

## University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

**REPRODUCTION OF ACOUSTIC FIELDS**

by

Ole Kirkeby

A thesis submitted for the award of Doctor of Philosophy

FACULTY OF ENGINEERING AND APPLIED SCIENCE  
INSTITUTE OF SOUND AND VIBRATION RESEARCH

March 1995

“That is a very good question, Billy. I’d better put on my patented stuponotron helmet. Prepare to activate...

(BANG!)

Well, the answer is simple really...”

Dr. Stupid from the Ren and Stimpy show

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE  
INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

REPRODUCTION OF ACOUSTIC FIELDS

by Ole Kirkeby

This thesis investigates the extent to which a sound field can be reproduced by using a relatively small number of loudspeakers. The digital signal processing techniques developed recently for the active control of sound are used to determine the loudspeaker inputs. The quality of the reproduced sound field is assessed by comparing it to a desired sound field. Three aspects of the sound reproduction problem are considered. The first is the recording of a sound field, in particular the consequences of using only a relatively small number of microphones. The second is the reproduction of a sound field, in particular the physical limitations of the quality the reproduced field when only a few loudspeakers are used. The third is the combined effect of the recording and reproduction process on the design of the digital filters. The emphasis is on basic principles, and for this reason the problems studied have been simplified greatly in order to limit the number of independent system parameters to a minimum. However, the mathematical models are general.

The sound reproduction problem is cast into the form of a linear equation system, which is solved in the least squares sense. The use of regularisation alleviates the undesirable effects of ill-conditioning, and it also makes it possible to solve any linear least squares problem regardless of its number of equations and unknowns.

A frequency domain analysis is used to investigate the physical limitations of sound field recording and reproduction by considering the problem of reproducing a plane wave, which is recorded by a microphone array inside a 'target' area, by optimally adjusting the complex amplitudes of a set of monopoles in a free field. As long as the microphone array is sufficiently 'dense', no information is lost by the recording process, and the sound field can be faithfully reproduced over the target area as long as the acoustical wavelength is not too much shorter than the 'size' of the target area.

A z-domain analysis illustrates the basic properties of the digital filters that, in the least squares sense, are given by the exact solution of the multi-channel sound reproduction system in discrete time. All these filters share a common set of poles which are readily calculated from the zeros of the determinant of a square matrix.

A time domain analysis is used to investigate the design of a set of finite impulse response filters that implement approximately the multi-channel sound reproduction problem in discrete time. The filters can be designed by direct - or adaptive inversion.

An analysis of the transient response of a sound reproduction system is used to assess the performance of the digital filters. The spatially reproduced sound field demonstrates the size of the area over which the sound can be controlled.

Some experimental results show that a matrix of measured electro-acoustic transfer functions can be inverted accurately. However, in terms of direction-of-arrival reproduction, the digital filters calculated from modelled data seem to perform just as well as the digital filters calculated from measured data.



# Thesis Contents

<b>1. An introduction to the reproduction of acoustic fields</b>	<b>1</b>
1.1. Introduction .....	1
1.2. A review of the reproduction of acoustic fields .....	1
1.3. The contribution of the thesis .....	11
<b>2. Least squares inversion theory</b>	<b>15</b>
2.1. Least squares problems .....	15
2.2. Singular value decomposition.....	16
2.3. Pseudo-inverses .....	18
2.4. Ill-conditioning and regularisation .....	22
2.5. Conclusions .....	25
<b>3. Frequency domain analysis</b>	<b>27</b>
3.1. Continuous time systems.....	28
3.2. Acoustic fields and sources .....	30
3.3. Reproduction of spatially varying sound fields .....	37
3.4. Sampling of spatially varying sound fields.....	42
3.5. Reproduction of spatially sampled sound fields.....	47

3.6.	Reproduction of plane wave sound fields .....	49
3.7.	Direction-of-arrival reproduction.....	53
3.8.	Conclusions .....	57
<b>4.</b>	<b><math>z</math>-domain analysis</b>	<b>87</b>
4.1.	Discrete-time systems .....	87
4.2.	Inversion of single-channel systems .....	91
4.3.	Inversion of multi-channel systems.....	93
4.4.	A simple class of multi-channel systems.....	98
4.5.	Least squares inversion of a 2-by-1 system .....	99
4.6.	Least effort inversion of a 1-by-2 system.....	101
4.7.	Exact inversion of a 2-by-2 system.....	102
4.8.	Least squares inversion of a 3-by-2 system .....	104
4.9.	Conclusions .....	105
<b>5.</b>	<b>Time domain analysis</b>	<b>115</b>
5.1.	Inversion of single-channel systems .....	115
5.2.	Inversion of multi-channel systems.....	121
5.3.	Direct inversion methods .....	126
5.4.	Adaptive inversion methods .....	128

5.5. Conclusions .....	132
<b>6. Transient response of multi-channel sound reproduction systems</b>	<b>142</b>
6.1. Modelling the electro-acoustic transfer function .....	143
6.2. General properties of the inverse filters .....	146
6.3. Loudspeaker-microphone layouts and conditioning .....	149
6.4. The Hanning pulse .....	151
6.5. Properties of the source inputs .....	151
6.6. Properties of the reproduced sound field .....	154
6.7. Conclusions .....	156
Appendix 6.1. Modelling the reproduced sound field .....	157
<b>7. Experimental results</b>	<b>190</b>
7.1. Measured data .....	190
7.2. Inverse filters calculated by inversion of measured transfer functions .....	192
7.3. Inverse filters calculated by inversion of modelled transfer functions .....	193
7.4. Conclusions .....	194
Appendix 7.1. Making measurements with the YDAP .....	195
<b>8. Conclusions</b>	<b>216</b>

## Acknowledgements

I am very grateful to the following Danish funds and institutions whose financial support made it possible for me to go to England and study acoustics: the Danish Research Academy, Knud Højgaards Fond, Otto Mønsted's Fond, Laurits Andersens Fond, Otto Bruuns Fond, Thomas B. Thrige's Uddannelseslegat, and Direktør Ib Henriksens Fond.

I have been supervised very competently by Phil Nelson who has always kept me on the right track, and I have had invaluable help from my fellow student Felipe Orduna-Bustamante.

I would also like to thank

- Nick Gant for the video animations he made from my data
- Dave Street for supplying "the voice on the tape"
- The people at ISVR who have helped me with technical questions and practical problems

# **1. An introduction to the reproduction of acoustic fields**

## **1.1 Introduction**

With the rapid developement of digital signal processing equipment, active control of sound and vibration continues to find new applications in the engineering sciences (Burdisso (editor) [84]). In the past, active control has been used mainly to reduce unwanted sound and vibration (Nelson and Elliott [9]). Nevertheless, a noise cancellation system is, in principle, also a sound reproduction system since, were the noise sources absent, it would reproduce a set of specified signals at a number of transducers. By specifying a set of desired, rather than undesired, signals at a number of positions, the techniques used in active control are directly applicable to sound reproduction problems. One general problem with active control systems is that the system's performance is monitored only at a number of discrete "target" positions. In practice, it is often difficult to predict what influence the system has on the sound field everywhere else, even in the vicinity of the targets, and it therefore sometimes happens that an active control system causes undesirable side-effects.

Since the output from any sound reproduction system is eventually judged by a number of listeners, such systems are usually designed from subjective criteria rather than physical criteria. Sound reproduction systems for use in the home, and in cinemas, are optimised by using the knowledge of how we perceive sound. Even if a reproduced sound field is very different from that intended, it might still "sound good" to a listener.

## **1.2 A review of techniques for the reproduction of acoustic fields**

Sound perceived by humans is inevitably filtered through the auditory system. Our brain uses sophisticated techniques to process the two signals that it receives from the ears to form a subjective impression of the physical disturbances in the medium that surrounds us (Moore [16]). By comparing those two signals, the auditory system can localise the spatial position of a sound source (Blauert [93]). The techniques it uses to achieve this are complex and still not understood in every detail (MacPherson [1],

Schroeder [60]) even though there is a vast amount of experimental data available. However, it is widely accepted that, to a first approximation, the auditory system uses three different mechanisms to localise sound, one for low-, mid-, and high frequencies respectively. At frequencies below 700Hz, the acoustical wavelength is relatively long compared to the size of the listener's head, and localisation is therefore derived from the interaural phase difference, or interaural time difference, between the sound at the two ears (Blauert [93] Section 2.4.1). At frequencies between 1500Hz and 5000Hz, where the acoustical wavelength is relatively short compared to the size of the listener's head, the head casts an acoustic shadow (Wiener [72]), and this causes an interaural level difference (Blauert [93] Section 2.4.2). At frequencies above 5000Hz, the acoustical wavelength is so short that the incoming sound is modified, or "coloured", by the outer ear according to the direction of arrival (Gerzon [101]). The transition areas between the three localisation mechanisms are relatively smooth since the auditory system combines all three types of directional cues before forming a subjective impression of the spatial sound field (Moore [16] Chapter 6 p.212). "Our acuity in locating sounds is greatest in the horizontal plane, fairly good in the vertical direction and poorest for distance" (quote from Moore [16] Chapter 6 p.227). In the horizontal plane, localisation is most accurate for sound sources in front of the listener, and least accurate for sound sources to the side of the listener (Blauert [93] Section 2.1).

When two qualitatively similar transient signals are perceived at the ears of a listener, the auditory system tends to localise the sound by using only the leading part of the transient. For example, it is usually not difficult to localise the position of a speaker in a reverberant room, because the direct sound is received before the room reflections. This effect is known as the precedence effect (Moore [16] Chapter 6, p.206), or the law of the first wavefront (Blauert [93] Section 3.1.2). A related effect, known as time-intensity trading, occurs when the signal at one ear is a scaled and delayed copy of the signal at the other ear (Moore [16] Chapter 6, p.209). In that case, the "summing localisation" (Blauert [93] Section 3.1.1) performed by the auditory system is a compromise that reflects the ratio between the level difference and the time delay. By varying the level- and time differences between the output from two loudspeakers, the effect of time-intensity trading makes it possible to give a listener the impression

that a sound originated from a physically non-existent source referred to as a phantom image (Eargle [91]). In order to be able to produce a stable phantom image in between two loudspeakers by appropriately adjusting the level and phase of their outputs, the angle between the loudspeakers, as seen by the listener, should preferably be less than 60 degrees, and certainly less than 90 degrees (de Boer [7], Theile and Plenge [55], Eargle [91] p.77). When phantom images in the centre between two loudspeakers are of poor quality, the system is said to suffer from the "hole-in-the-middle" effect.

The principle of using two loudspeakers to convey spatial information to a listener via a sound reproduction system was first described by A.D. Blumlein in his 1933 patent (Blumlein [49]). Blumlein's idea was to convert the directional information in a sound field into level differences only between the outputs from the two loudspeakers L and R (left and right) positioned in front of the listener. This simple scheme works remarkably well, because at low frequencies the level difference between the sound radiated by the loudspeakers surprisingly results in a phase difference at the ears of a centrally positioned listener, thus providing the auditory system with the desired directional cue (Lipshitz [95]). The inputs to the loudspeakers are derived from the signals registered by two closely spaced, or coincident, directional microphones. Blumlein suggested two ways to combine the output from the microphones. One, which is now referred to as XY stereo, is to point one microphone to the left of the sound stage and the other to the right of the sound stage. The recorded signals can then be fed directly to the left and right loudspeakers respectively. The other, which is now referred to as MS (middle-side) stereo, has one microphone M, typically with a cardioid directional pattern, pointing directly at the centre of the sound stage, and the other S, typically with a figure-eight directional pattern, pointing in a direction parallel to the sound stage. The L and R loudspeaker feeds are then given by  $M-S$  and  $M+S$  respectively. The XY and MS techniques are still popular. The following references describe some of the different aspects involved in sound recording and reproduction using a pair of coincident microphones: Bartlett and Billingsley [14], Julstrom [17], Gerzon [19], Streicher and Dooley [23], Dooley and Streicher [56], Eargle [91] Chapter 3, Lipshitz [95], Gerlach [97], and Hibbing [117]. The opposite extreme to the coincident microphone pair, namely a pair of widely spaced microphones, is far

less popular. Physically, the reproduced sound field is generally very different from the recorded sound field, so this method must be judged subjectively (Gerzon [31], Eargle [91] p.57, Lipshitz [95], Nisbett [108] p.68). Good results, subjectively, can also be obtained by mixing sound recorded by spaced microphones with sound recorded by a pair of coincident microphones (Wöhr et al [83]). One of the main problems with stereo reproduction over two loudspeakers is that if the listener is not positioned centrally with respect to the loudspeakers, the precedence effect tends to pull the phantom images towards the nearest loudspeaker (Moore [16] p.208-209, de Boer [98], Olson [99]). For this reason, good stereo reproduction is only experienced within a small region of the listening space, hence the notion of a "stereo seat". This effect can be alleviated by using more than two loudspeakers (Gerzon [75], Aoki et al [96]), or by optimising the directional patterns of the loudspeakers (Eargle [91] p.77-80).

In recent years, so-called binaural technology has made significant progress (Møller [45], Gierlich [105]). The basic idea is to reproduce the desired sound field exactly at the ears of the listener. The two desired signals are usually recorded with a "dummy-head" (Kleiner [78]), or they are artificially processed by using head-related transfer functions (HRTFs) to give an equivalent effect (Gierlich [79]). When the two signals are presented over headphones, the listener ideally experiences a very convincing reproduction of the spatial characteristics of the original, or artificially created, sound field. Theile [74] has suggested to use a diffuse field HRTF to convert "ordinary" recordings into "binaural format" suitable for headphone reproduction. In practice, binaural reproduction sometimes suffers from in-the-head localisation of certain sound sources, and poor localisation of frontal sources (Begault [25], Begault [76]). As opposed to loudspeaker reproduction, headphone reproduction also causes the whole sound stage to move with the listener's head. Dummy-head recordings do usually not sound good when they are played back over loudspeakers without being processed first (Theile [73]) because the sound perceived at the listeners ears have effectively been filtered by two HRTFs, and also because cross-talk spoils the desired relationship between the two signals at the listener's ears (Damaske [47]). Cross-talk occurs when a signal, which is required to be reproduced only at one location in space, is picked up at another location. Thus, when a dummy-head recording is played



back over two loudspeakers, the sound emitted from the right loudspeaker, and heard at the left ear, is cross-talk, and vice versa. A method for cancellation of cross-talk using analog techniques was patented by Atal et al [18] in 1966, and subjectively verified by Damaske [47] in 1971 and Schroeder [60] in 1975 with good results even for phantom images positioned outside the angle subtended by the loudspeakers. Since then, more sophisticated methods based on digital signal processing techniques have been developed (Nelson et al [10], Griesinger [40], Cooper and Bauck [80], Møller [82]). However, binaural reproduction over loudspeakers works only for a single listener who has to sit almost perfectly still, and it is also very sensitive to small changes in the acoustic environment.

The principle of quadraphony, in which four speakers in the corners of a square cover the full 360 degrees in the horizontal plane for a centrally positioned listener, was the first commercial "surround sound" system for domestic use (Woodward [81]). Quadraphony was never very successful, mainly because it provided phantom images of poor quality, especially to the side of the listener (Blauert [91] p.325). Theile and Plenge [55] suggest that at least six loudspeakers are necessary to provide full surround sound. At the time quadraphony was introduced in the late sixties, it was treated as an extension of two-speaker stereo, so phantom images were created by "pairwise mixing" of two of the four loudspeaker inputs. A phantom image was produced by optimally adjusting the inputs to a pair of adjacent loudspeakers. It was not long before it was discovered that pairwise mixing is not the optimal way to use the potential of the four loudspeakers (Nakabayashi [5], Cooper and Shiga [37]). Another difficulty was that, in practice, the media used for storage and transmission of the four loudspeaker signals could usually handle only two independent signals, and this made it necessary to use encoding-decoding schemes.

An encoding-decoding scheme is implemented by a matrix system. For example, a quadraphonic system that encodes four signal sources on to two tracks (or "transmission channels"), and then decodes those two signals into four loudspeaker feeds, is called a 4-2-4 matrix system. Matrix systems for surround sound reproduction were first considered by Cooper [37] in 1972. Cooper's idea was to expand the desired "phantom image distribution", specified as a function of azimuthal angle and therefore periodic, into a Fourier series. A sound field is encoded on to  $N$

“transmission channels” by storing only the first  $N$  coefficients of that Fourier series. For example, with just one transmission channel available, only the sound pressure as recorded by an omni-directional microphone would be stored since this signal corresponds to the “azimuthal DC-component” of the sound field. The  $N$  recorded signals are decoded by approximating the desired azimuthal phantom image distribution by the function given by its truncated Fourier series. Thus, the output from a loudspeaker positioned at an azimuthal angle  $\theta$  is the value of the resulting “interpolating” function at  $\theta$ . Since the encoding scheme is essentially a kind of spatial, or rather “angular”, low-pass filtering, “cross-talk” between the loudspeakers is inevitably introduced. This means that, for any position of a single desired phantom image, several loudspeakers will radiate sound simultaneously, the principle is therefore different from that of pairwise mixing.

The Ambisonics system (“ambi” being short for ambience), mainly developed by M. Gerzon and based on Cooper’s Fourier series encoding-decoding scheme, is probably the most ambitious attempt to create a commercial surround sound system that uses a standard format (Gerzon [13], Gerzon [70]). The sound field is recorded in so-called B-format using four closely spaced microphones arranged in a tetrahedon (Farrar [88], Farrar [89]). The four microphone outputs are filtered through an analog network whose four outputs X, Y, Z, and M, contain the directional information about the original sound field. X, Y, and Z are the signals as they would have been recorded by figure-of-eight microphones pointing forward, leftward, and upward respectively. M is the sound pressure as it would have been recorded by an omni-directional microphone at the centre of the tetrahedon. The B-format preserves the energy of the original sound field, meaning that  $X^2+Y^2+Z^2$  is a constant independent of the frequency and the direction of the incoming sound. The four B-format signals are encoded into UHJ format, which is convenient if it is not possible to transmit, or store, all four B format signals. A decoder recovers the relevant number of B-format signals, and filters them through “shelf filters”. The output from the shelf filters are filtered through an amplitude matrix, which depends on the specific loudspeaker setup, before eventually being fed to the loudspeakers. The shelf filters compensate for the change from time- to level difference localisation at 700Hz (Gerzon [101]), consequently they have different gains for frequencies below and above that

frequency. They also ensure that the energy is preserved for all reproduced directions. For full horizontal surround sound, a minimum of four loudspeakers are necessary, six are recommended. In 1968, Camras [20] experimentally verified that good results can be obtained when twelve loudspeakers are used to reproduce a sound field recorded with twelve directional microphones, and the position of each loudspeaker corresponds to the direction of the main lobe of one of the microphones.

Historically, most multi-channel sound systems have been developed for use in cinemas (Klapholz [42], Ampel [106]). The design criteria for cinema sound systems are quite different from those of sound reproduction systems for the home since, in a cinema, the sound system is there only to support the action on the screen, but it has to work well over a very large listening area (Holman [69]). In addition, the sound must be perfectly synchronized with the movie, the sound track therefore has to be physically registered, optically or magnetically, adjacent to the pictures on the film (Dolby Laboratories [26], Frayne [61]). Dolby Stereo, introduced in the mid seventies, is by far the most widespread cinema sound format in the world today (Dolman [102]). This format provides four signals, left, centre, right and surround. The four channels are encoded on to two analog optical sound tracks using "noise reduction" (Dolby Laboratories [39], Eargle [91] pp.246-254), and then decoded back into four signals. Consequently, the encoding-decoding scheme is implemented by a 4-2-4 matrix system. A decoder is also available for use with home video recorders (Julstrom [3]). Traditionally, the sound track for a film is split up into three components, dialogue, music, and effects (Dolman [102]). Dialogue is usually presented exclusively by the centre loudspeaker, thus providing everybody in the cinema with a "hard source" (Holman [69]) that corresponds to what is seen on the screen. Music is usually presented in conventional left-right stereo, sometimes the centre speaker is used as well. The effects channel is usually presented by several "surround" loudspeakers to the side and rear of the audience. The input to the surround speakers are processed, or "decorrelated", such that they are all slightly different, thus avoiding the precedence effect and the unintentional creation of stable phantom images for certain members of the audience (Holman [69]). As with the quadraphony format, the storage of four channels on only two tracks inevitably leads to a loss of information. The latest cinema sound format, Dolby Digital, stores the full

information about five channels, left, centre, right, left-surround, and right-surround, plus an optional sixth band-limited channel for low-frequency effects (Dolby Laboratories [38]). By using sophisticated data compression techniques, Dolby's new format, AC-3, codes these channels "into a single composite data stream with a lower bit rate than is required for just one channel on a CD" (quote from Dolby Laboratories [38]). This means that the media that are used for storage and transmission today, for example CD, LaserDisk, DAT, and VHS video tape, all have the capability to carry at least five independent sound channels. For this reason, the digital "5.1" format is likely to become widely used in the future, particularly with the so called home cinema systems and high definition television (Meares [63], Steinke [92]). The recent introduction of the "OMNIMAX theatres", which are cinemas with "hemispherical" screens has made it necessary also to consider reproduction of sound sources outside the horizontal plane (Heringa et al [103]).

In some situations, for example at live concerts, sound has to be distributed to an audience that occupies an area which is significantly larger than even the seating area of a large cinema (Gander and Eargle [59]). Sound reproduction systems for this purpose are often referred to as sound reinforcement systems. For practical reasons, the loudspeakers usually have to be grouped together in clusters, and under those circumstances, the best that can be done is to try to control the directional characteristics, in particular the main lobe, or "beam", of each of the loudspeaker clusters (see J. Audio Eng. Soc. 38 (4), April 1990; this issue is dedicated to loudspeaker arrays and their performance).

The problem of controlling the directional characteristics of an array of transducers is one commonly encountered in engineering (Haykin (editor) [22]). In underwater acoustics, arrays containing many hydrophones, frequently several hundred, are used for calibration and beamsteering. Trott [113], in 1964, suggested that an unknown transducer could be calibrated by placing it in a plane wave field created in the near-field of a "dense" planar hydrophone array (NFCA - near-field calibration array) by appropriately adjusting the gain of each hydrophone. This gain adjustment is referred to as a "shading" of the array elements. In 1973, the method was put on a firm mathematical basis by Van Buren [109] who used a least squares minimisation to determine both the gain and phase of a set of frequency independent "shading

coefficients". In Trott's original work, the direction of the plane wave was perpendicular to the array, in 1978 Van Buren [94] demonstrated how to use the NFCA for other directions of the plane wave as well. The working frequency bandwidth is limited by the size of the acoustical wavelength  $\lambda$  relative to the NFCA. From Van Buren [67], "...the element spacing must be no greater than  $0.8\lambda$  at the highest frequency of operation. Also the lateral dimensions of the NFCA should be greater than about  $4\lambda$  at the lowest frequency in order to obtain a significant plane-wave volume". The NFCA does not work well for plane waves almost parallel to the transducer array (Van Buren [94]). The principle used for beamforming is the same as that applied by the NFCA and the more advanced sound reinforcement systems. When a beamformer is placed in a sound field, the output from the many transducers are combined such that the output from the beamformer corresponds to the sound received from one, or more, directions (Halpheny and Childers [28], Flanagan [114]). This is traditionally done by adding all the transducer outputs together after each individual output has been scaled and delayed appropriately (see, for example, Flanagan et al [30]). The beam can be controlled very accurately with digital signal processing provided a suitable interpolation, or oversampling, scheme is used (Pridham and Mucci [6], Pridham and Mucci [32], Dudgeon [46]). As with the NFCA, the limitations of the beamformer are determined by the acoustical wavelength  $\lambda$  compared to the transducer array. At low frequencies the size, or aperture, of the transducer array must not be very much smaller than  $\lambda$ , at high frequencies the spacing between adjacent transducers must not be greater than  $\lambda/2$ , otherwise spatial aliasing occur (Van Veen and Buckley [24]). The beamformer is effectively a spatial filter that allows some plane wave components to pass through while other plane wave components are greatly attenuated.

Almost any conceivable sound field can be decomposed into plane waves by using near-field acoustic holography, or NAH (Maynard et al [68]). If a sound field is known on an infinite two-dimensional "hologram" plane, it is possible, in principle, to calculate the sound field on any "target" plane parallel to the hologram plane by propagating each plane wave component from the hologram plane to the target plane. This technique is useful for identifying sound sources on, for example, parts of vibrating machinery such as cars (Hald and Ginn [116]). In practice, however, the

sound field must somehow be sampled in the hologram plane, and this inevitably introduces errors (Hald [100]). Berkhout [4] has used the NAH principle to increase the reverberation time of small concert halls. A sparse microphone array near the stage picks up the sound of the performers, from those microphone outputs a digital signal processor predicts what the sound field would have been like at various positions in a large concert hall, and a number of loudspeakers are then used to create a "large hall" reverberant field. The system works well only at low frequencies because of the relatively large distance between adjacent microphones.

Sometimes, too much reverberance is a problem, for example, excessive room reflections tends to degrade speech intelligibility. The first "dereverberation" methods applied pbst-processing to a number of recorded signals in order to produce one "less reverberant" output (Allen et al [65], Flanagan and Lummis [85]). The succes of these methods relied on the specific room response, in particular the amount of "early reflections" and the length of the "reverberant tails". One fundamental problem with dereverberating, or deconvolving, a room response is that such a response is generally non-minimum phase, which means that an exact dereverberation algorithm is bound to be either unstable or non-causal, and therefore unrealisable in practice (Neely and Allen [53], Tohyama and Lyon [118]).

The deconvolution problem, and its approximate solution in the single-channel case, has been known for some time in digital signal processing (Widrow and Stearns [11] Chapter 10). If the impulse response of a system is known, it is possible to design a digital finite impulse response (FIR) filter, also referred to as an inverse filter, such that when the system's output is filtered through the inverse filter, the system's input is restored almost perfectly. The use of a modelling delay overcomes the problems caused by non-minimum phase components in the system response. The principle of inverse filtering is implemented in active control, which has been used mainly to reduce unwanted noise and vibration (Nelson and Elliott [9]). Active control of sound uses a set of digital filters to calculate the input to a number of loudspeakers in order to achieve a "desired" set of outputs from a number of microphones. In the case of noise reduction, the desired signals are zero. Miyoshi and Kanada [51] achieved perfect dereverberation of a room response at one microphone position by using two digital filters to calculate the input to two loudspeakers. Rather than cancelling out

unwanted sound, however, active control and the principle of inverse filtering can also be used for sound reproduction. For example, in stereo reproduction with two loudspeakers, a set of four digital filters can compensate for both the room response and the response of the loudspeakers, and also cancel the cross-talk (Nelson et al [10]). The method used for designing the digital filters is general, it works for any number of microphones and loudspeakers in any geometrical configuration. The optimal digital filters are chosen as the ones that, given a limited number of filter coefficients, do the best job in the least squares sense.

### **1.3 The contribution of the thesis**

The aim of this thesis is to investigate the extent to which a sound field can be reproduced by using the digital signal processing techniques developed recently for the active control of sound to determine the input to a relatively small number of loudspeakers. The quality of the reproduced sound field is assessed by comparing it to a desired sound field. Usually, the desired sound field is not known exactly; all that is available is a set of signals recorded by only a relatively small number of microphones. Three aspects of the sound reproduction problem are considered. The first is the recording of a sound field, in particular the consequences of using only relatively few microphones. The second is the reproduction of a sound field, in particular the physical limitations of the quality the reproduced field when only relatively few loudspeakers are used. The third is the combined effect of the recording and reproduction process on the design of the digital filters. The emphasis is on basic principles, and for this reason all the sound reproduction problems that are studied have been simplified greatly in order to limit the number of independent system parameters to a minimum. However, the mathematical models are general, the simplification is made to make the results easier to interpret and understand.

Chapter 2 describes the mathematical theory necessary for solving a linear equation system in the least squares sense. This is necessary in order to be able to deal with multi-channel sound reproduction systems. The theory of linear equation systems is readily available in textbooks on numerical methods, see for example Press et al [71]. The two standard methods used for solving linear least squares problems are singular value decomposition and pseudo-inversion of the coefficient matrix. Since the

problem is essentially to invert a non-square matrix, there are usually two definitions of both the singular value decomposition and the pseudo-inverse of a matrix, one for matrices with more columns than rows, and one for matrices with at least as many rows as columns. However, for the purpose of this work, the crucial property of a matrix is its number of linearly independent columns compared to its number of rows, and for this reason, the theory presented here is slightly different from that presented in Press et al [71]. The use of very small effort penalty factor makes it possible to define a least squares inversion formula which is applicable to any linear least squares problem regardless of its number of equations and unknowns. This technique is called regularisation (Press et al [71] Section 18.5).

Chapter 3 concerns the performance, in the frequency domain, of a sound reproduction system working in continuous time. First, the mathematical background necessary for analyzing linear time-invariant systems is presented. In particular, the principle of representing a filter by a transfer function is important. Next, some results from the theory of acoustics are presented. Of particular importance are the spherical- and plane wave fields, and also the concept of a monopole source. The remainder of the chapter deals with the problem of reproducing a desired sound field in the least squares sense by optimally adjusting the complex amplitude of a set of monopoles in a free field. The optimal complex amplitudes are calculated by pre-multiplying a vector, which contains the information about the desired sound field, by a matrix of inverse filters, which is calculated by inverting a matrix of electro-acoustic transfer functions in the least squares sense. The desired sound field is a plane wave, at first specified as a continuous function over a two-dimensional "target" area. This is equivalent to recording the plane wave with a very large number of microphones inside the target area. As long as the microphone array is sufficiently "dense", no information is lost by the recording process, and the sound field can be faithfully reproduced as long as the acoustical wavelength is not too much shorter than the "size" of the target area (depends on the number and positions of the sources). However, when the acoustical wavelength becomes shorter than half the distance between two adjacent microphones the effects of spatial aliasing modifies the spatial spectrum of the recorded signals, and this generally has an undesirable effect on the optimal source outputs. At low frequencies, an effort penalty is usually necessary, and this tends to distribute the



acoustic output evenly between the sources. Direction-of-arrival reproduction aims to make most of the acoustic output come from the sources whose positions correspond to the direction of the desired plane wave. This is desirable at high frequencies where the auditory system uses mainly intensity information to localise sound. Direction-of-arrival reproduction works best when there are more microphones than loudspeakers, but the microphones have to be very close together in order to avoid spatial aliasing.

Chapter 4 concerns the exact solution when the sound reproduction system is working in discrete time. First, some basic theory of discrete-time systems and digital filters are presented. Discrete-time systems are analyzed using the  $z$ -transform. In the single-channel case, the inversion of an electro-acoustic transfer function is exact (provided the transfer function does not have any zeros on the unit circle). In the multi-channel case, a matrix of exact least squares, or least effort, inverse filters is calculated by inversion of a matrix containing the  $z$ -transforms of the electro-acoustic transfer functions. A filter's frequency response is given uniquely by its  $z$ -transform, and its time response, calculated by taking the inverse  $z$ -transform, is made unique by assuming that the filter must be stable. It is demonstrated that all the inverse filters share a common set of poles, and that some of these poles are likely to be outside the unit circle. A pole outside the unit circle makes the filter non-causal since the filter is assumed to be stable. Some simple examples illustrate the properties of the inverse filters, and show the output from the sources when each of the recorded signals is an impulse.

Chapter 5 concerns an approximate least squares solution when the sound reproduction system is working in discrete time, and the inverse filters are constrained to be causal and have finite impulse responses. In both the single-channel and multi-channel case, the optimal filter coefficients are found by inversion of a large matrix containing the impulse responses of the electro-acoustic transfer functions. When there are at least as many microphones as loudspeakers, it is impossible to invert the system exactly. However, when there are more loudspeakers than microphones, and there are sufficiently many coefficients in each of the inverse filters, it is sometimes possible to invert the system exactly, in which case the desired signals are reproduced exactly at the microphones. By using adaptive algorithms, it is possible to calculate the inverse filters without explicitly building the large matrix that contains the impulse

responses of the electro-acoustic transfer functions. Adaptive inversion gives results that are almost identical to those obtained by direct inversion using an effort penalty, but it is necessary to decide subjectively when to stop the iterative algorithm.

Chapter 6 concerns the transient response of a sound reproduction system working in discrete time. In order to calculate the inverse filters, an accurate model of the electro-acoustical transfer function is necessary. The ideal electro-acoustic transfer function can be approximated to any degree of accuracy by using a fractional sample phase shifter, but in practice it is more convenient to use a digital filter that contains just one, or two, coefficients. Some examples illustrate how the set of common poles of the exact inverse filters is affected by the type of electro-acoustic transfer function, the microphone spacing, and the sampling frequency. Poles near the unit circle cause a set of inverse finite impulse response filters to perform badly at a range of frequencies corresponding to the position of the pole. Certain loudspeaker-microphone layouts, particularly highly symmetrical ones, tend to suffer from this problem much more than other layouts. The output from the loudspeakers are analyzed when the recorded signals are simple delayed copies of a "Hanning" pulse, a transient signal whose bandwidth can be controlled by varying its duration. Two examples showing the spatial reproduction of the sound field give a feel for the size of the area over which the sound field can be controlled.

Chapter 7 verifies some of the results from the previous chapters by two sets of experiments. By looking at the result of passing a set of measured recorded signals through the inverse filter matrix, it appears that, apart from at very low frequencies, the system works well as long as the acoustical wavelength is at least twice as long as the spacing between adjacent microphones. Two sets of inverse filters are analyzed, one calculated from measured electro-acoustic transfer functions, and one calculated from electro-acoustic transfer functions modelled by digital filters containing two coefficients. The frequency responses of the inverse filters calculated from measured data seem to fluctuate more wildly than the frequency responses calculated from a set of modelled electro-acoustic transfer functions.

## 2. Least squares inversion theory

In order to be able to solve a physical problem numerically, it must be cast into the form of an equation system. An equation system is a number of coupled equations, not necessarily linear, involving a number of unknowns, such as variables and functions. An equation system either has no solutions, a unique solution, or several solutions. If no exact solution exists, an approximation is used instead, and the best approximation is the one that minimises an error function defined for all available solutions. If there are several solutions that all achieve the minimum value of the error function, an additional criterion must be used in order to choose one particular solution from the family of “equally good” solutions.

The properties of linear equation systems, and their solutions, are well understood and well-documented in the literature. This chapter lists some properties of such equation systems, and gives some techniques for how to solve them. A method, called zeroth-order regularisation, is particularly useful because it works well for problems with any number of equations and any number of unknowns.

### 2.1 Least squares problems

The least squares method is by far the most common choice for solving problems that do not have an exact solution (Press et al [71]). We are only going to consider linear least squares problems in a finite number of unknown variables, that is, equation systems of the type

$$\mathbf{C}\mathbf{v} = \mathbf{d}, \tag{2.1.1}$$

where  $\mathbf{C}$  is an  $R$ -by- $S$  rectangular matrix,  $\mathbf{d}$  is a column vector containing the  $R$  “desired” right hand sides, and  $\mathbf{v}$  is a column vector containing the  $S$  unknown variables (the variables  $S$  and  $R$  are later used to denote the number of sources, or loudspeakers, and the number of receivers, or microphones, respectively). If there are more rows than columns in  $\mathbf{C}$ , the equation system is referred to as *over-determined* because it is usually not possible to find an exact solution. If there are more columns than rows, it is referred to as *under-determined* because there are usually many exact solutions. If there are equally many columns and rows, then  $\mathbf{C}$  is square, and it has a

unique solution if  $\mathbf{C}$  is not singular. A square matrix  $\mathbf{C}$  is singular if its determinant is zero; this determinant is denoted by  $|\mathbf{C}|$ . It is convenient to use the terms over-determined, square, and under-determined to classify linear equation systems, but it is not only the number of rows and columns that determine whether an equation system has a solution or not. The important property is the number of rows compared to the number of *linearly independent* columns in  $\mathbf{C}$ . This information is readily available in the singular value decomposition of  $\mathbf{C}$ .

## 2.2 Singular value decomposition

Singular value decomposition, or SVD, is presented here to illustrate some important properties of linear equation systems. The method works by factorising an  $R$ -by- $S$  matrix  $\mathbf{C}$  into a product of the three matrices  $\mathbf{U}$ ,  $\mathbf{W}$ , and  $\mathbf{V}^T$ . When  $S$  is not greater than  $R$ ,  $\mathbf{U}$  and  $\mathbf{V}$  are *orthonormal*  $R$ -by- $S$  and  $S$ -by- $S$  matrices respectively, and  $\mathbf{W}$  is a diagonal matrix that contains the  $S$  *singular values*  $w_s$  of  $\mathbf{C}$ , so

$$\mathbf{C} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T = \mathbf{U} \cdot \begin{bmatrix} w_1 & & \\ & \ddots & \\ & & w_S \end{bmatrix} \cdot \mathbf{V}^T, \quad (2.2.1)$$

where  $w_1 \geq w_2 \geq \dots \geq w_S$ . When  $S$  is greater than  $R$ ,  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal  $R$ -by- $R$  and  $S$ -by- $R$  matrices respectively, and  $\mathbf{W}$  is a diagonal matrix that contains the  $R$  largest singular values  $w_r$  of  $\mathbf{C}$ . A matrix is orthonormal when any two of its columns are orthogonal to each other, which means their dot product is zero, and, in addition, when the Euclidean length of each of its column vectors is one. The singular values  $w_r$  are the square roots of the eigenvalues of  $\mathbf{C}^H \mathbf{C}$ , and since  $\mathbf{C}^H \mathbf{C}$  is Hermitian, they are always real and positive (Kreyszig [12] p.353). A singular value of zero indicates that  $\mathbf{C}$  does not have *full rank*, there is a linear dependence between its columns. The rank of a matrix is the number of linearly independent columns, and it can be found by counting the number of its non-zero singular values. It is useful to interpret this in terms of vector-spaces, in particular, the concepts of *nullspace* and *range* are important (Press et al [71] p.61, Keener [77] p.105). The equation  $\mathbf{C}\mathbf{v} = \mathbf{d}$  defines a linear mapping from the vector space  $\mathbf{v}$  to the vector space  $\mathbf{d}$ . If there is some subspace of  $\mathbf{v}$  that is mapped to zero,  $\mathbf{C}\mathbf{v} = \mathbf{0}$ , that subspace is called the nullspace. The dimension of the nullspace (the number of linearly independent vectors  $\mathbf{v}$  that can

be found in it) is called the nullity of  $\mathbf{C}$ . The subspace of  $\mathbf{d}$  that can be “reached” by  $\mathbf{C}$ , in the sense that there exists some  $\mathbf{v}$  which is mapped there, is called the range of  $\mathbf{C}$ . The dimension of the range is called the rank of  $\mathbf{C}$ . For any matrix  $\mathbf{C}$ , its rank plus its nullity equals  $S$ , the number of columns in  $\mathbf{C}$ . For each singular value that is equal to zero, the rank is reduced by one, the nullity is increased by one, and so a dimension in the output space is “lost”. A matrix with singular values of zero is said to be singular. A matrix with no zero singular values is said to have full rank. The concepts introduced above are now used to characterise square, over-determined, and under-determined equation systems.

First, we consider square matrices. The 2-by-2 identity matrix has full rank two and nullity zero. However, the matrix

$$\mathbf{C} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \mathbf{U}\mathbf{W}\mathbf{V}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & \\ & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad (2.2.2)$$

is singular, it has rank one and nullity one. This means that even though  $\mathbf{v}$  can be chosen as a point in a two-dimensional plane,  $\mathbf{C}\mathbf{v}$  always lies on a straight line. A 3-by-3 matrix containing all ones has also rank one since its singular values are three, zero and zero. Thus, this matrix is even “more singular” than the 2-by-2 matrix containing all ones in the sense that two dimensions in the output space, rather than just one, are lost. When a square matrix  $\mathbf{C}$  has linearly dependent columns, the equation system  $\mathbf{C}\mathbf{v} = \mathbf{d}$  generally does not have an exact solution. An exact solution exists only if  $\mathbf{d}$  happens to lie in the range of  $\mathbf{C}$ .

When  $\mathbf{C}$  has more rows than columns, the equation system is over-determined. The dimension of the output vector is  $R$ , which is greater than  $S$ , but it is clear that the maximum dimension of the output space is the number of linearly independent columns which cannot be greater than  $S$ . Thus, even if  $\mathbf{C}$  has full rank  $S$ , no exact solutions exist unless  $\mathbf{d}$  happens to lie in the range of  $\mathbf{C}$ .

When  $\mathbf{C}$  has more columns than rows, the equation system is under-determined. As opposed to over-determined and square problems, the columns of  $\mathbf{C}$  are always linearly dependent when the problem is under-determined. The maximum number of

linearly independent columns that  $\mathbf{C}$  can have is  $R$ , the number of rows. For example, at first sight, the matrix

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 6 \end{bmatrix}, \quad (2.2.3)$$

looks "more singular" than the matrix

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad (2.2.4)$$

because  $\mathbf{C}_1$  has two identical columns. However, this is not the case since  $\mathbf{C}_1$  has the two non-zero singular values 6.97 and 0.61, and  $\mathbf{C}_2$  has the two non-zero singular values 9.51 and 0.77. Both  $\mathbf{C}_1$  and  $\mathbf{C}_2$  have full rank two because they can "reach" every two-dimensional vector  $\mathbf{d}$ . An example of a 2-by-3 matrix that does not have full rank is

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (2.2.5)$$

The only non-zero singular value of  $\mathbf{C}_3$  is 2.45, so the rank of this matrix is one. This means, just as in the case of a 2-by-2 matrix containing all ones, that not every two-dimensional vector  $\mathbf{d}$  can be reached because the endpoint of the vector  $\mathbf{C}_3\mathbf{v}$  always falls on a straight line. In general, a matrix with more columns than rows "compresses" the vector space  $\mathbf{v}$  into the vector space  $\mathbf{d}$  which has fewer dimensions. Consequently, there must be several vectors from  $\mathbf{v}$  that are all mapped to the same vector in  $\mathbf{d}$ . If a vector  $\mathbf{v}_1$  is mapped to  $\mathbf{d}_1$ , then so are all the vectors  $\mathbf{v}_1$  plus any linear combination of the vectors from the nullspace. Note that the nullspace cannot be empty because the rank of  $\mathbf{C}$ , which is never greater than  $R$ , is smaller than  $S$ . Therefore, if a solution exists, it is not unique, even if the rank of  $\mathbf{C}$  is equal to the dimension of  $\mathbf{d}$ .

## 2.3 Pseudo-inverses

A square equation system of the type  $\mathbf{C}\mathbf{v} = \mathbf{d}$  has an exact solution  $\mathbf{v}_0$  for any right hand side  $\mathbf{d}$ , provided that  $\mathbf{C}$  is not singular. The exact solution  $\mathbf{v}_0$  is calculated by pre-multiplying  $\mathbf{d}$  by the inverse of  $\mathbf{C}$  such that

$$\mathbf{C}^{-1}\mathbf{d} = \mathbf{v}_0. \quad (2.3.1)$$

We want to be able to determine  $\mathbf{v}_0$  uniquely even when  $\mathbf{C}$  is not square and non-singular. Since a rectangular matrix  $\mathbf{C}$  generally does not have an exact and unique inverse  $\mathbf{C}^{-1}$ , the term "pseudo-inverse" is used instead. The pseudo-inverse of  $\mathbf{C}$  is denoted by  $\mathbf{C}_{\text{pinv}}$ , and the optimal solution  $\mathbf{v}_0$ , which generally does not satisfy  $\mathbf{C}\mathbf{v}_0 = \mathbf{d}$  exactly, is given by

$$\mathbf{C}_{\text{pinv}}\mathbf{d} = \mathbf{v}_0. \quad (2.3.2)$$

The pseudo-inverse is determined by two criteria. The first is that  $\mathbf{C}\mathbf{v}_0$  must be as close as possible to  $\mathbf{d}$  in the least squares sense. The second is that if there are several vectors  $\mathbf{v}$  for which  $\mathbf{C}\mathbf{v}$  are equally close to  $\mathbf{d}$ , then  $\mathbf{v}_0$  is the shortest of those vectors. It is now demonstrated how to find the pseudo-inverse when the equation system is square, over-determined, and under-determined.

The singular value decomposition of a square  $S$ -by- $S$  matrix  $\mathbf{C}$  is given by Equation (2.2.1). If  $\mathbf{C}$  is not singular, its exact inverse is given by

$$\mathbf{C}^{-1} = [\mathbf{U} \mathbf{W} \mathbf{V}^T]^{-1} = \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T = \mathbf{V} \begin{bmatrix} 1/w_1 & & \\ & \ddots & \\ & & 1/w_S \end{bmatrix} \mathbf{U}^T. \quad (2.3.3)$$

Since both  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal square matrices, their inverses are equal to their transposes (Press et al [71] p.61). However, if  $\mathbf{C}$  is singular, then there is at least one  $w_s$  equal to zero, and this makes it impossible to calculate all the diagonal elements of  $\mathbf{W}$  using Equation (2.3.3). The pseudo-inverse of  $\mathbf{C}$  is then defined as the matrix obtained from Equation (2.3.3) by setting  $1/w_s$  to zero for all  $w_s$  that are zero, we denote this by using quotation marks around -1 for inversion,

$$\mathbf{C}_{\text{pinv}} = \mathbf{V} \mathbf{W}^{-1'} \mathbf{U}^T. \quad (2.3.4)$$

For example, the 2-by-2 matrix containing all ones, and whose singular value decomposition is given in Equation (2.2.2), has the pseudo-inverse

$$\mathbf{C}_{\text{pinv}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}^{-1'} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/2 & \\ & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (2.3.5)$$

This is not an exact inverse, so  $\mathbf{C}\mathbf{v}_0$  is generally not equal to  $\mathbf{d}$ . The difference between  $\mathbf{d}$  and  $\mathbf{C}\mathbf{v}$  is the performance error vector  $\mathbf{e}$ ,

$$\mathbf{e} = \mathbf{d} - \mathbf{C}\mathbf{v}, \quad (2.3.6)$$

and the performance cost  $E$  quantifies the size of  $\mathbf{e}$ . The value of  $E$  is a real positive scalar which is defined as the sum of the squares of the elements of  $\mathbf{e}$ , so

$$E = \sum_{r=1}^R e_r^2 = \mathbf{e}^T \mathbf{e}. \quad (2.3.7)$$

The cost function is readily generalised to include complex numbers, in that case it is defined as the sum of the squared moduli of the elements of  $\mathbf{e}$ , so

$$E = \sum_{r=1}^R |e_r|^2 = \mathbf{e}^H \mathbf{e}, \quad (2.3.8)$$

where the superscript  $H$  denotes Hermitian transpose, which is equivalent to transposition followed by complex conjugation.  $E$  can be written in a more compact form by using the 2-norm of  $\mathbf{e}$ , written as  $\|\mathbf{e}\|$ , which is the Euclidean length of  $\mathbf{e}$  (Wilkinson [104] Chapter 1 Section 52),

$$E = \|\mathbf{d} - \mathbf{C}\mathbf{v}\|^2 = \|\mathbf{e}\|^2. \quad (2.3.9)$$

The pseudo-inverse  $\mathbf{C}_{\text{pinv}}$  must be chosen such that  $\mathbf{v}_0$ , calculated from  $\mathbf{C}_{\text{pinv}}\mathbf{d}$ , minimises  $E$  with a minimum of effort  $V$  for any choice of  $\mathbf{d}$ . The effort  $V$  is defined in a way similar to  $E$ , it is given by

$$V = \sum_{s=1}^S |v_s|^2 = \mathbf{v}^H \mathbf{v} = \|\mathbf{v}\|^2. \quad (2.3.10)$$

The performance error  $E$  is minimised with a minimum of effort  $V$  by setting zero elements of  $\mathbf{W}$  to zero in the inverse of  $\mathbf{W}$  because this effectively removes the components of the nullspace from  $\mathbf{v}$ . Vectors from the nullspace can only increase the effort, but never decrease the performance cost.

When the equation system is over-determined (more rows than columns), there are usually no exact solutions. The objective is then to minimise  $E$  for a given right hand side  $\mathbf{d}$ , so by substituting the expression for  $\mathbf{e}$  given by Equation (2.3.6) into the expression for  $E$  given by Equation (2.3.8), we find



$$E = \mathbf{e}^H \mathbf{e} = (\mathbf{d} - \mathbf{C}\mathbf{v})^H (\mathbf{d} - \mathbf{C}\mathbf{v}) = \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{C}\mathbf{v} - \mathbf{v}^H \mathbf{C}^H \mathbf{d} + \mathbf{v}^H \mathbf{C}^H \mathbf{C} \mathbf{v}. \quad (2.3.11)$$

This is a Hermitian quadratic form (Nelson and Elliott [9] Appendix A.5), and it takes its minimum value

$$E_0 = \mathbf{d}^H (\mathbf{I} - [\mathbf{C}^H \mathbf{C}]^{-1} \cdot \mathbf{C}^H) \mathbf{d}, \quad (2.3.12)$$

when  $\mathbf{v}_0$  is given by

$$\mathbf{v}_0 = [\mathbf{C}^H \mathbf{C}]^{-1} \cdot \mathbf{C}^H \mathbf{d}, \quad (2.3.13)$$

provided  $\mathbf{C}$  has full rank. Note that this is the solution to the equation system,

$$\mathbf{C}^H \mathbf{C} \cdot \mathbf{v} = \mathbf{C}^H \cdot \mathbf{d}, \quad (2.3.14)$$

referred to as the normal equations (Press et al [71] p.34), which are obtained by multiplying both sides of the equation  $\mathbf{C}\mathbf{v} = \mathbf{d}$  by  $\mathbf{C}^H$ . Thus, for the over-determined case,

$$\mathbf{C}_{\text{pinv}} = [\mathbf{C}^H \mathbf{C}]^{-1} \cdot \mathbf{C}^H. \quad (2.3.15)$$

If  $\mathbf{C}$  does not have full rank, it is necessary to use singular value decomposition to remove the non-empty nullspace that makes the square Hermitian matrix  $\mathbf{C}^H \mathbf{C}$  singular. Diagonal elements of  $\mathbf{W}$  that are zero must be set to zero in the "inverse" of  $\mathbf{W}$  as indicated by Equation (2.3.4).

When the equation system is under-determined (more columns than rows), there are usually many exact solutions. The nullity of a matrix with more columns than rows is at least  $S$  minus  $R$ . However, these  $S$  minus  $R$  dimensions of the nullspace do not appear in the singular value decomposition of  $\mathbf{C}$  since we have defined  $\mathbf{W}$  to be an  $R$ -by- $R$  matrix rather than an  $S$ -by- $S$  matrix. Therefore, the pseudo-inverse is calculated formally by Equation (2.3.3), the same expression as the one used for the square and the over-determined case. If  $\mathbf{C}$  has full rank  $R$ , then this is equivalent to

$$\mathbf{C}_{\text{pinv}} = \mathbf{C}^H [\mathbf{C} \cdot \mathbf{C}^H]^{-1}. \quad (2.3.16)$$

If  $\mathbf{C}$  does not have full rank, it is again necessary to use singular value decomposition to invert the square Hermitian matrix  $\mathbf{C}\mathbf{C}^H$  in the least squares sense.

The inversion problem has a visual interpretation. Since the performance error  $E$  is quadratic in  $\mathbf{v}$ , a plot of  $E$  against the individual elements of  $\mathbf{v}$  results in a "bowl-shaped" surface, referred to as an error surface (Nelson and Elliott [9] p.110). The qualitative properties of the error surface depend on  $\mathbf{C}$ , and its quantitative properties depend on  $\mathbf{d}$ . The bottom of the error surface represents the minimum value of  $E$  denoted by  $E_0$ . If the null-space of  $\mathbf{C}$  is empty, then the error surface is a "proper" bowl with a unique minimum. This is indicated qualitatively in Figure 2.1a. If  $\mathbf{C}$  has a non-trivial null-space, then  $E_0$  is not given uniquely by a single point, but for example by a line, this is indicated qualitatively in Figure 2.1b. However, by changing the minimisation problem slightly using a technique called regularisation, it is possible to avoid completely the problem of non-unique minima.

## 2.4 Ill-conditioning and regularisation

In theory, a singular value is either zero or not zero, but when singular values are calculated by a numerical algorithm, a true singular value of zero will usually come out as a very small number (Press et al [71] p.63). It can also happen that the equations are close to being linearly dependent, but without being exactly linearly dependent, that will also result in at least one very small singular value. A very small singular value  $w_s$  makes the corresponding element  $1/w_s$  in  $\mathbf{W}^{-1}$  very large. This phenomenon is called ill-conditioning, and it is usually an undesirable property, traditionally because it makes the solution very sensitive to small changes in the data (Wilkinson [104] Chapter 2 Section 28). In our case, ill-conditioning is undesirable because it tends to make  $\mathbf{v}_0$  very large for almost any choice of the right hand side  $\mathbf{d}$ . The ill-conditioning is quantified by a condition number  $K$ , sometimes referred to as the spectral condition number (Wilkinson [104] Chapter 2 Section 30), it is defined as the ratio between the maximum and minimum singular value of  $\mathbf{C}$ ,

$$K(\mathbf{C}) = \frac{w_{\max}}{w_{\min}}. \quad (2.4.1)$$

Note that the condition number is independent of the right hand side  $\mathbf{d}$ . Ill-conditioning makes the error surface have very long narrow valleys. At a position near the valley floor,  $E$  is quite close to  $E_0$ , but  $\mathbf{v}$  can still be very far away from  $\mathbf{v}_0$ . Singular value decomposition overcomes this problem by setting very small singular

values of  $\mathbf{W}$  to zero in  $\mathbf{W}^{-1}$ . This effectively makes the long valley of the error surface completely flat, allowing the pseudo-inverse to pick out the point on the valley floor that is closest to the origin. This point is sometimes referred to as the principal solution (Press et al [71] p.806). However, even when the influence of very small singular values is removed, this technique still tends to result in subjectively undesirable solutions. Furthermore, the singular value decomposition of a matrix can only be calculated by using relatively sophisticated numerical techniques, and it is generally quite expensive in terms of computation time and memory requirements (Press et al [71] p.60).

A better method is to force the inversion problem to have a unique solution by using a technique called zeroth-order regularisation (Press et al [71] Section 18.4). The basic idea is to modify the original cost function  $E$  slightly by adding another positive definite quadratic term proportional to  $V$ . The modified cost function  $J$  is given by

$$J = E + \beta V, \quad (2.4.2)$$

and from here on we will refer to  $\beta V$  as the effort cost, and  $J$  as the total cost. The positive real number  $\beta$  determines how much weight to assign to the effort term. By varying  $\beta$  from zero to infinity, the solution changes gradually from minimizing only the performance error to minimizing only the effort cost. As opposed to the singular value decomposition technique, regularisation alleviates the ill-conditioning by "widening" the narrow valleys of the error surface instead of "flattening" them. The total cost  $J$  can be expanded into a quadratic form whose value now also depends on  $\beta$ ,

$$J = \mathbf{e}^H \mathbf{e} + \beta \mathbf{v}^H \mathbf{v} = \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{C} \mathbf{v} - \mathbf{v}^H \mathbf{C}^H \mathbf{d} + \mathbf{v}^H [\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}] \mathbf{v}, \quad (2.4.3)$$

and takes its minimum value

$$J_0 = \mathbf{d}^H \left( \mathbf{I} - [\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}]^{-1} \cdot \mathbf{C}^H \right) \mathbf{d} \quad (2.4.4)$$

when

$$\mathbf{v}_0 = [\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}]^{-1} \cdot \mathbf{C}^H \mathbf{d}, \quad (2.4.5)$$

provided  $\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}$  is not singular. This expression is valid for all rectangular matrices, and it is applicable to all inversions problems, over-determined, square, and under-

determined, because the addition of the term  $\beta \mathbf{I}$  to  $\mathbf{C}^H \mathbf{C}$  guarantees that  $\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}$  is not singular when  $\beta$  is greater than zero. The pseudo-inverses defined in the previous section are the limiting cases of Equation (2.4.5) as  $\beta$  tends to zero. The term  $\beta \mathbf{I}$  effectively shifts all the eigenvalues of the matrix that has to be inverted to the right by exactly the same amount,  $\beta$ . For example, the matrix

$$\mathbf{C} = \mathbf{A} + \beta \mathbf{I} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \beta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (2.4.6)$$

has the two eigenvalues  $\beta$  and  $2+\beta$ , while the eigenvalues of  $\mathbf{A}$  are 0 and 2. This eigenvalue-shift is a consequence of the minimax principle described by Wilkinson [104] Chapter 2 Sections 38-44. The minimax principle makes it possible to calculate certain bounds for the eigenvalues of the matrix  $\mathbf{C} = \mathbf{A} + \mathbf{B}$ , where both  $\mathbf{A}$  and  $\mathbf{B}$  are Hermitian, in terms of the individual eigenvalues of  $\mathbf{A}$  and  $\mathbf{B}$ . In the special case where  $\mathbf{B}$  is equal to  $\beta \mathbf{I}$ , all the eigenvalues of  $\mathbf{A}$  are shifted by exactly the same amount because all the eigenvalues of  $\mathbf{B}$  are equal to  $\beta$ . Since  $\mathbf{C}^H \mathbf{C}$  is positive definite, the minimum eigenvalue of  $\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}$  can never be smaller than  $\beta$ . Thus, it is possible to specify exactly how "non-singular"  $\mathbf{C}$  should be. However, one has to keep in mind that a large value of  $\beta$  will minimise the effort cost at the expense of the performance error. The condition number  $K$  of  $\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}$  is given by

$$K(\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}) = \frac{w_{\max}^2 + \beta}{w_{\min}^2 + \beta} \approx \frac{w_{\max}^2}{\beta}, \quad (2.4.7)$$

where the approximation is valid when  $\mathbf{C}^H \mathbf{C}$  is close to singular. Note that the singular values are squared in this expression, so if  $\mathbf{C}$  is ill-conditioned, then  $\mathbf{C}^H \mathbf{C}$  is much more ill-conditioned. This is one of the reasons why direct solution of the normal equations is usually not recommended (Press et al [71] pp.34-35). A rule of thumb is to aim for a value of  $\beta$  that puts  $K$  in the range between 1,000 and 5,000. Consequently, if the largest singular value  $w_{\max}$  is known,  $\beta$  should be chosen according to

$$\frac{w_{\max}}{30} < \beta < \frac{w_{\max}}{70}. \quad (2.4.8)$$

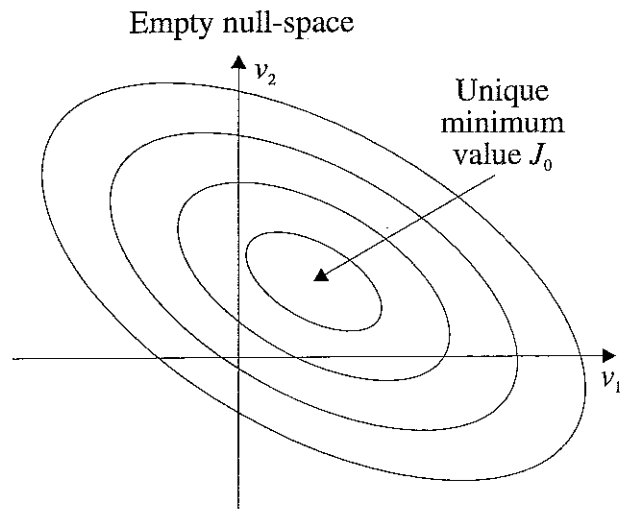
For such values of  $\beta$ , the solution tends to be well-behaved, judged by subjective standards.

Even though regularisation is usually a very efficient way of making the solution well-behaved without increasing the error too much, it still fails in some cases. It is intuitively obvious that it does not make sense to define the inverse of a matrix containing all zeros. The only way to deal with problems that are extremely badly conditioned is to modify the system that originally generated the equations.

## 2.5 Conclusions

A linear least squares inversion problem  $C\mathbf{v} = \mathbf{d}$  is usually classified as belonging to one of three groups according to the difference between its number of equations, which is the number of rows in  $C$ , and its number of unknowns, which is the number of columns in  $C$ . When  $C$  has more rows than columns, the problem is over-determined, when  $C$  has equally many rows and columns, the problem is square, and when  $C$  has more columns than rows, the problem is under-determined. Nevertheless, the difference between the number of columns and rows in  $C$  is not sufficient to tell whether its least squares inverse is unique or not, or whether it is exact or not. This information is contained in the singular value decomposition of  $C$ . The more singular values that are close to zero, the less likely it is that there exists, for a given  $\mathbf{d}$ , a  $\mathbf{v}_0$  such that  $C\mathbf{v}_0$  is close to  $\mathbf{d}$ . Such problems are said to be ill-conditioned. Apart from being an undesirable property because it makes accurate inversion difficult, ill-conditioning is undesirable because it tends to make the norm of the optimal solution  $\mathbf{v}_0$  very large for a given  $\mathbf{d}$ . This effect can be alleviated by setting small singular values of  $C$  to zero in its inverse. There is a better way, however, to avoid small singular values from corrupting the solution. The technique is called regularisation, and it is general in the sense that it gives a unique solution for both over-determined, square, and under-determined problems. It works by penalising large values of  $\mathbf{v}$  so that for a set of  $\mathbf{v}$ s that are constrained to be relatively small, the optimal solution  $\mathbf{v}_0$  is the one that gives the smallest possible performance error. Regularisation is easy to implement, and it has only negligible effect on the solution when the problem is well-conditioned.

a)



b)

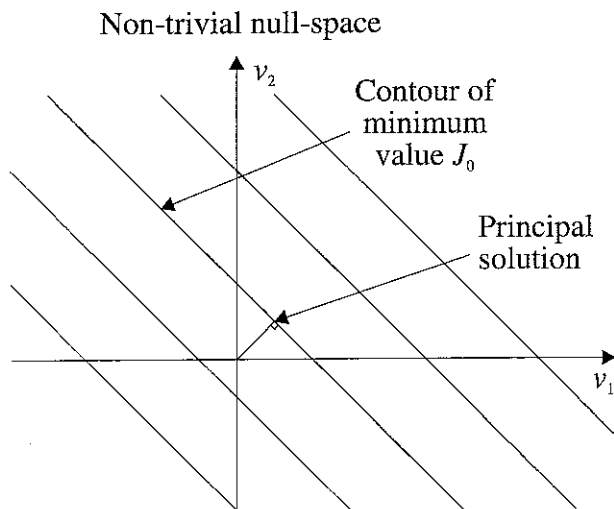


Figure 2.1. The contours of constant error for a quadratic function of two variables for a) an empty null space, and b) a non-trivial null-space

### 3. Frequency domain analysis

Many engineering problems are more easily described in the frequency domain than in the time domain. In particular, problems that involve convolution in the time domain “decouple” in the frequency domain. Translation from one domain to the other is achieved with the Fourier transform, which is a special case of the Laplace transform. By assuming that a sound field is created by a number of monopoles in a free field, it is straightforward to solve the sound reproduction problem analytically, and numerically, in the least squares sense at any one single frequency  $\omega$ , including zero (from here on, the terms “monopole”, “source”, and “loudspeaker” will be used interchangeably). The sound reproduction problem is the task of optimally reproducing a desired sound field at  $R$  receivers, or positions, distributed over a “target” area by controlling the complex input signals  $\mathbf{v}(\omega)$  to  $S$  sources. The desired signals  $\mathbf{d}(\omega)$  are defined, and optimally reproduced, in a “listening space”, but they are derived, conceptually, from a set of recorded signals  $\mathbf{u}(\omega)$  defined in a “recording space”. In this chapter the desired signals are taken to be exact copies of the recorded signals. The inputs to the sources are calculated by passing the recorded signals through a matrix  $\mathbf{H}(\omega)$  of inverse filters.  $\mathbf{H}(\omega)$  is calculated by inverting, in the least squares sense, a matrix  $\mathbf{C}(\omega)$  of electro-acoustic transfer functions. The signals that are reproduced at the  $R$  positions in the listening space are  $\mathbf{w}(\omega)$ , and the difference between the reproduced signals and the desired signals are the performance error signals  $\mathbf{e}(\omega)$ .

It is intuitively obvious that, when the size of the target area becomes too large compared to the acoustical wavelength, it is not possible to reproduce the sound field faithfully with only a few loudspeakers. Likewise, when the spacing between adjacent microphones becomes too large compared to the acoustical wavelength, or when the number of microphones is very small, it is not possible to record the sound field without losing information. This chapter concerns the physical limitations of sound field recording and reproduction.

### 3.1 Continuous time systems

The bilateral Laplace transform  $H(s)$  of a signal  $h(t)$  is defined as (Kraniauskas [112] Section 3.1)

$$H(s) = \int_{-\infty}^{+\infty} h(t) \exp(-st) dt. \quad (3.1.1)$$

The variable  $s$  is the complex frequency given by  $\sigma + j\omega$  where both  $\sigma$  and  $\omega$  are real. From  $H(s)$ , the original time signal  $h(t)$  is recovered uniquely by applying the Laplace inversion formula

$$h(t) = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} H(s) \exp(st) ds, \quad (3.1.2)$$

where the value of  $\sigma$  is a convergence factor, a real number that has to be large enough to ensure that the integral converges (Kraniauskas [112] Section 3.2). When  $H(s)$  is written as a rational polynomial,

$$H(s) = \frac{N(s)}{D(s)} = K \frac{\prod_{z=1}^{N_z} (s - s_z)}{\prod_{p=1}^{N_p} (s - s_p)}, \quad (3.1.3)$$

the numerator polynomial  $N(s)$  has  $N_z$  zeros, and the denominator polynomial  $D(s)$  has  $N_p$  zeros, referred to as poles. It is seen that, apart from the scaling factor  $K$ , the signal is determined only by the positions of its zeros,  $s_z$ , and poles,  $s_p$  in the complex  $s$ -plane (Nelson and Elliott [9] Section 3.4). If  $H(s)$  is not exactly a rational polynomial, it can be arbitrarily well approximated by one (Press et al [71] Section 3.2).

The Fourier transform is a slice of the Laplace transform along the imaginary axis. By substituting  $s$  by  $j\omega$ , it follows immediately that

$$H(j\omega) = \int_{-\infty}^{+\infty} h(t) \exp(-j\omega t) dt. \quad (3.1.4)$$

The inverse Fourier transform recovers the original time signal  $h(t)$ , and it is defined as

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} H(j\omega) \exp(j\omega t) d\omega, \quad (3.1.5)$$



which is a special case of Equation (3.2.1) with  $\sigma$  equal to zero.  $H(j\omega)$  is called the *spectrum* of  $h(t)$ , and  $H(j\omega)$  and  $h(t)$  are said to be a Fourier transform pair, this is denoted by

$$h(t) \leftrightarrow H(j\omega). \quad (3.1.6)$$

Often the imaginary unit  $j$  is omitted for notational convenience. In that case  $H(j\omega)$  is written  $H(\omega)$ . The total energy  $E_\omega$  of the spectrum  $H(j\omega)$  is proportional to the total energy  $E_t$  of the time signal. This result is known as Parseval's formula (Poularikas and Seely [86] p.182),

$$E_t = \int_{-\infty}^{+\infty} |h(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |H(j\omega)|^2 d\omega = \frac{1}{2\pi} E_\omega. \quad (3.1.7)$$

The square root of  $E_t$  is the 2-norm of  $h(t)$  in the function space  $L^2$ , and it is loosely referred to as the "length" of  $h(t)$  and written as  $\|h\|$  (Keener [77] p.61). Two Fourier transform pairs are particularly important; these are

$$h(t) = \begin{cases} 1 & \text{for } |t| < T/2 \\ 0 & \text{elsewhere} \end{cases} \leftrightarrow H(j\omega) = 2 \frac{\sin(\omega T)}{\omega}, \quad (3.1.8)$$

and

$$h(t) = 1 \leftrightarrow H(j\omega) = 2\pi \delta(\omega). \quad (3.1.9)$$

The delta-function  $\delta$  has got infinite energy, it is therefore not possible to realise in any physical system. In a linear time invariant input-output system, the output  $y(t)$  is the input  $x(t)$  convolved with the impulse response  $h(t)$  of the system,

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau) h(t - \tau) d\tau. \quad (3.1.10)$$

The delta-function  $\delta(t)$  has a "sifting" property (Kraniauskas [112] Section 1.1.5),

$$x(t) = x(t) * \delta(t) = \int_{-\infty}^{+\infty} x(\tau) \delta(t - \tau) d\tau, \quad (3.1.11)$$

which means that the delta function is the unit function for convolution, just as the constant one is the unit function for multiplication.

For linear time invariant systems, the output spectrum  $Y(\omega)$  is conveniently calculated in the frequency domain by multiplying the input spectrum  $X(\omega)$  by the transfer function  $H(\omega)$  which is the Fourier transform of the impulse response  $h(t)$ ,

$$Y(\omega) = X(\omega)H(\omega). \quad (3.1.12)$$

The system represented by  $h$  (or  $H$ ) is referred to as a *filter*. A realisable filter must be both causal and stable. A filter is causal if its impulse response is identically zero before time zero (Kraniauskas [112] p.10). This restriction ensures that the output of the filter does not depend on future values of the input. A filter is stable if a bounded input always results in a bounded output, the mathematical condition for this is that its impulse response must be absolutely integrable (Poularikas and Seely [86] p.93),

$$\int_{-\infty}^{+\infty} |h(t)| dt < +\infty. \quad (3.1.13)$$

This condition is satisfied if the Laplace transform of  $h(t)$  has all its poles in the left half of the complex plane.

### 3.2 Acoustic fields and sources

Acoustic fields are described by two physical quantities: the particle velocity vector  $\mathbf{u}$ ,

$$\mathbf{u} = \begin{bmatrix} u_{x_1} \\ u_{x_2} \\ u_{x_3} \end{bmatrix}, \quad (3.2.1)$$

whose unit is m/s, and the sound pressure  $p$ , which is a scalar whose unit is Pascal ( $\text{N/m}^2$ ). The particle velocity and the sound pressure depend on both the spatial position vector  $\mathbf{x}$ ,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad (3.2.2)$$

and the time  $t$ , but they are not independent of each other (Kinsler et al [44] Sections 5.1-5.6). The spatial variation of  $\mathbf{u}(\mathbf{x}, t)$  is coupled to the time variation of  $p(\mathbf{x}, t)$  by the linearized mass conservation equation

$$\frac{1}{c^2} \frac{\partial p(\mathbf{x}, t)}{\partial t} + \rho \nabla \cdot \mathbf{u}(\mathbf{x}, t) = 0, \quad (3.2.3)$$

provided that no mass is injected into the medium. Similarly, the spatial variation of  $p(\mathbf{x},t)$  is coupled to the time variation of  $\mathbf{u}(\mathbf{x},t)$  by the linearized momentum conservation equation

$$\rho \frac{\partial \mathbf{u}(\mathbf{x},t)}{\partial t} + \nabla p(\mathbf{x},t) = 0, \quad (3.2.4)$$

provided that no body forces are present in the medium. Two important properties of the medium in which the sound propagates appear in the two equations above. One is  $c$ , the speed of sound, which is the speed by which an acoustic disturbance travels through the medium. The speed of sound in air is approximately 340 m/s. The other is  $\rho$ , the density of the medium. The density of air is approximately 1.21 kg/m<sup>3</sup>. When Equations (3.2.3) and (3.2.4) are combined,  $\mathbf{u}$  can be eliminated. The resulting equation for the sound pressure  $p(\mathbf{x},t)$  is

$$\nabla^2 p(\mathbf{x},t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} = 0. \quad (3.2.5)$$

This expression is referred to as the homogeneous wave equation. The time variable can be eliminated by assuming that the sound field varies harmonically with time, which means that

$$p(\mathbf{x},t) = p(\mathbf{x}) \cdot e^{j\omega t} = p(\mathbf{x}) \cdot e^{j2\pi f t}, \quad (3.2.6)$$

and

$$\mathbf{u}(\mathbf{x},t) = \mathbf{u}(\mathbf{x}) \cdot e^{j\omega t} = \mathbf{u}(\mathbf{x}) \cdot e^{j2\pi f t}, \quad (3.2.7)$$

where  $\omega$  is the angular frequency in radians per second and  $f$  is the frequency in Hz, and  $p(\mathbf{x})$  and  $\mathbf{u}(\mathbf{x})$  are complex quantities that depend only on the position. The time dependent quantities  $p(\mathbf{x},t)$  and  $\mathbf{u}(\mathbf{x},t)$  are now also complex, their real parts contain the physical values of the sound pressure and the particle velocity respectively. Assuming harmonically varying sound fields leads to the homogeneous Helmholtz equation (Nelson and Elliott [9] Section 1.10),

$$(\nabla^2 + k^2)p(\mathbf{x}) = 0, \quad (3.2.8)$$

where  $k$  is the wave number, a variable proportional to the frequency given by

$$k = \frac{\omega}{c} = \frac{2\pi f}{c}. \quad (3.2.9)$$

The unit of  $k$  is  $\text{m}^{-1}$ . The acoustic wavelength  $\lambda$  is given by

$$\lambda = \frac{c}{f}, \quad (3.2.10)$$

so  $k$  is roughly six times the number of acoustic wavelengths per metre. Once  $p(\mathbf{x})$  is known,  $\mathbf{u}(\mathbf{x})$  is calculated from

$$\mathbf{u}(\mathbf{x}) = -\frac{1}{j\omega\rho} \nabla p(\mathbf{x}). \quad (3.2.11)$$

The complex sound pressure  $p(\mathbf{x})$  can be split into a magnitude part  $|p(\mathbf{x})|$  and a phase part  $\phi_p$ ,

$$p(\mathbf{x}) = |p(\mathbf{x})| e^{j\phi_p}. \quad (3.2.12)$$

The phase  $\phi_p$  is not uniquely defined by this expression since the complex exponential function of a purely imaginary argument is periodic with the period  $2\pi$ . An unfortunate implication of this is that the contours of constant phase of  $p(\mathbf{x})$  cannot be calculated directly by solving Equation (3.2.12) with respect to  $\phi_p$ . However, it is clear that a small change in  $\mathbf{x}$  causes only a small change in  $p(\mathbf{x})$ , and therefore also only a small change in  $\exp(j\phi)$ . This suggests that  $\phi_p$  itself does not change too much when  $\mathbf{x}$  is changed only slightly and that  $\phi_p$  is a continuous function (Oppenheim and Schaffer [35] p.346). By summing up all the small changes in  $\phi_p$  as  $\mathbf{x}$  moves from some arbitrary point, say  $\mathbf{x}_0$ , to  $\mathbf{x}$ , we can obtain a uniquely defined value for  $\phi_p$  at  $\mathbf{x}$ . The reason why the contours of constant phase are important is that the energy travels perpendicular to those contours. Transport of sound energy is defined by the vector quantity  $\mathbf{I}(\mathbf{x}, t)$ , it is defined as (Fahy [90] Section 4.3)

$$\mathbf{I}(\mathbf{x}, t) = p(\mathbf{x}, t) \mathbf{u}(\mathbf{x}, t). \quad (3.2.13)$$

$\mathbf{I}(\mathbf{x}, t)$  is the instantaneous sound intensity, its unit is  $\text{W/m}^2$ . For harmonically varying sound fields, the complex sound intensity is given by

$$\mathbf{I}(\mathbf{x}) = \mathbf{I}_a(\mathbf{x}) + j \mathbf{I}_r(\mathbf{x}) = \frac{1}{2} p(\mathbf{x}) \mathbf{u}^*(\mathbf{x}), \quad (3.2.14)$$

where  $\mathbf{I}_a(\mathbf{x})$ , the real part of  $\mathbf{I}(\mathbf{x})$ , is the active sound intensity, and  $\mathbf{I}_r(\mathbf{x})$ , the imaginary part of  $\mathbf{I}(\mathbf{x})$ , is the reactive sound intensity. The active and reactive sound intensities

are linked to the magnitude of  $p(\mathbf{x})$ ,  $|p(\mathbf{x})|$ , and the phase of  $p(\mathbf{x})$ ,  $\phi_p$ , by the two expressions

$$\mathbf{I}_a(\mathbf{x}) = -\frac{1}{2\omega\rho} |p(\mathbf{x})|^2 \nabla \phi_p(\mathbf{x}), \quad (3.2.15)$$

and

$$\mathbf{I}_r(\mathbf{x}) = -\frac{1}{4\omega\rho} \nabla (|p(\mathbf{x})|^2). \quad (3.2.16)$$

The active sound intensity  $\mathbf{I}_a(\mathbf{x})$  indicates the size and direction of the net sound energy that is transported through the medium,  $\mathbf{I}_a(\mathbf{x})$  corresponds to the product of the components of  $p(\mathbf{x})$  and  $\mathbf{u}(\mathbf{x})$  that are in phase with each other. It is seen that the net transport of energy is proportional to the spatial gradient of the phase, and this gradient is perpendicular to the contours of constant phase. The reactive sound intensity  $\mathbf{I}_r(\mathbf{x})$  indicates the size and direction of the energy that is stored in the sound field without contributing to the net transport of sound energy,  $\mathbf{I}_r(\mathbf{x})$  corresponds to the product of the components of  $p(\mathbf{x})$  and  $\mathbf{u}(\mathbf{x})$  that are out of phase (in quadrature) with each other. The reactive sound intensity is a measure of how much energy is converted between kinetic and potential energy during one period  $1/f$ . The instantaneous kinetic energy density  $E_{\text{kin}}(\mathbf{x}, t)$  is given by

$$E_{\text{kin}}(\mathbf{x}, t) = \frac{1}{2} \rho u^2(\mathbf{x}, t), \quad (3.2.17)$$

where  $u^2(\mathbf{x}, t)$  is the squared Euclidean length of  $\mathbf{u}(\mathbf{x}, t)$ . The instantaneous potential energy density  $E_{\text{pot}}(\mathbf{x}, t)$  is given by

$$E_{\text{pot}}(\mathbf{x}, t) = \frac{1}{2\rho c^2} p^2(\mathbf{x}, t). \quad (3.2.18)$$

The sum of  $E_{\text{kin}}(\mathbf{x}, t)$  and  $E_{\text{pot}}(\mathbf{x}, t)$  is the total energy density  $E_{\text{tot}}(\mathbf{x}, t)$ ,

$$E_{\text{tot}}(\mathbf{x}, t) = E_{\text{kin}}(\mathbf{x}, t) + E_{\text{pot}}(\mathbf{x}, t). \quad (3.2.19)$$

The unit of energy density is  $\text{J/m}^3$ . The specific acoustic impedance  $z(\mathbf{x})$  of a sound field is defined as the ratio between the sound pressure  $p(\mathbf{x})$  and the magnitude of the particle velocity  $u(\mathbf{x})$ , so

$$Z(\mathbf{x}) = Z_{\text{re}}(\mathbf{x}) + j Z_{\text{im}}(\mathbf{x}) = \frac{p(\mathbf{x})}{u(\mathbf{x})}, \quad (3.2.20)$$

where the real part of  $Z(\mathbf{x})$ ,  $Z_{\text{re}}(\mathbf{x})$ , is the specific acoustic resistance, and the imaginary part of  $Z(\mathbf{x})$ ,  $Z_{\text{im}}(\mathbf{x})$ , is the specific acoustic reactance. Note that specific acoustic resistance and reactance are conventionally denoted by the letters  $R$  and  $X$  respectively (Pierce [8] Section 3-3, Kinsler et al [44] Section 5.10).

The two most basic three-dimensional sound fields are the plane wave and the spherical wave. The complex sound pressure  $p(\mathbf{x})$  of a three-dimensional plane wave is given by

$$p(\mathbf{x}) = Ae^{-j\mathbf{k} \cdot \mathbf{x}}, \quad (3.2.21)$$

where  $A$  is a complex constant and  $\mathbf{k}$  is a propagation vector

$$\mathbf{k} = \begin{bmatrix} k_{x_1} \\ k_{x_2} \\ k_{x_3} \end{bmatrix}, \quad (3.2.22)$$

whose Euclidean length is  $\|\mathbf{k}\|$ . The contours of constant phase (surfaces in three dimensional space, lines in two-dimensional space) of a plane wave are perpendicular to the propagation vector  $\mathbf{k}$ . Since plane waves are easy to understand physically, it is often useful to decompose a sound field into plane wave components (Maynard et al [68]).

