The Department of Physics
The University of Auckland
New Zealand

Microphone Phased Array 3D Beamforming and Deconvolution

Mathew Legg

April 2012

Supervisor: Prof. Stuart Bradley

A thesis submitted in fulfillment of the requirements of Doctor of Philosophy in Physics

Abstract

Microphone phased arrays are used to generate acoustic maps showing the position and magnitude of sound sources. Deconvolution of these acoustic maps, which are generated using beamforming, is commonly performed to remove sidelobe artifacts so that it is possible to accurately describe the position and magnitude of the sound source distribution. Traditionally beamforming and deconvolution have used a 2D scanning surface, which is orientated perpendicular to the array axis, but errors can arise when imaging 3D objects. The work in this thesis investigates the use of a deconvolution algorithm for 3D beamformed maps and compares the results to those obtained using traditional 2D acoustic scanning surfaces.

Microphone phased array hardware and 3D objects were designed and built. Acoustic maps were generated by attaching mini speakers onto the surface of an object and performing beamforming and deconvolution for both traditional 2D scanning surfaces and 3D scanning surfaces corresponding to the 3D surface geometry of the object. The 3D surface geometry was obtained using computer vision techniques. For more complex objects or where no CAD model of the object exists, structured light scanning was used to obtain an accurate scan of the 3D surface of the object. The scan points obtained using the above two methods were in the reference frame of the primary optical camera in the array. To enable these scan points to be used for beamforming and deconvolution, a microphone position calibration technique was developed which automatically found the coordinates of the microphones, in the reference frame of the primary camera in the array, using computer vision techniques and audio time of flight measurements. This technique was extended to enable dense point clouds of experimental deconvolution errors to be automatically obtained as a function of the frequency and location of the sound sources. These point clouds were used to analyse the deconvolution errors for 3D and traditional 2D scanning surfaces. The data obtained showed that using the 3D scanning surface corresponding to the surface geometry of the object gave more accurate sound pressure levels and, at higher frequencies, more accurate positioning of sound sources than the 2D case.

Acknowledgements

Most of all I would like to thank my supervisor Prof. Stuart Bradley for his help, encouragement, and support throughout this study. While always being ready to help, he has allowed and encouraged me to pursue my own research, and for this I am very grateful. I would also especially like to thank Associate Prof. Robert Dougherty of OptiNav and the *University of Washington* for his advice and encouragement at the initial stage of this study. Thanks also to Gunnar Heilmann of *GFaI* for suggesting the idea of deconvolution of 3D beamforming maps as a topic of this thesis. I would like to thank my colleagues Alex Strehz, Paul Behrens, and especially Brice Vallés for their encouragement and for proof reading of this thesis. Thanks also to Steve Warrington, Owen Caughley, and Trevor Waldmeyer of the *Faculty of Science Workshop* for the construction of some of the hardware used in this study. I would also like to thank Bryan Davis, Murray Hollis, and Harry Oudenhoven of the *Physics Electronic Workshop* for their advice in relation to the preamp circuit board construction. Thanks too to Francie Norman who, as the Physics Department Manager, always makes administration matters run smoothly. Finally, thanks to my mother for her encouragement and help.

Contents

List of Key Symbols

Notation

a	Scalar value.
$\boldsymbol{a}$	Array or matrix of values.
$\boldsymbol{a}_n$	The n^{th} component of the array of elements $\boldsymbol{a}$.
$\boldsymbol{a}_{mn}$	The m^{th} row, n^{th} column component of the matrix $\boldsymbol{a}$.
$\vec{\boldsymbol{a}}$	Position vector notation.
$\bar{\mathsf{a}}$	Homogeneous vector notation.
$\boldsymbol{a}^\dagger$	Hermitian complex conjugate transpose.
$\boldsymbol{a}^\intercal$	Transpose.
$\langle \boldsymbol{a} \rangle$	Time average for time domain data and $STFT$ data.
$\boldsymbol{a} \cdot \boldsymbol{b}$	Dot product.
$\|\boldsymbol{a}\|$	Euclidean norm.
$\|\boldsymbol{a}\|$	Norm.
$\boldsymbol{I}_{(a,a)}$	Identity matrix of size $(a \times a)$.
$\mathbf{0}_{(a,b)}$	Zero matrix of size $(a \times b)$.
$a^{\{b\}}$	Loop iteration b.

Acoustic Imaging Symbols

$\mathbf{A}$	*DAMAS* matrix.
$\boldsymbol{b}(\vec{\boldsymbol{\xi}})$	Beamformed map evaluated at point $\vec{\boldsymbol{\xi}}$.
$\boldsymbol{b}_{max}^{\{l\}}$	*CLEAN-SC* peak intensity value for iteration l.
BW	-3dB beamwidth of main lobe.
c	Speed of sound.

$\boldsymbol{C}(\vec{\boldsymbol{X}}_s)$	Array propagation vector for a point source at $\vec{\boldsymbol{X}}_s$.
d	Sound source separation.
D	Diameter of the array.
$\mathbf{D_{cal}}$	Diagonal eigenvalue calibration matrix.
$\boldsymbol{E}$	Uncorrelated channel noise.
η_m	Microphone sensitivity.
f	Frequency.
F	Figure of merit of an array.
$\mathbf{G}$	Cross-spectral matrix.
$\boldsymbol{h}$	Coherent source component in $\mathbf{G}$ used in the *CLEAN-SC* algorithm.
$\mathbf{H}$	Diagonal matrix used in the *CLEAN-SC* algorithm if diagonal removal of $\mathbf{G}$ is used.
j	Index of data blocks used in CSM.
J	Number of data blocks used in CSM.
k	Frequency bin index.
K	Data block length of $STFT$.
$\vec{\boldsymbol{\kappa}}$	Unit vector in direction of propagation of a sound wave or in steering direction.
l	*CLEAN-SC* iteration loop number.
λ	Wavelength.
$\mathbf{\Lambda}$	Propagation amplitude factor.
m	Microphone index.
M	Number of array microphones.
n	Acoustic scan point index.
N	Number of acoustic scan points.
$\mathbf{psf}(\vec{\boldsymbol{\xi}}, \vec{\boldsymbol{X}}_s)$	Point spread function for source located at $\vec{\boldsymbol{X}}_s$ for scan points located at $\vec{\boldsymbol{\xi}}$.
$\mathbf{\Phi}$	*CLEAN-SC* normalized clean beam.
$\mathbf{\Psi}$	General cross correlation weight function.
$q(\vec{\boldsymbol{X}}_s)$	Time-domain acoustic source strength for position $\vec{\boldsymbol{X}}_s$.
$Q(\vec{\boldsymbol{X}}_s)$	Frequency-domain acoustic source strength for position $\vec{\boldsymbol{X}}_s$.
q_o	Source amplitude.
$\boldsymbol{\mathcal{Q}}$	*CLEAN-SC* clean map.
$\mathcal{R}$	General cross correlation.
$\sigma(\vec{\boldsymbol{X}}_m, \vec{\boldsymbol{X}}_s)$	Acoustic propagation time from a sound source at $\vec{\boldsymbol{X}}_s$ to the m^{th} microphone.
s	Sound source index.

S	Total number of sound sources.
t	Time.
TOF	Time of flight.
$TDOF$	Time difference of flight.
τ_{mn}	Time delay used for beamforming for grid point n and microphone m.
Θ	Maximum likelihood estimator.
$\boldsymbol{u}$	Time-domain microphone array signal.
$\boldsymbol{U}$	Frequency-domain microphone voltage signal.
μ	Number of unique vector spacings in an array.
$\boldsymbol{v}$	Eigenvector of $\mathbf{G}$.
$\boldsymbol{\varepsilon}$	Position error of deconvolution peak.
$\boldsymbol{\varphi}$	*CLEAN-SC* loop gain.
ω	Angular frequency.
$\mathbf{w}$	Array steering vector.
W_m	Windowing function used for microphone m in cross spectral matrix.
$\mathcal{W}$	The Rayleigh criterion.
$\vec{\boldsymbol{X}}_m$	Location of microphone m.
$\vec{\boldsymbol{\xi}}_{max}^{\{l\}}$	Coordinates of the l peak in the *CLEAN-SC* deconvolution map.
$\vec{\boldsymbol{X}}_p$	Vector spacings between microphones in the array.
$\vec{\boldsymbol{X}}_s$	Location of source.
$\vec{\boldsymbol{\xi}}$	Acoustic scanning surface coordinate vector.
$\mathbf{Y}$	*DAMAS* beamformed map matrix.

