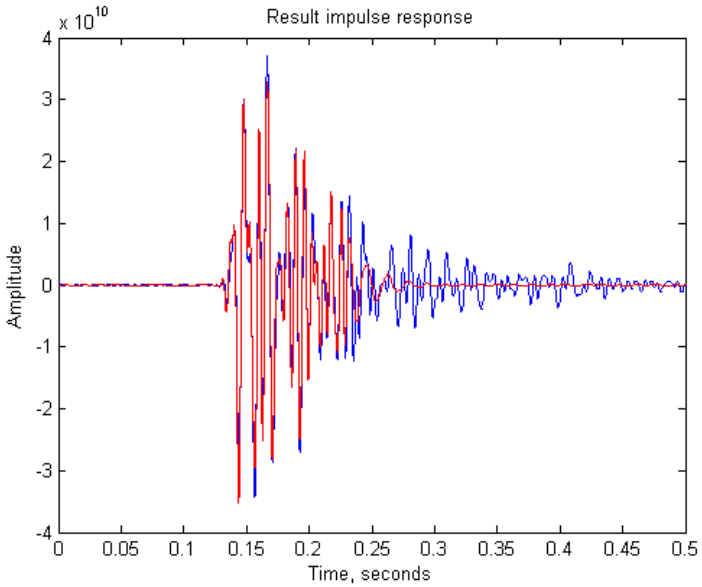


Figure 70: Time domain results from ML search, experiment 6, array position D, measured data. Actual (blue) and estimated (red)

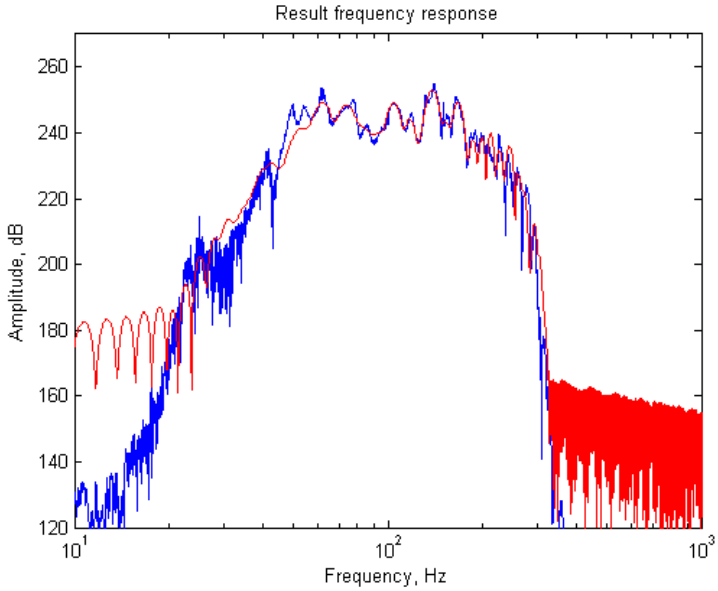


Figure 71: Frequency response results from ML search. Experiment 6, array position D, measured data. Actual (blue) and estimated (red)

7.4.1 Problem identification

Processing the data from both array positions in each experiment allowed the impulse response to be estimated at different locations within the room. The method used to combine the results from both arrays and calculate the predicted impulse responses is described in Chapter 8. The predictions were very poor and deviated greatly from expected.

In an attempt to isolate which part of the processing had failed, the frequency estimation was performed with a test set of images. A room simulation was used, as before, but rather than the estimated response being generated from a set of image positions estimated by the temporal and spatial processing, the exact locations from the simulator were used. Therefore, the only unknown was the set of frequency responses. An estimated response of individual images could be compared to the actual response for individual images, allowing the performance of the search to be assessed. This confirmed that the spectral processing was the problem.

Following this, investigations were performed in a testbed program where only three sources were present, allowing faster analysis of the problem. This proved that the implementation was correct; the filter which minimised the time domain difference was being selected. However when a phase shift was introduced, either by slightly changing the arrival time of an estimated image, or by summing responses which overlapped in time, the filter being selected had a different frequency response to the actual frequency response. Figure 72 shows an example test with the three signals, all modelled as 1st order reflections from plaster walls. The resulting estimate after

spectral processing can be compared to the actual signal in Figure 73, where it can be seen that the time domain responses look very similar; however, the frequency responses of the filters were very different. Figure 74 show the frequency responses for the actual filter, the estimated filter, by the ML processing and the optimum filter taken from the filter library. Clearly the ML search was not selecting the correct filter to match the frequency response. A modified maximum likelihood search was performed, selecting the best filter in the frequency domain: this revealed that there was a phase difference between the estimated and actual response.

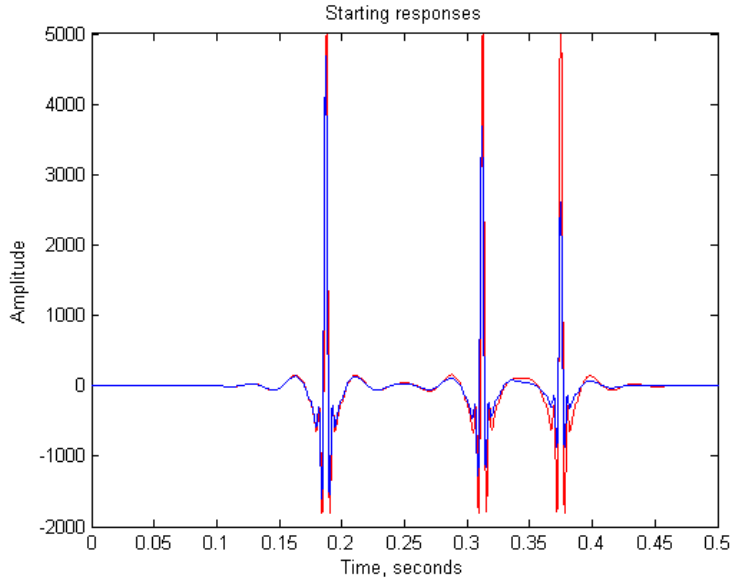


Figure 72: Wall ML testbed starting signals. Actual (blue) and estimated (red)

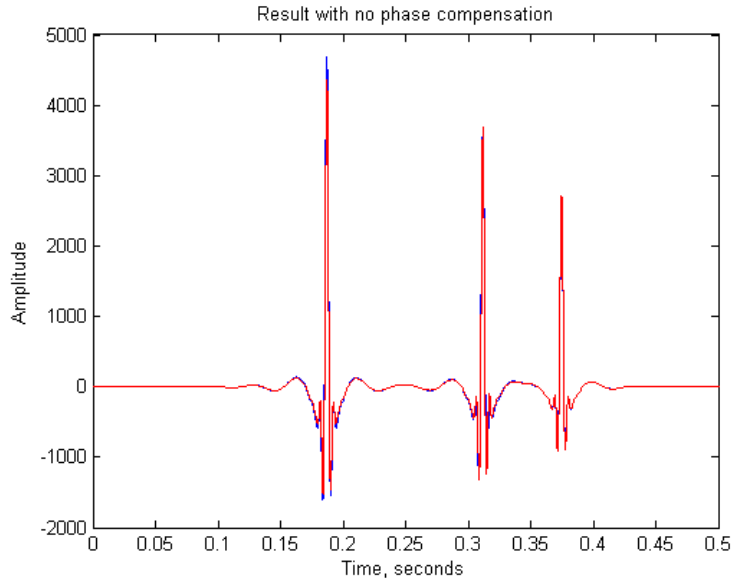


Figure 73: Wall ML testbed results, with no phase compensation. Actual (blue) and estimated (red)

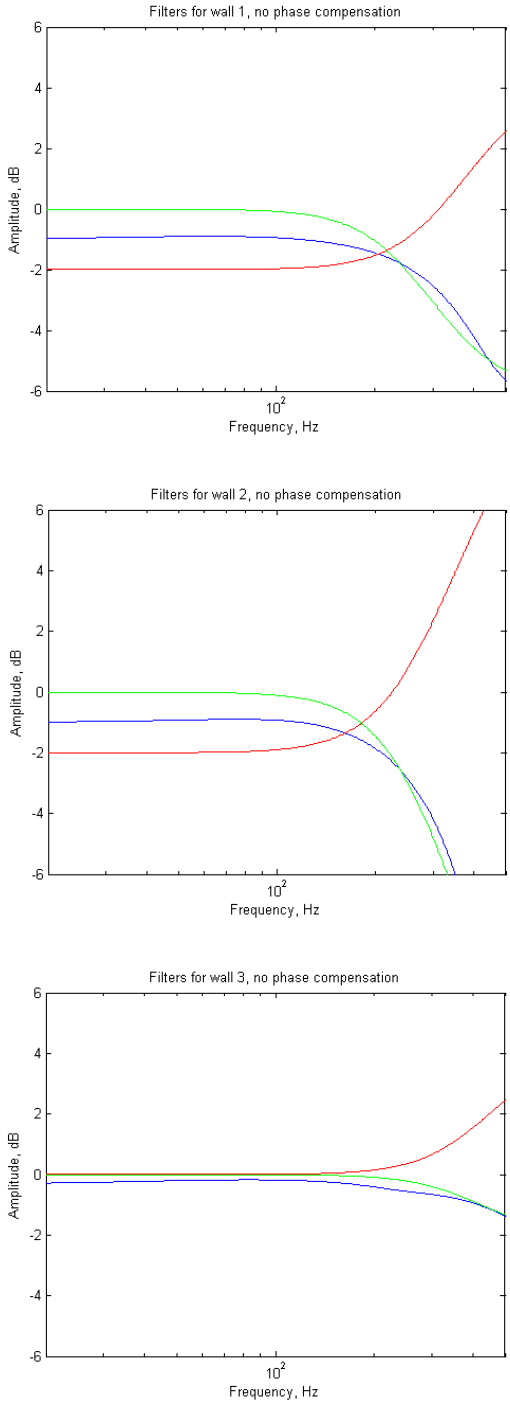


Figure 74: Wall ML testbed, results for walls 1 to 3 with no phase compensation. Actual (blue), optimum (green) and estimated (red)

7.4.2 Phase enhancements

The filter selection could not be performed in the frequency domain for a room response because there would be no relationship to time, and therefore to individual images. When selecting the best filter in the time domain the phase shift therefore also needed to be optimized along with the filter's parameters. Rather than add a phase shift to the filter library it was more computationally efficient to estimate the phase shift which would best line up the filtered estimated response and the actual response for that particular filter. This was implemented by filtering the single image's response, as done previously, but then cross correlating the estimated and actual responses over a small range; from the maximum the required shift was calculated and the estimated response shifted accordingly before the error was calculated.

In test programs this did reduce the problem, as shown in Figure 75, and whilst the filter estimates are not ideal, they were much closer to the optimum responses.

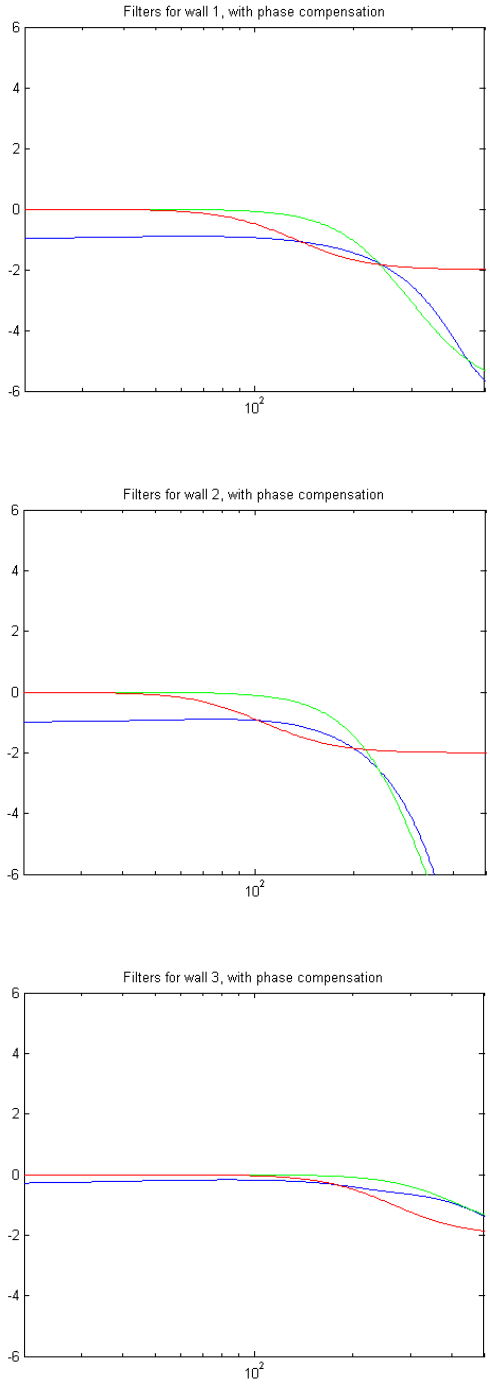


Figure 75: Wall ML testbed, results for wall 1 to 3 with phase compensation. Actual (blue), optimum (green) and estimated (red)

Even with efficient programming the modifications greatly increased the computation time. The computer used was an Intel quad core 2 duo running at 3GHz with 8Gbytes of RAM. Matlab's parallel for loop was used to distribute the search over all cores. Without the phase computation each filter took about 25 minutes; with the phase modification this increased to 3 hours. To verify the process with the exact image locations 384 images were required thus taking 48 days of continuous processing per array setup. This high computational time was prohibitive in running a full set of room data. It was therefore chosen to run one setup with exact image locations to prove the concept, allowing the individual image responses to be compared and therefore assess the algorithm's accuracy.