The sound intensity  $\mathbf{I}(\mathbf{x})$  of a plane wave field is purely active and independent of the position,

$$\mathbf{I}(\mathbf{x}) = \frac{|A|^2}{2\rho c} \frac{\mathbf{k}}{\|\mathbf{k}\|} = \frac{|A|^2}{2\rho c} \mathbf{e}_k, \quad (3.2.23)$$

where

$$\mathbf{e}_k = \frac{\mathbf{k}}{\|\mathbf{k}\|} \quad (3.2.24)$$

is the unit vector in the direction of propagation (Pierce [8] p.27). This result shows that  $p(\mathbf{x})$  and  $\mathbf{u}(\mathbf{x})$  are perfectly in phase everywhere in a plane wave. The specific impedance of the field is a constant,

$$Z(\mathbf{x}) = \rho c. \quad (3.2.25)$$

The product  $\rho c$  is often referred to as the characteristic impedance of the medium (Kinsler et al [44] Section 5.10). The inverse Fourier transform of the frequency expression for a plane wave is a function of time given by

$$p(\mathbf{x}, t) = A \delta\left(t - \frac{\mathbf{e}_k \cdot \mathbf{x}}{c}\right). \quad (3.2.26)$$

At time  $t_0$ ,  $p(\mathbf{x}, t_0)$  is a three-dimensional delta function for the positions  $\mathbf{x}$  that make the dot product of  $\mathbf{e}_k$  and  $\mathbf{x}$  equal to  $ct_0$ . In particular, when  $t_0$  is zero,  $p(\mathbf{x}, t)$  is a delta function for the values of  $\mathbf{x}$  that are orthogonal to  $\mathbf{e}_k$ . This type of signal is sometimes referred to as a line impulse (Dudgeon and Mersereau [64] Section 1.1.1, Evans et al [115] p.6). Note that since  $p(\mathbf{x}, t)$  contains delta functions, it has infinite energy. The contours of constant phase and the Fourier transform of a plane wave whose  $x_3$ -component is zero are illustrated in Figure 3.1a.

Another very basic acoustic field is the three-dimensional spherical field (Dowling and Ffowcs-Williams [87] Section 2.2). A spherical wave is the output from a monopole source. When the monopole is positioned at  $\mathbf{x}_0$ , the sound pressure  $p(\mathbf{x})$  of a spherical wave is defined by

$$p(\mathbf{x}) = A \frac{e^{-jk\|\mathbf{x}-\mathbf{x}_0\|}}{\|\mathbf{x}-\mathbf{x}_0\|}, \quad (3.2.27)$$

where  $\|\mathbf{x}-\mathbf{x}_0\|$  is the distance from the source to the field point  $\mathbf{x}$ , and  $A$  is a constant ( $k$  is a scalar here, not a vector). The contours of constant phase are concentric circles which indicate that the sound propagates away from the source in the centre. The particle velocity  $\mathbf{u}(\mathbf{x})$  of the wave is given by

$$\mathbf{u}(\mathbf{x}) = p(\mathbf{x}) \frac{1}{\rho c} \left( 1 - \frac{j}{k\|\mathbf{x}-\mathbf{x}_0\|} \right) \cdot (\mathbf{x} - \mathbf{x}_0), \quad (3.2.28)$$

and the sound intensity  $\mathbf{I}(\mathbf{x})$  is given by

$$\mathbf{I}(\mathbf{x}) = \frac{|A|^2}{\rho c} \frac{1}{\|\mathbf{x}-\mathbf{x}_0\|^2} \left[ 1 + j \frac{1}{k\|\mathbf{x}-\mathbf{x}_0\|} \right] \cdot (\mathbf{x} - \mathbf{x}_0). \quad (3.2.29)$$

Both  $p(\mathbf{x})$ ,  $\mathbf{u}(\mathbf{x})$ , and  $\mathbf{I}(\mathbf{x})$  are attenuated with the distance  $\|\mathbf{x}-\mathbf{x}_0\|$  because of the spherical spreading of the sound. It is seen that in the near-field, where  $k$  times  $\|\mathbf{x}-\mathbf{x}_0\|$  is small, the reactive component dominates, and in the far-field, where  $k$  times  $\|\mathbf{x}-\mathbf{x}_0\|$

is large, the active component dominates. Apart from the spherical spreading term, the sound intensity in the far-field is the same as for the plane wave (Blauert [93] Section 1.3.2). For that reason, the wavefronts are said to be locally plane. The inverse Fourier transform of the frequency expression for a spherical wave is given by

$$p(\mathbf{x}, t) = \frac{A}{\|\mathbf{x} - \mathbf{x}_0\|} \delta\left(t - \frac{\|\mathbf{x} - \mathbf{x}_0\|}{c}\right), \quad (3.2.30)$$

which shows that for a given value of  $t$ ,  $p(\mathbf{x}, t)$  is a sphere, with radius  $ct$ , of delta functions whose amplitudes decrease as  $t$  increases. Note that since  $p(\mathbf{x}, t)$  contains delta functions, it has infinite energy. Figure 3.1b shows two-dimensional slices of the contours of constant phase and the Fourier transform of a three-dimensional spherical wave.

A point monopole source at  $\mathbf{x}_0$  radiates a spherical field of the type given by Equation (3.2.27). The spherical field has a singularity at  $\mathbf{x}_0$ , the position of the source. A point monopole source can be thought of as a very small pulsating sphere whose acoustic output is proportional to its volume velocity, or source strength  $q$ . The source strength  $q$  is the product of the surface area  $A$  and the surface velocity  $u_A$ ,

$$q = Au_A, \quad (3.2.31)$$

its unit is  $\text{m}^3/\text{s}$ . It is clear that for a very small sphere,  $u_A$  must be very large to ensure that  $q$  is significantly greater than zero. For this reason, a point monopole is a hypothetical source which is impossible to build in practice. The sound pressure  $p(\mathbf{x})$  emitted from a point monopole is given by (Nelson and Elliott [9] Section 1.12)

$$p(\mathbf{x}) = j\omega q \frac{\rho}{4\pi} \frac{e^{-jk\|\mathbf{x} - \mathbf{x}_0\|}}{\|\mathbf{x} - \mathbf{x}_0\|}, \quad (3.2.32)$$

and it is seen that  $p(\mathbf{x})$  is proportional to the frequency. This frequency dependence can be avoided by using the time derivative of  $q$  rather than  $q$  itself. We call the time derivative of  $q$  the source acceleration, and use the symbol  $V$  to denote its value, so

$$V = j\omega q. \quad (3.2.33)$$

The unit of  $V$  is  $\text{m}^3/\text{s}^2$ . The introduction of  $V$  leads to a simpler expression for  $p(\mathbf{x})$ ,



$$p(\mathbf{x}) = V \frac{\rho}{4\pi} \frac{e^{-jk\|\mathbf{x}-\mathbf{x}_0\|}}{\|\mathbf{x}-\mathbf{x}_0\|}, \quad (3.2.34)$$

this expression can be used to define the transfer function from a point monopole to a field point  $\mathbf{x}$ .

Acoustic sources are formally included in the wave equation as a non-zero right hand side, the resulting expression is the inhomogeneous wave equation (Kinsler et al [44] Section 5.14)

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = -\frac{\partial G(\mathbf{x}, t)}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{x}, t). \quad (3.2.35)$$

$G(\mathbf{x}, t)$  is the rate of mass being injected into the medium per unit volume, its unit is  $\text{kg}/(\text{s m}^3)$ .  $\mathbf{F}(\mathbf{x}, t)$  is the body force per unit volume; its unit is  $\text{N}/\text{m}^3$ . A pulsating sphere in the medium makes  $G(\mathbf{x}, t)$  non-zero, and when the source is a point, it can be represented by a delta function. The inhomogeneous wave equation then becomes

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = -4\pi A \delta(\mathbf{x} - \mathbf{x}_0) e^{j\omega t} \quad (3.2.36)$$

and its solution is the spherical wave given by Equation (3.2.27) (Kinsler et al [44] Section 5.15).

### 3.3 Reproduction of spatially varying sound fields

The objective is to reproduce a desired sound field  $d(\mathbf{x}, \omega)$ , which is specified over a target area, or volume,  $\Omega$ , by using  $S$  sources. The transfer function from source  $s$  to a field point  $\mathbf{x}$  is denoted by  $C_s(\mathbf{x}, \omega)$ . This is illustrated in Figure 3.2. The reproduced sound field  $W(\mathbf{x}, \omega)$  is a linear function of the complex amplitudes of the source outputs  $V_s(\omega)$ ,

$$W(\mathbf{x}, \omega) = \sum_{s=1}^S C_s(\mathbf{x}, \omega) \cdot V_s(\omega) = \mathbf{c}(\mathbf{x}, \omega) \cdot \mathbf{v}(\omega), \quad (3.3.1)$$

where

$$\mathbf{c}(\mathbf{x}, \omega) = [C_1(\mathbf{x}, \omega) \quad C_2(\mathbf{x}, \omega) \quad \cdots \quad C_S(\mathbf{x}, \omega)], \quad (3.3.2)$$

and

$$\mathbf{v}(\omega) = \begin{bmatrix} V_1(\omega) \\ V_2(\omega) \\ \vdots \\ V_s(\omega) \end{bmatrix}. \quad (3.3.3)$$

The difference between the desired sound field  $D(\mathbf{x}, \omega)$  and the reproduced sound field  $W(\mathbf{x}, \omega)$  is the performance error  $E(\mathbf{x}, \omega)$ ,

$$E(\mathbf{x}, \omega) = D(\mathbf{x}, \omega) - W(\mathbf{x}, \omega) = D(\mathbf{x}, \omega) - \mathbf{c}(\mathbf{x}, \omega) \mathbf{v}(\omega). \quad (3.3.4)$$

This is illustrated in Figure 3.3. In the special case where the desired sound field is a plane wave whose amplitude is unity, the complex sound pressure  $D(\mathbf{x}, \omega)$  of the plane wave is always on the unit circle in the complex plane for all values of  $\mathbf{x}$  and  $\omega$ . The performance cost  $E_x(\mathbf{x}, \omega)$  measures the absolute value of the squared error, so

$$E_x(\mathbf{x}, \omega) = |E(\mathbf{x}, \omega)|^2, \quad (3.3.5)$$

and the total performance cost  $E_\Omega(\omega)$  is defined as the integral of  $E_x(\mathbf{x}, \omega)$  over  $\Omega$ ,

$$E_\Omega(\omega) = \int_\Omega |D(\mathbf{x}, \omega) - \mathbf{c}(\mathbf{x}, \omega) \mathbf{v}(\omega)|^2 d\Omega = \int_\Omega |E(\mathbf{x}, \omega)|^2 d\Omega. \quad (3.3.6)$$

It is convenient to be able to refer to a normalised error; we therefore define a root mean squared error  $E_{\text{rms}}(\omega)$  by

$$E_{\text{rms}}(\omega) = \frac{\sqrt{E_\Omega(\omega)}}{\Omega}, \quad (3.3.7)$$

which is a dimensionless measure of the average error per unit area (or volume). The solution for  $\mathbf{v}(\omega)$  that minimises  $E(\omega)$ , and therefore also  $E_{\text{rms}}(\omega)$ , is given as the solution to the linear equation system (Keener [77] Section 2.2)

$$\mathbf{C}(\omega) \mathbf{v}(\omega) = \mathbf{d}(\omega), \quad (3.3.8)$$

where  $\mathbf{C}(\omega)$  is a square Hermitian matrix,

$$\mathbf{C}(\omega) = \begin{bmatrix} \langle C_1(\omega), C_1(\omega) \rangle & \langle C_1(\omega), C_2(\omega) \rangle & \cdots & \langle C_1(\omega), C_s(\omega) \rangle \\ \langle C_2(\omega), C_1(\omega) \rangle & \langle C_2(\omega), C_2(\omega) \rangle & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \langle C_s(\omega), C_1(\omega) \rangle & \cdots & \cdots & \langle C_s(\omega), C_s(\omega) \rangle \end{bmatrix}, \quad (3.3.9)$$

and

$$\mathbf{d}(\omega) = \begin{bmatrix} \langle C_1(\omega), d(\omega) \rangle \\ \langle C_2(\omega), d(\omega) \rangle \\ \vdots \\ \langle C_s(\omega), d(\omega) \rangle \end{bmatrix}. \quad (3.3.10)$$

The elements of  $\mathbf{C}(\omega)$  and  $\mathbf{d}(\omega)$  are inner products. The inner product of the two functions  $f$  and  $g$  is defined as

$$\langle f, g \rangle = \int_{\Omega} f^*(\mathbf{x}) \cdot g(\mathbf{x}) d\Omega, \quad (3.3.11)$$

and it is effectively a generalisation of the dot product between vectors (Rice [21] Section 2.2, Mautz and Harrington [29], Körner [43] Section 32; the exact definition varies in the literature). The inner product of a function with itself is its squared Euclidean length,

$$\langle f, f \rangle = \|f\|^2, \quad (3.3.12)$$

and if the inner product between two functions is zero, the functions are said to be orthogonal. If  $\mathbf{C}(\omega)$  is non-singular, then the solution for  $\mathbf{v}(\omega)$  to Equation (3.3.8) is given uniquely by

$$\mathbf{C}^{-1}(\omega)\mathbf{d}(\omega) = \mathbf{v}(\omega). \quad (3.3.13)$$

We now illustrate the method with two examples. First consider the geometry shown in Figure 3.4a. There are two sources,  $S_1$  and  $S_2$  positioned at  $(-1.4\text{m}, -2.4\text{m})$  and  $(1.4\text{m}, -2.4\text{m})$  respectively, and the target area is a  $0.5\text{m} \times 0.5\text{m}$  square whose centre is at the origin. Thus, the angle covered by the sources is approximately  $60^\circ$ . If the sources are simple monopoles, then  $C_1(\mathbf{x}, \omega)$  and  $C_2(\mathbf{x}, \omega)$  are both of the type given by Equation (3.2.34). The frequency  $f$  is 500Hz, so  $\omega$  is approximately 3142 rad/s. The corresponding value for  $k$  is approximately  $9.24\text{m}^{-1}$ , which gives an acoustic wavelength of approximately 68cm when the speed of sound is assumed to be 340m/s. The desired sound field  $d(\mathbf{x}, \omega)$  is a plane wave with an amplitude of unity propagating along the positive direction of the  $x_2$  axis, so from Equation (3.2.21)

$$D(\mathbf{x}, \omega) = e^{-jk \cdot \mathbf{x}} = e^{-jkx_2} = e^{-j\frac{\omega x_2}{c}}, \quad (3.3.14)$$

The optimal source accelerations in  $\mathbf{v}(\omega)$  are given by multiplying the inverse of  $\mathbf{C}(\omega)$  by  $\mathbf{d}(\omega)$  as stated in Equation (3.3.13). The elements of  $\mathbf{C}(\omega)$  and  $\mathbf{d}(\omega)$  are all

inner products, and since the target area is two-dimensional, they must be calculated by surface integrations. Unfortunately, these integrals cannot be solved analytically, but they can be solved numerically (for acoustical wavelengths longer than the side-length of the target area, it usually takes a few minutes to perform a surface integration by using the software package Mathematica (Wolfram [57]) running on a 486 PC. For acoustical wavelengths shorter than the side-length of the target area, numerical integration rapidly becomes an extremely difficult task). For the example given above,  $\omega$  is 3142 rad/s, and so the element  $C_{12}(3142)$  in first row and second column of  $\mathbf{C}(3142)$  is given by (Equation(3.3.15))

$$C_{12}(3142) = \left(\frac{\rho}{4\pi}\right)^2 \int_{-0.25}^{0.25} \int_{-0.25}^{0.25} \frac{e^{j\frac{3142}{340}\sqrt{(x_1+1.4)^2+(x_2+2.4)^2}}}{\sqrt{(x_1+1.4)^2+(x_2+2.4)^2}} \cdot \frac{e^{-j\frac{3142}{340}\sqrt{(x_1-1.4)^2+(x_2+2.4)^2}}}{\sqrt{(x_1-1.4)^2+(x_2+2.4)^2}} dx_1 dx_2.$$

The value of the integral is approximately 0.0100, so  $C_{12}(3142)$  is approximately  $93.17 \times 10^{-6}$ . Equivalently, the first element  $D_1(3142)$  of  $\mathbf{d}(3142)$  is given by

$$D_1(3142) = \left(\frac{\rho}{4\pi}\right)^2 \int_{-0.25}^{0.25} \int_{-0.25}^{0.25} \frac{e^{j\frac{3142}{340}\sqrt{(x_1+1.4)^2+(x_2+2.4)^2}}}{\sqrt{(x_1+1.4)^2+(x_2+2.4)^2}} \cdot e^{-j\frac{3142}{340}x_2} dx_1 dx_2, \quad (3.3.16)$$

where the value of the integral is approximately  $0.0594 + j0.0368$ , so  $D_1(3142)$  is approximately  $(5.72 + j3.54) \times 10^{-3}$ . The matrix  $\mathbf{C}(3142)$  and the vector  $\mathbf{d}(3142)$  written out in full are

$$\mathbf{C}(3142) = \left(\frac{\rho}{4\pi}\right)^2 \begin{bmatrix} \frac{153}{4699} & \frac{197}{19604} \\ \frac{197}{19604} & \frac{153}{4699} \end{bmatrix} \quad (3.3.17)$$

and

$$\mathbf{d}(3142) = \frac{\rho}{4\pi} \begin{bmatrix} \frac{457}{7689} + j\frac{449}{12208} \\ \frac{457}{7689} + j\frac{449}{12208} \end{bmatrix}. \quad (3.3.18)$$

For notational convenience, the elements of  $\mathbf{C}(3142)$  and  $\mathbf{d}(3142)$  are given as rational approximations to the numerical values, the absolute error on the

approximations is less  $10^{-8}$ . It is seen that because of the symmetry of the geometry,  $C_{11} = C_{22}$ ,  $C_{12} = C_{21}$ , and  $D_1 = D_2$ . Note also that  $\mathbf{C}$  is entirely real and that the equation system is well-conditioned,  $K(\mathbf{C})$  being less than 1.9. The solution for  $\mathbf{v}_0(3142)$  is given by

$$\mathbf{v}_0(3142) = \frac{4\pi}{\rho} \begin{bmatrix} \frac{25835}{18521} + j \frac{15267}{17687} \\ \frac{25835}{18521} + j \frac{15267}{17687} \end{bmatrix}, \quad (3.3.19)$$

and the corresponding root mean squared error  $E_{\text{rms}}$  is approximately 0.29. When both sources are silent,  $E_{\text{rms}}$  is exactly one, so the reduction in the root mean squared error is just over 70%.

Now consider the geometry shown in Figure 3.4b. There are four sources,  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_4$  positioned at  $(-2\text{m}, -2\text{m})$ ,  $(-0.7\text{m}, -2.4\text{m})$ ,  $(0.7\text{m}, -2.4\text{m})$ , and  $(2\text{m}, -2\text{m})$  respectively. The target area is still a  $0.5\text{m} \times 0.5\text{m}$  square whose centre is at the origin. Thus, the angle covered by the sources has increased to 90 degrees, but since there are now four sources, it is intuitively obvious that this system ought to be able to perform better than the previous system with only two sources. By calculating the surface integrals equivalent to the previous example, we find (Equation 3.3.20)

$$\mathbf{C}(3142) = \left( \frac{\rho}{4\pi} \right)^2 \begin{bmatrix} \frac{119}{3788} & -\frac{155}{5493} + j \frac{49}{15318} & -\frac{303}{27661} + j \frac{5}{3207} & -\frac{5}{4073} \\ -\frac{155}{5493} - j \frac{49}{15318} & \frac{1447}{35934} & \frac{421}{14124} & -\frac{303}{27661} - j \frac{5}{3207} \\ -\frac{303}{27661} - j \frac{5}{3207} & \frac{421}{14124} & \frac{1447}{35934} & -\frac{155}{5493} - j \frac{49}{15318} \\ -\frac{5}{4073} & -\frac{303}{27661} + j \frac{5}{3207} & -\frac{155}{5493} + j \frac{49}{15318} & \frac{119}{3788} \end{bmatrix},$$

and

$$\mathbf{d}(3142) = \frac{\rho}{4\pi} \begin{bmatrix} \frac{83}{2984} - j \frac{973}{23309} \\ -\frac{362}{9341} - j \frac{499}{5894} \\ -\frac{362}{9341} - j \frac{499}{5894} \\ \frac{83}{2984} - j \frac{973}{23309} \end{bmatrix}. \quad (3.3.21)$$

The condition number  $K(\mathbf{C})$  of  $\mathbf{C}(3142)$  is close to 100, so the equation system is still well-conditioned. The solution for  $\mathbf{v}_0(3142)$  is

$$\mathbf{v}_0(3142) = \frac{4\pi}{\rho} \begin{bmatrix} \frac{620}{12241} - j \frac{13382}{36133} \\ -\frac{2906}{5817} - j \frac{7430}{5263} \\ -\frac{2906}{5817} - j \frac{7430}{5263} \\ \frac{620}{12241} - j \frac{13382}{36133} \end{bmatrix}, \quad (3.3.22)$$

and the corresponding root mean squared error  $E_{\text{rms}}$  is approximately 0.037, which is about eight times smaller than the value for  $E_{\text{rms}}$  in the previous example. This confirms, not surprisingly, that the system comprising four sources is superior to the system comprising only two sources for this particular choice of target area and desired sound field. Figure 3.5 illustrates this for each of the two cases by showing how well the plane wave is reproduced spatially in the  $(x_1, x_2)$ -plane. Figures 3.5a, b and c show the sound field reproduced by the two sources whose inputs are given by Equation (3.3.19), Figures 3.5d, e, and f show the sound field reproduced by the four sources whose inputs are given by Equation (3.3.22). All six plots show the reproduced sound field inside a  $2\text{m} \times 2\text{m}$  square as indicated in Figure 3.4. Figures 3.5a, b, d, and e are “snapshots” of the instantaneous value of  $w(\mathbf{x})$  at one particular time during a cycle. Figures 3.5c and f show the contours of constant phase, which indicate the direction of the active energy flow. It is seen that the plane wave is reproduced more accurately, and over a larger area, by the four sources than by the two sources.

### 3.4 Sampling of spatially varying sound fields

Even though in theory it is possible to calculate the optimal source accelerations that are necessary in order to reproduce a desired sound field, the properties of the desired field are not known exactly in practice. Furthermore, as the frequency increases, it becomes increasingly difficult to evaluate the numerical integrals that determine the elements of  $\mathbf{C}(\omega)$  and  $\mathbf{d}(\omega)$ . In practice, the desired sound field is recorded by a number of microphones. This is equivalent to a spatial sampling of the field. Spatial sampling is a phenomenon that is well known in other engineering disciplines such as digital picture processing (Rosenfeld and Kak [33]) and acoustical holography (Hildebrand and Brenden [15]).

If a sound field  $d(\mathbf{x})$ , for example a plane wave, is observed only over a square whose sidelength is  $l$ , then the “windowed” plane wave  $d_{\text{win}}(\mathbf{x})$  is given by

$$d_{\text{win}}(\mathbf{x}) = d(\mathbf{x})f_{\text{win}}(\mathbf{x}), \quad (3.4.1)$$

where

$$f_{\text{win}}(\mathbf{x}) = f_{\text{win}}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{cases} 1 & \text{for } |x_1| < l/2, |x_2| < l/2 \\ 0 & \text{elsewhere} \end{cases} \quad (3.4.2)$$

The function  $f_{\text{win}}(\mathbf{x})$  that modifies the original sound field is referred to as a "window", or weighting, function (Dudgeon and Mersereau [64] p.118). This function can be thought of as a filter whose effect on any spatial signal can be quantified by analysing the modification of the spatial spectrum of the original signal. The spatial spectrum  $F(\mathbf{k})$  of a two-dimensional spatial signal  $f(\mathbf{x})$  is equivalent to the frequency spectrum of a time signal, and it is defined as a two-dimensional Fourier transform of  $f(\mathbf{x})$  (Hildebrand and Brenden [15] Section 5.1.2),

$$F(\mathbf{k}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{x}) e^{-j\mathbf{x}\cdot\mathbf{k}} d\mathbf{x} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) e^{-j(k_1 x_1 + k_2 x_2)} dx_1 dx_2. \quad (3.4.3)$$

The two-dimensional Fourier transform effectively decomposes a sound field into plane wave components, each propagating in a direction given by the two-dimensional wave vector  $\mathbf{k}$  and a spatial frequency given by the Euclidean length of  $\mathbf{k}$  (we will not consider evanescent waves as discussed by Maynard et al [68]). As with one dimensional signals, multiplication in one domain corresponds to convolution in the other, so

$$D_{\text{win}}(\mathbf{k}) = D(\mathbf{k}) * F_{\text{win}}(\mathbf{k}), \quad (3.4.4)$$

where capital letters denote two-dimensional Fourier transforms. Two-dimensional convolution is a natural generalisation of one-dimensional convolution, it is expressed as

$$D_{\text{win}}(\mathbf{k}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D\left(\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}\right) F_{\text{win}}\left(\begin{bmatrix} k_1 - \alpha_1 \\ k_2 - \alpha_2 \end{bmatrix}\right) d\alpha_1 d\alpha_2. \quad (3.4.5)$$

The spatial spectrum of a plane wave  $D(\mathbf{k})$  is a single delta function whose position corresponds to its wave vector  $\mathbf{k}_D$ ,

$$D(\mathbf{k}) = \delta(\mathbf{k} - \mathbf{k}_D). \quad (3.4.6)$$

Thus, when the spatial spectrum of a plane wave is convolved with that of a weighting function, the net result is that the spectrum of the weighting function is shifted by the amount  $\mathbf{k}_D$ ,

$$D_{\text{win}}(\mathbf{k}) = F_{\text{win}}(\mathbf{k} - \mathbf{k}_D), \quad (3.4.7)$$

this is illustrated in Figure 3.6. The spatial spectrum of a square with sidelength  $l$  is given by (Rosenfeld and Kak [33] Section 2.1.3)

$$W(\mathbf{k}) = 4 \frac{\sin(k_1 \cdot \frac{l}{2})}{k_1} \frac{\sin(k_2 \cdot \frac{l}{2})}{k_2}. \quad (3.4.8)$$

The spectrum of a square weighting function is a product of two  $\sin(x)/x$  functions (Rosenfeld and Kak [33] p.13), the zeros of the  $\sin(x)/x$  function are located at odd multiples of  $2\pi/l$ . The larger the square, the closer to the origin is the first zero of its spectrum. In the limit, when the square covers the whole  $(x_1, x_2)$ -plane, its spectrum is a single delta function positioned at the origin. Since the delta function is the unit function for convolution, this confirms that, if the square is sufficiently large, no significant information about the original sound field is lost. However, in the case of a  $0.5\text{m} \times 0.5\text{m}$  square, the first zero of the spectrum of the weighting function occurs when  $\|\mathbf{k}\|$  is approximately equal to  $12.6\text{m}^{-1}$  which corresponds to a frequency of 680Hz. This means that the spatial spectrum of a sound field containing only one frequency component, such as a plane wave, is quite severely modified. This smearing of the true spectrum is well-known from the time-frequency analysis, where a suitable window function is usually used to reduce the undesirable spectral effects caused by the multiplication by the weighting function (Oppenheim and Schaffer [35] Section 5.5).

When only a finite number of microphones are used to record the sound field within a finite area, the sound field is sampled spatially. The spatial sampling is described mathematically by a weighting function  $f_w(\mathbf{x})$ , which is a weighted sum of  $R$  delta functions, each one positioned at the location  $\mathbf{x}_r$  of a microphone, therefore

$$f_w(\mathbf{x}) = \frac{\Omega}{R} \sum_{r=1}^R \delta(\mathbf{x} - \mathbf{x}_r), \quad (3.4.9)$$



where  $\Omega$  is the size of the area over which the sound field is sampled. The spectrum  $F_{\delta}(\mathbf{k})$  of a single delta function at  $\mathbf{x}_r$  is given by

$$F_{\delta}(\mathbf{k}) = e^{-j\mathbf{k} \cdot \mathbf{x}_r}, \quad (3.4.10)$$

and since the Fourier transform is linear, the spectrum of  $f_w(\mathbf{x})$  is given by

$$F_w(\mathbf{k}) = \frac{\Omega}{R} \sum_{r=1}^R e^{-j\mathbf{k} \cdot \mathbf{x}_r}. \quad (3.4.11)$$

This result can also be derived by sampling the function on an infinite lattice after windowing it (Hildebrand and Brenden [15] Section 5.1.2, Rosenfeld and Kak [33] Sections 4.1.2-3, Dudgeon and Mersereau [64] Section 1.4). One of the advantages of using a sampling lattice is that it reveals the periodic structure of the spectrum of a sampled signal. Here, we illustrate the spectral effects of spatial sampling with a few examples. Figure 3.7 shows the magnitudes of four spatial spectra in the range from  $\|\mathbf{k}\| = -20\text{m}^{-1}$  to  $20\text{m}^{-1}$  ( $\|\mathbf{k}\|$  equal to  $20\text{m}^{-1}$  is equivalent to a frequency of 1080Hz). Each of the four spectra corresponds to a weighting function defined on a  $0.5\text{m} \times 0.5\text{m}$  square. Figure 3.7a shows the spectrum of the continuous weighting function given by Equation (3.4.2). Figure 3.7b shows the spectrum for a 5-by-5 "microphone" array, it is seen that it is quite similar to that of the continuous weighting function. Figures 3.7c and d illustrate the periodic structure of the spectrum for a 3-by-3 and a 2-by-2 microphone array respectively. Note the many repetitions of the part of the spectrum that is concentrated near the origin for the 2-by-2 array. These repetitions are referred to as "images". As the number of microphones decreases, the images move closer to the origin. When there is only one microphone, the magnitude of the spectrum is constant for all values of  $\mathbf{k}$ , which means that there is no spatial information in the sampled sound field. When the number of microphones is very large, the images are very far away from the origin, and the part of the spectrum of the weighting function that is near the origin is almost identical to that of the continuous weighting function. The normalisation factor  $\Omega/R$  in Equation (3.4.11) ensures that the sum converges as  $R$  tends towards infinity. The convergence can be proved formally by applying the scaling property of two-dimensional Fourier transforms to a windowed comb function (Rosenfeld and Kak [33] Sections 2.1.3-4). The windowed comb function is a set of delta functions defined over the square, and

the scaling property gives an expression for the spectrum as a function of the "density" of the delta functions.

When a two-dimensional sound field is known only over a finite area  $\Omega$ , it is intuitively obvious that some information about the original sound field is inevitably lost. The information in a sound field is defined as the number of samples of the field that are necessary in order to reconstruct the field perfectly from its sampled values (Hildebrand and Brenden [15] Section 5.2). The basic rule is that no information is lost if at least two samples are taken per wavelength. The general formulation is that no information is lost if the number  $I$  of samples taken is at least

$$I = F \Omega, \quad (3.4.12)$$

where  $F$  is the size of the support of the spatial spectrum (Dudgeon and Mersereau [64] p.176). The support of a spectrum is the set of arguments for which the spectrum is non-zero. For a two-dimensional square whose sidelength is  $l$ ,  $I$  is given by

$$I = \left( \frac{2f_{\max} l}{c} \right)^2 = \left( \frac{2l}{\lambda_{\min}} \right)^2, \quad (3.4.13)$$

where  $f_{\max}$  is the signal's highest frequency component, and  $\lambda_{\min}$  is its corresponding acoustical wavelength. For a  $0.5\text{m} \times 0.5\text{m}$  square,  $I$  is approximately 2.16 when  $f_{\max}$  is 500Hz. For a three-dimensional cube whose sidelength is  $l$ ,  $I$  is given by

$$I = \left( \frac{2f_{\max} l}{c} \right)^3 = \left( \frac{2l}{\lambda_{\min}} \right)^3. \quad (3.4.14)$$

When  $l$  is one meter and  $f_{\max}$  is 17kHz,  $I$  is one million (1.000.000). Consequently, one million microphones are necessary in order to record a sound field in a  $1\text{m} \times 1\text{m} \times 1\text{m}$  cube over the full audio bandwidth without losing any information (Gerzon [13]). It is worth pointing out that the expression for  $I$  does not take into account any a priori information about the sound field. For example, if the sound field is known to be a plane wave, then the sound field is specified completely not just in  $\Omega$ , but in the entire three-dimensional space, by its wave vector  $\mathbf{k}_D$  plus a complex amplitude. In that case, only a few samples are necessary to determine the sound field completely.

### 3.5 Reproduction of spatially sampled sound fields

It is now assumed that a spatially varying sound field is sampled by  $R$  microphones in a "recording" space (equivalent to "source space" in Cooper and Shiga [37]). The objective is to use these recorded signals to reproduce the sound field at geometrically similar positions in a listening space. This is illustrated in Figure 3.8. Since there are  $R$  recorded signals  $U_r(\omega)$ , there are also  $R$  reproduced signals  $W_r(\omega)$  and  $R$  error signals  $E_r(\omega)$ , so

$$\mathbf{u}(\omega) = \begin{bmatrix} U_1(\omega) \\ U_2(\omega) \\ \vdots \\ U_R(\omega) \end{bmatrix}, \quad \mathbf{w}(\omega) = \begin{bmatrix} W_1(\omega) \\ W_2(\omega) \\ \vdots \\ W_R(\omega) \end{bmatrix}, \quad \mathbf{e}(\omega) = \begin{bmatrix} E_1(\omega) \\ E_2(\omega) \\ \vdots \\ E_R(\omega) \end{bmatrix}. \quad (3.5.1)$$

The reproduced signals are calculated by adding the contributions from all the sources together, so

$$\mathbf{w}(\omega) = \mathbf{C}(\omega)\mathbf{v}(\omega), \quad (3.5.2)$$

where

$$\mathbf{C}(\omega) = \begin{bmatrix} C_{11}(\omega) & C_{12}(\omega) & \cdots & C_{1S}(\omega) \\ C_{21}(\omega) & C_{22}(\omega) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ C_{R1}(\omega) & \cdots & \cdots & C_{RS}(\omega) \end{bmatrix}. \quad (3.5.3)$$

The  $R$  times  $S$  matrix  $\mathbf{C}(\omega)$  contains the transfer functions from each source to each microphone. There is a row in  $\mathbf{C}(\omega)$  for each microphone, and a column for each loudspeaker. The source input vector  $\mathbf{v}(\omega)$  is a column vector with  $S$  components as given by Equation (3.3.3). Equation (3.5.2) is equivalent to Equation (3.3.1). The desired signals  $\mathbf{d}(\omega)$  are exact copies of the recorded signals  $\mathbf{u}(\omega)$ , so

$$\mathbf{d}(\omega) = \mathbf{u}(\omega). \quad (3.5.4)$$

The error signals  $\mathbf{e}(\omega)$  are given by the difference between the desired signals and the reproduced signals, and by substituting Equation (3.5.4) into Equation (3.5.2), we find

$$\mathbf{e}(\omega) = \mathbf{d}(\omega) - \mathbf{w}(\omega) = \mathbf{d}(\omega) - \mathbf{C}(\omega)\mathbf{v}(\omega), \quad (3.5.5)$$

this expression is equivalent to Equation (3.3.4). It is now possible to solve the linear equation system

$$\mathbf{C}(\omega)\mathbf{v}(\omega) = \mathbf{d}(\omega) \quad (3.5.6)$$

in the least squares sense as a function of  $\mathbf{v}(\omega)$ . A performance cost  $E(\omega)$  equivalent to Equation (3.3.5) is defined according to Equation (2.3.8),

$$E(\omega) = \sum_{r=1}^R |e_r(\omega)|^2 = \mathbf{e}^H(\omega)\mathbf{e}(\omega), \quad (3.5.7)$$

and a root mean squared error  $E_{\text{rms}}(\omega)$  equivalent to Equation (3.3.7) is defined as

$$E_{\text{rms}}(\omega) = \sqrt{\frac{E(\omega)}{R}}, \quad (3.5.8)$$

which is a measure of the average error per microphone. Equation (2.4.5) gives the solution  $\mathbf{v}_0(\omega)$  that minimises  $E(\omega)$  as a function of  $\mathbf{v}(\omega)$ ,

$$\mathbf{v}_0(\omega) = [\mathbf{C}^H(\omega)\mathbf{C}(\omega) + \beta \mathbf{I}]^{-1} \cdot \mathbf{C}^H(\omega)\mathbf{d}(\omega) \quad (3.5.9)$$

for any choice of the desired signals  $\mathbf{d}(\omega)$ . Consequently, if the desired signals are passed through a filter matrix  $\mathbf{H}_0(\omega)$  given by

$$\mathbf{H}_0(\omega) = [\mathbf{C}^H(\omega)\mathbf{C}(\omega) + \beta \mathbf{I}]^{-1} \cdot \mathbf{C}^H(\omega) \quad (3.5.10)$$

before they are input to the loudspeakers, the error is minimised for any set of recorded signals  $\mathbf{u}(\omega)$ . The  $S$  times  $R$  matrix  $\mathbf{H}_0(\omega)$  is referred to as the inverse filter matrix, its input is a vector of recorded signals, its output is a vector containing the inputs to the loudspeakers. It is seen that  $\mathbf{H}_0(\omega)$  is the pseudo-inverse of  $\mathbf{C}(\omega)$  with an optional regularisation factor  $\beta$ . The problem is shown in block diagram form in Figure 3.9. Trott [113], in 1964, was the first to use this type of filter design to reproduce a single-frequency plane wave in the least squares sense. Van Buren [109] later refined Trott's technique so that any desired sound field could be reproduced optimally in the least squares sense. However, the optimal filter matrix only implemented simple phase and gain adjustments, referred to as "shadings", the optimal filters were therefore independent of the frequency. In 1978, Van Buren [94] also included a regularisation factor in his filter design. This technique was already known from work on antenna patterns (Mautz and Harrington [29]).

It was demonstrated in the previous section that the spatial spectrum of a sampled sound field converges to the spatial spectrum of the continuous sound field as the number of microphones tends towards infinity. Consequently, the solution  $\mathbf{v}_0(\omega)$  must also converge to the solution for the continuous case given by Equation (3.3.13), and so must the root mean squared error  $E_{\text{rms}}$ . An example of the convergence of  $E_{\text{rms}}$  is illustrated in Figure 3.10 for the geometry shown earlier in Figure 3.4a (two point monopoles at (-1.4m, -2.4m) and (1.4m, -2.4m), 0.5m  $\times$  0.5m square target area). The desired sound field is a plane wave travelling along the positive  $x_2$ -axis, and its frequency is 500Hz. Figure 3.10 shows  $E_{\text{rms}}$  as a function of  $R^{1/2}$ , which is the square root of the number of microphones used inside the target area. The microphones are evenly spread over the target area as indicated by Figure 3.7 for the cases  $R$  equal to 4, 9, and 25. The sources are assumed to be point monopole sources, so all the electro-acoustic transfer functions contained in  $\mathbf{C}(\omega)$  are of the form given by Equation (3.2.34). It is seen that  $E_{\text{rms}}$  converges relatively slowly to the limit value 0.29 for the continuous case. The convergence is guaranteed because the scaled normal equations

$$\frac{\Omega}{R} [\mathbf{C}^H(\omega) \mathbf{C}(\omega)] \mathbf{v}(\omega) = \frac{\Omega}{R} \mathbf{C}^H(\omega) \mathbf{d}(\omega) \quad (3.5.11)$$

approach Equation (3.3.8) since the vector dot product approximates the inner product with increasing accuracy as  $R$  becomes larger (Keener [77] p.5).

### 3.6 Reproduction of plane wave sound fields

We now investigate quantitatively the extent to which a number of monopole sources can reproduce a harmonic plane wave as a function of its frequency  $f$  and its angle of incidence  $\theta_{\text{aoi}}$ . The parameters used to specify the plane wave are shown in Figure 3.11. As indicated in Figure 3.11a, when  $\theta_{\text{aoi}}$  is measured in degrees,  $\theta_{\text{aoi}} = 0^\circ$  means that the plane wave is going to the 'right' in the positive direction of the  $x_1$ -axis,  $\theta_{\text{aoi}} = 90^\circ$  is 'up',  $\theta_{\text{aoi}} = 180^\circ$  is 'left', and  $\theta_{\text{aoi}} = 270^\circ$  is 'down'. Thus, for the plane wave considered in the previous examples,  $\theta_{\text{aoi}}$  is  $90^\circ$  and  $f$  is 500Hz. A component of the plane wave in the  $x_3$ -direction is indicated by a separate parameter,  $\theta_{\text{ver}}$  (subscript ver is for vertical). When the value of  $\theta_{\text{ver}}$  is not given explicitly, it is assumed to be zero.

Since the solution to the discrete sound reproduction problem converges to the solution of the continuous problem, it is reasonable to assume that there is a certain “density” of the microphone array beyond which the solution no longer depends on the number of microphones. As mentioned in Section 3.4, no information is lost if at least two microphones per wavelength are used, so a crude estimate for the maximum spacing between adjacent microphones is half a wavelength. The four plots in Figure 3.12 illustrate that this simple result is remarkably accurate. The four plots show the optimal total source input  $\|v_0\|$  calculated without regularisation (see Equation (2.3.10)), for four microphone arrays when the frequency and angle of incidence of the plane wave are in the ranges  $0\text{Hz} < f < 3\text{kHz}$  and  $0^\circ < \theta_{\text{aoi}} < 180^\circ$ . The sources and the target area are the same as in the previous examples, the geometry is shown in Figure 3.4a (two point monopoles at  $(-1.4\text{m}, -2.4\text{m})$  and  $(1.4\text{m}, -2.4\text{m})$ ,  $0.5\text{m} \times 0.5\text{m}$  square target area). The results shown in Figure 3.12a are calculated by using a “dense” 9-by-9 microphone array, it can be considered to be a “reference” because it is effectively uncontaminated by spatial aliasing. Figures 3.12b, c and d show the equivalent results for a 5-by-5, 3-by-3, and 2-by-2 microphone array respectively. It is seen that compared to Figure 3.12a, spurious peaks start to appear at approximately 2.8kHz for the 5-by-5 array, 1.4kHz for the 3-by-3 array, and 700Hz for the 2-by-2 array. The acoustical wavelength at these frequencies is almost exactly twice the distance between adjacent microphones in each case. Consequently, two microphones per wavelength is not quite enough to avoid the effects of spatial aliasing completely. A rule of thumb is to use at least three microphones per acoustical wavelength.

The three different layouts of loudspeakers and microphones shown in Figures 3.13a, b, and c are used to illustrate the general behaviour of the error  $E_{\text{rms}}$  and the solution  $v_0$ . These three arrangements have been chosen to correspond approximately to a) a “conventional stereo” loudspeaker layout, b) a four loudspeaker layout which roughly mimics “quadraphony”, and c) a “narrow-arc” four-loudspeaker arrangement. A “dense” 9-by-9 microphone array is positioned in a target area which is a  $0.5\text{m} \times 0.5\text{m}$  square centred at the origin of the coordinate system, consequently the effects of spatial aliasing are insignificant. These arrangements model the situation in a normal living room quite well, the microphone array represents a listener sitting approximately 2.5m away from each of the speakers (a realistic model of a living

room would also have to include room reflections). If the sound field is well reconstructed over the entire microphone array, the listener would not have to sit completely still, but he can move his head moderately and still experience a good spatial reproduction of the sound field.

Figures 3.14a, b, and c illustrate some of the limitations that inevitably apply to any sound reproduction system. The three figures show  $E_{rms}$  for each of the configurations in Figure 3.13 as a function of the frequency  $f$  and angle of incidence  $\theta_{aoi}$  of the plane wave. Consider first the effect of changing the angle of incidence. When the angle corresponds to the plane wave coming directly from one loudspeaker,  $E_{rms}$  is small. As the angle moves away from one loudspeaker and towards another,  $E_{rms}$  goes up, reaches a local maximum, and then goes down again as the angle approaches the adjacent loudspeaker. The greater the angle between the two speakers is as seen from the centre of the target area, the greater is the peak value of  $E_{rms}$ . When the angle moves outside the loudspeakers,  $E_{rms}$  approaches 1 rapidly. These results agree well with general experience gained from listening tests, they show that if the angle between two loudspeakers is greater than  $60^\circ$ , it sounds as if there is a "hole" in the middle, while location is excellent when the sound comes directly from one loudspeaker only (de Boer [7], Theile and Plenge [55], Eargle [91] p.77). Now consider the frequency dependence. At very low frequencies,  $E_{rms}$  is small and almost independent of the angle, which indicates that there is very little directional information in the recorded signals. This happens because the target area is very small compared to the acoustical wavelength; therefore there is not enough phase variation over the target area to discriminate the direction of arrival of the plane wave. As the frequency increases, the "holes" between adjacent loudspeakers start to appear, and it is evident that any loudspeaker arrangement eventually fails above a certain frequency because the angle between adjacent loudspeakers is too large. In this respect, quadraphony looks almost useless, and even the conventional stereo arrangement works well only up to about 600Hz. The four-loudspeaker arrangement covers  $90^\circ$  up to over 1000Hz with only moderate errors ( $E_{rms} = 0.5$  at 1200Hz). In practice, it is difficult to increase this upper frequency limit further because the sound field is too complex to reconstruct with only a few loudspeakers when the target area is more than just a few wavelengths long. In theory, this problem can be overcome by

increasing the number of loudspeakers, but this is not an attractive alternative, first because it is impractical, and secondly because it is well known from experience with active control systems that the harder one tries to control the sound in one region, the more unpredictable and undesirable effects one gets outside that region (Nelson and Elliott [9]).

The least squares error is the most important measure of whether a given plane wave can be reproduced at all. If the least squares error is close to 1, it is a waste of effort to even try to make use of the solution. The model should reflect this fact by giving small values for the total source input  $\|v_0\|$  when  $E_{rms}$  is large. This is a way of saying that silence is the worst approximation to a plane wave. If the least squares error is small,  $\|v_0\|$  should have a moderate value, almost independent of the angle and the frequency, indicating that the effort goes into creating the plane wave and not into producing unwanted side-effects somewhere else outside the target area. Figures 3.15a, b, and c show how  $\|v_0\|$  varies for the "stereo"- , "quadraphony"- and the "narrow-arc" arrangements respectively. At high frequencies, the results are exactly as described above. There is an almost perfect negative correlation between  $E_{rms}$  and  $\|v_0\|$ . However, for moderately high frequencies,  $\|v_0\|$  increases outside the interval that the loudspeakers are supposed to cover, and it even takes its peak value outside this interval. This is because the loudspeakers try hard to make an interference field that matches the plane wave and, as seen from the corresponding plots of  $E_{rms}$ , they succeed quite well in doing so at low frequencies but not so well in the mid- and high-frequency range.

The conditioning of the equation system is illustrated in Figure 3.16. The left column shows the largest and smallest singular value of  $C^H C + \beta I$ , the right column shows the condition number of  $C^H C + \beta I$ . It is seen that the largest singular value is largely independent of the frequency, whereas the smallest singular value takes its minimum value when the frequency is zero. Once the largest singular value is known, the value of the regularisation parameter  $\beta$  can be calculated by using Equation (2.4.8). According to Equation (2.4.7), a value of 0.0002 for  $\beta$  should give a condition number of approximately 2500 if the largest singular value is 0.5, and the results shown in Figure 3.16 are in good agreement with this prediction. Since the



regularisation has a significant influence on the smallest singular value, but only negligible effect on the largest singular value, the net effect is a reduction of the condition number at low frequencies where the ill-conditioning is worst. The ill-conditioning is worst at low frequencies because there is very little variation between the phase of the different elements of  $\mathbf{C}(\omega)$ . They only differ by their amplitudes. In the limit when the frequency is zero, all the elements of  $\mathbf{C}(\omega)$  are real. It is difficult to make any other general statements about the condition number for the equation system generated by a certain layout of loudspeakers and microphones. Addition of more loudspeakers tends to make the conditioning worse, and it also looks like layouts that have a high degree of geometrical "symmetry" tend to be badly conditioned. Certain symmetric loudspeaker-microphone layouts make  $\mathbf{C}(\omega)$  singular for all frequencies. This happens when the "inverse distance matrix", which contains the inverse distances from the loudspeakers to the microphones, is singular (see later discussion in Section 6.3).

### 3.7 Direction-of-arrival reproduction

It is not surprising that, with only a few loudspeakers, it is not possible to control the sound field around a single listener's head at frequencies higher than a few kHz unless the angle of incidence corresponds to the position of a loudspeaker. Consequently, in order to reproduce a plane wave with a high frequency, the best that can be done is to radiate sound only from the loudspeaker whose position correspond to the angle of incidence of the plane wave. We refer to this type of reproduction as direction-of-arrival reproduction. The principle is illustrated in Figure 3.17.

As demonstrated in the previous section, a large mean squared error indicates a poor reproduction of the sound field, which means that the optimal solution is almost entirely useless for the purpose it is designed for. On the other hand, a small mean squared error does not necessarily mean that the "right" speaker switches on for the "right" angle. This is indicated in Figure 3.15c by some large source inputs for angles of incidence well outside the range of angles that the sources cover. We now look at the individual source outputs that add up to the results shown in Figure 3.15c. Figure 3.18 shows the mean squared error  $E_{\text{rms}}$ , the total source input  $\|\mathbf{v}_0\|$ , and the four individual source inputs  $|V_1|$ ,  $|V_2|$ ,  $|V_3|$  and  $|V_4|$  in polar form for the two frequencies

1kHz (left column) and 500Hz (right column). The individual source inputs for  $f$  equal to 1kHz confirm that the source “closest” to the direction of the plane wave dominates, and that the sources do not radiate too much sound as the angle moves outside the loudspeakers. However, this is no longer true when the frequency is decreased to 500Hz. A simple way to avoid making the system try to reproduce angles of incidence in the recorded field that are not covered by the speakers is to use an additional number of “imaginary” speakers for the design of the filter matrix. The additional speakers must be positioned in such a way that all angles are covered by either a real or an imaginary speaker. For each imaginary speaker, there is an extra column in  $\mathbf{C}$ , and consequently also an extra row in  $\mathbf{H}$ . The extra rows and columns must then be discarded before passing a set of recorded signals through the system. An example of this technique is given in Figure 3.19. However, since phase variation is not taken into account, this approach might not work. From here on, we only consider loudspeaker setups that cover the full  $360^\circ$  in the  $(x_1, x_2)$ -plane.

The physical limitations that apply to sound field reproduction also apply to direction-of-arrival reproduction. Compared to the wavelength, the target area must still not be very small or very large, and the spacing between adjacent microphones must still be relatively small. In order to make the direction-of-arrival system work at frequencies above 1kHz, the target area must be quite small, for example a  $5\text{cm} \times 5\text{cm}$  square. When the target area is that small, it is difficult in practice to use a large number of microphones. Since it is not desirable to use a large number of microphones anyway, it is natural also to consider the use of fewer microphones than loudspeakers. The minimum number of microphones for direction-of-arrival reproduction is three because a microphone array consisting of only two microphones is effectively one-dimensional. Until now, all microphone arrays have been arranged on a square grid. This sampling scheme has the advantage that it is easy to analyze mathematically. However, the microphone array can also be arranged on a circle, or a set of circles, this sampling scheme sometimes performs better than the square grid (Hildebrand and Brenden [15] Section 5.3).

The four different microphone arrays shown in Figure 3.20 are now compared. The four arrays are referred to as a) 3-by-3, b) 2-by-2, c) circular 3 plus centre, and d) circular 3 respectively. Each microphone array covers an area of approximately

25cm<sup>2</sup>. The loudspeaker arrangement is the same in each case, eight loudspeakers are positioned in a “perfect” circle whose radius is approximately 2.83m. Figure 3.21 shows the largest and smallest singular value of  $\mathbf{C}$  and the condition number  $K$  of  $\mathbf{C}^H\mathbf{C}$  as a function of the frequency for each of the four microphone arrays. The value of  $K(f)$  is the squared ratio between the largest and the smallest singular value of  $\mathbf{C}(f)$ . It is seen that the largest singular value is almost independent of the frequency whereas the smallest singular value tends to increase with the frequency, and it takes its minimum at 0Hz. Note that the smallest singular value decreases significantly at low frequencies when a centre microphone is added to the circular 3 array. This effectively means that adding the centre microphone does not increase the rank of  $\mathbf{C}(f)$  at low frequencies, and it is an example of how a high degree of symmetry in the layout of loudspeakers and microphones can lead to ill-conditioning. According to Equation (2.4.8), a regularisation factor  $\beta$  of 0.00002 is chosen for all four microphone arrays to avoid the ill-conditioning at low frequencies.

Figure 3.22 shows the input  $|V_1|$  to source number one, which is positioned at (0m,-2.83m), as a function of the frequency  $f$  and the angle of incidence  $\theta_{aoi}$  of the plane wave for each of the four microphone arrays. Each of the four plots has a “main lobe” at  $\theta_{aoi} = 90^\circ$ , this value corresponds to the angle that is covered by the source. Ideally, the main lobe is narrow and much larger than the “sidelobes”. The most desirable input  $|V_1|$  is achieved with the 3-by-3 array. Not only is the frequency at which spatial aliasing start to occur significantly higher than for the other three microphone arrays, the main lobe is also significantly narrower. The reason for this has to do with the compromise made by minimising the sum of the performance cost  $E$  and the effort cost  $\beta V$ . When there are more microphones than loudspeakers, it is not possible to drive  $E$  to zero; consequently the effort cost  $\beta V$  is relatively small compared to the performance cost  $E$ . The regularisation therefore works very well in the over-determined case; it prevents the loudspeaker outputs from becoming very large, but it still makes the sound come mainly from the loudspeaker that corresponds to the angle of incidence of the plane wave in order to prevent the performance cost from becoming too large. When there are fewer microphones than loudspeakers, there are many solutions that all give zero performance cost, consequently  $E$  is relatively small compared to  $\beta V$ . Since the effort cost is the dominating term, the system tends

to “distribute” its power to several loudspeakers rather than just a single one. For example, if four loudspeakers each give an output of magnitude 1, then the total effort  $V$  is 4. However, if one loudspeaker gives an output of magnitude 4, then the total effort  $V$  is 16. Furthermore, at very low frequencies the inversion problem is so ill-conditioned that the total effort goes down because of the use of regularisation. This phenomenon is demonstrated by Figure 3.23 which shows the total loudspeaker input  $\|v_0\|$  for a) the 3-by-3 microphone array and b) the circular 3 array. It is seen that  $\|v_0\|$  is largely independent of  $\theta_{aoi}$ , apart from at very high frequencies, but that  $\|v_0\|$  decreases with the frequency. Figure 3.24 shows  $|V_1|$  as a function of frequency for a fixed angle of incidence of  $90^\circ$  for a) the 3-by-3 microphone array and b) the circular 3 array (same arrays as used in Figure 3.20). Thus, the two plots are slices of Figures 3.22a and 3.22d along the frequency axis. The output is small at low frequencies in both cases because the inversion problem is ill-conditioned. There does not seem to be a simple way of telling exactly how ill-conditioned the problem is at low frequencies (see later discussion in Section 6.3).

Finally, we consider the effect of plane waves that come in from outside the horizontal  $(x_1, x_2)$ -plane. As shown in Section 3.6, attempting to reproduce a plane wave whose angle of incidence is not covered by the loudspeakers usually leads to undesirable results, particularly in terms of direction-of-arrival reproduction. Figure 3.25 shows the input  $|V_1|$  to loudspeaker number one for the three values a)  $30^\circ$ , b)  $45^\circ$ , and c)  $70^\circ$  of the vertical component  $\theta_{ver}$  of the plane wave (see Figure 3.11b) when the microphone array is the circular 3 (see Figure 3.20d). It is seen that  $|V_1|$ , as expected, becomes less dependent on  $\theta_{aoi}$  as  $\theta_{ver}$  increases. When  $\theta_{ver}$  is  $30^\circ$ ,  $|V_1|$  is very similar to the case where the plane wave is in the horizontal plane (see Figure 3.22d). This means that even a relatively large vertical component has little influence on the source inputs. The reason for this is that since the microphone array is “flat”, having all the microphones positioned in the  $(x_1, x_2)$ -plane, the system does not have any “resolution” in the vertical direction. Thus, the influence of unwanted vertical components is not likely to cause any problems.

### 3.8 Conclusions

The results of the frequency domain analysis demonstrate some of the physical limitations that apply to sound field reproduction. The quality of the reproduced sound field is mainly determined by three parameters. The first is the size of the target area, the second is the angles between the loudspeakers as seen from the centre of the target area, and the third is the distance between adjacent microphones. The maximum dimension of the target area is of the order of a few acoustical wavelengths. Specifically, it is possible to reproduce a sound field over a  $0.5\text{m} \times 0.5\text{m}$  square for frequencies below 1kHz with four loudspeakers per  $90^\circ$ , assuming that spatial aliasing is not introduced by using a too large microphone spacing. In order to avoid spatial aliasing, a rule of thumb is to use three microphones per wavelength. Nothing is gained by using a denser microphone array. For direction-of-arrival reproduction, a desired sound field recorded with three, or four, microphones can be produced over a  $5\text{cm} \times 5\text{cm}$  target area at frequencies up to approximately 4kHz. Here the limit is set by spatial aliasing. At very low frequencies, the inversion problem is usually ill-conditioned, so in order to avoid very large loudspeaker outputs it is necessary to use regularisation.

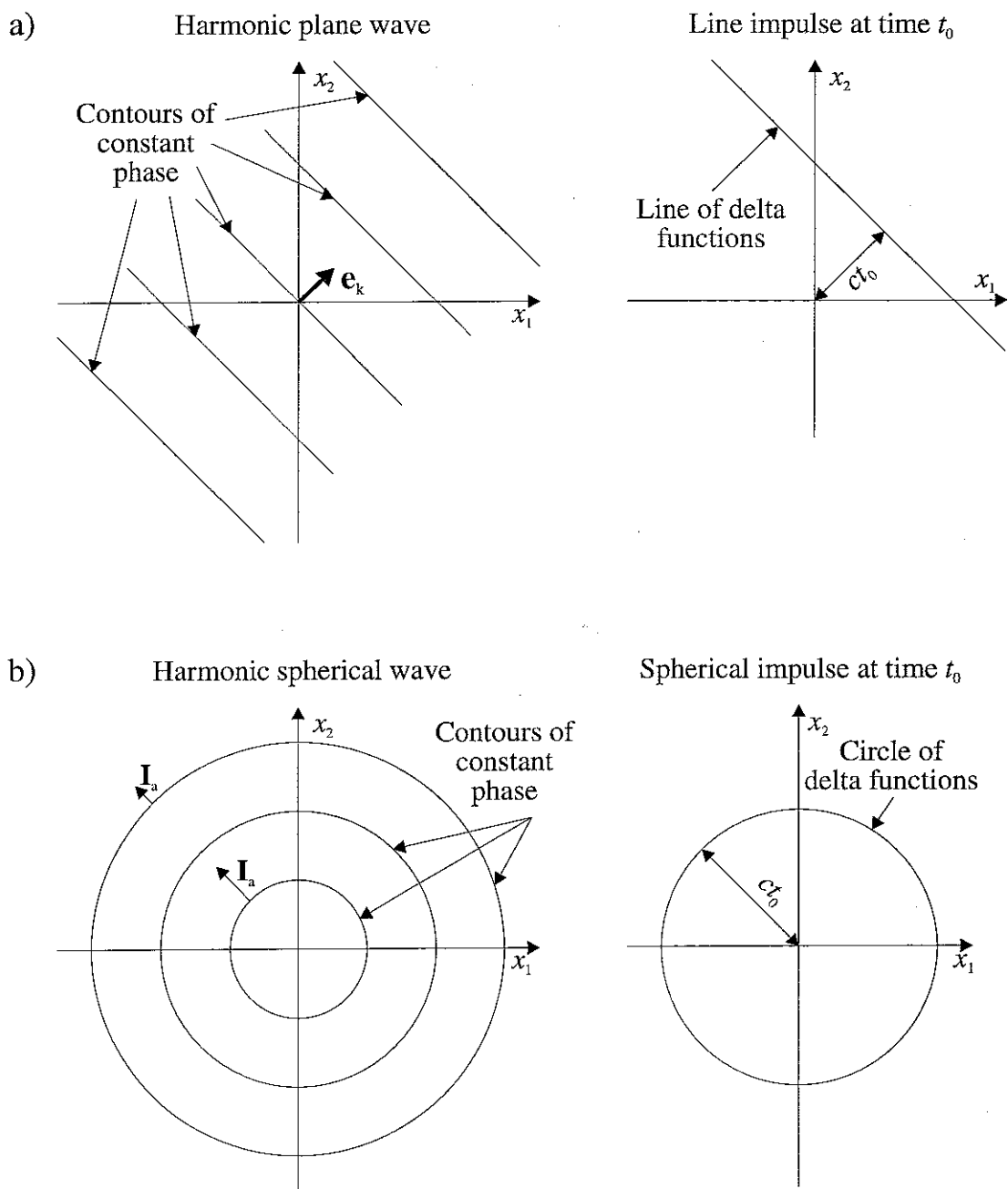


Figure 3.1. Two-dimensional slices of the contours of constant phase and the time domain representation of a) a plane wave whose  $x_3$ -component is zero, and b) a three-dimensional spherical wave

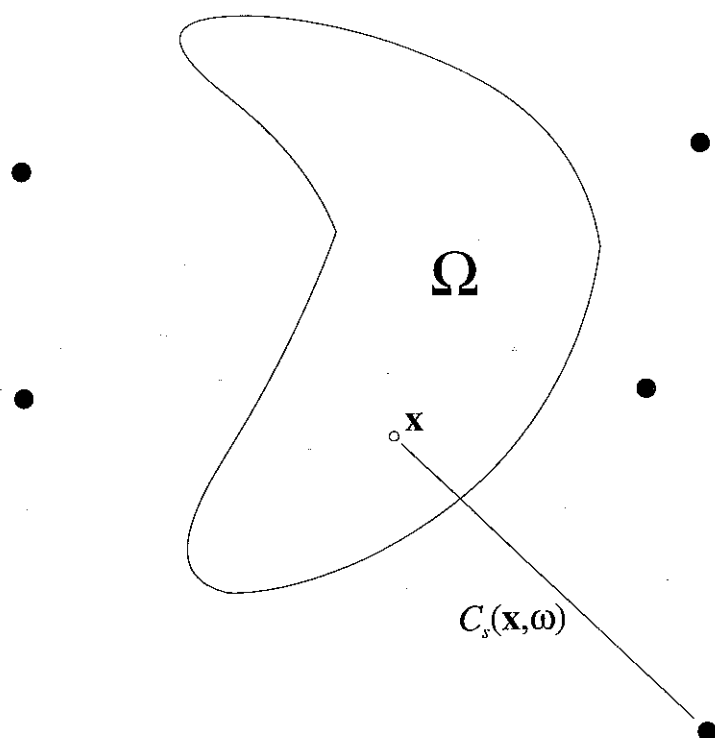


Figure 3.2. The transfer function  $C_s(x, \omega)$  from a source to a field point  $x$ . The sources are filled black spheres

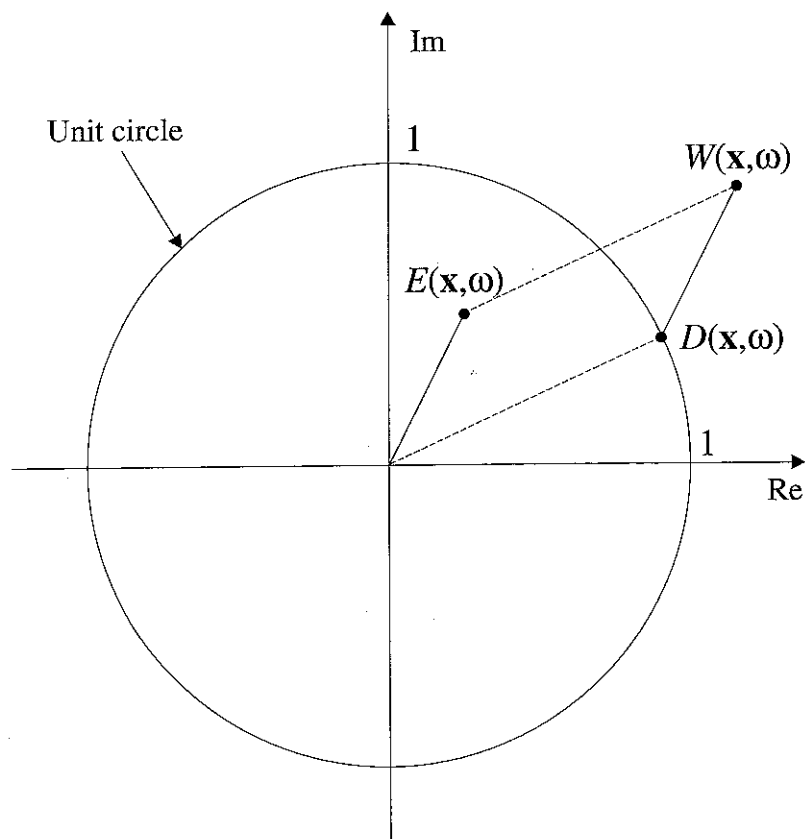


Figure 3.3. The relationship between the desired signal  $D(\mathbf{x}, \omega)$ , the reproduced signal  $W(\mathbf{x}, \omega)$ , and the error  $E(\mathbf{x}, \omega)$  at a field point  $\mathbf{x}$ . When the desired signal is a plane wave whose amplitude is one,  $D(\mathbf{x}, \omega)$  is on the unit circle



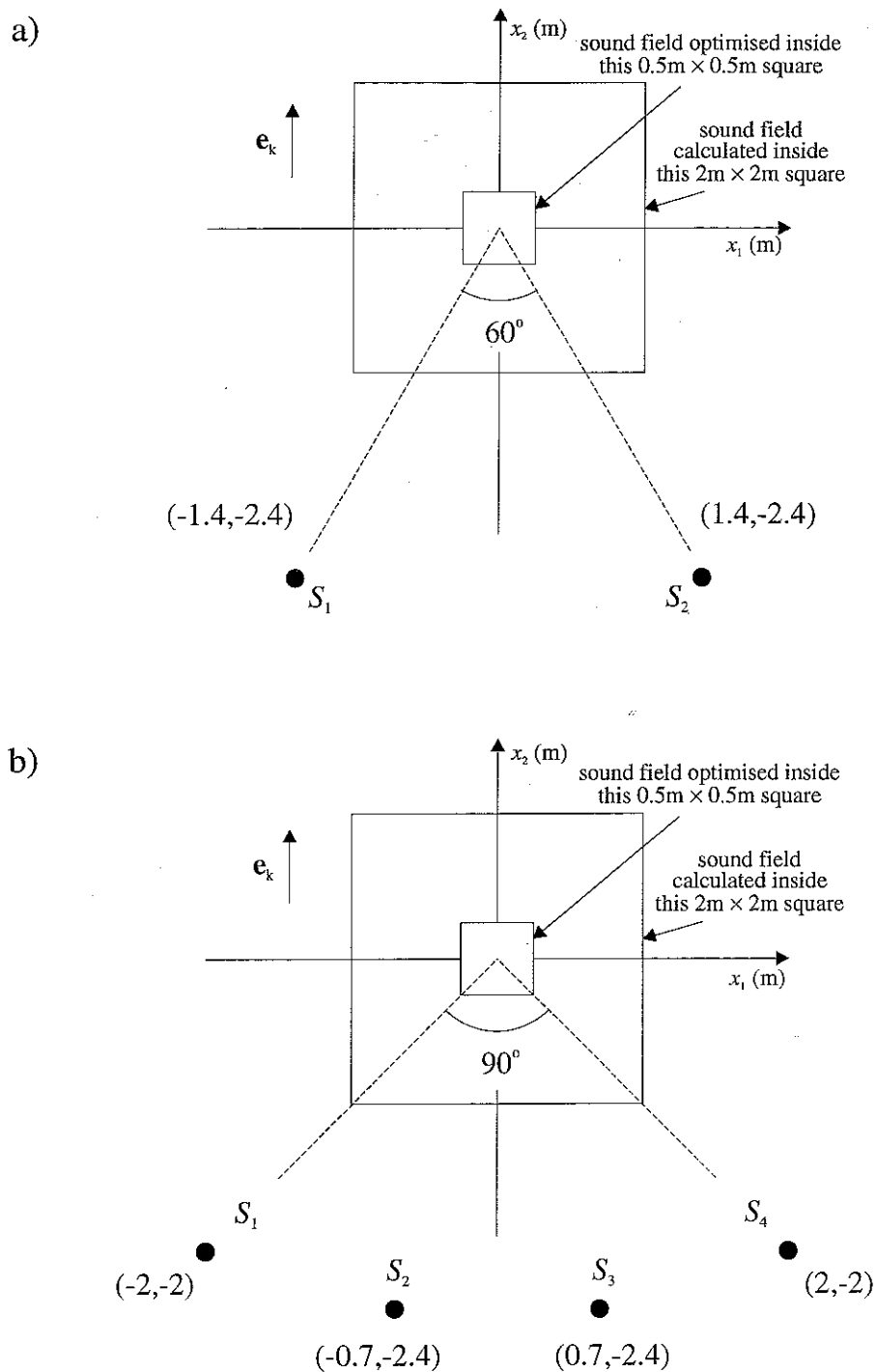
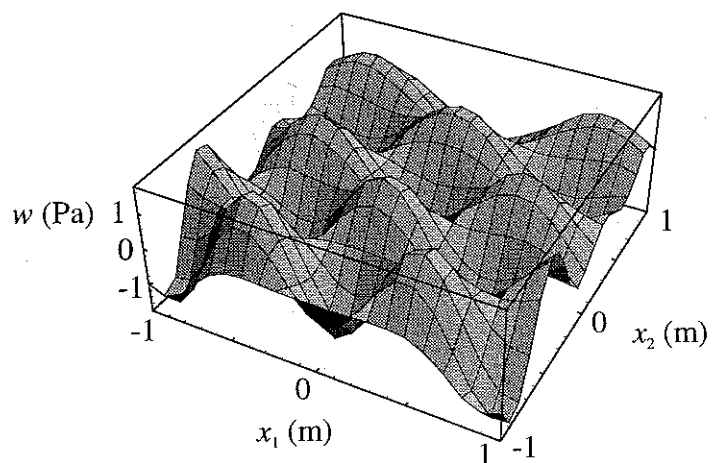
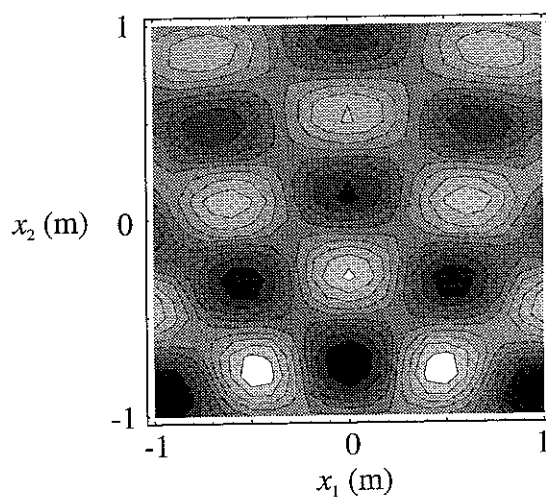


Figure 3.4. Two geometries used to calculate the optimal inputs  $\mathbf{v}_0$  to the sources and the reproduced sound field  $w(\mathbf{x})$ . The target area is a  $0.5\text{m} \times 0.5\text{m}$  square, the desired sound field is a plane wave with a frequency of  $500\text{Hz}$  travelling parallel to the positive direction of the  $x_2$ -axis

a)



b)



c)

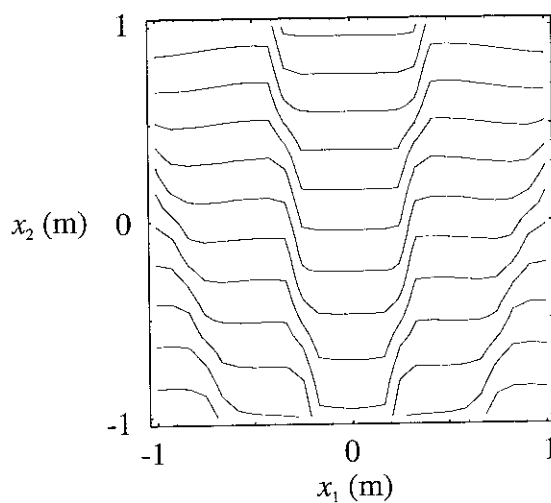
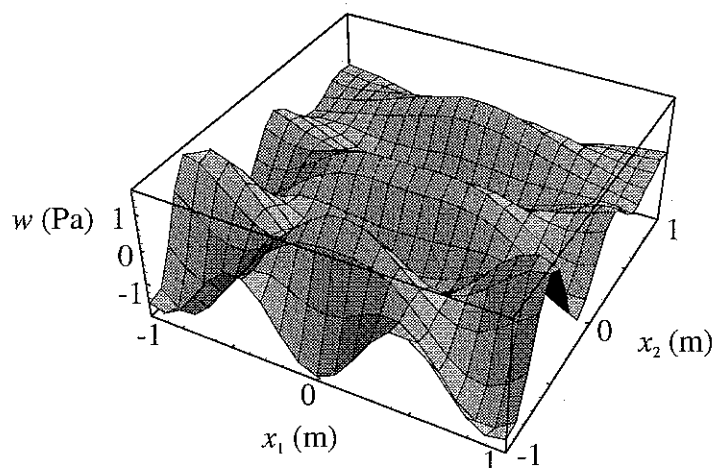
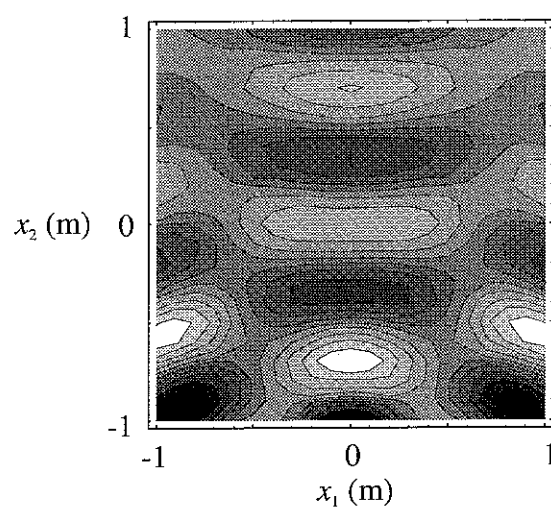


Figure 3.5 (first page of two). The sound field  $w(\mathbf{x})$  reproduced by optimally adjusting the two sources shown in Figure 3.4a. The frequency is 500Hz, a) is a 3D time "snapshot" of  $w(\mathbf{x})$ , b) is a 2D contourplot of a). c) shows the contours of constant phase, the distance between adjacent contours is  $180^\circ$

d)



e)



f)

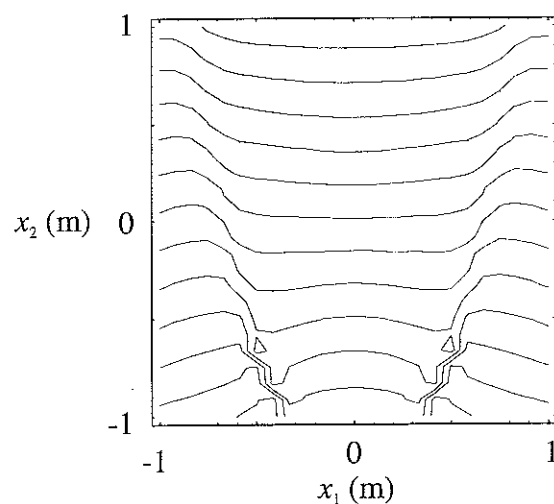


Figure 3.5 (second page of two). The sound field  $w(\mathbf{x})$  reproduced by optimally adjusting the four sources shown in Figure 3.4b. The frequency is 500Hz, d) is a 3D time "snapshot" of  $w(\mathbf{x})$ , e) is a 2D contourplot of d). f) shows the contours of constant phase, the distance between adjacent contours is  $180^\circ$

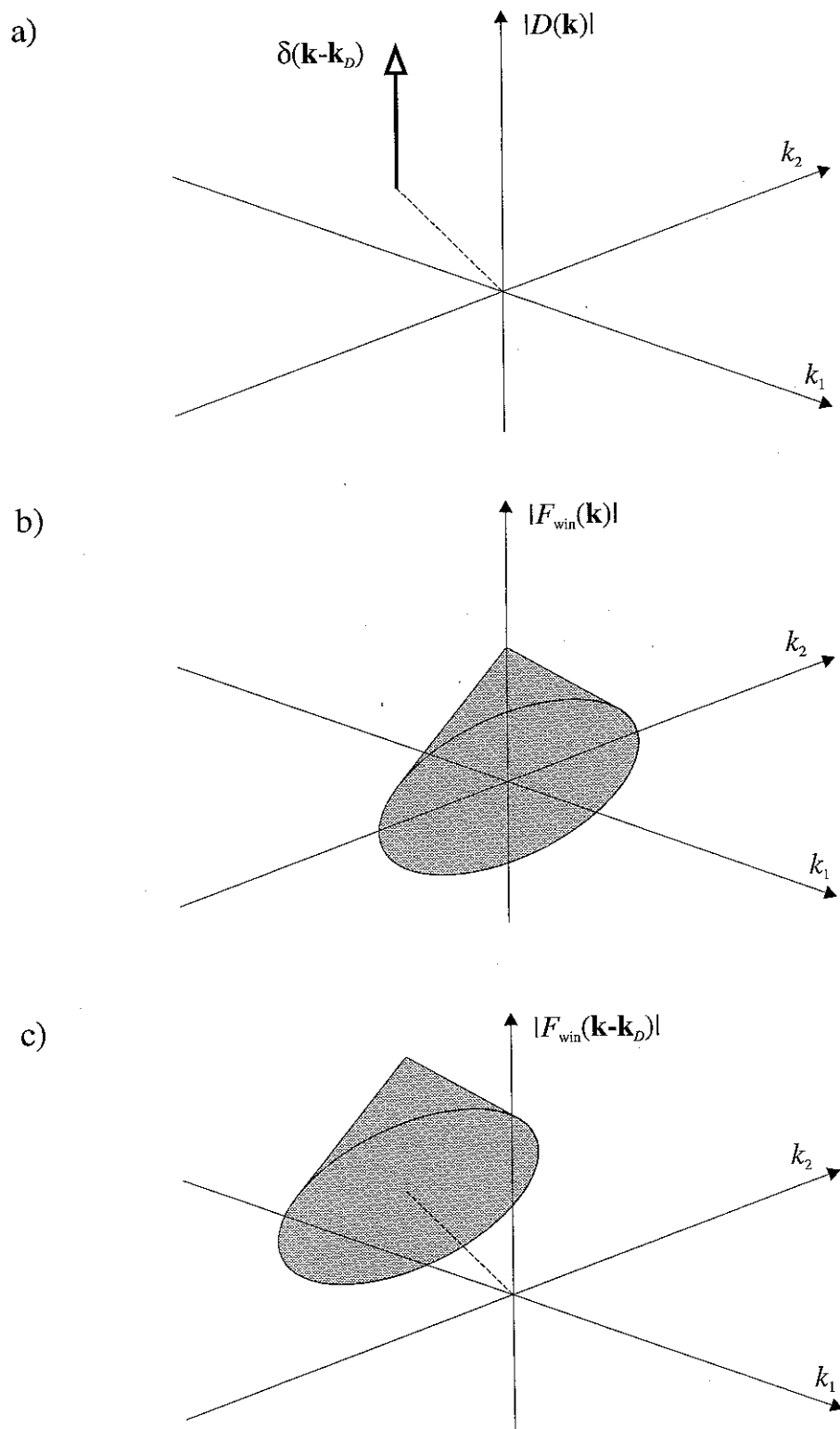
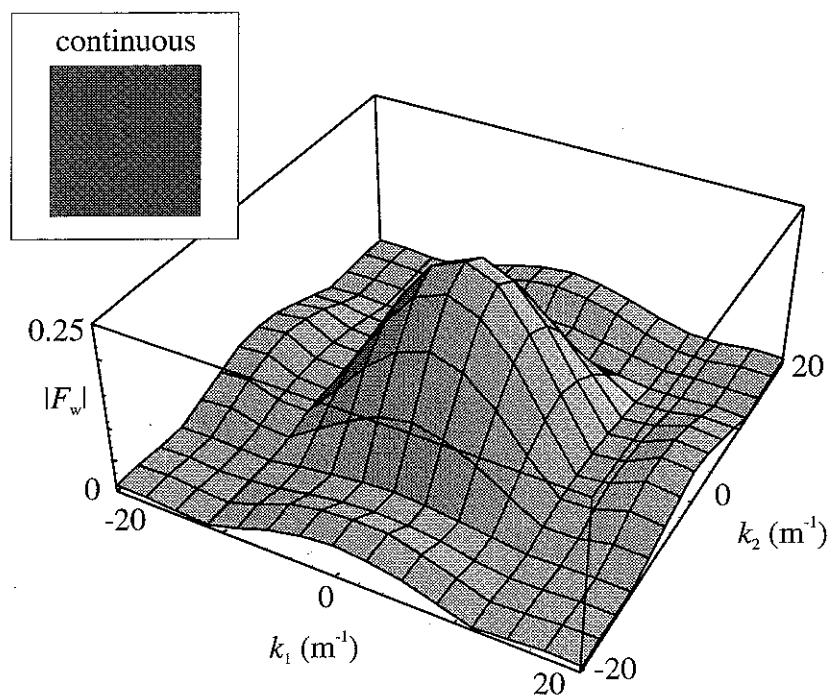


Figure 3.6. An example of the modification of the spatial spectrum  $D(\mathbf{k})$  of a plane wave by the window (or weighting) function  $F_{\text{win}}(\mathbf{k})$

a)



b)

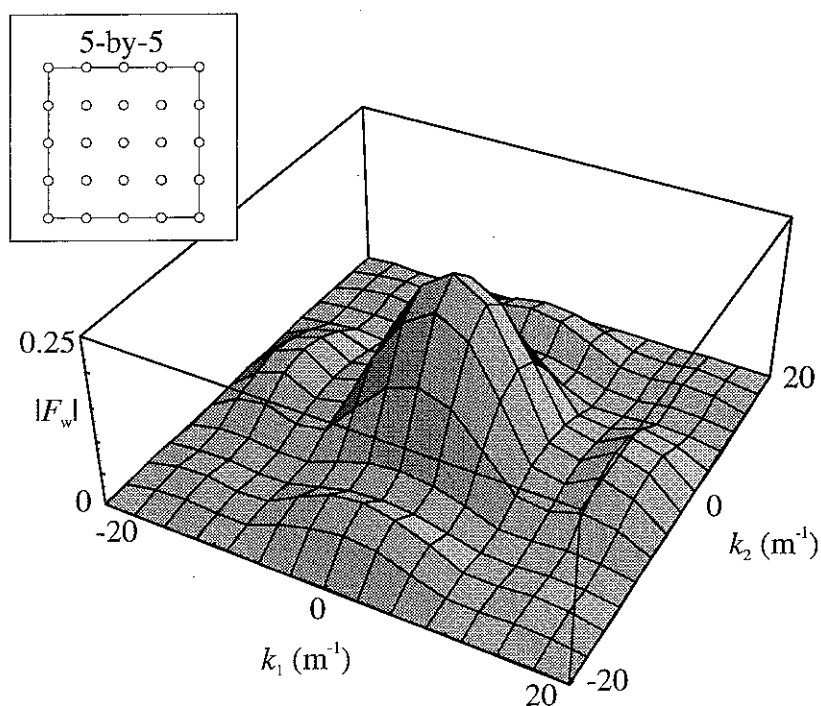


Figure 3.7 (first page of two). Spatial spectra for a) a continuous, or very dense, microphone array, and b) a 5-by-5 microphone array.