Computer Vision Symbols

α_c	Camera skew coefficient corresponding to the angle between the normalised x and y pixel axes.
$\boldsymbol{c}_c$	Camera principal point: normalised x and y-axis components.
dx, dy	Dimension of a pixel: normalised x and y-axis components.
dX, dY	Dimensions of a single checkerboard square: x and y-axis components.
f_o	Camera focal length.
$\boldsymbol{f}_c$	Camera focal length in pixels.
$\mathcal{F}$	Reference frame.

g	Rigid body motion transformation matrix.
$\mathbf{I}$	Identity matrix.
$\boldsymbol{\zeta}$	Camera image distortion coefficients (radial and tangential distortions) 5x1 vector.
$\boldsymbol{\mathcal{K}}$	Camera matrix.
$\boldsymbol{\ell}$	Variable for scaling projection vector.
O	Origin of reference frame.
$\vec{\boldsymbol{\Omega}}$	Rotation vector.
$\vec{\boldsymbol{p}}$	Pixel coordinates in image plane.
p	Point or set of points on image plane.
P	Point or set of points in 3D Euclidean space.
$\mathbf{R}$	Rotation matrix.
$\vec{\boldsymbol{T}}$	Translation vector.
$\vec{\boldsymbol{x}}$	Pinhole projection coordinate vector on image plane.
$\vec{\boldsymbol{x_d}}$	Pinhole projection vector with lens distorted included.
X, Y, Z	X,Y and Z-axis components of position vector.
$\vec{\boldsymbol{X}}$	Coordinate vector in 3D Euclidean space.

Abbreviations

CLEAN-SC	CLEAN algorithm based on Source Coherence.
CSM	Array Cross Spectral Matrix.
DAMAS	Deconvolution Approach for the Mapping of Acoustic Sources.
DAS	Delay And Sum beamforming.
DR	Diagonal Removal.
FFT	Fast Fourier Transform.
GCC	General Cross Correlation.
kSPS	Sampling rate unit (kilo Samples Per Second.)
LORE	Localization and Optimization of array REsults.
PHAT	PHAse Transformation.
RELAX	Relaxation parametric method.
SADA	Small Aperture Directional Array.
SONAH	Statistically Optimized Near-field Acoustic Holography.
SPL	Sound Pressure Level.
SPR	Steered Response Power.
STFT	Short Time Fourier Transform.
TOA	Time Of Arrival.

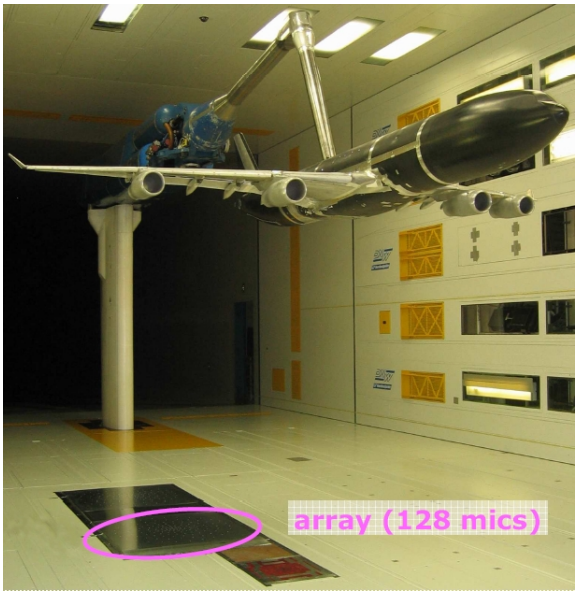
1

Introduction

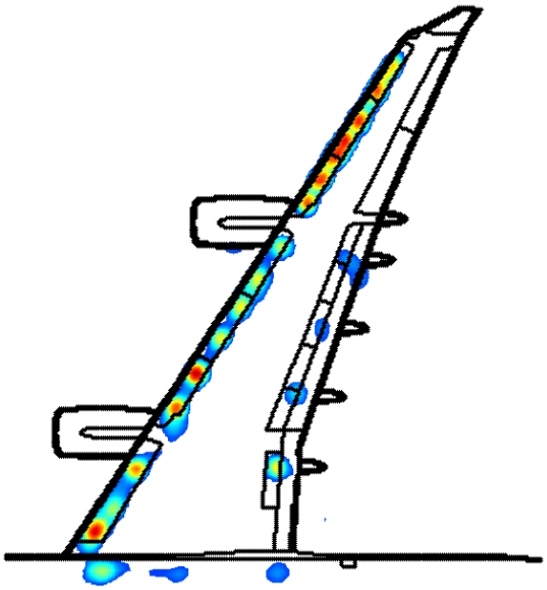
1.1 Background

The microphone phased array has been developed as a tool to enable the position and magnitude of sound sources to be identified. Arrays are commonly used by industries such as aeroplane and automotive manufacturers to identify sound sources. During the design process of an aeroplane, for example, a model of part of an aeroplane, such as a wing or landing gear, might be placed in a wind tunnel. A microphone phased array, located in the wall of the wind tunnel, is used to image the location and magnitude of sound sources on the model, see Figure 1.1.

A microphone phased array is composed of multiple microphones, data acquisition hardware, and software. The acoustic signals, received by the spatially separated microphones, are used to generate acoustic maps of the sound source distribution. To enable sound sources to be identified, an acoustic map is commonly overlayed as a transparency over an image captured by a camera located at the centre of the array. For this reason, microphone phased arrays are often referred to as 'acoustic cameras'. To avoid any confusion with optical cameras, the term 'microphone phased array' or just 'microphone array' will be used in this work. There are three main methods used to generate acoustic maps



(a) Microphone phased array and A340 scale model in wind tunnel. Figure from Sijtsma (2008).



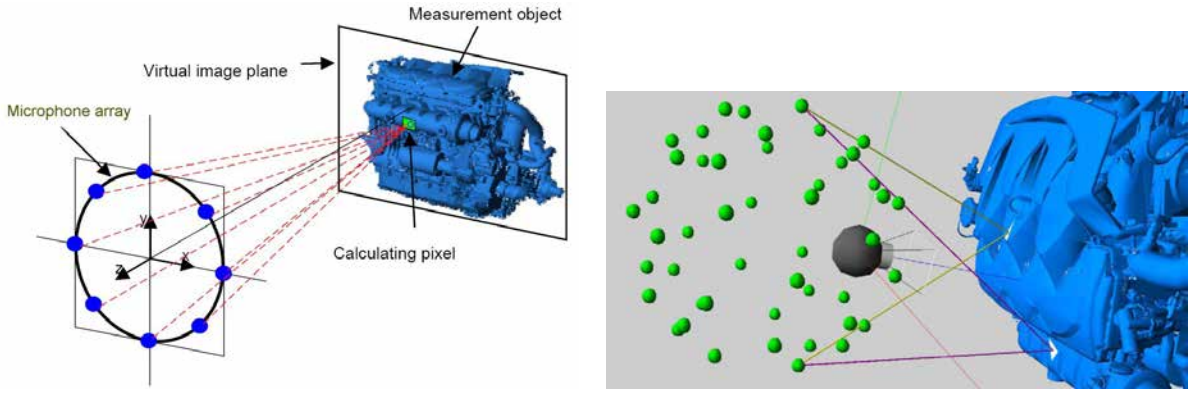
(b) Acoustic map showing noise sources on the wing. This acoustic map was generated using *CLEAN-SC* deconvolution of a planar beamformed map. Figure from Sijtsma (2007a).

Figure 1.1: Example of an application of microphone phased array in aeroacoustics.

from microphone phased array data. These are beamforming (Humphreys et al., 1998; Dougherty, 2004), acoustic holography (Maynard and Williams, 1985; Hald, 2004; Deblauwe et al., 2007; Lanslots et al., 2010), and inverse methods (Suzuki, 2008; Lanslots et al., 2010; Schmitt et al., 2010).