The final time and frequency domain results are shown in Figure 76 and Figure 77. It can be seen that the actual and estimated responses are not an exact match. Looking at the responses of the individual image filters it was clear that although adding the phase compensation had improved the situation, the search was still not selecting the correct filter response. Figure 78 shows the filter response for the first 6 images estimated. It is thought that there are still two factors which are limiting the accuracy. First, were the limited options in the filter library, since there was a trade-off between processing time and resolution, such that if more processing power had been available different library options could have enhanced the results. The second limiting factor identified is the overlap of signals. Because the algorithm used one filter for the image under investigation and one for all the rest of the un-estimated images it is likely that the effect of the un-estimated images was not being suppressed sufficiently for an exact individual filter to be estimated for the particular image under

investigation. Due to time constraints this problem was not solved, and it is quite possible that the problem cannot be solved, particularly with existing computers. Nevertheless, the following are some suggested possibilities that could be tried:

- Employ multiple filters to estimate, and therefore further suppress the effect of the un-estimated images.
- Refine the estimated filters by processing the data multiple times, gradually homing in with a finer resolution filter library to increase the resolution and allow a filter to be re-estimated after the other images have been estimated.
- Find a way of identifying an image's order, therefore simplifying the search because higher order responses could be calculated from an estimate of the lower order images which created them.

Looking at the individual filter responses in Figure 78 it is likely that the second option would improve the situation most significantly, since by re-estimating a filter's response it is likely that the effect of other images are suppressed and a more accurate estimation can be make.

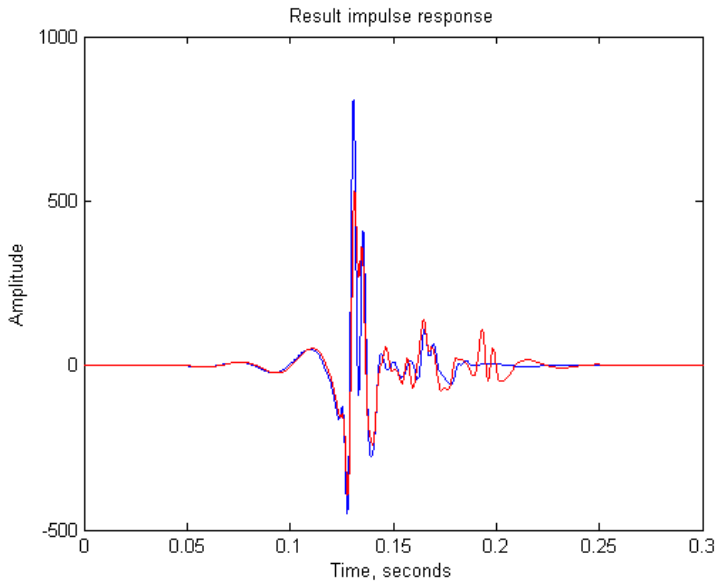


Figure 76: Time domain results from ML search with exact image positions and phase compensation. Actual (blue) and estimated (red)

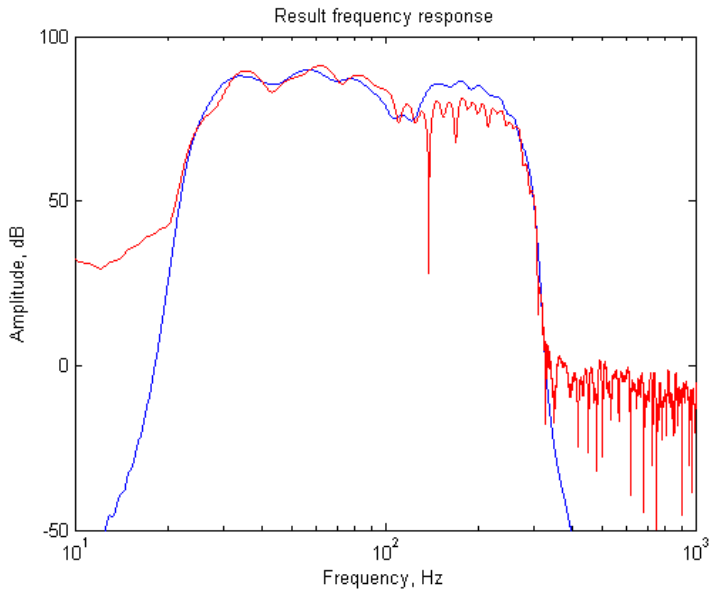


Figure 77: Frequency domain results from ML search with exact image positions and phase compensation. Actual (blue) and estimated (red)

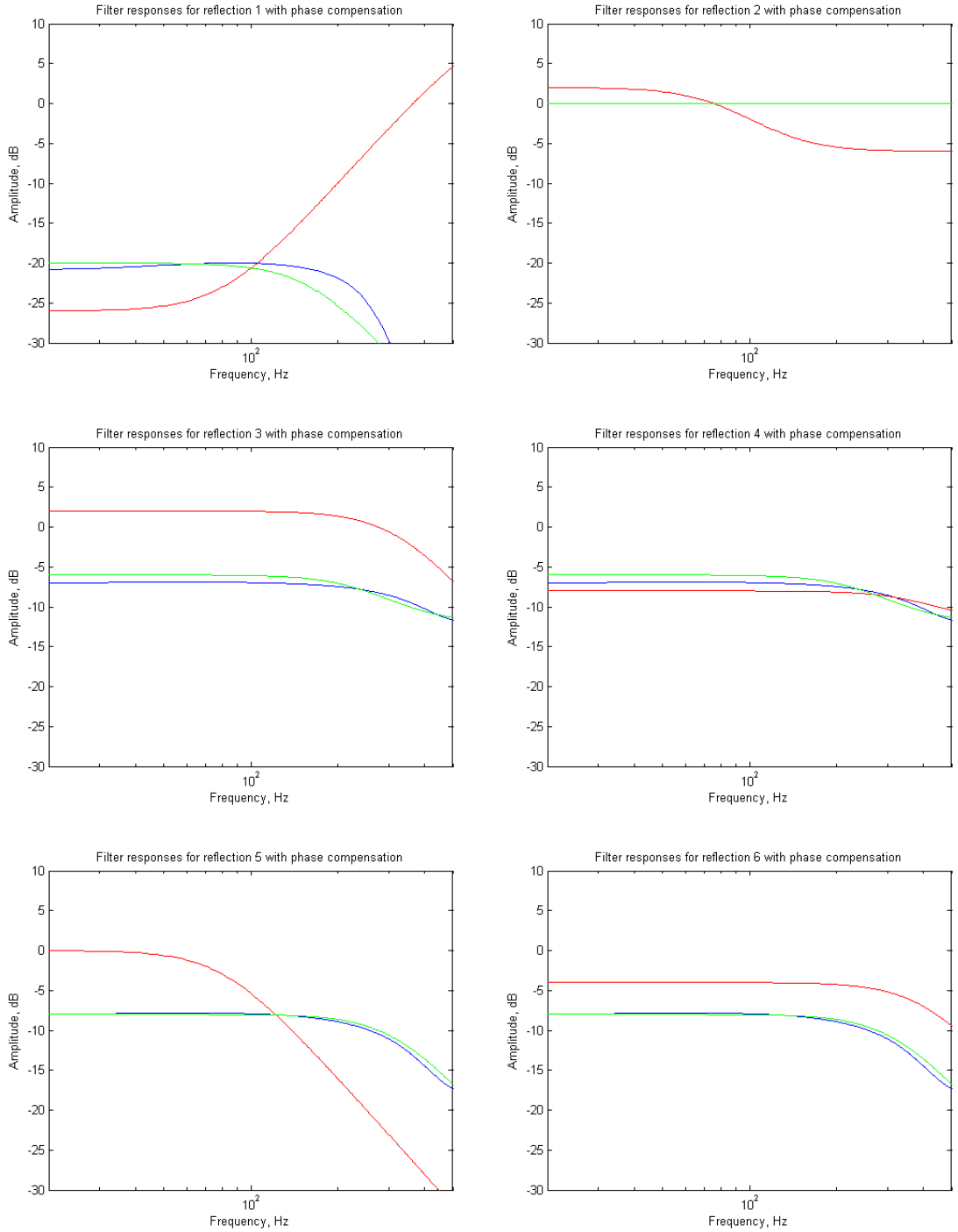


Figure 78: Filter responses for the first 6 images estimated: actual (blue), optimum (green) and estimated (red).

7.5 Conclusions

The first part of this chapter covered methods of measuring the low frequency response of the room. A logarithmic swept sine wave was selected as the most appropriate method. This was used as the target function in subsequent processing that minimised the difference between the estimated and measured responses at the array locations.

Different methods of estimating the responses of individual reflections were investigated, but it proved more complicated than had first been thought. Solving the problem using adaptive filters would have been computationally efficient and would have been ideal for making periodic updates of the estimations because the search could have started with the last estimation. However, this method was unsuccessful when the time domain responses from different filters overlapped. Attempts to remedy the situation with windows applied to the weighting function and filter implementation changes failed.

A maximum likelihood search was, therefore, performed. A two-filter approach was taken, one for the individual reflection and one for the remaining reflections to improve the efficiently of the search. A library of filter options was generated using a simple filter model with 3 parameters. Tests without the inclusion of a phase parameter also failed, despite promising first impressions. In hindsight further verification of the algorithm was required before all the simulated and measured data was processed. The inclusion of a phase parameter improved the accuracy of the

individual filter's frequency response estimate; a similar response to the optimum was selected in the testbed simulator. When a room response was analysed the quality of the estimated degraded, suggesting that the approach of using just 2 ML filters with one pass was not sufficient. Unfortunately the very high computational cost of the algorithm prevented it from being refined or real room data being reanalysed. The main difficulty here was that there was simply insufficient time available.

Efforts were made to optimise the maximum likelihood search, both in terms of memory management to allow parallel processing and in reducing the computational load within each iteration. However, the solution was not fast enough to allow the extensive testing which would be required before the scope of the search could be reduced. Further work is therefore required to increase the accuracy of the algorithm and gain confidence in the maximum likelihood search prior to investigating optimised search algorithms.

Chapter 8 System Performance

It had been intended that this chapter would cover the integration of the temporal, spatial and spectral processing stages as well as the calculation of an estimated, band limited, impulse response at different locations within the room. These could then have been compared to measurements at spot locations or exact simulations to judge the performance of the complete system. Unfortunately problems arose within the spectral processing, (see Chapter 7) such that although a solution has been proposed and partially demonstrated it has not been possible to process the room data. This chapter therefore covers the concepts which could be employed if future processing were to prove successful.

8.1 Combining array estimates

The data captured from the two arrays placed within the room has been processed independently for the temporal spatial and spectral processing. The next stage would therefore be to combine the estimations from both arrays. Since each array can see the images in front of it, but not behind it, the images behind array A are estimated by the opposite array, B, and vice versa. Figure 79 shows a rectangular room with the two arrays and the regions which are coved by only one array.