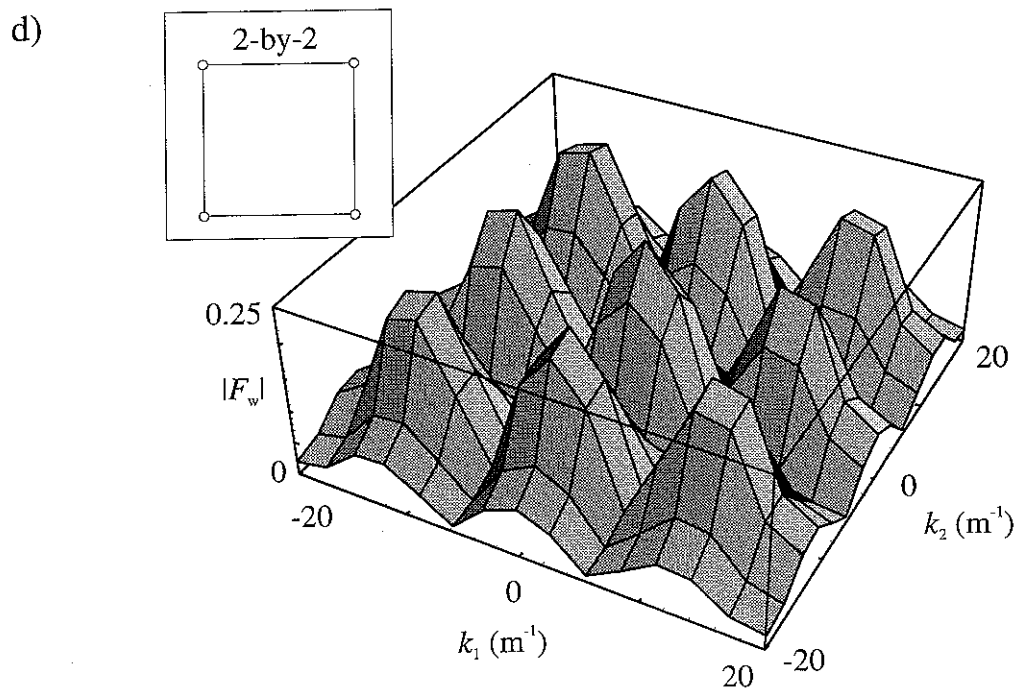
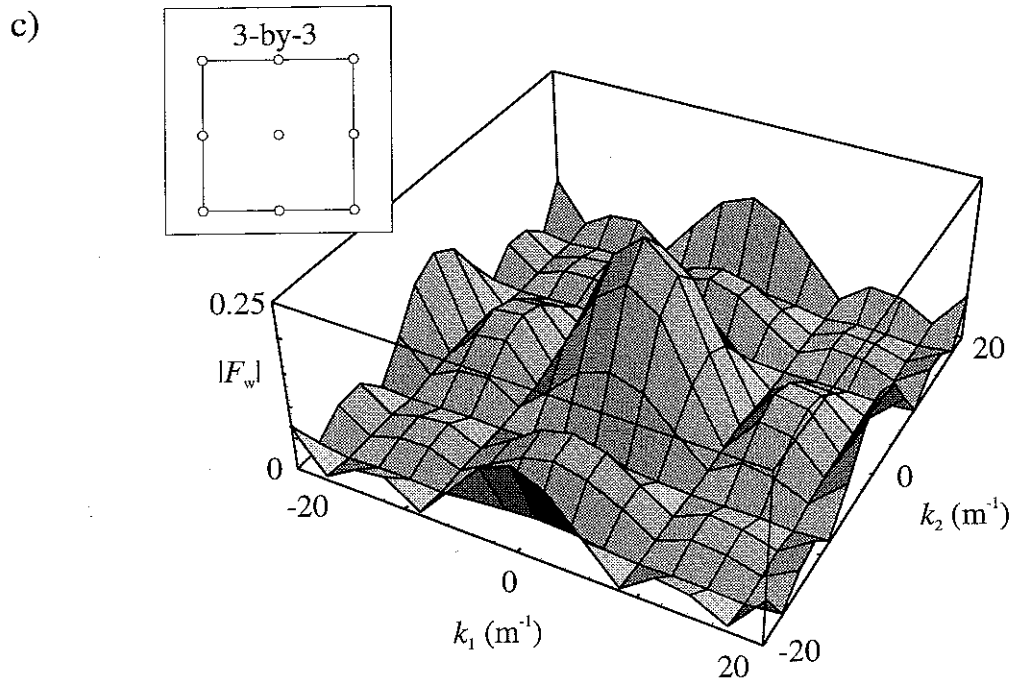
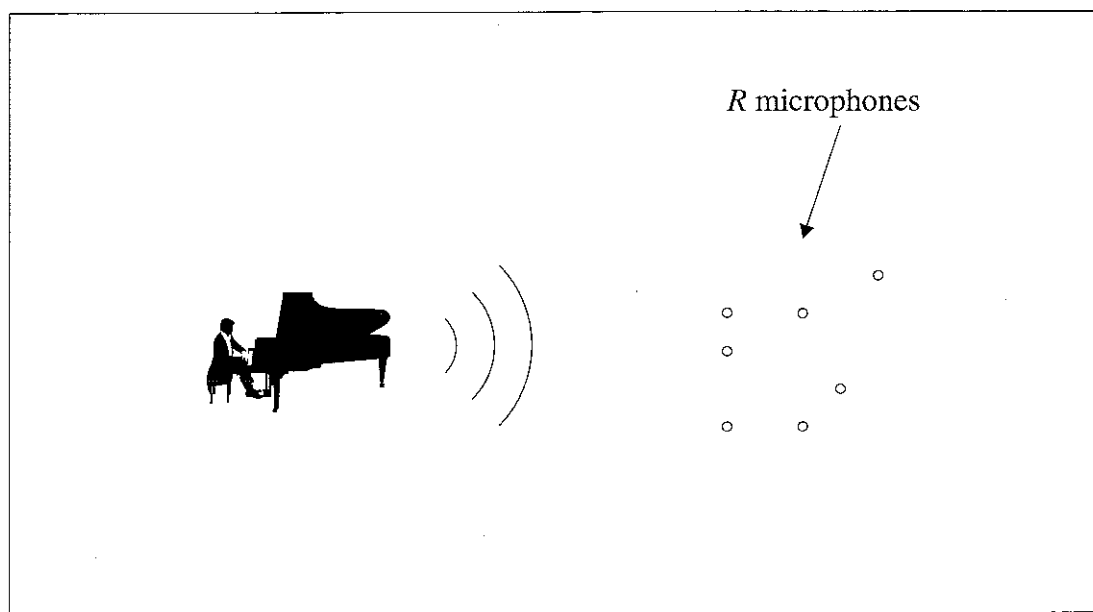


Figure 3.7 (second page of two). Spatial spectra of c) a 3-by-3 microphone array, and d) a 2-by-2 microphone array. The repetitive structure of the spectra are images resulting from the spatial sampling

"Recording" space



"Listening" space

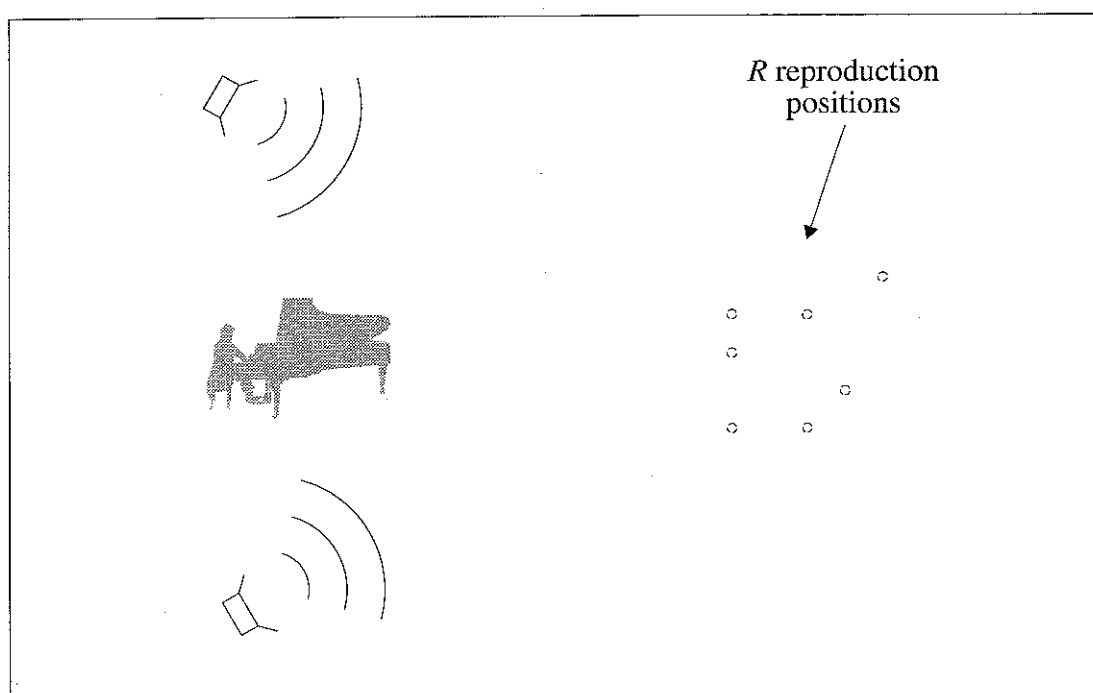


Figure 3.8. An illustration of the concepts of "recording space" and "listening space"

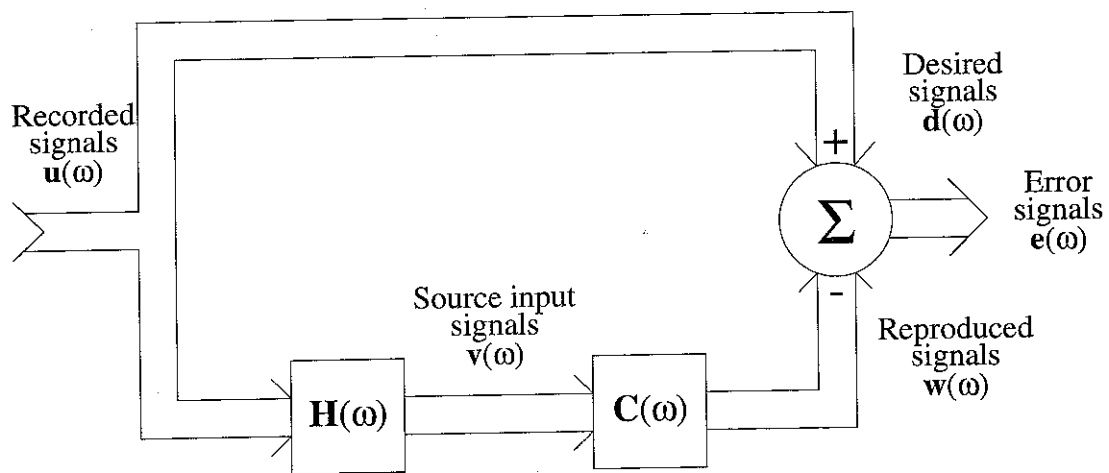


Figure 3.9. The sound reproduction problem in block diagram form when the desired signals  $\mathbf{d}(\omega)$  are exact copies of the recorded signals  $\mathbf{u}(\omega)$



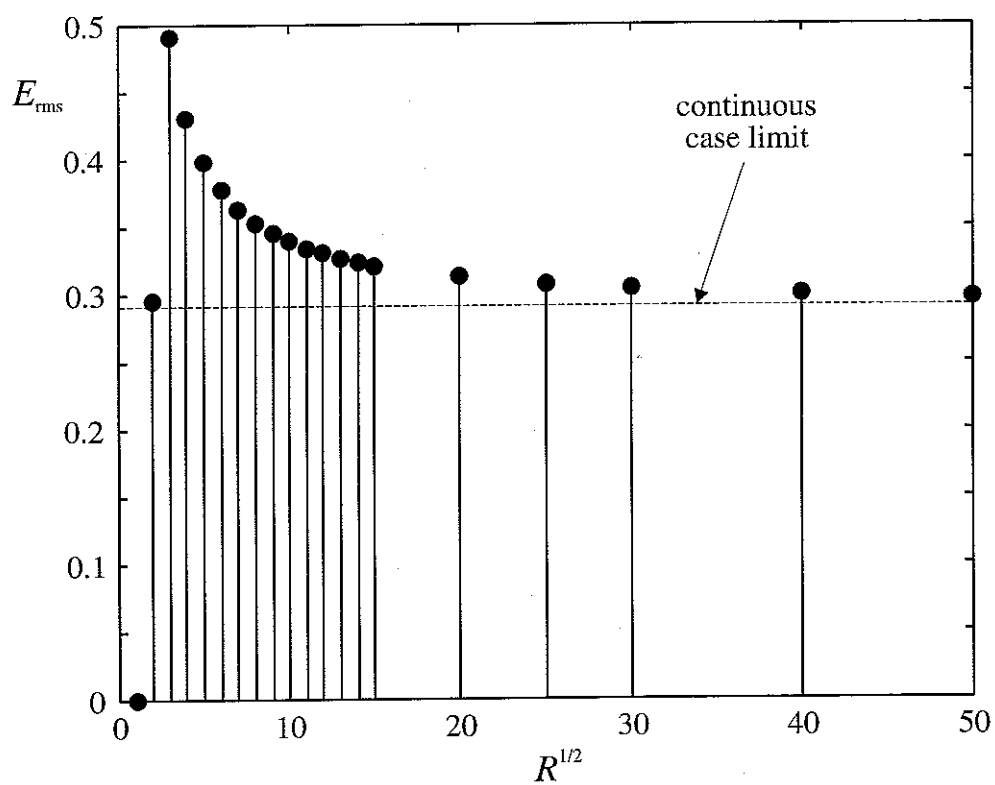
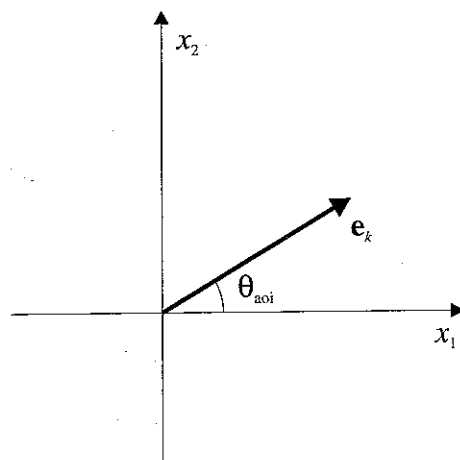


Figure 3.10. An example of the convergence of the root mean squared error  $E_{rms}$  to the solution to the continuous case problem as the number of microphones  $R$  increases

a)

Plane wave in  $(x_1, x_2)$ -plane



b)

Plane wave with non-zero  $x_3$ -component

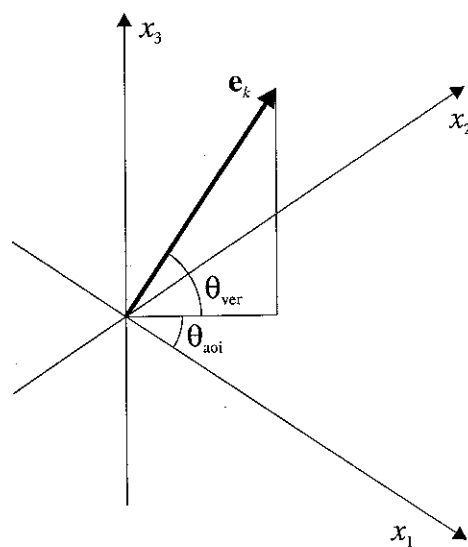


Figure 3.11. Definitions of the parameters a)  $\theta_{aoi}$  and b)  $\theta_{ver}$  that specify the direction of the desired plane wave

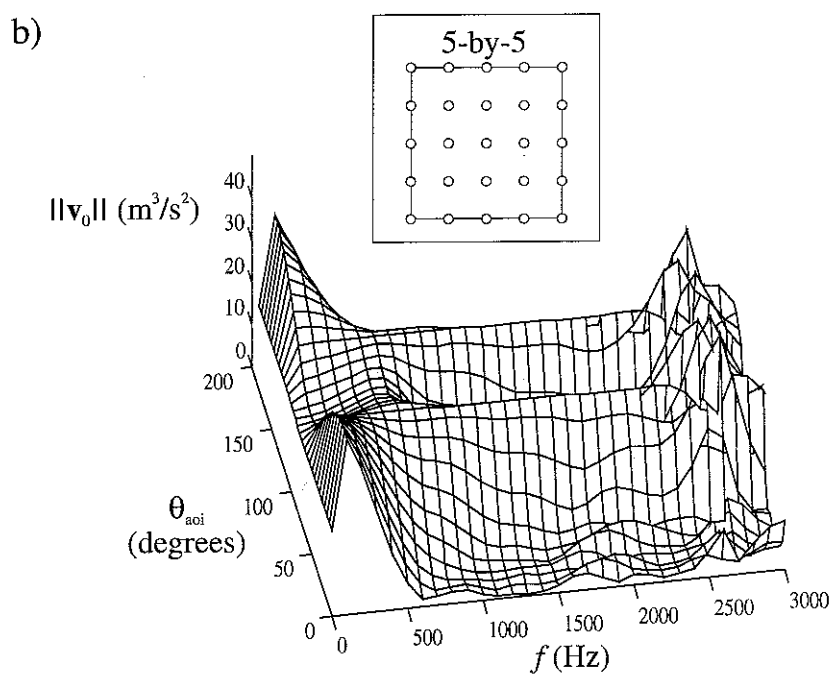
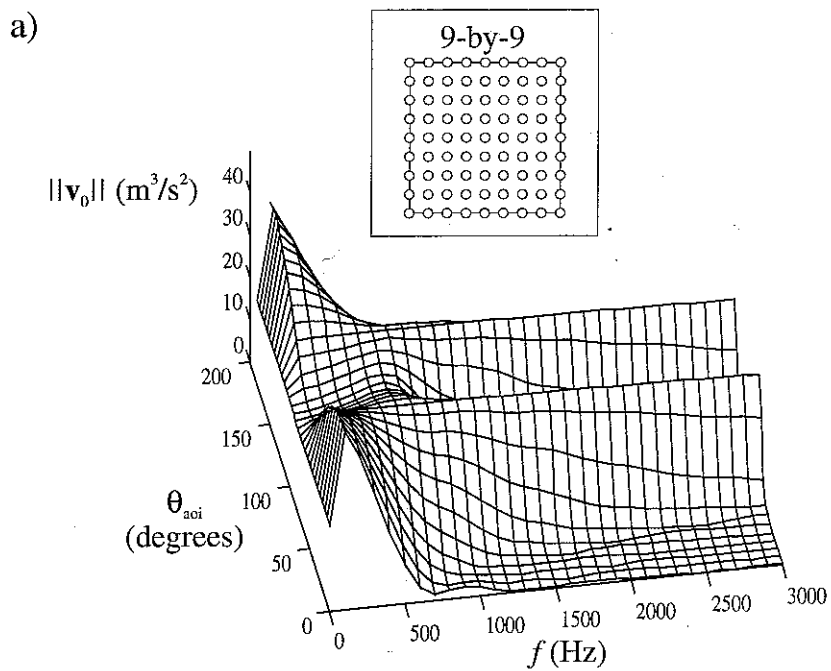


Figure 3.12 (first page of two). The influence of the microphone spacing on the total source input  $\|v_0\|$ , a) for a 9-by-9 microphone array, and b) a 5-by-5 microphone array

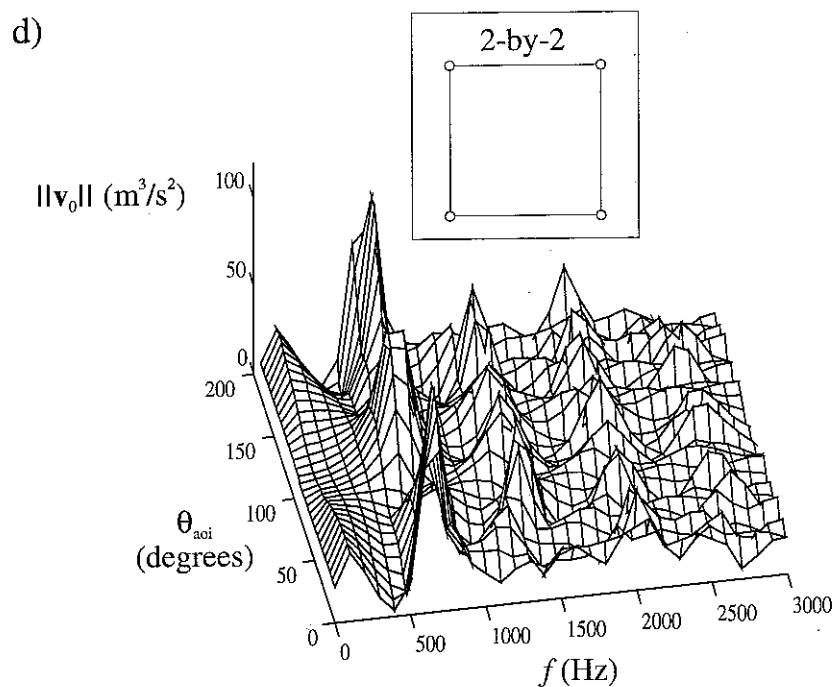
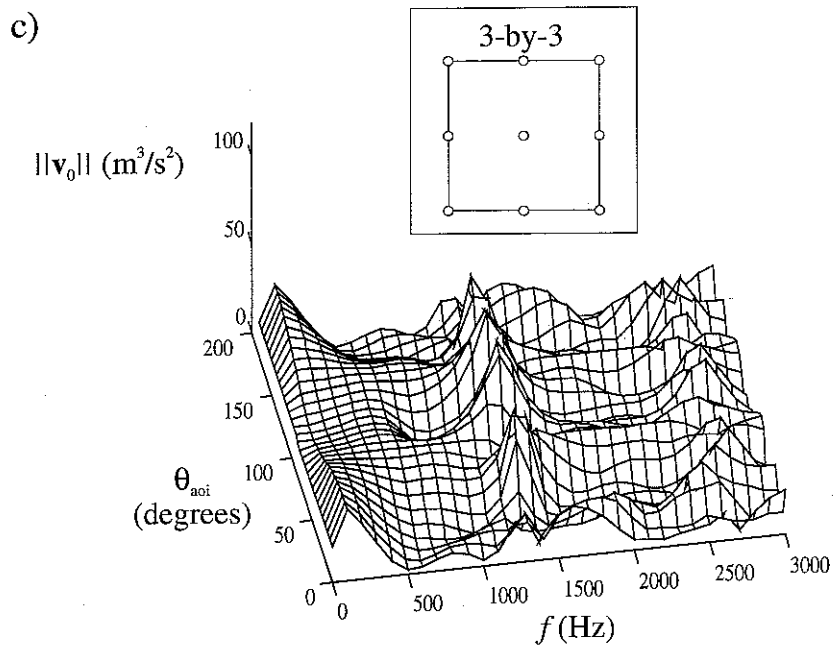
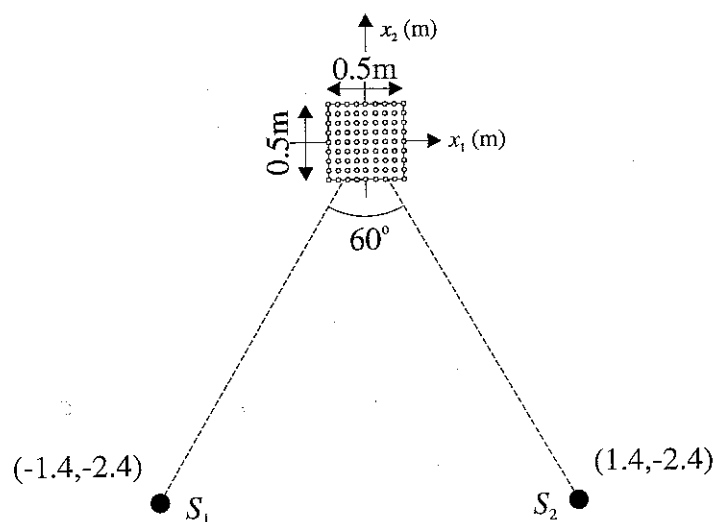


Figure 3.12 (second page of two). The influence of the microphone spacing on the total source input  $\|v_0\|$ , c) for a 3-by-3 microphone array, and b) a 2-by-2 microphone array. The frequency at which the peaks start to appear decrease with increasing microphone spacing

a) "Stereo"



b) "Quadraphony"



c) "Narrow-arc"

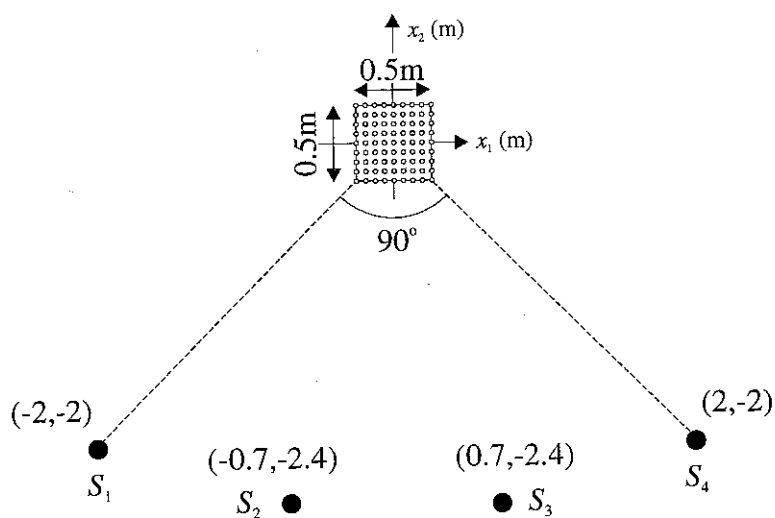
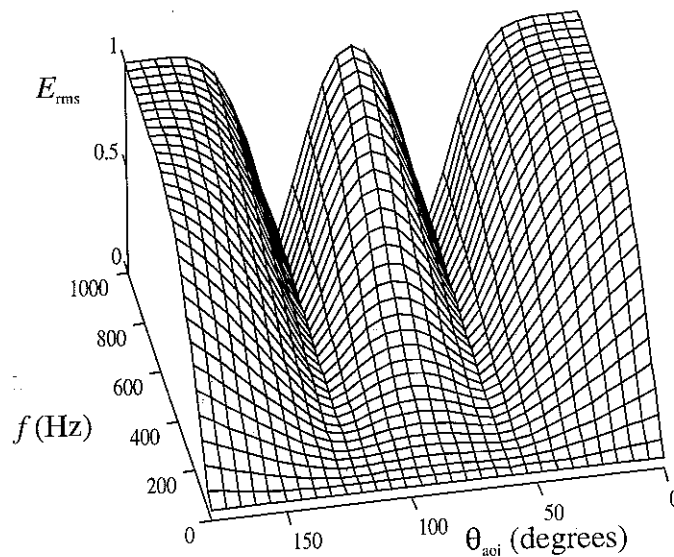
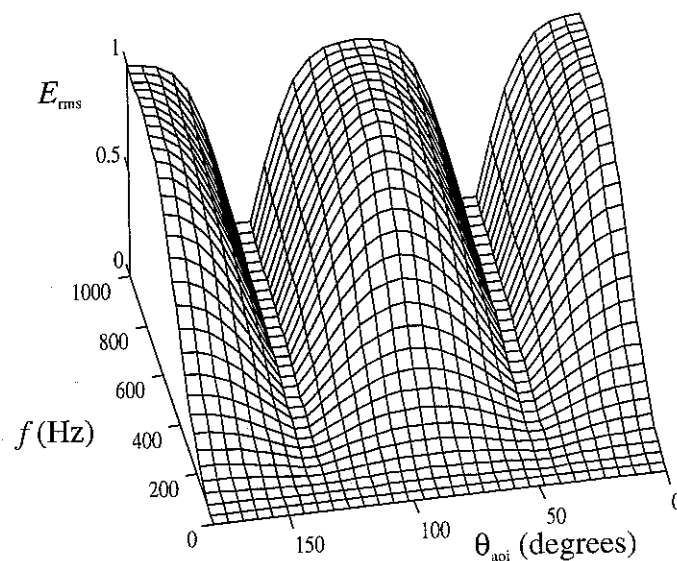


Figure 3.13. Three layouts of sources (black spheres) and microphones (white spheres) that roughly mimic the listening situation for a single listener

a) "Stereo"



b) "Quadraphony"



c) "Narrow-arc"

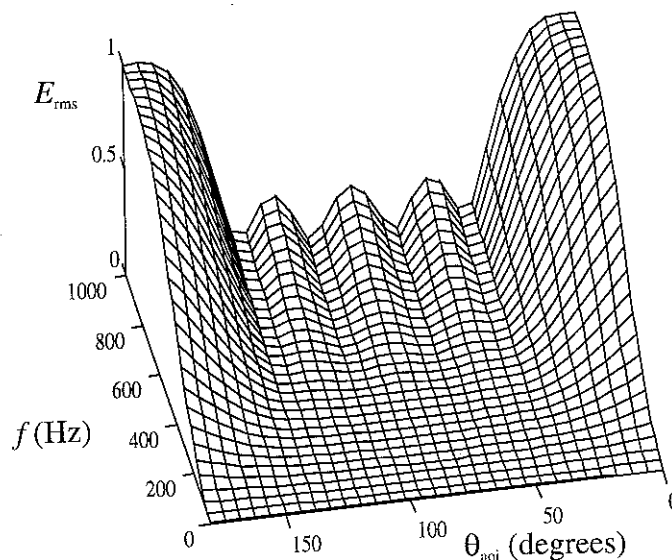
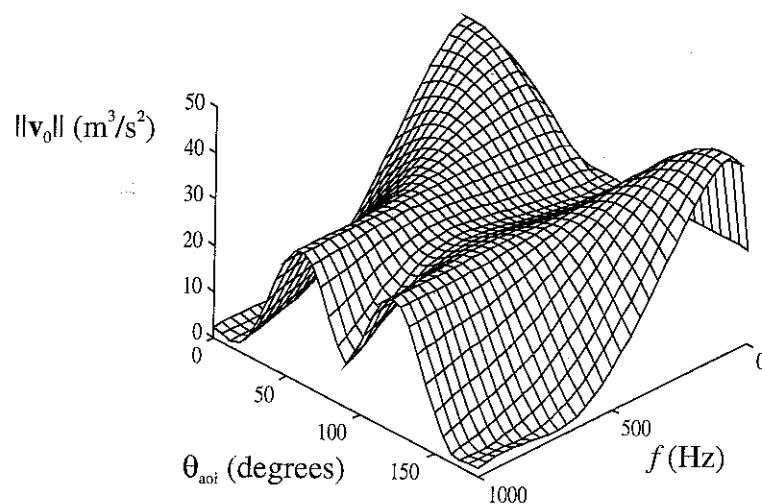
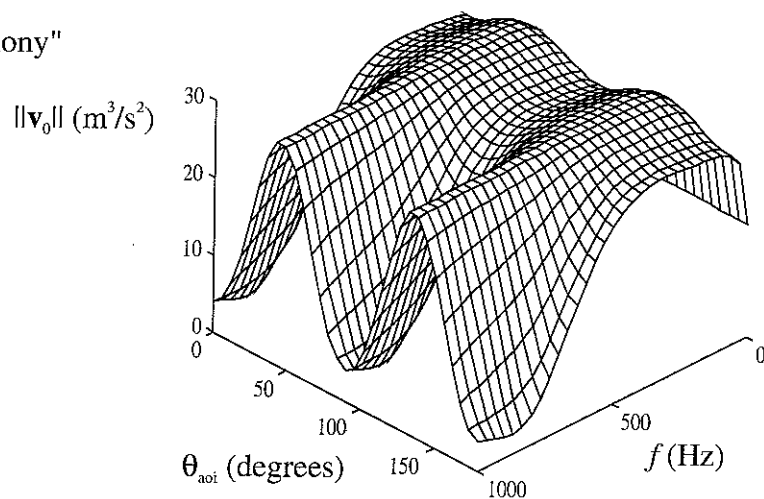


Figure 3.14. The root mean squared error  $E_{rms}$  as a function of the frequency  $f$  and the angle of incidence  $\theta_{aoi}$  of the plane wave for the three source-microphone layouts shown in Figure 3.13

a) "Stereo"



b) "Quadraphony"



c) "Narrow-arc"

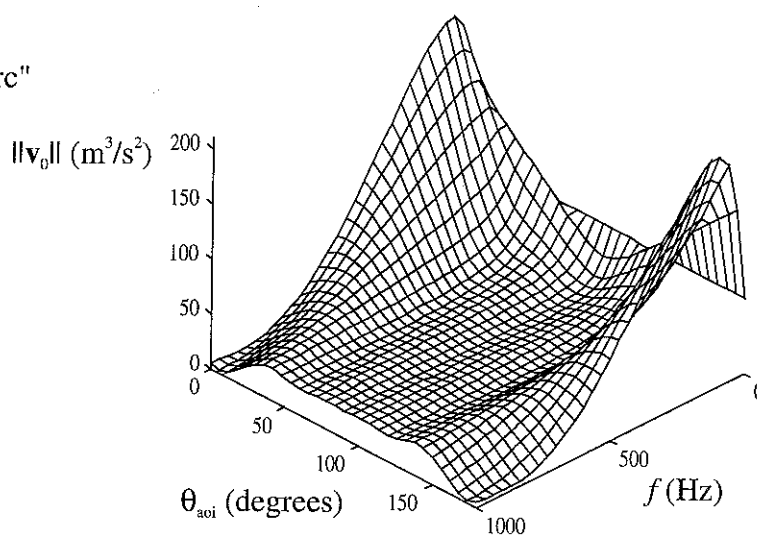


Figure 3.15. The total source input  $\|v_0\|$  as a function of the frequency  $f$  and the angle of incidence  $\theta_{aoi}$  of the plane wave for the three source-microphone layouts shown in Figure 3.13

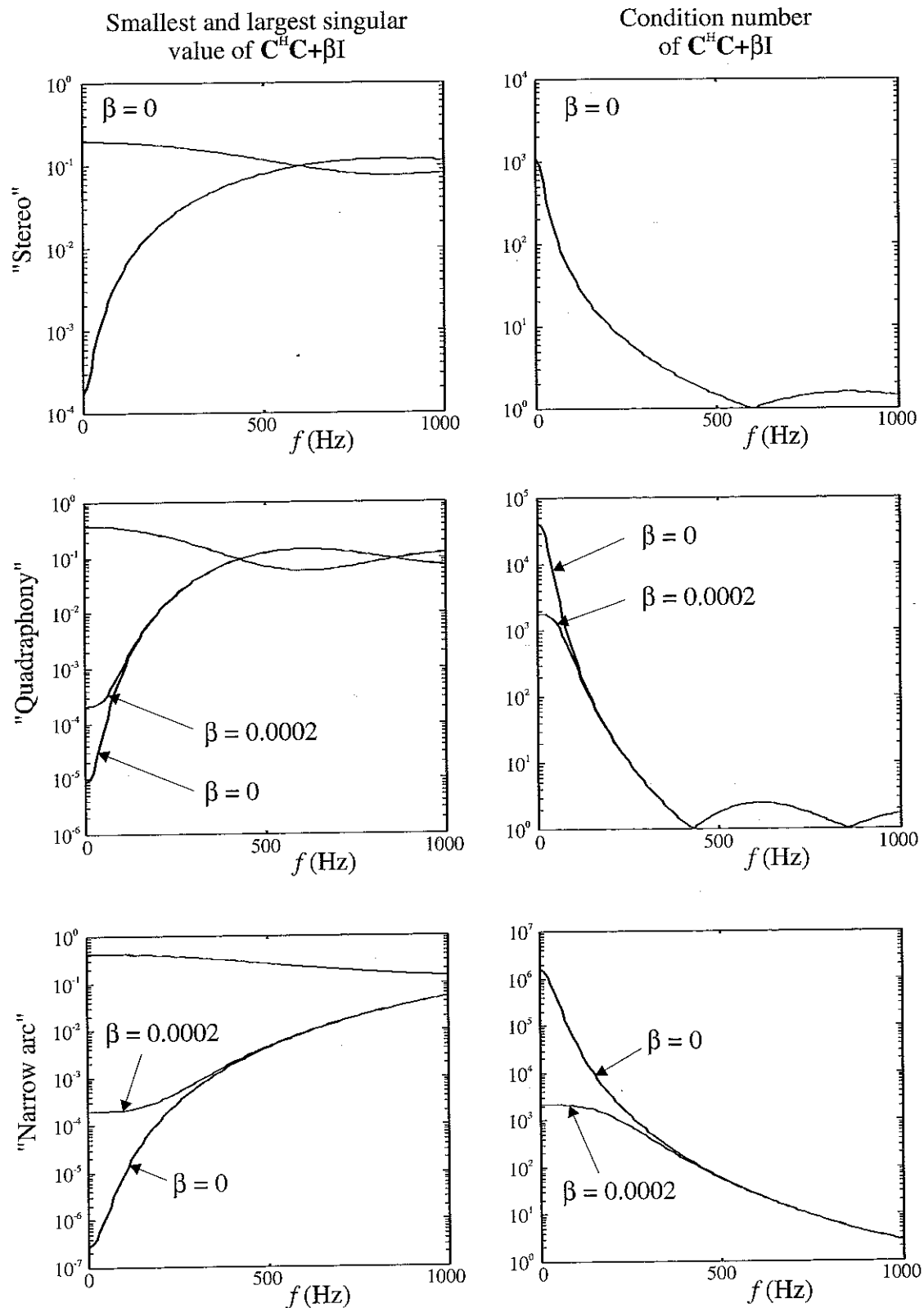
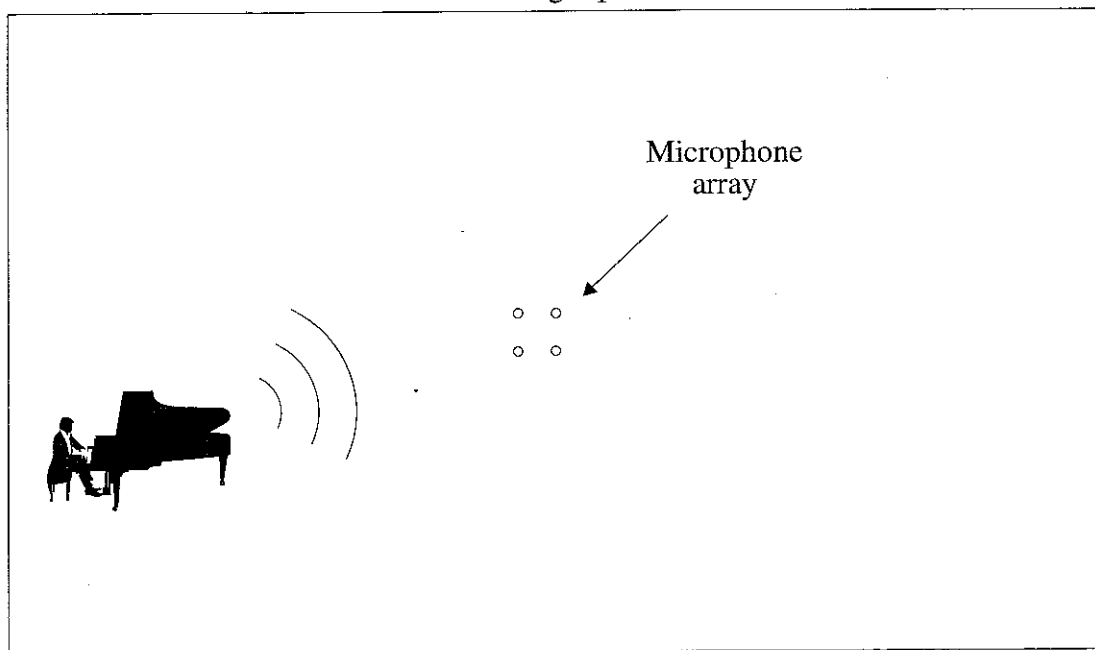


Figure 3.16. Singular values and condition numbers of the equation systems that are used to calculate the results that are shown in Figures 3.14 and 3.15



"Recording" space



"Listening" space

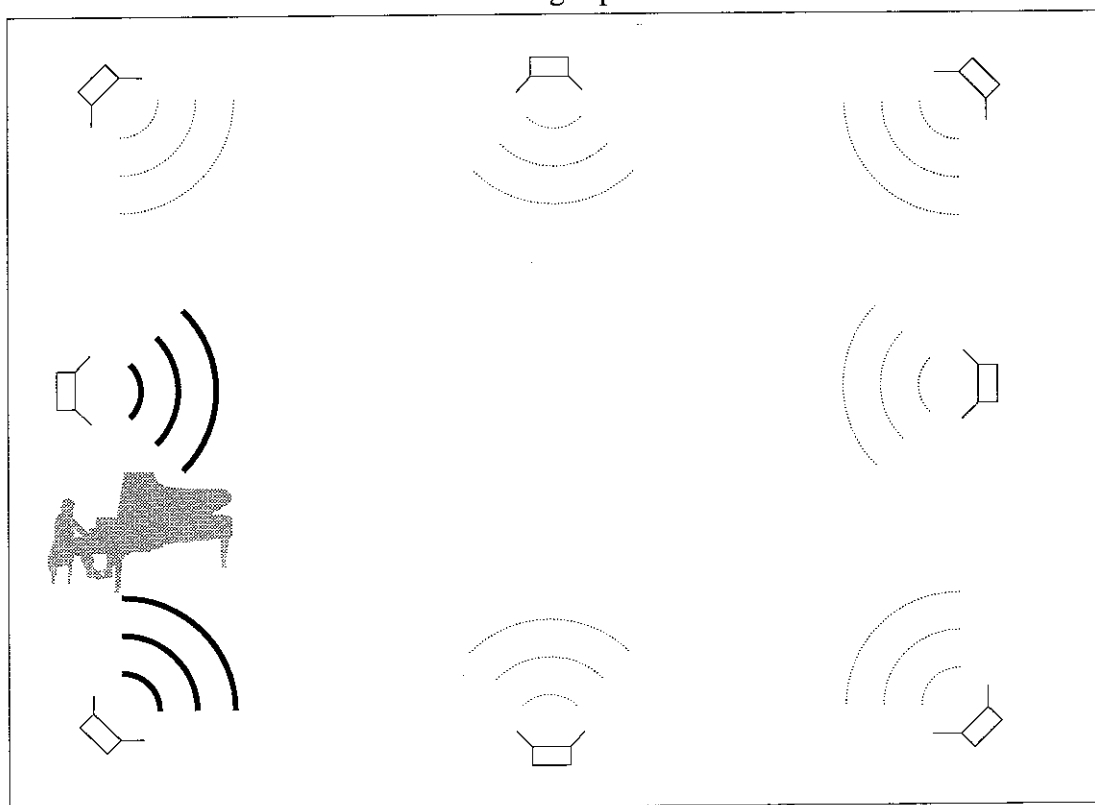


Figure 3.17. The principle of direction-of-arrival reproduction. Most of the sound must be emitted from the loudspeakers that are closest to the direction of the original sound source

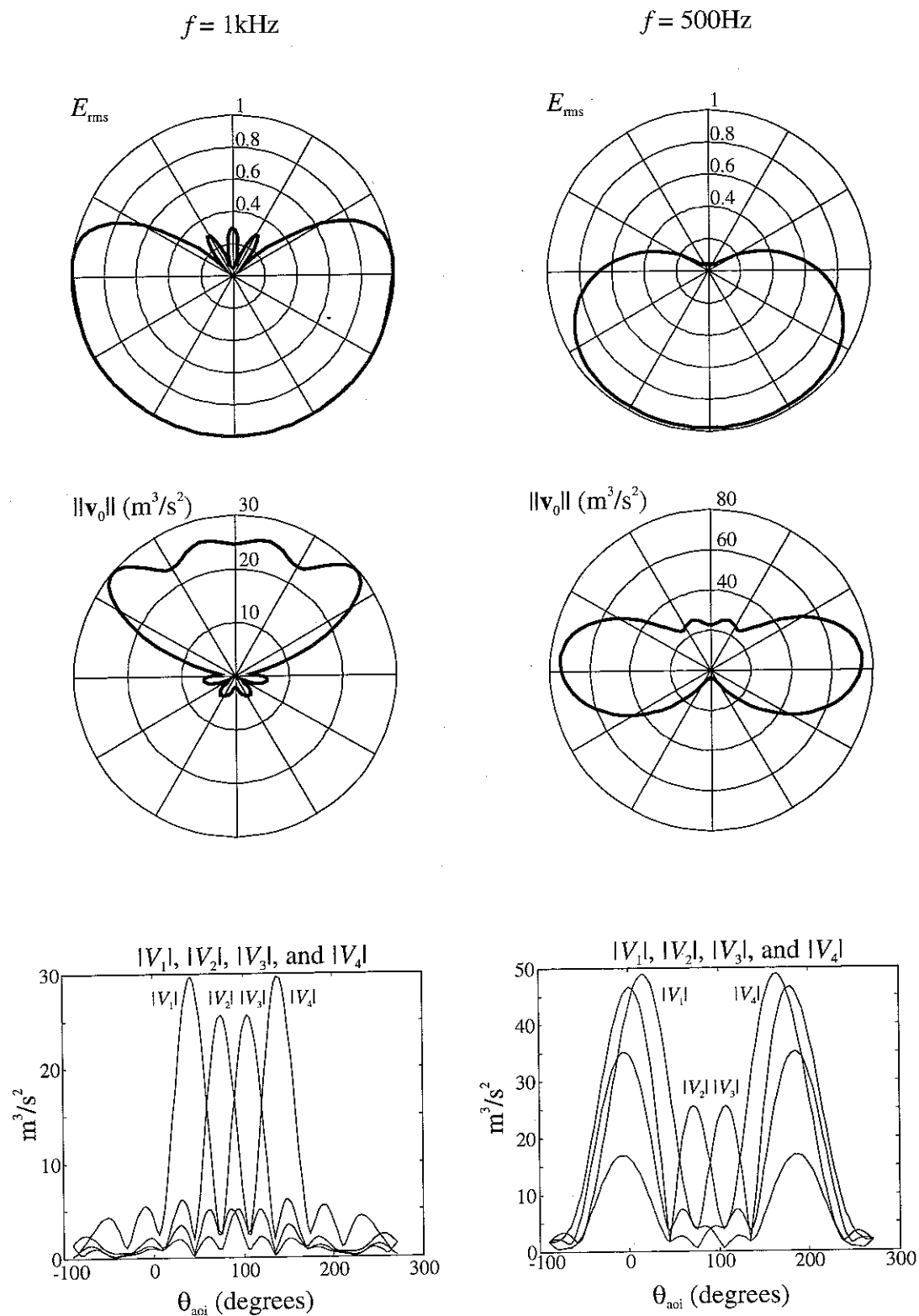
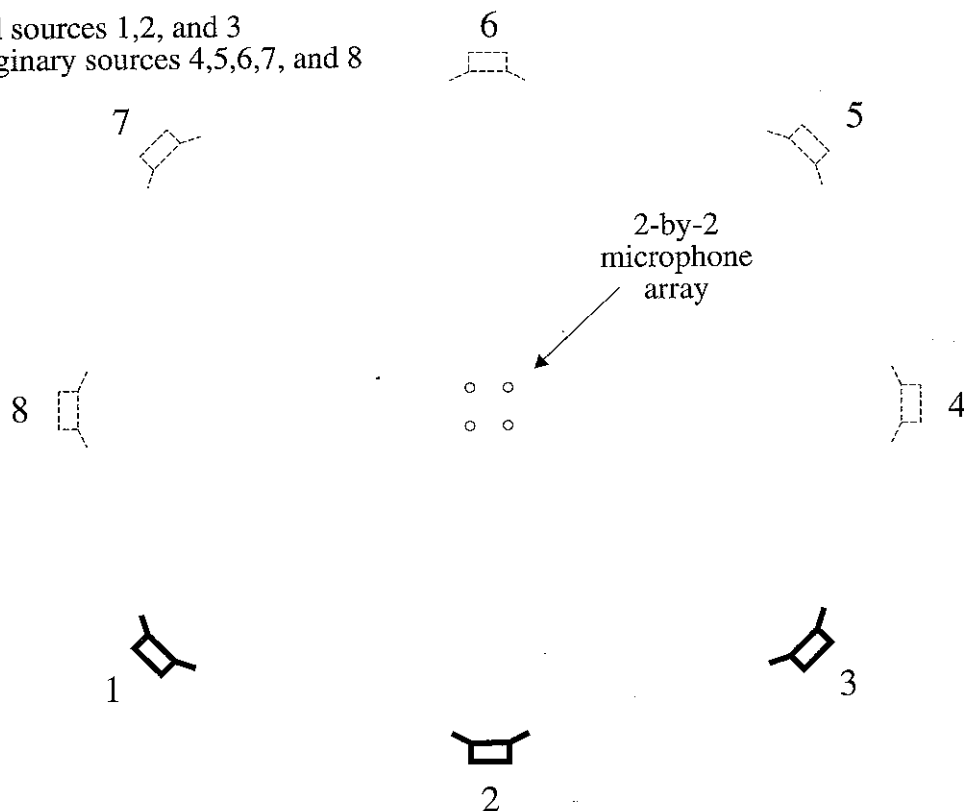


Figure 3.18. The root mean squared error  $E_{rms}$ , total source input  $\|v_0\|$ , and individual source inputs as a function of the angle of incidence  $\theta_{aoi}$  at the frequencies 500Hz and 1kHz. The geometry of the problem is the narrow-arc layout shown in Figure 3.13c

Real sources 1,2, and 3  
Imaginary sources 4,5,6,7, and 8



Use full **C**-matrix  
to calculate **H**

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} & C_{18} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} & C_{27} & C_{28} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} & C_{37} & C_{38} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} & C_{47} & C_{48} \end{bmatrix}$$

Use only top three  
rows of **H** to  
calculate the three  
source inputs

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{21} & H_{22} & H_{23} & H_{24} \\ H_{31} & H_{32} & H_{33} & H_{34} \\ \hline H_{41} & H_{42} & H_{43} & H_{44} \\ H_{51} & H_{52} & H_{53} & H_{54} \\ H_{61} & H_{62} & H_{63} & H_{64} \\ H_{71} & H_{72} & H_{73} & H_{74} \\ H_{81} & H_{82} & H_{83} & H_{84} \end{bmatrix}$$

Figure 3.19. An example of how to use "imaginary" loudspeakers to avoid large outputs from the real loudspeakers for angles of incidence outside the range that are covered by the real loudspeakers

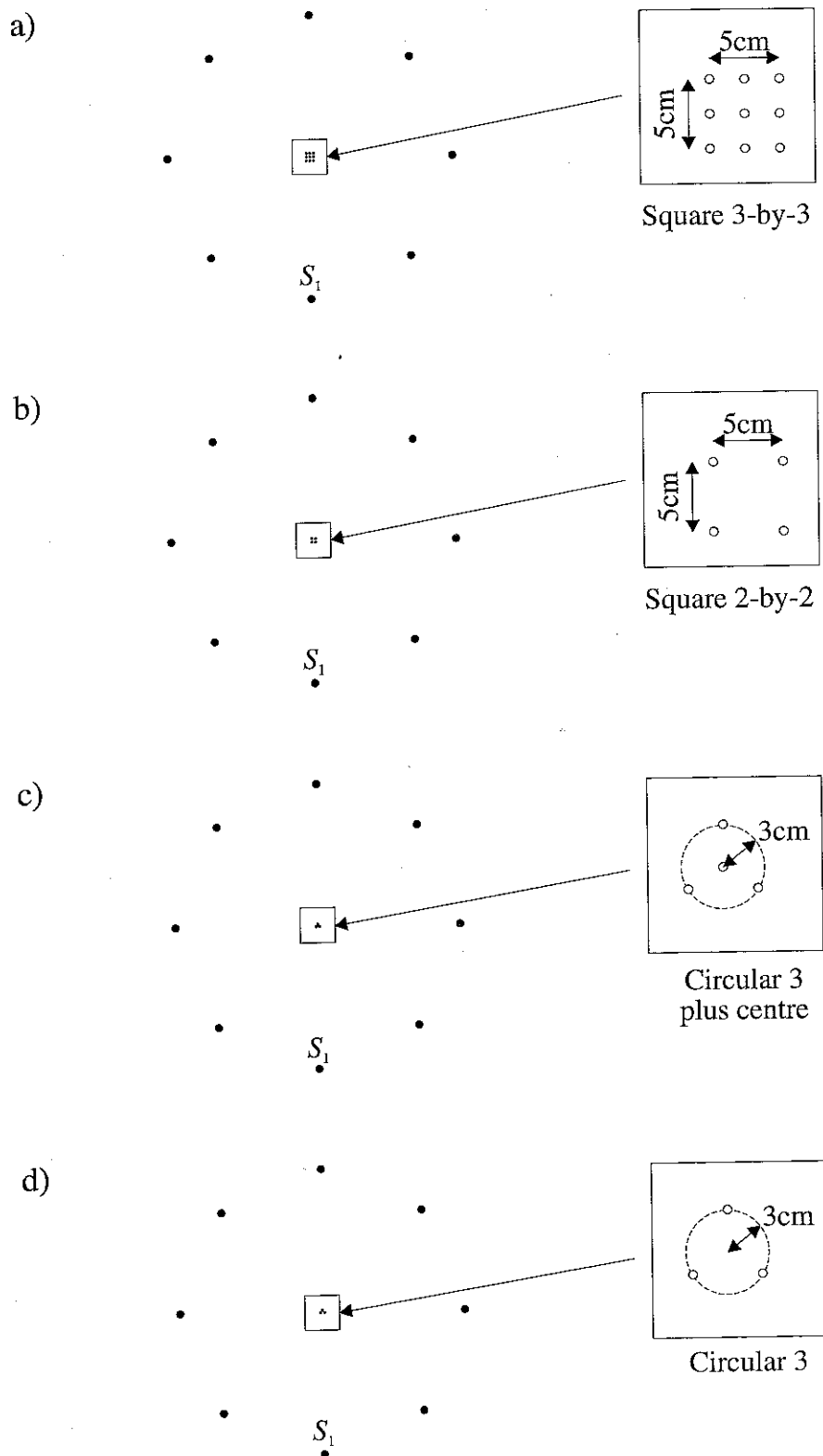


Figure 3.20. Four different microphone arrays that all cover an area of approximately  $25\text{cm}^2$ . The eight sources are positioned in a "perfect" circle whose radius is approximately  $2.83\text{m}$ . The input  $V_1$  to the source  $S_1$  is studied later

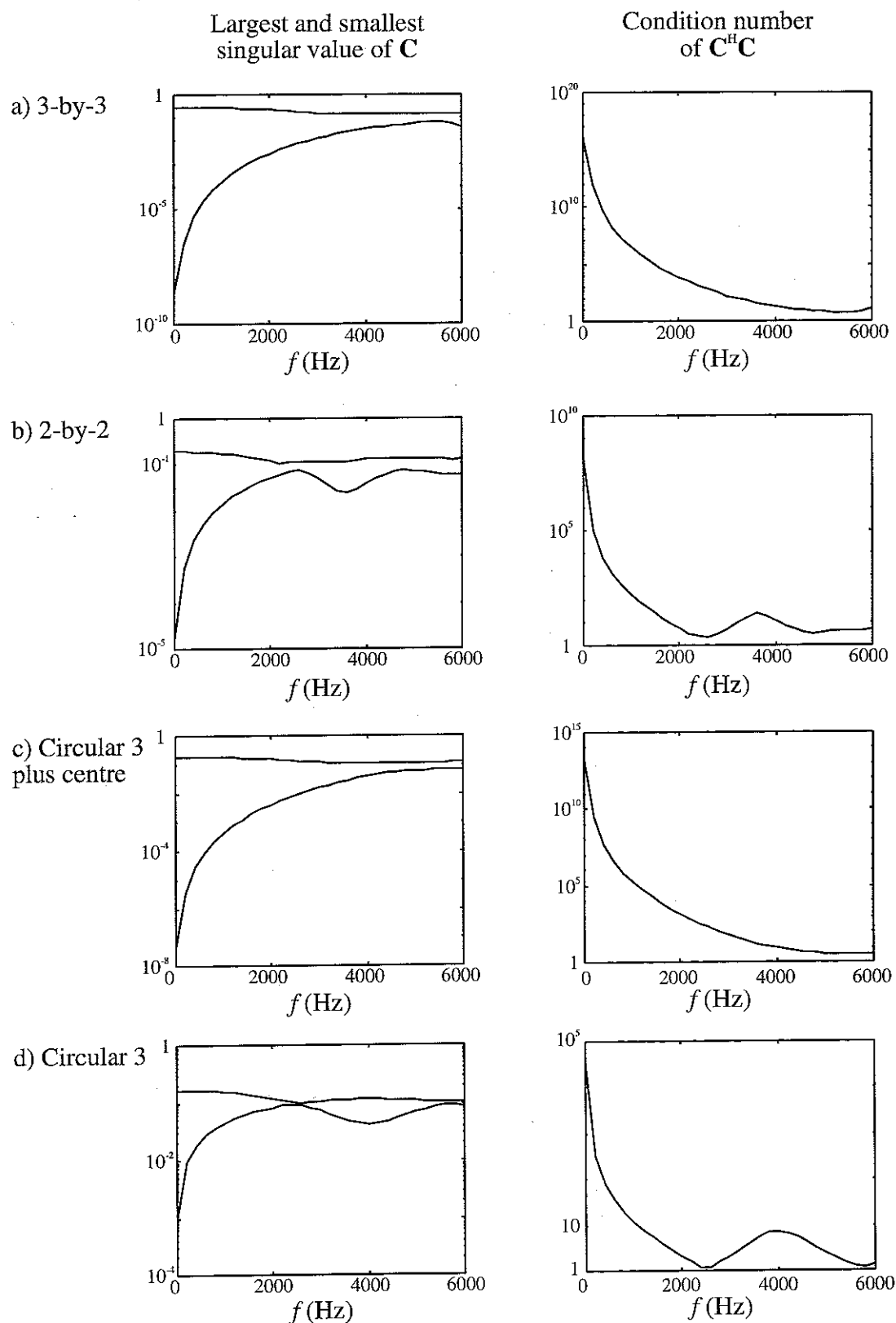
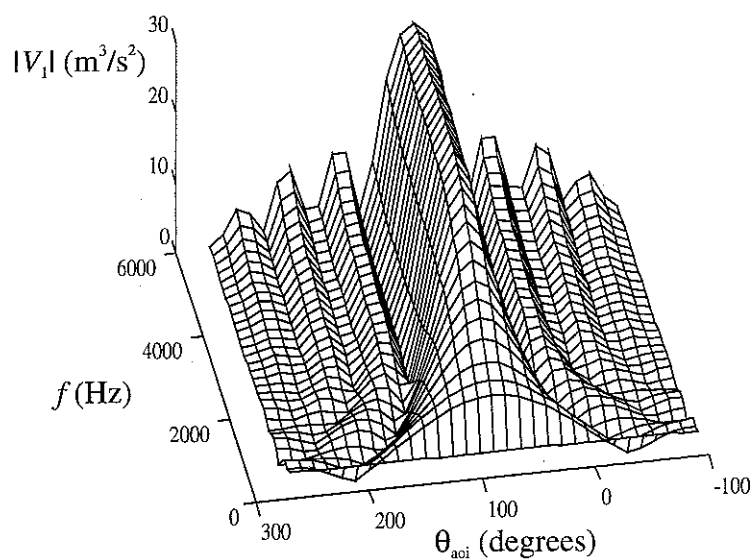


Figure 3.21. As a function of frequency, the largest and the smallest singular value of  $\mathbf{C}$ , and the condition number of  $\mathbf{C}^H \mathbf{C}$ , for the four microphone arrays shown in Fig.3.20

a) 3-by-3



b) 2-by-2

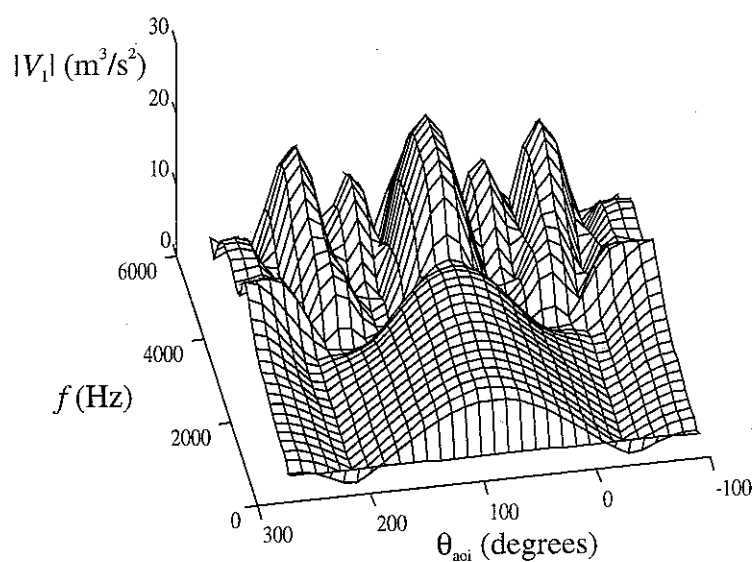
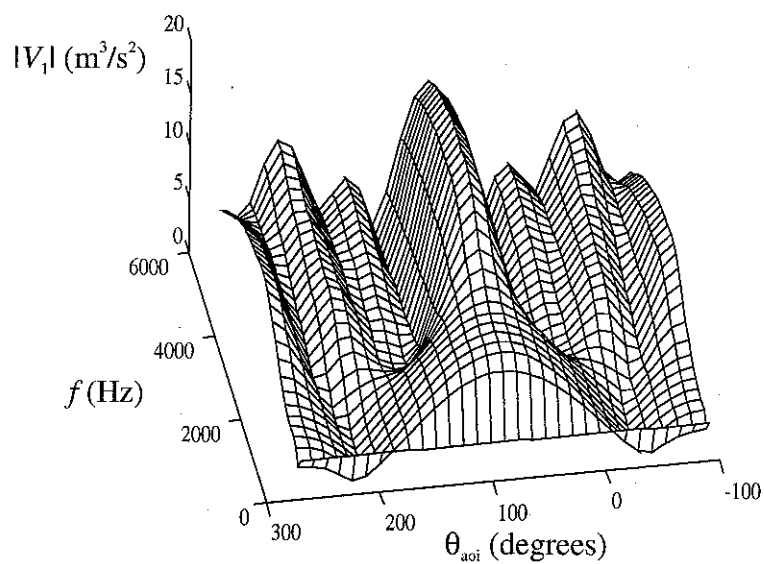


Figure 3.22 (first page of two). The input  $|V_i|$  to source number one for each of the four microphone arrays shown in Figure 3.20. The regularisation factor  $\beta$  is 0.00002

c) Circular 3 plus centre



d) Circular 3

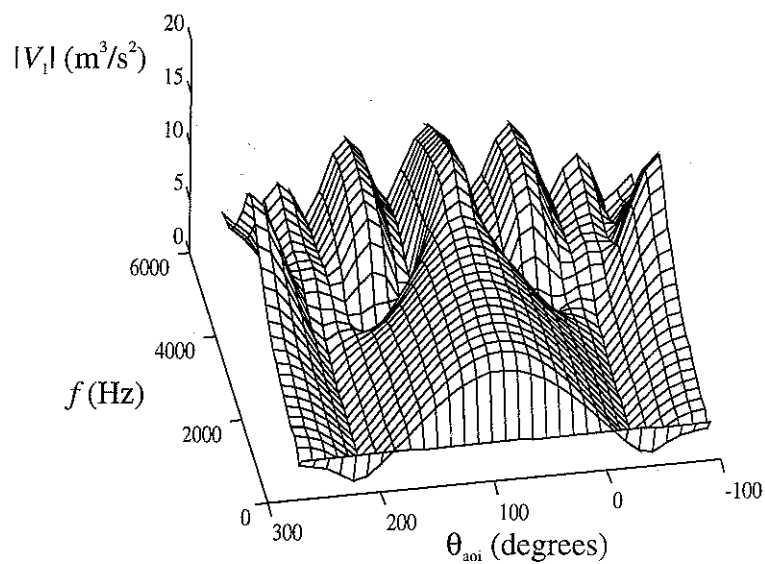
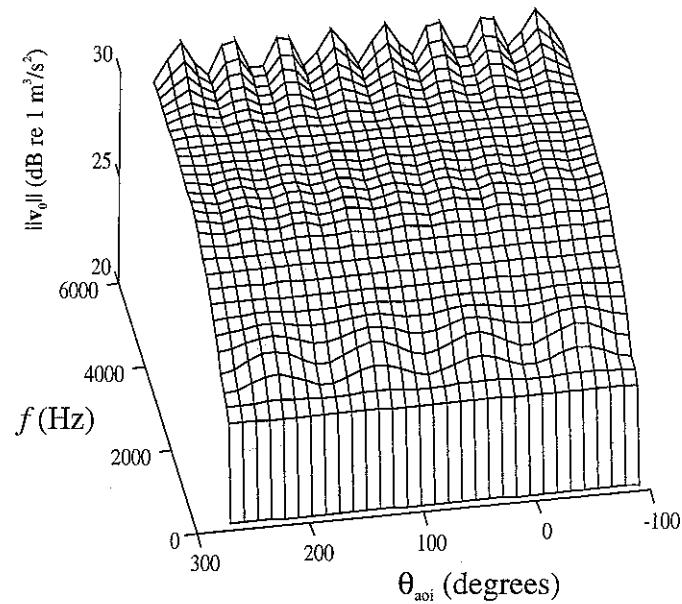


Figure 3.22 (second page of two). The input  $|V_1|$  to source number one for each of the four microphone arrays shown in Figure 3.20. The regularisation factor  $\beta$  is 0.00002

a) 3-by-3



b) Circular 3

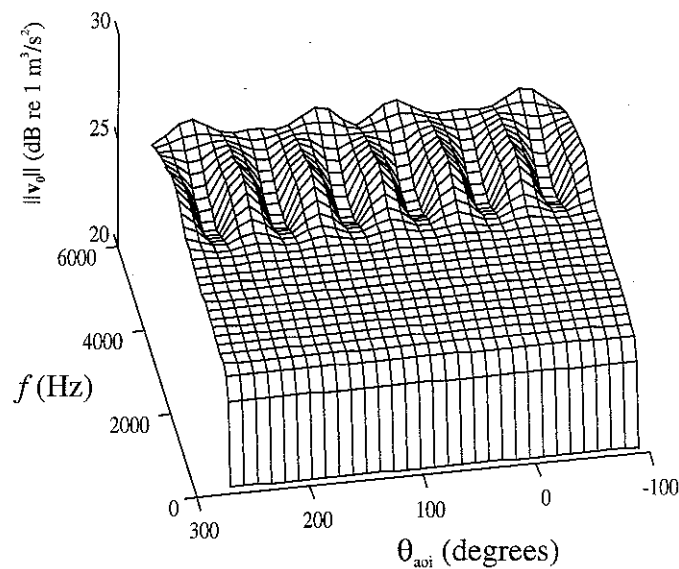
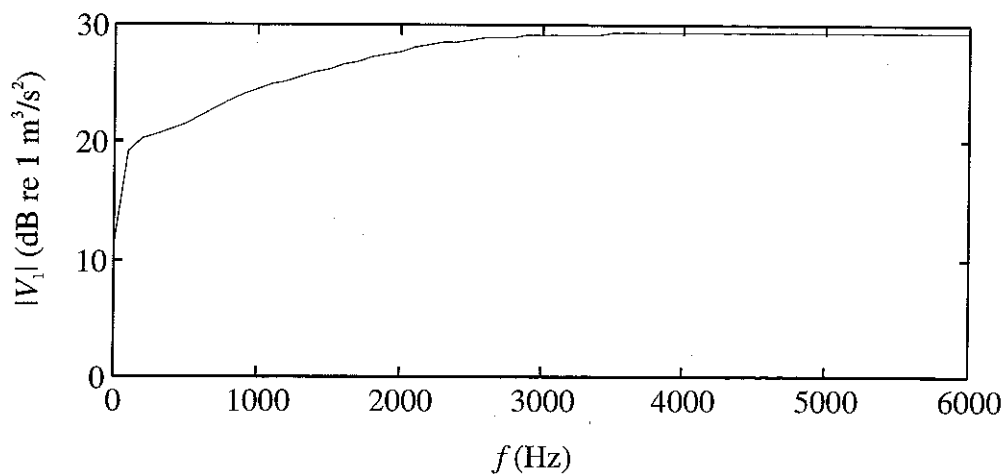


Figure 3.23. Total source effort  $\|v_0\|$  for a) the 3-by-3 microphone array shown in Figure 3.20a, and b) the circular 3 microphone array shown in Figure 3.20d. Note that  $\|v_0\|$  is roughly independent of  $\theta_{aoi}$ .



a) 3-by-3



b) Circular 3

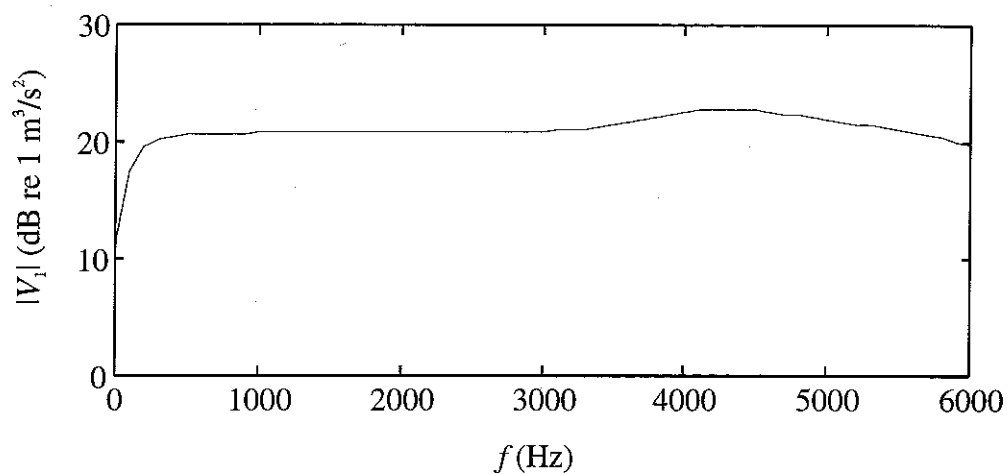


Figure 3.24. For a fixed value of  $90^\circ$  for the angle of incidence  $\theta_{aoi}$ , the input  $|V_1|$  to source number one for the 3-by-3 microphone array shown in Figure 3.20a, and the circular 3 microphone array shown in Figure 3.20d. Note the small input at low frequencies

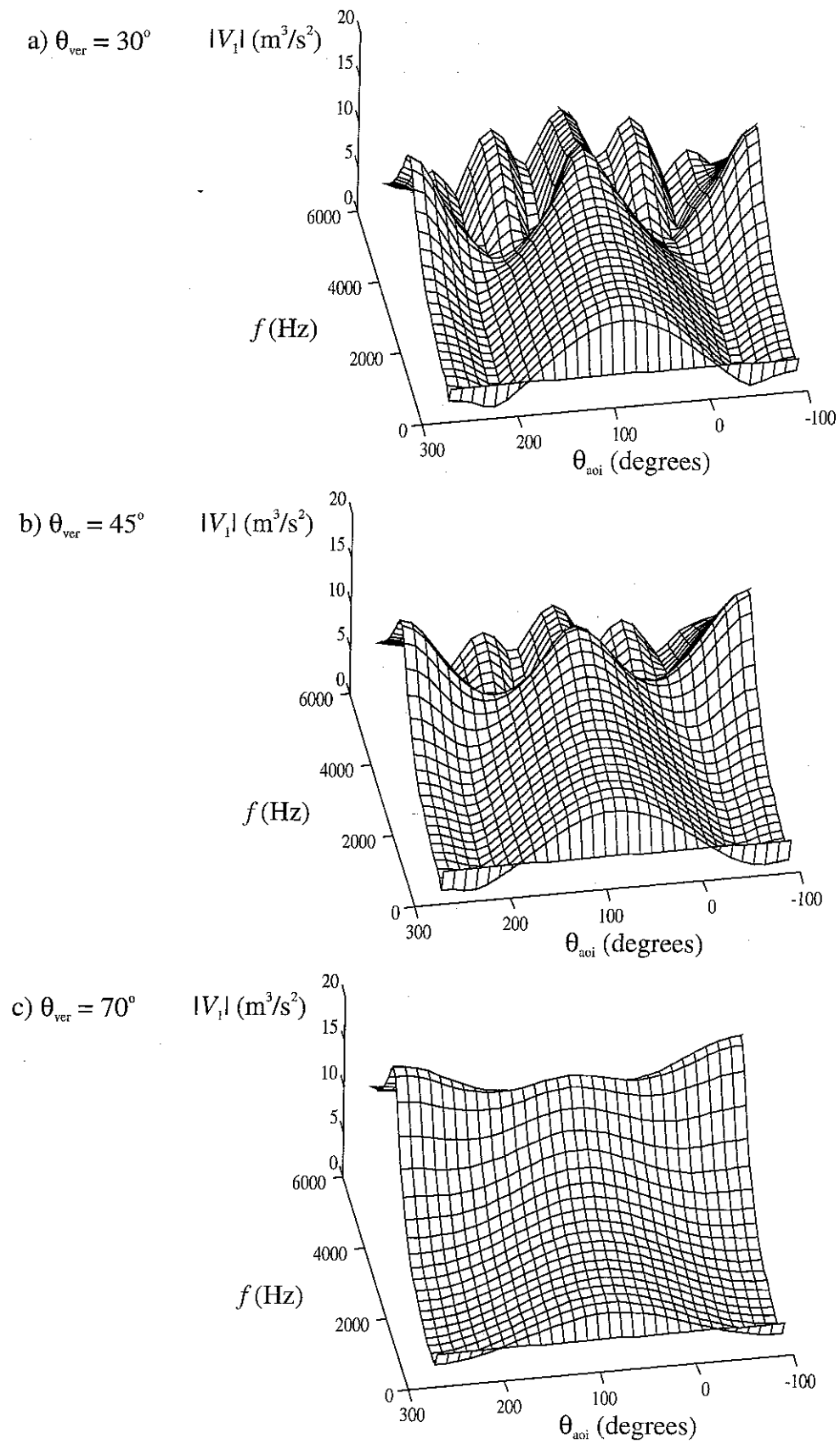


Figure 3.25. For the three values  $30^\circ$ ,  $45^\circ$ , and  $70^\circ$  of  $\theta_{\text{ver}}$ , the input  $|V_1|$  to source number one for the circular 3 microphone array shown in Fig.3.20d

## 4. $z$ -domain analysis

In the same way that frequency domain analysis can provide an understanding of the basic physics of sound reproduction,  $z$ -domain analysis can provide an understanding of the basic properties of the digital filters necessary to implement the inverse filter matrix in discrete time. As in the chapter on frequency domain analysis, it is straightforward to calculate a matrix  $\mathbf{H}(z)$  of exact least squares inverse filters from a matrix  $\mathbf{C}(z)$  of electro-acoustic transfer functions. The inputs  $\mathbf{v}(z)$  to a number of loudspeakers are calculated by pre-multiplying a set of recorded signals  $\mathbf{u}(z)$ , which are still taken to be exact copies of  $\mathbf{d}(z)$ , by  $\mathbf{H}(z)$ . The performance error  $\mathbf{e}(z)$  is the difference between the desired signals  $\mathbf{d}(z)$  and the reproduced signals  $\mathbf{w}(z)$ .

The qualitative behaviour of a filter's time response, or impulse response, is given by the positions in the complex plane of the poles of its  $z$ -transform. A pole away from the origin makes the filter's impulse response infinitely long, and when the pole is very close to the unit circle, the impulse response decays only very slowly with time. A pole outside the unit circle makes the filter's impulse response either non-causal or unstable, and therefore unrealisable. The inverse filters in  $\mathbf{H}(z)$  almost always have poles away from the origin, and this indicates that the least squares inversion of  $\mathbf{C}(z)$  inevitably is achieved through a recursive mechanism. The physical explanation for this is most easily demonstrated by considering a few simple examples. The principles are applicable also to more complex systems, but it is impractical to analyze such systems using only pure physical reasoning.