Beamforming uses delaying and summing of microphone channel data to obtain acoustic maps. It can be used to acoustically image objects located near or far from the microphone array and objects larger than the diameter of the microphone array. A disadvantage of this method, however, is that the beamformed maps contain an interference pattern referred to as sidelobes. For example, if beamforming is used to image an acoustic point source, the resulting beamformed map may consist of a main lobe and unwanted additional sidelobes. The angular width of the main lobe is proportional to the wavelength of the acoustic signal, resulting in reduced spatial resolution of the beamformed maps for lower frequencies. Algorithms have been developed which perform deconvolution of the beamformed maps to remove these sidelobes and attempt to obtain the true sound source distribution. Examples of these algorithms are *DAMAS* (Deconvolution Approach for the Mapping of Acoustic Sources) (Brooks and Humphreys, 2004) and *CLEAN-SC* (CLEAN algorithm based on Source Coherence.) (Sijtsma, 2007b). These algorithms are discussed in more detail in Section 2.4.

Acoustic phased array beamforming and deconvolution have traditionally used planar scanning surfaces to image the noise emitted by an object, see Figure 1.2a. This 2D



(a) Traditional 2D scanning surface.

(b) Scanning surface corresponding to the 3D surface geometry of the object.

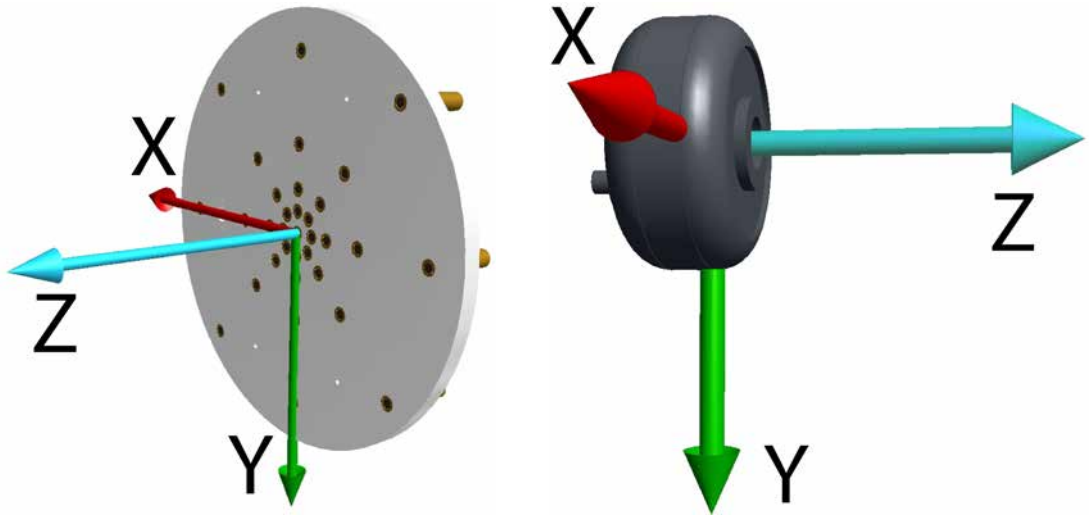
Figure 1.2: Diagram of a 2D and 3D of scanning surface for use with beamforming. Figures from Heilmann et al. (2008a)

scan surface is usually oriented perpendicular to the array's Z axis, refer to Figure 1.3a, and located roughly at the same distance from the array as the object. This method is simple to use since the only information required to be known about the object is the approximate distance of the object from the array. Often, however, only a few of the scan points will correspond to the actual 3D surface of the object. This can lead to errors in the resulting 2D beamforming and deconvolution maps. These errors appear as incorrect measured sound pressure (SPL) levels, incorrect location of sound sources, and distortions in the plotting of the acoustic maps due to incorrect focus being used for the beamforming (Heilmann et al., 2008b; Döbler et al., 2008; Maffei and Bianco, 2008; Irimia et al., 2009). Dougherty (2005b) refers to this issue.

> "Phased arrays are usually used for 2D imaging, although severe errors can occur if the source is not in the beamforming plane. These include parallax errors in source location and the possibility that sources can be missed entirely or that sidelobes in the form of out of plane sources can be confused with real sources in the plane. Expanded use of 3D imaging could improve the results."

1.2 3D Grids and Scanning Surfaces

Beamforming and deconvolution have been performed using a 3D grid (Brooks and Humphreys, 2005; Ravetta, 2005; Xenaki et al., 2010; Dougherty, 2010a). However, most practical arrays have a poor beamforming and deconvolution resolution in the direction along the array axis, which is the Z axis for the 2D array shown in Figure 1.3a. Another problem is that these 3D grids can contain a large number of scan points which can make deconvolution of these 3D grid beamformed maps very computationally expensive.



(a) Microphone array reference frame for a 2D array.

(b) Camera reference frame for a web camera.

Figure 1.3: Diagrams illustrating a microphone array and camera reference frames.

An alternative technique that has been developed is to use a scanning surface for beamforming which corresponds to the 3D surface geometry of the object that is being acoustically imaged, see Figure 1.2b. For example, the microphone array manufacturing company, *GFaI*, uses a CAD model of an object to generate this 3D scanning surface for 3D beamforming (Heilmann et al., 2008a), see Figure 1.2b and Figure 1.5. This method is said to provide the correct beamforming focus (time delays) for sound propagating from the surface of the object and should provide more accurate measured sound pressure levels, positions, and plotting (Döbler et al., 2008; Maffei and Bianco, 2008; Irimia et al., 2009).

Previous attempts at handling 3D image sharpening all use a 3D grid for deconvolution scan points and not the 3D surface geometry of an object. The deconvolution algorithm *DAMAS*, (Brooks and Humphreys, 2004), was investigated for scan points obtained using multiple 2D planes forming a 3D grid (Brooks and Humphreys, 2005). It was shown that, for planar arrays considered to be practical, the beamforming resolution was poor in the direction along the array axis (longitudinal direction) compared to that in a direction parallel to the array (lateral direction). It was shown that *DAMAS* was able to resolve sources in 3D space but that its longitudinal resolution was limited by the poor longitudinal beamforming resolution of the arrays.

The *DAMAS* algorithm is computationally expensive. To increase the speed of computation, Dougherty (2005a,b) developed *DAMAS2* and *DAMAS3*. These are iterative Fourier-based deconvolution algorithms for beamformed maps using the assumption of a shift-invariant point-spread function (the point spread function is the beamformed map

from a point source). This limits the 3D region to which the deconvolution may be applied. A change of coordinates was used to expand the range of 3D cases to which the convolutional model applies. This resulted in the scanning surfaces being non-planar. It appears that multiple scanning surfaces were used. Dougherty's change of coordinates for *DAMAS2* was investigated by Xenaki et al. (2010) to identify its effect on extending the 3D range over which the shift-invariant point spread function assumption was valid. Dougherty (2010a) used the deconvolution algorithms *CLEAN-SC*, *DAMAS*, *TIDY*, and eigenvalue cancellation for 3D imaging of turbofan engine jet noise using a cage array and a 3D grid composed of transverse planes.

Ravetta (2005) developed the deconvolution algorithm *LORE* for the deconvolution of beamformed acoustic maps. It was applied to the 3D beamformed maps generated using a 3D grid scan of points. It was shown that sources located in 3D space could be distinguished. However, the resolution was limited by the poor longitudinal resolution of the 2D array at low frequencies and the poor calibration of the array at mid to high frequencies.

Sarradj (2010b) used orthogonal beamforming to generate 3D maps of sound sources using a 3D grid of scan points. This method uses eigenvalue decomposition to generate acoustic maps. Sound sources were able to be distinguished in 3D space but it appears, from the plots provided, that a relatively poor resolution was achieved in a direction parallel to the array axis. While this is not deconvolution of beamformed maps, it is included here since it gives similar results and is faster than deconvolution methods. Also see related work by Sarradj (2010a). Wang et al. (2004b) presented the parametric methods wideband *RELAX* (*WB-RELAX*) and wideband *CLEAN* (*WB-CLEAN*) algorithms which were used to construct 3D images of acoustic sources using a 3D grid of scan points. These methods were able to locate sound sources in 3D space under certain conditions but their resolution was limited by the poor resolution of the array in the direction parallel to the array axis. They are not deconvolution of beamformed images but are included here for completeness.

Döbler et al. (2008) of *GFaI* investigated the depth of field of several array types, see Figure 1.4. The depth of field is a measure of the reduction in measured sound pressure level as a 2D scanning surface is offset from the true sound source location in the direction along the array axis. This effect is caused by incorrect focus (time delays) being used in the beamforming and is dependent on the array type, sound frequency, and the position of sound sources in relation to the array. Döbler et al. (2008) also mentioned a position error but gave no mention of the magnitude of this error.