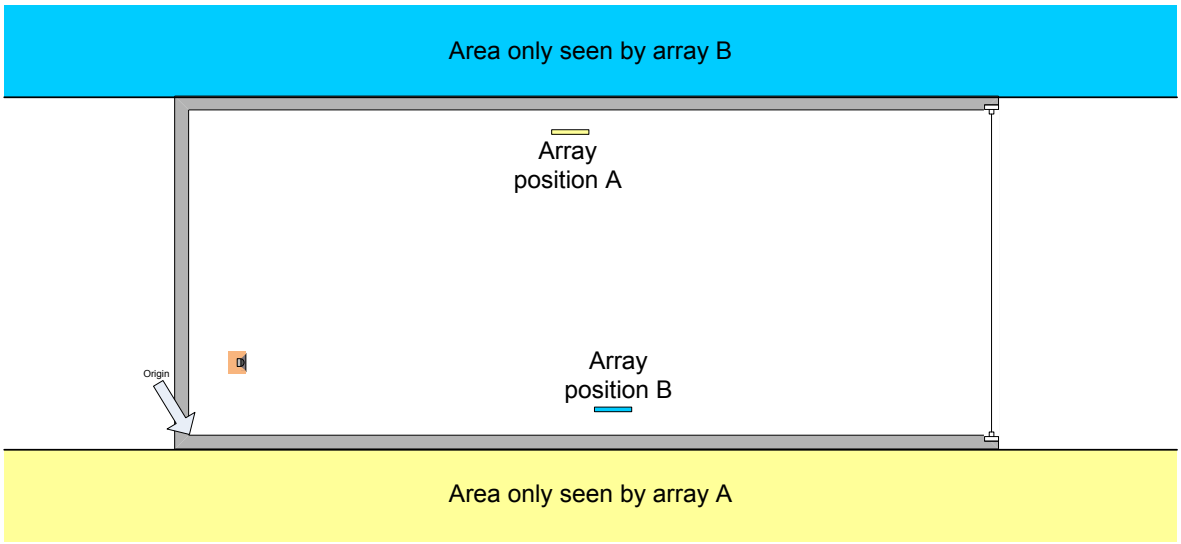


Figure 79: Areas covered by only one array

The images which fall within the white area should have been identified and estimated by both arrays. The simplest method of combining these results is to divide the room in half. It is known that the direction of arrival estimate accuracy is greatest when the signal arrives at broadside and weakest when the signal arrives at endfire, (see section 6.4.1). Therefore the array furthest away from the images will, theoretically, give the highest accuracy. The images identified by the further away array could therefore be kept and the images nearer array could be discarded. This concept is illustrated in Figure 80.

An alternative approach could be to classify the images which have been identified by both arrays and originated from the same reflection source. The reflection's parameters could then be either averaged, or the strongest image selected and the others discarded. However identifying which estimates are pairs is likely to be a difficult task, so the simple method described above is a recommended starting point.

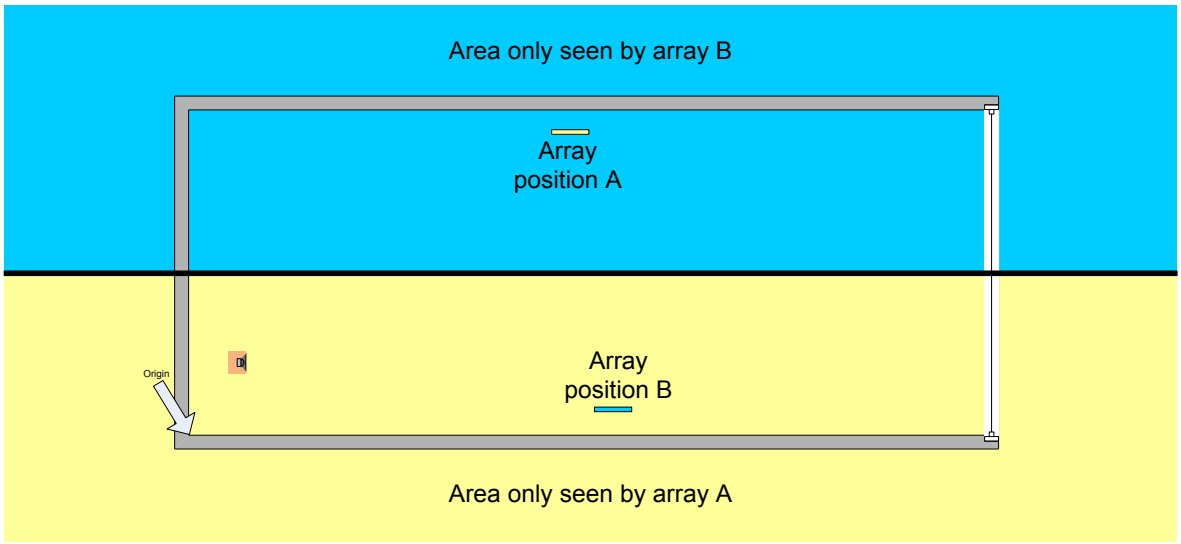


Figure 80: Areas seen by array A and B, using a simple method of dividing the room

8.2 Predicting the response at a remote location

Once the arrays have been combined all the information is available to create an acoustic model of the room, based on the estimated parameters of the reflections. Referring to the image source method, described in Chapter 4, if the room was rectangular all images can be seen at all locations within the room, and thus all images contribute to the impulse response no matter where in the room the sound is being predicted. If the room's shape is more complex then the same criteria for identifying contributing images can be used as in the image source method. The band limited response at locations anywhere within the room could then be calculated.

8.3 Using the predictions

The accuracy of the system could be assessed by comparing the estimated response at spot locations within the room to real measurements. Assuming that the estimated sound is considered accurate enough it can be used in a room correction system. A variety of room correction methods were discussed in Chapter 2: essentially the signals to the loudspeakers in the room are filtered in some way that will improve the perceived sound quality at the listeners' locations, thus increasing their enjoyment of the music or film they are listening to.

8.4 Conclusions

Unfortunately, because of problems in the reflection's frequency response it was not possible to fully verify the system. Instead this chapter has briefly discussed how successfully estimated data could be used to calculate the sound at any location within the room. The process starts by combining the estimates from the two arrays, and a basic method for doing this has been presented. Calculating the response within the room follows the same procedure as in the image source method of simulating the acoustics of a room, as discussed in Chapter 4.

Chapter 9 Conclusions

9.1 The project's aim

The aim of this project was to design a measurement tool which could be used to predict the sound at any location within a room, using small microphone arrays placed on the walls of the room. The design of the arrays was to be as aesthetically unobtrusive as possible. The application of the predicted sound was an adaptive room correction system, for domestic use, which could increase the perceived sound quality of music and films. A review of the literature in Chapter 2 revealed that it was the low frequencies which were most in need of electronic filtering to reduce the effect of room modes, from 20Hz to about 500Hz. This was therefore the bandwidth over which the prediction was required to be accurate

9.2 Summary of the thesis

As such a problem has not been tackled before, this thesis is therefore unique and the solution has had to be developed from the ground up. A system was proposed in Chapter 3, which was designed to meet the aims of the project. By analysing the reverberation, to estimate the parameters of individual reflections, a mathematical model of the room could be generated from acoustic measurements. A room simulation was produced, based on the image source method described in Chapter 4, which allowed the algorithms and concepts to be quickly evaluated and developed,

with all developments backed up with real measured data. Three processing stages were required: temporal to separate the received signals in time; spatial to estimate the location of the reflections' origin; and spectral to estimate the individual reflection's frequency response. Although the project has not been able to achieve its main overall goal it has nevertheless been successful in many respects. Firstly the temporal processing successfully identified the arrival time of the reflections, deep into the reverberation, through careful design of the transmit signal and its post processing as discussed in Chapter 5. Following this, Chapter 6 reviews the spatial processing where the majority of the research effort was focussed.

Secondly, I designed the microphone arrays with a bespoke geometry to minimise the hardware requirements and produce a compact array. This has been implemented to allow measured data to be recorded and processed, a significant differentiation from a large amount of the published research on direction of arrival, which typically only presents simulated results. The captured reverberation signals were highly correlated and overlapped in time. This was a significant challenge which is not often covered by the spatial processing literature where reverberation is often ignored, or is far less severe than is encountered in a domestic room.

The third major achievement was to successfully analyse the reverberant signals, resulting in the majority of reflections' angles of arrival being estimated with good accuracy. Fourthly, an efficient processing architecture was used to ensure that the whole two dimensional space did not have to be searched, thus reducing the computational demands significantly. The results demonstrate that the azimuth angle was consistently estimated with good accuracy, but that sometimes the elevation angle

could not be resolved correctly. A number of possible changes are suggested in Chapter 6 which are expected to increase the elevation accuracy significantly. The problem of estimating the frequency response of a reflection was tackled in Chapter 7 and a number of solutions were investigated. However, the complexity of the problem meant that an extensive search algorithm was required, but due to its high computational load the development and verification of this solution has been limited. Finally, methods of combining the results from multiple arrays and the use of the estimated data were introduced in Chapter 8.

This research project has been presented at two Audio Engineering Society (AES) conferences, AES31 as a poster and AES32 as a talk, both covering the proposed system, the temporal and spatial processing, including preliminary results. Copies of the papers can be found in Appendix C and D.

9.3 Future work

The weakest part of the spatial processing was the consistency of the estimation of the elevation angle. It is thought that either the 24 microphones in the circle array were not sufficient, meaning a larger circle of more microphones, or a planar array might be required to increase consistency. The limitation of the existing design is evident in the simulated and measured case. The accuracy of the measured results is likely to have been increased if array calibration had been successful. With the equipment available, however, it was not possible to accurately measure the microphones' angular responses. This problem might be sidestepped by increasing the overall size

of the array and using a lower transmit frequency, to say 4kHz so that the microphones' responses are more omnidirectional.

A greater limitation of the project was, however, the spectral processing, limited by the processing power required. It is thought that a more efficient solution than the maximum likelihood search is possible, but without the maximum likelihood search working and the complexities fully understood, a reduced search is risky because the optimum solution might be missed. It might be that this problem cannot be solved, at least until faster computers are available. Suggested avenues for investigation are as follows:

- Employ multiple filters to estimate the responses of the un-estimated images.
- Refine the estimates filters by processing the data multiple times
- Gradually home in with a finer resolution filter library to increase the resolution of the filter
- Find a way of identifying an images order, thereby simplifying the search because higher order responses could be calculated from an estimate of the lower order images which created them.

Successfully estimating the frequency response of the reflections would allow the full system to be verified. It is expected that at this stage other issues would arise and further refinements in the processes would be required, for example filtering out false estimates. However it is believed that the work done so far means that if the spectral processing can be achieved with enough accuracy, predicting the sound within the room up to at least 150Hz will be possible and that refinements to the spatial processing could increase this towards the target of 500Hz. This would be sufficient

to design a set of room correction filters to successfully correct the room response and therefore improve the users' listening experience.

References

Acoustics and Sonar Group, 2008, *Continuing education course in Underwater Acoustics and Sonar Systems*, University of Birmingham.

Aichner, R., Buchner, H. and Kellermann, W., 2005, *On the Causality Problem in Time-Domain Blind Source Separation and Deconvolution Algorithms*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, PA, USA.

Allen, J.B. and Berkley, D.A., 1979, Image method for efficient simulating small-room acoustics, *Journal of the Acoustical Society of America*, vol 65, no 4, pp. 943-950.

Ambisonics (online), Available at www.ambisonic.net/ (Accessed 2nd February 2009).

Ando, Y., 1998, *Architectural Acoustics Blending Sound Sources, Sound Fields, and Listeners*, Springer-Verlag New York Inc.

Angus, J.A.S., 2001, The effects of specular versus diffuse reflections on the frequency response at the listener, *Journal of the Audio Engineering Society*, vol. 49, no. 3, pp. 125-133.

Audyssey (online), Available at www.audyssey.com (Accessed 15th August 2008).

Avis, M.R., Fazenda, B.M. and Davis, W.J., 2006, Thresholds of detection for changes to the Q factor of low frequency modes in listening environments, *Journal of the Audio Engineering Society*, vol. 55, no. 7/8, pp. 611-622.

Beard, A.A. and Atkins, P.R., 1998, Position measurement of vertical line arrays using multi-dimensional scaling, Proceedings of the Institute of Acoustics, Sonar Signal Processing Conference December, pp.137-145.

Bell, A.J. and Sejnowski, T.J., 1995, An information maximization approach to blind separation and blind deconvolution, *Neural Computation*, vol. 7, no. 6, pp. 1004-1034.

Beranek, L.L., 1996, *Concert and Opera Halls: How they sound*, Published for the Acoustical Society of America, through AIP Press.

Berkhout, A.J., Vries, D., Hemingway, J.R. and Griffioen, A., 1988, *Experience with the Acoustical Control System ACS,* Audio Engineering Society, 6th International Conference.

Berkhout, A.J., Vries, D.de and Sonke, J.J., 1997, Array technology for acoustic wave field analysis in enclosures, *Journal of the Acoustical Society of America*, vol. 102, no. 5, pp. 2757-2770.

Bharitkar, S. and Kyriakakis, C., 2001, *A cluster centroid method for room response equalization at multiple locations*, IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics.

Bharitkar, S. and Kyriakakis, C., 2002, *Perceptual multiple location equalization with clustering*, IEEE 36th Asilomar conference on Signals, Systems and Computers.

Bharitkar, S. and Kyriakakis, C., 2004, *Multirate signal processing for multiple listener low frequency room acoustic equalization*, IEEE 38th Asilomar conference on Signals, Systems and Computers.

Blunt, S.D. and Gerlach, K., 2004, Adaptive pulse compression, *Proceedings of the IEEE 2004 Radar Conference*, pp. 271-276.

Bolt, R.H., 1946, Note on the normal frequency statistics in rectangular rooms**,** *The Journal of the Acoustical Society of America*, 18, pp. 130-133.

Borish, J., 1984, Extension of the image model to arbitrary polyhedral, *Journal of the Acoustical Society of America*, vol. 75, pp. 1827-1836.