### 4.1 Discrete-time systems

Discrete-time signals are sequences of real numbers. Time sequences are divided into four groups: finite length sequences, right-sided sequences, left-sided sequences, and two-sided sequences (Oppenheim and Schaffer [35] Section 2.1). A finite length sequence has only a finite number of non-zero elements. A right-sided sequence is one for which all of its elements are zero for all  $n$  less than some integer  $n_1$ . Similarly, a left-sided sequence is one for which all of its elements are zero for all  $n$  greater than some integer  $n_2$ . A two-sided sequence has infinitely many non-zero elements for both  $n$  smaller than  $n_1$  and  $n$  greater than  $n_2$ . Thus, only some of the finite length sequences

and some of the right-sided sequences are causal, and all left-sided and two-sided sequences are non-causal. However, all finite length sequences and all right-sided sequences can be made causal by shifting them sufficiently far towards the positive end of the time axis. This means that all systems with an impulse response of one of those two types can be made causal by delaying the output by an appropriate amount,  $m$ . The positive integer  $m$  is measured in sampling intervals, and it is later interpreted as a “modelling delay”. When the sequence is the impulse response of a filter, the filter is said to be stable if every bounded input produces a bounded output (Oppenheim and Schaffer [35] Section 1.3). The properties of stability and causality for the four different types of sequences are illustrated in Figure 4.1.

The output sequence  $y(n)$ , which results from passing an input sequence  $x(n)$  through a causal digital filter, satisfies the linear difference equation (Oppenheim and Schaffer [35] Section 1.4)

$$\sum_{k=0}^N a(k)y(n-k) = \sum_{r=0}^M b(r)x(n-r). \quad (4.1.1)$$

The coefficient  $a(0)$  is usually assumed to be unity. The filter is naturally divided into two parts: a recursive part with the  $N+1$  coefficients  $a(k)$ , and a non-recursive part with the  $M+1$  coefficients  $b(r)$ . The behaviour of the filter is greatly simplified if it does not contain a recursive part. For this reason, digital filters are divided into two main groups: FIR (finite impulse response) filters, which do not contain a recursive part, and IIR (infinite impulse response) filters, which do. The impulse response  $h(n)$  of a digital filter is traditionally defined to be the causal sequence  $y(n)$  that satisfies the difference Equation (4.1.1) when  $x(n)$  is a unit-sample sequence  $\delta(n)$  (Oppenheim and Schaffer [35] Section 1.1). Without the causality restriction imposed, however, the impulse response  $h(n)$  is generally not uniquely determined. This ambiguity can also be resolved by constraining  $h(n)$  to be a stable sequence rather than a causal sequence. This problem is examined in more detail later in this section.

The mathematical tool that is used to analyse discrete time systems is the  $z$ -transform. The  $z$ -transform of a time sequence  $h(n)$  is defined as (Oppenheim and Schaffer [35] p.45)

$$H(z) = \sum_{n=-\infty}^{+\infty} h(n)z^{-n}, \quad (4.1.2)$$

so  $H(z)$  is a power series that can have infinitely many terms. When the number of terms in  $H(z)$  is finite, we will refer to  $H(z)$  as a polynomial even if it contains negative powers of  $z$ . The  $z$ -transform is a generalisation of the Fourier transform for discrete time signals since the two transforms yield identical results when they are evaluated on the unit circle defined by  $|z| = 1$  (Oppenheim and Schaffer [35] p.46). When  $H(z)$  is the  $z$ -transform of the impulse response  $h(n)$  of a digital filter,  $H(z)$  is often referred to as the system function (Oppenheim and Schaffer [35] Section 2.4). The system function evaluated on the unit circle is referred to as the frequency response of the filter. The system function  $H(z)$  that relates the output  $y(n)$  of a digital filter to its input  $x(n)$  is a rational polynomial in  $z$ ,

$$H(z) = Y(z) / X(z) = \left( \sum_{r=0}^M b(r)z^{-r} \right) / \left( \sum_{k=0}^N a(k)z^{-k} \right), \quad (4.1.3)$$

or, written out in full,

$$H(z) = \frac{b(0) + b(1)z^{-1} + b(2)z^{-2} + \dots + b(M)z^{-M}}{a(0) + a(1)z^{-1} + a(2)z^{-2} + \dots + a(N)z^{-N}}. \quad (4.1.4)$$

A more compact way of writing this expression is to use  $B(z)$  for the non-recursive part of the filter, and  $A(z)$  for the recursive part,

$$H(z) = B(z) / A(z). \quad (4.1.5)$$

The zeros of  $B(z)$  are the zeros of  $H(z)$ , and the zeros of  $A(z)$  are the poles of  $H(z)$ . The positions in the complex plane of the poles and zeros of the system function  $H(z)$  of a digital filter are central to the discussion of the qualitative properties of the filter's impulse response. Thus, it is common practice to illustrate the properties of a digital filter by plotting, in the complex plane, its poles as crosses ( $\times$ ) and its zeros as circles ( $\circ$ ). Such plots are referred to as pole-zero maps. Pole-zero maps also provide a quick way to assess the frequency response of the filter. A pole close to the unit circle indicates a large frequency response at frequencies corresponding to the location of the pole, and a zero close to the unit circle indicates a small frequency response at frequencies corresponding to the location of the zero (Oppenheim and Schaffer [35] pp.70-73).

Most of the useful results concerning manipulations of Laplace transforms are also applicable to  $z$ -transforms (Oppenheim and Schaffer [35] p.67). The  $z$ -transform  $H(z)$  of the sequence  $h(n)$  obtained by sampling a continuous time signal  $h(t)$  can be found formally by making the substitution

$$z = e^{sT_s}, \quad (4.1.6)$$

in the expression for the Laplace transform of the sampled continuous time signal (Kraniauskas [112] Section 7.2.1) (sampling in continuous time is equivalent to multiplication by a train of delta functions, see Kraniauskas [112] Chapter 6). Here,  $T_s$  is the sampling interval which is one divided by the sampling frequency  $f_s$ . It is important to know that if  $h(t)$  contains energy at frequencies higher than half the sampling frequency  $f_s/2$ , there will be frequency components in the spectrum of the sampled signal that were not present in the signal before sampling. The frequency  $f_s/2$  is referred to as the Nyquist frequency  $f_{Nyq}$ . The "contamination" of the true spectrum caused by frequency components above  $f_{Nyq}$  is known as aliasing (Kraniauskas [112] Section 6.2.3). The practical implication is that if the sampling rate is too slow, some information about the original signal is lost, and it cannot be recovered from its sampled values. Poles in the Laplace transform  $H(s)$  of  $h(t)$  are mapped directly to the  $z$ -plane according to Equation (4.1.6) (Kraniauskas [112] Section 11.3.2). This means, loosely stated, that the issue of stability is unaffected by the sampling process. An equivalent result does not exist for the zeros of  $H(s)$ , they do not map directly to the  $z$ -plane like the poles. In the remaining part of this chapter, we will assume that  $T_s$  is unity for notational convenience.

In order to find the impulse response  $h(n)$  corresponding to  $H(z)$ , it is necessary to apply the inverse  $z$ -transform to  $H(z)$ . In the time domain,  $h(n)$  is the result of the convolution of  $a_{inv}(n)$  and  $b(n)$ ,

$$h(n) = a_{inv}(n) * b(n), \quad (4.1.7)$$

where  $b(n)$   $z$ -transformed is  $B(z)$  and  $a_{inv}(n)$   $z$ -transformed is  $1/A(z)$ . There is only one sequence  $b(n)$  that has the  $z$ -transform  $B(z)$ , this sequence is unique and of finite length. However, there are usually several different sequences  $a_{inv}(n)$  that have the  $z$ -transform  $1/A(z)$ . The qualitative behaviour of  $h(n)$  is determined by the positions in the complex plane of the roots of the denominator polynomial  $A(z)$ , these roots being

the poles of the system. The contribution to  $a_{\text{inv}}(n)$  from each pole is an exponential term (Oppenheim and Schaffer [35] p.70). If the pole is inside the unit circle, the exponential term is either stable and right-sided, or unstable and left-sided. If the pole is outside the unit circle, the exponential term is either stable and left-sided, or unstable and right-sided. The inverse  $z$ -transform is defined by a contour integral,

$$h(n) = \frac{1}{2\pi j} \oint_C H(z) z^{n-1} dz, \quad (4.1.8)$$

where  $C$  is a counterclockwise closed contour in the region of convergence of  $H(z)$  and encircling the origin of the  $z$ -plane (Oppenheim and Schaffer [35] Section 2.2). In practice, inversion of a  $z$ -transform is usually calculated using the theory of residues. Since  $h(n)$  is effectively a Laurent power series expansion of  $H(z)$ , each power series has a corresponding ring of convergence in the complex  $z$ -plane, which cannot contain any poles, in which  $h(n)$  is uniquely determined (Churchill and Brown [27] Section 46). By forcing  $h(n)$  to be stable, the unit circle is guaranteed to be in the region of convergence, and consequently the Fourier transform of  $h(n)$  always exists (Oppenheim and Schaffer [35] p.22). Equivalently, by choosing the unit circle as the integration contour,  $h(n)$  is guaranteed to be stable with a well-defined frequency response. Therefore, the ambiguity of the inverse  $z$ -transform can be resolved by requiring that  $h(n)$  must be the impulse response of a stable system. It is worth noting that the roots of  $B(z)$ , which are the zeros of the system, can in some cases cancel out exactly some of the poles. If a pole outside the unit circle is not cancelled by a zero, it is impossible to realize  $h(n)$  as a causal and stable filter. If all the poles are inside the unit circle, it is always possible to realize  $h(n)$  as a causal and stable filter, but it might be necessary to use a modelling delay.

## 4.2 Inversion of single-channel systems

We now look at the problem of inverting exactly a single-channel system given by its impulse response  $c(n)$ . Assume that  $c(n)$  is a causal finite length sequence containing  $N$  coefficients, such as a measured or modelled electro-acoustic transfer function. In that case, the  $z$ -transform of  $c(n)$  is of the form

$$C(z) = \sum_{n=0}^{N-1} c(n) z^{-n} = c(0) + c(1)z^{-1} + c(2)z^{-2} + \dots + c(N-1)z^{-(N-1)}, \quad (4.2.1)$$

where  $N$  is some non-negative integer. If an input sequence  $d(n)$  is passed through the system, a corresponding output sequence,  $w(n)$ , is generated. In the  $z$ -domain

$$W(z) = C(z)D(z). \quad (4.2.2)$$

If a filter with the  $z$ -transform  $H(z) = 1/C(z)$  is added to the system, the net output is the same as the input,

$$W(z) = [1/C(z)]C(z)D(z) = D(z), \quad (4.2.3)$$

which shows that the ideal inverse filter  $H(z)$  is  $1/C(z)$ . This is illustrated in Figure 4.2. Therefore

$$H(z) = \frac{1}{C(z)} = \frac{1}{c(0) + c(1)z^{-1} + c(2)z^{-2} + \dots + c(N)z^{-N-1}}, \quad (4.2.4)$$

which is a special case of Equation (4.1.5) with  $B(z)$  equal to one and  $A(z)$  equal to  $C(z)$ . A system is said to be *minimum phase* if the impulse response of its inverse is both causal and stable (Oppenheim and Schaffer [35] p. 347). Consequently, a sequence  $c(n)$  is a minimum phase sequence only if all the zeros and poles of its  $z$ -transform are inside the unit circle. Minimum phase sequences have a minimal delay property (Oppenheim and Schaffer [35] Section 7.2), which effectively guarantees that the sequence has most of its energy concentrated at the start. A signal that contains echos, as for example the response of a room to an impulsive sound, is not likely to be a minimum phase signal due to the late arrivals of reflections (Neely and Allen [53]). If  $c(0)$  is zero, then  $c(n)$  cannot be minimum phase because the  $z$ -transform of a pure delay  $z^{-N}$  has its  $N$  zeros at infinity. If all the zeros of the  $z$ -transform of  $c(n)$  are outside the unit circle, it is called a maximum phase sequence. If  $C(z)$  has a zero very close to the unit circle, then  $H(z)$  has a pole very close to the unit circle. This makes the impulse response of the inverse filter very long. The time constant  $\tau$ , in samples, associated with a single pole close to the unit circle is approximately proportional to the reciprocal of the distance  $r$  from the pole to the unit circle, so

$$\tau = \frac{1}{r} \quad (4.2.5)$$

when  $r \ll 1$  (Bellanger [34] p.149). The time constant  $\tau_N$  associated with a pole of multiplicity  $N$  is approximately (Bellanger [34] p.153)



$$\tau_N = \sqrt{N} \cdot \tau. \quad (4.2.6)$$

Figure 4.3 illustrates qualitatively some of the results mentioned above. The left column of Figure 4.3 shows a minimum phase sequence  $c_1(n)$ , the position of the zero of  $C_1(z)$  in the complex plane, the position of the pole of the inverse filter  $H_1(z)$ , and finally the two sequences that both have the  $z$ -transform  $H_1(z)$ . The right column of Figure 4.3 shows the equivalent plots for a maximum phase sequence.

### 4.3 Inversion of multi-channel systems

We now look at the problem of inverting a multi-channel system. We wish to reproduce  $\mathbf{d}(z)$ , a vector of desired signals specified at  $R$  microphones. There are  $S$  sources, and the inputs  $\mathbf{v}(z)$  to the sources are the result of passing  $\mathbf{d}(z)$  through a matrix of inverse filters,  $\mathbf{H}(z)$ ,

$$\mathbf{v}(z) = \mathbf{H}(z)\mathbf{d}(z). \quad (4.3.1)$$

The signals that are reproduced at the microphones,  $\mathbf{w}(z)$ , are the result of passing  $\mathbf{v}(z)$  through a matrix of electro-acoustic transfer functions,  $\mathbf{C}(z)$ ,

$$\mathbf{w}(z) = \mathbf{C}(z)\mathbf{v}(z). \quad (4.3.2)$$

The difference between  $\mathbf{d}(z)$  and  $\mathbf{w}(z)$  is the error  $\mathbf{e}(z)$ ,

$$\mathbf{d}(z) - \mathbf{w}(z) = \mathbf{e}(z). \quad (4.3.3)$$

This multi-channel reproduction problem is illustrated schematically in Figure 4.4. The electro-acoustic transfer paths are given by the  $R$  times  $S$  impulse responses  $c_{rs}(n)$ . These impulse responses are assumed to be causal finite length sequences so their  $z$ -transforms are all of the same form as Equation (4.2.1). The matrix  $\mathbf{C}(z)$  contains the  $z$ -transforms of the individual impulse responses,

$$\mathbf{C}(z) = \begin{bmatrix} C_{11}(z) & C_{12}(z) & \cdots & C_{1S}(z) \\ C_{21}(z) & & & \\ \vdots & & \ddots & \\ C_{R1}(z) & & & C_{RS}(z) \end{bmatrix}. \quad (4.3.4)$$

The inverse filter matrix  $\mathbf{H}(z)$  has  $S$  rows and  $R$  columns,

$$\mathbf{H}(z) = \begin{bmatrix} H_{11}(z) & H_{12}(z) & \cdots & H_{1R}(z) \\ H_{21}(z) & & & \\ \vdots & & \ddots & \\ H_{S1}(z) & & & H_{SR}(z) \end{bmatrix}. \quad (4.3.5)$$

The simplest case is when  $R$  is equal to  $S$ , then both  $\mathbf{C}(z)$  and  $\mathbf{H}(z)$  are square. If  $\mathbf{C}(z)$  is also regular, it has a unique inverse. However,  $\mathbf{C}(z)$  is singular for the values of  $z$  that make the determinant of  $\mathbf{C}(z)$  equal to zero. This suggests a way of revealing the structure of  $\mathbf{H}(z)$ . The matrix  $\mathbf{C}(z)$  can be inverted formally by dividing its adjoint by its determinant  $|\mathbf{C}(z)|$  (Kreyszig [12] p.338),

$$\mathbf{H}(z) = \mathbf{C}^{-1}(z) = \frac{\text{adj}[\mathbf{C}(z)]}{|\mathbf{C}(z)|}. \quad (4.3.6)$$

The adjoint of a square matrix is another square matrix with dimensions identical to those of the original matrix. It contains the cofactors of the original matrix. Since the elements of  $\mathbf{C}(z)$  are polynomials in  $z^{-1}$ , the elements of the adjoint of  $\mathbf{C}(z)$  and the determinant of  $\mathbf{C}(z)$  are also polynomials in  $z^{-1}$ . It is clear from Equation (4.3.6) that all the inverse filters have the common denominator  $|\mathbf{C}(z)|$ . Thus, all the inverse filters have a common set of poles which are given by the zeros of the determinant of  $\mathbf{C}(z)$ . Methods exist for assessing the distribution of the roots of a polynomial as a function of its coefficients (Jury [110] Chapter 3), but they are impractical if the polynomial has many coefficients, so instead we look at the number of roots  $N_z$  of  $|\mathbf{C}(z)|$ .  $N_z$  depends on the number of microphones and loudspeakers, and on the number of coefficients in the impulse responses  $c_{rs}(n)$ , and it is calculated as follows. The determinant of  $\mathbf{C}(z)$  is a sum of products, each of which is the result of a multiplication of  $S$  elements of  $\mathbf{C}(z)$  (one from each column and row). If each  $c_{rs}(n)$  contains  $N$  coefficients, then the highest possible order (strictly speaking the maximum difference between the highest and lowest power with a non-zero coefficient) of such a product is  $S$  times  $N-1$ . Consequently,  $|\mathbf{C}(z)|$  is a polynomial of the same order, and it therefore has at most  $S$  times  $N-1$  zeros away from the origin,

$$N_z(|\mathbf{C}(z)|) \leq S(N-1). \quad (4.3.7)$$

Similar reasoning shows that each element of the adjoint of  $\mathbf{C}(z)$  has at most  $S-1$  times  $N-1$  zeros away from the origin,

$$N_z(\text{adj}(\mathbf{C}(z))) \leq (S-1)(N-1). \quad (4.3.8)$$

Consequently, the exact inverse filters generally have more poles than zeros. For example, if  $R = S = 4$  and  $N = 128$ , then each of the inverse filters has at most 508 poles and 381 zeros. If  $R = S = 4$  and  $N = 15$ , then each of the inverse filters has at most 56 poles and 42 zeros. The greater  $N$  is, the greater is the number of poles in the exact inverse. This confirms the well-known fact that practical inversion of multi-channel systems is much more difficult to achieve when the individual  $c_{rs}(n)$  are long than when they are short.

It is interesting to note that the poles of the inverse filters usually do not have anything to do with the zeros of the individual elements of  $\mathbf{C}(z)$ . For example, consider the 2-by-2 system given by the four transfer functions  $C_{11}(z)$ ,  $C_{12}(z)$ ,  $C_{21}(z)$ , and  $C_{22}(z)$  which each contains two coefficients,

$$\mathbf{C}(z) = \begin{bmatrix} 1+z^{-1} & 1+0.5z^{-1} \\ 1+0.5z^{-1} & 1+z^{-1} \end{bmatrix}. \quad (4.3.9)$$

$C_{11}(z)$  and  $C_{22}(z)$  both have the single zero  $z = -1$ , and  $C_{12}(z)$  and  $C_{21}(z)$  both have the single zero  $z = -0.5$ . However, the determinant of  $\mathbf{C}(z)$ ,

$$|\mathbf{C}(z)| = z^{-1} + 0.75z^{-2}, \quad (4.3.10)$$

has the two zeros  $z = \text{infinity}$  and  $z = -0.75$  which are not zeros of any of the individual elements of  $\mathbf{C}(z)$ .

When there are more microphones than loudspeakers, it is generally not possible to invert  $\mathbf{C}(z)$  exactly, the problem is then referred to as over-determined. For continuous time systems, a least squares solution is calculated as a function of frequency. For discrete time systems, the same approach is used, but instead of minimising the squared absolute value of the Laplace transform of the error at each point on the imaginary axis, the squared absolute value of the  $z$ -transform of the error is minimised at each point on the unit circle. Parseval's relation gives a useful way of interpreting this property in the time domain. Parseval's relation states that

$$\sum_{n=-\infty}^{\infty} x(n)y^*(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})Y^*(e^{j\omega}) d\omega \quad (4.3.11)$$

if  $X(z)$  and  $Y(z)$  converge on the unit circle (Oppenheim and Schaffer [35] Section 2.3.10). In the special case where  $X(z) = Y(z) = E(z)$ , the equivalent relation is

$$\sum_{n=-\infty}^{\infty} |e(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(e^{j\omega})|^2 d\omega, \quad (4.3.12)$$

The performance cost  $E$  measures the squared error when it is added up for all the microphones and integrated around the unit circle. Parseval's relation then gives

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{e}^H(e^{j\omega}) \cdot \mathbf{e}(e^{j\omega}) d\omega = \sum_{r=1}^R \sum_{n=-\infty}^{\infty} |e_r(n)|^2, \quad (4.3.13)$$

This means that even though the solution is found by minimising the error in the frequency domain, the solution also minimises the error in the time domain. Thus,  $E$  can be calculated by adding up the energies of the “error sequences” at all the microphones.

When there are more microphones than loudspeakers, the problem is over-determined. The inverse filter matrix  $\mathbf{H}(z)$  is found by calculating the pseudo-inverse  $\mathbf{C}_{\text{pinv}}(z)$  of  $\mathbf{C}(z)$  using the substitution  $z = e^{j\omega}$ ,

$$\mathbf{H}(e^{j\omega}) = \mathbf{C}_{\text{pinv}}(e^{j\omega}) = [\mathbf{C}^H(e^{j\omega})\mathbf{C}(e^{j\omega})]^{-1} \cdot \mathbf{C}^H(e^{j\omega}). \quad (4.3.14)$$

Since  $e^{j\omega}$  conjugated is equal to  $e^{-j\omega}$ , this is equivalent to

$$\mathbf{H}(e^{j\omega}) = \mathbf{C}_{\text{pinv}}(e^{j\omega}) = [\mathbf{C}^T(e^{-j\omega})\mathbf{C}(e^{j\omega})]^{-1} \cdot \mathbf{C}^T(e^{-j\omega}). \quad (4.3.15)$$

Thus, the  $z$ -transform of the matrix of inverse filters that minimise the error at each point on the unit circle is

$$\mathbf{H}(z) = [\mathbf{C}^T(z^{-1})\mathbf{C}(z)]^{-1} \cdot \mathbf{C}^T(z^{-1}). \quad (4.3.16)$$

Since  $\mathbf{H}(z)$  is chosen such that the frequency response is optimal on the unit circle  $|z| = 1$ , the inverse filters must be chosen such that the region of convergence of their  $z$ -transforms includes the unit circle. As stated in Section 4.1, this is equivalent to constraining all the inverse filters to be stable. Just as with the square systems, the

pole-zero structure of  $\mathbf{H}(z)$  can be revealed by dividing the adjoint of  $\mathbf{C}^T(z^{-1})\mathbf{C}(z)$  with its determinant,

$$\mathbf{H}(z) = \frac{\text{adj}[\mathbf{C}^T(z^{-1})\mathbf{C}(z)]}{|\mathbf{C}^T(z^{-1})\mathbf{C}(z)|} \cdot \mathbf{C}^T(z^{-1}), \quad (4.3.17)$$

which again shows that the inverse filters share a common set of poles given by the zeros of the determinant of  $\mathbf{C}^T(z^{-1})\mathbf{C}(z)$ . These zeros always appear in structured patterns. Since the determinant of a square matrix is equal to the determinant of its transpose (Kreyszig [12] p.327), and since

$$[\mathbf{C}^T(z^{-1})\mathbf{C}(z)]^T = \mathbf{C}^T(z)\mathbf{C}(z^{-1}), \quad (4.3.18)$$

it follows that if  $z_0$  is a zero of  $|\mathbf{C}^T(z^{-1})\mathbf{C}(z)|$ , then so is  $1/z_0$ . This means that for each zero outside the unit circle there is a corresponding zero inside the unit circle. Furthermore, since the coefficients of the polynomials  $C_{rs}(z)$  are real, the zeros of the determinant always occur in complex conjugate pairs (Press et al [71] p.369). The net effect is that zeros that are not on the real axis always occur in groups of four because if  $z_0$  is a zero, so are  $1/z_0^*$ ,  $z_0^*$ , and  $1/z_0$ , this is illustrated in Figure 4.5a. Zeros on the real axis occur in groups of two because  $z_0 = z_0^*$  and  $1/z_0 = 1/z_0^*$ , this is illustrated in Figure 4.5b. The numbers  $z_0$  and  $1/z_0^*$  are said to be at conjugate reciprocal positions. One of them is inside the unit circle, the other is outside, and both lie on the same radial line going through the origin; they are "reflections" of each other in the unit circle. This also applies to the pair  $z_0^*$  and  $1/z_0$  because they are the reflection of the pair  $z_0$  and  $1/z_0^*$  in the real axis. Consequently,  $|\mathbf{C}^T(z^{-1})\mathbf{C}(z)|$  is usually non-minimum phase, and the exact inverse filters therefore cannot be both causal and stable unless the zeros of  $|\mathbf{C}^T(z^{-1})\mathbf{C}(z)|$  that are outside the unit circle are cancelled out exactly by the zeros of each element of the numerator, which is the adjoint of  $\mathbf{C}^T(z^{-1})\mathbf{C}(z)$  times  $\mathbf{C}^T(z^{-1})$ . Although this cancellation is possible in theory, it has never been observed for any numerical example. If each  $c_{rs}(n)$  contains  $N$  coefficients, then  $|\mathbf{C}^T(z^{-1})\mathbf{C}(z)|$  has at most  $2S$  times  $N-1$  zeros away from the origin. Each of the inverse filters has at most  $2S-1$  times  $N-1$  zeros away from the origin, so there are generally more poles than zeros. Note that the number of poles and zeros do not depend on the number of microphones.

When there are more loudspeakers than microphones,  $\mathbf{H}(z)$  is not uniquely determined because many solutions give zero error, so the problem is under-determined. In that case, the pseudo-inverse is usually chosen as the solution that achieves zero error with a minimum of effort (Keener [77] Section 1.5). By using the same arguments as above,  $\mathbf{H}(z)$  is derived to be

$$\mathbf{H}(z) = \mathbf{C}^T(z^{-1}) \cdot [\mathbf{C}(z)\mathbf{C}^T(z^{-1})]^{-1}. \quad (4.3.19)$$

This solution for the under-determined problem is closely linked to the solution to the over-determined problem given by Equation (4.3.14). The solution to the under-determined problem is the transpose of the solution to the over-determined system that results from replacing  $\mathbf{C}(z)$  by its transpose,  $\mathbf{C}^T(z)$ .  $\mathbf{C}^T(z)$  is the electro-acoustic transfer matrix of the system that is derived from the system that has  $\mathbf{C}(z)$  as its electro-acoustic transfer matrix, by replacing all sources with microphones, and replacing all microphones with sources. To verify that this is true, transpose both sides of Equation (4.3.16), then replace  $\mathbf{C}(z)$  by  $\mathbf{C}^T(z)$ . Loosely stated, this means that when  $\mathbf{C}(z)$  is replaced by  $\mathbf{C}^T(z)$ ,  $\mathbf{H}(z)$  is replaced by  $\mathbf{H}^T(z)$ . Consequently, the inverse filters share a common set of poles, given by the zeros of the determinant of  $\mathbf{C}(z)\mathbf{C}^T(z^{-1})$ , which has the same symmetry properties as the common set of poles for an over-determined problem.

#### 4.4 A simple class of multi-channel systems

In order to investigate the behaviour of some basic multi-channel systems, it is convenient to assume that both the electro-acoustic transfer matrix  $\mathbf{C}(z)$  and the desired signals  $\mathbf{d}(z)$  are defined by simple expressions. Therefore, each electro-acoustic transfer function  $c_{rs}(n)$  is approximated with a simple one-coefficient digital filter. The value of its single element depend only on the distances  $l_{rs}$  between the sources and the microphones measured in sampling intervals. The individual  $l_{rs}$  are generally positive real numbers, but in order to keep the systems as simple as possible, all the examples in this section are chosen so that the  $l_{rs}$  are integers. The distance between a source and a microphone results in a delay and an attenuation proportional to the distance, so the  $z$ -transform of  $c_{rs}(n)$  is given by

$$C_{rs}(z) = \frac{z^{-l_{rs}}}{l_{rs}}. \quad (4.4.1)$$

Strictly speaking, it is also necessary to include a constant in this expression, but this has also been omitted for convenience. The desired signals  $\mathbf{d}(z)$  are chosen to contain all unit values, so

$$\mathbf{d}(z) = \begin{bmatrix} D_1(z) \\ D_2(z) \\ \vdots \\ D_R(z) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (4.4.2)$$

In the time domain, this corresponds to a single impulse at time zero at each of the microphones.

#### 4.5 Least squares inversion of a 2-by-1 system

The simplest multi-channel system that does not have an exact inverse comprises one source,  $S_1$ , and two microphones,  $R_1$  and  $R_2$ . The microphone  $R_1$  is the one closest to the source,  $R_2$  is the one furthest away. Their distances to the source, measured in sampling intervals, are  $l_{11}$  and  $l_{21}$  respectively, so  $\mathbf{C}(z)$  is given by the 2-by-1 matrix

$$\mathbf{C}(z) = \begin{bmatrix} z^{-l_{11}} / l_{11} \\ z^{-l_{21}} / l_{21} \end{bmatrix}. \quad (4.5.1)$$

The inverse filter matrix  $\mathbf{H}(z)$  is calculated from Equation (4.3.16), so

$$\mathbf{H}(z) = \left( \begin{bmatrix} \frac{z^{l_{11}}}{l_{11}} & \frac{z^{l_{21}}}{l_{21}} \end{bmatrix} \begin{bmatrix} \frac{z^{-l_{11}}}{l_{11}} \\ \frac{z^{-l_{21}}}{l_{21}} \end{bmatrix} \right)^{-1} \begin{bmatrix} \frac{z^{l_{11}}}{l_{11}} & \frac{z^{l_{21}}}{l_{21}} \end{bmatrix} = \frac{l_{11}^2 l_{21}^2}{l_{11}^2 + l_{21}^2} \begin{bmatrix} \frac{z^{l_{11}}}{l_{11}} & \frac{z^{l_{21}}}{l_{21}} \end{bmatrix}, \quad (4.5.2)$$

which shows that the two inverse filters have no poles because  $|\mathbf{C}^T(z^{-1})\mathbf{C}(z)|$  is a scalar. The desired signals  $\mathbf{d}(z)$  are given by

$$\mathbf{d}(z) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (4.5.3)$$

As mentioned in the previous section, in the time domain this corresponds to a single impulse at time zero at each of the two microphones. It is obvious that it is only

possible to make the two pulses appear simultaneously at the two microphones if  $R_1$  and  $R_2$  are exactly equally far away from the source; otherwise a pulse emitted from  $S_1$  will inevitably reach  $R_1$  before  $R_2$ . The single input  $V_1(z)$  to the loudspeaker is readily calculated, and is given by

$$V_1(z) = \mathbf{H}(z) \cdot \mathbf{d}(z) = \frac{1}{l_{21}/l_{11} + l_{11}/l_{21}} (l_{21}z^{l_{11}} + l_{11}z^{l_{21}}). \quad (4.5.4)$$

It is seen that the source sends out two pulses: one that will reach  $R_1$  at time zero and another that will reach  $R_2$  also at time zero. The source output is clearly non-causal. The important quantities are the relative amplitudes of the two pulses. Figure 4.6 shows the amplitudes of the two pulses for three different choices of the pair  $(l_{11}, l_{21})$ . The three pairs are a)  $l_{11} = 10, l_{21} = 12$ , b)  $l_{11} = 10, l_{21} = 20$ , and c)  $l_{11} = 10, l_{21} = 40$ . The simplest case is when  $l_{11} = l_{21}$  (not shown in the figure). If  $l_{11} = l_{21}$ , the source emits only one pulse that reaches both microphones at the same time; this is a special case where the error is zero and the pulses at  $R_1$  and  $R_2$  are reproduced perfectly by a single source. When  $R_2$  is much further away from the source than  $R_1$ , the first pulse is close to zero and the second pulse is close to ten. It might seem strange that the pulse that reaches  $R_2$  at time zero is emitted at all, because this pulse will first reach the nearer microphone with a higher amplitude at the "wrong" time. The reason for this can be understood by using the relation between the error calculated in the time domain and in the frequency domain as stated in Equation (4.3.13). The error that is minimised in the frequency domain is given by Equation (4.3.12). The error vector  $\mathbf{e}(z)$  depends not only on  $\mathbf{H}(z)$  and  $\mathbf{C}(z)$ , but also on the desired signal  $\mathbf{d}(z)$ ,

$$\mathbf{e}(z) = \mathbf{d}(z) - \mathbf{C}(z)\mathbf{H}(z)\mathbf{d}(z) = [\mathbf{I} - \mathbf{C}(z)\mathbf{H}(z)] \cdot \mathbf{d}(z). \quad (4.5.5)$$

The matrix  $\mathbf{I} - \mathbf{C}(z)\mathbf{H}(z)$  can be thought of as an "error filter matrix" whose output is  $\mathbf{e}(z)$  when the input is  $\mathbf{d}(z)$ . For the 2-by-1 example, we find, after rearranging terms,

$$\mathbf{I} - \mathbf{C}(z)\mathbf{H}(z) = \frac{1}{l_{21}/l_{11} + l_{11}/l_{21}} \begin{bmatrix} 1 - l_{21}/l_{11} & -z^{l_{21}-l_{11}} \\ -z^{l_{11}-l_{21}} & 1 - l_{11}/l_{21} \end{bmatrix} \quad (4.5.6)$$

and so, with  $\mathbf{d}(z)$  given by Equation (4.4.2),

$$\mathbf{e}(z) = \begin{bmatrix} e_1(z) \\ e_2(z) \end{bmatrix} = \frac{1}{l_{21}/l_{11} + l_{11}/l_{21}} \begin{bmatrix} l_{11}/l_{21} - z^{l_{21}-l_{11}} \\ l_{21}/l_{11} - z^{l_{11}-l_{21}} \end{bmatrix}. \quad (4.5.7)$$



It is seen that  $\mathbf{e}(z)$  is zero when  $l_{11}$  is equal to  $l_{21}$ . When  $l_{11}$  is not equal to  $l_{21}$ , the performance cost  $E$ , which is the sum of the energies of the two time sequences  $e_1(n)$  and  $e_2(n)$  (see Equation 4.3.13), is given by

$$E = 1 + \frac{2}{\left(l_{11}/l_{21}\right)^2 + \left(l_{21}/l_{11}\right)^2}, \quad (4.5.8)$$

The performance cost is always greater than one and less than two. For example, for  $l_{11} = 10$  and  $l_{21} = 12$ ,  $E = 1.94$ , and for  $l_{11} = 10$  and  $l_{21} = 40$ ,  $E = 1.12$ . In the case where all the sources are silent,  $E$  is equal to the sum of the energies in  $\mathbf{d}(z)$ . When  $\mathbf{d}(z)$  contains all ones, that sum is  $R$ , the number of microphones. This is the worst case,  $E$  can never be greater than  $R$ . When the sources are optimally adjusted,  $E$  is less than  $R$ . The time domain interpretation of this is that a pulse has to come along at a microphone at the "right" time, which is at time zero in the special case when  $\mathbf{d}(z)$  contains all ones. If a pulse arrives at any other time at that particular microphone,  $E$  goes up. Therefore, we can loosely refer to "wrong" and "right" times of arrival of a pulse at a microphone. Seen from the point of view of a microphone  $r$ , the time can be right for two reasons. The first is that the desired signal  $d_r$  is different from zero at time  $n$ . When  $\mathbf{d}(z)$  contains all unit values, this happens only once for each microphone. The other reason is that a source might try to cancel or attenuate a pulse which has arrived at the wrong time. A pulse emitted from a source will not only reach one microphone, it will eventually reach all of them, so it is possible to reduce  $E$  by attenuating or cancelling a pulse that has arrived at a microphone at the wrong time. When there are only two microphones, each pulse emitted creates only one additional unwanted pulse. Therefore, it is possible to predict the behaviour of such a system by using purely physical reasoning. When there are more than two microphones, this quickly becomes impractical, if not impossible, which makes it necessary to use numerical methods to analyse such systems.

#### 4.6 Least effort inversion of a 1-by-2 system

The simplest multi-channel system that has many solutions that all give zero error comprises two sources,  $S_1$  and  $S_2$ , and one microphone,  $R_1$ . The distances from the

microphone to the sources, measured in sample intervals, are  $l_{11}$  and  $l_{12}$  (note index 12 rather than 21), so  $\mathbf{C}(z)$  is given by the 1-by-2 matrix

$$\mathbf{C}(z) = \begin{bmatrix} \frac{z^{-l_{11}}}{l_{11}} & \frac{z^{-l_{12}}}{l_{12}} \end{bmatrix}. \quad (4.6.1)$$

As demonstrated in section 4.3, the inverse filter matrix  $\mathbf{H}(z)$  is the transpose of the solution given in Equation (4.3.16), so

$$\mathbf{H}(z) = \frac{l_{11}^2 l_{12}^2}{l_{11}^2 + l_{12}^2} \begin{bmatrix} z^{l_{11}} / l_{11} \\ z^{l_{12}} / l_{12} \end{bmatrix}. \quad (4.6.2)$$

When  $D(z)$ , which is a scalar in this case, is one,  $\mathbf{v}(z)$  is equal to  $\mathbf{H}(z)$ , which means that  $S_1$  and  $S_2$  send out one pulse each, both of these pulses reach  $R_1$  at time zero. The desired signal is reproduced exactly with a minimum of total output from the two sources.

#### 4.7 Exact inversion of a 2-by-2 system

The simplest multi-channel system that has an exact inverse comprises two sources,  $S_1$  and  $S_2$ , and two microphones,  $R_1$  and  $R_2$ . The matrix  $\mathbf{C}(z)$  is given by

$$\mathbf{C}(z) = \begin{bmatrix} z^{-l_{11}} / l_{11} & z^{-l_{12}} / l_{12} \\ z^{-l_{21}} / l_{21} & z^{-l_{22}} / l_{22} \end{bmatrix}, \quad (4.7.1)$$

and the adjoint of  $\mathbf{C}(z)$  is

$$\text{adj}[\mathbf{C}(z)] = \begin{bmatrix} z^{-l_{22}} / l_{22} & -z^{-l_{21}} / l_{21} \\ -z^{-l_{12}} / l_{12} & z^{-l_{11}} / l_{11} \end{bmatrix}. \quad (4.7.2)$$

The determinant of  $\mathbf{C}(z)$  is

$$|\mathbf{C}(z)| = \frac{z^{-(l_{11}+l_{22})}}{l_{11}l_{22}} \left[ 1 - \frac{l_{11}l_{22}}{l_{21}l_{12}} z^{(l_{11}+l_{22})-(l_{21}+l_{12})} \right], \quad (4.7.3)$$

and since the individual elements of the adjoint of  $\mathbf{C}(z)$  do not have any zeros away from infinity, the zeros of the determinant of  $\mathbf{C}(z)$  cannot be cancelled by the numerator. Therefore, all the zeros of  $|\mathbf{C}(z)|$  are poles of all the inverse filters. The value of each of the poles must satisfy

$$\frac{l_{21}l_{12}}{l_{11}l_{22}} = z^{(l_{11}+l_{22})-(l_{21}+l_{12})}, \quad (4.7.4)$$

therefore all the poles are the same distance away from the origin (Churchill and Brown [27] Section 7). When  $l_{11} = 4$ ,  $l_{12} = 6$ ,  $l_{21} = 6$ , and  $l_{22} = 4$ , the poles of the inverse are all inside the unit circle. This is illustrated in Figure 4.7a. When  $l_{11} = 8$ ,  $l_{12} = 2$ ,  $l_{21} = 17$ , and  $l_{22} = 7$ , the poles of the inverse are all outside the unit circle, this is illustrated in Figure 4.7b. Most 2-by-2 systems have inverses whose poles are all inside the unit circle, in particular, this is the case when  $R_1$  is closest to  $S_1$  and  $R_2$  is closest to  $S_2$ , which is the case for the example shown in Figure 4.7a. It is useful to consider the behaviour of this system in some detail, because it demonstrates how the digital signal processing and the physics of the problem are linked together. Once again, we assume that the desired signals are of the form given by Equation (4.5.3) (two impulses at time zero). Since the two microphones and the two sources are positioned symmetrically around the origin of the coordinate system, the time responses  $v_1(n)$  and  $v_2(n)$  are identical. The  $z$ -transforms of the two loudspeaker inputs are

$$V_1(z) = V_2(z) = 12z^4 \frac{1}{3 + 2z^{-2}} \quad (4.7.5)$$

Before time  $n = -4$  the sources are silent, at time  $n = -4$  they start to emit pulses. First  $S_1$  sends out a pulse that reaches  $R_1$  at time zero, and at the same time  $S_2$  sends out a pulse that reaches  $R_2$  also at time zero. Now consider the situation at the microphones. The pulses that arrive at time zero are the only pulses that the microphones will see, this has to be so because the inversion of the 2-by-2 system is exact. However, the pulse emitted from  $S_1$  will also reach  $R_2$  and the pulse emitted from  $S_2$  will also reach  $R_1$ . These pulses are unwanted and they must be cancelled by pulses of opposite sign emitted from the opposite source. When a pulse is emitted with the purpose of cancelling another pulse, it will create an additional unwanted pulse, which is then cancelled by the other source, and so this cancelling goes on forever. It is the spherical spreading of the sound that ensures that the amplitude of the series of pulses emitted from each source is exponentially decaying. This type of geometry is that studied by Atal, Hill and Schroeder in their patent application from 1966 (Atal et al [18]) describing a cross-talk canceller, and they also used the physical

argument given above to justify that the system has a causal and stable inverse. Note, however, that this argument is only valid when the transfer functions contained in  $\mathbf{C}(z)$  are of the single-coefficient type given by Equation (4.4.1). If just one of the four  $c_n(n)$  contains more than one coefficient, the inverse filters can have poles both inside and outside the unit circle. An example of that is when  $\mathbf{C}(z)$  is given by

$$\mathbf{C}(z) = \begin{bmatrix} z^{-4}/4 & z^{-6}/6 \\ z^{-6}/6 & z^{-4}/4 + z^{-5}/5 \end{bmatrix}, \quad (4.7.6)$$

the determinant of this matrix has zeros both inside and outside the unit circle. Since the inverse of the system shown in Figure 4.7b has all its poles outside the unit circle, the stable time responses  $v_1(n)$  and  $v_2(n)$  are qualitatively "time-reversed" versions of the time responses shown in Figure 4.7a. As before, the cancellation of unwanted pulses guarantees that only one pulse at time zero is seen at each of the microphones. However, the pulse that is seen at the microphone is now emitted as the last of an exponentially increasing sequence rather than as the first of an exponentially decreasing sequence.

#### 4.8 Least squares inversion of a 3-by-2 system

The simplest "proper" multi-channel system that cannot be inverted exactly comprises two sources,  $S_1$  and  $S_2$ , and three microphones,  $R_1$ ,  $R_2$ , and  $R_3$ . The electro-acoustic transfer matrix  $\mathbf{C}(z)$  for a 3-by-2 system is

$$\mathbf{C}(z) = \begin{bmatrix} z^{-l_{11}}/l_{11} & z^{-l_{12}}/l_{12} \\ z^{-l_{21}}/l_{21} & z^{-l_{22}}/l_{22} \\ z^{-l_{31}}/l_{31} & z^{-l_{32}}/l_{32} \end{bmatrix}, \quad (4.8.1)$$

so

$$\mathbf{C}^T(z^{-1})\mathbf{C}(z) = \begin{bmatrix} \frac{1}{l_{11}^2} + \frac{1}{l_{21}^2} + \frac{1}{l_{31}^2} & \frac{z^{l_{11}-l_{12}}}{l_{11}l_{12}} + \frac{z^{l_{21}-l_{22}}}{l_{21}l_{22}} + \frac{z^{l_{31}-l_{32}}}{l_{31}l_{32}} \\ \frac{z^{l_{12}-l_{11}}}{l_{11}l_{12}} + \frac{z^{l_{22}-l_{21}}}{l_{21}l_{22}} + \frac{z^{l_{32}-l_{31}}}{l_{31}l_{32}} & \frac{1}{l_{12}^2} + \frac{1}{l_{22}^2} + \frac{1}{l_{32}^2} \end{bmatrix}. \quad (4.8.2)$$

The determinant of  $\mathbf{C}^T(z^{-1})\mathbf{C}(z)$  is

$$|C^T(z^{-1})C(z)| = \left( \frac{1}{l_{11}^2} + \frac{1}{l_{21}^2} + \frac{1}{l_{31}^2} \right) \left( \frac{1}{l_{12}^2} + \frac{1}{l_{22}^2} + \frac{1}{l_{32}^2} \right) + \left( \frac{1}{l_{11}^2 l_{12}^2} + \frac{1}{l_{21}^2 l_{22}^2} + \frac{1}{l_{31}^2 l_{32}^2} \right) \\ + \frac{z^{l_{12}-l_{11}+l_{21}-l_{22}} + z^{l_{11}-l_{12}+l_{22}-l_{21}}}{l_{11}l_{12}l_{21}l_{22}} + \frac{z^{l_{12}-l_{11}+l_{31}-l_{32}} + z^{l_{11}-l_{12}+l_{32}-l_{31}}}{l_{11}l_{12}l_{31}l_{32}} + \frac{z^{l_{22}-l_{21}+l_{31}-l_{32}} + z^{l_{21}-l_{22}+l_{32}-l_{31}}}{l_{21}l_{22}l_{31}l_{32}} \quad (4.8.3)$$

This expression looks complicated, but it can be simplified considerably by substituting the three powers of  $z$  with  $d_1$ ,  $d_2$ , and  $d_3$ , and the four coefficients with  $a_0$ ,  $a_1$ ,  $a_2$ , and  $a_3$ . It is then evident that the determinant of  $C(z^{-1})^T C(z)$  is of the form

$$|C^T(z^{-1})C(z)| = a_0 + a_1(z^{d_1} + z^{-d_1}) + a_2(z^{d_2} + z^{-d_2}) + a_3(z^{d_3} + z^{-d_3}), \quad (4.8.4)$$

which illustrates the symmetry in  $z$  and  $1/z$ . An example of a 3-by-2 system is shown in Figure 4.8a. The six distances in samples are  $l_{11} = 3$ ,  $l_{12} = 7$ ,  $l_{21} = 5$ ,  $l_{22} = 5$ ,  $l_{31} = 6$ , and  $l_{32} = 4$ , and the exact determinant of  $C(z^{-1})^T C(z)$  is given by

$$|C^T(z^{-1})C(z)| = \frac{289}{17640} - \frac{1}{600}(z^2 + z^{-2}) - \frac{1}{525}(z^4 + z^{-4}) - \frac{1}{504}(z^6 + z^{-6}). \quad (4.8.5)$$

It is a coincidence that only even powers of  $z$  appear in this polynomial. The polynomial has twelve roots away from the origin. Figure 4.8b shows the pole-zero maps of the 6 inverse filters. It is seen that the poles appear in pairs at conjugate reciprocal positions, there are six poles inside the unit circle and six poles outside the unit circle. Note that the six patterns of zeros do not appear to be related to each other, and they do not seem to share any common feature. Note also that the zeros do not tend to be particularly close to the poles.

## 4.9 Conclusions

The exact least squares inverse filters in  $H(z)$  generally have infinite impulse responses, and they share a common set of poles. These poles are readily calculated as the zeros of the determinant of the square matrix  $C(z)$  when there are equally many loudspeakers and microphones,  $C^T(z^{-1})C(z)$  when there are more microphones than loudspeakers, and  $C(z)C^T(z^{-1})$  when there are more loudspeakers than microphones. When the number of loudspeakers and microphones is not the same, the poles appear in pairs at conjugate reciprocal positions which means that for every pole inside the unit circle there is a corresponding pole outside the unit circle. The number of poles is generally greater than the number of zeros, but it is still possible, in theory, that all the

poles outside the unit circle are cancelled out exactly by the zeros of each of the inverse filters. However, this can only happen by an extraordinary coincidence and has never been observed with any numerical examples. Thus, the exact least squares inverse filters are generally unrealisable in the time domain. The recursive nature of the inverse filters have a physical interpretation since minimising the error in the  $z$ -domain is equivalent to minimising the sum of the energy of the error sequences in the time domain. Each pulse emitted from a loudspeaker has the purpose of either reproducing a desired pulse at a microphone, or cancelling out an unwanted pulse at a microphone. This principle can be used to demonstrate the important application of cross-talk cancellation in a system comprising two loudspeakers and two microphones.

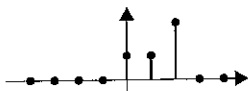
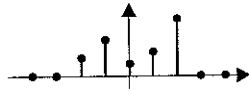
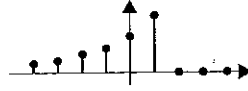

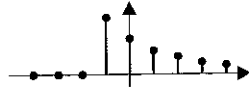

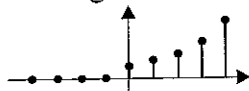
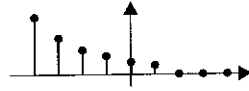
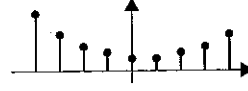
	Causal	Non-Causal	
Stable	Finite length 	Finite length 	Left-sided 
	Right-sided 	Right-sided 	Double-sided 
Unstable	Right-sided 	Left-sided 	Double-sided 

Figure 4.1. Summary of stability and causality properties of the four different types of time sequences, finite length, right-sided, left-sided, and double-sided

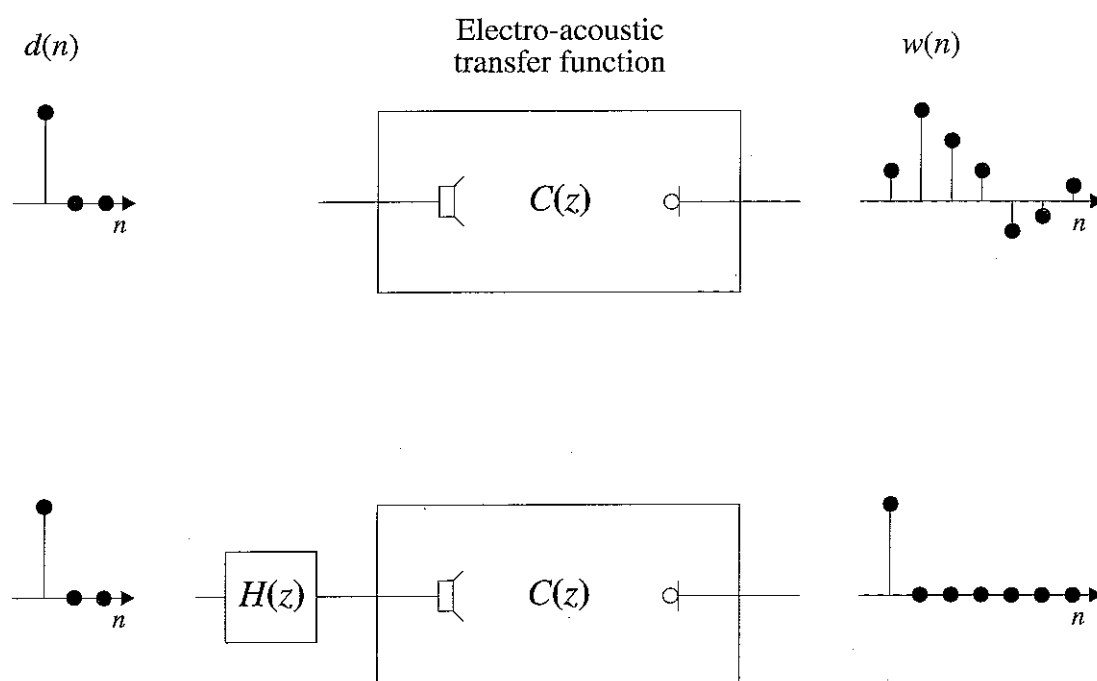


Figure 4.2. The principle of single-channel inverse filtering. The effect of  $C(z)$  is undone by  $H(z)$



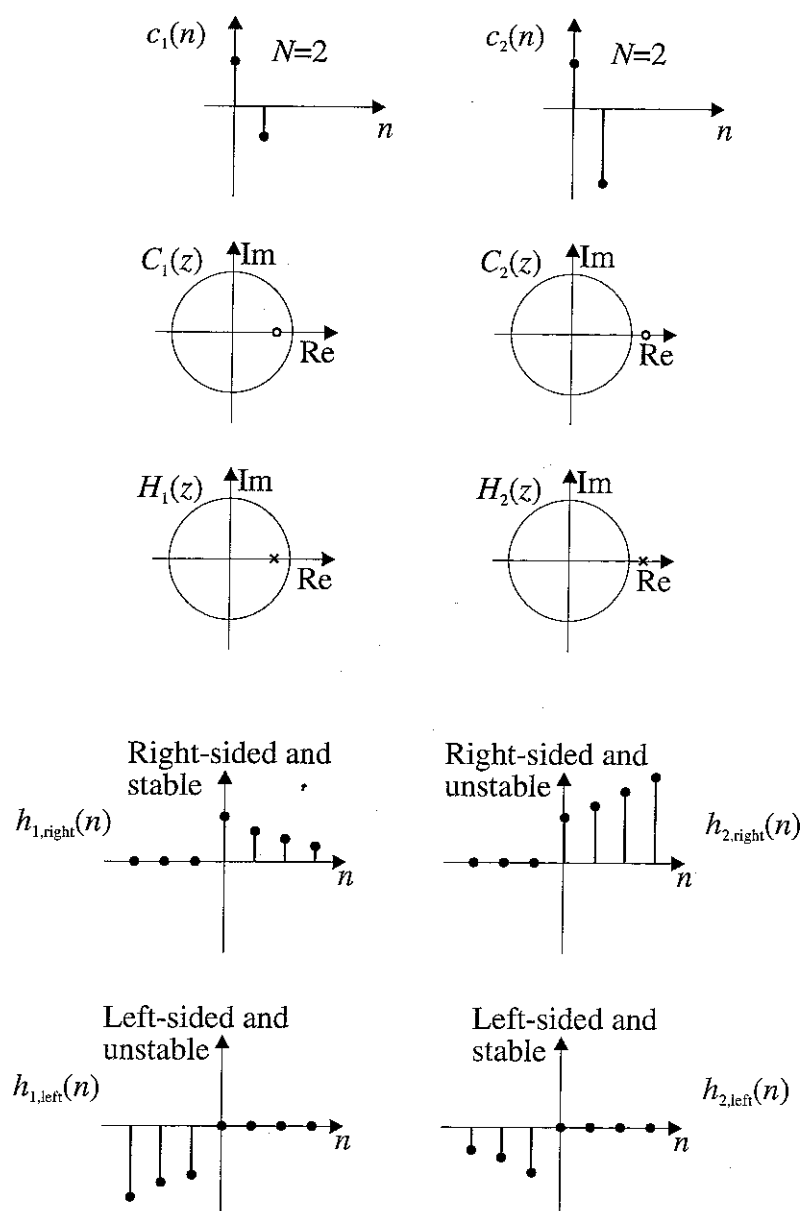


Figure 4.3. Two examples of a filter and its two exact inverses

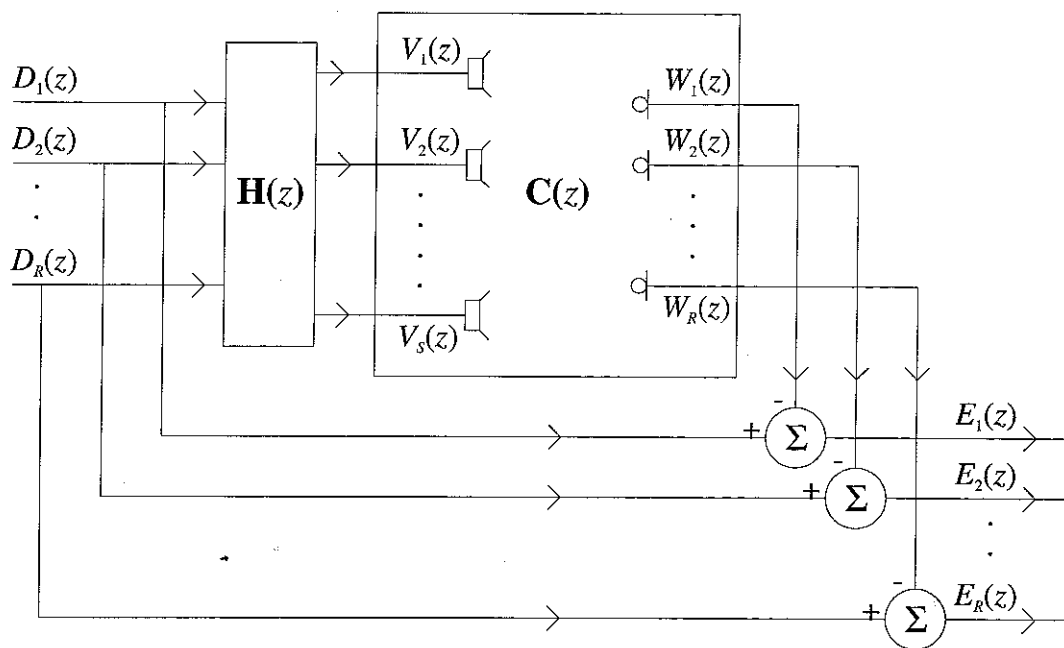


Figure 4.4. The multi-channel sound reproduction problem when the desired signals  $\mathbf{d}(z)$  are exact copies of the recorded signals  $\mathbf{u}(z)$

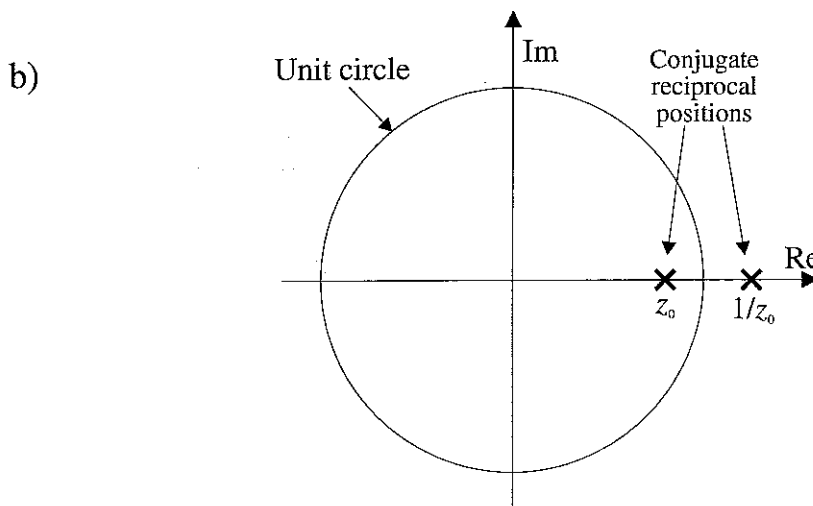
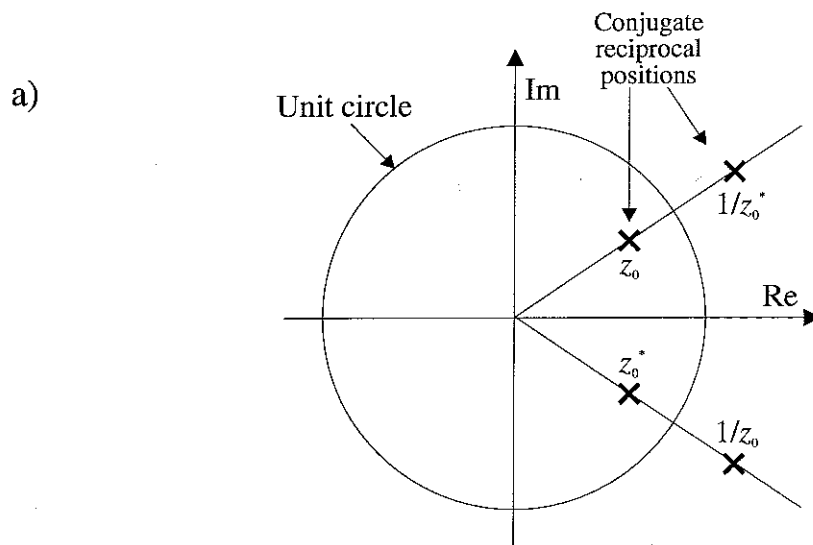


Figure 4.5. The zeros of  $|C^T(z^{-1})C(z)|$  always appears in a) groups of four or b) groups of two. Note the symmetry in the unit circle

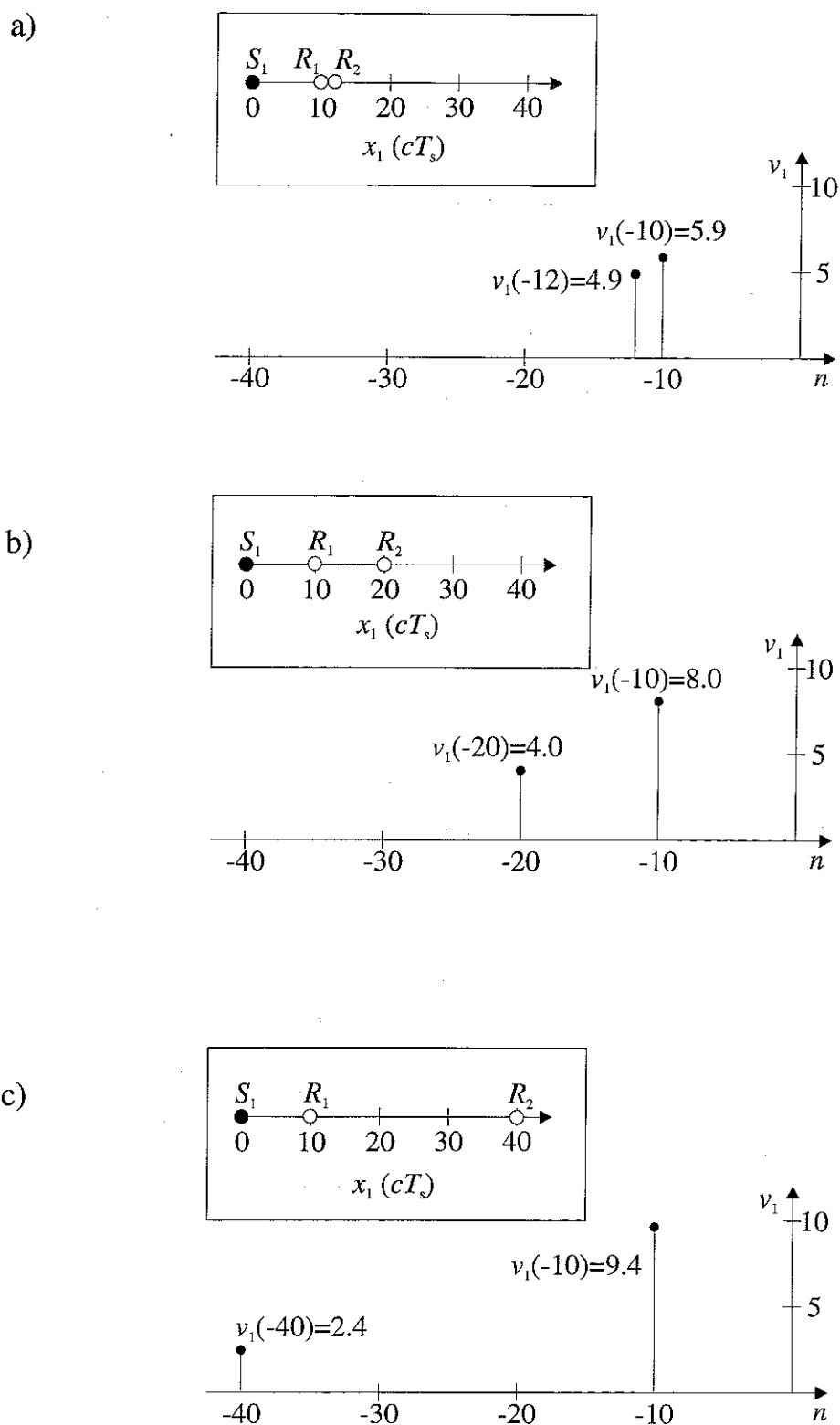


Figure 4.6. Three examples of the optimal output  $v_1(n)$  from a single monopole source when there are two microphones,  $R_1$  and  $R_2$ . Note how the first pulse becomes weaker as  $R_2$  is moved away from the source

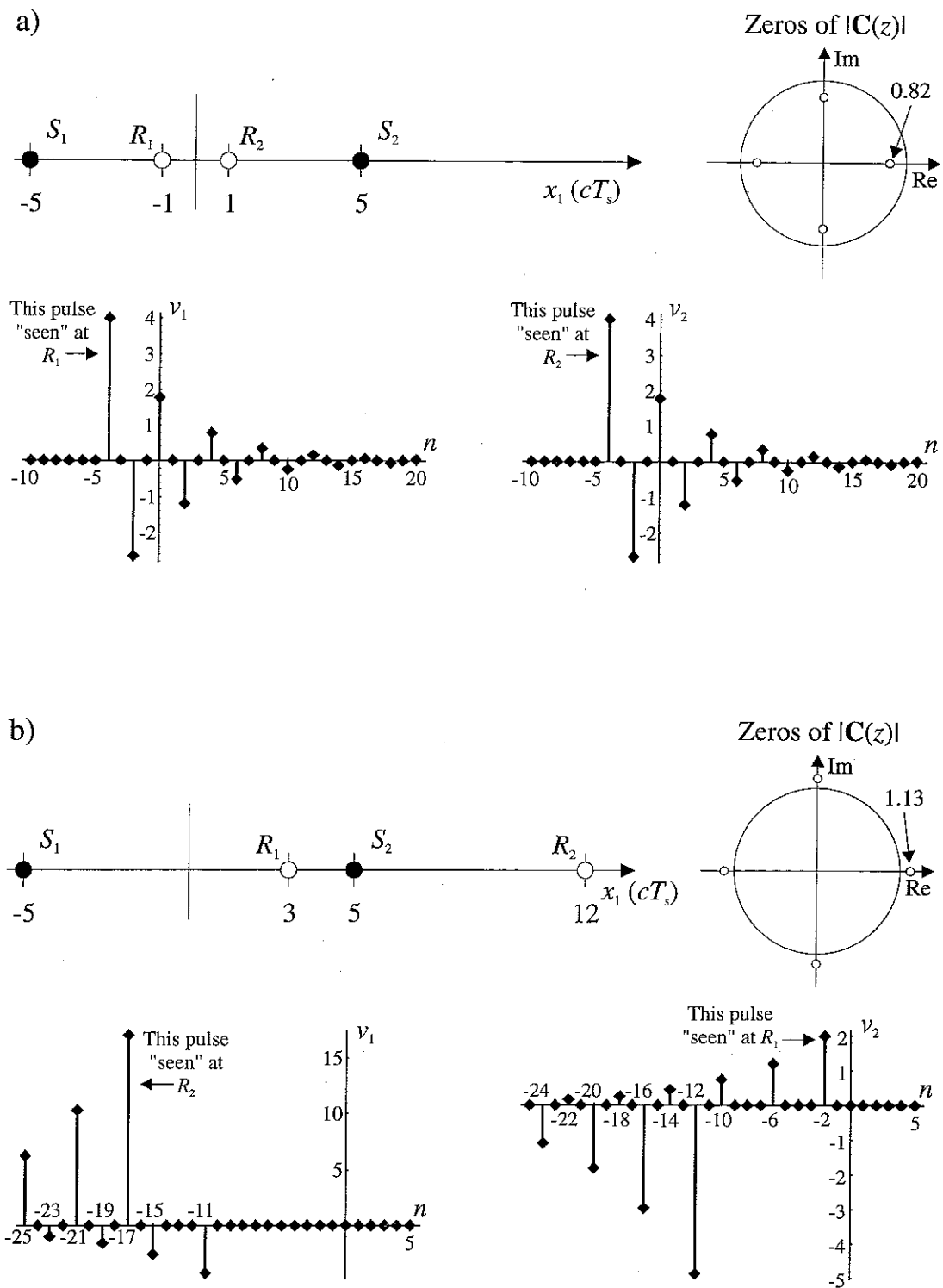


Figure 4.7. An example of a 2-by-2 system whose inverse is a) stable in forward time and b) unstable in forward time. Consequently, the stable inverse of b) must be exponentially decaying in backward time. The inputs  $v_1$  and  $v_2$  to the two loudspeakers are calculated by passing the recorded signals  $\mathbf{u}(z)=[1 \ 1]^T$  through the inverse filter matrix

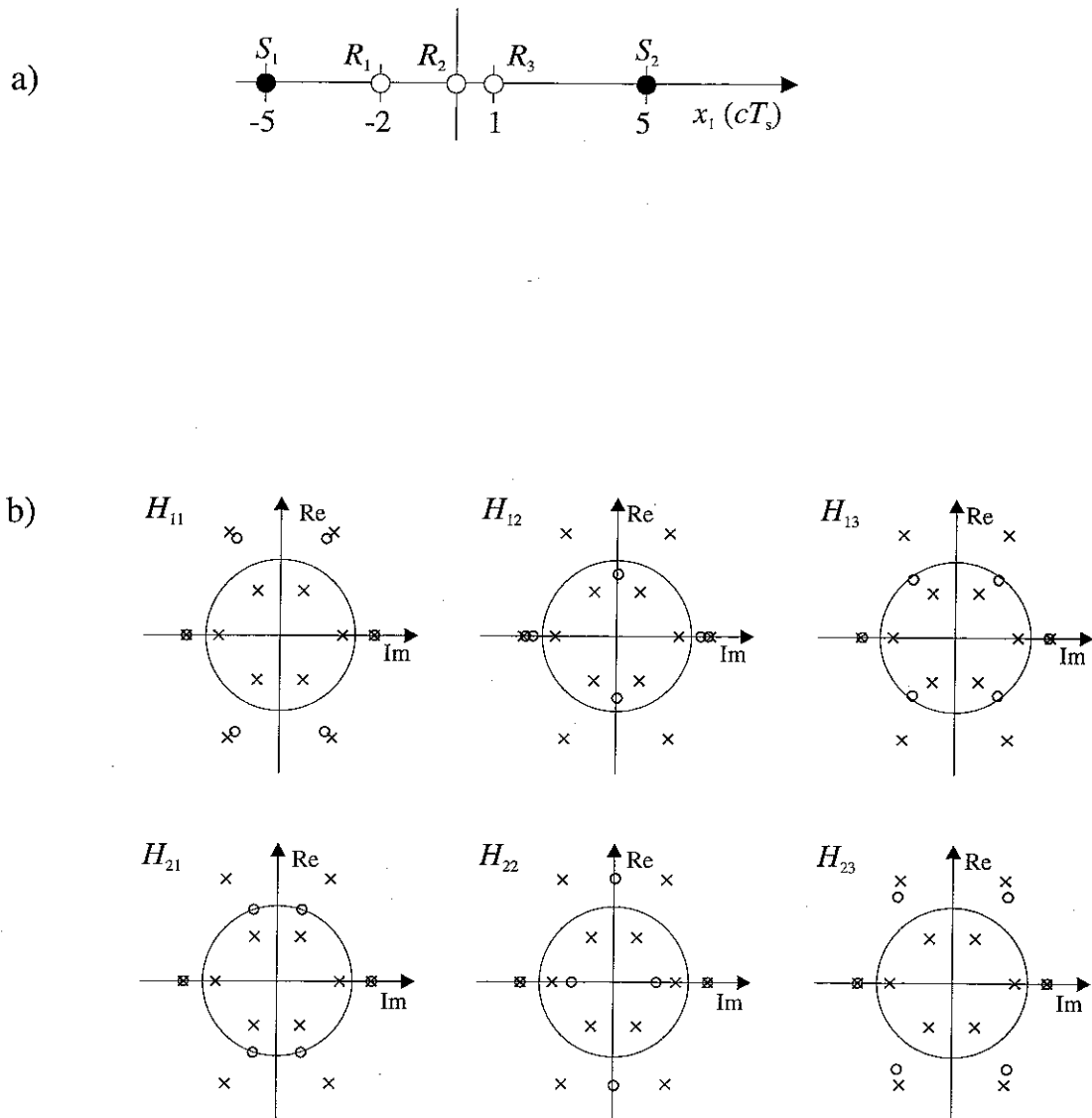


Figure 4.8. a) An example of a 3-by-2 system and b) the pole-zero maps of its exact least squares inverse. Note the symmetry of the poles in the unit circle

## 5. Time domain analysis

The exact least squares inverse filters are now approximated by causal finite impulse response (FIR) filters. The mathematical formulation of the inversion problem using FIR filters is considerably more complicated than the equivalent inversion problem formulated in the frequency domain and in the  $z$ -domain. However, the approach is fundamentally the same; the inverse FIR filters, denoted by  $h_{st}(n)$ , are determined by inverting a usually very large matrix  $\mathbf{C}$  that contains the electro-acoustic transfer functions. When  $\mathbf{C}$  is extremely large, it is advantageous to be able to calculate the inverse filters without building  $\mathbf{C}$  explicitly. This can be done by using adaptive methods. The definitions of the recorded signals  $\mathbf{u}$ , the loudspeaker inputs  $\mathbf{v}$ , and the reproduced signals  $\mathbf{w}$  are essentially the same as in Chapters 3 and 4. However, in order to make the performance of the FIR filters comparable to that of the exact least squares inverse filters, it is crucial to introduce a modelling delay. The modelling delay is implemented by a target matrix  $\mathbf{A}$ , which specifies the desired signals  $\mathbf{d}$  to be simple delayed copies of the recorded signals  $\mathbf{u}$ . The use of a modelling delay serves two purposes. One is to account for the initial delay, which is the time it takes for the sound to propagate from the loudspeaker to the microphone, the other is to allow the inverse filter to include in its response a part of the non-causal components of the exact least squares inverse (Widrow and Stearns [11] Chapter 10). Here,  $\mathbf{A}$  is used only for implementing the modelling delay, but it is included in the model as a general matrix of FIR filters. With the introduction of  $\mathbf{A}$ , another variable,  $T$ , is introduced.  $T$  is the number of recorded signals, or "tracks", and  $\mathbf{A}$  generally specifies the desired signals  $\mathbf{d}$  in terms of the recorded signals  $\mathbf{u}$ .

### 5.1 Inversion of single-channel systems

A modelling delay  $m$  defines the desired signal  $d(n)$  to be a copy of the input signal  $u(n)$  delayed by  $m$  samples. This is illustrated in the blockdiagram of the system shown in Figure 5.1. The system function  $H(z)$  of the ideal inverse filter is then given by

$$H(z) = z^{-m} \frac{1}{C(z)} \quad (5.1.1)$$

which shows that the net effect on the impulse response  $h(n)$  of the filter is a delay of  $m$  samples. For example, consider a single-channel system whose impulse response contains only two non-zero elements,  $c(5) = 1$ , and  $c(6) = -0.8$ . This sequence is meant to be a crude approximation to an electro-acoustic transfer function. The initial delay of five samples indicates the finite distance from the loudspeaker to the microphone, and the two non-zero coefficients indicate the "ringing" of the loudspeaker. The  $z$ -transform of  $c(n)$  is  $C(z)$ ,

$$C(z) = z^{-5} - 0.8z^{-6}, \quad (5.1.2)$$

which has one zero at  $z = 0.8$ . With no modelling delay ( $m = 0$ ), the exact inverse is stable and right-sided, but non-causal, given by

$$h(n) = z^5 \sum_{n=0}^{\infty} (0.8)^n z^{-n}, \quad (5.1.3)$$

the first non-zero element of this sequence occurs at time  $n = -5$ . With a modelling delay  $m = 5$ ,  $h(n)$  is shifted five samples towards the right, so

$$h(n) = \sum_{n=0}^{\infty} (0.8)^n z^{-n}, \quad (5.1.4)$$

which is stable, right-sided, and causal. However, if  $C(z)$  has zeros outside the unit circle, it is non-minimum phase and it is therefore not possible to make the ideal inverse filter  $h(n)$  causal by the use of a modelling delay. For example, if  $c(n)$  is the sequence shown in Figure 5.2a, it has the three non-zero elements,  $c(3) = -1$ ,  $c(4) = 2.05$ , and  $c(5) = -1$ , then  $C(z)$  is given by

$$C(z) = -z^{-3} + 2.05z^{-4} - z^{-5}. \quad (5.1.5)$$

There are two zeros of  $C(z)$  away from the origin and infinity, both are on the real axis at the conjugate reciprocal positions  $z = 0.8$  and  $z = 1.25$ . The ideal inverse  $H(z)$  of  $C(z)$  is found, after partial fraction expansion and rearranging terms, to be

$$H(z) = \frac{1}{C(z)} = \frac{-z^3}{1 - 2.05z^{-1} + z^{-2}} = \frac{20}{9} z^3 \left[ \frac{0.8}{1 - 0.8z^{-1}} + \frac{z}{1 - 0.8z} \right]. \quad (5.1.6)$$

The stable sequence  $h(n)$  with the  $z$ -transform  $H(z)$  is given by



$$h(n) = \frac{20}{9} \begin{cases} 1.25^{4+n} & \text{for } n \leq -4 \\ 0.8 \cdot 0.8^{3+n} & \text{for } n \geq -3 \end{cases} \quad (5.1.7)$$

this sequence is shown in Figure 5.2b. It is seen that the ideal inverse cannot be made causal by delaying  $h(n)$  by a finite amount.

Now assume that both  $h(n)$  and  $c(n)$  are causal finite length sequences of length  $N_h$  and  $N_c$  respectively. The convolution of  $c(n)$  with  $h(n)$  is then a causal finite length sequence of length  $N_h + N_c - 1$ . This convolution can be written as a product of the matrix  $\mathbf{C}$  and the column vector  $\mathbf{h}$ ,

$$c(n) * h(n) = \mathbf{C} \cdot \mathbf{h}, \quad (5.1.8)$$

when  $\mathbf{C}$  and  $\mathbf{h}$  are defined in the following way. The coefficients of the inverse filters are contained in  $\mathbf{h}$ , the element at the top is  $h(0)$ , the element at the bottom is  $h(N_h-1)$ ,

$$\mathbf{h} = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(N_h-1) \end{bmatrix}. \quad (5.1.9)$$

The matrix  $\mathbf{C}$  is often referred to as a convolution matrix, its columns containing the coefficients of the electro-acoustic transfer function  $c(n)$ . The element in the upper left corner is  $c(0)$ , the element in the lower right corner is  $c(N_c-1)$ ,

$$\mathbf{C} = \begin{bmatrix} c(0) & & & & \\ c(1) & c(0) & & & \\ \vdots & c(1) & \ddots & & \\ c(N_c-1) & \vdots & \ddots & c(0) & \\ & c(N_c-1) & \ddots & c(1) & \\ & & \ddots & \vdots & \\ & & & c(N_c-1) & \end{bmatrix}. \quad (5.1.10)$$

The time index increases as one moves down a column. Each column is shifted one position down relative to its left neighbour, the matrix therefore has a Toeplitz structure which means that elements are identical along diagonals (Press et al [71] p.92). Note that  $\mathbf{C}$  is not square, its number of rows  $N_{\text{rows}}$  is

$$N_{\text{rows}} = N_h + N_c - 1, \quad (5.1.11)$$

and its number of columns  $N_{\text{cols}}$  is

$$N_{\text{cols}} = N_h, \quad (5.1.12)$$

so apart from the trivial case where  $N_c$  is one, the number of rows is always greater than the number of columns. However, since the columns are clearly linearly independent,  $\mathbf{C}$  is guaranteed to have maximum rank,  $N_h$ . The number of non-zero elements  $N_{\text{nz}}$  in  $\mathbf{C}$  is at most  $N_c$  times  $N_h$ ,

$$N_{\text{nz}} = N_c N_h, \quad (5.1.13)$$

so if  $N_h$  is much greater than  $N_c$ , then  $\mathbf{C}$  is sparse. This can be quantified by defining a "density"  $\rho_c$  as the ratio between the number of non-zero elements and the total number of elements. When  $\rho_c$  is close to zero,  $\mathbf{C}$  contains almost exclusively zeros. When  $\rho_c$  is close to one,  $\mathbf{C}$  contains only very few zeros. For the single-channel case,  $\rho_c$  is given by

$$\rho_c = \frac{N_c}{N_c + N_h - 1}. \quad (5.1.14)$$

The sequence given by  $\mathbf{C}$  times  $\mathbf{h}$  is identical to the reproduced signal  $\mathbf{w}$  when the input sequence  $u(n)$  is an impulse. The column vector  $\mathbf{w}$  is of the same form as  $\mathbf{h}$ , so

$$\mathbf{w} = \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(N_h + N_c - 2) \end{bmatrix}. \quad (5.1.15)$$

When the reproduced sequence  $\mathbf{w}$  is compared to a desired sequence  $\mathbf{d}$ , the difference between the two gives a measure of the performance error  $\mathbf{e}$ ,

$$\mathbf{d} - \mathbf{C}\mathbf{h} = \mathbf{e}. \quad (5.1.16)$$

Since  $\mathbf{C}$  has more rows than columns, the equation system is over-determined. The optimal inverse filter  $\mathbf{h}_0$  that minimises a cost function of the type given in Equation (4.3.13) is determined uniquely by multiplying the pseudo-inverse of  $\mathbf{C}$  and the desired signal  $\mathbf{d}$ ,

$$\mathbf{h}_0 = [\mathbf{C}^T \mathbf{C}]^{-1} \mathbf{C}^T \mathbf{d}. \quad (5.1.17)$$

Without the use of a modelling delay, the desired signal  $\mathbf{d}$  is the same as the input signal  $\mathbf{u}$ , a single impulse at time zero. If the modelling delay is  $m$  samples, the desired signal  $\mathbf{d}$  is a single impulse at time  $m$ . Consequently,  $\mathbf{h}_0$  is the  $m$ 'th column of the pseudo-inverse of  $\mathbf{C}$ . It is important to realise that, even though  $\mathbf{h}_0$  is calculated purely in the time domain, the error is also minimised in the frequency domain as a consequence of Parseval's theorem which is stated in Equation (4.3.11).