In order to correct for these effects, *GFaI* use a CAD model of an object to generate 3D scanning points, located on the surface of the object, for beamforming. Figure 1.5 gives

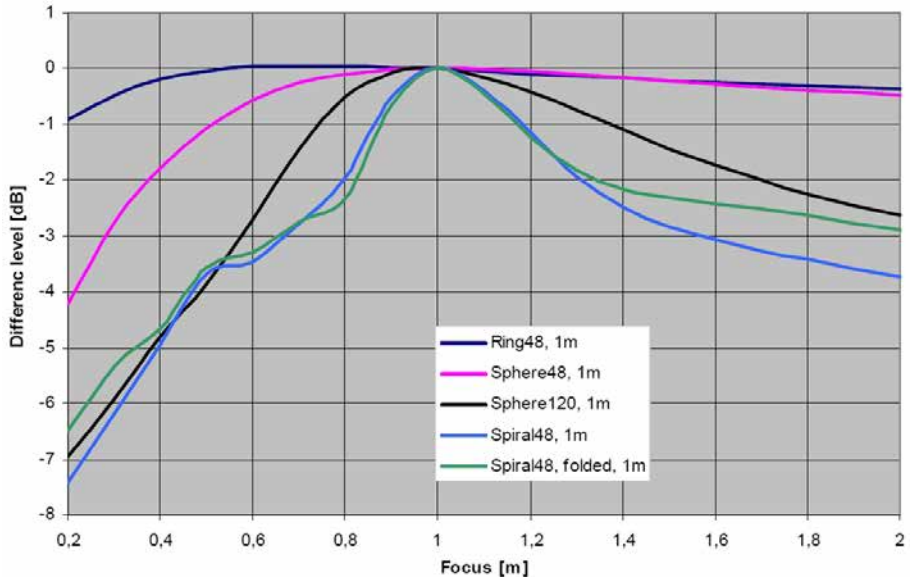


Figure 1.4: Difference in measured sound pressure level as a function of focus distance for beamforming using several array types and a sound source located a metre from the array. Figure is from Döbler et al. (2008).

an example where beamforming has been performed using a CAD model of the interior of a car. *GFaI*state that errors in these 3D acoustic maps can occur if the alignment of the array and CAD model is different from the real alignment of the array and the 3D object, if the CAD model is inaccurate or has missing parts, or if there are scan points included in the beamforming which are 'hidden' from view of the array (Meyer and Döbler, 2006; Schröder, 2009). Such a system is limited to static scenes. It appears that *GFaI* have also used a laser scanner for imaging the interior of rooms for acoustic imaging (GFaI, a,b). They do not appear to have performed 3D deconvolution of these 3D beamformed maps.

Maffei and Bianco (2008) also showed that using the real 3D surface geometry of an object for beamforming gives more accurate SPL (Sound Pressure Levels)measurements compared to using 2D scanning surfaces. Beamforming was performed using the 3D geometry of an object, such as a car, in the *Pininfarina Full Scale Wind Tunnel*, see Figure 1.6. This 3D surface geometry was obtained using stereoscopic imaging from multiple cameras fixed to an arm able to move along one axis. To increase the ability to achieve matches between camera images, which is a main problem with stereoscopic imaging, a static pattern is projected onto the object using projectors attached to the arm. No deconvolution of these 3D beamformed maps appears to have been carried out.

Irimia et al. (2009) ($MicrodB/LMS$) showed that using the real 3D surface geometry of an object as a scanning surface for spherical beamforming gave more accurate sound pressure levels measurements. A laser scanner, attached to a solid spherical array, was used to obtain the 3D geometry of the interior of cars for acoustic imaging, see Figure

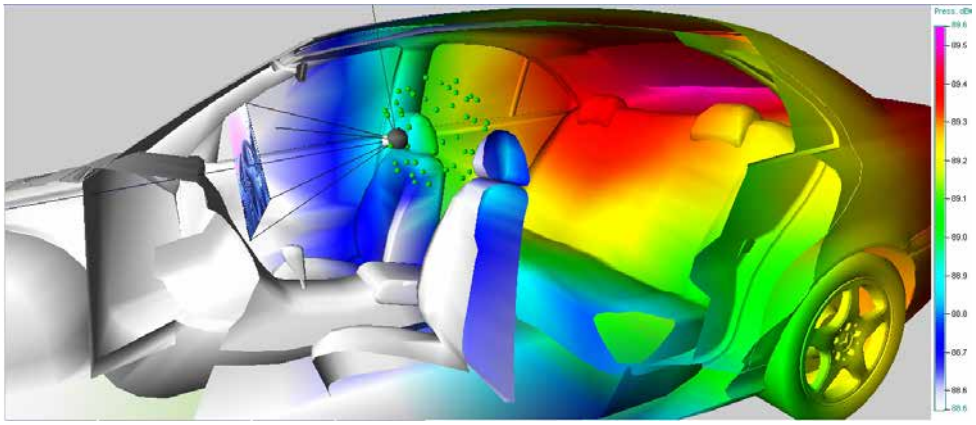


Figure 1.5: Beamformed acoustic map where the 3D scanning surface has been generated by *GFaI* using a CAD model of a car. Figure from Heilmann et al. (2008a).

1.7. It appears that a striped laser system is used which is able to rotate in two axes. Camera images, captured during the scan, were used to generate a 3D photo realistic image of the object. Such a system is limited to static scenes. The 3D surface is then used for acoustic imaging using spherical beamforming and for inverse methods (Irimia et al., 2009; Béguet and Robin, 2010; Lanslots et al., 2010; Lamotte and Deblauwe, 2010). Note that inverse methods do not use deconvolution of beamformed maps but instead attempt to solve for sound pressure levels directly from the microphone pressure data. Also see related literature by Robin and Béguet (2008); Deblauwe and Robin (2009); Lamotte et al. (2009); Schmitt et al. (2010).

Another acoustic imaging technique that uses the surface geometry of an object is the acoustic holography technique referred to as conformal mapping. While this is not 3D beamforming or deconvolution, it is mentioned here for completeness. *Brüel & Kjær* measure the surface geometry of an object using a sonic contact measurement tool. This is used with the *SONAH* algorithm to obtain acoustic maps of sound pressure, sound intensity, or particle velocity, plotted onto the measured 3D surface geometry. This process can be slow and, therefore, generally only applied to small areas where sound sources have been identified and are wanted to be investigated in more detail (Upton et al., 2008; Hald et al., 2008). Li et al. (2008); Yang et al. (2011) use a stereo camera pair and triangulation to determine the distance of a car from the array. A traditional 2D plane, located at this distance, is then used for acoustic holography. This does not use the actual 3D surface geometry of the object. A more detailed description of reports and publications relating to deconvolution is given in Appendix 11.

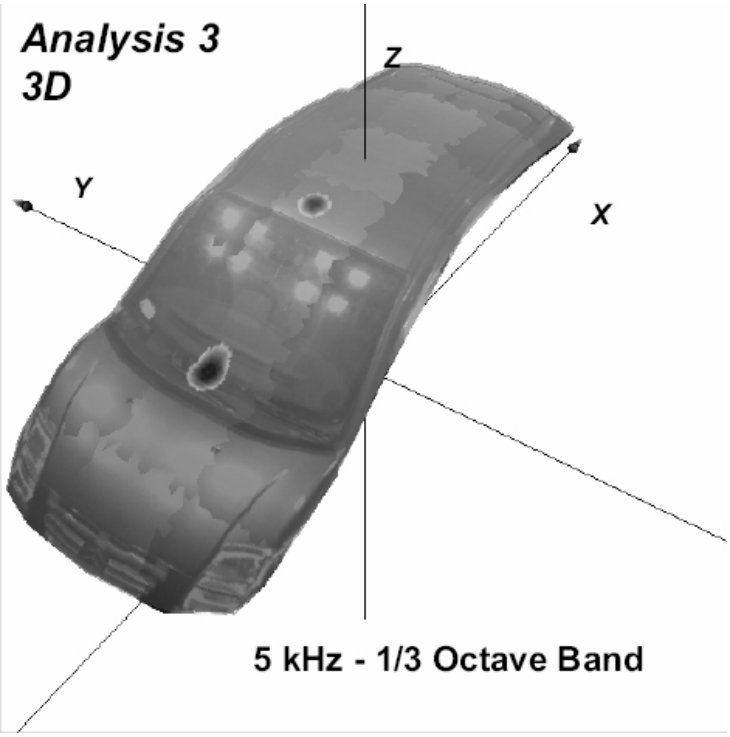
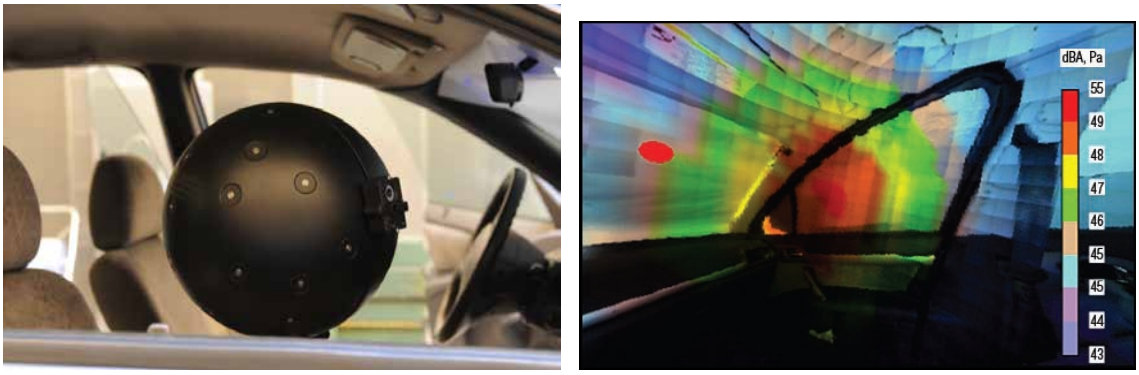