Bourennane, S. and Bendjama, A., 2000, Locating wide band acoustic sources using higher order statistics, *Applied Acoustics Journal*, no. 63, pp. 235-251.

Bradwood, D.H., 1989, Multiple singal DF with coherent signals using a general array, *IEEE Colloquium on Passive Direction Fnding*, pp. 5/1-5/4.

Bresler, Y. and Macovski, A., 1986, Exact maximum likelihood parameter estimation of superimposed exponential signals in noise, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 5, pp. 1081-1089.

Campbell, D., 2004, *Roomsim,* version 3p4p1 (A downloadable Matlab program), University of the West of Scotland, http://media.paisley.ac.uk/~campbell/Roomsim/

Cardoso, J-F, 1998, Blind Signal Separation: Statistical Principles, *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009-2025.

Carr, J.J., 2001, *Practice Antenna Handbook*, McGraw-Hill Professional. 4th edition.

CEA (Comsumer Electronis Assosiation) (online), Available at www.ce.org/Press/CEAPubs/929.asp (Accessed 17th July 2008).

Cedervall, M. and Moses, R.L., 1997, Efficient maximum likelihood DOA estimation for signals with known waveforms in the presence of multipath, *IEEE Transactions on Signal Pprocessing*, vol. 45, no. 3, pp. 808-812.

Chen Y-M., 1997, On spatial smoothing for two-dimensional direction-of-arrival estimation of coherent signals, *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1689-1696.

Chen, J.C., Yao, K. and Hudson, R.E., 2002a, Source Localization and beamforming, *IEEE Signal Processing Magazine*, vol. 2, pp. 30-39.

Chen, J.C., Hudson, R.E. and Yao, K., 2002b, Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field, *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1843-1854.

Clarke, I.J., 1989, Efficient Maximum Likelihood Using Higher Rank Spectral Estimation, IEEE *Workshop on Higher-Order Spectral Analysis*, pp. 229-234.

Clarke, I.J., 1992, Multi-signal time-frequency model fitting using an approximate maximum likelihood algorithm, *Proceedings of the Time-Frequency and Time-Scale Analysis, IEEE-SP International Symposium* , pp. 269-272.

Collins, T. and Atkins, P., 1999, Nonlinear frequency modulation chirps for active sonar, Radar, Sonar and Navigation, *IEE Proceedings,* vol. 146, pp. 312-316.

Collins, T., 2004, *Implementation of a Non-Linear room impulse response estimation algorithm*, Presented at the 116th Audio Engineering Society Convention.

Colloms, M., 1999, *High performance loudspeakers*, 5th edition. Wiley.

Cops, A., Vanhaecht, J. and Leppens, K., 1995, Sound absorption in a room: Causes of discrepancies on measured results, *Applied Acoustics*, 46, pp. 215-232.

Cortell, E., 2006, Equalization in an extended area using multichannel inversion and Wave Field Synthesis, *Journal of the Audio Engineering Society*, vol. 54, no. 12, pp. 1140-1161.

Cox, T.J., D'Antonio, P. and Avis, M.R., 2004, Room sizing and optimization at low frequencies, *Journal of the Audio Engineering Society*, vol. 52, no. 6, pp. 640-651.

Cremer, L., Muller, H.A. and Schultz, T.J., 1990, *Principles and Applications of Room Acoustics: Geometrical, Statistical and Psychological Room Acoustics,* Volume I, Applied Science/ Elsevier.

Diaz, C. and Pedrero, A., 2005, The Reverberation time of furnished rooms in dwellings, *Applied Acoustics*, 66 pp. 945-956.

Dobbins, P. F., 1984, *The use of non-ideal microphones in directional arrays*, Presented at the 75th AES convention, no. 2062.

Eargles, J., 2001, *The Microphone Book*, Focal Press.

Evans, J.E., Johnson, J.R. and Sun, D.F., 1981, High resolution angular spectrum estimation techniques for terrain scattering analysis and angle of arrival estimation of coherent signals, in *Proceedings of 1st Acoustic, Speech, Signal Processing Workshop on Spectrum Estimation*, Hamilton, Canada, pp. 134-139.

Everest, A. F., 2002, *Master Handbook of Acoustics*, 4th edition, McGraw-Hill/TAB Electronics**.**

Farina, A., 2000, *Simultaneous measurements of impulse response and distortion with a swept-sine technique*, Audio Engineering Society 108th International Convention, Paris, paper no 5093.

Featherstone, W. and Strangeways, H.J., 1999, Improved multipath resolution using joint space-time maximum likelihood estimation, *IEE National conference on Antennas and Propagation,* publication no. 461, pp. 204-208.

Feng, M. and Kammeyer, K-D., 1998, Blind source separation for communication signals using antenna arrays, *IEEE 1998 International Conference on Universal* Personal Communications, vol. 1, pp. 665-669.

Feng, A., Zhao, Z. and Yin, Q., 2001, Wideband direction of arrival estimation using chirplet based adaptive signal decomposition algorithm, *IEEE 54th Vehicular Technology Conference,* pp. 1432-1436.

Fielder, L.D., 2003, Analysis of Traditional and Reverberation-Reducing Methods if Room Equalization, *Journal of the Audio Engineering Society*, vol. 51, no. 1/2, pp. 3-26.

Fistas, N. and Manikas, A., 1994, A new general global array calibration method, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP-94, vol. 4, pp. 73-76.

Forsyth, M., 1985, *Buildings for Music*, Cambridge, MA: MIT Press.

Fuchs, H.V., Zha, X. and Pommerer, M., 2000, Qualifying freefield and reverberation rooms for frequencies below 100Hz*, Applied Acoustics* 59, p. 303-322.

Gerganova, D. and Christov, D., 1990, *Computer-aided room acoustical design*, Audio Engineering Society 88th International Convention paper, No 2900.

Gerzon, M., 1990, *Why do equalisers sound different?*, Studio Sound, July.

Goodwin, M., 1997, Matching pursuit with damped sinusoids, IEEE conference on Acoustics, *Speech, and Signal Processing,* vol. 3, pp. 2037-2040.

Greenfield, R., and Hawksford, M. O., 1991, Efficient Filter Design for Loudspeaker Equalization, *Journal of the Audio Engineering Society*, vol. 39, no. 10, pp. 739-751.

Harris, F.J., 1978, On the use of windows for harmonic analysis with the discrete fourier transform, *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51-83.

Hatziantoniou, P.D., Mourjopoulos, J. N., *2003, Results for Room Acoustics Equalisation Based on Smoothed Responses*, Audio Engineering Society 114th Convention, paperNo 5779.

Haubrich, R.A., 1968, Array design, *Bulletin of the Seismological society of America*, vol. 58, pp. 977-991.

Howard, D.M. and Angus, H., 2001, *Acoustics and Psychoacoustics*, 2nd edition, Focal Press.

Hruska, G., Magrab, E.B. and Penzes, W.B., 1977, Enviromental effects on microphones of various constructions, *Journal of the Acoustical Society of America*, vol. 61, no. 1, pp. 206-210.

Institute Of Acoustics - Speaker unknown, 2007, discussion at the Institute of Acoustics 23rd annual conference, Reproduced Sound, The Sage Gateshead, 29-30 November 2007.

iTrax, by Waldrep, M., www.iTrax.com (Accessed 8th August 2008).

ITU-R Recommendation BS. 775-1, 1992-1994, Multi-channel stereophonic sound system with or without accompanying picture. International Telecommunications Union

Jackson, G.M. and Leventhall, H.G., 1972, The acoustics of Domestic Rooms, *Applied Acoustics*, 5, pp. 265-277.

Johnson, D.H. and Dudgeon D.E, 1993, *Array Signal Processing Concepts and techniques*, Prentice Hall.

Karjalainen, M., Piirila, E., Jarvinen, A. and Houpaniemi, J., 1999, Comparison of Loudspeaker Equalization Methods Based on DSP Techniques, *Journal of the Audio Engineering Society*, vol. 47, no. 1/2, pp. 14-31.

Karjalainen, M., Antsalo, P. and Makivirta, A., 2003, *Modal equalization by temporal shaping of room response*, Audio Engineering Society 23rd International Conference paper, no 10.

Knudsen, V.O., 1929, The hearing of speech in auditorium, *The Journal of the Acoustical Society of America*, 1, pp. 56-82.

Kobayashi, K., Furuya, K. and Kataoka, A., 2002, *A Talker-tracker microphone array for teleconferencing systems*, 113th Audio Engineering Society Convention, Paper 5642.

Kopuz, S. and Lalor, N., 1995, Analysis of interior acoustic fields using the finite element method and the boundary element method, *Applied Acoustics* 45, pp. 193-210.

Krim, J., and Viberg, M., 1996, Two decades of array signal processing research: The parametric approach, *IEEE Signal Processing Magazine*, vol. 13, pp. 67-94.

Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M. and Shafer, S., 2000, *Multi-Camera Multi-Person Tracking for EasyLiving*, Third IEEE International Workshop on Visual Surveillance.

Laakso, T.I., Valimaki, V., Karjalinen, M. and Laine, U.K., 1996, Splitting the unit delay, *IEEE Signal Processing Magazine*, January, pp. 30-60.

Laborie, A., Bruno, R. and Montoya, S., 2005, *Reproducing multichannel sound on any speaker layout*, Audio Engineering Society 118rd Convention paper, no 6375.

Levi, M. and Messer, H., 1990, Sufficient conditions for array calibration using sources of mixed tapes, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, no. 5, pp.2943-2946.

Lindsay, B.R., 1966, The Story of Acoustics, *The Journal of Acoustical Society of America*, vol. 39, no. 4, pp. 629-644.

Linkwitz, S., 2007. *Room Reflections Misunderstood*, Audio Engineering Society 123rd Convention paper, no 7162.

Lissek, H. and Meynial, X., 2003, A preliminary study of an isodynamic transducer for use in active acoustic materials*, Applied Acoustics* 64, pp. 917-930.

Liu, J., Yan, M., Peng, Y. and Li, C., 1996 DOA estimation based on array manifold calibration and weighted subspace fitting, *IEEE 3rd International Conference on Signal Processing*, vol. 1, pp. 489-492.

Long, M., Levy, M. and Stern, R., 2005, *Architectural Acoustics (Applications of Modern Acoustics)*, Academic Press.

MacLaverty, R. and Furlong, D.J., 1992, *Room acoustical simulation algorithms compared*, Audio Engineering Society 93rd International Convention paper, no 3444.

MacNair, W. A., 1930, Optimum reverberation time for auditoriums, *The Journal of the Acoustical Society of America*, 1, pp. 242-248.

Makivirta, A., Antsalo, P., Karjalainen, M. and Valimaki, V., 2001, *Low-frequency modal equalization of loudspeaker-room responses*, Audio Engineering Society 111th Convention paper, no. 5480.

Mallat, S.G., 1993, Matching pursuit with time-frequency dictionaries, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397-3415.

Manikas, A.N., Turner, L.F., 1991, Adaptive signal parameter estimation and classification technique, *Radar and Signal Processing, IEE Proceedings F*, vol. 138, no. 3, pp. 267-277.

Manikas, A., Ratnarajah, T. and Lee, J., 1997, Evaluation of superresolution array-techniques as applied to coherent sources, *International Journal of Electronics,* vol. 82, no. 1, pp. 77-105.

Mather, J.L., 1990, *The incremental multi-parameter algorithm, Conference Record of The Twenty-Fourth Asilomar Conference on Signals, Systems and Computers,* vol. 1, pp. 368–372.

Merimaa, J., 2002, *Applications of a 3-D microphone array*, Presented at the 112th Audio Engineering Society Convention, Paper no. 5501.
Meynial, X. and Vian, J.P., 1998, *Active walls for room acoustics,* Tonmeistertagung, Karsruhe.

Morrison, A., Sharif, B.S., and Hinton, O.R., 2000, An iterative DOA algorithm for a space-time DS-CDMA rake receiver, *IEE 3G Mobile Communication Technologies*, Conference Publication no. 471, pp. 208-212.

Moyle, D.E. and Warrington, E.M., 1997, *Some super-resolution DF measurements within the HF band*, 10th IEE International Conference on Antennas and Propagation, no. 436.

Muller, S. and Massarani, P., 2001, Transfer-function measurements with sweeps, *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443-471.

Nielsen, S.B. and Celestinos, A., 2007, *Time based room correction system for low frequencies using multiple loudspeakers*, Audio Engineering Society 32nd International Conference paper, no. 19.

Nikishkov, G.P., 2007, *Introductory of the finite element method, Lecture notes*, University of Aizu, Aizu-Wakamatsu, Japan.

Norcross, S.G., Soulodre G.A. and Lavoie, M. C., 2004, Subjective investigation of inverse filtering**,** *Journal of the Audio Engineering Society*, vol. 52, no. 10, pp. 1003-1028.

Northwood, T.D., 1968, *Transmission loss of plasterboard walls*, Building Research Note BRN66, Canada.