We now use an FIR filter to invert the system whose impulse response is shown in Figure 5.2a. The inverse filter  $h(n)$  is chosen to have eight coefficients, so  $N_h = 8$ . The electro-acoustic transfer function  $c(n)$  has only three non-zero elements, the last of which occurs at time 5, so  $N_c = 6$ , consequently  $\mathbf{C}$  written out in full is

$$\mathbf{C} = \begin{bmatrix} 0 & & & & & & & \\ 0 & 0 & & & & & & \\ 0 & 0 & 0 & & & & & \\ -1 & 0 & 0 & 0 & & & & \\ 2.05 & -1 & 0 & 0 & 0 & & & \\ -1 & 2.05 & -1 & 0 & 0 & 0 & & \\ & -1 & 2.05 & -1 & 0 & 0 & 0 & \\ & & -1 & 2.05 & -1 & 0 & 0 & 0 \\ & & & -1 & 2.05 & -1 & 0 & 0 \\ & & & & -1 & 2.05 & -1 & 0 \\ & & & & & -1 & 2.05 & -1 \\ & & & & & & -1 & 2.05 \\ & & & & & & & -1 \end{bmatrix}. \quad (5.1.18)$$

where only coefficients of  $c(n)$  are included. A row of zeros in  $\mathbf{C}$  does not affect the solution to the problem, because  $\mathbf{C}$  still has rank  $N_h$ , but it is nevertheless numerically inconvenient. For example, it causes the corresponding column in the pseudo-inverse of  $\mathbf{C}$  to contain all zeros, and it also causes  $\mathbf{C}$  to have a singular value of zero which makes it impossible to estimate the condition number of  $\mathbf{C}$ . Thus, it is usually advantageous to discard the rows of  $\mathbf{C}$  that contain all zeros because of the initial delay of  $c(n)$ . This suggests that  $N_c$  can be thought of as an "effective" duration of  $c(n)$  rather than the length of the whole sequence including the initial delay, so for the above example, the effective duration of  $c(n)$  is three. For a modelling delay  $m$  of seven,  $\mathbf{d}$  is given by

$$\mathbf{d} = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]^T. \quad (5.1.19)$$

The inverse filter  $\mathbf{h}_0$  is calculated according to Equation (5.1.17). The inverse filter  $\mathbf{h}_0$  is shown in Figure 5.3a. Note that it has the same structure as the ideal inverse shown in Figure 5.2, and, in particular, that it includes "a part of" the non-causal component of the ideal inverse. Figure 5.3b shows  $\mathbf{C}$  times  $\mathbf{h}_0$  which should ideally be a single impulse at time seven. It is seen that  $\mathbf{C}$  times  $\mathbf{h}_0$  is quite close to the desired signal considering that the length of the inverse filter is very short. Figure 5.3c shows the total performance cost  $E_0 = \mathbf{e}^T \mathbf{e}$  as a function of the modelling delay, and Figure 5.3d shows  $\mathbf{C}$  times  $\mathbf{h}_0$  for three choices of the modelling delay. When  $m$  is zero, one, or two, the error is unity and the inverse filter contains all zeros. As demonstrated in section 3.6, this is the worst case since there is no output from the inverse filter. This result indicates that the modelling delay must be at least as long as the initial delay inherent in the electro-acoustical transfer path, otherwise it is not possible to reduce the total error. The reason for this is that when  $m$  is less than three, any signal that is passed through  $\mathbf{C}$  is delayed too long to arrive at the microphone at the "right time",  $m$ . There is an efficient, and simple, rule of thumb for how to choose the modelling delay; take the initial delay and add half the filter length. In this example, this is  $3+8/2 = 7$ . The exact value of the modelling delay is not critical, and there is usually quite a wide range of modelling delays that will all work almost equally well.

Once the coefficients of the inverse filter are determined, the reproduced sequence  $\mathbf{w}$  can be calculated for any input sequence  $\mathbf{u}$  by the matrix-matrix-vector multiplication

$$\mathbf{w}_{N_c+N_h+N_u-2} = \mathbf{C}_{N_c+N_h+N_u-2, N_h+N_u-1} \mathbf{H}_{N_h+N_u-1, N_u} \mathbf{u}_{N_u}, \quad (5.1.20)$$

where the subscripts indicate the dimensions of the vectors and matrices.  $\mathbf{H}$  is a convolution matrix of the same form as  $\mathbf{C}$ . The column vector  $\mathbf{u}$  is of the same form as  $\mathbf{w}$ , and contains the  $N_u$  most recent samples of the input. Written out in full, Equation (5.1.20) becomes

$$\begin{bmatrix} w(0) \\ \vdots \\ w(N_c+N_h+N_u-3) \end{bmatrix} = \begin{bmatrix} c(0) & & \\ \vdots & \ddots & \\ c(N_c-1) & & \\ & \ddots & \\ & & c(N_c-1) \end{bmatrix} \begin{bmatrix} h(0) & & \\ \vdots & \ddots & \\ h(N_h-1) & & h(0) \\ & \ddots & \\ & & h(N_h-1) \end{bmatrix} \begin{bmatrix} u(0) \\ \vdots \\ u(N_u-1) \end{bmatrix}. \quad (5.1.21)$$

The desired sequence can be calculated for any input in a similar way to  $\mathbf{w}$ . The target matrix  $\mathbf{A}$  must be cast into the form of a convolution matrix whose columns contain the  $N_d$  coefficients of the impulse response from the input  $\mathbf{u}$  to the desired signal  $\mathbf{d}$ . By using the same notation as in Equation (5.1.20),  $\mathbf{d}$  can be written as

$$\mathbf{d}_{N_d} = \mathbf{A}_{N_d, N_u} \mathbf{u}_{N_u}. \quad (5.1.22)$$

Note that if the objective is to compare  $\mathbf{d}$  directly to  $\mathbf{w}$ , then  $N_d$  must be equal to the number of elements in  $\mathbf{w}$ ,  $N_u + N_c + N_h - 2$ . Even though it is notationally convenient to be able to express convolutions as matrix multiplications, it is a very inefficient way to convolve sequences in practice.

## 5.2 Inversion of multi-channel systems

The multi-channel sound reproduction problem is illustrated in block diagram form in Figure 5.4. A filter matrix is added to the system; it is the target matrix  $\mathbf{A}$  which specifies the  $R$  desired signals  $\mathbf{d}(z)$ , one for each microphone, in terms of the  $T$  recorded signals  $\mathbf{u}(z)$ , one for each track. We will use  $\mathbf{A}$  for implementing the modelling delay only, but  $\mathbf{A}$  can also be used for more sophisticated purposes, such as virtual source imaging (Kirkeby et al [50], Cooper and Bauck [80]) and compensation for head-related transfer functions (Orduna-Bustamente et al [41]). As in the previous section, the aim is to be able to calculate  $\mathbf{h}$  by minimising a cost function of the type  $\mathbf{e}^T \mathbf{e}$  in the least squares sense. It is convenient to define, and minimise, the error  $\mathbf{e}_t$  separately for each track rather than for the system as a whole. When the input to the system is a single impulse on track  $t$ ,  $R$  error sequences  $\mathbf{e}_{rt}$  are generated, one at each microphone. These error sequences are “stacked” on top of each other to form a composite error vector  $\mathbf{e}_t$ ,

$$\mathbf{e}_t = \begin{bmatrix} \mathbf{e}_{1t} \\ \mathbf{e}_{2t} \\ \vdots \\ \mathbf{e}_{Rt} \end{bmatrix}. \quad (5.2.1)$$

This error vector is the difference between the desired signals  $\mathbf{d}_t$  and the reproduced signals  $\mathbf{w}_t$ , so

$$\mathbf{e}_t = \mathbf{d}_t - \mathbf{w}_t, \quad (5.2.2)$$

where  $\mathbf{d}_t$  and  $\mathbf{w}_t$  are defined in a similar way to  $\mathbf{e}_t$ ,

$$\mathbf{d}_t = \begin{bmatrix} \mathbf{d}_{1t} \\ \mathbf{d}_{2t} \\ \vdots \\ \mathbf{d}_{Rt} \end{bmatrix}, \quad \mathbf{w}_t = \begin{bmatrix} \mathbf{w}_{1t} \\ \mathbf{w}_{2t} \\ \vdots \\ \mathbf{w}_{Rt} \end{bmatrix}. \quad (5.2.3)$$

In order to be able to calculate  $\mathbf{w}_t$ , it is necessary to define  $\mathbf{H}$  and  $\mathbf{C}$  as "composite" matrices. Both  $\mathbf{H}$  and  $\mathbf{C}$  contain a number of digital filters. In  $\mathbf{C}$ , each digital filter is a convolution matrix of the form given by Equation (5.1.10). Each convolution matrix  $\mathbf{C}_{rs}$  has a double index that indicates the row and column number of its position in its "parent" matrix. By using this notation, we can write  $\mathbf{C}$  as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \cdots & \mathbf{C}_{1S} \\ \vdots & \mathbf{C}_{rs} & \vdots \\ \mathbf{C}_{R1} & \cdots & \mathbf{C}_{RS} \end{bmatrix}. \quad (5.2.4)$$

Equivalently for  $\mathbf{H}$ , each digital filter is a column vector  $\mathbf{h}_{st}$  whose double index indicates its position in  $\mathbf{H}$ ,

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{11} & \cdots & \mathbf{h}_{1T} \\ \vdots & \mathbf{h}_{st} & \vdots \\ \mathbf{h}_{S1} & \cdots & \mathbf{h}_{ST} \end{bmatrix}. \quad (5.2.5)$$

The column  $t$  of  $\mathbf{H}$  corresponds to track  $t$ , so

$$\mathbf{h}_t = \begin{bmatrix} \mathbf{h}_{1t} \\ \mathbf{h}_{2t} \\ \vdots \\ \mathbf{h}_{St} \end{bmatrix}, \quad (5.2.6)$$

and  $\mathbf{w}_t$  is now calculated by the matrix-vector multiplication

$$\mathbf{w}_t = \mathbf{C}\mathbf{h}_t, \quad (5.2.7)$$

which means that the problem is now in the familiar form of a linear equation system. This formulation of the multi-channel sound reproduction problem was derived, in a slightly different form, independently by Nelson and Elliott [54] and Miyoshi and Kanada [51]. In Section 5.1 it was demonstrated that in the single-channel case an initial delay in the electro-acoustic transfer function causes  $\mathbf{C}$  to have rows of zeros.

This also occurs in the multi-channel case when there is an initial delay in all the transfer functions  $C_{r1}, C_{r2}, \dots, C_{rS}$  that make up one row of convolution matrices in  $\mathbf{C}$ . Such a row of convolution matrices contains the transfer functions from the  $r$ 'th microphone to all the loudspeakers, and it is the minimum initial delay of these  $S$  transfer functions that determines how many rows of zeros there are in the  $r$ 'th row of convolution matrices in  $\mathbf{C}$ . Consequently, it is necessary to define an effective duration  $N_{cr}$  for each microphone in order to calculate the number of "non-zero" rows in  $\mathbf{C}$ . The number of non-zero rows  $N_{\text{rows}}$  in  $\mathbf{C}$  is given by

$$N_{\text{rows}} = \sum_{r=1}^R (N_h + N_{cr} - 1) = R(N_h - 1) + \sum_{r=1}^R N_{cr}, \quad (5.2.8)$$

and the number of columns  $N_{\text{cols}}$  in  $\mathbf{C}$  is given by

$$N_{\text{cols}} = SN_h. \quad (5.2.9)$$

In the special case where all the  $N_{cr}$  have the same value  $N_c$ , the number of non-zero rows is

$$N_{\text{rows}} = R \cdot (N_h + N_c - 1), \quad (5.2.10)$$

and the density  $\rho_C$  is then the same as in the single-channel case: the value of  $\rho_C$  is given by Equation (5.1.14). Thus,  $\mathbf{C}$  generally has more rows than columns for all systems that comprise at least as many microphones as loudspeakers. All such systems lead to an over-determined least squares problem but, as opposed to the single-channel problem, it is possible that the columns of  $\mathbf{C}$  are linearly dependent. If there are more loudspeakers than microphones, it is possible for  $\mathbf{C}$  to have more columns than rows, provided that the number of filter coefficients  $N_h$  in each of the inverse filters is sufficiently large. When  $N_{\text{cols}}$  is equal to  $N_{\text{rows}}$ ,  $\mathbf{C}$  is a square matrix that can be inverted exactly provided it is not singular, a result first published by Miyoshi and Kanada [51] in 1988. When  $\mathbf{C}$  has more rows than columns, the least squares problem is over-determined as in the single-channel case. In the special case where all the  $N_{cr}$  have the same value  $N_c$ , the difference between the number of columns and rows in  $\mathbf{C}$  is given by

$$N_{\text{cols}} - N_{\text{rows}} = N_h(S - R) - R(N_c - 1). \quad (5.2.11)$$

The desired signal  $\mathbf{d}_t$  is the result of passing a single impulse on track  $t$  through the target matrix  $\mathbf{A}$  which is generally a matrix of the same form as  $\mathbf{H}$ . When each element of  $\mathbf{A}$  is used only to implement a modelling delay  $m_{rt}$ , its  $z$ -transform is of the form

$$\mathbf{A}(z) = \begin{bmatrix} z^{-m_{11}} & \dots & z^{-m_{1T}} \\ \vdots & z^{-m_{rt}} & \vdots \\ z^{-m_{R1}} & \dots & z^{-m_{RT}} \end{bmatrix}. \quad (5.2.12)$$

In the special case where the objective is to reproduce  $R$  recorded signals at the  $R$  microphones, it is usually sufficient to use only one modelling delay  $m$  that is common to all the elements along the main diagonal of  $\mathbf{A}$ . Therefore,

$$\mathbf{A}(z) = \mathbf{I} z^{-m}, \quad (5.2.13)$$

where  $\mathbf{I}$  is the identity matrix of order  $R$ .

A simple, and purely hypothetical, example illustrates the dimensions of the different matrices and vectors. There are two tracks, two sources, and two microphones, so  $T = S = R = 2$ . Each of the four electro-acoustic transfer functions  $c_{rs}(n)$  has an impulse response whose two coefficients take the values  $r$  and  $s$ , so for example  $c_{11}(1) = 1$ ,  $c_{11}(2) = 1$ , and  $c_{21}(1) = 2$  and  $c_{21}(2) = 1$ . The two inverse filters each contain three coefficients. The target matrix is of the "diagonal" type given by Equation (5.2.13) with a common modelling delay of two. For these choices of system parameters, the error  $\mathbf{e}_1$ , which corresponds to track one, is given by

$$\mathbf{e}_1 = \mathbf{d}_1 - \mathbf{C}\mathbf{h}_1 = \begin{bmatrix} \mathbf{d}_{11} \\ \mathbf{d}_{21} \end{bmatrix} - \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{h}_{11} \\ \mathbf{h}_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ - \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & & & 1 \\ 1 & 1 & & 2 \\ & 1 & 1 & 2 \\ & & 1 & 2 \\ 2 & & & 2 \\ & 1 & 2 & 2 \\ & & 1 & 2 \end{bmatrix} \begin{bmatrix} h_{11}(0) \\ h_{11}(1) \\ h_{11}(2) \\ h_{21}(0) \\ h_{21}(1) \\ h_{21}(2) \end{bmatrix}, \quad (5.2.14)$$

where zeros have been left out for clarity. For track number two, the single non-zero element of  $\mathbf{d}_2$  would be in row seven.



The expression for each column of  $\mathbf{H}$ ,  $\mathbf{h}_t$ , is derived using standard least squares theory. Formally,  $\mathbf{h}_t$  is calculated by pre-multiplying  $\mathbf{d}_t$  with the pseudo-inverse of  $\mathbf{C}$ ,  $\mathbf{C}_{\text{pinv}}$ . In practice, it is usually not desirable to calculate the pseudo-inverse of  $\mathbf{C}$  explicitly. However, if  $\mathbf{C}_{\text{pinv}}$  is available, then it is possible to calculate all the inverse filters in  $\mathbf{H}$  by a single matrix-matrix multiplication. A matrix  $\mathbf{D}$  is needed, formed by storing each of the desired signals  $\mathbf{d}_t$  in column number  $t$  of  $\mathbf{D}$ . The inverse filter matrix  $\mathbf{H}$  is then given by

$$\mathbf{H} = \mathbf{C}_{\text{pinv}} \mathbf{D}. \quad (5.2.15)$$

Once the coefficients in the inverse filters are known, it is possible to write down an expression for the output sequence  $\mathbf{w}$  caused by any input sequence  $\mathbf{u}$ . The coefficients of the inverse filters must be arranged in a convolution matrix  $\mathbf{H}$  of the same form as  $\mathbf{C}$  given by Equation (5.1.10). As in the single-channel case, the number of rows and columns in  $\mathbf{u}$ ,  $\mathbf{H}$ , and  $\mathbf{C}$  must match each other. The dimensions of  $\mathbf{u}$ ,  $\mathbf{H}$ , and  $\mathbf{C}$  depend on the number of tracks, microphones and loudspeakers in addition to the filter lengths  $N_h$ ,  $N_c$ , and the length of the recorded sequences,  $N_u$ . We will use the notation  $\mathbf{H}_{(r,c),(sr,sc)}$  to denote a filter matrix that consists of  $r$  times  $c$  convolution matrices that each have  $sr$  rows and  $sc$  columns ( $sr$  for "sub-rows",  $sc$  for "sub-columns"). The notation  $\mathbf{u}_{(r),(sr)}$  denotes a column vector that consists of  $r$  column vectors, each containing  $sr$  coefficients, stacked on top of each other. The resulting expression for  $\mathbf{w}$  is very similar to the single-channel expression given by Equation (5.1.20),

$$\mathbf{w}_{(R),(N_c+N_h+N_u-2)} = \mathbf{C}_{(R,S),(N_c+N_h+N_u-2,N_h+N_u-1)} \mathbf{H}_{(S,T),(N_h+N_u-1,N_u)} \mathbf{u}_{(T),(N_u)}, \quad (5.2.16)$$

the only difference being the additional indices that indicate the relevant numbers of tracks, microphones and loudspeakers. Equivalent to Equation (5.1.22), the expression for  $\mathbf{d}$  is

$$\mathbf{d}_{(R),(N_d)} = \mathbf{A}_{(R,T),(N_d,N_u)} \mathbf{u}_{(T),(N_u)}. \quad (5.2.17)$$

Equations (5.2.16) and (5.2.17) are notationally convenient but are very inefficient for convolving sequences in practice.

### 5.3 Direct inversion methods

Once the electro-acoustic transfer matrix  $\mathbf{C}$  is known, it is in principle straightforward to calculate  $\mathbf{H}$  using the standard least squares method. However, a direct least squares inversion of  $\mathbf{C}$  is complicated in practice for two reasons. The first is that  $\mathbf{C}$  is usually a very large matrix, typically of the order 1000-by-1000, but since  $\mathbf{C}$  is usually also sparse, it is not necessary to store and operate on the full matrix. The second is that  $\mathbf{C}$  is usually extremely ill-conditioned which makes its exact pseudo-inverse almost entirely useless. The undesirable effects of ill-conditioning are alleviated by the use of regularisation.

An example using the loudspeaker-microphone layout shown in Figure 5.5 is now used to demonstrate the properties of a "typical"  $\mathbf{C}$  matrix. The four loudspeakers are positioned in the "quadraphony" arrangement shown in Figure 3.13b, and the microphone arrangement is similar in shape to the circular 3 shown in Figure 3.20d. The difference between the two microphone arrays is that the microphone at the top on the  $x_2$ -axis is moved up slightly, and the radius of the circle is increased from 3cm to 10cm. When the sampling frequency is 34kHz and the speed of sound is 340m/s, the sound travels 1m in 100 sampling intervals. The distance matrix  $\mathbf{L}$  in samples is then

$$\mathbf{L} = \begin{bmatrix} 273 & 273 & 293 & 293 \\ 280 & 293 & 273 & 286 \\ 293 & 280 & 286 & 273 \end{bmatrix}, \quad (5.3.1)$$

where the values are rounded to the nearest integers. The smallest value in each row is the initial delay for each microphone. This delay must be compensated for in order to avoid rows of zeros in  $\mathbf{C}$ . Since the initial delay is 273 for each microphone, 273 rows of zeros can be removed from each row of convolution matrices in  $\mathbf{C}$  as explained in Section 5.2. Thus, the distance matrix  $\mathbf{L}$  can also be written

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 20 & 20 \\ 7 & 20 & 0 & 13 \\ 20 & 7 & 13 & 0 \end{bmatrix} + 273 \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \quad (5.3.2)$$

Since the maximum difference between the elements in each of the three rows of  $\mathbf{L}$  is 20, the three effective durations  $N_{c1}$ ,  $N_{c2}$ , and  $N_{c3}$  all have the same value,  $N_c = 21$ . This is the reason for modifying the circular 3 microphone array slightly. The difference between the number of columns and rows can now be calculated from Equation (5.2.11); the result is

$$N_{\text{cols}} - N_{\text{rows}} = N_h - 60. \quad (5.3.3)$$

Figure 5.6 shows the non-zero elements in  $\mathbf{C}$  and  $\mathbf{C}^T\mathbf{C}$  for the four different values 15, 30, 60, and 120 of the filter length  $N_h$  when  $\mathbf{C}$  is calculated using the single-coefficient approximation given by Equation (4.4.1). The number of non-zero elements  $N_{\text{nz}}$  in the matrices is indicated below each plot, and it is seen that  $\mathbf{C}^T\mathbf{C}$  contains more non-zero elements than  $\mathbf{C}$ . The difference between the number of non-zero elements of  $\mathbf{C}^T\mathbf{C}$  and  $\mathbf{C}$  is sometimes referred to as fill-ins (Press et al [71] Section 2.7), and as  $\mathbf{C}$  becomes larger, the “relative” amount of fill-ins increases. Note that each of the 12 convolution matrices in  $\mathbf{C}$  is identified by a diagonal line, and that  $\mathbf{C}^T\mathbf{C}$  is square and symmetric with non-zero elements along its diagonal. When  $N_h$  is 15, there are rows of zeros in the middle of the first row of convolution matrices in  $\mathbf{C}$  because the effective duration  $N_c$  is larger than the filter length  $N_h$ . This problem of “displaced” convolution matrices in the same row is more serious than it looks at first sight because the same phenomenon also occurs when there is a large difference between the initial delays of the transfer functions relating to one particular microphone. Thus, when two loudspeakers have different distances to the microphone array, their contributions to  $\mathbf{C}$  are bound to be displaced because the initial delay for each row is always given by the nearest loudspeaker only. This phenomenon is examined in some detail by Orduna-Bustamante et al [62]. There is no way to avoid this problem with the techniques described here. One possible solution, however, is to delay the input to the nearest loudspeakers such that the shortest initial delays are artificially increased (Gerzon [75]).

Now the properties of the  $\mathbf{C}$ -matrices shown in the left column of Figure 5.6 are investigated. Figure 5.7 shows the singular values of  $\mathbf{C}$  for the four values 15, 30, 60, 120 of  $N_h$  on a linear scale. Surprisingly, the four plots are very similar. The largest 80%-90% of the singular values are quite evenly spread within approximately one

decade, then they drop off to zero (within machine precision) very rapidly. An exact inversion of any of these  $\mathbf{C}$  matrices would be disastrous, and a regularisation factor must be used to avoid the effect of the small singular values on the solution. This suggests that for this particular geometry there is no particular reason to prefer any particular kind of least squares problem (over-determined, square, or under-determined), but it is intuitively obvious that the performance of the inverse filters is best when  $N_h$  is large. An interpretation in terms of vector spaces indicates that there are always some dimensions in the desired output space that cannot be reached by  $\mathbf{C}$  no matter how large  $N_h$  is. The singular values of zero, which indicate linear dependence between the columns of  $\mathbf{C}$ , are caused by the layout of the loudspeakers and microphones in this specific example. The "cause" of the linear dependence is not obvious, for example, if any two loudspeakers are removed, then  $\mathbf{C}$  has full rank. It is the combination of two pairs of loudspeakers that makes  $\mathbf{C}$  singular. This phenomenon is examined in more detail in Section 6.3.

## 5.4 Adaptive inversion methods

Adaptive filters are useful for many applications (Nelson and Elliott [9], Widrow and Stearns [11]), and they are useful for the active control of sound mainly for three reasons. The first is that an adaptive algorithm is relatively simple to implement, the second is that the algorithm does not need any information about the desired signal as long as the error signal is known, and the third is that the algorithm can "track" relatively slow changes in the system parameters in real-time. Adaptive filters are digital filters whose coefficients change with time depending on the input. By comparing the output of the filter with a desired output signal, the adaptive filter is able to adjust its filter coefficients in order to improve its performance. The vital information that is necessary for the filter to converge to the unknown optimal solution is supplied by the error signal. The error signal is crucial; without it, the filter has no way of assessing its own performance. In order to account for the dependence on time, the theory of adaptive filters is usually described in terms of statistical results rather than deterministic results. For example, instead of defining a deterministic input signal, a correlation matrix is used to indicate the average behaviour of a whole family of input signals.

The adaptive filter  $\mathbf{H}$  assesses its own performance by comparing its output to a desired signal  $\mathbf{d}$ . However, as illustrated in Figure 5.4, the output of  $\mathbf{H}$  is filtered by  $\mathbf{C}$ , the electro-acoustic transfer matrix, before it is compared to  $\mathbf{d}$ . There is a simple and straightforward way, in principle, to overcome this problem. The idea is to interchange the two filtering operations of  $\mathbf{H}$  and  $\mathbf{C}$ . If both  $\mathbf{H}$  and  $\mathbf{C}$  are linear, it does not make any difference whether  $\mathbf{H}$  is applied directly to the recorded signals  $\mathbf{u}$ , or to the acoustic signals at the microphone positions after the recorded signals have been played back through the reproduction system. The latter signals are called the “filtered reference signals”, and are denoted by the letter  $\mathbf{r}$ . The only difficulty with this formulation is a purely practical one. The objective is effectively to make the “new” matrix product  $\mathbf{CH}$  equal to the “old” matrix product  $\mathbf{HC}$  regardless of the dimensions of  $\mathbf{C}$  and  $\mathbf{H}$ , and this inevitably leads to quite a complicated mathematical formulation. It is possible to derive an expression for the error  $\mathbf{e}(n)$  of the type (Nelson and Elliott [9] Section 12.8)

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{w}(n) = \mathbf{d}(n) - \mathbf{R}(n) \mathbf{h}, \quad (5.4.1)$$

where  $\mathbf{R}(n)$  is a matrix containing all the filtered reference signals, and  $\mathbf{h}$  contains all the coefficients of the inverse filters (not just the coefficients relating to one track). In the statistical formulation, the optimal filter  $\mathbf{h}_0$  is defined as the filter that does the best job, on average, in the least squares sense. “On average” refers to the average over an ensemble of input sequences, or recordings, so  $\mathbf{h}_0$  depends on the average behaviour of  $\mathbf{u}$ . Usually,  $\mathbf{u}$  is assumed to be the outcome of a stationary stochastic process with zero mean and variance  $\sigma^2$ , and a white (constant as a function of frequency) spectrum (Widrow and Stearns [11] Chapter 10). The cost function  $J$  is defined as

$$J = E[\mathbf{e}^T(n) \mathbf{e}(n) + \beta \mathbf{v}^T(n) \mathbf{v}(n)]. \quad (5.4.2)$$

where  $E$  is an expectation operator (not performance cost) and  $\beta$  is a regularisation factor that penalises large outputs. The least squares solution  $\mathbf{h}_0$  that minimizes  $J$  is

$$\mathbf{h}_0 = \{E[\mathbf{R}^T(n) \mathbf{R}(n)] + \beta \sigma^2 \mathbf{I}\}^{-1} E[\mathbf{R}^T(n) \mathbf{d}(n)]. \quad (5.4.3)$$

One of the most common methods for updating the coefficients of  $\mathbf{h}$  is the steepest descent algorithm (Widrow and Stearns [11] p.56, Nelson and Elliott [9] Section 12.4). The name of the method is quite descriptive of the way it works: after each

new sample,  $\mathbf{h}$  is adjusted in such a way that it takes a step from its current position on the quadratic error surface down the locally steepest slope. The length of the step is proportional to the gradient, so the steeper the slope is, the longer is the step. The update equation for  $\mathbf{h}$  is

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \mu \frac{\partial J}{\partial \mathbf{h}}, \quad (5.4.4)$$

where  $\mu$  is a positive convergence factor, and the minus sign on the right hand side indicates that the step is in the opposite direction of the gradient. The value of  $\mu$  is crucial for the convergence properties of the adaptive algorithm; if  $\mu$  is too big, the algorithm is unstable and the error grows without bound, if  $\mu$  is too small, the algorithm converges extremely slowly because it takes a very large number of short steps to get to  $\mathbf{h}_0$ . The best way to see how quickly the algorithm converges is by plotting  $J$  as a function of the number of iterations. Such a plot is often referred to as a learning curve (see Widrow and Stearns [11] p.108, for a typical example). The learning curve is ideally a sum of exponentials, each decaying with its own time constant, and  $\mu$  is effectively a common divider of all these time constants. So, the larger  $\mu$  is, the quicker the solution converges. There are useful general guidelines for how to choose a sensible initial value for  $\mu$  (Widrow and Stearns [11] p.103), but in practice it is always necessary to make trial-and-error experiments. The gradient as a function of  $\mathbf{h}$  is given by

$$\frac{\partial J}{\partial \mathbf{h}} = 2E[\mathbf{R}^T(n)\mathbf{R}(n)\mathbf{h} - \mathbf{R}^T(n)\mathbf{d}(n)] + \sigma^2\beta\mathbf{h}, \quad (5.4.5)$$

which means that the desired signal,  $\mathbf{d}(n)$ , has to be known. However, Equation (5.4.5) can also be expressed as

$$\frac{\partial J}{\partial \mathbf{h}} = \sigma^2\beta\mathbf{h} - 2E[\mathbf{R}^T(n)\mathbf{e}(n)], \quad (5.4.6)$$

which shows that if the error signal,  $\mathbf{e}(n)$ , is known, it is not necessary to know  $\mathbf{d}(n)$ . Equation (5.4.6) also shows that the gradient's exact value depends on the average behaviour of  $\mathbf{e}(n)$ , and not just its instantaneous value. This means that it is not possible to update  $\mathbf{h}$  after each sample without having some information about the statistical properties of the error signal (Widrow and Stearns [11]). One way to avoid

this problem is simply to remove the expectation operator from Equation (5.4.6). The update equation for  $\mathbf{h}$  then becomes

$$\mathbf{h}(n+1) = (1 - \mu\sigma^2\beta)\mathbf{h}(n) + 2\mu\mathbf{R}^T(n)\mathbf{e}(n), \quad (5.4.7)$$

or, after simplifying the notation by setting  $1 - \mu\sigma^2\beta = \gamma$  and  $2\mu = \alpha$ ,

$$\mathbf{h}(n+1) = \gamma\mathbf{h}(n) + \alpha\mathbf{R}^T(n)\mathbf{e}(n). \quad (5.4.8)$$

This update equation uses an estimate of the gradient instead of the exact value as given by Equation (5.4.6). Note how the regularisation factor enters the update equation. At each iteration, the  $\beta$ -term makes  $\mathbf{h}$  take a step back towards the origin before it continues down the local gradient. The larger  $\mathbf{h}(n)$  is, the longer is the step backwards. This is how  $\mathbf{h}$  is prevented from becoming very large. If  $\alpha$  is suddenly set to zero,  $\mathbf{h}$  decays exponentially to zero, so  $\gamma$  is effectively the time constant of the regularisation.

The steepest descent algorithm has its own “built-in” regularisation. Since the change made to  $\mathbf{h}$  is proportional to the gradient of the error surface, it is clear that the closer the cost function  $J$  is to its minimum  $J_0$ , the smaller are the changes made to  $\mathbf{h}$  at each update. If the problem is ill-conditioned, then the error surface has very long valleys, so even though  $J$  is very close to  $J_0$ ,  $\mathbf{h}$  can be very far away from its optimal value  $\mathbf{h}_0$ . The steepest descent algorithm is very efficient for ill-conditioned problems because the local gradient guides  $\mathbf{h}$  towards the nearest point on the valley floor rather than towards  $\mathbf{h}_0$ . As  $J$  approaches  $J_0$ ,  $\mathbf{h}$  approaches  $\mathbf{h}_0$  very slowly. For example, it might take 100 iterations to make the normalised cost  $(J - J_0)/J_0$  less than 1%, but it would take another 10,000 iterations to make  $(J - J_0)/J_0$  less than 0.25%. At the same time, the length of  $\mathbf{h}$  might increase by a factor of 10, and still be a factor of 1000 smaller than  $\mathbf{h}_0$ . In practice, it is not possible to obtain an exact solution to an ill-conditioned problem with an adaptive algorithm; instead the algorithm is stopped when  $J$  no longer decreases significantly and the solution “looks good”. Therefore, the algorithm can be stopped when the learning curve is flat. When the learning curve is flat, it is impossible to tell whether the solution has actually converged to the exact solution or it has just found a point close to the valley floor, but in either case there is no point in continuing the adaptation. Figure 5.8 shows the nine learning curves for the system

illustrated in Figure 5.5. The adaptive algorithm is a steepest descent algorithm, without regularisation ( $\gamma = 1$ ), trained with a deterministic impulse rather than random white noise (this algorithm is developed by Orduna-Bustamante and described in [36]). There are nine learning curves  $J_r$  in total, each one relating to the contribution to  $J$  from track  $t$  at microphone  $r$ . Each of the twelve adaptive filters contains 60 coefficients, and the modelling delay is 30. It is seen that the learning curves for the three tracks are qualitatively similar. Since the adaptive filters are initialised with zeros, there is no output from the filters at first. Therefore, the cost is large for the elements corresponding to the diagonal elements in the target matrix, and small for the elements corresponding to the off-diagonal elements in the target matrix. As the filters start to adapt for a given track, the distance between the individual learning curves becomes smaller, and after about ten iterations (impulse inputs), they are almost flat. The value of the total cost  $J$  after 20 iterations agrees well with results obtained by direct inversion of  $\mathbf{C}$  using a regularisation factor of 0.000001. For track number one, direct inversion gives a value of 0.1362 for  $J$ , adaptive inversion gives a value of 0.1391 for  $J$  (found as the sum of  $J_{11}$ ,  $J_{21}$ , and  $J_{31}$ ). For track number two and three, direct inversion gives a value of 0.2295 for  $J$ , adaptive inversion gives a value of 0.2329 for  $J$  (found as the sum of  $J_{12}$ ,  $J_{22}$ , and  $J_{32}$ ). The two sets of inverse filters are almost exactly identical, and this indicates that the steepest descent method's "built-in" regularisation works just as well as the corresponding regularisation of direct inversion problems provided the adaptive algorithm is given enough time to converge.

## 5.5 Conclusions

In order to be able to deconvolve a system using inverse filters  $h_{st}(n)$  that have finite impulse responses, a modelling delay  $m$ , which specifies the desired signals  $\mathbf{d}$  to be simple delayed copies of the recorded signals  $\mathbf{u}$ , is crucial. The filter coefficients in  $h_{st}(n)$  can be calculated either by a least squares inversion of a large matrix  $\mathbf{C}$ , which contains the finite impulse responses of the electro-acoustic transfer functions, or by using an adaptive algorithm that uses a steepest descent method to approximate the exact least squares solution.  $\mathbf{C}$  usually has more rows than columns, but when there are more loudspeakers than microphones, it is possible to invert  $\mathbf{C}$  exactly provided



the number of filter coefficients in the inverse filters is sufficiently large. However, an exact inverse does not seem to offer any obvious advantages over a close-to-exact least squares solution. In fact, if  $\mathbf{C}$  happens to be square and is inverted exactly without regularisation, the solution is likely to be very large because the inversion problem is usually ill-conditioned. Since the condition number of  $\mathbf{C}$ , which is the ratio between its largest and smallest singular values, is not a very good measure for how ill-conditioned the inversion problem is, it is necessary to look at the distribution of all the singular values of  $\mathbf{C}$ . If many of these are very small compared to the largest singular value, then  $\mathbf{C}$  is very singular and extremely difficult to invert.

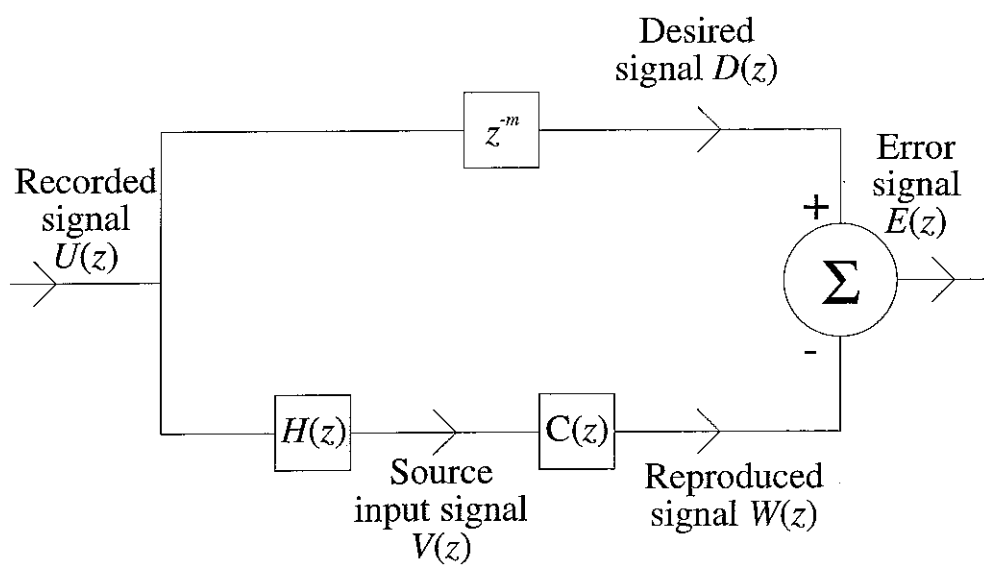
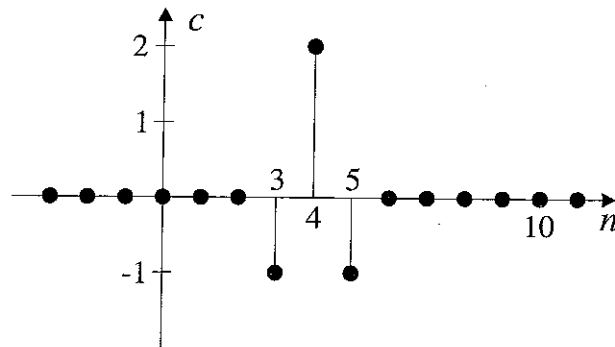


Figure 5.1. The single-channel inversion problem in block diagram form including a modelling delay  $m$

a)



b)

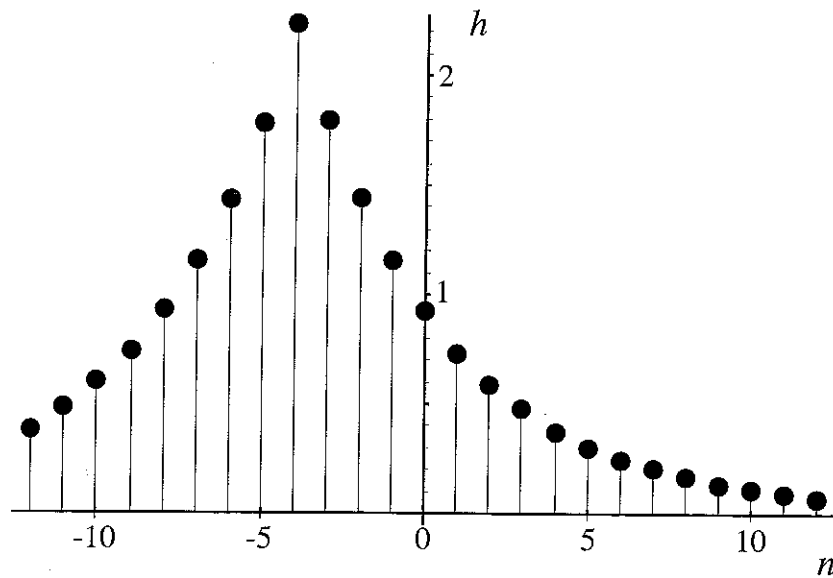


Figure 5.2. a) An example of a system which has zeros both inside and outside the unit circle. Consequently, its exact stable inverse, which is shown in b), is double-sided

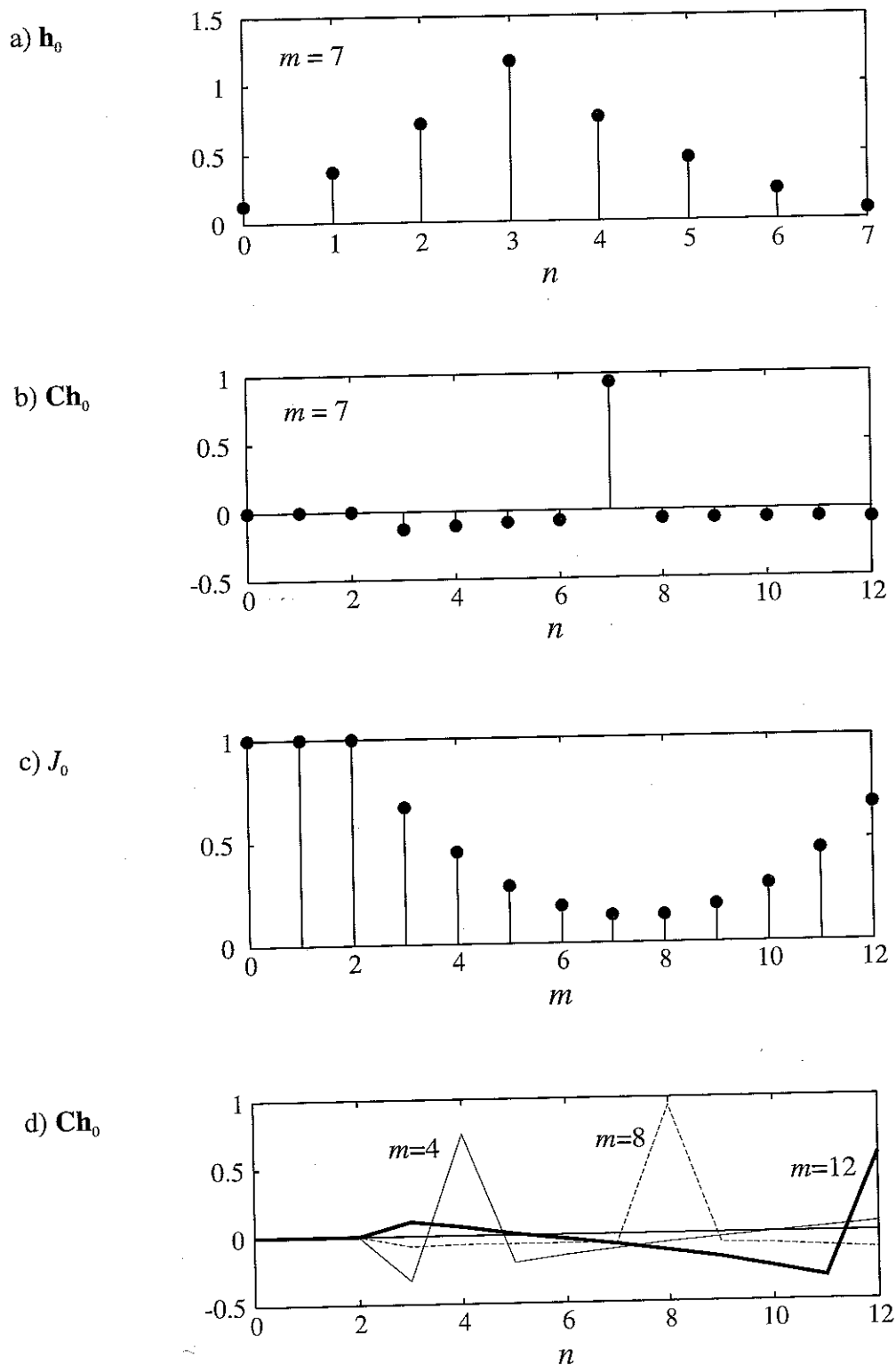


Figure 5.3. The impulse response of a) the eight-coefficient inverse filter  $h_0$ , calculated using a modelling delay  $m$  of 7 and b)  $Ch_0$  (the inverse filter  $h_0$  convolved with c). c) shows the cost function  $J_0$  as a function of  $m$  and d) shows  $Ch_0$  for three different choices of  $m$

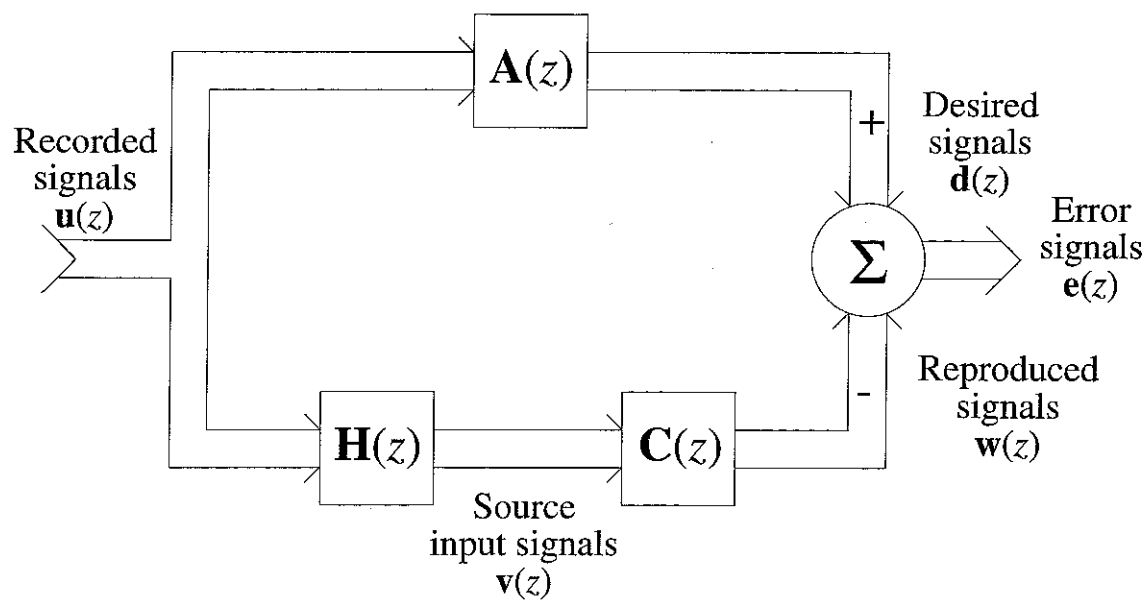


Figure 5.4. The discrete-time multi-channel sound reproduction problem in block diagram form including a general target matrix  $\mathbf{A}$

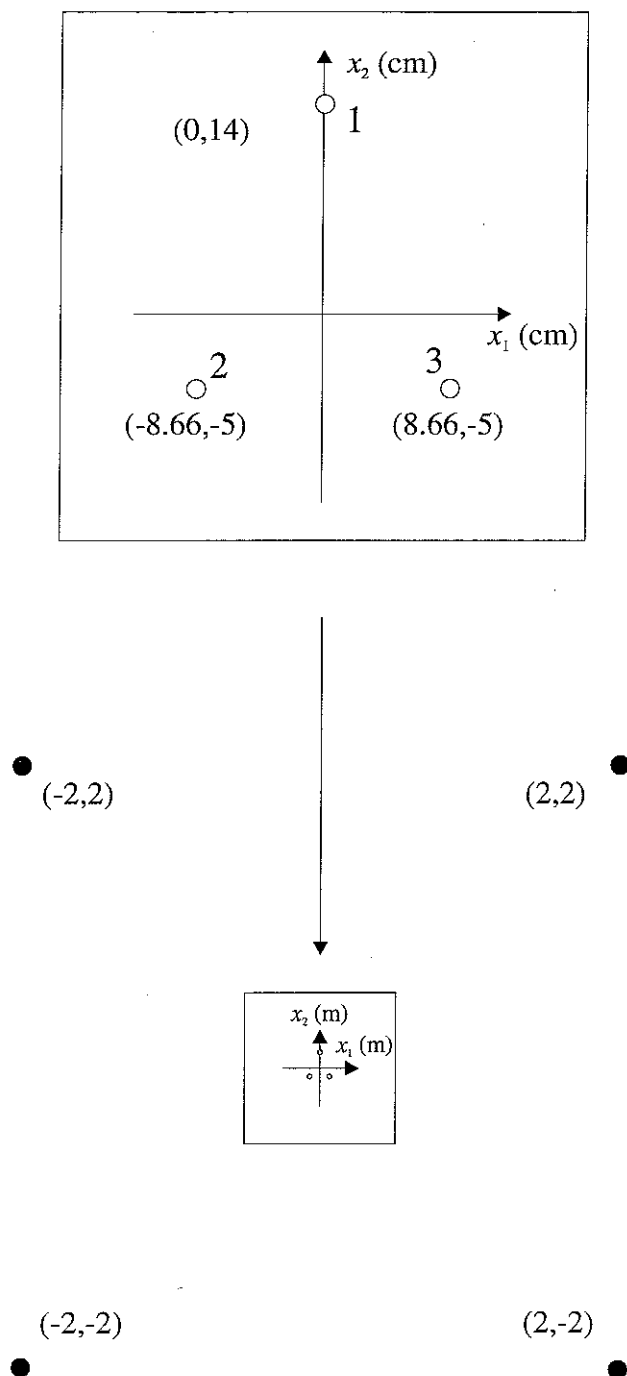


Figure 5.5. Four sources positioned in a "quadraphony" arrangement and a microphone array which is similar in shape to the circular 3 shown in Figure 3.20d

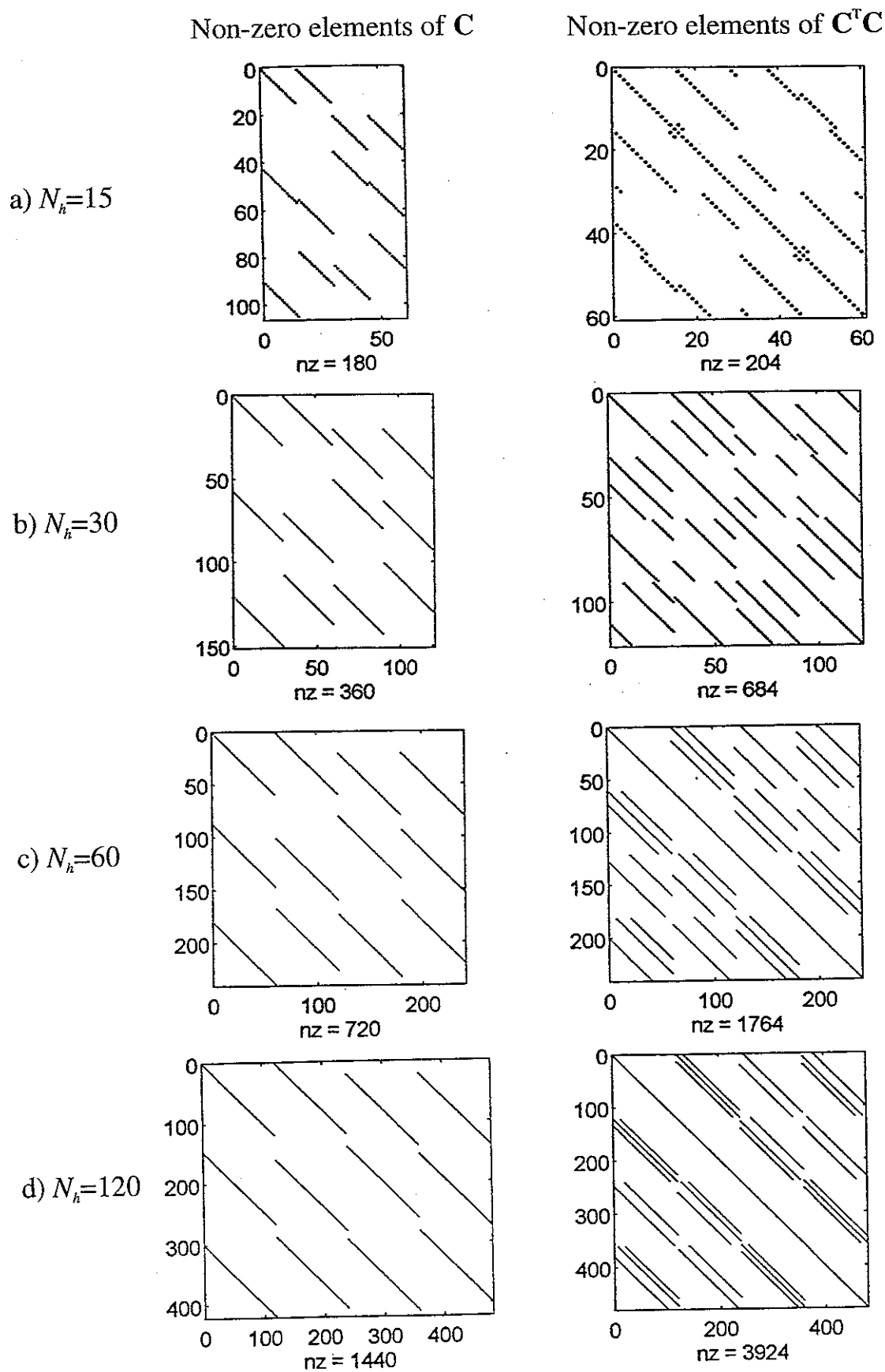


Figure 5.6. An illustration of the  $N_{nz}$  non-zero elements in  $\mathbf{C}$  and  $\mathbf{C}^T \mathbf{C}$  for the four different values 15, 30, 60, and 120 of the filter length  $N_h$ .

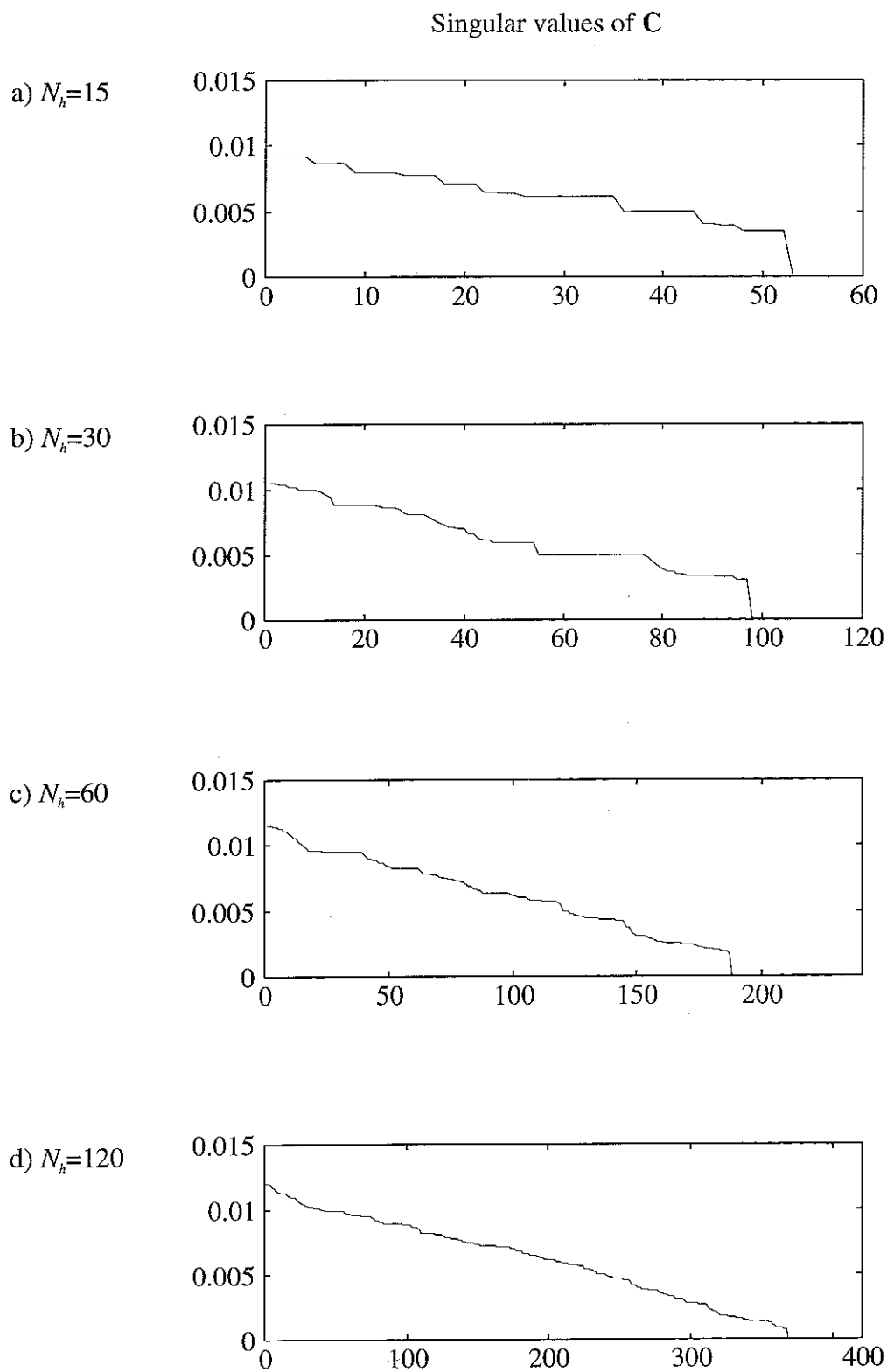


Figure 5.7. The singular values of the four  $\mathbf{C}$ -matrices shown in the left column of Figure 5.6



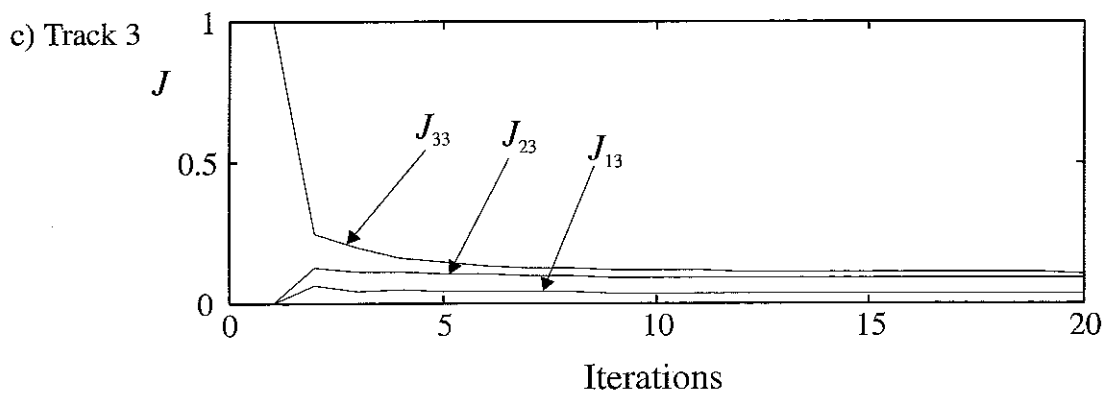
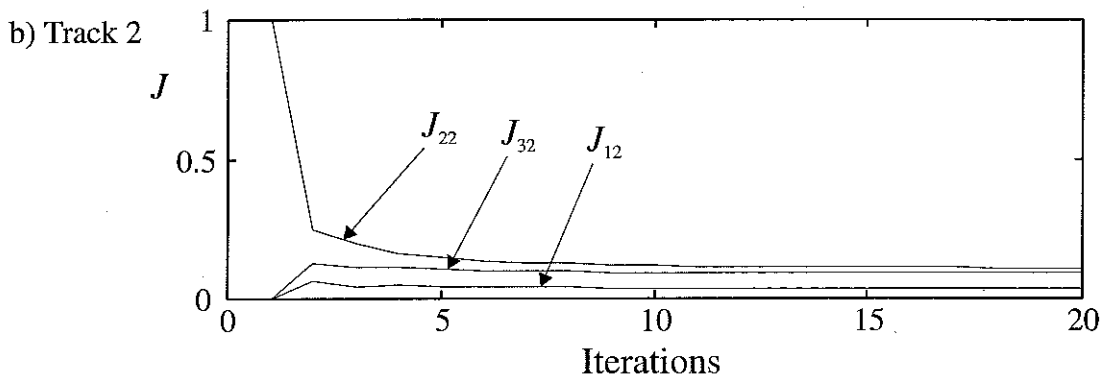
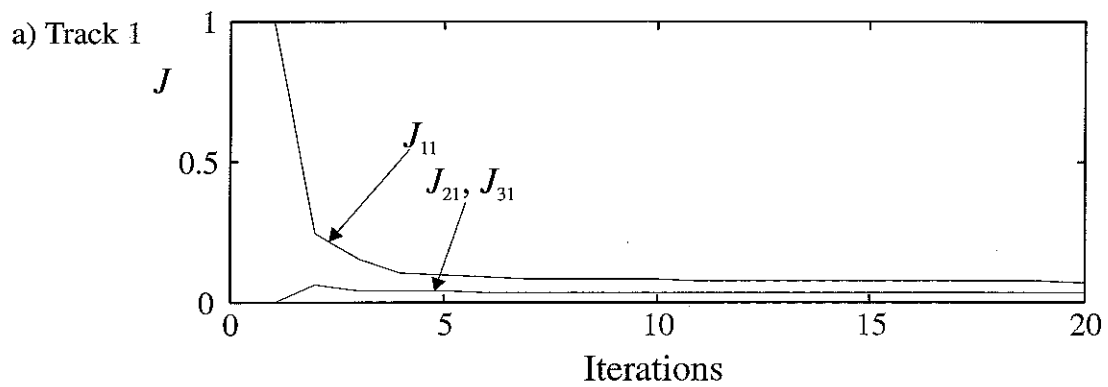


Figure 5.8. Learning curves for a steepest descent adaptation of the system shown in Figure 5.5. The length  $N_h$  of each of the inverse filters is 60, the modelling delay is 30

## **6. Transient response of multi-channel sound reproduction systems**

The theory developed in the previous chapters is now used to model the transient response of some simple multi-channel systems working in discrete time. In order to make the model realistic, it is necessary to consider how to approximate an electro-acoustic transfer function given only the distance between the source and the receiver. If the time it takes for the sound to travel that distance is not an integer number of sampling intervals, a suitable interpolation scheme can be used to obtain a more accurate approximation. Some examples illustrate how the properties of the inverse filters depend on the electro-acoustic transfer function, the distance between adjacent microphones, and the sampling frequency. One has to be careful about generalising the results from these examples, however, because even very subtle changes made to a system's parameters can make the behaviour of that system change quite dramatically. Certain "symmetrical" layouts of microphones and loudspeakers lead to extremely ill-conditioned inversion problems, which means that in practice they are impossible to invert even with an extremely large number of coefficients in the inverse filters. The transient response of a sound reproduction system is best illustrated by using an input that has a short duration. However, the duration of the input cannot be too short, otherwise spatial aliasing occurs. It is convenient to use a "Hanning pulse", which is a digital Hanning window, to define a desired "plane wave transient" sound field, because the Hanning pulse looks like a "smooth" pulse, and by varying its duration, its bandwidth can be controlled. When the desired field is recorded and passed through the inverse filter matrix, the input to each of the sources is the sum of the outputs from several inverse filters. For this reason, the source inputs can be qualitatively different from the individual outputs from the inverse filters. By looking at the spatially reproduced sound field, it is possible to get a physical feel for the size of the area over which the sound can be controlled.

## 6.1 Modelling the electro-acoustic transfer function

In the previous chapters, the continuous-time transfer function from a monopole source to a microphone position is approximated in the discrete-time  $z$ -domain by

$$C(z) = \frac{z^{-l}}{l}, \quad (6.1.1)$$

where  $l$  is the number of sampling intervals it takes for the sound to travel from the loudspeaker to the microphone (given earlier in Equation (4.4.1), repeated here for convenience without the indices  $r$  and  $s$ ). If the exact value of  $l$  is not an integer  $l_{\text{int}}$ , then it has a non-zero fractional part  $l_{\text{frac}}$ , so

$$l = l_{\text{int}} + l_{\text{frac}}. \quad (6.1.2)$$

For example, if  $l = 11.3$ , then  $l_{\text{int}} = 11$  and  $l_{\text{frac}} = 0.3$ . Since Equation (6.1.1) requires that  $l$  is an integer, a non-zero value of  $l_{\text{frac}}$  will cause an approximation error. The error as a function of frequency caused by an approximation  $C_{\text{approx}}(e^{j2\pi f/f_s})$  to the exact electro-acoustic transfer function  $C_{\text{exact}}(e^{j2\pi f/f_s})$  is split into a “distance-normalised” amplitude error

$$E_{\text{amp}}(e^{j2\pi f/f_s}) = l \cdot \left| C_{\text{exact}}(e^{j2\pi f/f_s}) - C_{\text{approx}}(e^{j2\pi f/f_s}) \right| \quad (6.1.3)$$

and a phase error

$$E_{\text{phase}}(e^{j2\pi f/f_s}) = \left| \angle C_{\text{exact}}(e^{j2\pi f/f_s}) - \angle C_{\text{approx}}(e^{j2\pi f/f_s}) \right| \quad (6.1.4)$$

Since the quantisation of  $l$  is equivalent to a quantisation of the distance between the source and the microphone, the approximation can be improved by increasing the sampling frequency. An example of this technique, referred to as over-sampling, is given in Figure 6.1. The integer part  $l_{\text{int}}$  of  $l$  is removed since the phase error depends only on  $l_{\text{frac}}$ . The value of  $l_{\text{frac}}$  determines the slope of the exact phase response, which is linear. When  $l$  is rounded to the nearest integer, the phase response of the approximation is also linear, consequently the phase error is a linearly increasing function of frequency. As seen in Figure 6.1a, when the sampling frequency is  $f_1$ , the phase error is largest when  $l_{\text{frac}}$  is 0.5, in that case  $E_{\text{phase}}$  is 90 degrees at  $f_{\text{Nyq}}$ . An increase of the sampling frequency  $f_1$  by a factor of 10 to  $f_{10}$  effectively increases the

phase resolution of the approximation by a factor of 10. This is illustrated in Figure 6.1b.

If the system is going to be capable of using the differences between the arrival times at the different microphones of a sound wave, it must take the sound several sampling intervals, for example 10, to travel from any one microphone to its nearest neighbour. However, in order to avoid the undesirable effect of spatial aliasing as described in Section 3.6, the acoustic wavelength corresponding to  $f_{\max}$ , the highest frequency component in the recorded signals, must be at least three times longer than the distance between adjacent microphones. Consequently, the sampling frequency  $f_s$  must be a factor of 30 higher than  $f_{\max}$ . So, if the microphone spacing is 10 cm, then the sampling frequency must be greater than 34kHz, but  $f_{\max}$  is only about 1kHz. Thus, without affecting frequencies lower than  $f_{\max}$ , it is possible to lowpass filter the signal and then to decimate it in order to obtain a shorter digital filter that runs at a much lower sampling rate. For example, if one of the optimal filters contains 200 coefficients and runs at 34kHz, then after lowpass filtering and decimation by a factor of 10, the new filter runs at 3.4kHz and contains only 20 coefficients. This "over-sampling, lowpass-filtering, decimation" - scheme is simple to implement, and has been used for high-resolution beam-steering (Pridham and Mucci [6], Dudgeon [46]). The principle is illustrated qualitatively in the frequency domain in Figure 6.2, and in the time domain in Figure 6.3.

The transfer function approximation problem is essentially an interpolation problem and it can be approached in different ways (see for example Schafer and Rabiner [2] or Pridham and Mucci [32]). One such approach is to use multirate digital signal processing. The over-sampling technique described above is a special case of a more general method for implementing a fractional delay in a discrete time system (Crochiere and Rabiner [66] Section 6.3), and the resulting filter is referred to as a "fractional sample phase shifter". The basic idea is to approximate the real number  $l_{\text{frac}}$  by the fraction  $N/D$  where the numerator  $N$  and the denominator  $D$  are positive integers. A delay by  $N/D$  of  $c(n)$  is implemented in three steps. First,  $c(n)$  is interpolated by a factor of  $D$ , this is the opposite operation of decimation, it effectively increases the sampling frequency from  $f_s$  to  $D$  times  $f_s$  by "filling in the gaps" between the samples of the original signal. Secondly, the new signal is shifted

by  $N$  samples to the right. Finally, the signal is lowpass filtered and decimated by a factor of  $D$ , this changes the sampling frequency back to  $f_s$ . Figure 6.4 illustrates the three steps in the time domain for a delay of 0.4. In practice, errors are caused by the interpolation and the decimation because a lowpass-filtering is included in both operations (see Schafer and Rabiner [2] for details). Figure 6.5 shows a comparison of the quantization error  $e_{\text{quant}}$  as a function of the fractional delay  $l_{\text{frac}}$  for three different approximation schemes, a) rounding to the nearest integer, b) rounding to the nearest integer after increasing the sampling frequency by a factor of 10, and c) rounding to the nearest fraction  $N/D$  where the maximum value of  $D$  is 10. In all three cases, the phase error is a linear function of frequency as illustrated in Figure 6.1, and the maximum phase error, which occurs at the Nyquist frequency  $f_{\text{Nyq}}$ , is  $e_{\text{quant}}$  times  $180^\circ$  (this assumes that the lowpass-filters are ideal). It is seen that  $l_{\text{frac}}$  is overall well approximated by a fraction apart from near zero and one where the error is identical to that of rounding after increasing the sampling frequency by a factor of 10.

Another way to model a fractional delay is to approximate the transfer function with a digital FIR filter containing two coefficients. When the two coefficients are found by linear interpolation, the  $z$ -transform of the filter becomes

$$c(l, z) = \frac{1}{l} \left[ (1 - l_{\text{frac}}) z^{-l_{\text{int}}} + l_{\text{frac}} z^{-(l_{\text{int}}+1)} \right]. \quad (6.1.5)$$

where  $l$  is the exact delay. For example, if  $l = 5.4$ , then  $c(5) = 0.6/5.4$  and  $c(6) = 0.4/5.4$ . This simple approximation is remarkably efficient because it is very accurate at low frequencies. The use of linear interpolation for the purpose of approximating an electro-acoustic transfer function was originally an idea of Orduna-Bustamante [36]. The amplitude error  $E_{\text{amp}}$  and the phase error  $E_{\text{phase}}$  as a function of the normalised frequency and the fractional delay are shown in Figures 6.6 and 6.7 respectively. For frequencies lower than  $0.5f_{\text{Nyq}}$ , the phase error is roughly equal to the phase error obtained by using the more sophisticated fractional sample phase shifter. The linear interpolation is easy to implement, and it only doubles the number of elements in the  $\mathbf{C}$ -matrix, therefore the sparse structure of  $\mathbf{C}$ , which is illustrated in Figure 5.6, is preserved.

## 6.2 General properties of the inverse filters

Some simple examples are now presented in order to give a feel for how the inverse filters depend on three system parameters: the distance between adjacent microphones, the sampling frequency, and the type of electro-acoustic transfer function (rounding or linear interpolation).

The system that is investigated comprises four sources in a  $4\text{m} \times 4\text{m}$  square, and four microphones in a square whose sidelength is either 10cm or 5cm. Both the sources and the microphones are positioned in the corners of a square, both squares are positioned symmetrically around the origin of the coordinate system. There are 16 elements in the transfer function matrix  $\mathbf{C}$ , each of them depend only on the distance from the source to the microphone, but due to the symmetry of the geometrical arrangement of the sources and the microphones, only three of the sixteen distances are different. These three distances, measured in sampling intervals, are denoted by  $l_1$ ,  $l_2$ , and  $l_3$ , where  $l_1 < l_2 < l_3$ . The corresponding three acoustic transfer functions are denoted  $C_1(z)$ ,  $C_2(z)$ , and  $C_3(z)$  so

$$\mathbf{C}(z) = \begin{bmatrix} C_1(z) & C_2(z) & C_2(z) & C_3(z) \\ C_2(z) & C_1(z) & C_3(z) & C_2(z) \\ C_2(z) & C_3(z) & C_1(z) & C_2(z) \\ C_3(z) & C_2(z) & C_2(z) & C_1(z) \end{bmatrix}. \quad (6.2.1)$$

The exact inverse of this matrix also contains only three different elements, they are denoted  $H_1(z)$ ,  $H_2(z)$ , and  $H_3(z)$ , consequently

$$\mathbf{C}^{-1}(z) = \mathbf{H}(z) = \begin{bmatrix} H_1(z) & H_2(z) & H_2(z) & H_3(z) \\ H_2(z) & H_1(z) & H_3(z) & H_2(z) \\ H_2(z) & H_3(z) & H_1(z) & H_2(z) \\ H_3(z) & H_2(z) & H_2(z) & H_1(z) \end{bmatrix}. \quad (6.2.2)$$

By dividing the adjoint of  $\mathbf{C}(z)$  by its determinant  $D(z)$  we find

$$H_1 = (C_1^3 - 2C_1C_2^2 + 2C_2^2C_3 - C_1C_3^2) / D, \quad (6.2.3)$$

$$H_2 = (-C_1^2C_2 + 2C_1C_2C_3 - C_2C_3^2) / D, \quad (6.2.4)$$

and

$$H_3 = (2C_1C_2^2 - C_1^2C_3 - 2C_2^2C_3 + C_3^2) / D, \quad (6.2.5)$$

where

$$D = C_1^4 - 4C_1^2C_2^2 + 8C_1C_2^2C_3 - 2C_1^2C_3^2 - 4C_2^2C_3^2 + C_3^4. \quad (6.2.6)$$

In the Equations (6.2.3-6), the variable  $z$  has been omitted for notational convenience. It is clear that, because of the symmetry, this geometry is not representative of a large class of systems. On the other hand, the behaviour of the system is determined by only a few system parameters, this makes it relatively easy to relate certain properties of the solution to certain system parameters. When the sampling frequency is either 34kHz or 12kHz, there are eight different sets of transfer functions, one for each sampling frequency, microphone spacing and type of electro-acoustic transfer function. Table 6.1. lists the three integer approximations  $l_{\text{round}}$  that are used to approximate  $C(z)$  by Equation (6.1.1), and the three values of  $l_{\text{int}}$  and  $l_{\text{frac}}$  that are used to approximate  $C(z)$  by Equation (6.1.5).

Sampling-frequency	Microphone-array	$l_{\text{round}}$	$l_{\text{int}}$	$l_{\text{frac}}$
34kHz	10cm×10cm	276	275	0.77
		283	282	0.93
		290	289	0.91
	5cm×5cm	279	279	0.31
		283	282	0.86
		286	286	0.38
12kHz	10cm×10cm	97	97	0.35
		100	99	0.87
		102	102	0.34
	5cm×5cm	99	98	0.60
		100	99	0.85
		101	101	0.09

Table 6.1.

Figure 6.8 shows the positions in the complex  $z$ -plane of the poles of the exact inverse filters (exact in the frequency domain least squares sense) of each of the eight systems derived from the data in table 6.1. The set of poles are the zeros of the determinant of  $C(z)$  as given by Equation (6.2.6). The number of poles increases with both the sampling frequency  $f_s$  and the spacing between adjacent microphones, because the

difference between the maximum and minimum values of  $l$  increases with  $f_s$  and the microphone spacing. When that difference is large, the order of the determinant polynomial, which is calculated by multiplying and adding together elements from  $C(z)$ , is also large (see also Section 4.3). The number of poles therefore becomes larger when the spatial resolution is increased. This phenomenon is essentially caused by spatial aliasing of the elements of  $C(z)$ . For example, in Figure 6.8a there are 56 poles which appear in 14 groups of four, in Figure 6.8c there are 28 poles which appear in seven less clearly defined groups of four. In Figure 6.8a, the maximum value of the difference between the maximum and minimum values of  $l$  is 14 samples, in Figure 6.8c it is 7 samples. The regular pattern in Figure 6.8a appears because the difference between the rounded values of  $l_1$  and  $l_2$  is the same as the difference between the rounded values of  $l_2$  and  $l_3$ . This is not the case when the microphone spacing is reduced by a factor of two (see Table 6.1), and the poles in Figure 6.8c are therefore more "spread out". Clearly, a large number of poles makes it difficult to approximate the exact inverse filters with FIR filters, in particular when the poles are close to the unit circle. There are always poles of the exact inverse close to the unit circle on the real axis, which indicate the ill-conditioning of the inversion problem at low frequencies. Since the approximation of the electro-acoustic transfer function is accurate only at relatively low frequencies, there is no point in trying to minimise the error above, say, half the Nyquist frequency. Unfortunately, as shown in Section 4.3, the least squares method minimises the integral of the error around the unit circle. This means that a significant fraction of the coefficients in the inverse filters are inevitably used for minimising the error at frequencies above the upper working frequency limit of the system. When the acoustic wavelength at the sampling frequency is comparable to the spacing between adjacent microphones, there are relatively few poles in the exact inverse, and they tend to be scattered over mainly the left hand side of the complex plane as shown in Figure 6.8h. This is highly desirable, and it means that the FIR approximation to the exact inverse is likely to be very accurate, particularly at low frequencies (apart from very close to 0Hz). The approximation to the electro-acoustic transfer function by rounding the distance in sampling intervals is not only less accurate than the linear interpolation technique, it also tends to make the poles appear in clusters, which is undesirable.



Figure 6.9 shows the frequency response of an FIR approximation to one of the exact inverse filters,  $H_1(z)$ , for each of the eight systems whose poles are shown in Figure 6.8. The FIR filters  $H_1(z)$ ,  $H_2(z)$ , and  $H_3(z)$  each contain 120 coefficients, and the modelling delay is 60. Note how the peaks in the frequency response correspond well to the pole positions in Figure 6.8. Figure 6.10 shows the frequency responses, using the FIR approximations to the exact inverse filters of each of the eight systems whose poles are shown in Figure 6.8, of two of the signals that are reproduced at the microphones. One is  $|W_{11}|$ , the frequency response from track number one to microphone number one, which is ideally 1 (0dB), the other is  $|W_{12}|$ , the frequency response from track number two to microphone number one, which is ideally 0 (minus infinity dB). The performance of the filters is overall very good. The agreement with the desired signals is worst near a pole close to the unit circle, in particular at 0Hz, at those frequencies there is a dip in  $|W_{11}|$ . The FIR approximations to the exact inverse filters are all calculated without regularisation, the reason being that this particular geometry always leads to well-conditioned least squares problems.