Figure 1.6: Beamformed acoustic map where the scanning surface has been obtained by the *Pininfarina Full Scale Wind Tunnel* using stereoscopic imaging of the 3D geometry of the exterior of a car. Figure from Maffei and Bianco (2008).

1.3 Objectives and Approach Taken in this Work

The deconvolution of beamformed maps, which use a scanning surface corresponding to the 3D surface geometry of an object, has not been performed previously. An advantage of using this method for deconvolution is potential increased accuracy in the measured magnitude and positioning of sound sources, reduced computational cost compared to a 3D scanning grid, and independence of results from the deconvolution resolution in the direction along the array Z axis. Another potential advantage is allowing scattering effects to be included in beamforming and deconvolution theory. A disadvantage of this method, however, is the increased complexity introduced by having to obtain the 3D surface geometry of the object. Also, errors could occur if the sound sources are not located on the surface of the object.

In the work in this thesis, the deconvolution algorithm *CLEAN-SC* is used to compare the accuracy obtained using traditional 2D scanning surfaces and 3D scanning surfaces corresponding to the 3D geometry of an object. The approach taken is to use computer vision techniques to provide accurate measurement of the coordinates of sound sources in the reference frame of a camera in the array. This has enabled a new calibration technique to be developed for obtaining the position of microphones in the reference frame of a camera in the array. This is the first technique to provide combined microphone/camera



(a) Solid spherical microphone array with laser scanning system.

(b) Spherical beamforming 3D acoustic map.

Figure 1.7: Example of spherical beamforming where the scanning surface has been generated by *MicrodB*/*LMS* using a 3D laser scan of the interior of a car. Figures respectively from Irimia et al. (2009) and Lanslots et al. (2010).

calibration. No prior knowledge of microphone coordinates is required and microphones and cameras may be placed in arbitrary positions. This removes the need for fixed array structures and time consuming microphone position measurements and camera alignment. It enables the 3D surface geometry of an object, obtained using computer vision techniques such as structured light scanning, to be utilised directly as a 3D scanning surface for beamforming and deconvolution.

Building upon this new calibration methodology, a new technique has been developed which allows a large number of 2D and 3D experimental deconvolution maps to be generated and automatically processed to form dense point clouds of deconvolution errors. This is in marked contrast to previous error analysis, in which fewer than about six sound sources have been included. This new technique is used to compare the magnitude and position errors obtained using traditional 2D scanning surfaces with those obtained using a scanning surface corresponding to the **real** surface geometry of an object.

1.4 Organisation of Remainder of Thesis

- In Chapter 2, beamforming and deconvolution theory is presented.
- Chapter 3 provides details of the hardware designed and built to obtain experimental data.
- Chapter 4 describes the computer vision methods used extensively in the work covered in this thesis.
- Chapter 5 describes the methodology used to obtain the surface geometry of an object in the reference frame of the microphone array's main camera. The first

method uses known geometry of the object and computer vision techniques. The second method automatically measures the 3D surface geometry using a structured light scanning system.

- Chapter 6 presents a new automatic microphone position and phase calibration technique which uses computer vision to obtain the microphone positions in a camera's reference frame. No *a priori* knowledge of the microphone or camera positions is required. This is the first calibration technique which has been developed to obtain the microphone positions in a camera reference frame. This enables the surface geometry obtained in Chapter 5 to be used for beamforming and deconvolution.

- Chapter 7 presents the methodology used to obtain the SPL and position errors for *CLEAN-SC*. The methodology enables large point clouds of these errors to be obtained. A comparison is made between traditional 2D scanning surfaces and 3D scanning surfaces corresponding to the surface geometry of the object.

- Chapter 8 summarises the conclusions and discusses future work.

- Appendix 9 gives examples of *CLEAN-SC* plots for Chapter 7.

- Appendix 10 contains microphone coordinate information for the arrays used in this work.

- Appendix 11 provides a literature review related to the deconvolution of beamformed acoustic maps.

2

Microphone Phased Array Theory

2.1 Data Model

The acoustic signal received by an array containing M microphones can be modelled as being due to a distribution of monopole sources and uncorrelated channel noise $\boldsymbol{e}$ (Dougherty, 2002, 2005b). The signal at the m^{th} microphone can be modelled, for an non-dispersive medium, in the time domain as

$$\boldsymbol{u}_m(t) = \sum_{s=1}^{S} \frac{q(\vec{\boldsymbol{X}}_s, t - \sigma(\vec{\boldsymbol{X}}_m, \vec{\boldsymbol{X}}_s))}{\boldsymbol{\Lambda}(\vec{\boldsymbol{X}}_m, \vec{\boldsymbol{X}}_s)} + \boldsymbol{e}_m(t), \tag{2.1}$$

where $q(\vec{\boldsymbol{X}}_s, t)$ is the amplitude at time t of the signal emitted by a point source located at $\vec{\boldsymbol{X}}_s$, $\vec{\boldsymbol{X}}_m$ is the coordinate of the m^{th} microphone, $\boldsymbol{\Lambda}(\vec{\boldsymbol{X}}_m,\vec{\boldsymbol{X}}_s)$ is an amplitude factor, and $\sigma(\vec{\boldsymbol{X}}_m,\vec{\boldsymbol{X}}_s)$ is the acoustic propagation time from $\vec{\boldsymbol{X}}_s$ to $\vec{\boldsymbol{X}}_m$, see Figure 2.1. In wind tunnel measurements, $\sigma(\vec{\boldsymbol{X}}_m,\vec{\boldsymbol{X}}_s)$ may be different from that expected from the radial path length due to corrections that 'account for phase changes due to convected and refracted sound transmission through the shear layer to each microphone' (Brooks and Humphreys, 2004).