Pedersen, J.A., 2003, *Adaptive Bass Control – the ABC room adaptation system*, Audio Engineering Society 23rd International Conference paper, no. 3852.

Pedersen, J.A., 2006, *RoomPerfect DSP room correction*, Audio Engineering Society UK lecture, 14th November.

Pedersen, J.A., and Thomsen, K., 2007, *Fully automatic loudspeaker-room adaptation*, Audio Engineering Society 32nd International Conference paper, no 6.

Peterson, P.M., 1986, Simulating the response of multiple microphones to a single acoustic source in a reverberant room, *Journal of the Acoustical Society of America*, vol. 80, no. 5 pp. 1527-1529

Pham, T. and Sadler, B.M., 2004, Vehicle tracking using a network of small acoustic arrays**,** *Aerospace Conference Proceedings, 2004 IEEE***,** vol. 3, pp. 1842-1850.

Philips, G.P., 2006, *Cyclostationary direction finding*, PhD thesis, University of Birmingham.

Pillai, U.S. and Kwon B.H., 1989, Forward/backward spatial smoothing techniques for coherent signal identification, *IEEE Transactions on acoustics, Speech and Signal Processing*, vol. 37, no. 1, pp. 8-15.

Qi C., Zhang Y., Han Y. and Chen X., 2004, An algorithm on high resolution DOA estimation with unknown number of signal sources, *ICMMT 4th International Conference on Microwave and Millimeter Wave Technology, Proceedings*, pp. 227-230.

Raichel, D.R., 2000, *The Science and Applications of Acoustics*, AIP Press.

Ramos, G. and Lopez, J.J., 2006, Filter Design Method for Loudspeaker Equalization Based on IIR Parametric Filters, *Journal of the Audio Engineering Society*, vol. 54, no. 12, pp. 1162-1178.

Rindel, J.H., 2000, The use of computer modelling in room acoustics, *Journal of Vibroengineering,* no. 3(4) index 41-77, pp. 219-224.

Rindel, J.H., 2002, Modelling in Auditorium Acoustics. From Ripple Tank and Scale Models to Computer Simulations, *Revista de Acústica*, vol. XXXIII, no. 3-4, pp. 31-35.

Roy, R., Paulraj, A. and Kailath, T., 1986, Direction-of-arrival estimation by subspace rotation methods - ESPRIT, ICASSP '86, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.11, pp. 2495-2498.

Sabine, W.C., 1912, Reverberation, Reprinted from *Collected papers on acoustics*, Dover Publications, Inc. 1964, pp. 3-42.

Saruwatari, H., Kawamura, T., Sawai, K., Kaminuma, A. and Sakata, M., 2002, Blind source separation based on fast-convergence algorithm using ICA and beamforming for real convolutive mixture, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol 1. pp. 921-924.

Schmidt, R.O., 1979, Multiple emitter location and signal parameter estimation, *IEEE proceedings of the DADC Spectrum Estimation Workshop*, pp. 243-258.

See, C.M.S., 1995, Method for array calibration in high-resolution sensor array processing, *IEE Proceedings, Radar, Sonar and Navigation*, vol. 142, no. 3, pp. 90-96.

Ser, W., Waing, P., and Zhang, M., 2003, Loudspeaker Equalization Design for Near-Sound-Field Applications, *Journal of the Audio Engineering Society*, vol. 51, no. 3, pp. 156-161.

Shan, T-J., Wax, M. and Kailath, T., 1985, On spatial smoothing for direction-of-arrival estimation of coherent signals, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-33, no. 4, pp. 806-811.

Shannon, S., 1998, *Array calibration and surface wave tomography*, Ph.D. Dissertation, University of Birmingham, UK.

Skolnik, M.L., 2008, *Radar Handbook,* 3rd edition, McGrew-Hill Professional.

Spence, G. and Clarke, I. J., 1999, A novel wideband direction estimator of multiple periodic signals, *High Resolution Radar and Sonar* (ref. no. 1999/051), IEE Colloquium, pp. 8/1-8/6.

Stan, G-B, Embrechts, J-J. and Archambeau, D., 2002, Comparison of different impulse response measurement techniques, *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 249-262.

Stefanakis, N., Sarris, J. and Cambourakis, G., 2008, Source placement for equalization in small enclosures**,** *Journal of the Audio Engineering Society*, vol. 56, no. 5, pp. 357-371.

Stewart, J.L. and Westfielf, E.C., 1959, A theory of active sonar detection, *Proceedings of the IRE*, 47, pp. 872-881.

Stoica, P., Ottersten, B., Viberg, M. and Moses, R.L., 1996, Maximum likelihood array processing for stochastic coherent sources, *IEEE Transactions on Signal Processing,* vol. 44, no. 1, pp. 96-105.

Svensson, U.P., Kristiansen, U.R., 2002, Computational modelling and simulation of acoustic spaces, *Proceedings of the 22nd International AES Conference, Helsinki, Finland,* , pp. 1-20.

Takahashi, Y., Otsuru, T. and Tomiku, R., 2005, In situ measurements of surface impedance and absorption coefficients of porous materials using two microphones and ambient noise, *Applied Acoustics*, 66, pp. 845-865.

Toole, F.E., 2006, Loudspeakers and Rooms for Sound Reproduction – A Scientific Review, *Journal of the Audio Engineering Society*, vol. 54, no. 6, pp. 451-476.

Trinnov Audio, 2008, *5.0 Sound recording in high spatial resolution*, white paper downloaded from www.trinnov-audio.com/downloads.php on 15th August 2008.

Valaee, S. and Kabal, P., 1995, Wideband array processing using a two-sided correlation transform, *IEEE Transactions of Signal Processing,* vol. 43, no. 1, pp. 160-172.

van der Veen, A.-J., 1998, Algebraic methods for deterministic blind beamforming, *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1987 – 2008.

Van Trees, H.L., 2002, *Optimum Array Processing*, Wiley publishing.

Varshney, L.R. and Thomas, D., 2003, Sidelode reduction for matched filter range processing, *Proceedings of the IEEE 2003 Radar Conference*, pp. 446-451.

Volander, M., 1989, Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm, *Journal of the Acoustical Society of America*, vol. 86, no. 1, pp. 172-178.

Vries, D.de, Berkhout, A.J. and Sonke, J.J., 1996, *Array technology for measurement and analysis of sound fields in enclosures*, presented at the 100th Audio Engineering Society Conference, paper no. 4266.

Vries, D.de, Bann, J. and Sonke, J.J., 1997, *Multi-channel wave field analysis as a specification tool for wave field synthesis*, presented at the 102nd Audio Engineering Society Conference, paper no. 4453.

Wang, H. and Kaveh, M., 1984, Estimation of angle of arrival for wideband sources, IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 751-754.

Warrington, E.M., 1995a, A comparison of MUSIC and an iterative null steering technique for the resolution of multi-moded ionospherically propagated HF radio signals, *Antennas and Propagation*, Ninth International Conference (Conf. Publ. no. 407), vol. 2, pp. 220-224.

Warrington, E.M., 1995b, Measurements of direction of arrival of HF sky wave signals by means of a wide aperture antenna array and two super resolution direction finding algorithms, *Microwaves, Antennas and Propagation*, IEE Proceedings , vol. 142, no. 2, pp. 136-144.

Warrington, E.M. and Moyle, D.E., 1995c, Super-resolution HF DF measurements with simultaneous chirp sounding for a path within the UK, *HF Antennas and Propagation*, IEE Colloquium on, pp. 4/1-4/6.

Welti, T., 2002, *How many subwoofers are enough*, Audio Engineering Society 112th International Convention paper, no. 5602.

Welti, T. and Devantier, A., 2006, Low-frequency optimization using multiple subwoofers, *Journal of the Audio Engineering Society*, vol. 54, no. 5, pp. 347-364.

Wilson, R., 1991, Equalization of Loudspeaker Drive Units Considering Both On- and Off-Axis Responses, *Journal of the Audio Engineering Society*, vol. 39, no. 3, pp. 127-139.

Wilson, R., Capp, M. and Stuart, R., 2003, *The loudspeaker-room interface – Controlling excitation of room modes*, Audio Engineering Society 23rd International Conference paper, no 8.

Young, F.W., *Multidimensional scaling*, (online) http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html, (Accessed on the 28th August 2008).

Zatman, M.A., Strangeways, H.J. and Warrington, E.M., 1993, Resolution of multimoded HF transmissions using the DOSE superresolution direction finding algorithm, *IEEE Antennas and Propagation*, Eighth International Conference, vol. 1, pp. 415-417.

Zatman, M.A. and Strangeways, H.J., 1995, An efficient joint direction of arrival and frequency ML estimator, *IEEE Antennas and Propagation Society International Symposium*, vol. 1, pp. 431 – 434.

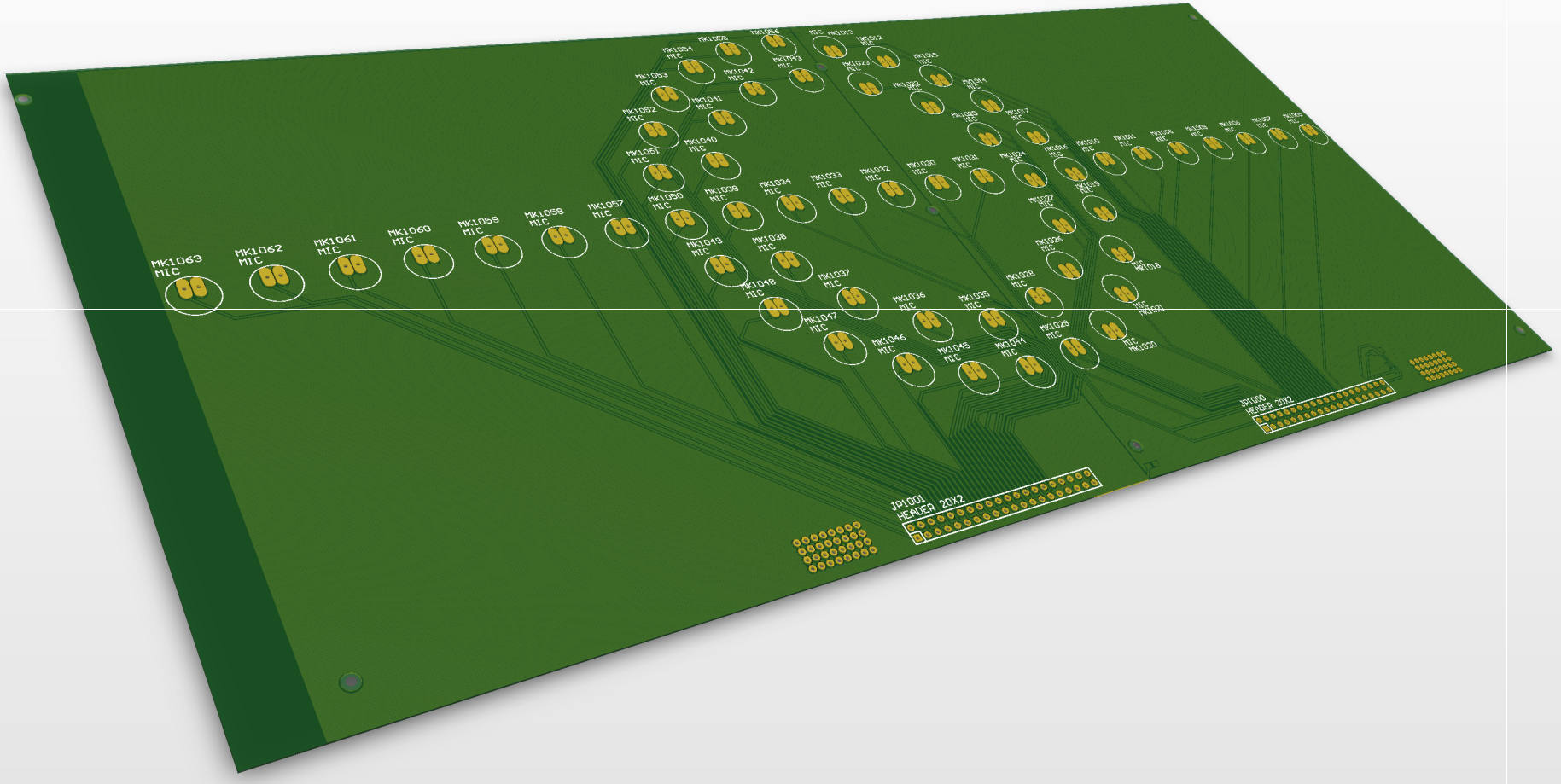
Zhang, Z., Potamianos, G., Chu, S.M., Tu, J. and Huang, T.S., 2006, *Person tracking in smart rooms using dynamic programming and subspace learning,* International Conference Multimedia Expo.