### 6.3 Loudspeaker-microphone layouts and conditioning

The conditioning of the inversion problem seems to depend more strongly on the layout of loudspeakers and microphones than any other system parameter. The geometry investigated in the previous section leads to well-conditioned inversion problems for almost any choice of system parameters. However, if the loudspeaker array (or the microphone array) is rotated 45 degrees, any choice of system parameters leads to a singular, and not just ill-conditioned, inversion problem. The electro-acoustic transfer matrix  $C(z)$  for such a system is of the type

$$C(z) = \begin{bmatrix} C_1(z) & C_1(z) & C_2(z) & C_2(z) \\ C_1(z) & C_2(z) & C_2(z) & C_1(z) \\ C_2(z) & C_1(z) & C_1(z) & C_2(z) \\ C_2(z) & C_2(z) & C_1(z) & C_1(z) \end{bmatrix} \quad (6.3.1)$$

assuming that each element in  $C(z)$  depends only on the distance from the loudspeaker to the microphone. The matrix  $C(z)$  contains only two different elements, and its determinant is zero for all values of  $z$ . Many other geometries that are much less trivial also lead to highly singular inversion problems (Swanson [107]). Even though

there is no direct way to find out whether the  $\mathbf{C}$ -matrix used for the design of the inverse FIR filters is highly singular, there is a way to find out whether a geometry leads to a singular  $\mathbf{C}(z)$ -matrix. At the frequency 0Hz,  $z$  is equal to one, so both the rounding approximation and the linear interpolation technique give each of the elements in  $\mathbf{C}(z)$  a real value of  $1/l_{rs}$ , where  $l_{rs}$  is the distance from loudspeaker  $s$  to microphone  $r$ . Thus, if the "inverse distances matrix", which is the matrix containing the values  $1/l_{rs}$ , does not have full rank, then  $\mathbf{C}(z)$  cannot be inverted exactly because the frequency response at 0Hz is infinite. Consequently, the  $\mathbf{C}$ -matrix that is generated in order to calculate the FIR approximation to the exact inverse is bound to have many zero singular values. However, even when the exact inverse is well-behaved it is still possible for the  $\mathbf{C}$ -matrix to have many zero singular values. This phenomenon occurred with the geometry used for the example in Section 5.3. Thus, a certain geometry usually leads to either a well-conditioned or ill-conditioned  $\mathbf{C}$ -matrix for all choices of sampling frequency, microphone spacing, and filter length. An ill-conditioned  $\mathbf{C}$ -matrix can be caused by either a singular "inverse distances matrix", in which case  $\mathbf{C}(z)$  is singular and the inversion of  $\mathbf{C}$  is bound to be virtually impossible, or, seemingly, by the block Toeplitz structure of  $\mathbf{C}$ , in which case a "good" inversion of  $\mathbf{C}$  might be possible. Figure 6.11 lists some geometries and the different types of inversion problems they can lead to.

Since the conditioning of the inversion problem depends on the layout of the microphones and the loudspeakers in a seemingly unpredictable way, it is not possible to design a good universal microphone array containing only three or four microphones. However, Figure 6.11 suggests that a combination of two square microphone arrays, each containing four microphones, could work for most loudspeaker setups. Such a microphone array is shown in Figure 6.12. This microphone array seems to lead to well-conditioned least squares problems for all conceivable loudspeaker setups (at least as long as the number of loudspeakers is not greater than eight), and it has a very good resolution with respect to different angles of incidence. The acoustic wavelength at the sampling frequency should be approximately equal to the sidelength of the "outer" square, this ensures that there are not too many poles in the exact inverse and that the system is not too ill-conditioned at low frequencies.

One final point needs to be made about the conditioning of inversion problems. Two types of ill-conditioning can occur. One is caused by poles near the unit circle in the exact least squares inverse. This leads to very long impulse responses of the inverse filters, so without the use of regularisation, many values of the coefficients in the FIR approximations are different from zero even if the inverse filters contain many coefficients. The second type of ill-conditioning is caused by linearly dependent columns in the **C**-matrix that determines the optimal values of the coefficients in the FIR approximations. There is no obvious reason why the two types of ill-conditioning should be related because they occur in different domains. One is in a complex "pole-space", the other is in a finite-dimensional coefficient space. The pole map of the exact inverse is the most important indicator of ill-conditioning because it is related to the physics of the problem. The conditioning of the **C**-matrix relates to the optimal values of the coefficients in the FIR filters, and this is relative only to the available filter coefficients, not directly to the physical problem that generated the equations.

#### 6.4 The Hanning pulse

Before considering the outputs from the inverse filters, and eventually the reproduced sound field, it is necessary to find a sequence that is suitable for modelling a recorded transient discrete-time signal. A Hanning pulse, which is a digital Hanning window, seems to be a good choice not only because it "looks" like a transient, but also because its bandwidth can be controlled by varying its duration. Figure 6.13 shows four digital Hanning pulses and their corresponding zero-maps. The lengths of the pulses are 5, 10, 20, and 50 samples. The zeros of the Hanning pulse are all on the unit circle, the high frequencies are therefore attenuated very efficiently. Almost all of the energy is concentrated in the main lobe near DC, and the longer the pulse is, the narrower is its main lobe. Figure 6.14 shows the frequency spectrum of each of the four pulses, the positions of the first zero as a fraction of the sampling frequency  $f_s$  are 0.3333, 0.1818, 0.0952, and 0.0392 respectively.

#### 6.5 Properties of the source inputs

The output from a loudspeaker is the sum of  $T$  convolutions, one for each track. In the  $z$ -domain, the output  $V_s$  from source  $s$  is given by

$$V_s(z) = \begin{bmatrix} H_{s1}(z) & \cdots & H_{sT}(z) \end{bmatrix} \cdot \begin{bmatrix} U_1(z) \\ \vdots \\ U_T(z) \end{bmatrix}. \quad (6.4.1)$$

Thus, the set of recorded signals  $\mathbf{u}(z)$  are first filtered by the elements of one row of  $\mathbf{H}(z)$ , then they are added together. If the properties of both  $H_{st}(z)$  and  $U_i(z)$  are known, it is straightforward to derive the properties of the convolution of the two. If the  $z$ -transforms of two signals are given by  $N_1(z)/D_1(z)$  and  $N_2(z)/D_2(z)$  respectively, then the  $z$ -transform of their product  $N_{\text{prod}}(z)/D_{\text{prod}}(z)$  is given by

$$\frac{N_{\text{prod}}(z)}{D_{\text{prod}}(z)} = \frac{N_1(z)N_2(z)}{D_1(z)D_2(z)}, \quad (6.4.2)$$

which shows that the poles and zeros of  $H_{st}(z)$  and  $U_i(z)$  are also the poles and zeros of  $H_{st}(z)*U_i(z)$  (unless a pole and a zero cancel out exactly). An equivalent result does not exist for the sum of two signals. The  $z$ -transform of the sum  $N_{\text{sum}}(z)/D_{\text{sum}}(z)$  of two signals is given by

$$\frac{N_{\text{sum}}(z)}{D_{\text{sum}}(z)} = \frac{N_1(z)}{D_1(z)} + \frac{N_2(z)}{D_2(z)} = \frac{N_1(z)D_2(z) + N_2(z)D_1(z)}{D_1(z)D_2(z)}, \quad (6.4.3)$$

which shows that the zeros of the sum are generally not directly related to the zeros of either of the two signals, not even when they are both finite length sequences ( $D_1(z) = D_2(z) = 1$ ). In the special case where all the recorded signals are simple delayed copies of each other, the output  $V_s(z)$  from source  $s$  is given by

$$V_s(z) = U(z) \left[ z^{-n_1} H_{s1}(z) + z^{-n_2} H_{s2}(z) + \dots + z^{-n_T} H_{sT}(z) \right]. \quad (6.4.4)$$

provided the delays  $n_1, n_2, \dots, n_T$  between the tracks are all integers. The “reference” for the recorded signals,  $U(z)$ , acts as an overall filter, so if  $U(z)$  is a long Hanning pulse, it effectively performs a very efficient lowpass-filtering of the elements of  $\mathbf{H}(z)$ .

The following example illustrates that the poles in the inverse filters can be cancelled almost exactly by combining the recorded signals with the inverse filters according to Equation (6.4.4). Consider the geometry studied earlier: four speakers are positioned in a  $4\text{m} \times 4\text{m}$  square, and four microphones are positioned in a  $10\text{cm} \times 10\text{cm}$  square. When the sampling frequency is 34kHz, and the electro-acoustic transfer function is approximated by the rounding technique described in Section 6.1, each of the inverse

filters have 56 poles which appear in 14 groups of four as shown in Figure 6.8a. These poles are all just inside the unit circle, and the distance from each group of poles to the unit circle is approximately 0.0035. Consequently, the exact inverse filters are all stable in forward time, but they all have long impulse responses; the contribution from each group of poles has a time constant of approximately 600 samples (see Equation (4.2.6)). When the outputs from the system are decimated by a factor of 10, the influence of the groups of poles away from the positive half of the real axis is eliminated, but there are still four poles near one, and they are not cancelled by corresponding zeros of the numerator. Thus, the impulse responses of the optimal filters are very long, even after decimation. When the recorded signals are delayed copies of a Hanning pulse of length 50, each of the outputs from the inverse filters in  $\mathbf{H}(z)$  are approximately 20 times longer than the input. Nevertheless, after decimation, the duration of the inputs to the loudspeakers are still comparable to the length of the Hanning pulse (this will be demonstrated in Section 6.6). The reason for this is that when the contributions from each of the inverse filters are added up, the four poles near one are almost exactly cancelled out by zeros.

A high sampling frequency and a wide microphone spacing generally leads to least squares problems whose exact solutions have many poles. Occasionally, some, or all, of these poles are cancelled out by zeros of the numerator, or by zeros that are generated by the addition of the individual filter outputs. In Figure 6.8, there are eight examples that demonstrate the influence of three system parameters on the number, and positions, of the poles of the exact inverse. The three system parameters are the approximation to the electro-acoustic transfer function, the microphone spacing, and the sampling frequency. The example above represented one extreme with many poles and a high degree of "symmetry" in the pole map as shown in Figure 6.8a. The other extreme is shown in Figure 6.8h. The sampling frequency is reduced from 34kHz to 12kHz, the microphone spacing is reduced from 10cm to 5cm, and the approximation to the acoustic transfer function is improved from the rounding technique to the linear interpolation technique. The inverse filters are calculated by using a filter length  $N_h$  of 60, a modelling delay  $m$  of 30, and a regularisation factor of  $10^{-6}$  (although use of regularisation is not crucial in this case). Figure 6.15a shows the geometry of the system, and Figure 6.15b shows the input  $|V_1|$  to source number one (positioned at

(-2m,2m)) as a function of the frequency  $f$  and the angle of incidence  $\theta_{aoi}$  of the recorded plane wave. Figure 6.15c shows five slices of Figure 6.15b along the  $\theta_{aoi}$ -axis. Figure 6.15b can be compared directly to Figure 3.22b, but note that the results plotted in Figure 3.22 are all obtained using eight loudspeakers rather than four. It is seen that  $|V_1|$  takes its largest values along the slice  $\theta_{aoi} = -45^\circ$  as it ideally should, and that it becomes smaller as  $\theta_{aoi}$  approaches a value that corresponds to the position of one of the other loudspeakers. The system does not work well above 3kHz, which is half the Nyquist frequency, but this is not surprising since the approximation to the electro-acoustic transfer function is known not to be accurate above this frequency.

## 6.6 Properties of the reproduced sound field

The transient responses of two sound reproduction systems are now simulated. In the first example, the reproduced sound field is illustrated in photographs of a few rendered frames that are taken from an animation. In the second example, the reproduced sound field is illustrated by a more comprehensive set of frames. Each frame is shown both as a basic “wire-frame plot” using only relatively few spatial sampling points, and as a “shaded density-plot”.

The geometry of the first example is shown in Figure 6.16. In the photograph in Figure 6.16c, the white spheres are the sources, and the red square is the “target area” which has a microphone in each of its corners. In the following photographs, the sources are excluded in order to be able to see more clearly what the sound field looks like near the target area. The sound field is calculated at 30 times 30 points over the yellow area. Two angles of incidence of a plane wave Hanning pulse are used to show first the desired sound field that is recorded by the microphones, and then the sound field that is caused by the sound radiated by the sources as a result of passing the recorded signals through the inverse filters. The instantaneous sound pressure is plotted as a three-dimensional surface which is rendered using artificial light-sources (the animations were originally done by Nick Gant who used the software 3D-Studio with the interface utility 3D-Surf). The sound pressure is calculated as a time sequence at a number of points on a two-dimensional grid assuming that the sampling frequency is 34kHz and that the speed of sound is 340m/s. The transfer function from

a source to a microphone, and from a source to a grid point, is approximated by rounding the distance in sampling intervals to the nearest integer. All the poles of the exact inverse filters are inside the unit circle; their impulse responses are therefore stable in forward time. Thus, it is straightforward to implement the inverse filters exactly as IIR filters. The time sequences observed at the 900 grid points are eventually low-pass filtered and decimated by a factor of ten to give the frames that are plotted and rendered as 3D-surfaces. Figure 6.17a shows the desired sound field for the first angle of incidence. It is a plane wave having a Hanning pulse time history of duration 50 samples, and it comes in from the upper left corner propagating down towards the lower right corner ( $\theta_{aoi} = -45^\circ$ ). The source  $S_1$  can easily reproduce waves coming from this direction, and Figure 6.17b shows the reproduced sound field when the recorded signals are passed through the inverse filter matrix. It is seen that only  $S_1$  gives a significant output, and this indicates that the system reconstructs the direction-of-arrival information in the recorded sound field very accurately when the original sound comes in from a direction that corresponds to a source position. Figure 6.18a shows the desired sound field for the second angle of incidence, it is a Hanning pulse of duration 50 samples that comes from a direction perpendicular to the line joining the sources  $S_1$  and  $S_2$  ( $\theta_{aoi} = -90^\circ$ ). Figures 6.18b and c show the reproduced sound field at two different times. Figure 6.18b shows the sound field at a time just before the interfering wavefronts reach the target area, and it illustrates how the sound emitted by  $S_1$  and  $S_2$  creates an approximately plane wavefront. Figure 6.18c shows the sound field a while later, and it is seen that  $S_3$  and  $S_4$  also emit sound, although of an amplitude much smaller than  $S_1$  and  $S_2$ . Note that for both  $\theta_{aoi} = -45^\circ$  and  $\theta_{aoi} = -90^\circ$ , the output from the sources is of short duration even though the poles of the system are close to the unit circle (see Figure 6.8a and also Section 6.5). This happens because the four poles near one are almost exactly cancelled by zeros of the sum of the filter outputs from each row of  $\mathbf{H}(z)$ . The influence of all other poles is eliminated by the lowpass-filtering preceding the decimation, and the many zeros of the Hanning pulse on the unit circle.

The geometry of the second example is shown in Figure 6.19. There are four loudspeakers in a narrow-arc arrangement as shown in Figure 3.4b, and three microphones in a circle whose radius is 10cm as shown in Figure 3.20d (for a radius

of 3cm). Figure 6.19b is equivalent to Figure 6.16b. It views the geometry from the same position as used in the frames that show the desired and the reproduced sound field. The sampling frequency is 12kHz, and the electro-acoustic transfer function is approximated by the linear interpolation technique described in Section 6.1. There are 120 coefficients in each of the inverse filters. They are calculated using a modelling delay of 60. Even though the positions of the 22 zeros of the determinant of  $\mathbf{C}^T(z^{-1})\mathbf{C}(z)$  do not appear to indicate an ill-conditioned inversion problem, the  $\mathbf{C}$ -matrix used for the design of the twelve inverse filters have singular values of zero, and therefore a regularisation factor  $\beta$  of  $10^{-7}$  is necessary. The recorded sound field is a Hanning pulse of length 25, which means that the main part of its energy is distributed over frequencies below 920Hz. Figures 6.20-25 show nine frames of a) the desired sound field, and nine frames of b) the reproduced sound field for the three angles of incidence  $90^\circ$  (Figures 6.20 and 6.21),  $45^\circ$  (Figures 6.22 and 6.23), and  $0^\circ$  (Figures 6.24 and 6.25). For each angle of incidence, the sound field is illustrated by a 3D wire-frame plot (Figures 6.20, 6.22 and 6.24) and a shaded density-plot (Figures 6.21, 6.23 and 6.25). There are twelve sampling intervals between each frame, this is indicated by a frame index which is shown above each 3D wire-frame (the frame index is relative to a reference that is not unique, it is specific to each sequence of frames). The recorded sound field is shown in the range from zero to one, the dynamic range of the reproduced frames is limited by using an exponential clipping function (the parameter values are  $y_{\text{clip}} = 1.5$  and  $y_{\text{max}} = 2.0$ , see Equation A6.1.1 in Appendix 6.1). When  $\theta_{\text{aoi}}$  is  $90^\circ$  or  $45^\circ$ , the reproduced field corresponds well to the recorded field over an area which is quite large compared to the area covered by the microphones. When  $\theta_{\text{aoi}}$  is  $0^\circ$ , the recorded plane wave comes in from a direction that is not covered by the loudspeakers. It is seen that the four loudspeakers succeed in making the wavefront bend near the microphones in order to obtain the desired effect, but it is clearly a very localised phenomenon.

## 6.7 Conclusions

A two-coefficient digital filter is very efficient for modelling an electro-acoustic transfer function from a monopole source to a field point. It is extremely accurate at low frequencies, and it is simple to implement. It also tends to break up the symmetry



of the poles in the exact least squares inverse, which makes it easier to invert the system with finite impulse response filters. When the microphones are moved closer together, or the sampling frequency is decreased, the poles decrease in number, and they also tend to move round the unit circle towards the left half of the complex plane. However, some "symmetrical" layouts of microphones and loudspeakers cannot be inverted properly because they always result in extremely ill-conditioned inversion problems. This is sometimes indicated by a singular "inverse distance matrix". Consequently, it is impossible to design a microphone array with only three or four microphones that will work well with almost any loudspeaker setup, but it might be possible with more microphones, for example eight. When a transient, such as a Hanning pulse, is recorded by a microphone array that covers a relatively small area, the recorded signals are to a good approximation delayed copies of the same "reference" pulse. After the reference pulse has been passed through the inverse filter matrix, the outputs from one row of inverse filters are added together to form the input to one loudspeaker. This addition can sometimes eliminate the influence of a pole, and thus make the sequence that is input to a source much shorter than expected. The spatially reproduced sound field confirms the results from the frequency analysis; the size of the area over which the sound field can be controlled is of the order of a few acoustical wavelengths.

## **Appendix 6.1 Modelling the spatially reproduced sound field**

Once the source outputs in  $v(z)$  are known, all the information necessary for calculating the spatially reproduced sound field is available. At any point in the listening space, the contributions from the individual sources are added together, and the result is the sound pressure at that point. This is straightforward in principle, but there are two problems with the practical implementation. One is that the source outputs are known only in discrete time, and it is generally not possible to calculate the reproduced sound field without knowing the source outputs as a continuous function of time. The other problem is that if the spatial grid is too coarse compared to the spatial variations in the reproduced field, the visual appearance of the sampled field varies from unpleasing at best to deceptive at worst.

The hypothetical example shown in Figure 6.26 illustrates the problem caused by  $v(z)$  being known only in discrete time. The output from two sources,  $S_1$  and  $S_2$ , are the

two finite length sequences  $v_1(n)$  and  $v_2(n)$ . At a listening point, the two sequences are delayed by an amount equivalent to the time it takes for the sound to travel from each of the sources to the listening point, and they are also attenuated by the spherical spreading of the sound (see Equation (6.1.1)). The delay is generally not an integer number of sampling intervals. It is therefore necessary to interpolate in the delayed and attenuated versions of  $v_1(n)$  and  $v_2(n)$  in order to find the values at the sampling times. The two contributions are then added together, and the result is the sequence  $w(n)$  that is reproduced at the listening point, and it is sampled at the same time instances as  $v_1(n)$  and  $v_2(n)$ . The interpolation can be made with any curve-fitting method, for example with splines or polynomial approximation (Press et al [71] Chapter 3), or the fractional delay technique demonstrated in Section 6.1. The linear interpolation technique given by Equation (6.1.3) is a special case of a polynomial approximation, and it is an efficient and simple way to approximate a non-integer delay.

Once it is decided how to calculate the reproduced sound field at any position in the listening space, it must be decided how many listening points that are necessary to represent the spatial variations in the field. The set of listening points are also referred to as grid points ([58] pp.246-247). If the field varies very quickly with the spatial position, many grid points are necessary. If too few points are used, important detail is lost due to spatial aliasing. When the objective is to achieve a faithful "snapshot" of the sound field, it is not sufficient to use the sampling theorem to determine the minimum spacing between grid points. A subjective assessment of the plot of the sound field must be made in order to determine when the plot "looks good". When sequences of plots are used for animations, it is also necessary to determine the time interval between consecutive plots. If that time interval is too long, the animation will not be smooth, instead "jumps" will appear. Each plot (or "snapshot") in an animation is referred to as a frame, and it is assigned an index that indicates its number in the sequence. Figure 6.27 illustrates how the quality of an animation depends on the spatial sampling grid. Three sampling grids are used to show what the sound field looks like after linear interpolation between the sampling points. The left column in Figure 6.27 is the reference sequence. A pulse, which is four spatial sampling intervals wide, is sampled in seven frames without loss of information. When the grid spacing is

doubled (middle column), some detail is lost because the peak of the original pulse falls in between the spatial sampling points in some of the frames. This phenomenon becomes even more apparent when the grid spacing is increased further to three times the original spacing (right column); in this case the pulse starts to look wider than it really is. Another way of assessing the interpolation error is to look at the observed time sequence at a position between the grid points, this is shown in the bottom row of Figure 6.27 for the position  $x = 5$ . The interpolated signal looks qualitatively similar to the original pulse only when the pulse extends over two, or more, sampling intervals, which is in agreement with the sampling theorem. As a rule of thumb, a sound field looks acceptable for animation as long as the spatial extent of its finest detail is no smaller than four spatial sampling intervals. Equivalently, an animation does not "jump" as long as there are at least four frames to cover the spatial variation of the finest detail in the reproduced field as indicated in the left column of Figure 6.27. Consequently, the time interval between two consecutive frames must be approximately equal to the time it takes for the sound to travel from one spatial sampling point to its nearest neighbour. For single frequency fields, this implies that at least eight frames are necessary for animation of one period, twelve are recommended. For animation of transient responses, a sensible starting point is to generate a minimum of 30 frames on a square grid consisting of 30 times 30 points. These choices, along with the size of the area over which the sound field is calculated, set a limit for the highest frequency component in the reproduced signal. For example, if the sound field is calculated over a  $2\text{m} \times 2\text{m}$  square, then the shortest wavelength in the reproduced signal must be less than  $4 \times 2\text{m} / 30$  which is approximately 27cm which corresponds to a frequency of 1275Hz (assuming the speed of sound is 340m/s). These guidelines are based on experience with basic mesh plot routines (linear interpolation between grid points). The details of three-dimensional surfaces are greatly enhanced by rendering them using artificial light-sources and sophisticated smoothing algorithms, and it also makes them look visually pleasing.

One final problem relating to the modelling of spatially reproduced sound fields is worth mentioning. Sometimes it is desirable to have sources close to the area over which the sound field is calculated, or perhaps even inside that area. This causes a problem with the dynamic range of the plot because the sound pressure close to a

source is very high. Software clipping must therefore be used. A rule of thumb is to aim for a ratio of three between the highest value at any grid point and the typical value of the acoustical phenomenon that we wish to observe. For example, if the desired value at a microphone is one, then there should be no grid points where the value of the sound field is greater than three. The simplest way to clip the values at the grid points is to set all values outside a certain range equal to the closest member of that range (for example, all values greater than three are set to three, and all values smaller than minus three are set to minus three). However, this method tends to result in very "brutal" clipping. It is better to compress the signal gradually, for example by using an exponential clipping function of the type

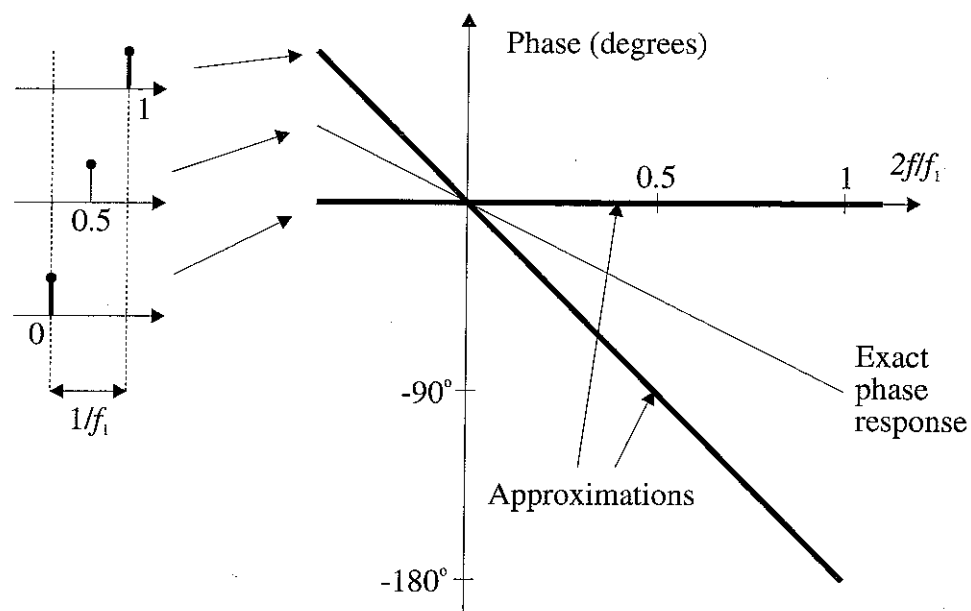
$$y(x) = \begin{cases} x < -y_{\text{clip}} & -y_{\text{clip}} - (y_{\text{max}} - y_{\text{clip}}) \left[ 1 - e^{(x+y_{\text{clip}})/(y_{\text{max}}-y_{\text{clip}})} \right] \\ -y_{\text{clip}} < x < y_{\text{clip}} & x \\ x > y_{\text{clip}} & y_{\text{clip}} + (y_{\text{max}} - y_{\text{clip}}) \left[ 1 - e^{-(x-y_{\text{clip}})/(y_{\text{max}}-y_{\text{clip}})} \right] \end{cases}, \quad (\text{A6.1.1})$$

where  $y_{\text{clip}}$  is the magnitude of the value at which clipping starts (for example 2.5), and  $y_{\text{max}}$  is the highest value the output can possibly take (for example 3). A function that compresses its argument even more gradually is

$$y(x) = \frac{2}{\pi} y_{\text{max}} \text{ArcTan} \left( \frac{\pi}{2} \frac{x}{y_{\text{max}}} \right). \quad (\text{A6.1.2})$$

The first derivative of both the exponential clipping function and the ArcTan compression function is continuous everywhere. The paper by Krause and Petersen [48] demonstrates how software clipping can increase the dynamic range of digital recordings.

a)  $f_s = f_1$



a)  $f_s = f_{10} = 10f_1$

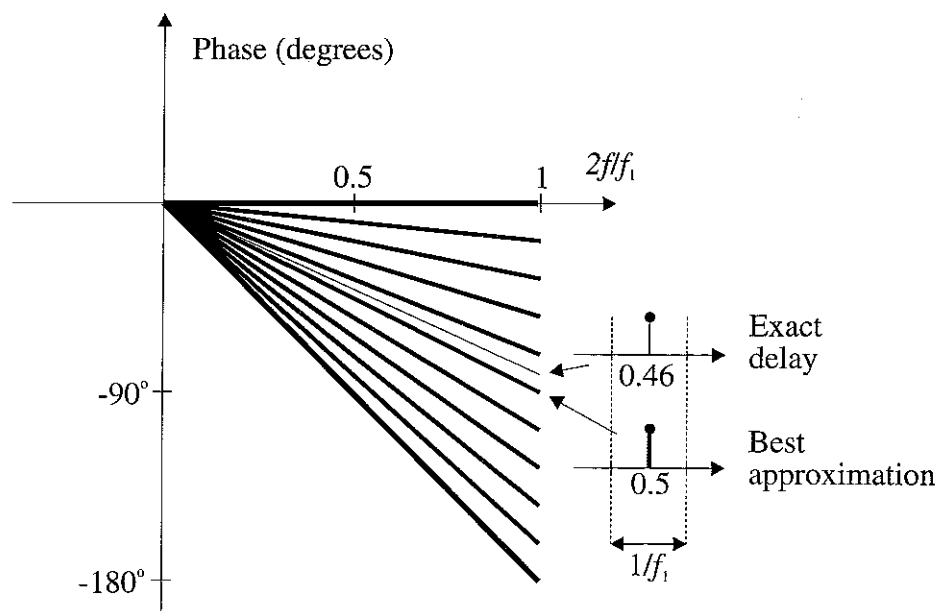


Figure 6.1. Illustration of phase quantisation. The higher the sampling frequency, the more accurate is the quantisation at low frequencies

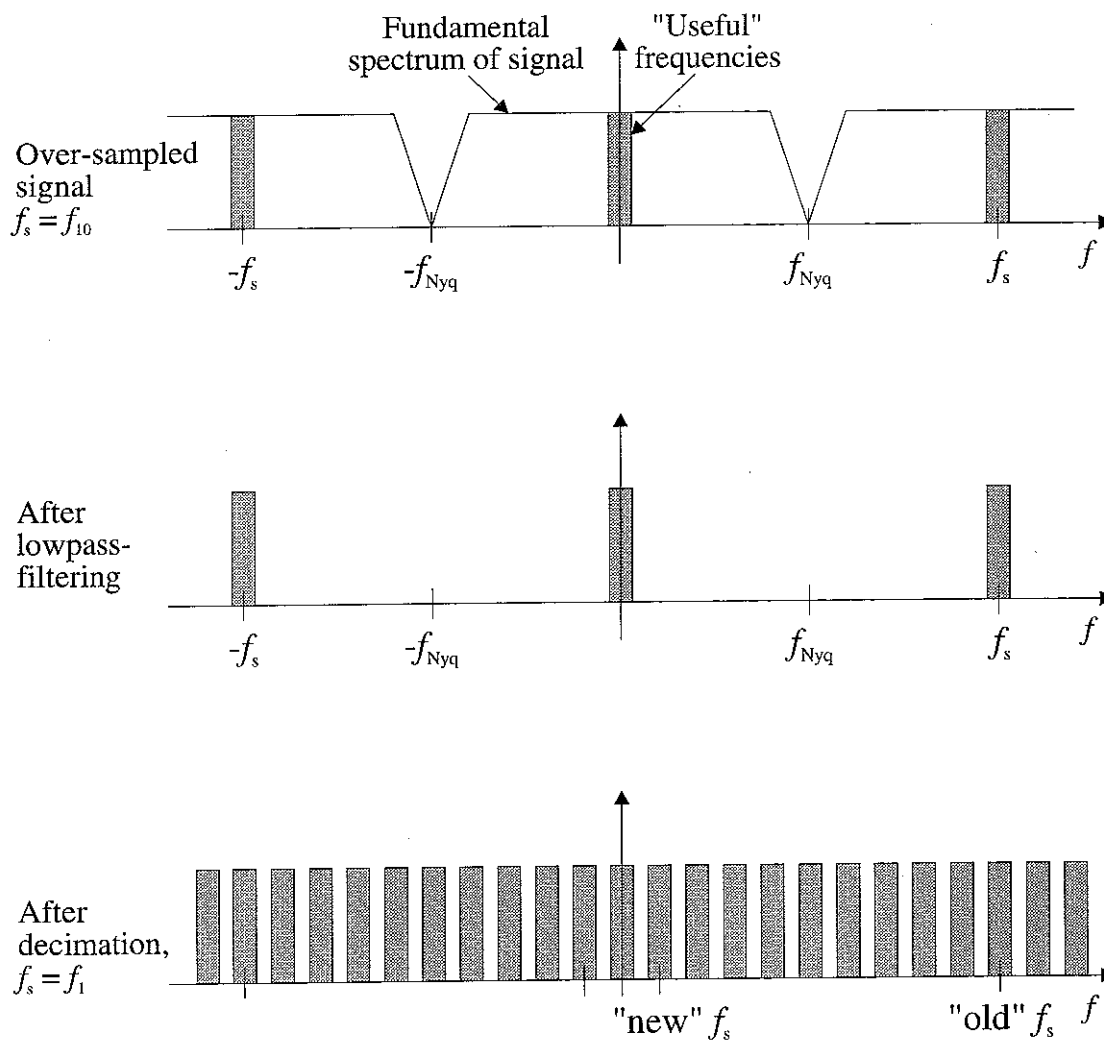
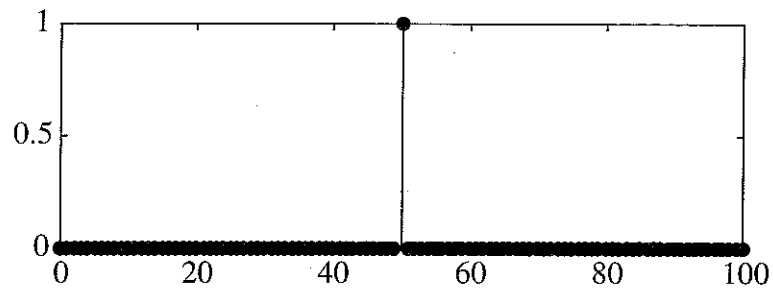
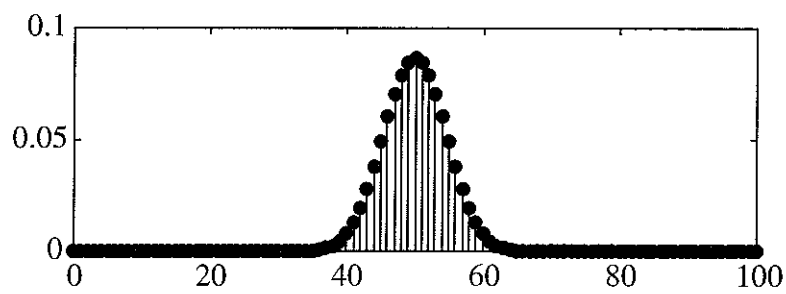


Figure 6.2. Frequency domain interpretation of the principle of increasing the phase resolution by using over-sampling

Over-sampled  
signal  
 $f_s = f_{10}$



After  
lowpass-  
filtering



After  
decimation  
 $f_s = f_1$

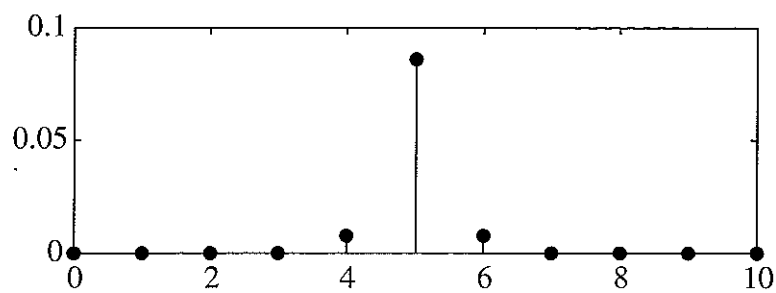


Figure 6.3. Time domain interpretation of the principle of increasing the phase resolution by using over-sampling

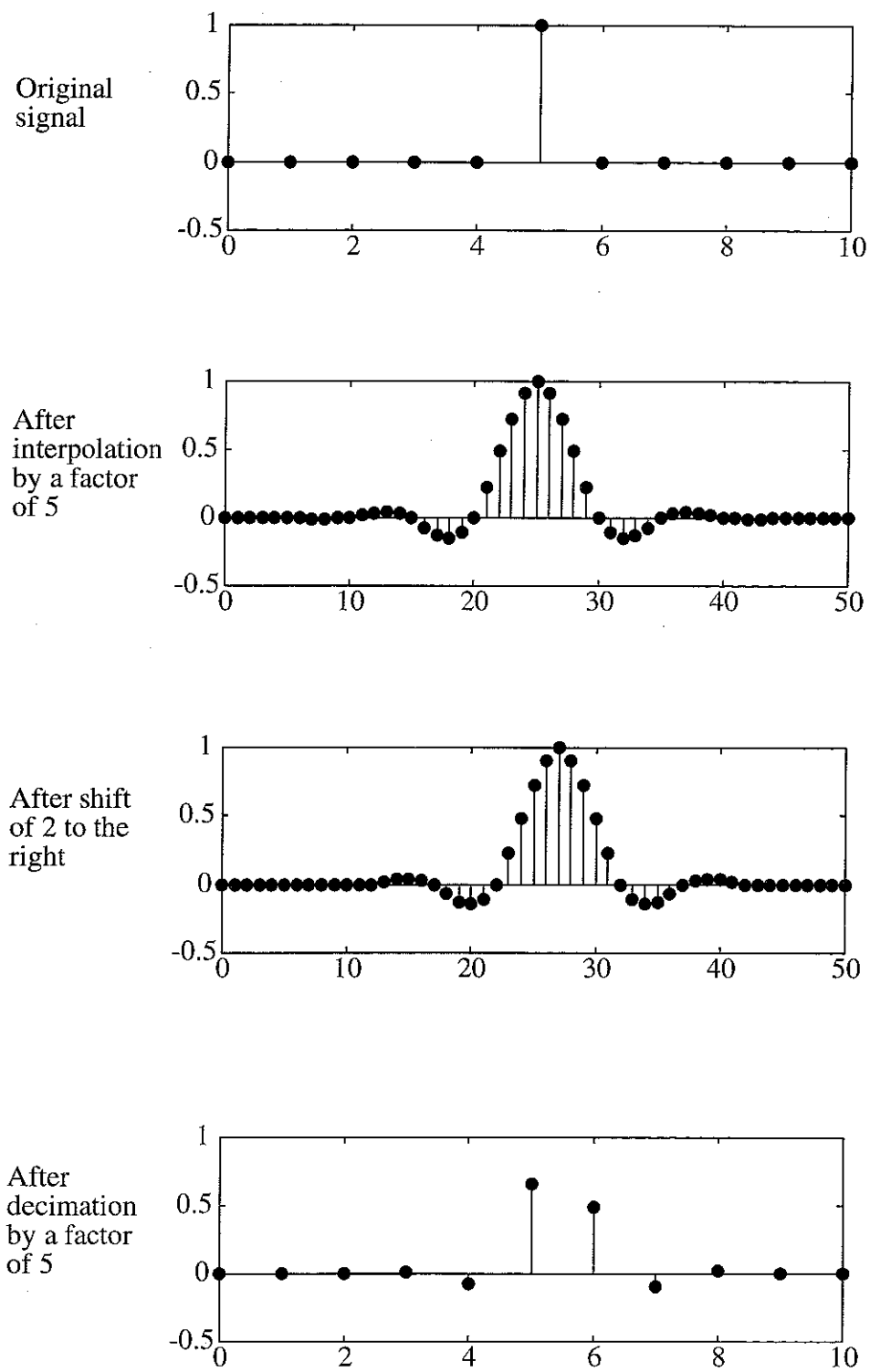
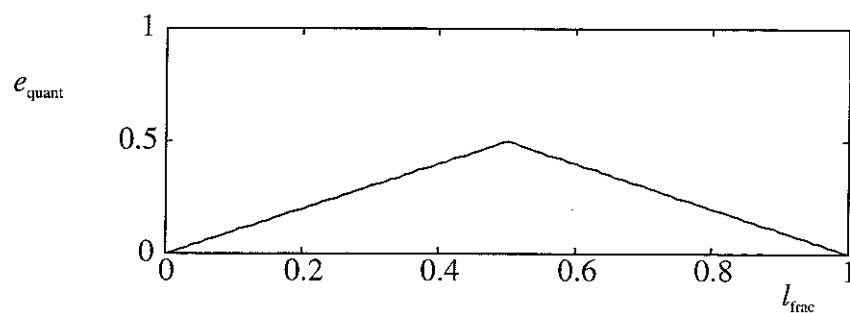


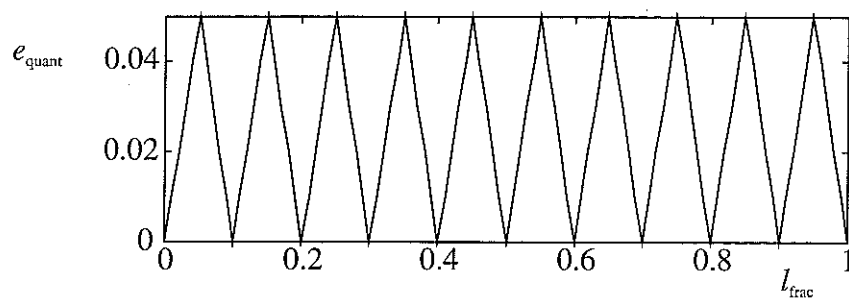
Figure 6.4. An example showing how to realise a delay of 0.4 samples with the fractional sample phase shifter



a) Rounding,  $f_s = f_1$



b) Rounding after 10 times over-sampling,  $f_s = f_{10}$



c) Fractional delay  $N/D$ ,  $D \leq 10$

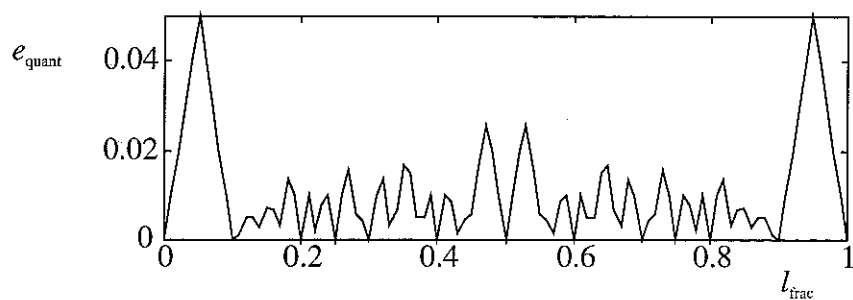


Figure 6.5. Comparison of quantisation errors using three different modelling techniques

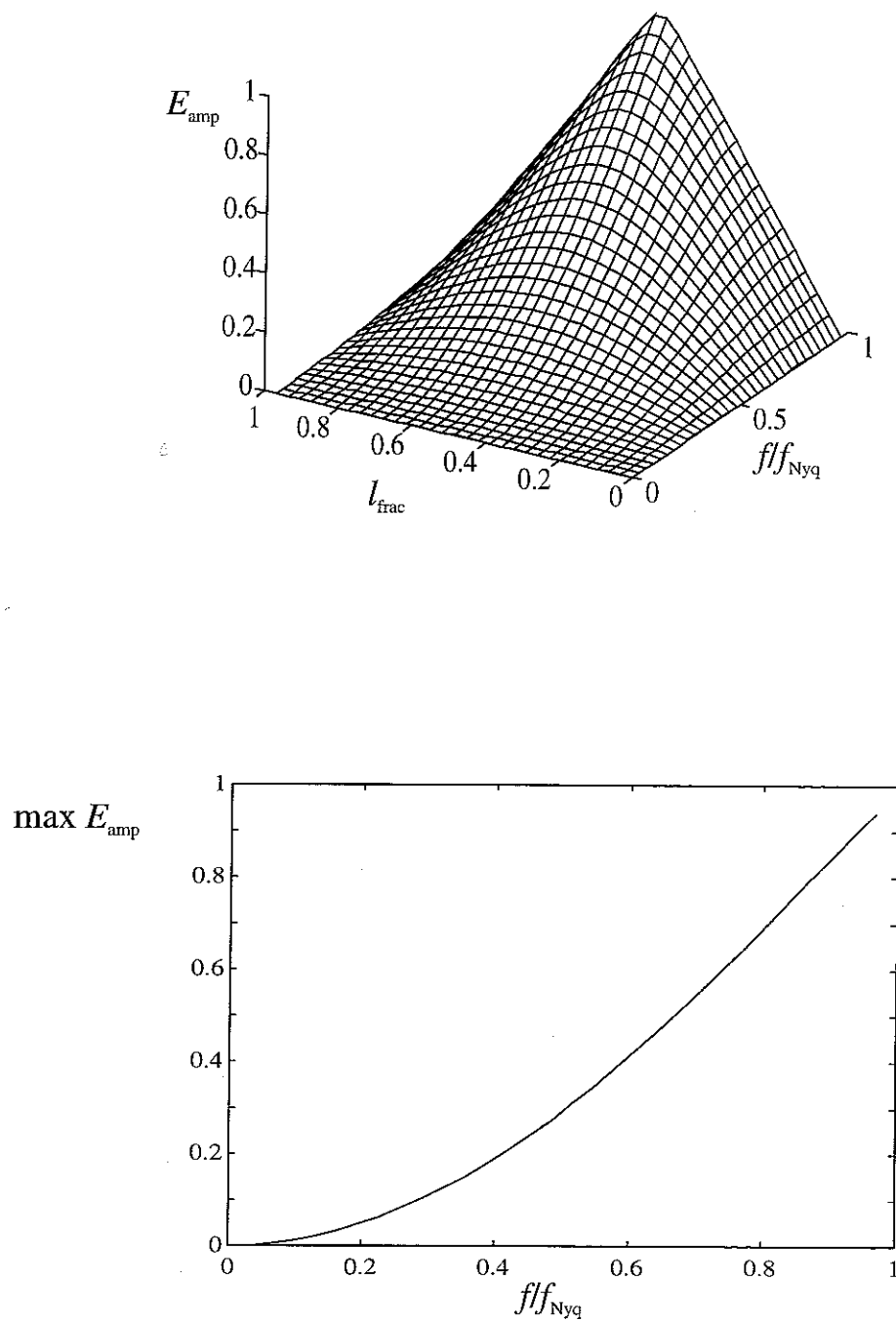


Figure 6.6. Amplitude error using the linear interpolation modelling technique

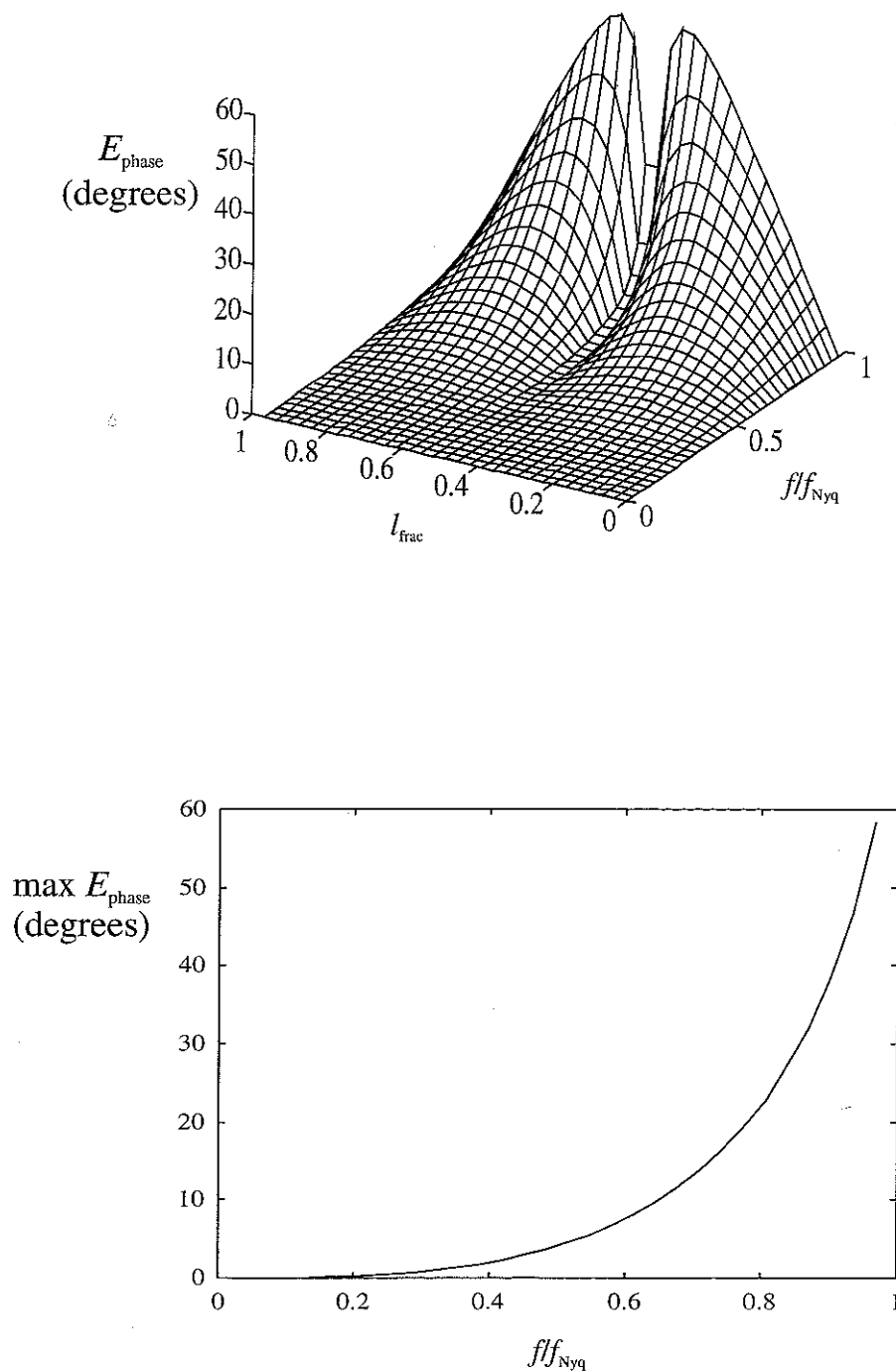
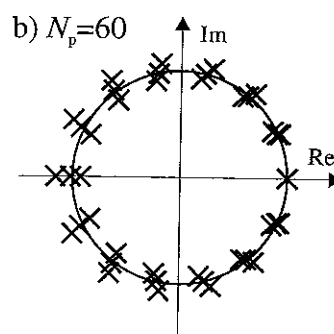
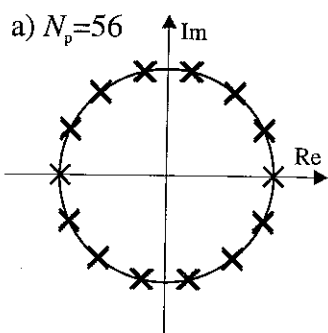
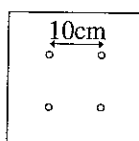


Figure 6.7. Phase error using the linear interpolation modelling technique

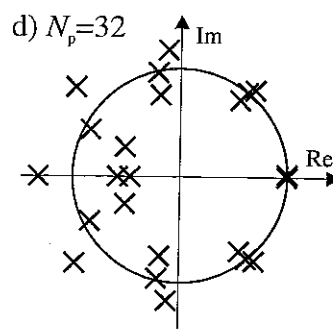
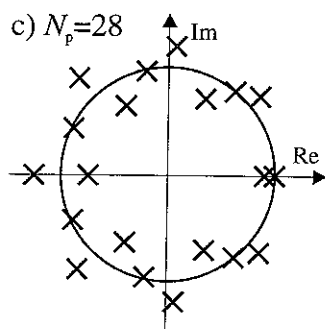
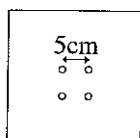
# Rounding

# Linear interpolation

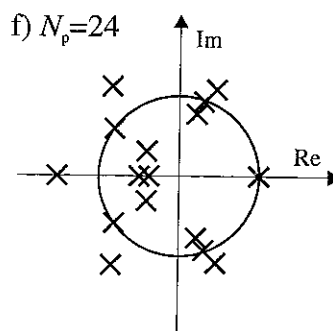
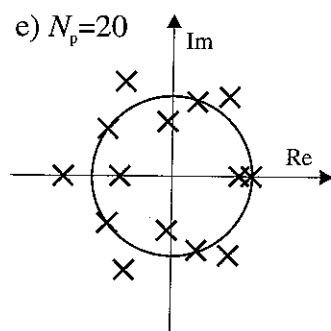
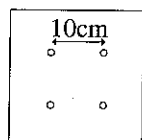
$$f_s = 34 \text{ kHz}$$



$$f_s = 34 \text{ kHz}$$



$$f_s = 12 \text{ kHz}$$



$$f_s = 12 \text{ kHz}$$

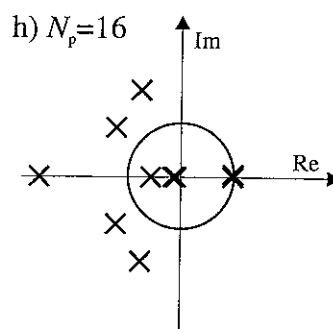
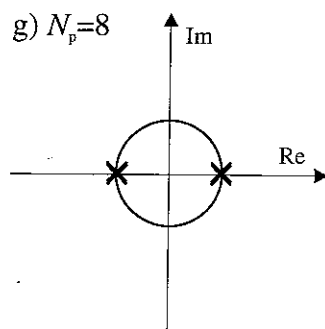
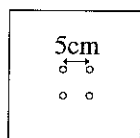


Figure 6.8. Pole-maps and the number of poles  $N_p$  of the exact inverse filters for the eight systems derived from the data in Table 6.1 (four poles at approximately -3.33 are omitted from b))

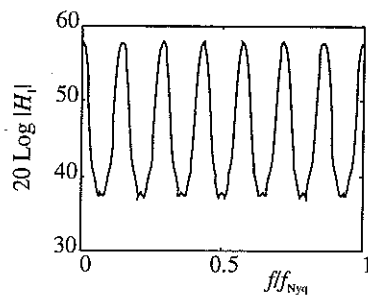
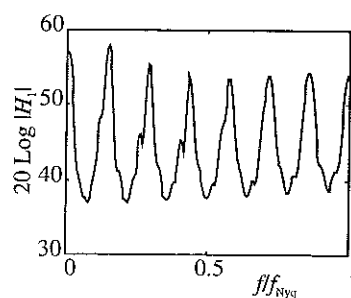
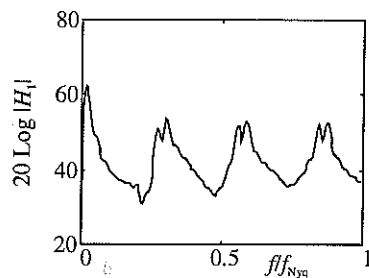
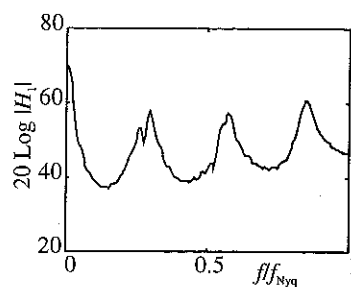
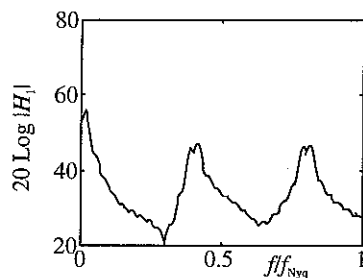
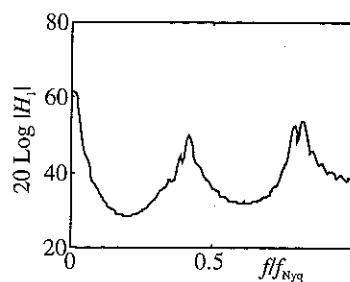
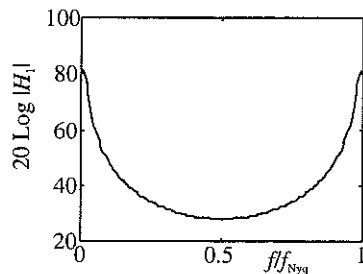
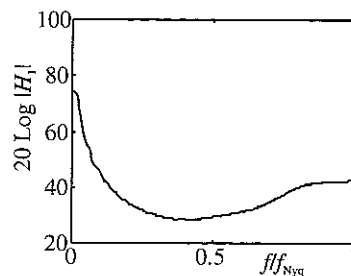
a)  $N_p=56$ b)  $N_p=60$ c)  $N_p=28$ d)  $N_p=32$ e)  $N_p=20$ f)  $N_p=24$ g)  $N_p=8$ h)  $N_p=16$ 

Figure 6.9. Frequency response of the optimal filter  $|H_1|$  for each of the eight systems whose poles are shown in Figure 6.8. Each of the optimal filters contain 120 coefficients, the modelling delay is 60 (no regularisation is used). Note the good agreement with the frequency response expected from the pole-maps in Figure 6.8

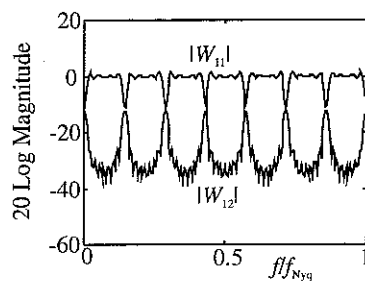
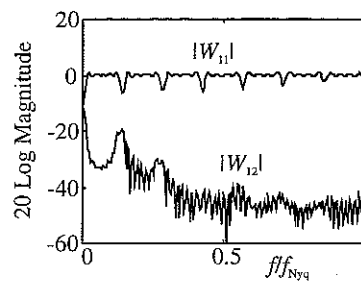
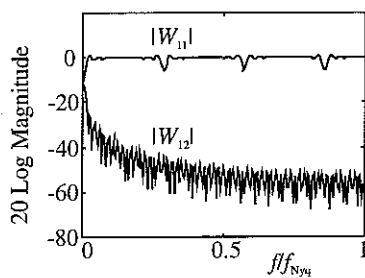
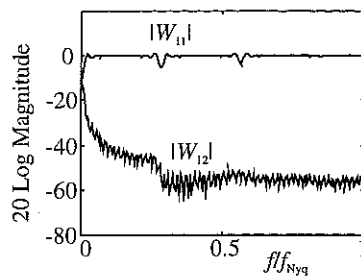
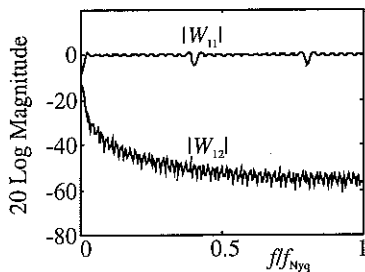
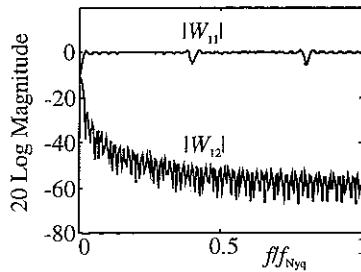
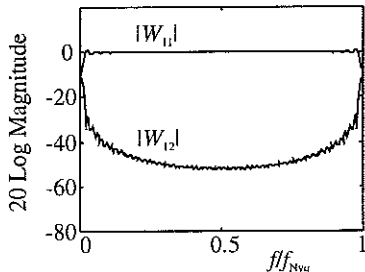
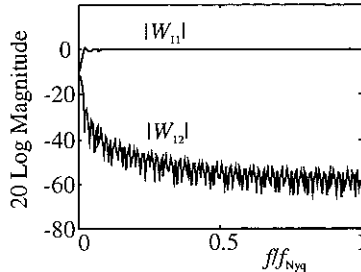
a)  $N_p=56$ b)  $N_p=60$ c)  $N_p=28$ d)  $N_p=32$ e)  $N_p=20$ f)  $N_p=24$ g)  $N_p=8$ h)  $N_p=16$ 

Figure 6.10. Frequency response of the reproduced signals  $|W_{11}|$  and  $|W_{12}|$  for each of the eight systems whose poles are shown in Figure 6.8, and whose FIR approximations are shown in Figure 6.9. Note the dips in  $|W_{11}|$  at the frequencies corresponding to the positions of poles close to the unit circle

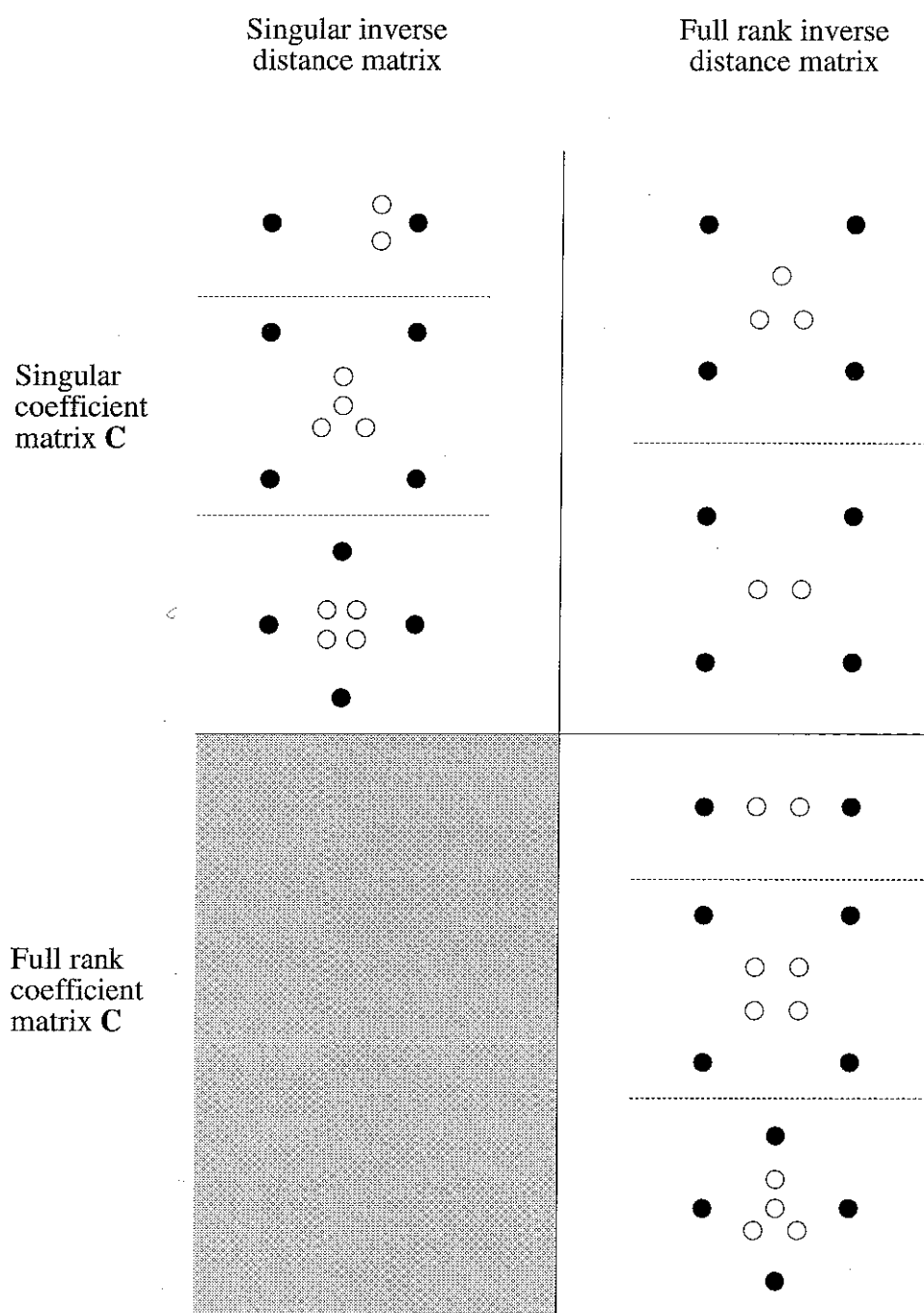


Figure 6.11. Some layouts of sources (black spheres) and microphones (white spheres), and the types of inversion problems they can lead to

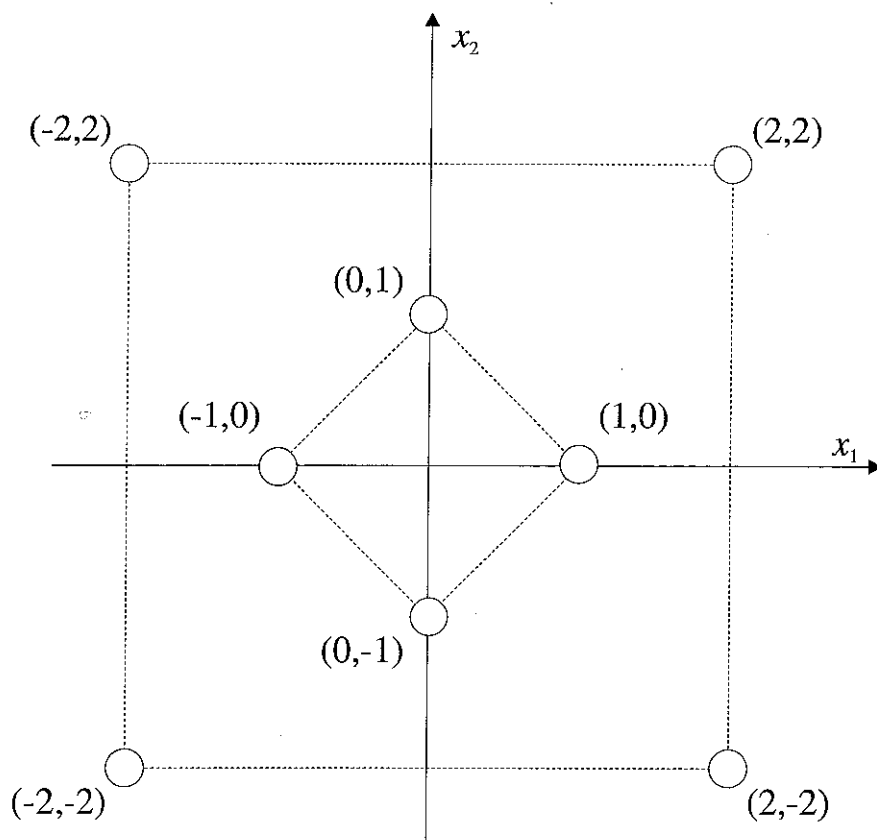


Figure 6.12. A microphone array comprising eight microphones arranged in two squares that are rotated  $45^\circ$  relative to each other



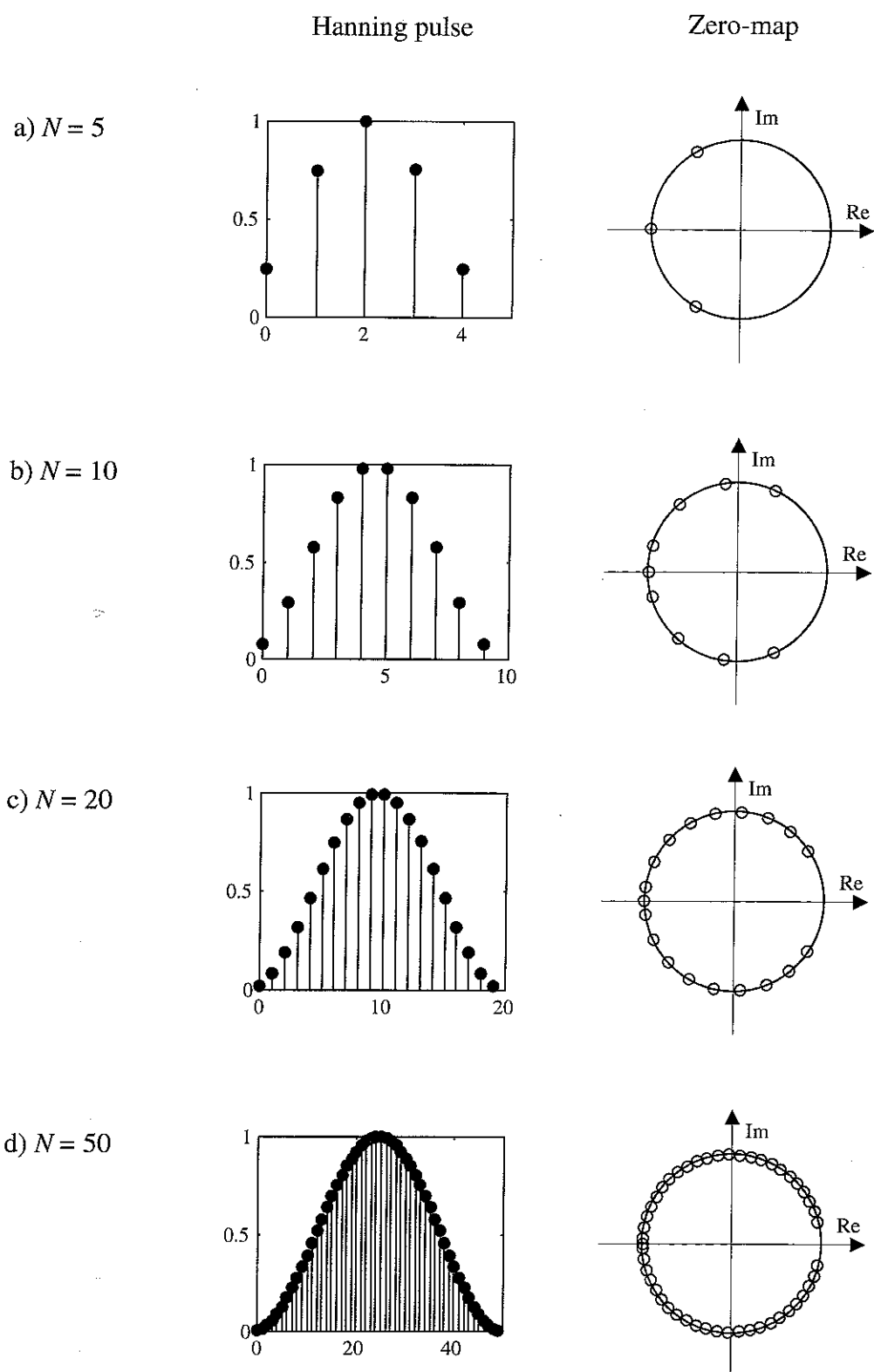
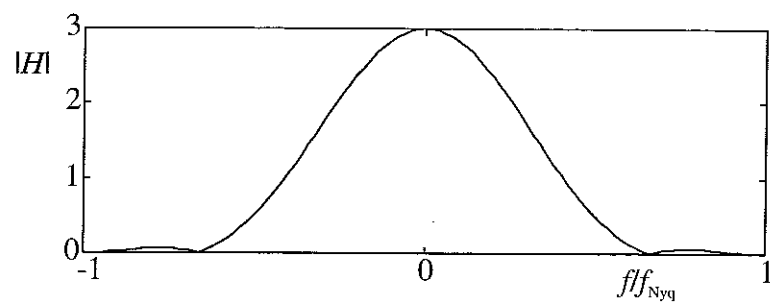
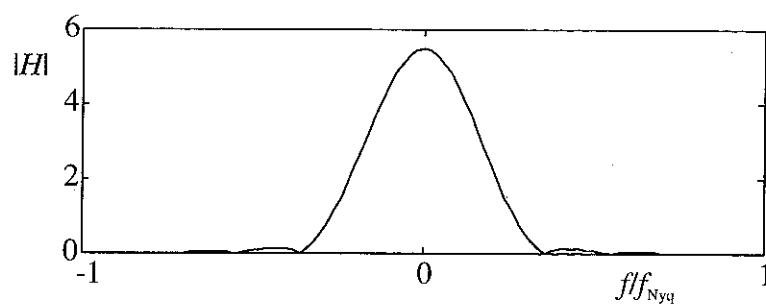


Figure 6.13. Four Hanning pulses of different lengths, and their corresponding zero-maps

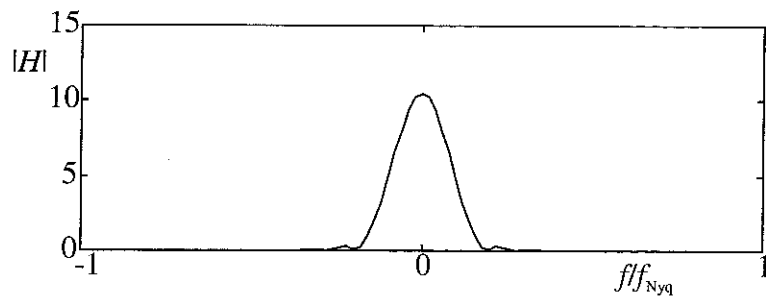
a)  $N = 5$



b)  $N = 10$



c)  $N = 20$



d)  $N = 50$

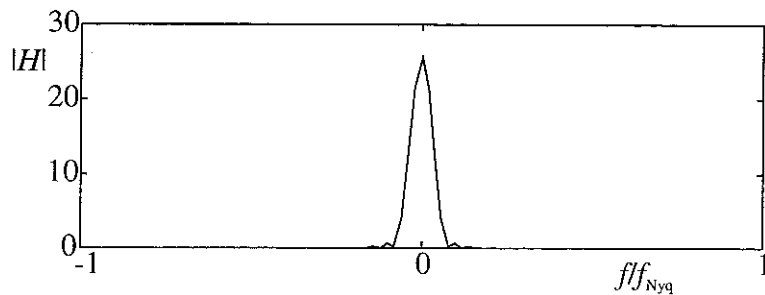


Figure 6.14. Frequency spectrum of the four Hanning pulses shown in Figure 6.13.