The time domain data $\boldsymbol{u}_m(t)$ may be converted to the frequency domain using a Short Time Fourier Transform ($STFT$). In this method, the data are divided into J blocks of

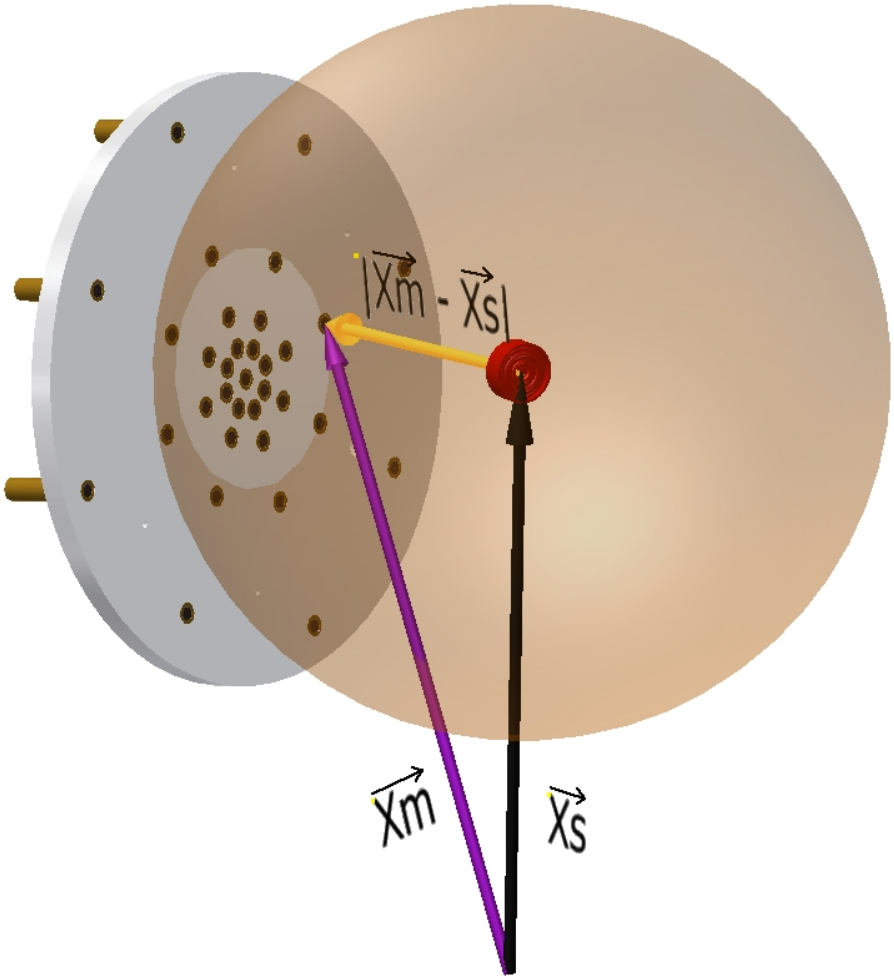


Figure 2.1: Illustration of a microphone at $\vec{\boldsymbol{X}}_m$, in a microphone array, a point source at $\vec{\boldsymbol{X}}_s$, and the spherical wavefront from the source.

length K and the Fast Fourier Transform (FFT) is completed for each data block. For each data block, this gives K frequency bins. For a single frequency bin ω, let $\boldsymbol{U}_m(\vec{\boldsymbol{X}}_s, j)$, $Q(\vec{\boldsymbol{X}}_s, j)$, and $\boldsymbol{E}(j)$ be the Fourier coefficients for the j^{th} data block of $\boldsymbol{u}_m(\vec{\boldsymbol{X}}_s, t)$, $q(\vec{\boldsymbol{X}}_s, t)$, and $\boldsymbol{e}(t)$ respectively. The frequency domain microphone data may be modelled as

$$\boldsymbol{U}(j) = \sum_{s=1}^{S} \boldsymbol{C}(\vec{\boldsymbol{X}}_s)\, Q(\vec{\boldsymbol{X}}_s, j) + \boldsymbol{E}(j), \quad j = 1 \ldots J \tag{2.2}$$

where

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \vdots \\ \boldsymbol{U}_M \end{bmatrix}, \; \boldsymbol{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_M \end{bmatrix}, \; \boldsymbol{E} = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_M \end{bmatrix}$$

and

$$C_m(\vec{\boldsymbol{X}}_s) = \frac{e^{-i\,\omega\,\sigma(\vec{\boldsymbol{X}}_m,\vec{\boldsymbol{X}}_s)}}{\boldsymbol{\Lambda}(\vec{\boldsymbol{X}}_m,\vec{\boldsymbol{X}}_s)} \tag{2.3}$$

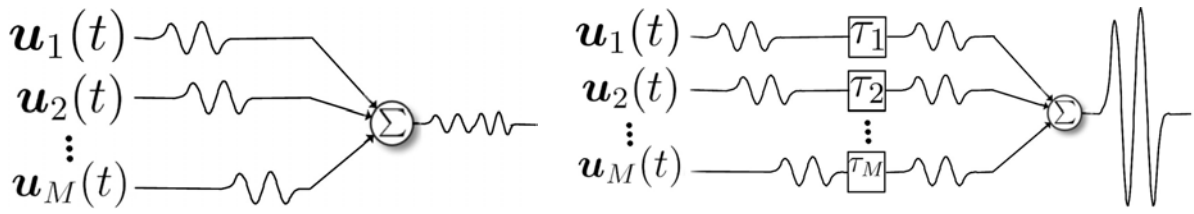
is the m^{th} component of the narrowband array propagation vector $\boldsymbol{C}(\vec{\boldsymbol{X}}_s)$. Note that $\boldsymbol{C}$, $\boldsymbol{U}$, Q, and $\boldsymbol{E}$ are all frequency dependant but the convention is to omit the ω as an index. The array propagation vector is the theoretical Green's function for acoustic propagation from $\vec{\boldsymbol{X}}_s$ to $\vec{\boldsymbol{X}}_m$ (Dougherty, 2004). Factors which effect the propagation of the signal include spherical spreading, air movement, temperature gradients, transmission through a non-homogeneous medium, scattering of the sound by solid objects and atmospheric absorption.

A common expression for the array propagation vector is

$$C_m(\vec{\boldsymbol{X}}_s) = \frac{e^{-i\,\omega\,\|\vec{\boldsymbol{X}}_m-\vec{\boldsymbol{X}}_s\|/c}}{\|\vec{\boldsymbol{X}}_m-\vec{\boldsymbol{X}}_s\|} \tag{2.4}$$

where the term $1/(\|\vec{\boldsymbol{X}}_m-\vec{\boldsymbol{X}}_s\|)$ allows for attenuation due to spherical spreading and c is the speed of sound. This assumes that the sound propagates through still, homogeneous air, with no scattering effects and minimal atmospheric attenuation. Note that in some literature a 4π term is added to the denominator of the propagation vector.

2.2 Beamforming



(a) Summing of out of phase microphone channel data resulting in destructive interference.

(b) Microphone data are delayed before summing resulting in constructive interference.

Figure 2.2: Diagram illustrating the delay and sum beamforming process.

2.2.1 Time Domain Beamforming

The acoustic signals arriving at the M microphones are converted to analog voltage signals and then into M channels of sampled digital data. Depending on the position of the sources and the geometry of the microphones in an array, the signals received will be out

of phase with each other. In Delay and Sum (DAS) beamforming, the signals are delayed in the time domain and then summed. If the delays are chosen so that the delayed signals from a monopole source at $\vec{\boldsymbol{X}}_s$ are in phase, the summed signals from this location will interfere constructively and amplification of the total received signal from this point will occur. If, however, the summed signals are out of phase, destructive interference will occur, see Figure 2.2.

To generate an acoustic map, a scanning surface is defined which has N grid points located at positions $\vec{\boldsymbol{\xi}}$. For the n^{th} grid point and the m^{th} microphone, the acoustic delay time τ_{mn} is calculated, using the theoretical propagation time from grid point n to microphone m. Delay and sum beamforming is then used to obtain a beamformed intensity value

$$\boldsymbol{b}(\vec{\boldsymbol{\xi}}_n) = \frac{1}{M} \left\langle \left| \sum_{m=1}^{M} w_m \, \boldsymbol{u}_m(t - \tau_{mn}) \right|^2 \right\rangle, \quad \vec{\boldsymbol{\xi}}_n = \vec{\boldsymbol{\xi}}_1 \ldots \vec{\boldsymbol{\xi}}_N \tag{2.5}$$

where there are M microphones, w_m are spatial shading weights, and $\langle\rangle$ is the time average of the square of the summed delayed signals. This is repeated for all grid points to form the acoustic map (Dougherty, 2004; Jaeckel, 2006). Assuming propagation in a uniform non-moving medium, the time delays can be chosen to be

$$\tau_{mn} = \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|/c. \tag{2.6}$$

For a sound source located in the far field, the delay times are

$$\tau_{mn} = \frac{\vec{\boldsymbol{\kappa}}_n \cdot \vec{\boldsymbol{X}}_m}{c} \tag{2.7}$$

where $\vec{\boldsymbol{\kappa}}$ is a unit vector in the direction of propagation and the '$\cdot$' symbol is the dot product (Christensen and Hald, 2004).