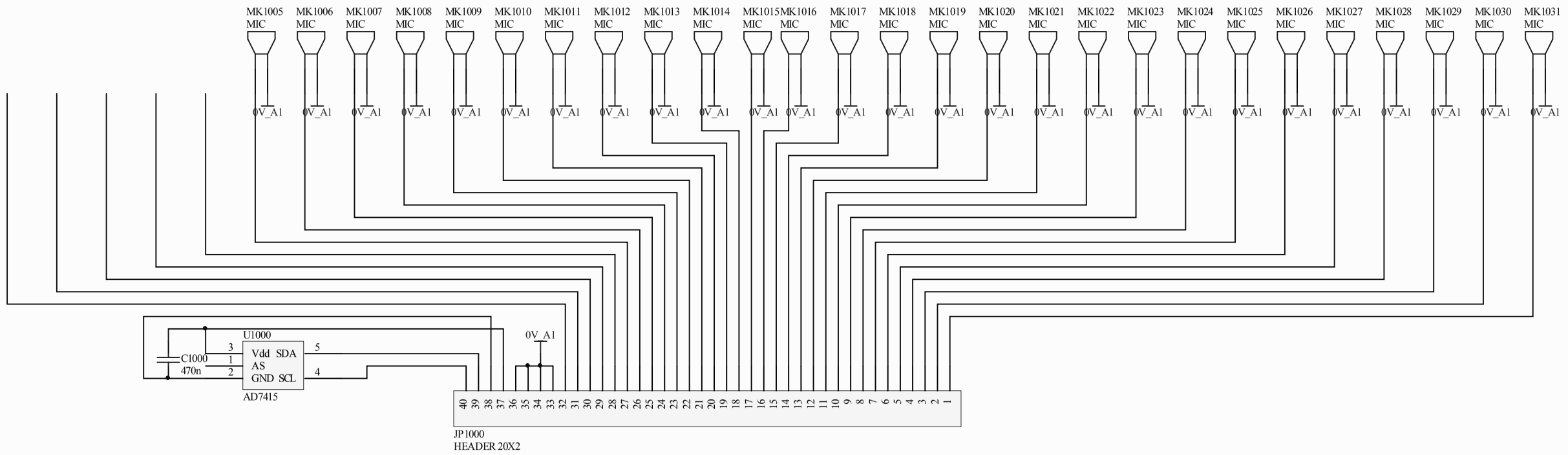
Zhou, K., Proakis, J.G. and Ling, F., 1990, Decision-feedback equalization of time-dispersive channels with coded modulation, *IEEE Transactions on Communications*, vol. 38, no. 1, pp. 18-24.

Zolzer, U., 1997, *Digital Audio Processing*, Whiley, New York.

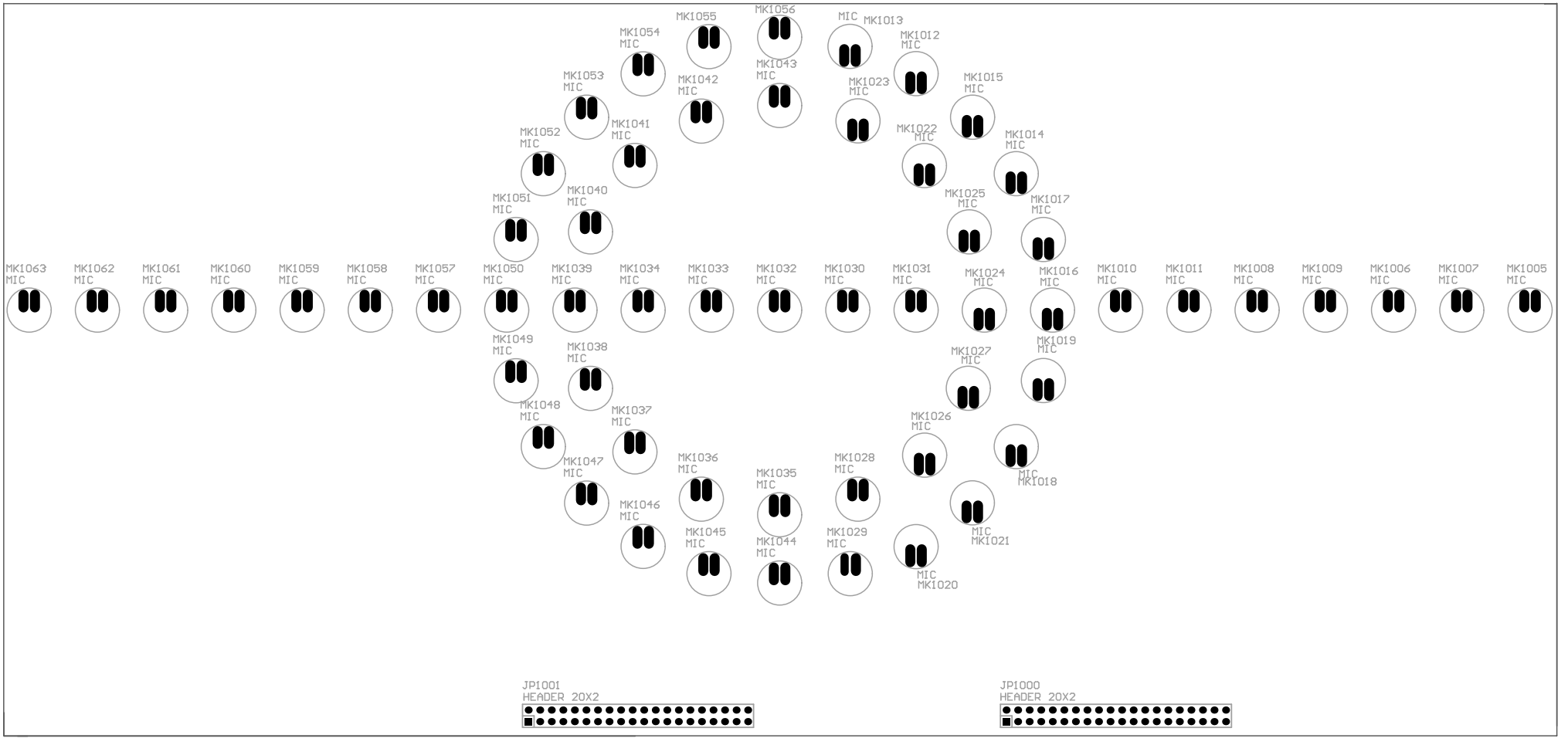
Appendix A:

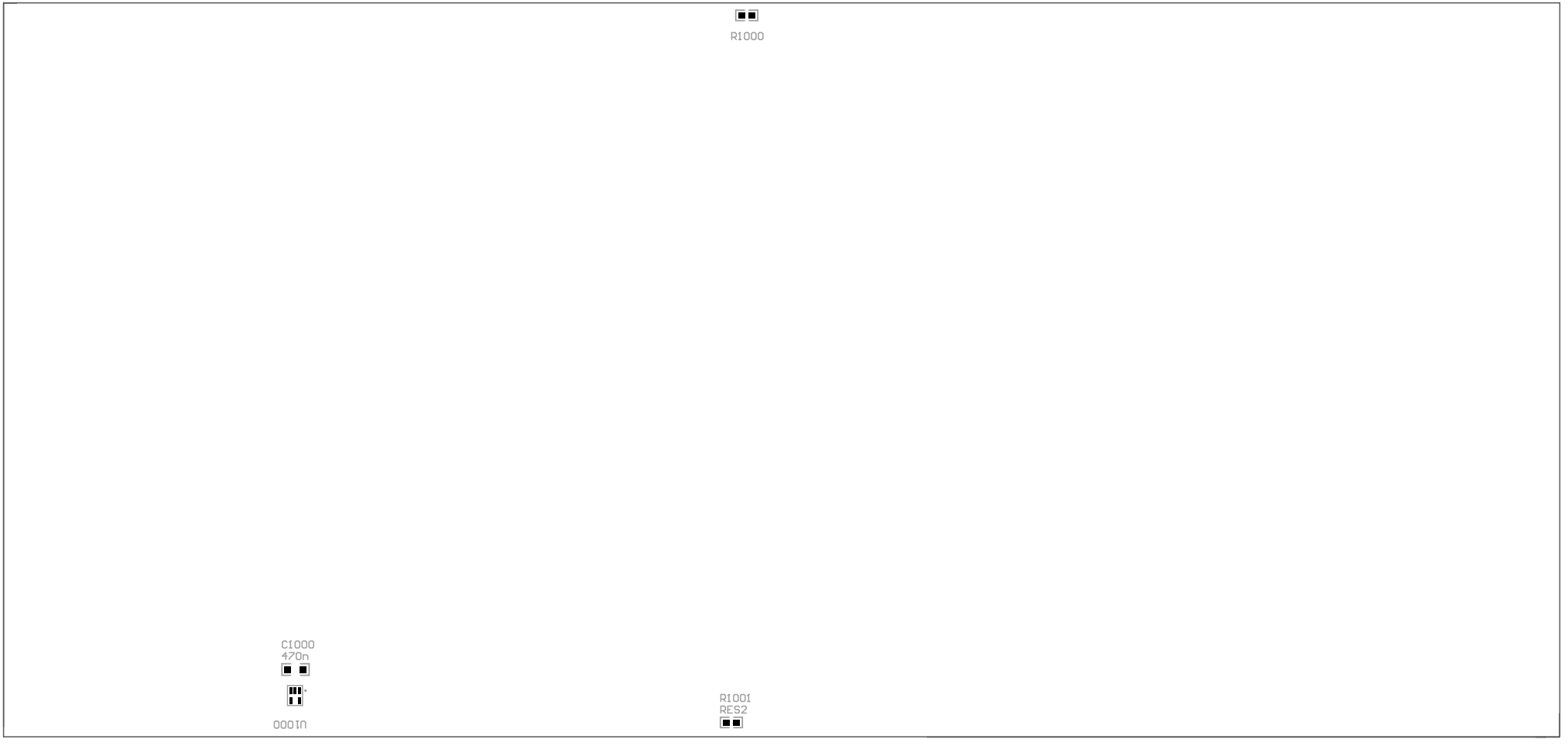
Microphone array schematic diagrams and PCB layout


MK1063 MIC
MK1062 MIC
MK1061 MIC
MK1060 MIC
MK1059 MIC
MK1058 MIC
MK1057 MIC
MK1050 MIC
MK1049 MIC
MK1048 MIC
MK1047 MIC
MK1046 MIC
MK1045 MIC
MK1044 MIC
MK1028 MIC
MK1029 MIC
MK1030 MIC
MK1031 MIC
MK1032 MIC
MK1033 MIC
MK1034 MIC
MK1039 MIC
MK1040 MIC
MK1041 MIC
MK1042 MIC
MK1043 MIC
MK1052 MIC
MK1053 MIC
MK1054 MIC
MK1055 MIC
MK1056 MIC
MK1057 MIC
MK1058 MIC
MK1059 MIC
MK1060 MIC
MK1061 MIC
MK1062 MIC
MK1063 MIC
JP1001
HEADER 20X2
JP1002
HEADER 20X2