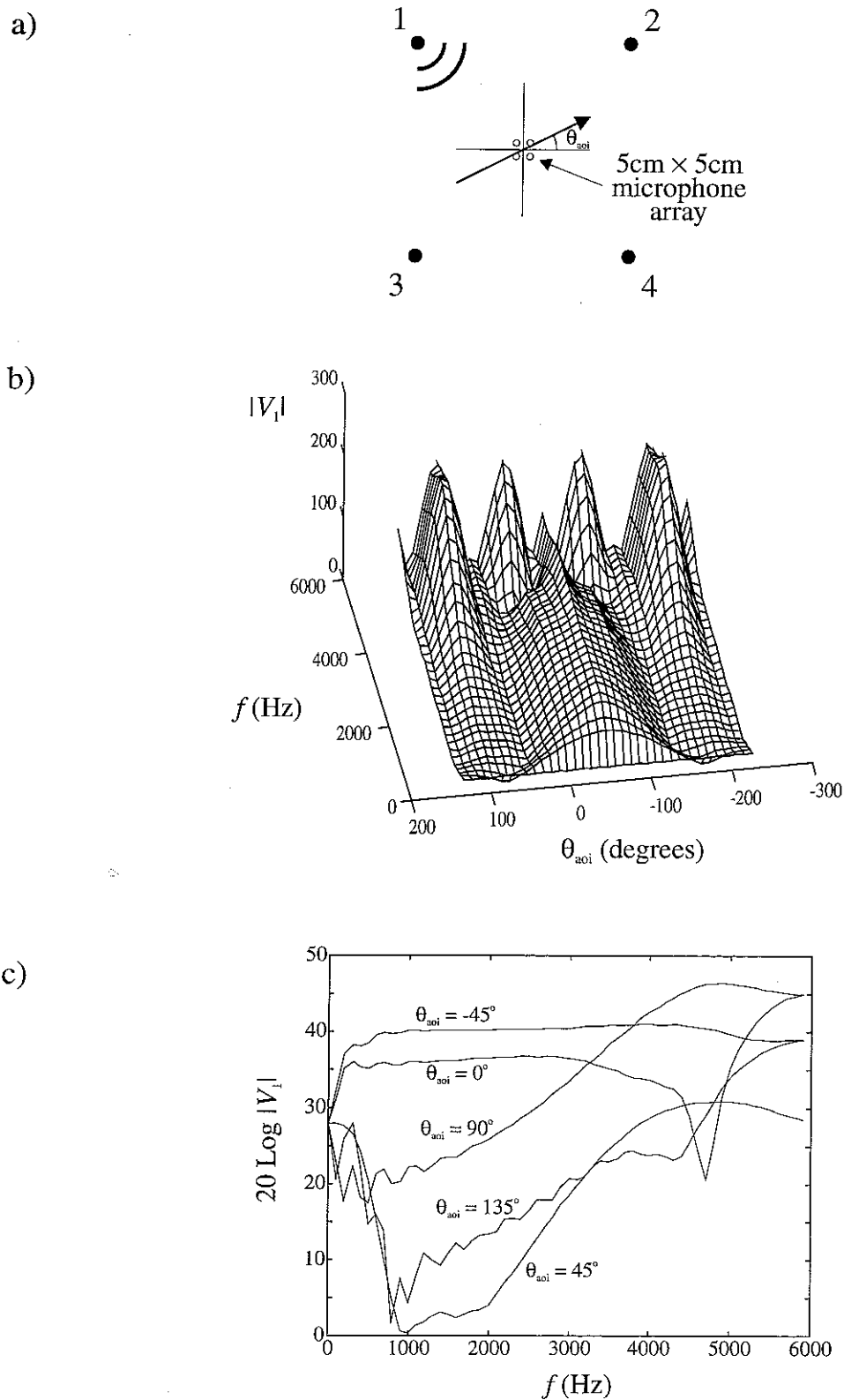


Figure 6.15. a) Geometry of a 4-by-4 system, b) the input  $|V_1|$  to loudspeaker number one as a function of the frequency and the angle of incidence of the recorded plane wave, and c) five slices of b) along the frequency axis

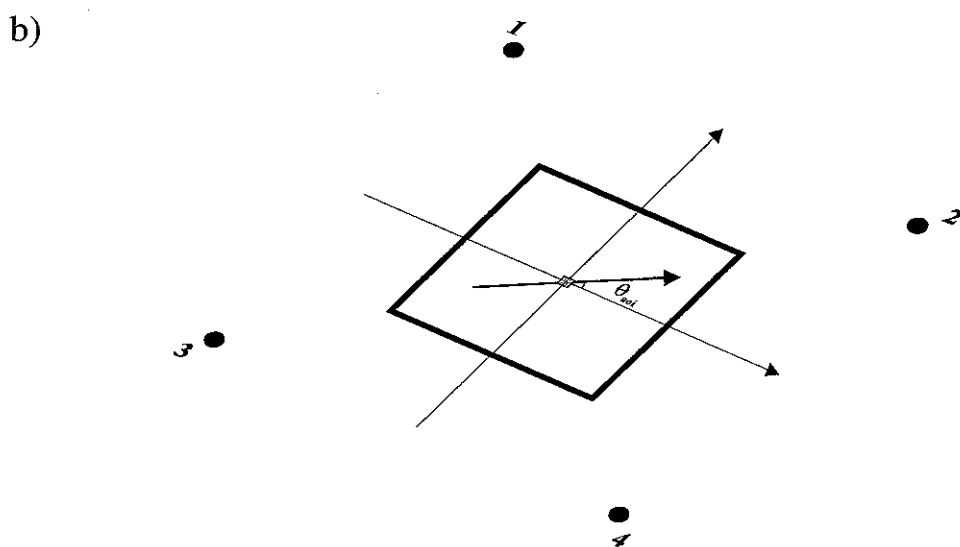
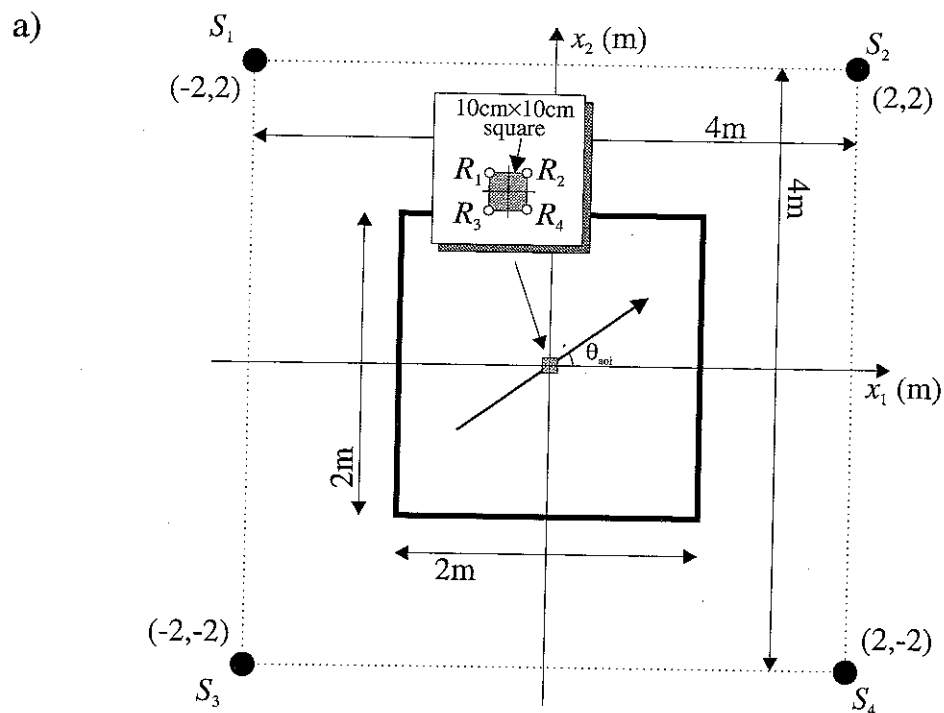


Figure 6.16 (first page of two). The first example of a geometry used for simulation of a transient response. a) is viewed from directly above, b) is viewed from the same position as used in the photographs of the sound field shown in Figures 6.17 and 6.18

c)

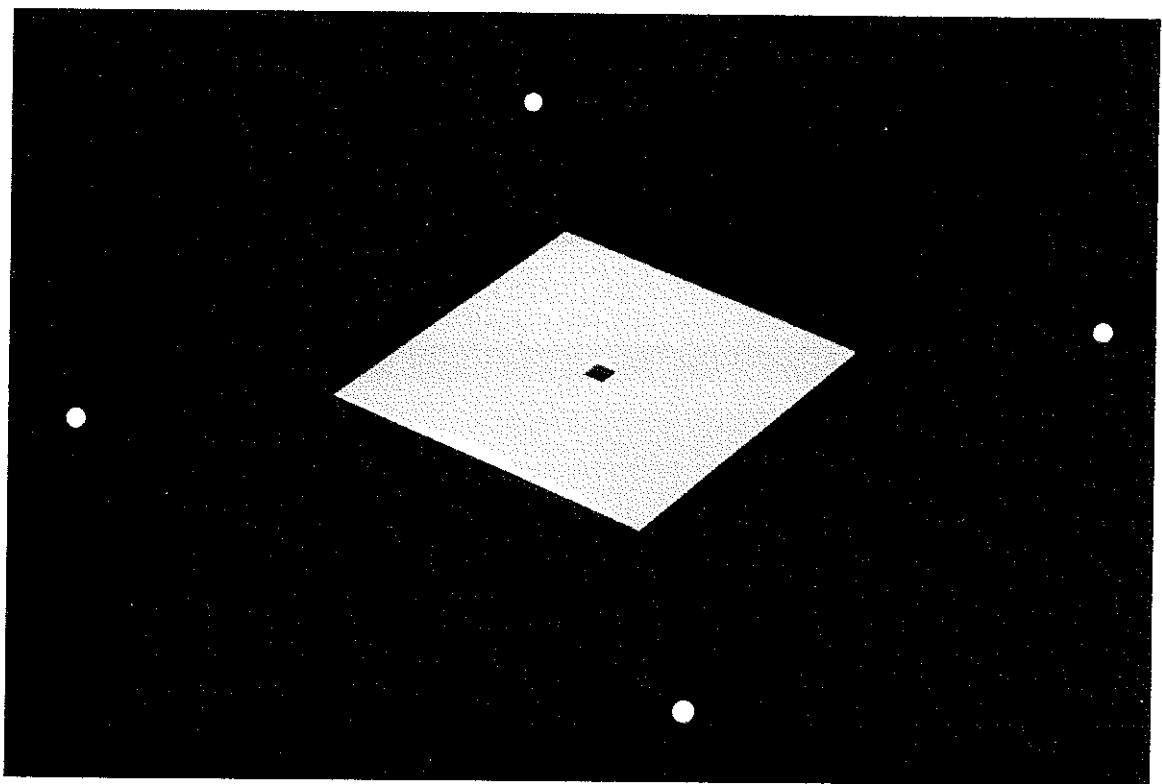
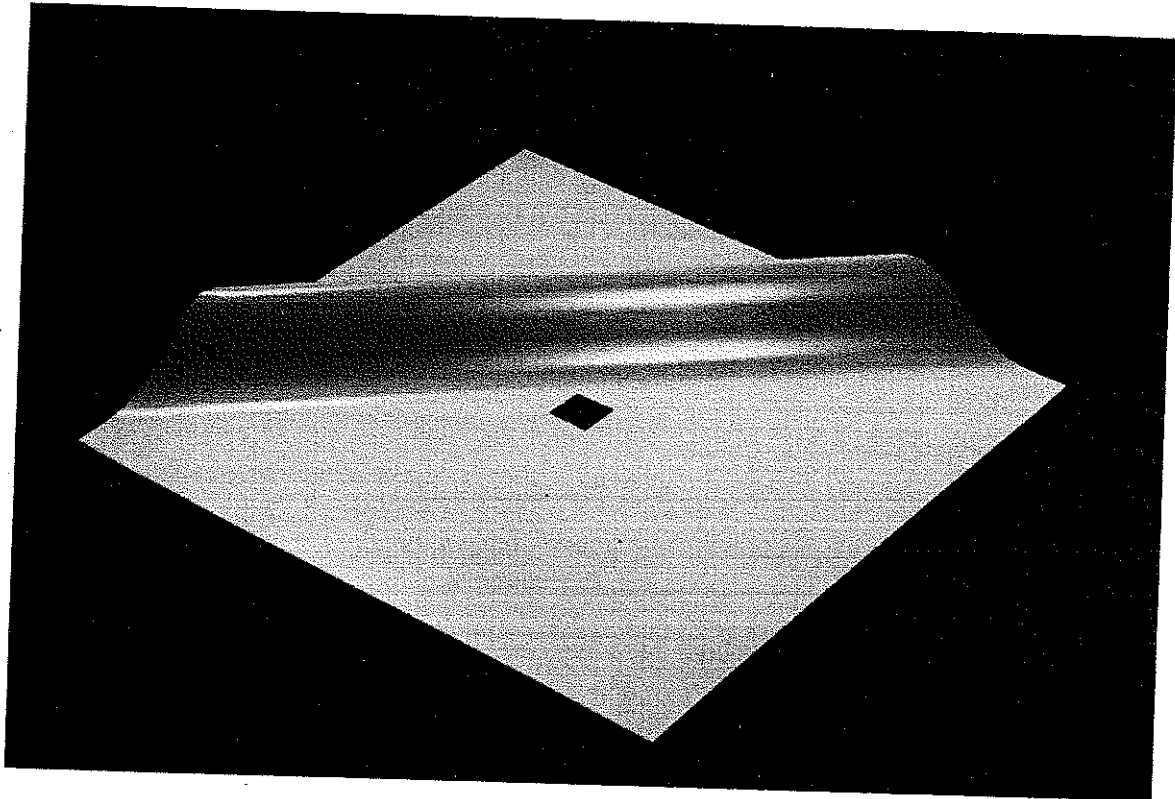


Figure 6.16 (second page of two). Photograph showing the geometry as illustrated in Figure 6.16b

a)



b)

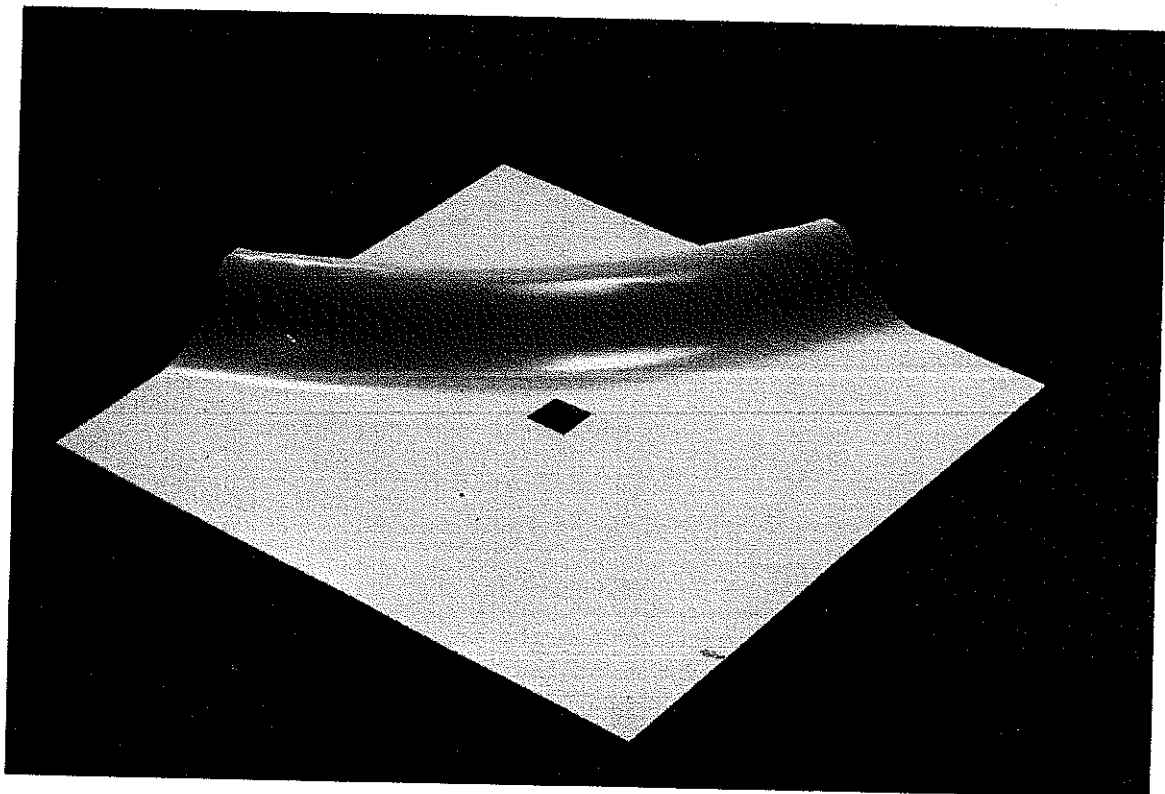


Figure 6.17. a) Desired plane wave Hanning pulse (duration 50 samples,  $\theta_{\text{des}} = -45^\circ$ ) and b) the reproduced sound field

a)

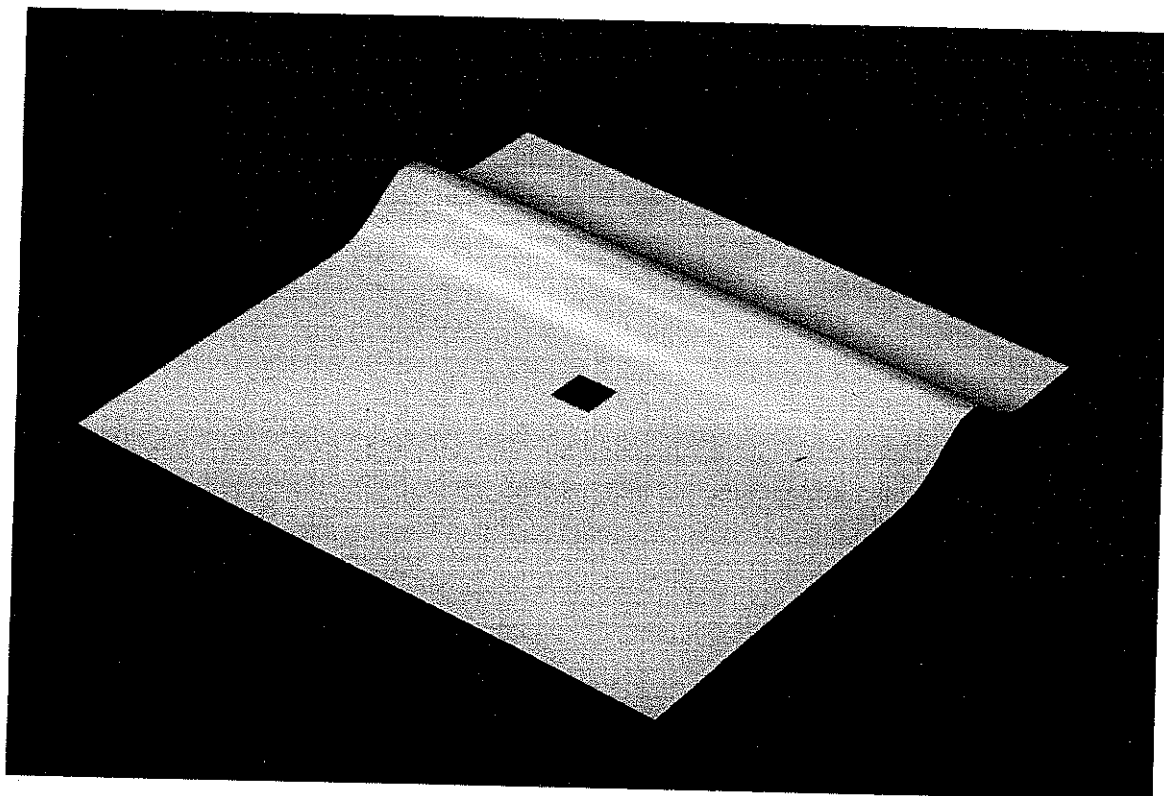
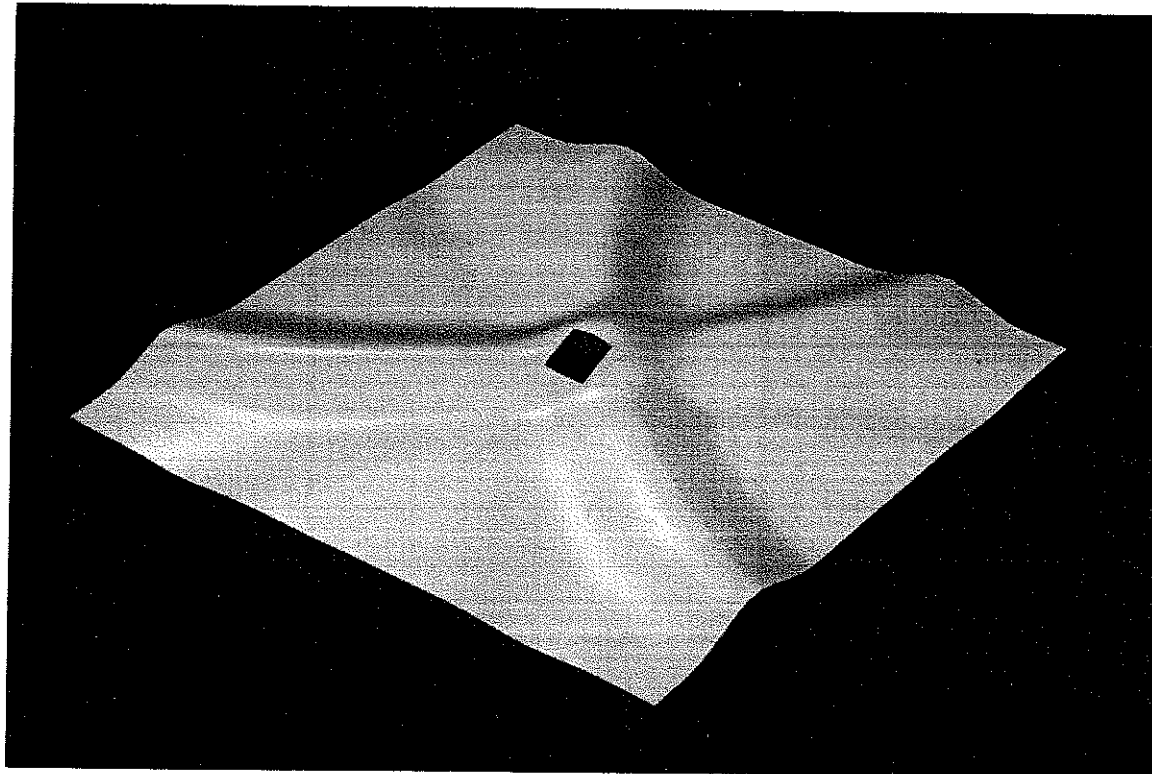


Figure 6.18 (first page of two). a) Desired plane wave Hanning pulse (duration 50 samples,  $\theta_{aoi} = -90^\circ$ ) at a time before the pulse reaches the target area

b)



c)

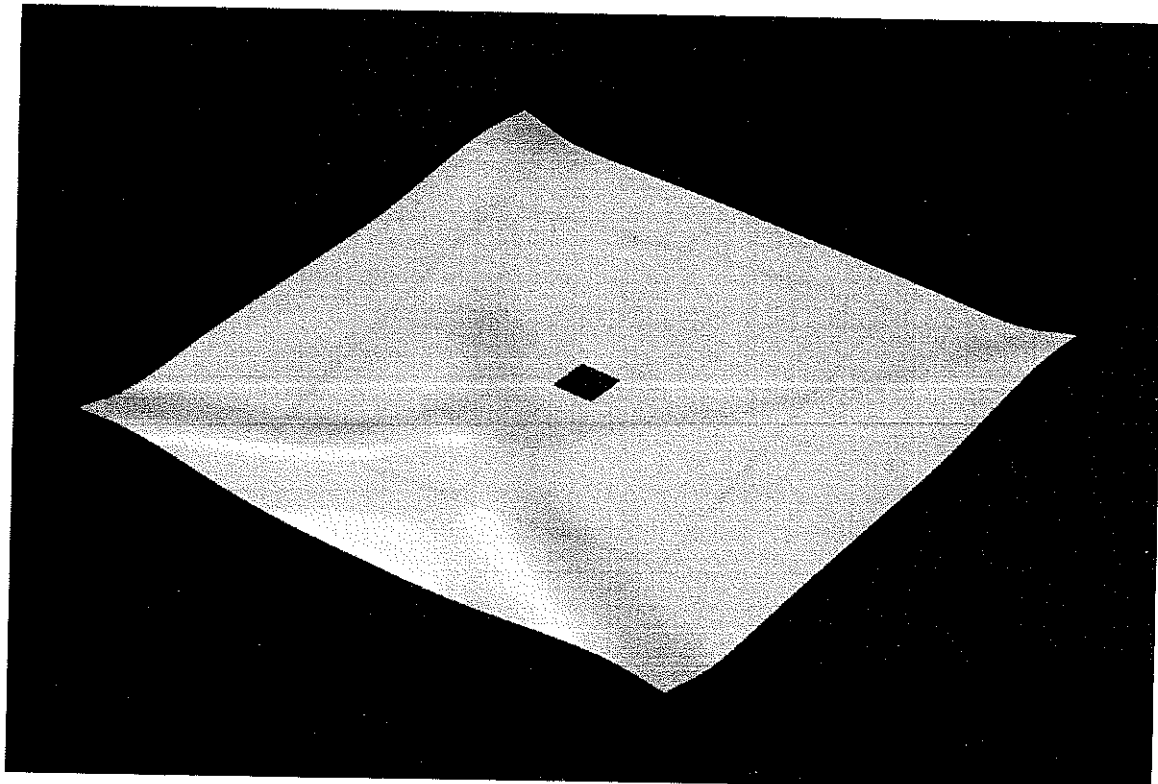


Figure 6.18 (second page of two). The reproduced sound field b) just before the pulse reaches the target area, and c) a while after it has passed through the target area



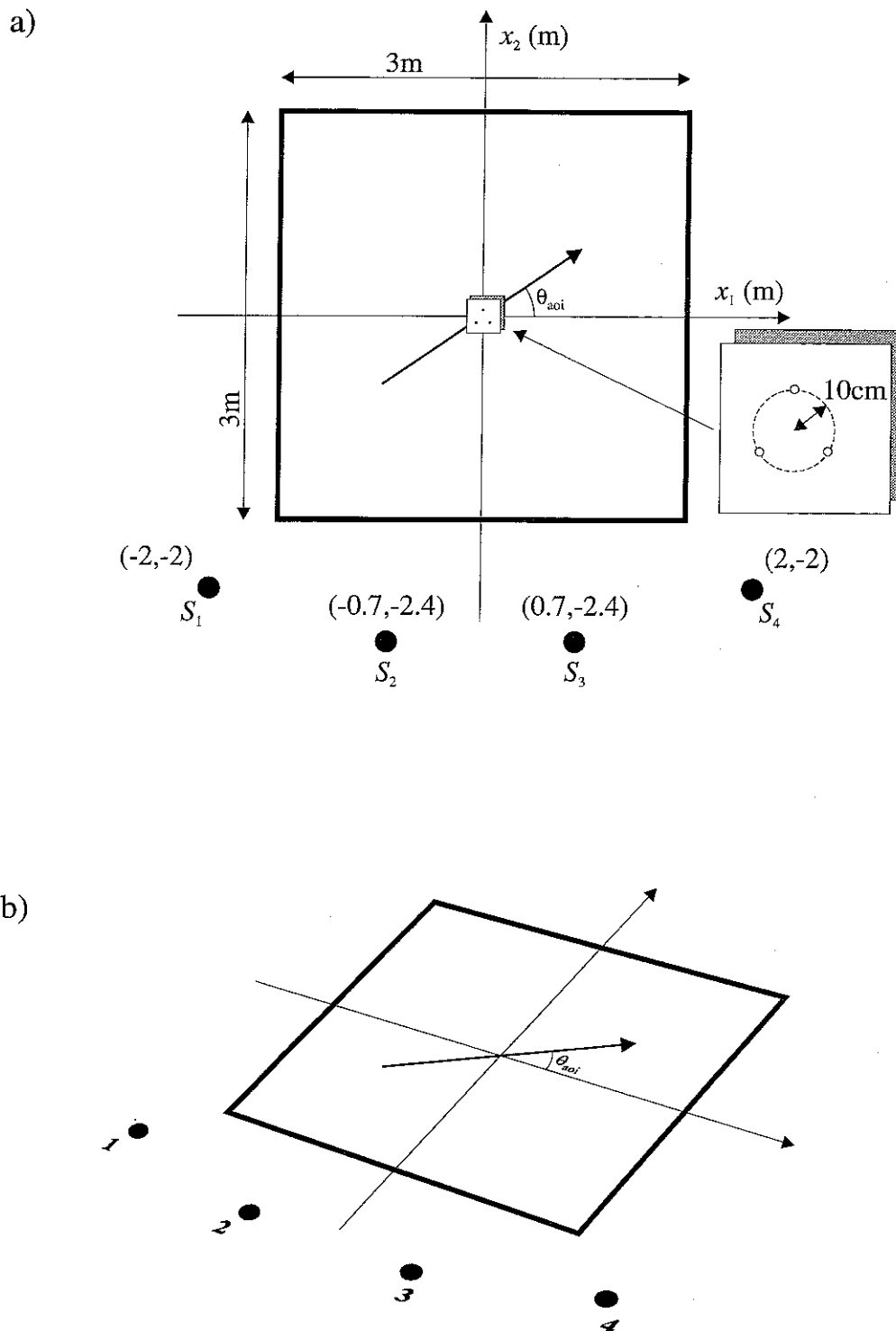


Figure 6.19. The second example of a geometry used for the simulation of a transient response. a) is viewed from directly above, b) is viewed from the same position as used for the frames shown in Figures 6.20, 6.22, and 6.24

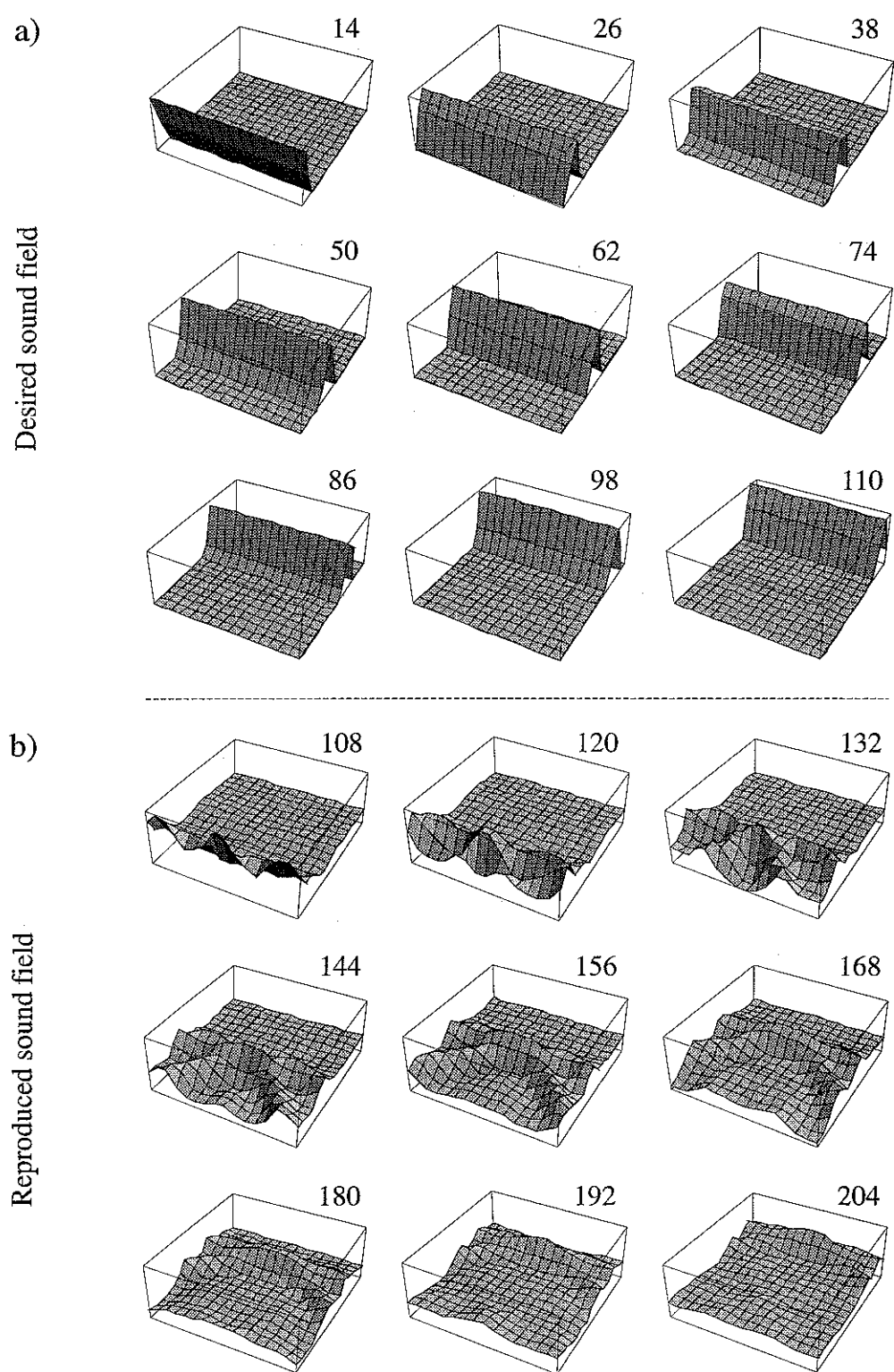


Figure 6.20.  $2 \times 9$  frames derived from the geometry shown in Figure 6.19. a) shows the desired plane wave Hanning pulse of duration 25 samples whose angle of incidence is  $90^\circ$ , and b) shows the reproduced sound field. The number above each frame is the frame index

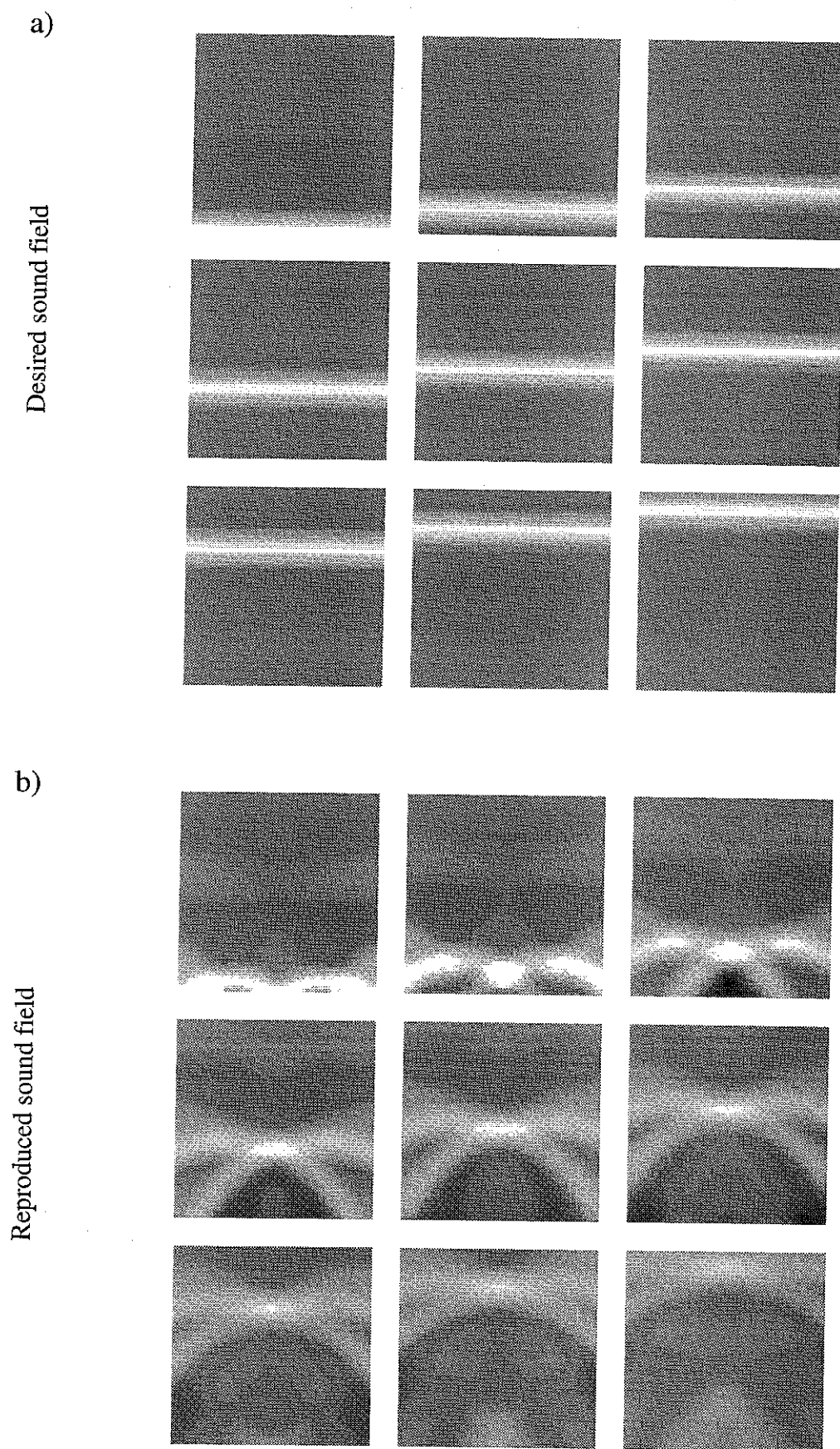


Figure 6.21. Shaded density plots of the  $2 \times 9$  frames shown in Figure 6.20

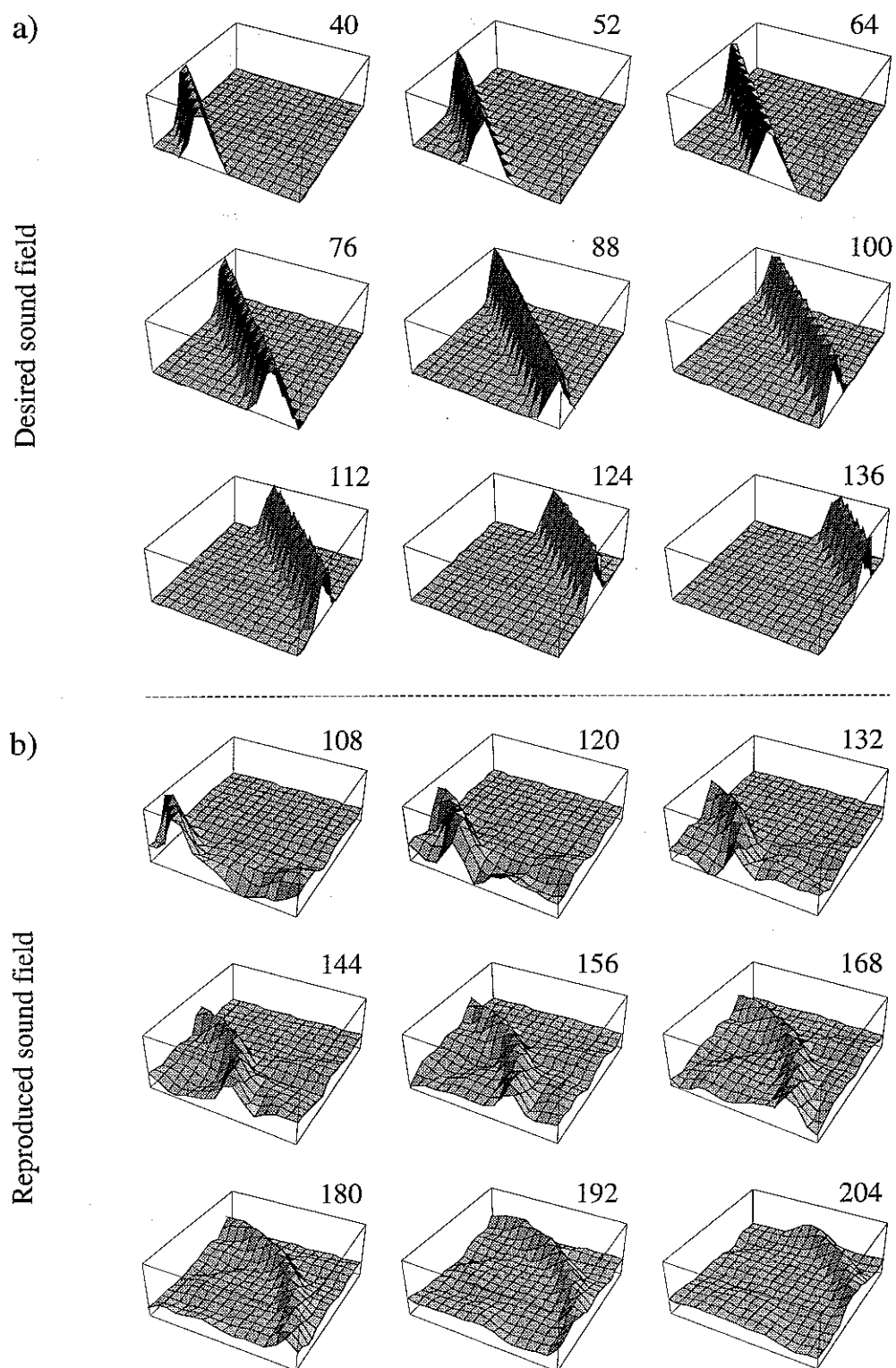
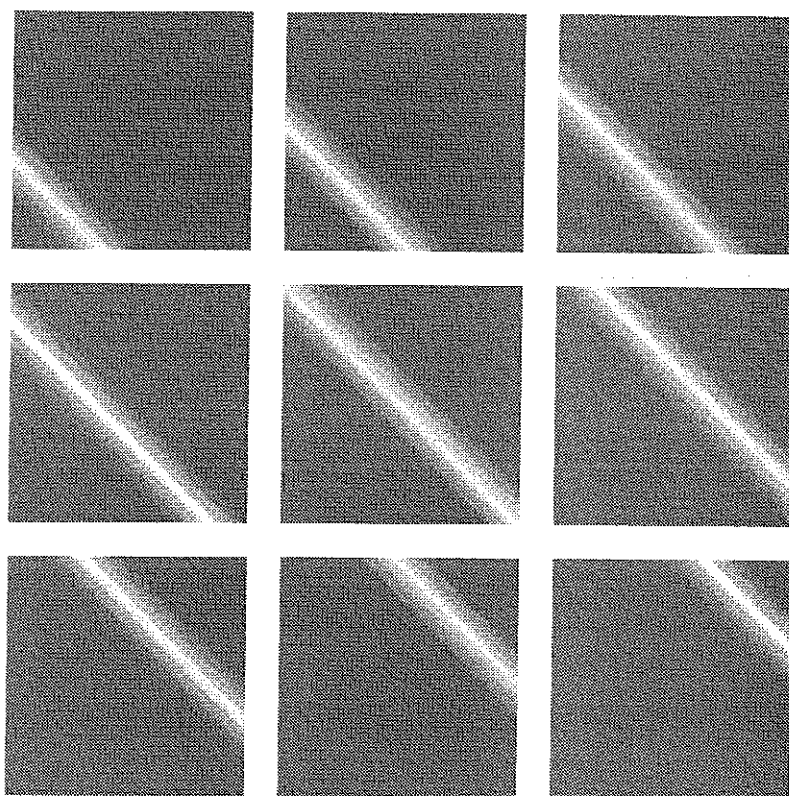


Figure 6.22.  $2 \times 9$  frames derived from the geometry shown in Figure 6.19. a) shows the desired plane wave Hanning pulse of duration 25 samples whose angle of incidence is  $45^\circ$ , and b) shows the reproduced sound field. The number above each frame is the frame index

a)

Desired sound field



b)

Reproduced sound field

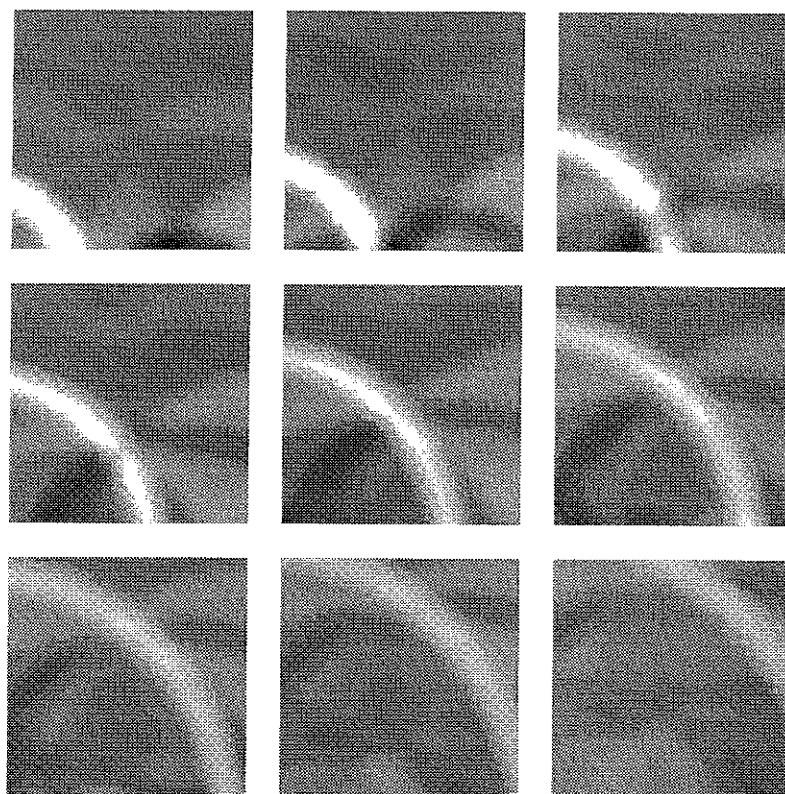


Figure 6.23. Shaded density plots of the  $2 \times 9$  frames shown in Figure 6.22

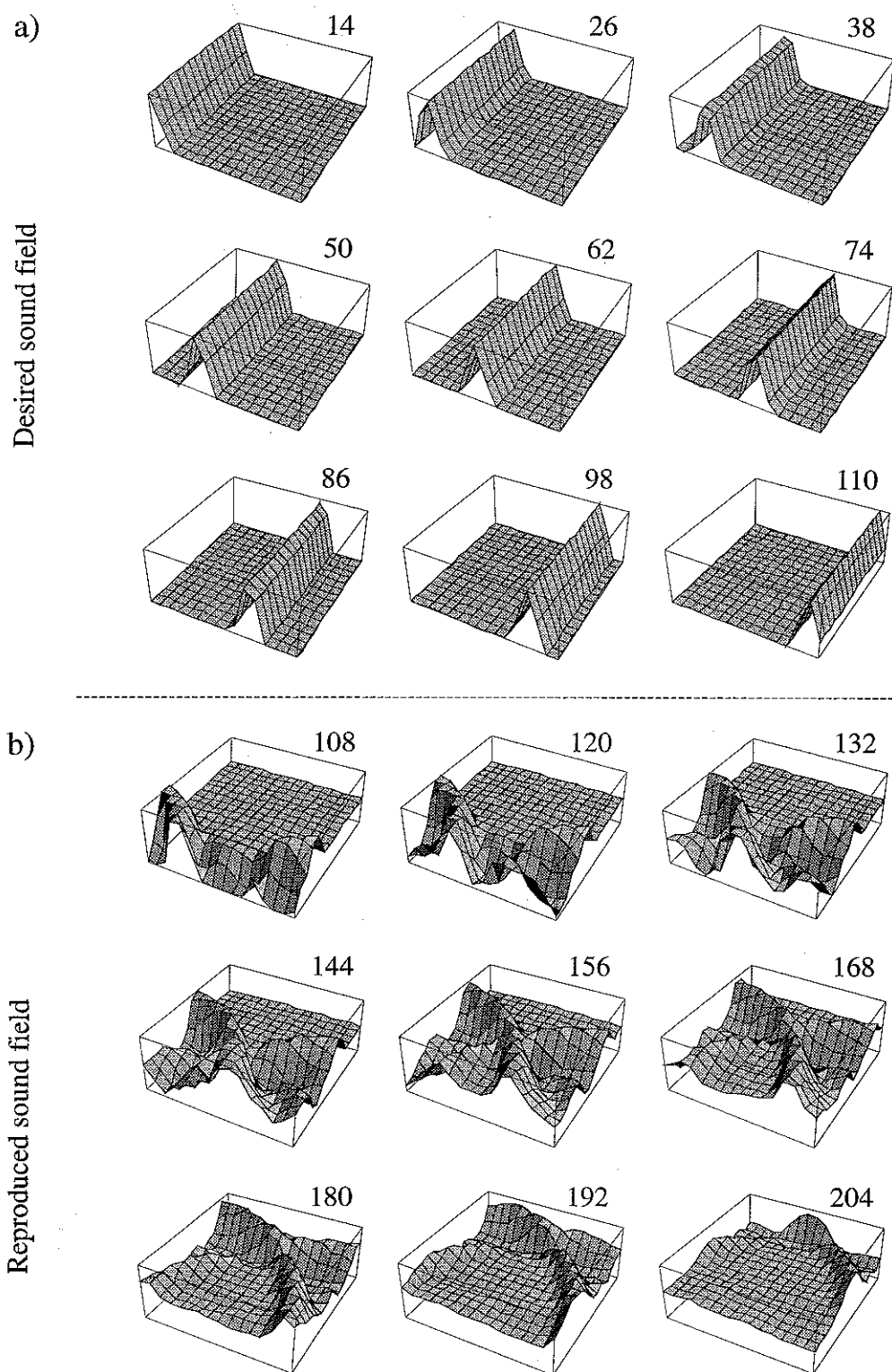
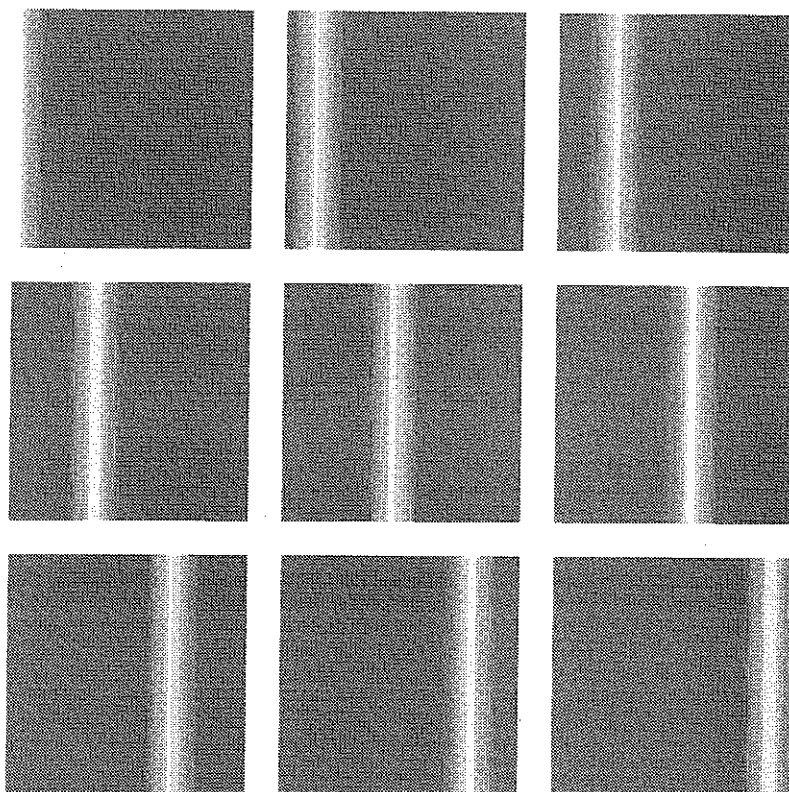


Figure 6.24.  $2 \times 9$  frames derived from the geometry shown in Figure 6.19. a) shows the desired plane wave Hanning pulse of duration 25 samples whose angle of incidence is  $0^\circ$ , and b) shows the reproduced sound field. The number above each frame is the frame index. Note the bend in the wavefront at the centre of frame 156 in b)



a)

Desired sound field



b)

Reproduced sound field

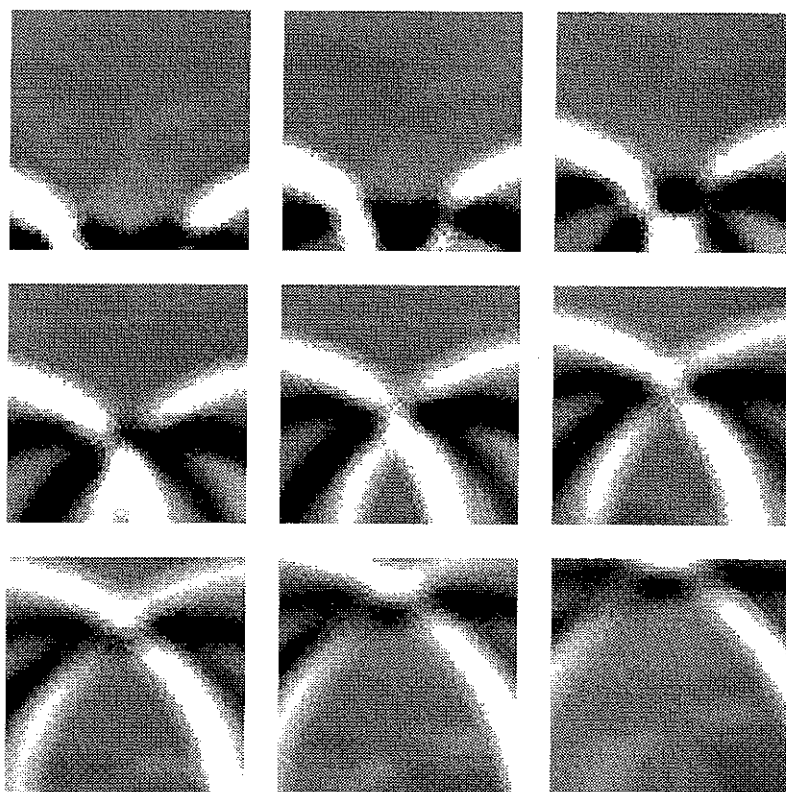


Figure 6.25. Shaded density plots of the  $2 \times 9$  frames shown in Figure 6.24

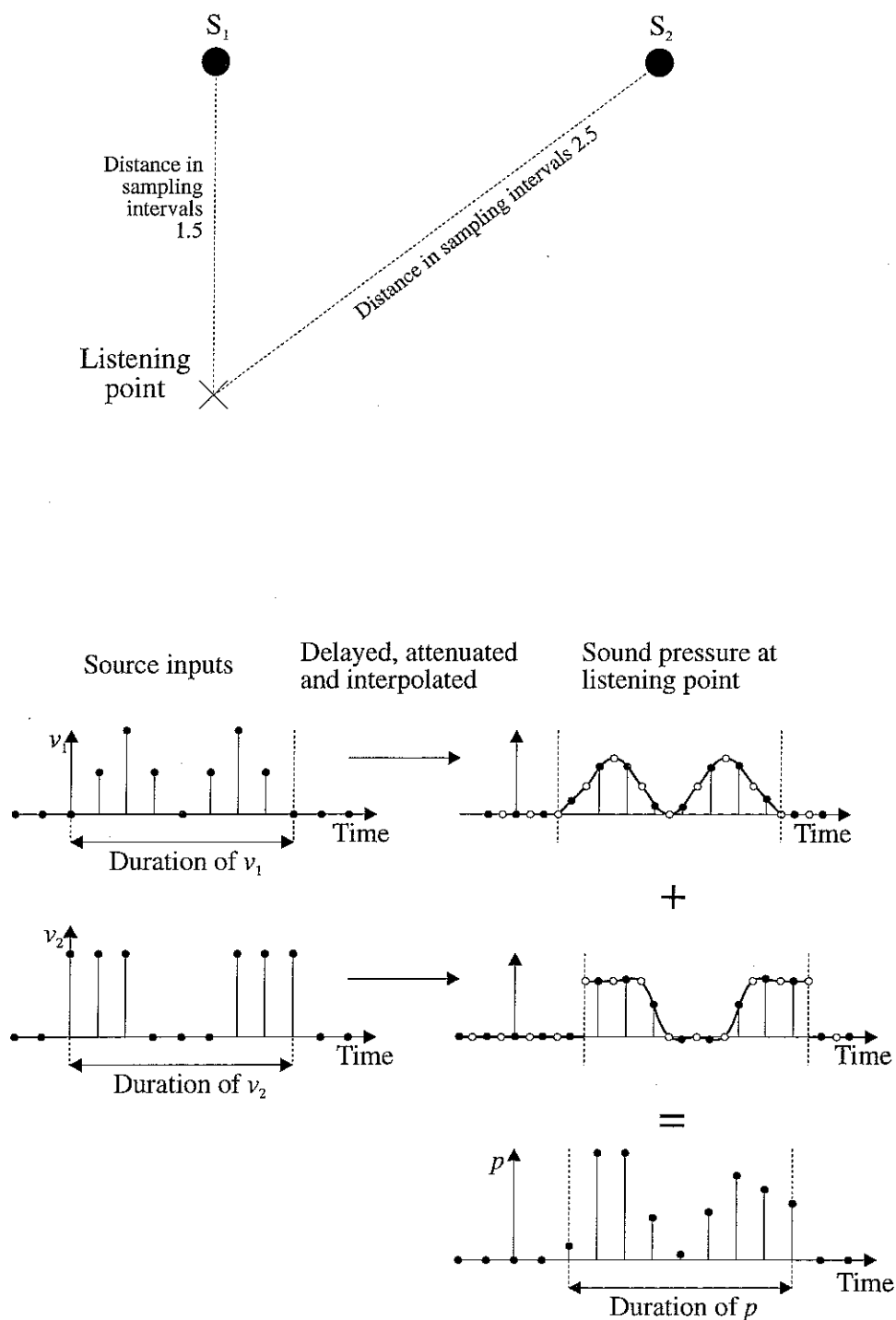


Figure 6.26. An example showing how to calculate the resulting sound pressure at a listening point by interpolating between the sampling times



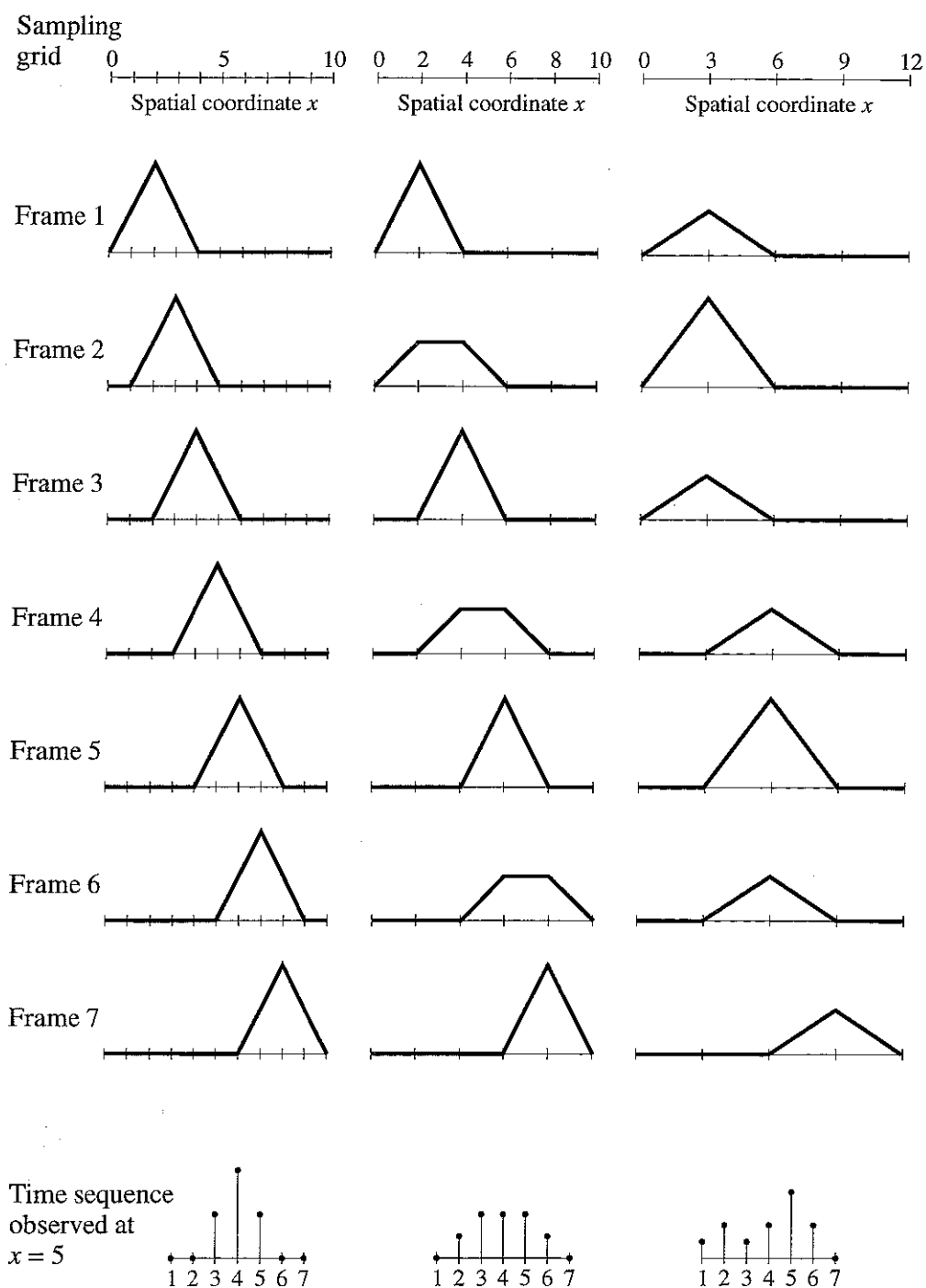


Figure 6.27. An example showing the influence on the linearly interpolated signal of the spacing between spatial sampling points

## 7. Experimental results

It is evident that the assumptions used throughout this work are unrealistic. In practice, loudspeakers do not behave like perfect monopoles, microphones are not ideal, and listening rooms are not anechoic. The results of two set of experiments made in an anechoic chamber are now compared to the simple numerical model developed in the previous chapters. For the first set of experiments, three microphones are used, for the second set, four microphones are used. The loudspeaker setup is the same for both microphone arrays, four loudspeakers are positioned in a  $4\text{m} \times 4\text{m}$  square. This is equivalent to the “quadraphony” setup described in Chapter 3. In each set of experiments, two sets of desired signals are measured, one set for an angle of incidence corresponding to the position of a loudspeaker, the other set for an angle of incidence corresponding to a position right in the middle between two of the loudspeakers. These desired signals are then passed through an inverse filter matrix, and the inputs to the loudspeakers are analyzed. Ideally, the input to the loudspeaker that is nearest to the direction of the source that generated the desired sound field must be the largest; if not, there is too much “cross-talk”. For each set of desired signals, two sets of inverse filters are used; one calculated from the measured electro-acoustic transfer functions, the other calculated from modelled electro-acoustic transfer functions. As mentioned in Chapter 4, we expect the system based on modelled data to be easier to invert than the system based on measured data.

### 7.1 Measured data

The two layouts of loudspeakers and microphones are shown in Figure 7.1. The first microphone array is a 2-by-2 array with a relatively wide microphone spacing of 10cm, the second array is a more compact circular 3 array with a radius of 3cm. The microphones are of the electret type, mounted in the top of an aluminium tube whose radius and height are 0.7cm and 9cm respectively. The loudspeaker layout is the same for the two microphone arrays: there are four loudspeakers in a  $4\text{m} \times 4\text{m}$  square. The measurements are made in an anechoic chamber with the Yamaha YDAP signal processing system (see Appendix 7.1). The YDAP is capable of measuring four sets

of very long impulse responses (more than 100,000 coefficients in each) simultaneously at a sampling frequency of 48kHz, so rather than using four different loudspeakers, a single KEF 101 loudspeaker is moved between each set of measurements. The YDAP controls a Yamaha DA8X digital-to-analog converter and a Yamaha AD8X analog-to-digital converter. The setup for both experiments are shown in Figure 7.2. The YDAP system can also implement a four-by-four FIR filter matrix, whose elements can contain up to 64,000 coefficients each, at a real time sampling rate of 48kHz. This makes it possible to make subjective tests on systems comprising up to four loudspeakers and four microphones.

For the 2-by-2 microphone array, there are 16 measured electro-acoustic transfer functions. For the circular 3 array, there are 12. The first 1024 coefficients of these 28 impulse responses are measured at 48kHz, with the results shown in "matrixform" in Figure 7.3. The performance of the two systems is assessed by using two sets of measured impulse responses as the recorded signals. The purpose is to confirm that, in practice, the relative magnitudes of the inputs to the loudspeakers correspond to the direction from which the recorded sound field originated. Two angles of incidence are used, one is  $\theta_{aoi} = -135^\circ$ , which corresponds to the position of loudspeaker number two, and  $\theta_{aoi} = -90^\circ$ , which corresponds to the direction in the middle between the positions of loudspeakers number one and two. The four sets of measured recorded signals are shown in Figure 7.4 (the initial delay does not correspond exactly to the distance between the loudspeakers and the microphones because the YDAP itself introduces an additional delay for unknown reasons).

In order to reduce the amount of measured data, all the measured impulse responses are decimated by a factor of four after being lowpass-filtered by a 30-coefficient FIR filter. This reduces the effective sampling frequency to 12kHz. The decimated impulse responses are then windowed by removing the initial delay and the "late" part of the ringing. Tohyama and Lyon [112] recommends to use an exponential weighting function when windowing time data, but this is not done here since the decimated impulse responses decay away quite quickly with time. Each of the impulse responses used for the analysis of the experimental data contains 128 coefficients. Figures 7.5 and 7.6 show the electro-acoustic transfer functions and the recorded signals after decimation and windowing of the original measured data shown in Figure 7.3 and 7.4.

## 7.2 Inverse filters calculated by inversion of measured transfer functions

The two sets of impulse responses shown in Figure 7.5 are deconvolved by an adaptive steepest descent algorithm trained with a deterministic impulse (described in Orduna-Bustamente [36]). The inverse filters, which each contain 256 coefficients, are trained with 100 "block impulse" inputs using a modelling delay  $m$  of 128 (for the 2-by-2 microphone array, the convergence coefficient  $\mu$  is 0.4, for the circular 3 array,  $\mu$  is 0.5). Figure 7.7 shows, in the time domain for the 2-by-2 microphone array, a) the 16 inverse filters  $h_n(n)$ , and b) the 16 "reproduced" signals  $w_n(n)$  that are calculated by convolving  $\mathbf{C}$  and  $\mathbf{H}$ . The impulse response  $w_n(n)$  is essentially an approximation to the element  $a_n(n)$  in the target matrix, and so Figure 7.7b indicates how well the inverse filters succeed in deconvolving the electro-acoustic transfer functions. Figure 7.8 shows the frequency responses a)  $|C_{11}|$  and b)  $|H_{11}|$ , and Figure 7.8c shows the frequency responses  $|W_{11}|$  and  $|W_{21}|$ .  $|W_{11}|$  is the contribution from track one to the signal reproduced at microphone one. This contribution is ideally 0dB because it corresponds to a diagonal element of the target matrix.  $|W_{21}|$  is the contribution from track one to the signal reproduced at microphone two. This contribution is ideally zero (minus infinity dBs) because it corresponds to an off-diagonal element of the target matrix. Figure 7.9 shows the frequency responses of the two sets of inputs to the four loudspeakers when the two sets of measured recorded signals are passed through the inverse filters shown in Figure 7.7a. Figures 7.10, 7.11, and 7.12 show the results for the circular 3 array that are equivalent to the results shown in Figures 7.7, 7.8 and 7.9 for the 2-by-2 array.

By comparing Figures 7.8c and 7.11c, it is seen that the deconvolution of the circular three array is overall more successful than for the 2-by-2 array. The 2-by-2 array starts to suffer from the effects of spatial aliasing at about 2kHz. This is indicated by a decrease of  $|W_{11}|$  and an increase in  $|W_{21}|$ . Since the microphones are closer together in the circular 3 array, spatial aliasing is not a problem at frequencies below 6kHz. Furthermore, since the number of loudspeakers is one greater than the number of microphones, we would expect the system to be easier to deconvolve even if the distance between adjacent microphones had been greater. In terms of direction-of-

arrival reproduction, however, the 2-by-2 microphone array seems to be better than the circular 3 array since it results in less "cross-talk" within its restricted working frequency range. When  $\theta_{aoi}$  is  $-135^\circ$ , the input  $|V_2|$  to loudspeaker number two should ideally be significantly greater than  $|V_1|$ ,  $|V_3|$ , and  $|V_4|$ . When  $\theta_{aoi}$  is  $-90^\circ$ , the inputs  $|V_1|$  and  $|V_2|$  should ideally be identical and significantly greater than  $|V_3|$  and  $|V_4|$ . Figure 7.9 indicates that the 2-by-2 array works well up to about 2kHz when  $\theta_{aoi}$  is  $-90^\circ$ , and up to about 2.5kHz when  $\theta_{aoi}$  is  $-135^\circ$ . The working frequency bandwidth of the system depends on the angle of incidence of the recorded sound field because the spatial sampling frequency as "seen" by the incident sound field is higher when the sound travels diagonally across the array (the distance between adjacent microphones is perceived to be 7cm) than when the sound travels straight across the array (the distance between adjacent microphones is perceived to be 10cm). Within the working frequency bandwidth,  $|V_2|$  is approximately 20dBs greater than  $|V_1|$ ,  $|V_3|$ , and  $|V_4|$  when  $\theta_{aoi}$  is  $-135^\circ$ , and  $|V_1|$  and  $|V_2|$  are approximately 10dBs greater than  $|V_3|$  and  $|V_4|$  when  $\theta_{aoi}$  is  $-90^\circ$ . For the circular 3 array, the results are qualitatively the same but the difference between the source inputs are not as large as for the 2-by-2 array. Note that for this particular geometry, the system, surprisingly, works better in terms of direction-of-arrival reproduction when  $\theta_{aoi}$  is  $-90^\circ$  than when  $\theta_{aoi}$  is  $-135^\circ$ .

### 7.3 Inverse filters calculated by inversion of modelled transfer functions

The performance of the inverse filters designed from experimental data is now compared to the performance of a set of inverse filters designed from electro-acoustic transfer functions that are modelled using the linear interpolation technique. Each of the inverse filters contain 120 coefficients, the modelling delay is 60, and the regularisation parameter is  $10^{-6}$ . Figures 7.13, 7.14, and 7.15 show the results for the 2-by-2 microphone array that are equivalent to those shown in Figures 7.7, 7.8 and 7.9 respectively. Figures 7.16, 7.17, and 7.18 show the results for the circular 3 microphone array that are equivalent to those shown in Figures 7.10, 7.11 and 7.12 respectively. It is seen that the results derived from modelled electro-acoustic transfer functions are qualitatively very similar to those derived from measured electro-acoustic transfer functions. The deconvolution of the circular 3 array is more

successful than for the 2-by-2 array because spatial aliasing is not a problem for the circular 3 system for frequencies below 6kHz. As illustrated in Figure 7.14b, for the 2-by-2 array there is a dip in  $|H_{11}|$  at the frequencies 0Hz, 2.4kHz and 4.8kHz. These frequencies correspond to the poles of the exact inverse that are close to the unit circle as illustrated in Figure 6.8f. It is surprising that at low frequencies the deconvolution of measured transfer functions works almost as well as the deconvolution of modelled transfer functions. The most notable difference between the results derived from modelled and measured data is that the frequency responses of the inverse filters and the loudspeaker inputs calculated from the modelled electro-acoustic transfer function oscillate less wildly. This happens because the inverse filters do not have to compensate for the frequency responses of the loudspeakers as well as realise the multi-channel deconvolution. As described in Section 4.3, the multi-channel deconvolution does not equalise the loudspeakers individually (see Equations (4.3.9) and (4.3.10)).

## 7.4 Conclusions

The results of the experiments confirm the expected limitations of the performance of the inverse filters. At very low frequencies, the ill-conditioning and the relatively short length of the inverse filters prevent the system from giving any output of significance. At high frequencies, it is quite easy to identify the frequency at which spatial aliasing sets in. The exact "spatial aliasing frequency" depends, in a non-trivial way, on both the spatial components of the recorded signals and the specific layout of loudspeakers and microphones. It is possible to obtain a good inversion of the matrix of electro-acoustic transfer functions even when the impulse responses of those transfer functions are measured and therefore of relatively long duration. The magnitude spectra of the loudspeaker inputs that are calculated by passing a set of measured recorded signals through the inverse filters are qualitatively similar regardless of whether the inverse filters are calculated from measured or modelled electro-acoustic transfer functions. Quantitatively, the power spectra of the loudspeaker inputs calculated from measured data oscillate more rapidly with frequency than the power spectra calculated from modelled data. Within the working frequency range of the system, the most of the input power to the loudspeakers goes to the ones that are

closest to the direction of the recorded sound source, this suggests that, as a simple direction-of-arrival reproduction system, the system seems to work in practice. It is encouraging that even with the use of cheap electret microphones, the system still behaves as expected, at least up to approximately 4kHz. However, at higher frequencies, the performance of the system is bound to get worse because of diffraction phenomena and the differences in both the phase and the amplitude response of the microphones.

## **Appendix 7.1. Making measurements with the YDAP**

This section gives the information that is absolutely vital for measuring the impulse response from one loudspeaker to four microphones. The "instruction manual" style is used for the benefit of future users of the equipment.

The YDAP, the AD8X, and the DA8X have all been supplied by YAMAHA in Japan, and they must be used only with the blue 100V AC power supply. Always check the mains connections before switching on any of the equipment. The switches on the back of the AD8X should be set to:

clk out	ON
mode	INT
internal sampling frequency	48kHz
emphasis	OFF
digital output	DMR8

The AD8X and DA8X are connected to the YDAP via the Digital In/Out connection. Now to the operation of the YDAP itself. Start up the system by first switching on the power supply, then push the power-switch on the back of the YDAP, and finally turn the key on the front of the YDAP. Give the user name dap, then select Irmec with the mouse. Return gives the options that are currently available and their settings, q moves up one directory. Go to the measurement menu and set the following parameters:

signal length 2048 (or whatever, this is in samples)  
signal interval 100,000 (or whatever, this is in samples)  
input select dsub

storage input si0 and si1

signal 3) time stretched pulse, 4) 32K (double, 4ch) (must  
be done last!)

Go to the generator menu, and use the load command to load the data files @tsp\_32k and @zero\_4k. Both files must be loaded, even though only one of them is actually used. Go to the storage menu and set trigger control, mode, to internal. Always make sure that trigger mode is internal before doing level adjust or averaging. If it isn't, the system hangs because it is waiting for a trigger it never receives, and the only way to recover is by rebooting. Go back to the measurement menu and set averaging times appropriately. Always do level adjust before a measurement unless you are confident that the input level is already matched to the internal scaling factor. Do not do level adjust if several sets of measurements need to be relative to a common reference. Use the wave monitor to look at the measured impulse responses, use the save menu to store the measurements on disk. To leave the Irmec menu, press control-z, then enter logout. Switch of the equipment in the reverse order it was switched on, so first turn the key on the front, push the power-switch on the back, and finally switch of the power supply.

Once the measurements are made, the data can be transferred from the YDAP to a SUN workstation over the RS232C interface using the program cp\_aux. Similarly, if a set of filters is to be implemented by the YDAP, it can be transferred from the SUN to the YDAP with the same program, cp\_aux. This program is run simultaneously in both the YDAP and the SUN. cp\_aux must always be run at the receiving end first. In order to transfer data from the YDAP to the SUN, change directory on the SUN to where the data is going to be stored. Enter the following command on the SUN:

```
cp_aux -p /dev/ttya -s 9600.
```

This sets the SUN in listening mode. Change directory on the YDAP to where the data is stored (irmec/data in the case of measured impulse responses). Enter the following command on the YDAP:

```
cp_aux <filename> -p /term -s 9600 -n.
```



The `-n` makes the program continue without quitting on the SUN after the transmission, and it should be omitted when transmitting the last filter. Transferring data from the SUN to the YDAP is very similar to the above procedure. Give the files on the SUN names of the type `h_1.tm4`, `h_2.tm4`....`h_g.tm4`. Using these names will later make it easier to load the filters in to the YDAP's working memory. Enter the following command on the YDAP:

```
cp_aux -p /term -s 9600.
```

This sets the YDAP in listening mode. Enter the following command on the SUN:

```
cp_aux <filename> -p /dev/ttya -s 9600 -n.
```

The `-n` should be omitted when transmitting the last filter.

The data received from the YDAP needs to be converted from YAMAHA's own binary format to text format, this data conversion is done on the SUN with the program `tmcvt`. From binary format to text format, the command is

```
tmcvt -N = 2048 -B -I -R -a -f -h -r source destination,
```

from text format to binary format, the command is

```
tmcvt -N = 2048 -A -F -H -R -b -i -r source destination
```

It is well worth the effort to create a few batch files to make the file transfer and conversion quicker to perform.