2.2.2 Frequency Domain Beamforming

Beamforming may also be performed in the frequency domain. While time domain beamforming generated wide frequency band acoustic maps, frequency domain beamforming generates beamformed maps at a single frequency or the central frequency of a frequency band such as $1/3^{rd}$ or $1/12^{th}$ octave band. Equation (2.5) may be written in the frequency domain as

$$\boldsymbol{b}_{FD}(\vec{\boldsymbol{\xi}}_n) = \left\langle \left| \, \mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_n) \, \boldsymbol{U} \right|^2 \right\rangle, \quad \vec{\boldsymbol{\xi}}_n = \vec{\boldsymbol{\xi}}_1 \ldots \vec{\boldsymbol{\xi}}_N, \tag{2.8}$$

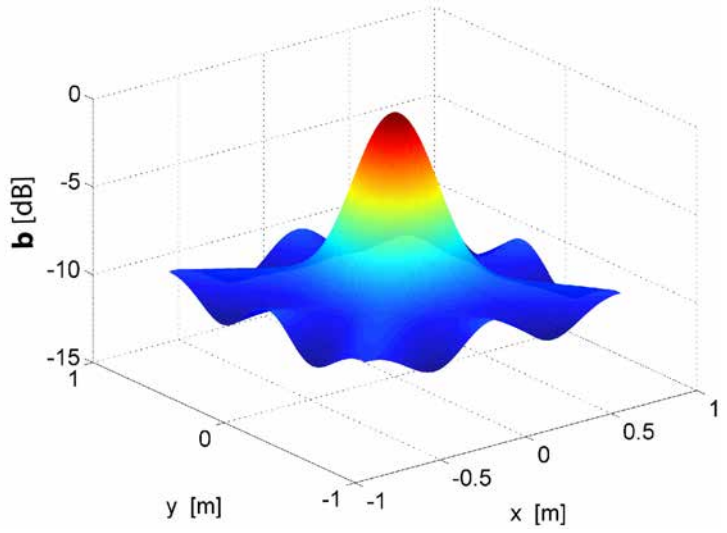


Figure 2.3: Example of a near-field normalised beamformed map for a $SADA$ microphone array at 3.3kHz.

where

$$\mathbf{w}(\vec{\boldsymbol{\xi}}_n) = \begin{bmatrix} \mathrm{w}_1(\vec{\boldsymbol{\xi}}_n) \\ \mathrm{w}_2(\vec{\boldsymbol{\xi}}_n) \\ \vdots \\ \mathrm{w}_M(\vec{\boldsymbol{\xi}}_n) \end{bmatrix},$$

is called the array steering vector, † is the complex conjugate transpose, and $\langle\ \rangle$ is the time average over data blocks, see Equation (2.12). Expanding Equation (2.8) gives

$$\boldsymbol{b}_{FD}(\vec{\boldsymbol{\xi}}_n) = \ \mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_n)\ \left\langle \boldsymbol{U}\,\boldsymbol{U}^{\dagger} \right\rangle\ \mathbf{w}(\vec{\boldsymbol{\xi}}_n) \tag{2.9}$$

(Dougherty, 2002, 2005b). Substituting in

$$\mathbf{G} = \left\langle \boldsymbol{U}\,\boldsymbol{U}^{\dagger} \right\rangle, \tag{2.10}$$

gives

$$\boldsymbol{b}(\vec{\boldsymbol{\xi}}_n) = \ \mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_n)\ \mathbf{G}\ \ \mathbf{w}(\vec{\boldsymbol{\xi}}_n), \quad \vec{\boldsymbol{\xi}}_n = \vec{\boldsymbol{\xi}}_1 \ldots \vec{\boldsymbol{\xi}}_N \tag{2.11}$$

where

$$\mathbf{G} = \begin{bmatrix} G_{11} & G_{12} & \ldots & G_{1M} \\ \vdots & G_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ G_{M1} & \ldots & \ldots & G_{MM} \end{bmatrix}$$

is referred to as the Cross Spectral Matrix CSM. Note that $\mathbf{w}(\vec{\boldsymbol{\xi}}_n)$ and $\mathbf{G}$ are different for each frequency and should, therefore, ideally be written as $\mathbf{w}(\vec{\boldsymbol{\xi}}_n, \omega)$ and $\mathbf{G}(\omega)$. However,

since frequency domain beamformed maps are calculated using a single frequency bin or using the central frequency of an octave frequency band the convention is to omit ω. Equation (2.11) gives a single value for each scan point. To generate a beamformed map this equation is repeated for all points (Dougherty, 2005b). Figure 2.3 shows an example of a beamformed map.

The steering vectors given above assume monopole sources. If dipole sources are imaged, different steering vectors are required. Examples of a dipole include a jet where back circulation occurs or an aeroplane wing viewed on its edge. Consider a planar array with the dipoles in a plane parallel to the array plane. Half of the array will see positive pressures and the other half negative. This results in a null in the acoustic map at the dipole location but two peaks at either side. The position of the peaks change with frequency. This can be corrected for by modifying the steering vectors. However, these steering vectors will not work for monopole sources (Liu et al., 2008; Jordan et al., 2002).

Cross Spectral Matrix

The CSM is obtained using the Short Time Fourier Transform ($STFT$). The time domain data are divided into J blocks of length K. The data blocks may be overlapped in time using a set percentage, such as 50%. A Fast Fourier Transform FFT is then obtained for each data block. A windowing function, such as a Hann window, may be applied to each data block before the FFT is computed. For each frequency bin, a CSM is then calculated. The components of each CSM are given by

$$\mathbf{G} = \frac{1}{W_m J} \sum_{j=1}^{J} \left[\boldsymbol{U}(j)\, \boldsymbol{U}^{\dagger}(j) \right] \tag{2.12}$$

where

$$\boldsymbol{U}(j)\, \boldsymbol{U}^{\dagger}(j) = \begin{bmatrix} \boldsymbol{U}_1(j)\, \boldsymbol{U}_1^{\dagger}(j) & \boldsymbol{U}_1(j)\, \boldsymbol{U}_2^{\dagger}(j) & \dots & \boldsymbol{U}_1(j)\, \boldsymbol{U}_M^{\dagger}(j) \\ \vdots & \boldsymbol{U}_2(j)\, \boldsymbol{U}_2^{\dagger}(j) & & \vdots \\ \vdots & & \ddots & \vdots \\ \boldsymbol{U}_M(j)\, \boldsymbol{U}_1^{\dagger}(j) & \dots & \dots & \boldsymbol{U}_M(j)\, \boldsymbol{U}_M^{\dagger}(j) \end{bmatrix} \tag{2.13}$$

and W_m is a windowing constant. The CSM is a Hermitian matrix with the lower triangular elements being the complex conjugates of the upper triangular elements (Brooks and Humphreys, 1999). The off-diagonal elements are the cross spectra of the microphone signals, while the diagonal elements are the auto spectra of the microphone signals. If a sufficient number of averaging data blocks are used, the channel self noise $\boldsymbol{E}$ will add

constructively in the diagonal components of the CSM but will be cancelled out in the off-diagonal components. To remove channel noise, these diagonal elements are often set equal to zero since they do not contribute to array resolution and introduce channel self noise. This method, known as Diagonal Removal (DR), is especially important for wind tunnel measurements where the wind noise, due to turbulence on the microphones, may be significantly greater than the acoustic signal being measured (Dougherty, 2002). See Dougherty (2004) for a time domain version of DR.

Array Steering Vector

The acoustic signal propagating from a sound source located at $\vec{\boldsymbol{X}}_s$ will arrive at each microphone in an array at different times and experience different amounts of attenuation, depending on the positions of the microphones relative to the sound source. The propagation vector $\boldsymbol{C}(\vec{\boldsymbol{X}}_s)$ mathematically describes these propagation effects. The role of the steering vector $\mathbf{w}(\vec{\boldsymbol{\xi}}_n)$ in frequency domain beamforming is to attempt to correct for these propagation effects by phase shifting and amplifying the microphone channel data.