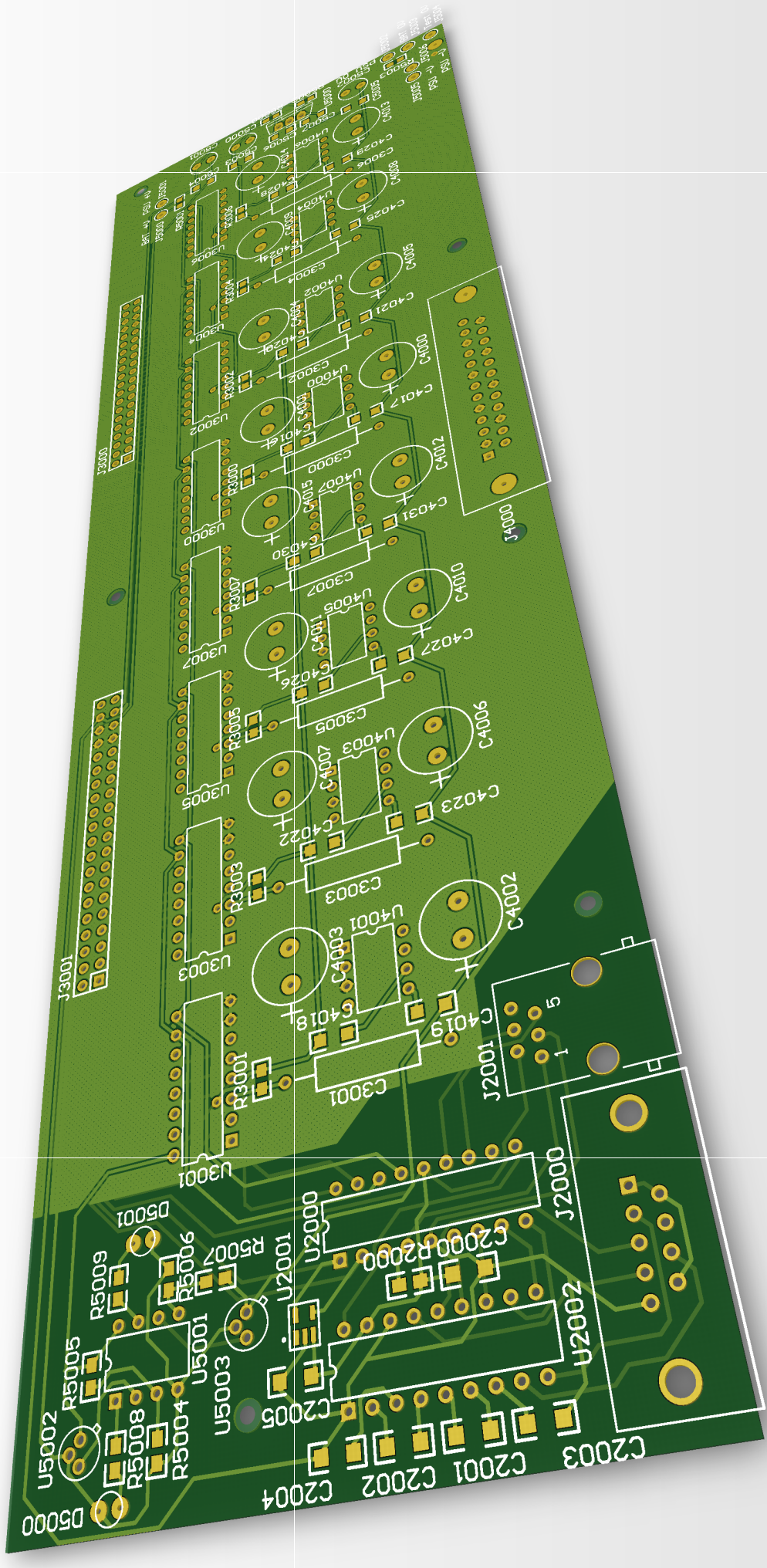
Title		
Size B	Number	Revision
Date: 25/11/2009		Sheet of
File: C:\Documents and Settings\..\Mics.Sch		Drawn By:

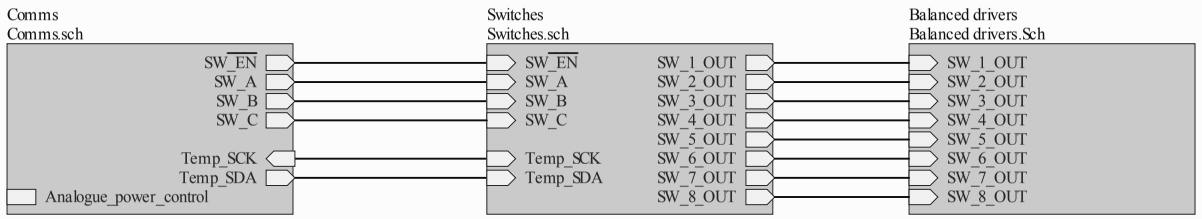

MK1063 MIC
MK1062 MIC
MK1061 MIC
MK1060 MIC
MK1059 MIC
MK1058 MIC
MK1057 MIC
MK1050 MIC
MK1039 MIC
MK1034 MIC
MK1033 MIC
MK1032 MIC
MK1030 MIC
MK1031 MIC
MK1024 MIC
MK1016 MIC
MK1010 MIC
MK1011 MIC
MK1008 MIC
MK1009 MIC
MK1006 MIC
MK1007 MIC
MK1005 MIC
MK1049 MIC
MK1038 MIC
MK1048 MIC
MK1037 MIC
MK1047 MIC
MK1036 MIC
MK1046 MIC
MK1045 MIC
MK1035 MIC
MK1044 MIC
MK1028 MIC
MK1029 MIC
MK1026 MIC
MK1027 MIC
MK1019 MIC
MK1018 MIC
MK1021 MIC
MK1020 MIC
MK1051 MIC
MK1040 MIC
MK1052 MIC
MK1041 MIC
MK1053 MIC
MK1042 MIC
MK1054 MIC
MK1055
MK1056
MK1043 MIC
MK1023 MIC
MK1022 MIC
MK1025 MIC
MK1017 MIC
MK1014 MIC
MK1015 MIC
MK1012 MIC
MK1013
MIC
JP1001
HEADER 20X2
JP1000
HEADER 20X2

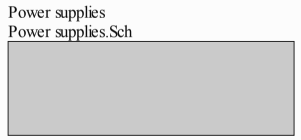

R1000
C1000
470n
U1000
R1001
RES2

Appendix B:

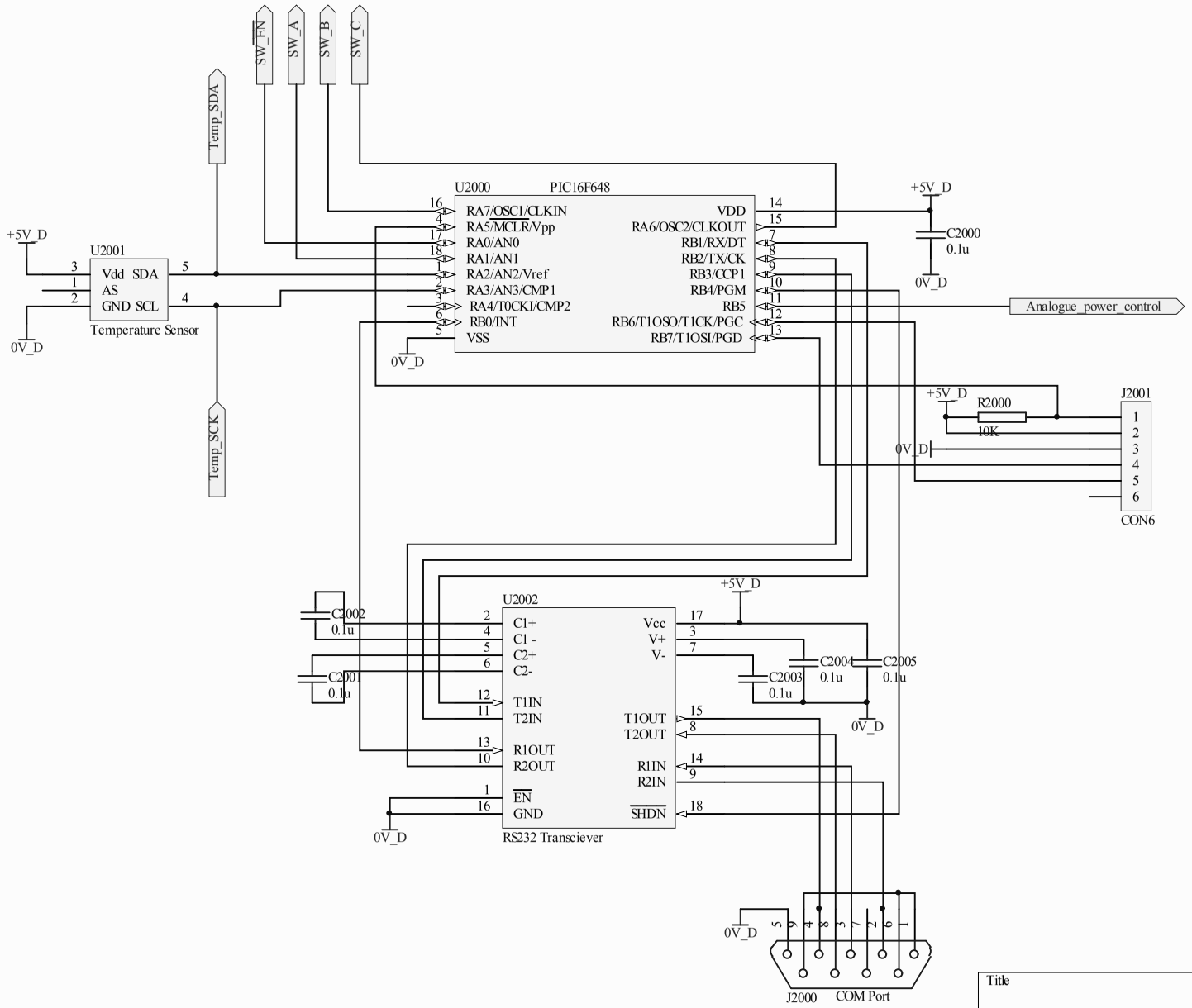
Microphone switch and preamplifier schematic diagrams and PCB layout


J3001
J4000
J2001
J2000
U3001
U3003
U3005
U3007
U3000
U3002
U3004
U3006
U4001
U4003
U4005
U4007
U4000
U4002
U4004
U4006
U2000
U2001
U2002
U5001
U5002
U5003
C3001
C3003
C3005
C3007
C3000
C3002
C3004
C3006
C4002
C4006
C4010
C4014
C4018
C4022
C4027
C4031
C4035
C4039
C4000
C4004
C4008
C4012
C4017
C4021
C4025
C4029
C2000
C2001
C2002
C2003
C2004
C2005
R3001
R3003
R3005
R3007
R3000
R3002
R3004
R3006
R5004
R5005
R5006
R5007
R5008
R5009
D5000
D5001