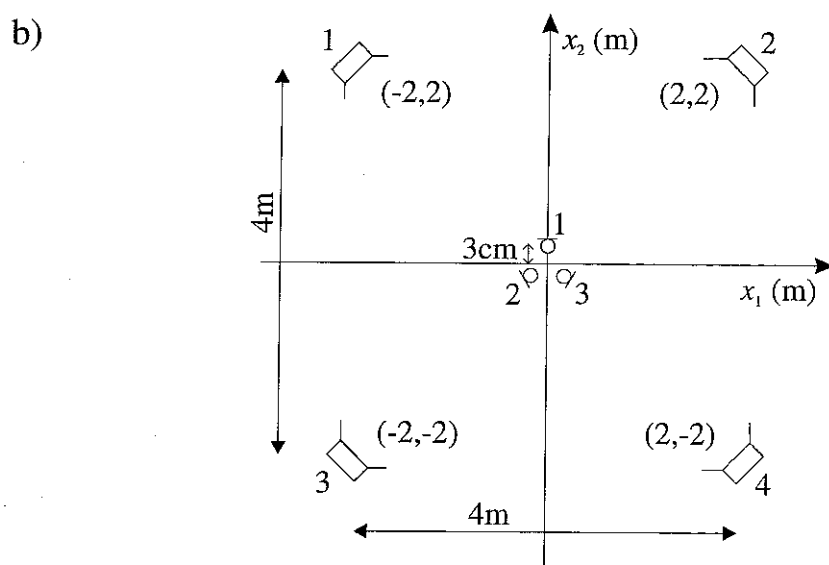
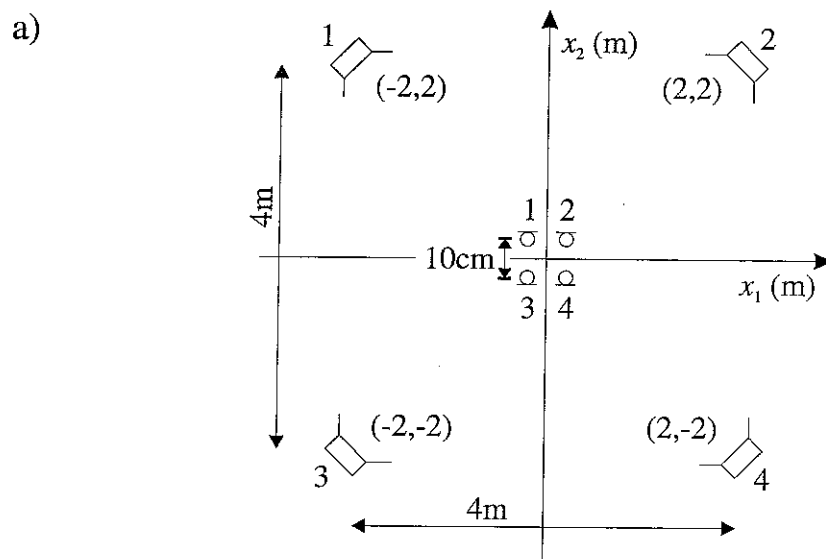


Figure 7.1. The two layouts of loudspeakers and microphones used for the experiments (microphone arrays not drawn to scale)

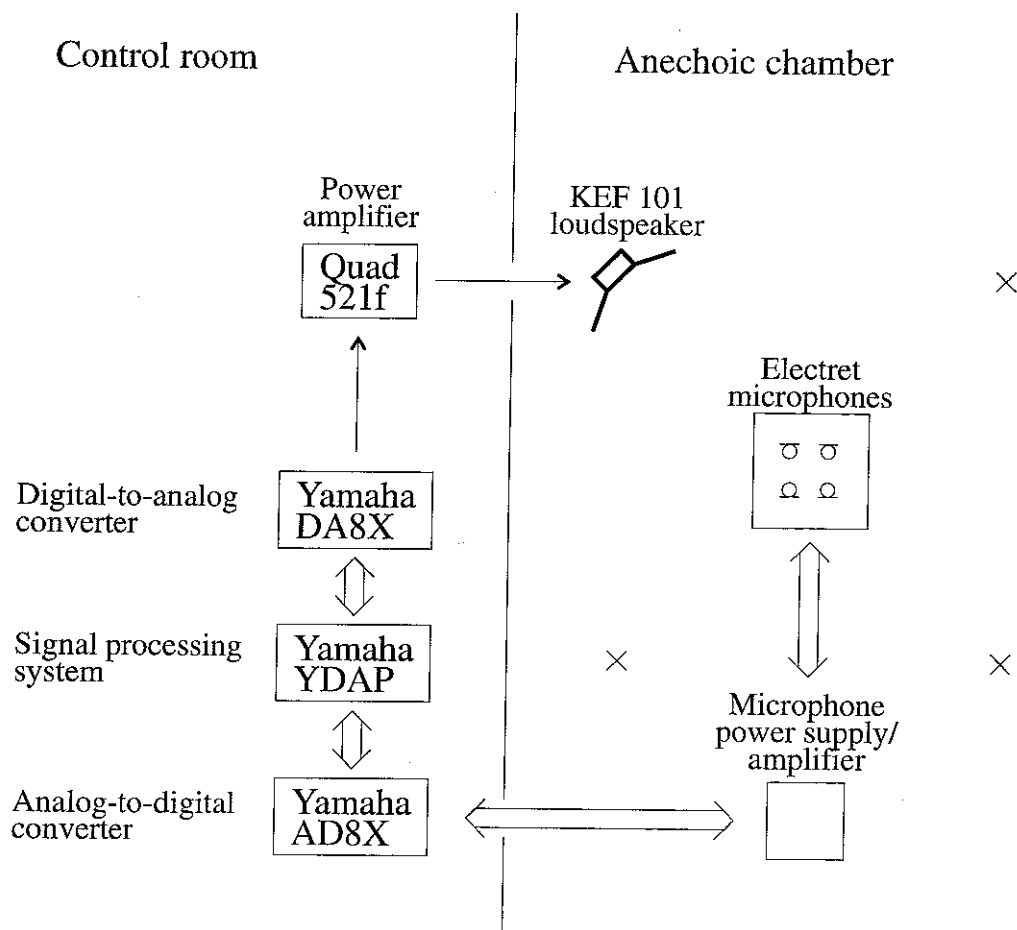


Figure 7.2. Schematic illustration of the experimental setup

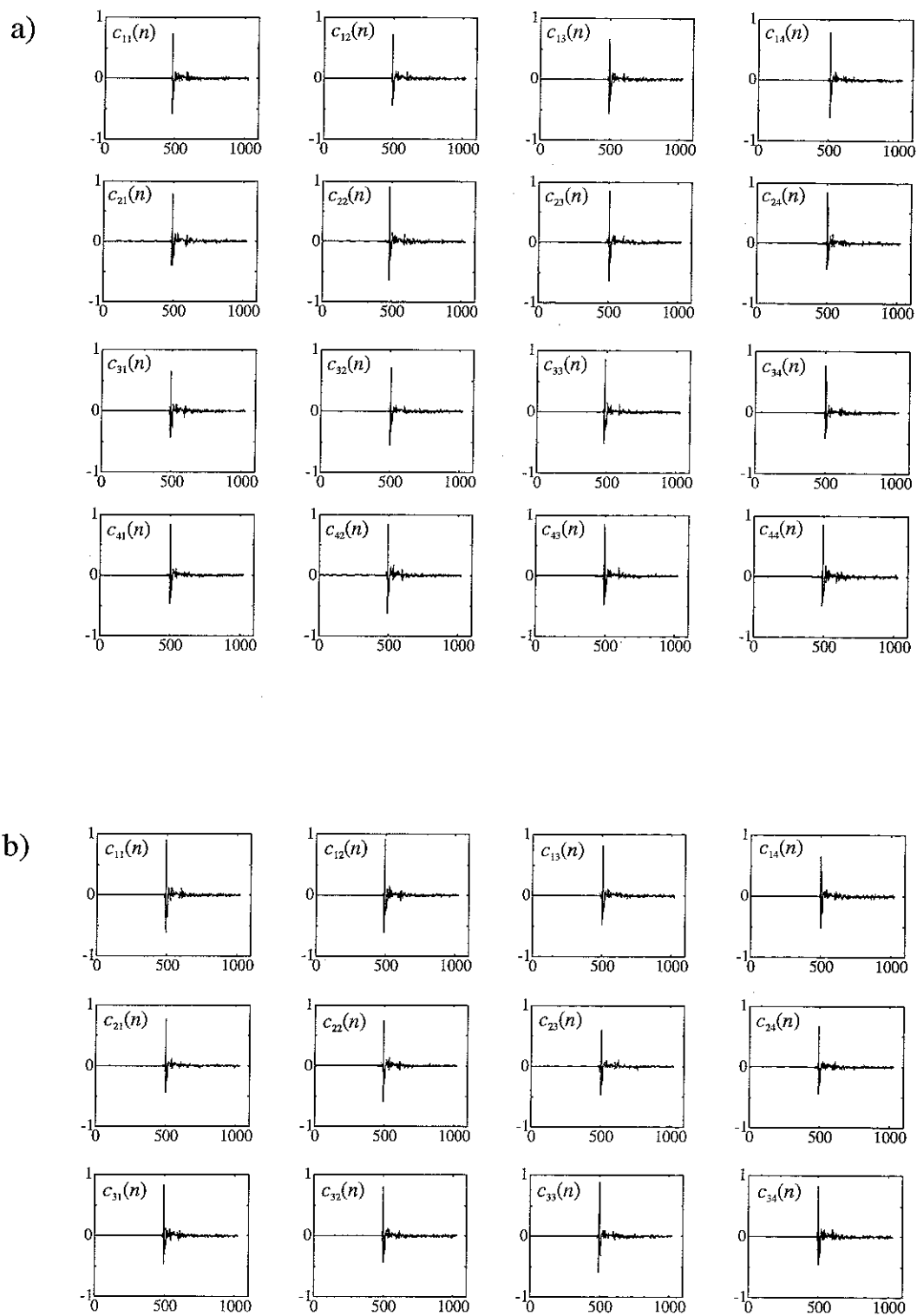
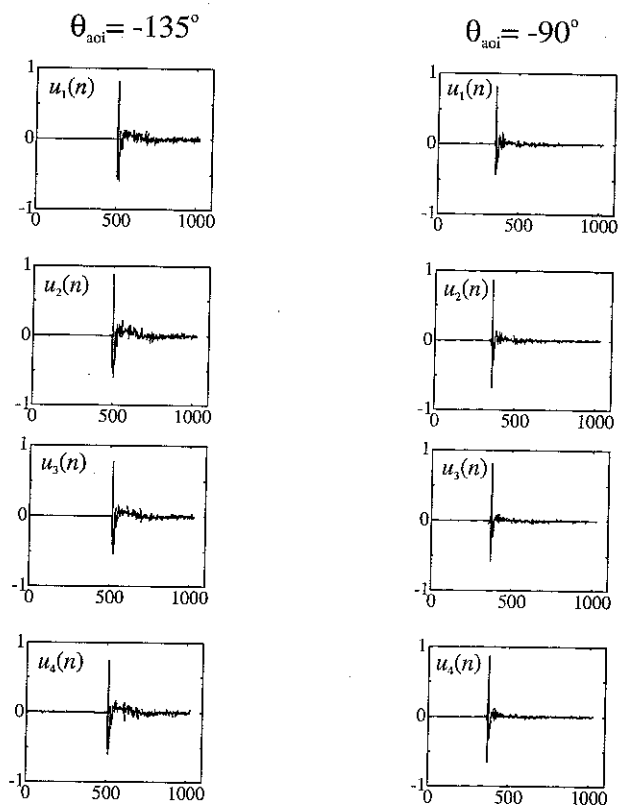


Figure 7.3. Measured electro-acoustic transfer functions for a) the 2-by-2 microphone array shown in Figure 7.1a and b) the circular 3 microphone setup shown in Figure 7.1b. The sampling frequency is 48kHz

a) 2-by-2



b) Circular 3

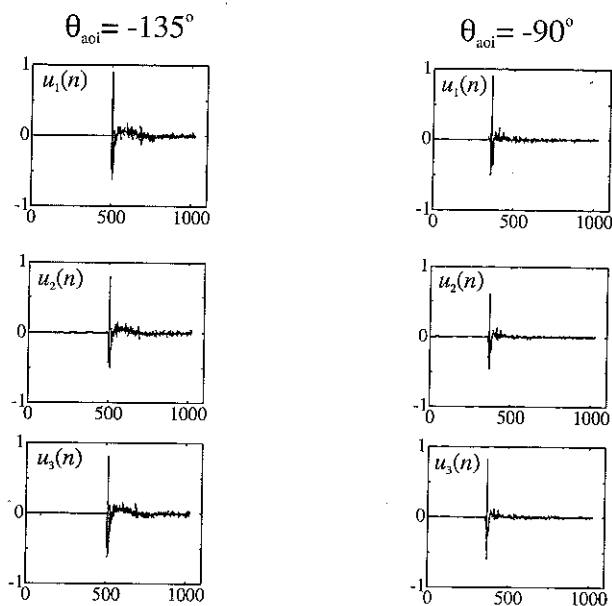


Figure 7.4. Measured recorded signals for two different angles of incidence a) for the 2-by-2 microphone array shown in Figure 7.1a, and b) for the circular 3 microphone array shown in Figure 7.1b. The sampling frequency is 48kHz

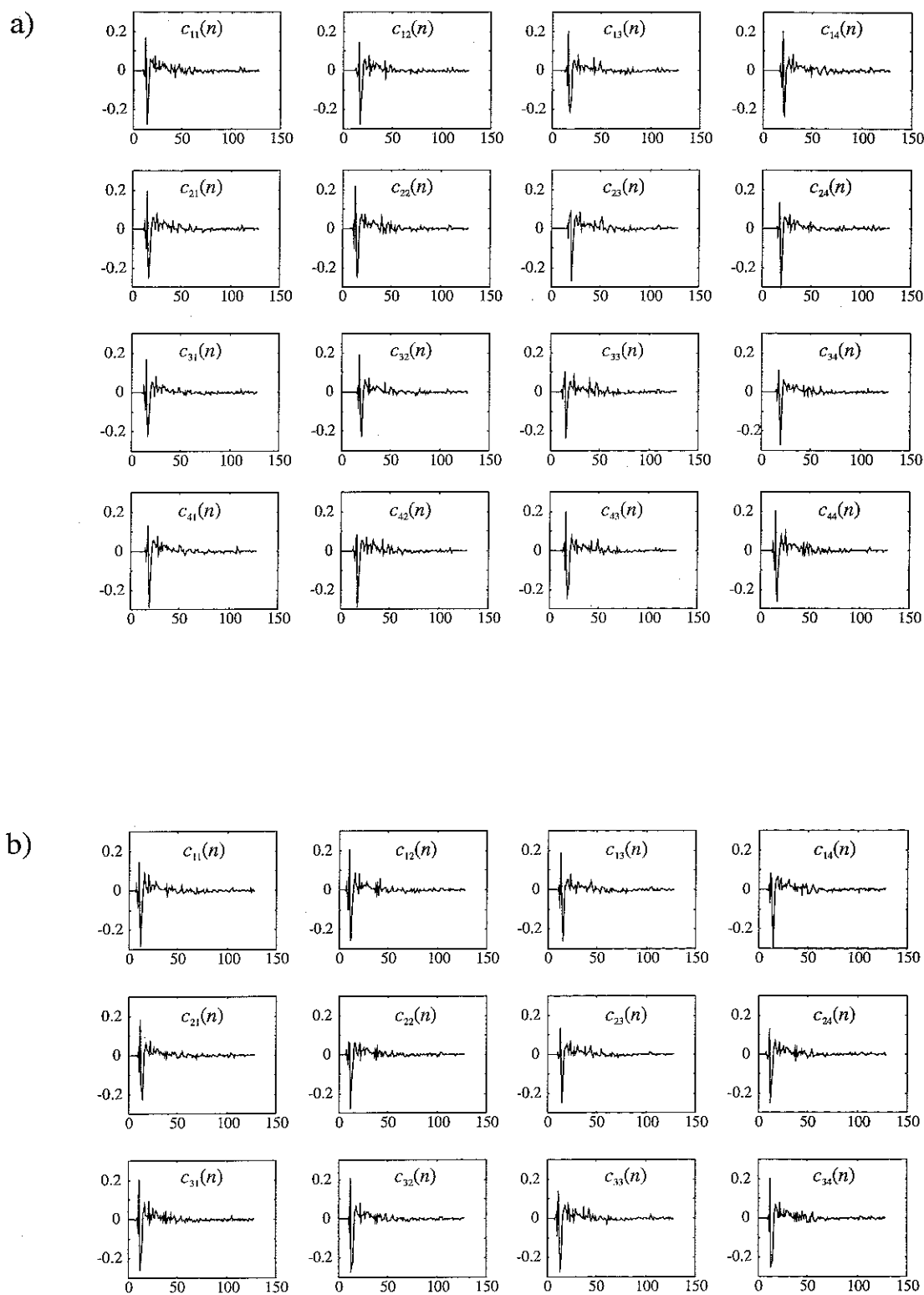
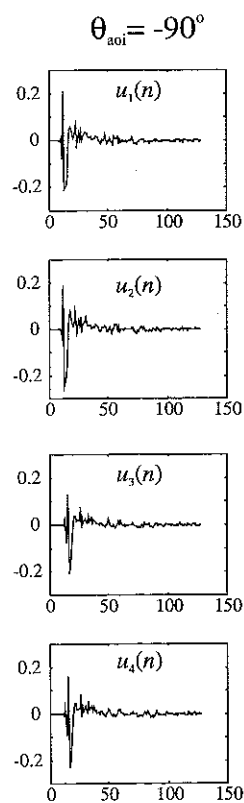
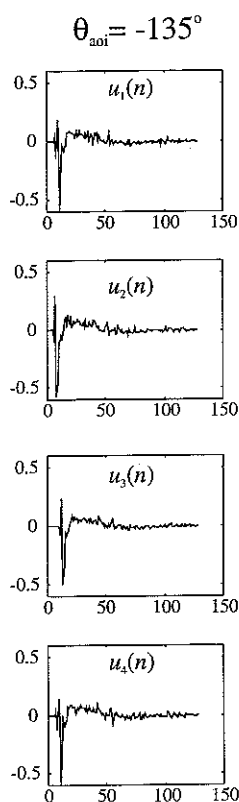


Figure 7.5. The measured impulse responses shown in Figure 7.3 after removal of a common initial delay and decimation by a factor of four. a) is for the 2-by-2 array, b) is for the circular 3 array. The sampling frequency is effectively 12kHz

a) 2-by-2



b) Circular 3

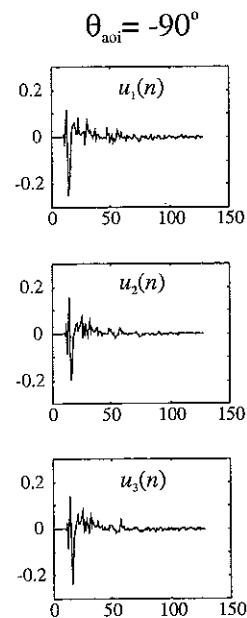
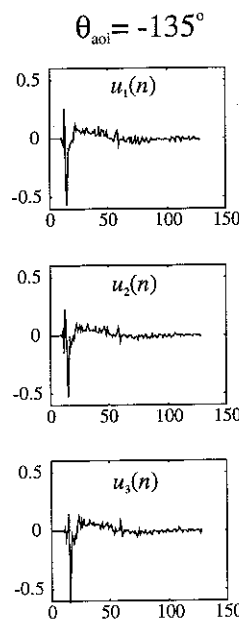


Figure 7.6. The measured recorded signals shown in Figure 7.4 after removal of an initial delay and decimation by a factor of four. a) is for the 2-by-2 array and b) is for the circular 3 array. The sampling frequency is effectively 12kHz

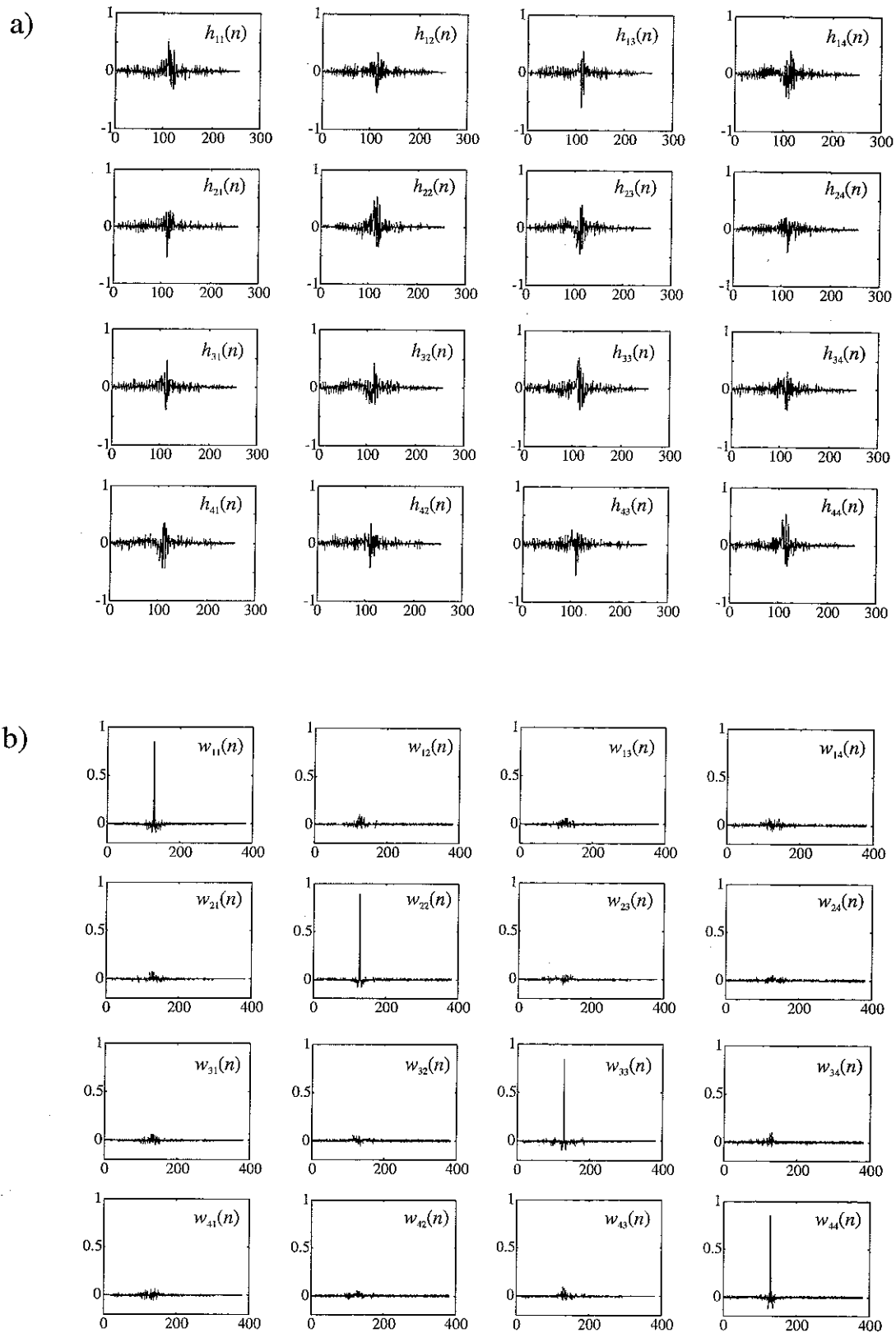


Figure 7.7. a) The 16 inverse filters in  $\mathbf{H}$  and b) the matrix  $\mathbf{W}$  of reproduced signals.  $\mathbf{W}$  is calculated by convolving  $\mathbf{C}$ , which is shown in Figure 7.5a, with  $\mathbf{H}$



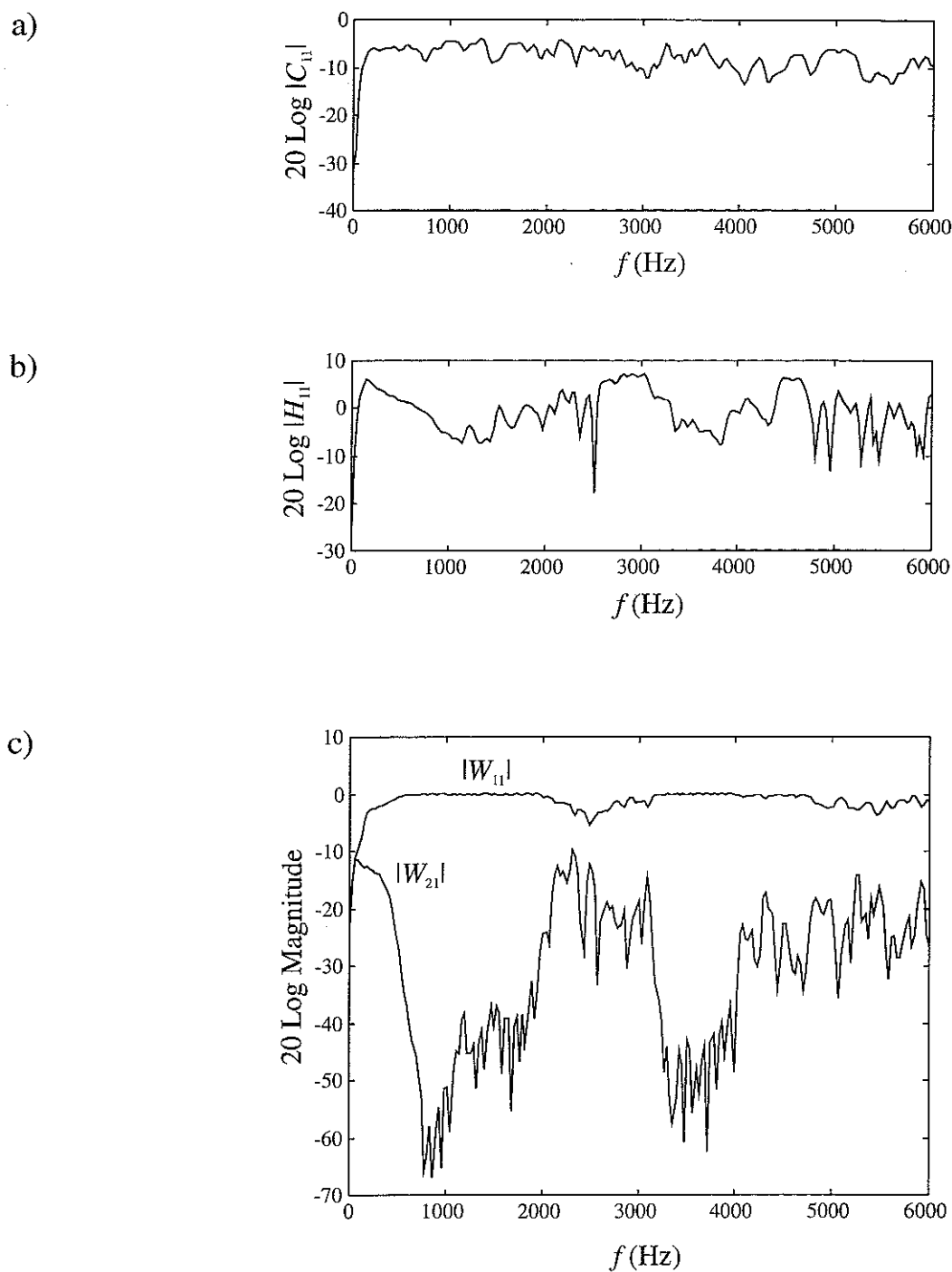
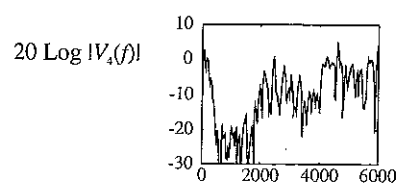
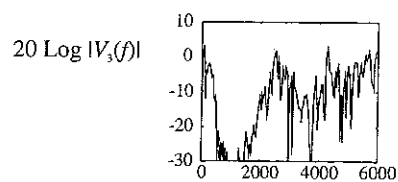
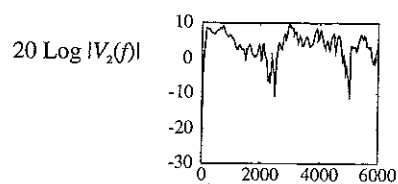
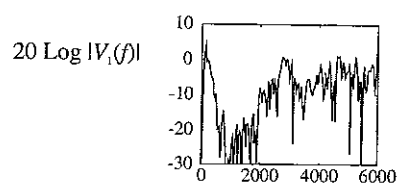
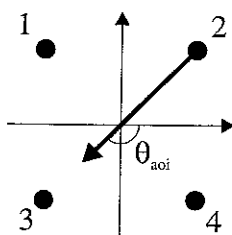


Figure 7.8. a) The frequency response  $|C_{11}|$  of  $c_{11}(n)$  which is shown in Figure 7.5a. b) The frequency response  $|H_{11}|$  of  $h_{11}(n)$  which is shown in Figure 7.7a. c) The frequency responses  $|W_{11}|$ , which is a diagonal element of  $\mathbf{C}^*\mathbf{H}$ , and  $|W_{21}|$ , which is an off-diagonal element of  $\mathbf{C}^*\mathbf{H}$ . The matrix  $\mathbf{C}^*\mathbf{H}$  is shown in Figure 7.7b

a)  $\theta_{aoi} = -135^\circ$



b)  $\theta_{aoi} = -90^\circ$

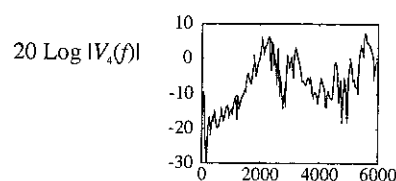
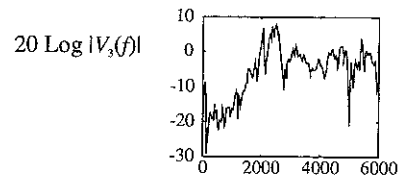
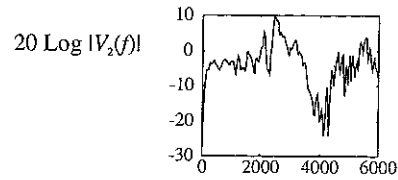
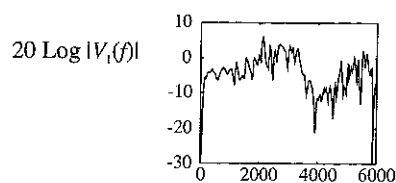
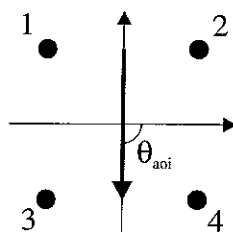
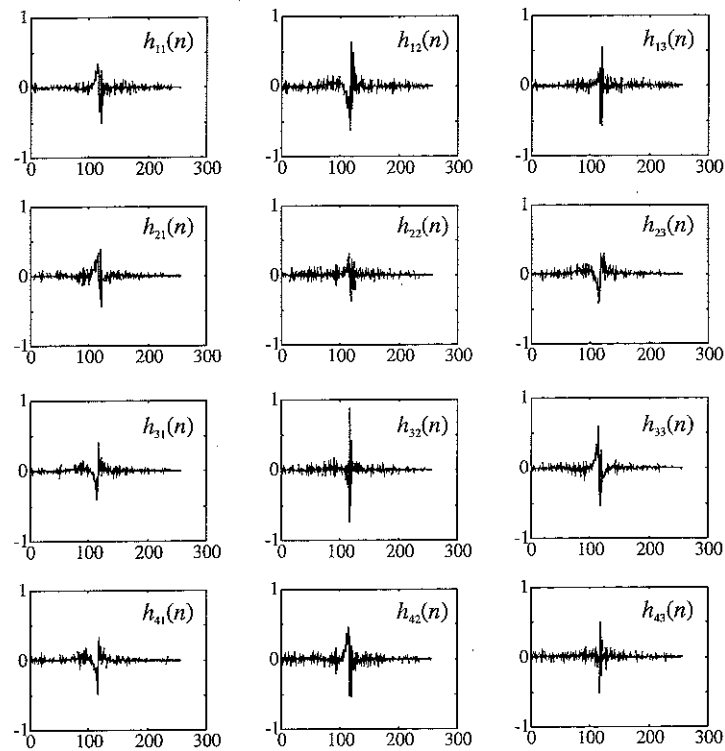


Figure 7.9. The input to the four loudspeakers when two different sets of 4 recorded signals, measured by the 2-by-2 microphone array shown in Figure 7.1a, are passed through the 16 inverse filters shown in Figure 7.7a

a)



b)

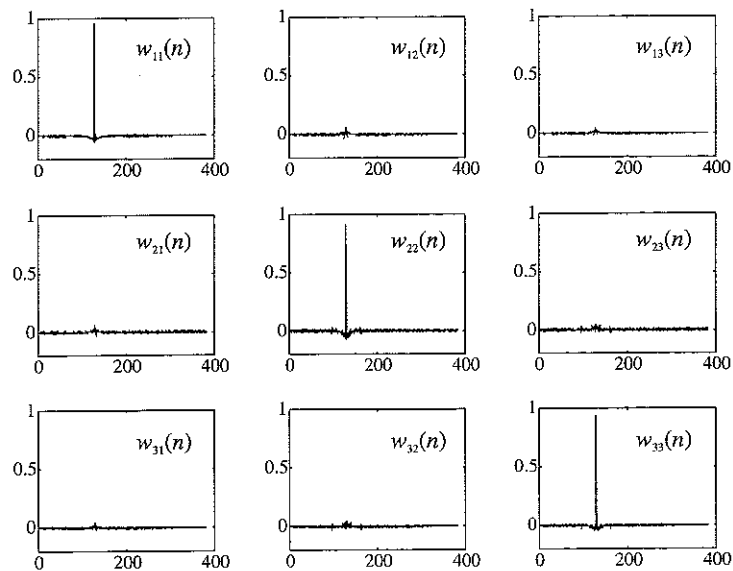


Figure 7.10. a) The 12 inverse filters in  $\mathbf{H}$  and b) the matrix  $\mathbf{W}$  of reproduced signals.  $\mathbf{W}$  is calculated by convolving  $\mathbf{C}$ , which is shown in Figure 7.5b, with  $\mathbf{H}$

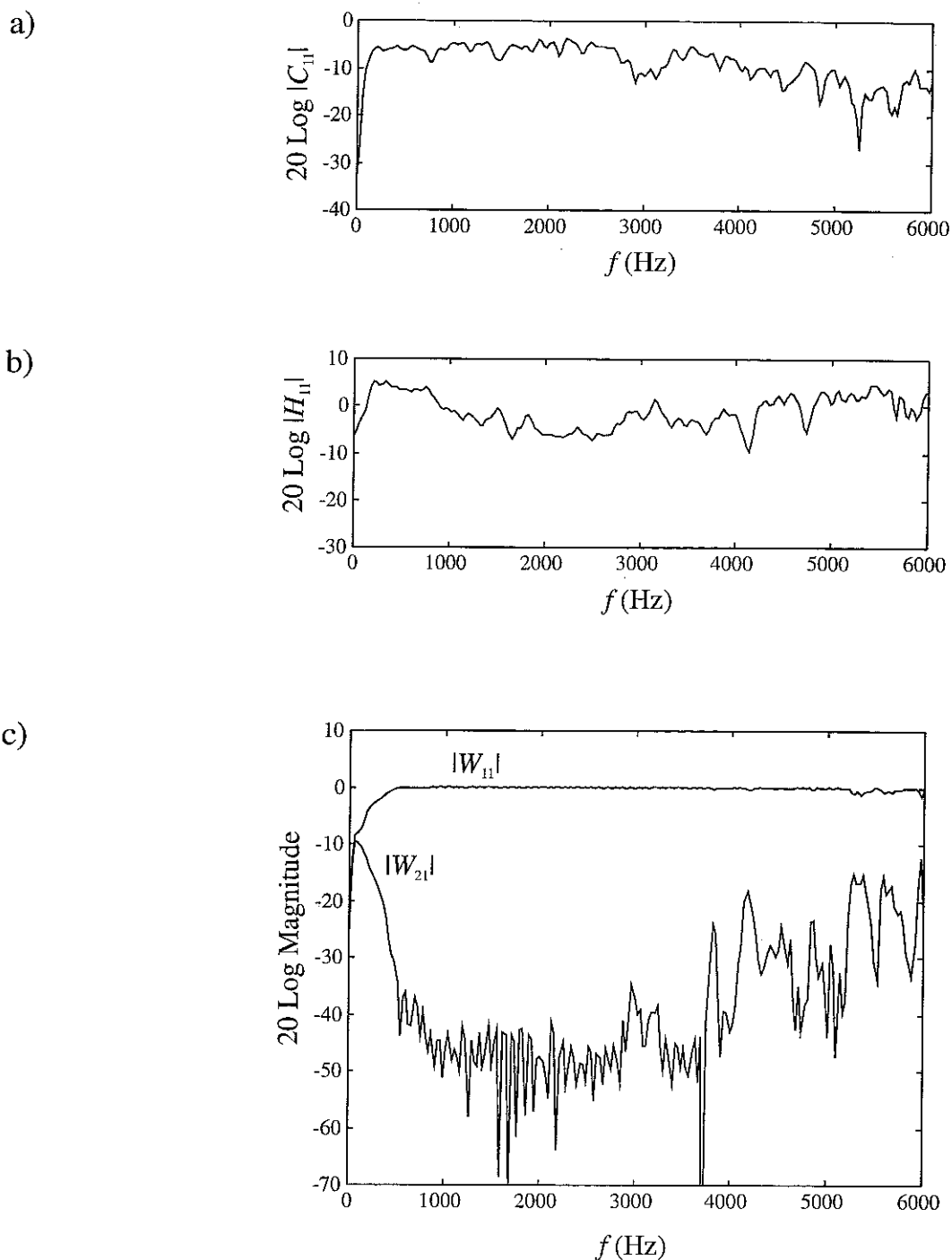
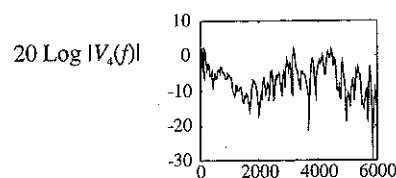
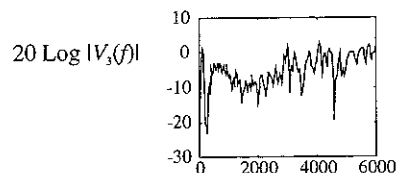
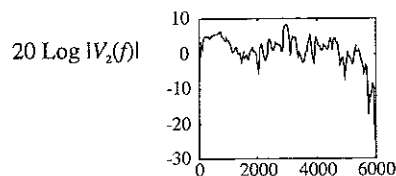
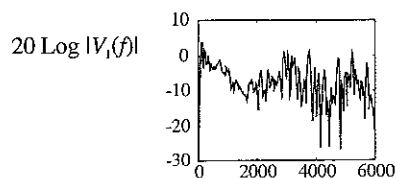
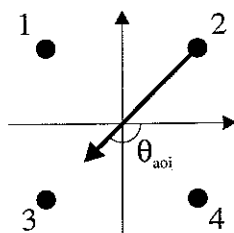


Figure 7.11. a) The frequency response  $|C_{11}|$  of  $c_{11}(n)$  which is shown in Figure 7.5b. b) The frequency response  $|H_{11}|$  of  $h_{11}(n)$  which is shown in Figure 7.10a. c) The frequency responses of  $|W_{11}|$ , which is a diagonal element of  $\mathbf{C}^*\mathbf{H}$ , and  $|W_{21}|$ , which is an off-diagonal element of  $\mathbf{C}^*\mathbf{H}$ . The matrix  $\mathbf{C}^*\mathbf{H}$  is shown in Figure 7.10b

a)  $\theta_{\text{aoi}} = -135^\circ$



b)  $\theta_{\text{aoi}} = -90^\circ$

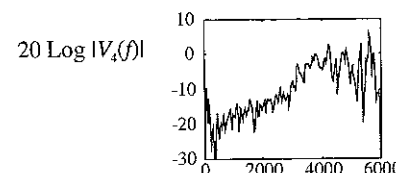
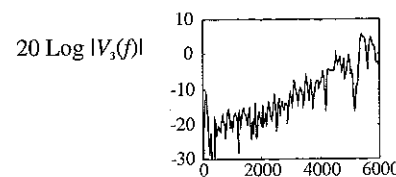
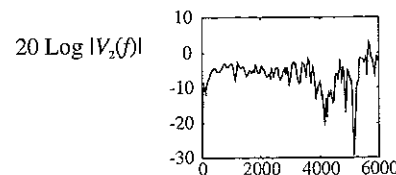
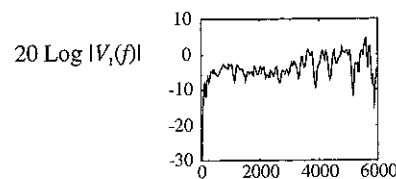
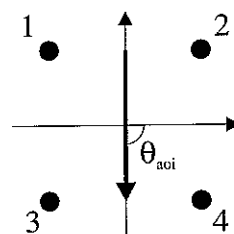
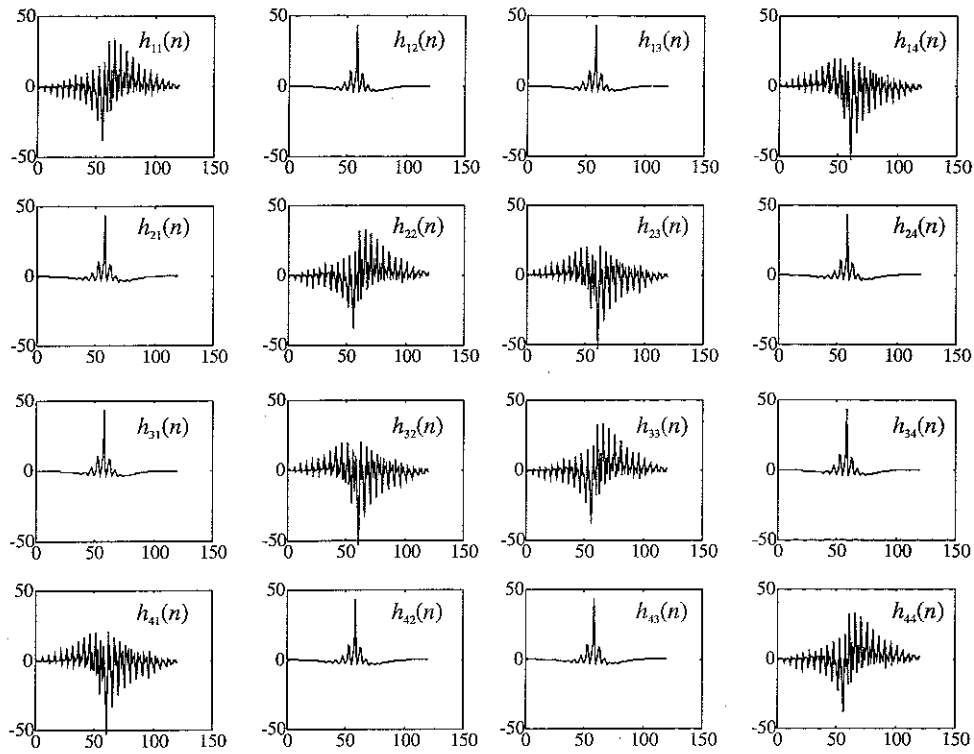


Figure 7.12. The input to the four loudspeakers when two different sets of 3 recorded signals, measured by the circular 3 microphone array shown in Figure 7.1b, are passed through the 12 inverse filters shown in Figure 7.10a

a)



b)

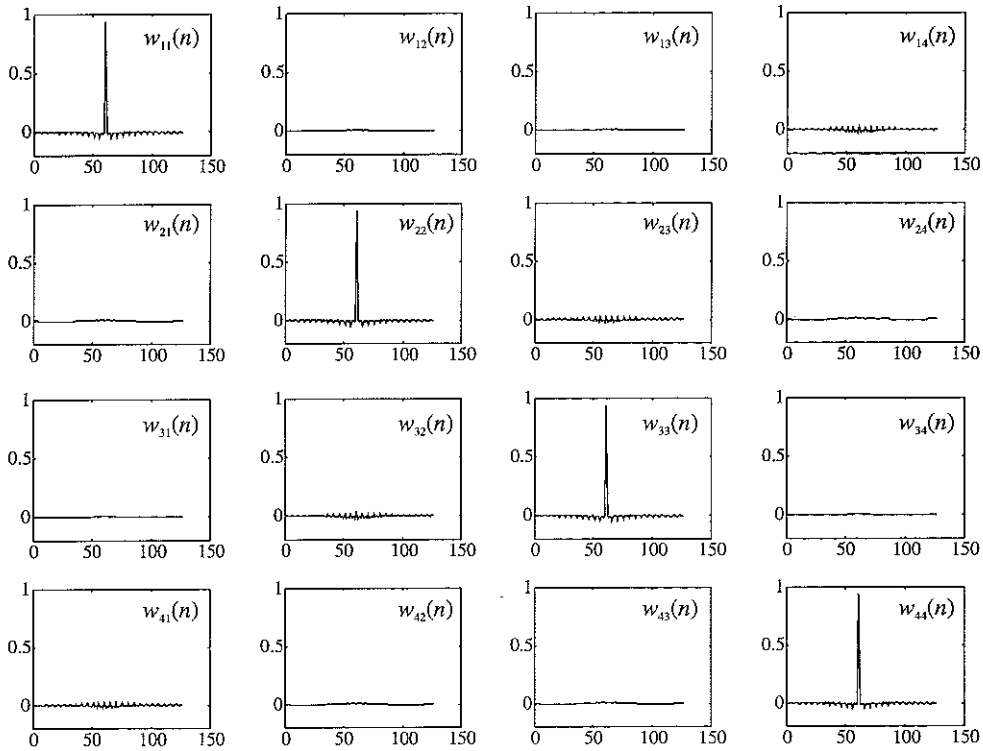


Figure 7.13. a) The 16 inverse filters in  $\mathbf{H}$  and b) the matrix  $\mathbf{W}$  of reproduced signals.  $\mathbf{W}$  is calculated by convolving  $\mathbf{C}$ , which is a 4-by-4 matrix of modelled electro-acoustic transfer functions, with  $\mathbf{H}$

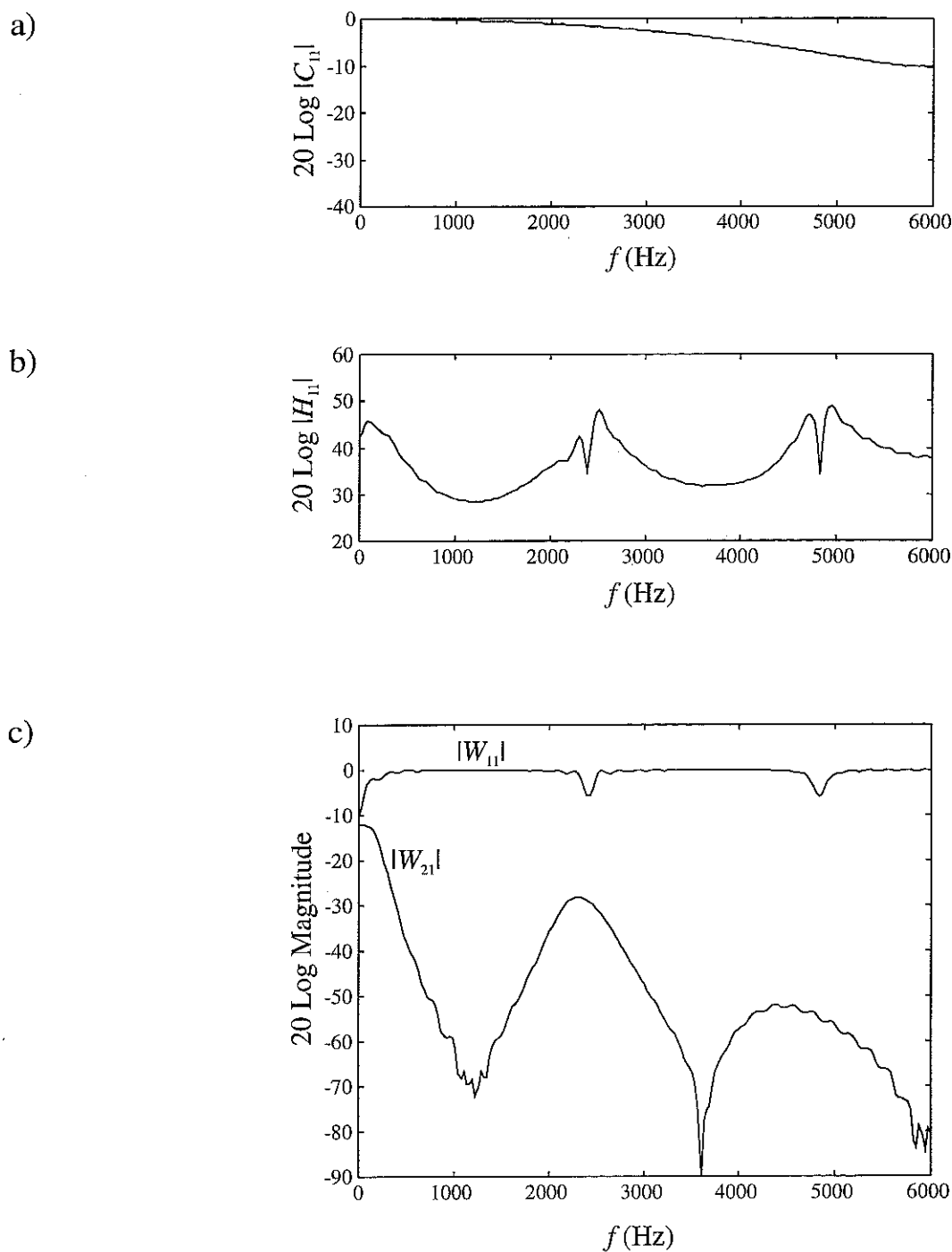
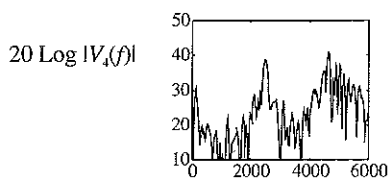
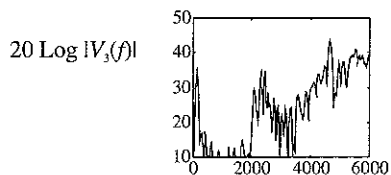
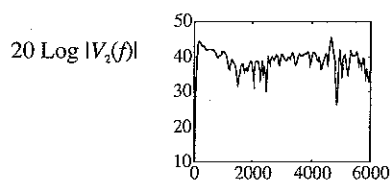
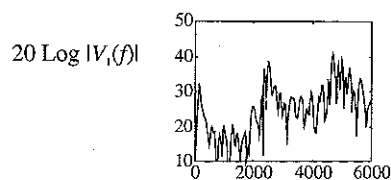
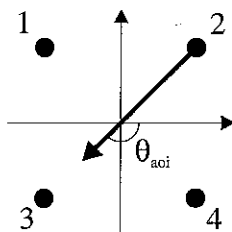


Figure 7.14. a) The frequency response  $|C_{11}|$  of  $c_{11}(n)$  which is a modelled electro-acoustic transfer function. b) The frequency response  $|H_{11}|$  of  $h_{11}(n)$  which is shown in Figure 7.13a. c) The frequency responses  $|W_{11}|$ , which is a diagonal element of  $\mathbf{C}^*\mathbf{H}$ , and  $|W_{21}|$ , which is an off-diagonal element of  $\mathbf{C}^*\mathbf{H}$ . The matrix  $\mathbf{C}^*\mathbf{H}$  is shown in Figure 7.13b

a)  $\theta_{\text{aoi}} = -135^\circ$



b)  $\theta_{\text{aoi}} = -90^\circ$

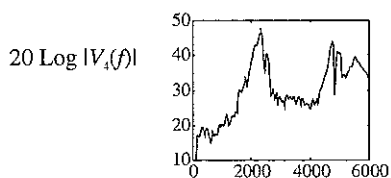
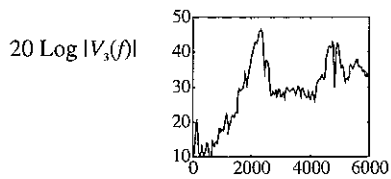
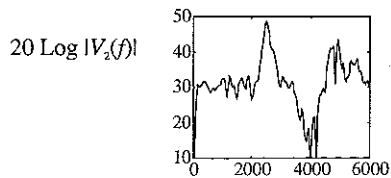
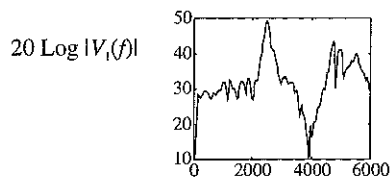
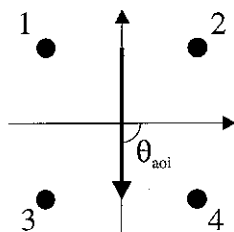
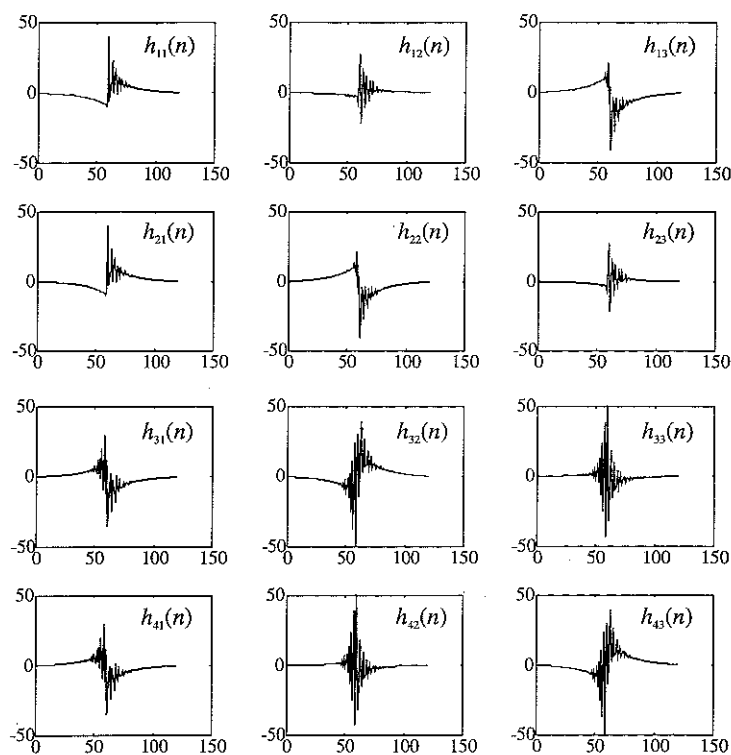


Figure 7.15. The input to the four loudspeakers when two different sets of 4 recorded signals, measured by the 2-by-2 microphone array shown in Figure 7.1a, are passed through the 16 inverse filters shown in Figure 7.13a



a)



b)

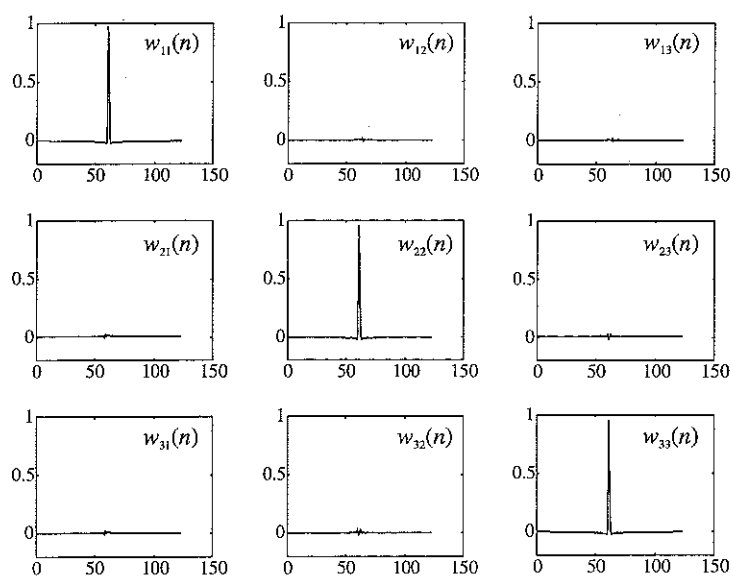


Figure 7.16. a) The 12 inverse filters in  $\mathbf{H}$  and b) the matrix  $\mathbf{W}$  of reproduced signals.  $\mathbf{W}$  is calculated by convolving  $\mathbf{C}$ , which is a 3-by-4 matrix of modelled electro-acoustic transfer functions, with  $\mathbf{H}$

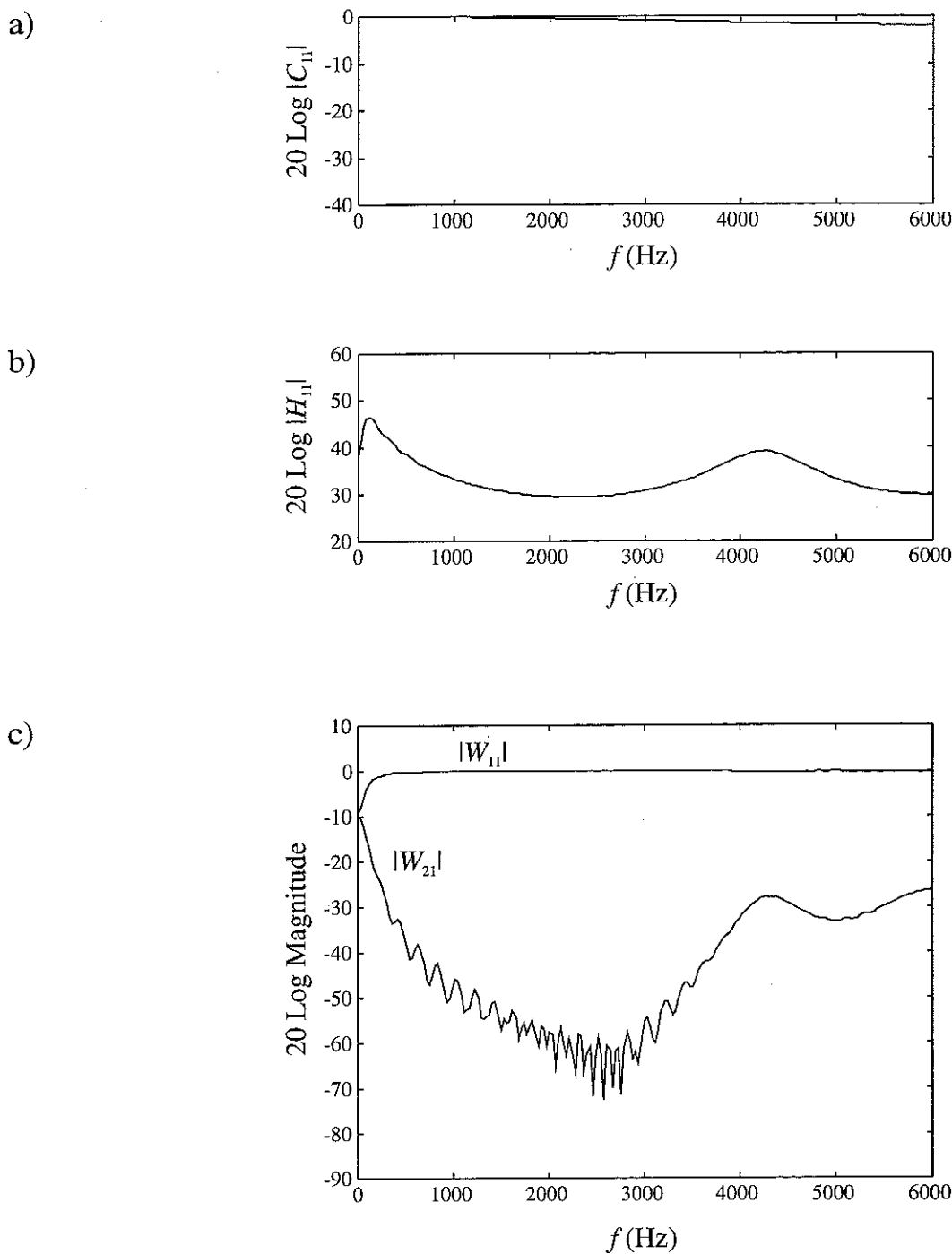
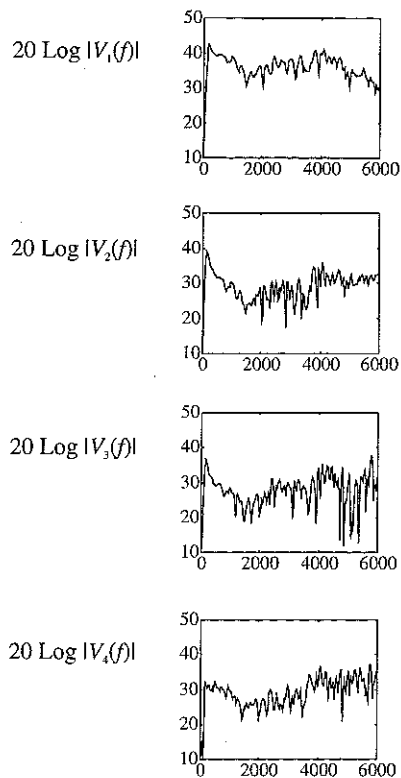
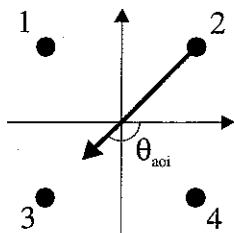


Figure 7.17. a) The frequency response  $|C_{11}|$  of  $c_{11}(n)$  which is a modelled electro-acoustic transfer function. b) The frequency response  $|H_{11}|$  of  $h_{11}(n)$  which is shown in Figure 7.16a. c) The frequency responses  $|W_{11}|$ , which is a diagonal element of  $\mathbf{C}^*\mathbf{H}$ , and  $|W_{21}|$ , which is an off-diagonal element of  $\mathbf{C}^*\mathbf{H}$ . The matrix  $\mathbf{C}^*\mathbf{H}$  is shown in Figure 7.16b

a)  $\theta_{aoi} = -135^\circ$



b)  $\theta_{aoi} = -90^\circ$

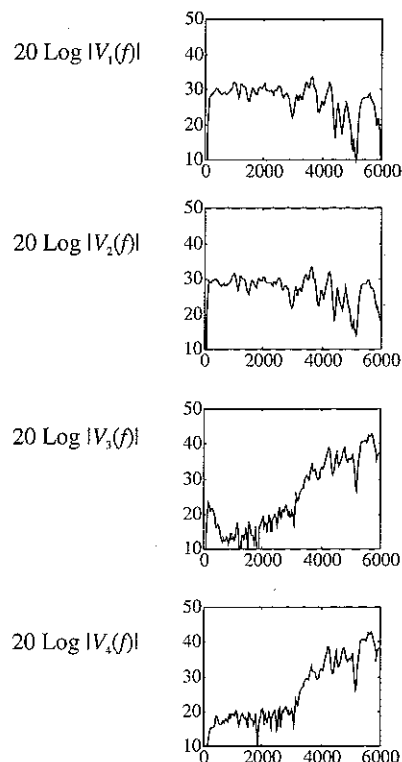
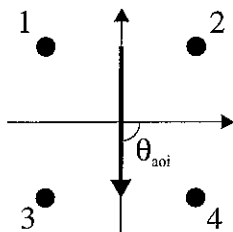


Figure 7.18. The input to the four sources when two different sets of 3 recorded signals, measured by the circular 3 microphone array shown in Figure 7.1b, are passed through the 12 inverse filters shown in Figure 7.16a

## 8. Conclusions

The least squares techniques known from the active control of sound are very suitable for assessing the quality of a reproduced sound field, and the physical limitations that apply to sound field reproduction are demonstrated to agree well with the results already known from listening tests. With only relatively few loudspeakers available, a sound reproduction system based on those techniques is capable of reproducing a sound field faithfully over a "target" area whose size is no greater than a few acoustical wavelengths. When the original sound field is recorded inside the target area with only relatively few microphones, the sound reproduction system still works well as long as the distance between adjacent microphones is no greater than half the acoustical wavelength.

The multi-channel sound reproduction problem is cast into the form of a linear equation system which is then solved in the least squares sense. The task is essentially to invert a matrix which is not necessarily square, and the crucial property of this matrix is its number of linearly independent columns compared to its number of rows. The use of regularisation alleviates the problem of ill-conditioning, and it also makes it possible to define a least squares inversion formula which is applicable to any linear least squares problem regardless of its number of equations and unknowns. The solution is a set of inverse filters specified in either the frequency domain, the  $z$ -domain, or the discrete-time domain. When the inverse filters are designed in the frequency domain, it is not straightforward to calculate the time response of those filters directly because that requires solution of divergent integrals, and the results can only be expressed using delta functions. When the inverse filters are designed in the  $z$ -domain, they are usually unrealisable in the discrete-time domain because they are either non-causal or unstable in forward time. Nevertheless, the frequency domain analysis is very useful for assessing the physical limitations of the reproduction of a sound field, and the  $z$ -domain analysis gives very useful information about the performance of a sound reproduction system implemented by a matrix of inverse finite impulse response filters.

In the frequency domain, and in the  $z$ -domain, the resulting equation systems are generally quite easy to solve as a function of the frequency  $f$ , or as a function of the complex variable  $z$ , since there are only relatively few unknowns, one for each microphone. However, in the discrete-time domain, the resulting equation system is usually very large, and there is an unknown for every single coefficient in each of the inverse filters. This makes the system difficult to solve numerically, particularly when the electro-acoustic transfer functions do not have very short impulse responses. Long impulse responses of the electro-acoustic transfer functions also tend to lead to very ill-conditioned equation systems. Thus, short impulse responses are advantageous numerically, and they can be realised by using modelled transfer functions rather than measured data. Since it is easy to calculate the inverse digital filters as a function of the complex variable  $z$ , it seems a good idea to try to design the inverse filters in the  $z$ -domain rather than in the time domain. For example, a very simple way of doing this would be to sample the  $z$ -transform of the exact least squares inverse filters as defined in Chapter 4 at  $N$  points around the unit circle, and then to use an  $N$  point discrete Fourier transform to obtain the  $N$  coefficients of the impulse responses of each of the inverse filters. An added advantage of the  $z$ -domain filter design is that it would be possible to introduce a frequency weighting of the error. Since the human hearing uses mainly frequencies below a few kHz for localisation, and since standard program material suitable for listening contains most of its energy at relatively low frequencies anyway, it is reasonable to assume that the performance of the system could be improved by optimising the performance of the inverse filters at low frequencies. As stated in Section 4.3, the minimisation of the error in the discrete-time domain is equivalent to a minimisation of the error averaged over the frequencies from 0 Hz to the Nyquist frequency  $f_{\text{Nyq}}$  and so this strategy does not seem to make the best use of the coefficients available in the inverse filters.

By looking at the spatial reproduction of a recorded "transient plane wave", it is possible to get a physical feel for the size of the target area over which the sound field can be controlled relative to the duration of the transient. Since any sound field can be decomposed into plane waves, and since plane waves are conceptually simple, this is a very efficient way to understand the physical limitations of sound field reproduction. This technique is particularly useful when the angle of incidence of the recorded plane

wave is "outside" the range of angles covered by the loudspeakers. For such angles of incidence it can be quite difficult to imagine how the loudspeakers can succeed in "bending" the wavefront near the target area in order to achieve the desired effect.

The results of the initial experiments are in good agreement with the results expected from the simple numerical model of the system: when a sound source is recorded and passed through the inverse filters, most of the sound is emitted from the loudspeaker closest to the position of the original source. This is true for frequencies up to approximately 3kHz when the original sound source is recorded by three microphones positioned in a triangle whose sidelength is 5cm. This performance is surprisingly good considering that the microphones used were of the relatively cheap electret type. In principle, the upper working frequency of the system can be increased by decreasing the distance between the microphones. However, diffraction around the microphones and the differences between the amplitude- and phase responses of the individual microphones inevitably limit this upper working frequency. Ultimately, the potential that the least squares techniques known from active control has for sound reproduction must be judged by listening tests. Such tests should be done with systems that preferably use more than four microphones for the recording and certainly more than four loudspeakers for the reproduction.

## Bibliography

- [1] *A computer model of binaural localization for stereo imaging measurement*, E.A. MacPherson, J. Audio Eng. Soc. 39 (9), 604-622 (1991)
- [2] *A digital signal processing approach to interpolation*, R.W. Schafer, L.R. Rabiner, Proceedings of the IEEE 61 (6), 692-702 (1973)
- [3] *A high-performance surround sound process for home video*, S. Julstrom, J. Audio Eng. Soc. 35 (7/8), 536-549 (1987)
- [4] *A holographic approach to acoustic control*, A.J. Berkhout, J. Audio Eng. Soc. 36 (12), 977-995, (1988)
- [5] *A method of analyzing the quadrasonic sound field*, K. Nakabayashi, J. Audio Eng. Soc. 23 (3), 187-193, (1975)
- [6] *A novel approach to digital beamforming*, R.G. Pridham and R.A. Mucci, J. Acoust. Soc. Am. 63, 425-434 (1978)
- [7] *A remarkable phenomenon with stereophonic sound reproduction*, K. de Boer, Philips Technical Review 9 (1), 8-13 (1947)
- [8] *Acoustics. An Introduction to Its Physical Principles and Applications*, Allan D. Pierce, McGraw-Hill, 1981
- [9] *Active Control of Sound*, P.A. Nelson and S.J. Elliott, Academic Press 1992
- [10] *Adaptive inverse filters for stereophonic sound reproduction*, P.A. Nelson, H. Hamada, and S.J. Elliott, IEEE Transactions on Signal Processing, 40 (7), 1621-1632 (1992)
- [11] *Adaptive Signal Processing*, Bernard Widrow and Samuel D. Stearns, Prentice-Hall, 1985
- [12] *Advanced Engineering Mathematics*, Erwin Kreyszig, John Wiley & Sons, 1983
- [13] *Ambisonics in multichannel broadcasting and video*, M.A. Gerzon, J. Audio Eng. Soc. 33 (11), 859-871 (1985)

- [14] *An improved stereo microphone array using boundary technology, theoretical aspects*, B. Bartlett and M. Billingsley, J. Audio Eng. Soc. 37 (7/8), 543-552, (1989)
- [15] *An Introduction to Acoustical Holography*, B.P. Hildebrand and B.B. Brenden, Plenum press, 1972
- [16] *An Introduction to the Psychology of Hearing (third edition)*, Brian C.J. Moore, Academic Press, 1989
- [17] *An intuitive view of coincident stereo microphones*, S. Julstrom, J. Audio Eng. Soc. 39 (9), 632-649 (1991)
- [18] *Apparent Sound Source Translator*, B.S. Atal, M. Hill, and M.R. Schroeder, United States Patent Office, No. 3,236,949, February 22, 1966
- [19] *Applications of Blumlein shuffling to stereo microphone techniques*, M.A. Gerzon, J. Audio Eng. Soc. 42 (7/8), 435-453 (1994)
- [20] *Approach to recreating a sound field*, M. Camras, J. Acoust. Soc. Am. 43 (6), 1425-1431 (1968)
- [21] *Approximations of Functions*, J.R. Rice, Addison Wesley, 1964
- [22] *Array Signal Processing*, S. Haykin (editor), J.H. Justice, N.L. Owsley, J.L. Yen, and A.C. Kak, Prentice-Hall, 1985
- [23] *Basic stereo microphone perspectives - a review*, P. Streicher and W. Dooley, J. Audio Eng. Soc. 33 (7/8), 548-556 (1985)
- [24] *Beamforming: A versatile approach to spatial filtering*, Barry D. Van Veen and Kevin M. Buckley, 4-24, IEEE ASSP Magazine, April 1988
- [25] *Challenges to the succesful implementation of 3-D sound*, D.R. Begault, J. Audio Eng. Soc. 39 (11), 864-870 (1991)
- [26] *Cinema sound and the evolution of Dolby stereo*, Dolby Laboratories, 1993
- [27] *Complex Variables and Applications (fourth edition)*, Ruel V. Churchill and James Ward Brown, McGraw-Hill, 1984



- [28] *Composite wavefront decomposition via multidimensional filtering of array data*, O.S. Halpheny and D.G. Childers, IEEE Transactions on Circuits and Systems, CAS-22 (6), 552-563 (1975)
- [29] *Computational methods for antenna pattern synthesis*, J.R. Mautz and R.F. Harrington, IEEE Transactions on Antennas and Propagation, AP-23, 507-512 (1975)
- [30] *Computer-steered microphone arrays for sound transduction in large rooms*, J.L. Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, J. Acoust. Soc. Am. 78, 1508-1518 (1985)
- [31] *Criteria for evaluating surround sound systems*, M.A. Gerzon, J. Audio Eng. Soc. (Communications) 25 (6), 400-408 (1977)
- [32] *Digital beamforming for low-pass and band-pass signals*, R.G. Pridham and R.A. Mucci, J. Acoust. Soc. Am. 67, 904-919 (1979)
- [33] *Digital Picture Processing*, Ariel Rosenfeld and Avinash C. Kak, Academic Press, 1982
- [34] *Digital Processing of Signals (second edition)*, Maurice Bellanger, John Wiley & Sons, 1989
- [35] *Digital Signal Processing*, Alan V. Oppenheim and Ronald W. Schaffer, Prentice-Hall, 1975
- [36] *Digital Signal Processing for Multi-Channel Sound Reproduction*, F. Orduna-Bustamante, Ph.D. dissertation, Institute of Sound and Vibration Research, Southampton University, UK, March 1995
- [37] *Discrete-matrix multichannel stereo*, D.H. Cooper and T. Shiga, Audio Eng. Soc. 20 (5), 346-360, (1972)
- [38] *Dolby AC-3 digital coding*, Dolby Laboratories 1993
- [39] *Dolby spectral recording*, Dolby Laboratories 1987
- [40] *Equalization and spatial equalization of dummy-head recordings for loudspeaker reproduction*, D. Griesinger, J. Audio Eng. Soc. 37 (1/2), 20-29, (1989)

- [41] *Evaluation of equalisation and source position compensation systems in two channel sound reproduction*, F. Orduna-Bustamante, P.A. Nelson, and H. Hamada, in Proceedings of Internoise '94, 1837-1840, Yokohama, Japan, August 29-31, 1994
- [42] *Fantasia: Innovations in sound*, J. Klapholz, J. Audio Eng. Soc. 39 (1/2), 66-70 (1991)
- [43] *Fourier Analysis*, T.W. Körner, Cambridge University Press, 1988
- [44] *Fundamentals of Acoustics (third edition)*, L.E. Kinsler, A.R. Frey, A.B. Coppens, and J.V. Sanders, John Wiley and Sons, 1982
- [45] *Fundamentals of binaural technology*, H. Møller, Applied Acoustics 36, 171-218 (1992)
- [46] *Fundamentals of digital array processing*, D.E. Dudgeon, Proceedings of the IEEE 65 (6), 898-904 (1977)
- [47] *Head-related two-channel stereophony with loudspeaker reproduction*, P. Damaske, J. Acoust. Soc. Am. 50, 1109-1115 (1971)
- [48] *How can the headroom of digital recordings be used optimally?*, M. Krause and H. Petersen, J. Audio Eng. Soc. 38 (11), 857-863 (1990)
- [49] *Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems*, A.D. Blumlein, British Patent Specification No. 394,325, June 14, 1933, reprinted in J. Audio Eng. Soc. 6 (2), 91-98 and 130, (1958)
- [50] *Inverse filtering in sound reproduction*, O. Kirkeby, F. Orduna-Bustamante, P.A. Nelson, and H. Hamada, Journal of Measurement + Control, 26 (9), 261-267 (1993)
- [51] *Inverse filtering of room acoustics*, M. Miyoshi and Y. Kanada, IEEE Transactions on Acoustics, Speech and Signal Processing 36 (2), 145-52 (1988)
- [52] *Inverse filters for multi-channel sound reproduction*, P.A. Nelson, H. Hamada, and S.J. Elliott, IEICE Trans. Fundamentals, E75-A, 11, 1468-1473, November 1992

- [53] *Invertibility of a room impulse response*, S.T. Neely and J.B. Allen, J. Acoust. Soc. Am. 66, 165-69 (1979)
- [54] *Least squares approximations to exact multiple point sound reproduction*, P.A. Nelson and S.J. Elliott, Institute of Sound and Vibration Research memorandum 683, Southampton University, UK (1988), also published in Proceedings of the Institute of Acoustics /\*need details\*/ (1988)
- [55] *Localization of lateral phantom sources*, G. Theile and G. Plenge, J. Audio Eng. Soc. (project notes/engineering briefs) 25 (4), 196-200, (1977)
- [56] *M-S stereo: A powerful technique for working in stereo*, W.L. Dooley and R.D. Streicher, J. Audio Eng. Soc. 30 (10), 707-717 (1982)
- [57] *Mathematica: A System for Doing Mathematics by Computer*, Stephen Wolfram, Addison-Wesley, 1991
- [58] *Matlab Reference Guide*, Math Works Inc., August 1992
- [59] *Measurement of large loudspeaker array performance*, M.R. Gander and J.M. Eargle, J. Audio Eng. Soc. 38 (4), 204-220 (1990)
- [60] *Models of hearing*, M.R. Schroeder, Proceedings of the IEEE 63 (9), 1332-1352 (1975)
- [61] *Motion picture sound recording - a capsule history*, J.G. Frayne, J. Audio Eng. Soc. 24 (6), 512-516, (1976)
- [62] *Multi-channel digital signal processing in sound field reproduction*, F. Orduna-Bustamante, P.A. Nelson, and H. Hamada, Second Conference on Recent Advances in Active Control of Sound and Vibration, edited by R.A. Burdisso, 981-992, Technomic publishing company, 1993
- [63] *Multichannel sound systems for HDTV*, D.J. Meares, Applied Acoustics 36, 245-57 (1992)
- [64] *Multidimensional Digital Signal Processing*, D.E. Dudgeon and R.M. Mersereau, Prentice-Hall, 1984

- [65] *Multimicrophone signal-processing technique to remove room reverberation from speech signals*, J.B. Allen, D.A. Berkley, and J. Blauert, J. Acoust. Soc. Am. 62, 912-915 (1977)
- [66] *Multirate Digital Signal Processing*, Ronald E. Crochiere and Lawrence R. Rabiner, Prentice-Hall, 1983
- [67] *Near-field calibration arrays for acoustic wavefield determination*, A.L. Van Buren, IEEE Transactions on Instrumentation and Measurement, 41 (1), 22-26 (1992)
- [68] *Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH*, J.D. Maynard, E.G. Williams, and Y. Lee, J. Acoust. Soc. Am. 78, 1395-1412 (1985)
- [69] *New factors in sound for cinema and television*, T. Holman, J. Audio Eng. Soc. 39 (7/8), 529-540 (1991)
- [70] *NRDC surround sound system*, M. Gerzon, Wireless World, 36-39, April 1977
- [71] *Numerical Recipes in C (second edition)*, William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, Cambridge University Press, 1992
- [72] *On the diffraction of a progressive sound wave by the human head*, F.M. Wiener, J. Acoust. Soc. Am. 19, 143-146 (1947)
- [73] *On the naturalness of two-channel stereo sound*, G. Theile, J. Audio Eng. Soc. 39 (10), 761-767 (1991)
- [74] *On the standardization of the frequency response of high-quality studio headphones*, G. Theile, J. Audio Eng. Soc. 34 (12), 956-969 (1986)
- [75] *Optimum reproduction matrices for multispeaker stereo*, M.A. Gerzon, J. Audio Eng. Soc. 40 (7/8), 571-589 (1992)
- [76] *Perceptual effects of synthetic reverberation on three-dimensional audio systems*, D.R. Begault, J. Audio Eng. Soc. 40 (11), 895-904 (1992)
- [77] *Principles of Applied Mathematics, Transformation and Approximation*, James P. Keener, Addison-Wesley, 1987

- [78] *Problems in the design and use of "dummy-heads"*, M. Kleiner, *Acustica*, 41, 183-193 (1978)
- [79] *Processing artificial-head recordings*, H.W. Gierlich and K. Genuit, *J. Audio Eng. Soc.* 37 (1/2), 34-39, (1989)
- [80] *Prospects for transaural recording*, D.H. Cooper and J.L. Bauck, *J. Audio Eng. Soc.* 37 (1/2), 3-19, (1989)
- [81] *Quadraphony - A Review*, J.G. Woodward, *J. Audio Eng. Soc.* 25 (10/11), 843-54 (1977)
- [82] *Reproduction of artificial head-recordings through loudspeakers*, H. Møller, *J. Audio Eng. Soc.* 37 (1/2), 30-33, (1989)
- [83] *Room-related balancing technique: A method for optimizing recording quality*, M. Wöhr, G. Theile, H.J. Goeres, and A. Persterer, *J. Audio Eng. Soc.* 39 (9), 623-631 (1991)
- [84] *Second conference on recent advances in active control of sound and vibration*, edited by R.A. Burdisso, Technomic publishing company, 1993
- [85] *Signal processing to reduce multipath distortion in small rooms*, J.L. Flanagan and R.C. Lummis, *J. Acoust. Soc. Am.* 47, 1475-1481 (1970)
- [86] *Signals and Systems*, Alexander D. Poularikas and Samuel Seely, Pws-Kent, 1985
- [87] *Sound and Sources of Sound*, A.P. Dowling and J.E. Ffowcs-Williams, Ellis Horwood Limited, 1983
- [88] *Sound field microphone*, K. Farrar, *Wireless World*, 48-51, October 1979
- [89] *Sound field microphone - 2*, K. Farrar, *Wireless World*, 99-104, November 1979
- [90] *Sound Intesity*, F.J. Fahy, Elsevier Applied Science, 1989
- [91] *Sound Recording*, J. Eargle, Van Nostrand Reinhold Company, 1976
- [92] *Sound reproduction for high-definition television* (in German), G. Steinke, *Fernseh- und Kino-Technik*, 47 (5), 305-313 and 316-317, May 1993

- [93] *Spatial Hearing (American Edition)*, Jens Blauert (translated by John S. Allen), MIT Press, Cambridge, Massachussets, 1983
- [94] *Steered planar near-field calibration arrays*, A.L. Van Buren, J. Acoust. Soc. Am. 50, 1052-1059 (1978)
- [95] *Stereo microphone techniques - are the purists wrong?*, S.P. Lipshitz, J. Audio Eng. Soc. (features) 34 (9), 716-744 (1986)
- [96] *Stereo reproduction with good localization over a wide listening area*, S. Aoki, H. Miyata, and K. Sugiyama, J. Audio Eng. Soc. 38 (6), 433-439, (1990)
- [97] *Stereo sound recording with shotgun microphones*, H. Gerlach, J. Audio Eng. Soc. 37 (10), 832-838, (1989)
- [98] *Stereophonic sound reproduction*, K. de Boer, Philips Technical Review 5 (4), 107-114 (1940)
- [99] *Stereophonic sound reproduction in the home*, H.F. Olson, Audio Eng. Soc. 6 (2), 80-90, (1958)
- [100] *STSF - a unique technique for scan based near field holography without restrictions on coherence*, J. Hald, Brüel & Kjær Technical Review (1), 1-50 (1989)
- [101] *Surround sound psychoacoustics*, M. Gerzon, Wireless World, 483-486, December 1974
- [102] *Surrounded by Sound*, Peter Dolman, Electronics World + Wireless World, 35-45, January 1990
- [103] *The acoustics and sound system for hemispherical film projection*, P.H. Heringa, B.H.M. Kok, and Y. Dekeyrel, J. Audio Eng. Soc. 35 (3), 119-29 (1987)
- [104] *The Algebraic Eigenvalue Problem*, J.H. Wilkinson, Oxford University Press, 1965
- [105] *The application of binaural technology*, H.W. Gierlich, Applied Acoustics 36, 219-243 (1992)

- [106] *The evolution of, and current developments in, non-commercial cinema/surround audio technology*, F.J. Ampel, Proceedings of the Institute of Acoustics, Vol. 15, Part 7, 165-173 (1993)
- [107] *The generalized multichannel filtered-X algorithm*, D.C. Swanson, Second Conference on Recent Advances in Active Control of Sound and Vibration, edited by R.A. Burdisso, 550-561, Technomic publishing company, 1993
- [108] *The Technique of the Sound Studio*, Alec Nisbett, Focal Press, 1979
- [109] *Theoretical design of nearfield calibration arrays*, A.L. Van Buren, J. Acoust. Soc. Am. 50, 192-199 (1973)
- [110] *Theory and Application of the z-Transform Method*, E.I. Jury, John Wiley & Sons, 1964
- [111] *Transfer function phase and truncated impulse response*, M. Tohyama and R.H. Lyon (letters to the editor) J. Acoust. Soc. Am. 86, 2025-2029 (1989)
- [112] *Transforms in Signals and Systems*, Peter Kraniuskas, Addison-Wesley, 1992
- [113] *Underwater sound transducer calibration from nearfield data*, W.F. Trott, J. Acoust. Soc. Am. 36, 1557-68 (1964)
- [114] *Use of acoustic filtering to control the beamwidth of steered microphone arrays*, J.L. Flanagan, J. Acoust. Soc. Am. 78, 423-428 (1985)
- [115] *User's guide to the signal processing packages and notebooks for implementing linear system theory in Mathematica 1.2 and 2.0*, B.L. Evans, J.H. McClellan, L.J. Karam, K.A. West, W.B. McClure, and J.A. Proctor, available from Mathsource Database via anonymous ftp
- [116] *Vehicle noise investigation using spatial transformation of sound fields*, J. Hald and K.B. Ginn, Sound and Vibration, 38-45, April 1989
- [117] *XY and MS microphone techniques in comparison*, M. Hibbing, J. Audio Eng. Soc. 37 (10), 823-831, (1989)
- [118] *Zeros of a transfer function in a multi-degree-of-freedom vibrating system*, M. Tohyama and R.H. Lyon, J. Acoust. Soc. Am. 86, 1854-63 (1989)

## Notation

Scalar variables are in italics (examples:  $x$ ,  $X$ )

Scalar functions of time variables are in lowercase italics (examples:  $x(n)$ ,  $x(t)$ )

Scalar functions of frequency variables are in uppercase italics (examples:  $X(z)$ ,  $X(\omega)$ )

Vectors are in lowercase bold (examples:  $\mathbf{x}$ ,  $\mathbf{x}(t)$ ,  $\mathbf{x}(\omega)$ )

Matrices are in uppercase bold (example:  $\mathbf{X}$ )

Subscript 0 indicates an optimal value (example:  $\mathbf{x}_0$ )

$a$	target response
$c$	electro-acoustic transfer function, or speed of sound
$d$	desired signal
$e$	performance error
$E_{\text{amp}}$	amplitude error
$E_{\text{phase}}$	phase error
$e_{\text{quant}}$	quantization error
$E_{\text{kin}}$	kinetic energy
$E_{\text{pot}}$	potential energy
$E_{\text{rms}}$	root mean squared performance error
$f$	frequency
$f_s$	sampling frequency
$f_{\text{max}}$	a signal's highest frequency component
$f_{\text{Nyq}}$	half the sampling frequency ( $f_s/2$ )
$h$	inverse filter
$I$	sound intensity
$I_a$	active sound intensity
$I_r$	reactive sound intensity



$J$	total cost
$k$	wave number
$K$	spectral condition number
$l$	distance measured in sampling intervals
$l_{\text{frac}}$	fractional part of $l$
$l_{\text{int}}$	integer part of $l$
$l_{\text{round}}$	$l$ rounded to nearest integer
$m$	modelling delay in samples
$n$	discrete time index
$N_c$	number of filter coefficients in an electro-acoustic impulse response
$N_{\text{cols}}$	number of columns in a matrix
$N_h$	number of filter coefficients in an inverse filter
$N_{\text{nz}}$	number of non-zero elements in a matrix
$N_p$	number of poles of a complex function (not including zero and infinity)
$N_{\text{rows}}$	number of rows in a matrix
$N_u$	number of filter coefficients in a recorded signal
$N_z$	number of zeros of a complex function (not including zero and infinity)
$q$	source strength
$R$	number of microphones (r for receivers)
$S$	number of loudspeakers (s for sources)
$t$	time
$T$	number of tracks
$T_s$	sampling interval
$u$	recorded signal, or particle velocity
$v$	source acceleration, or loudspeaker input signal

$w$	reproduced signal
$\mathbf{x}, x_1, x_2, x_3$	spatial coordinates
$Z$	specific acoustic impedance
$Z_{\text{re}}$	specific acoustic resistance
$Z_{\text{im}}$	specific acoustic reactance
$\beta$	regularisation parameter
$\lambda$	acoustical wavelength
$\phi$	phase
$\rho$	density of medium
$\rho_{\text{c}}$	density of elements in the matrix $\mathbf{C}$
$\tau$	discrete time constant
$\theta_{\text{aoi}}$	angle of incidence of plane wave
$\theta_{\text{ver}}$	angle of vertical component of plane wave
$\omega$	angular frequency
$\Omega$	target area or target volume
$ x $	Absolute value of the complex number $x$
$ \mathbf{X} $	Determinant of the matrix $\mathbf{X}$
$\ \mathbf{x}\ $	Euclidean length of the vector $\mathbf{x}$
$x^*$	Conjugate of $x$
$\mathbf{X}^T$	Transpose of the matrix $\mathbf{X}$
$\mathbf{X}^H$	Hermitian transpose of matrix $\mathbf{X}$ , equivalent to $(\mathbf{X}^T)^*$