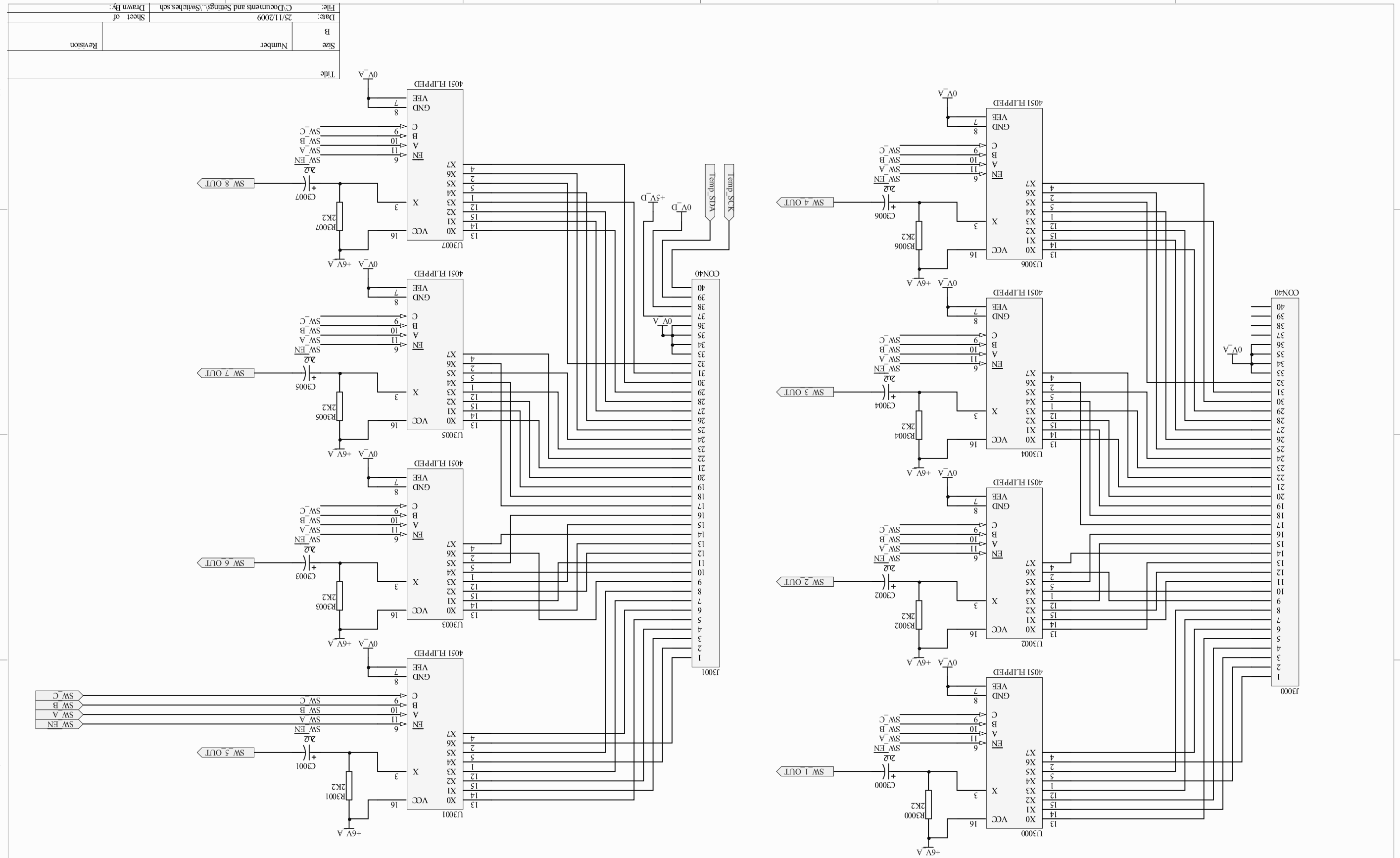




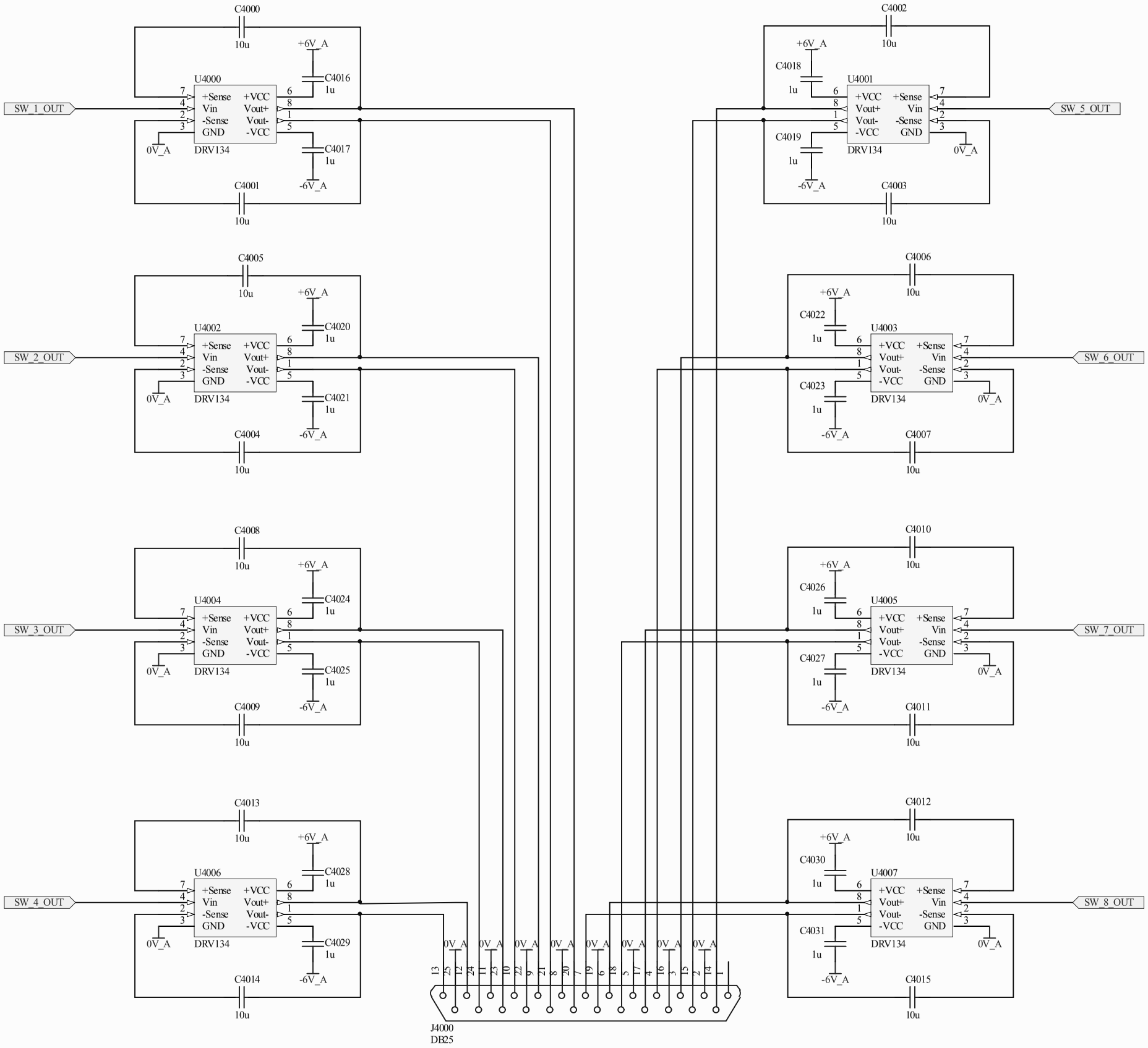
Title		
Size B	Number	Revision
Date: 25/11/2009		Sheet of
File: C:\Documents and Settings\...\Mic_Switch.SCHDOC		Drawn By:



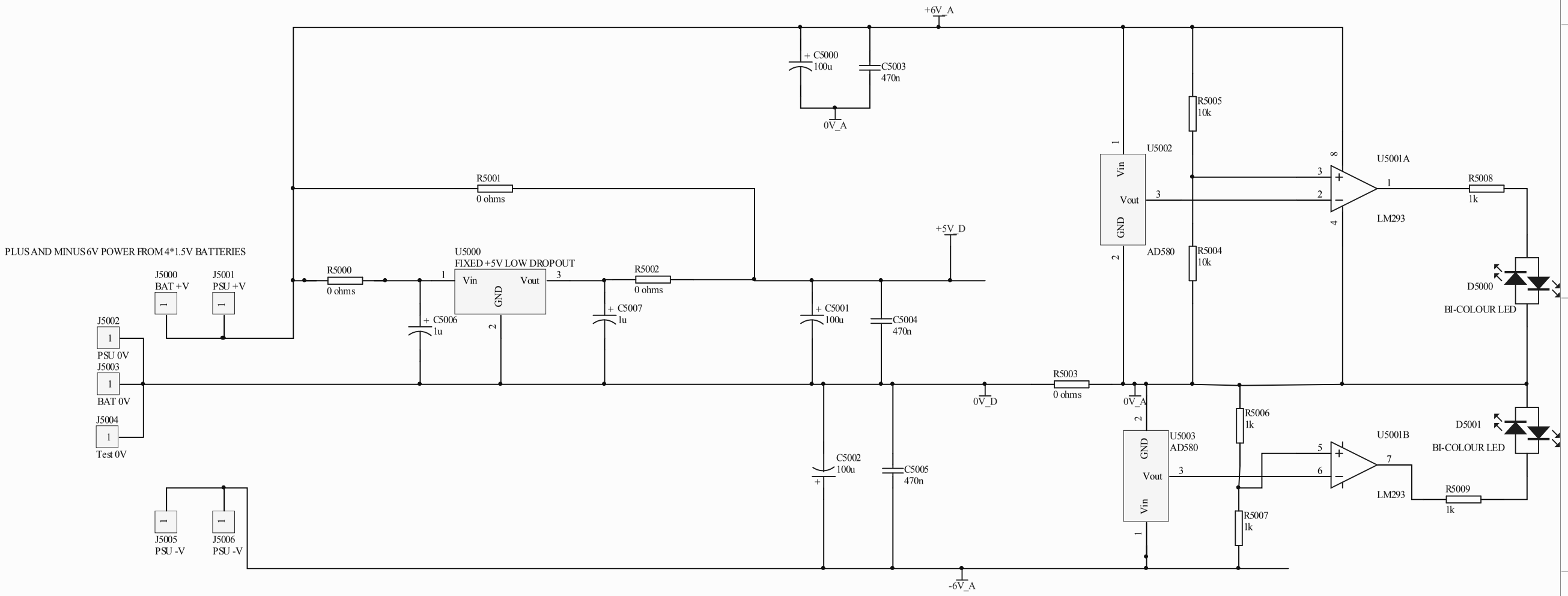
Title		
Size: A4	Number	Revision
Date: 25/11/2009		Sheet of
File: C:\Documents and Settings\..\Comms.sch		Drawn By:



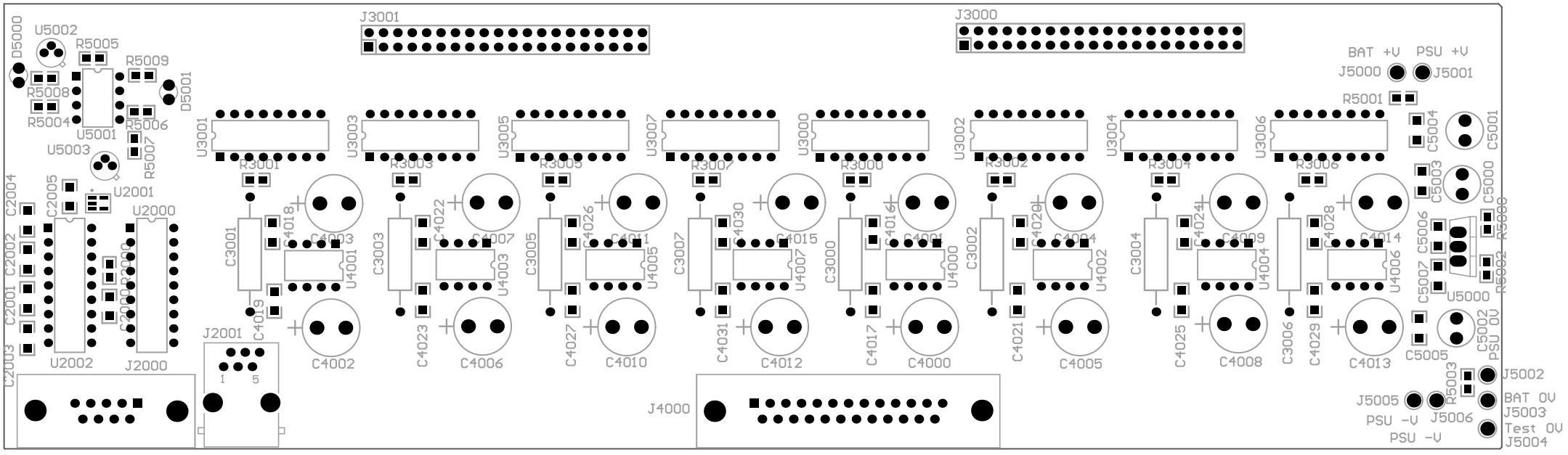
Title			
Size: B	Number:	Revision:	
Date:	25/11/2009	Sheet of	
File:	C:\Documents and Settings\...\Switches.sch	Drawn By:	


C4000
10u
+6V_A
C4016
1u
U4000
SW_1_OUT
+Sense
Vin
-Sense
GND
+VCC
Vout+
Vout-
-VCC
DRV134
0V_A
C4017
1u
-6V_A
C4001
10u
C4002
10u
+6V_A
C4018
1u
U4001
SW_5_OUT
C4019
1u
-6V_A
C4003
10u
C4005
10u
+6V_A
C4020
1u
U4002
SW_2_OUT
C4021
1u
C4004
10u
C4006
10u
+6V_A
C4022
1u
U4003
SW_6_OUT
C4023
1u
C4007
10u
C4008
10u
+6V_A
C4024
1u
U4004
SW_3_OUT
C4025
1u
C4009
10u
C4010
10u
+6V_A
C4026
1u
U4005
SW_7_OUT
C4027
1u
C4011
10u
C4013
10u
+6V_A
C4028
1u
U4006
SW_4_OUT
C4029
1u
C4014
10u
C4012
10u
+6V_A
C4030
1u
U4007
SW_8_OUT
C4031
1u
C4015
10u
J4000
DB25
Title
Size
B
Number
Revision
Date:
25/11/2009
Sheet of
File:
C:\Documents and Settings\...\Balanced drivers.Sch
Drawn By:

PLUS AND MINUS 6V POWER FROM 4*1.5V BATTERIES



Title		
Size: B	Number	Revision
Date: 25/11/2009		Sheet of
File: C:\Documents and Settings\...\Power supplies.SchDoc		Drawn By:


D5000
U5002
R5005
R5009
R5008
R5004
U5001
R5006
D5001
R5007
U5003
U2001
C2004
C2005
U2000
C2002
C2001
C2000
R2000
C2003
U2002
J2000
J2001
1
5
J4000
J3001
J3000
U3001
U3003
U3005
U3007
U3000
U3002
U3004
U3006
R3001
R3003
R3005
R3007
R3000
R3002
R3004
R3006
C3001
C3003
C3005
C3007
C3000
C3002
C3004
C3006
C4018
C4003
U4001
C4019
C4002
C4022
C4007
U4003
C4023
C4006
C4026
C4011
U4005
C4027
C4010
C4030
C4015
U4007
C4031
C4012
C4016
C4001
U4000
C4017
C4000
C4029
C4004
U4002
C4021
C4005
C4024
C4009
U4004
C4025
C4008
C4028
C4014
U4006
C4029
C4013
BAT +V
PSU +V
J5000
J5001
R5001
C5004
C5001
C5003
C5000
C5006
R5000
C5007
R5002
U5000
C5002
PSU 0V
C5005
R5003
J5002
J5005
BAT 0V
J5003
PSU -V
J5006
Test 0V
PSU -V
J5004

Appendix C:

The Localization of a Sound Source in a Reverberant Room using Arrays of Microphones, Roper S. and Collins T., 25-27th June 2007, AES 31st International Conference, London, UK

Paper not included due to copyright.

Paper available from the AES website, www.aes.org.

Appendix D:

A Sound Source and Reflection Localization method for Reverberant Rooms using Arrays of Microphones, Roper S. And Collins, T., 21-23rd September 2007, AES 32nd International Conference, Hillerød, Denmark.

Paper not included due to copyright.

Paper available from the AES website, www.aes.org.