The steering vector is commonly normalised such that
$\mathbf{w}^\dagger(\vec{\boldsymbol{\xi}}_n)\ \ \mathbf{w}(\vec{\boldsymbol{\xi}}_n) = 1$ or

$$\mathbf{w}(\vec{\boldsymbol{\xi}}_n) = \frac{\boldsymbol{C}(\vec{\boldsymbol{\xi}}_n)}{\|\boldsymbol{C}(\vec{\boldsymbol{\xi}}_n)\|}. \tag{2.14}$$

For free field propagation, Dougherty (2002) shows that

$$\mathrm{w}_m(\vec{\boldsymbol{\xi}}_n) = \frac{\frac{e^{-i\,\omega\,\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|/c}}{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|}}{\sqrt{\sum_{m=1}^{M} \frac{1}{(\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|)^2}}}. \tag{2.15}$$

The steering vector aims to counter the phase shifts resulting from the acoustic propagation. If diagonal removal of the CSM is used, different weighting of the steering vector may be used (Brooks and Humphreys, 2004; Sijtsma, 2007a). An alternative is to have

$$\mathbf{w}^\dagger(\vec{\boldsymbol{\xi}}_n)\ \boldsymbol{C}(\vec{\boldsymbol{\xi}}_n) = 1 \tag{2.16}$$

or, for free field propagation,

$$\mathrm{w}_m(\vec{\boldsymbol{\xi}}_n) = \frac{1}{M}\ \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|\ e^{-i\,\omega\,\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|/c}, \tag{2.17}$$

where † is the complex conjugate transpose (Brooks and Humphreys, 2004). In addition to correcting for phase shifts, Equation (2.17) also attempts to correct each scan point

for spherical spreading during propagation. This should allow the sound pressure levels of sound sources to be measured relative to the source locations as apposed to measuring the sound pressure level relative to the array. However, it does give more gain to scan points further from the array. It can cause some position errors by biassing main lobe peaks and side lobes towards scan points further from the array and allocates more gain on scan points with lower signal to noise ratio.

For a sound source located in the far field, the wavefronts arriving at the array are planar and the difference in attenuation between microphones due to spherical spreading is negligible. The propagation vector may, therefore, be written as

$$\mathrm{w}_m(\vec{\boldsymbol{\kappa}}_n) = \frac{1}{M} e^{-i\,\omega\,\vec{\boldsymbol{\kappa}}_n\,\cdot\,\vec{\boldsymbol{X}}_m/c} \tag{2.18}$$

where $\vec{\boldsymbol{\kappa}}_n$ is a unit vector in the direction of arrival (Underbrink, 2002; Christensen and Hald, 2004).

Note that, in the literature, different conventions exist for the sign of the exponential term in the steering vector. This is related to the way that the CSM is generated. A negative sign is used here since the CSM is generated using $\mathbf{G} = \boldsymbol{U}\ \boldsymbol{U}^{\dagger}$, where $\boldsymbol{U}$ is a $[M \times 1]$ column vector. However, some literature instead calculates the CSM using $\mathbf{G} = \boldsymbol{U}^{\dagger}\ \boldsymbol{U}$ where instead $\boldsymbol{U}$ is now a $[1 \times M]$ row vector. The correct sign must be used depending on which method is used to generate the CSM.

Relationship Between the Eigenvectors of a CSM and the Propagation Vector

Recall that in Equation (2.2), the frequency domain data were modelled as being due to monopole sources and channel noise using

$$\boldsymbol{U}(j) = \sum_{s=1}^{S} \boldsymbol{C}(\vec{\boldsymbol{X}}_s)\ Q(\vec{\boldsymbol{X}}_s, j) + \boldsymbol{E}(j), \quad j = 1 \ldots J.$$

For a unit magnitude source, $Q(\vec{\boldsymbol{X}}_s,j) = 1$, with no uncorrelated channel noise

$$\boldsymbol{U}(\vec{\boldsymbol{X}}_s, j) = \boldsymbol{C}(\vec{\boldsymbol{X}}_s). \tag{2.19}$$

Substituting this into Equation (2.12), the CSM may be modelled, for a unit source at $\vec{\boldsymbol{X}}_s$, as

$$\mathbf{G} = \boldsymbol{C}(\vec{\boldsymbol{X}}_s)\ \boldsymbol{C}(\vec{\boldsymbol{X}}_s)^{\dagger} \tag{2.20}$$

(with appropriate scaling). Note that this equation gives a theoretical CSM rather than the measured CSM given by Equation (2.12). The eigenvectors and eigenvalues of $\mathbf{G}$ are often used to classify sound sources. The eigenvectors corresponding to larger eigenvalues are assumed to be related to the propagation vectors $\boldsymbol{C}$ of sound sources, while the eigenvalues corresponding to small eigenvalues are assumed to be related to noise in the system. Consider a CSM due to a single unit source at $\vec{\boldsymbol{X}}_s$. Multiplying the cross spectral matrix $\mathbf{G}$ and the propagation vector $\boldsymbol{C}(\vec{\boldsymbol{X}}_s)$ gives

$$\mathbf{G}\,\boldsymbol{C}(\vec{\boldsymbol{X}}_s) = \left[\boldsymbol{C}(\vec{\boldsymbol{X}}_s)\,\boldsymbol{C}(\vec{\boldsymbol{X}}_s)^{\dagger}\right]\,\boldsymbol{C}(\vec{\boldsymbol{X}}_s) = \boldsymbol{C}(\vec{\boldsymbol{X}}_s)\,|\boldsymbol{C}(\vec{\boldsymbol{X}}_s)|^2$$

Using the definition of an eigenvector

$$\mathbf{G}\,\boldsymbol{C}(\vec{\boldsymbol{X}}_s) = \gamma\,\boldsymbol{C}(\vec{\boldsymbol{X}}_s), \tag{2.21}$$

where γ is an eigenvalue of $\mathbf{G}$. The eigenvalue corresponding to the only non-zero eigenvector of $\mathbf{G}$ is equal to the normalised propagation vector $\boldsymbol{C}(\vec{\boldsymbol{X}}_s)/\|\boldsymbol{C}(\vec{\boldsymbol{X}}_s)\|$ and, using Equation (2.14), also equal to the steering vector $\mathbf{w}(\vec{\boldsymbol{X}}_s)$ (Dougherty, 2002; Sarradj, 2010a,b).

Eigenvalue Calibration Using a Speaker

The actual beamformed map generated by a point source may differ from the theoretical beamformed map due to a number of factors. These include errors in microphone positions, temperature (and hence speed of sound estimation), the calculated CSM, and variation in the amplitude and phase between microphone channels (Yardibi, 2009). Eigenvalue calibration can help to correct for these errors. The calibration data are obtained by setting up a speaker at position $\vec{\boldsymbol{X}}_s$, in front of the array, in an anechoic-like environment. Acoustic foam can be set up around the array and speaker if required. White noise is played on the speaker and a CSM obtained. The leading eigenvalue of the CSM should be at least 10 times higher than the next highest eigenvalue (Dougherty, 2002). If it were nearly equal to the next biggest eigenvalue it would mean that a secondary source was almost as loud (perhaps a reflection or background noise), meaning that a poor calibration would be achieved. If this is the case, the speaker should be played louder or the octave bandwidth reduced. The dominant eigenvector $\boldsymbol{v}_{cal}$ of the CSM is taken to be the true normalised steering vector $\boldsymbol{C}(\vec{\boldsymbol{X}}_s)/\|\boldsymbol{C}(\vec{\boldsymbol{X}}_s)\|$. This eigenvector and the corresponding theoretical steering vector $\mathbf{w}(\vec{\boldsymbol{X}}_s)$ is used to calculate the calibration

matrix

$$\mathbf{D_{cal}} = \begin{bmatrix} \frac{(\boldsymbol{v}_{cal})_1}{\mathrm{w}_1} & 0 & \dots & 0 \\ 0 & \frac{(\boldsymbol{v}_{cal})_2}{\mathrm{w}_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \frac{(\boldsymbol{v}_{cal})_M}{\mathrm{w}_M} \end{bmatrix} \tag{2.22}$$

The diagonal terms contain amplitude and phase correction factors. This is used to correct the CSM using

$$\mathbf{G}_{cal} = \mathbf{D_{cal}}^{\dagger}\ \mathbf{G}\ \mathbf{D_{cal}}. \tag{2.23}$$

(Dougherty, 2002; Underbrink, 2002; Ravetta, 2005; Yardibi et al., 2009).

Octave Frequency Bands

Octave frequency bands, such as $1/3^{rd}$ or $1/12^{th}$ octave bands, are generally used for analysis. This can be achieved by performing narrow band beamforming for each spectral line and summing the beamformed map into octave bands. A faster method, however, is to sum the CSMs into octave frequency bands and use the central frequencies for the array steering vector. The central, lower, and upper frequency range of the $1/b^{th}$ octave frequency bands can be respectively given by

$$f_{c_n} = 2^{n/b} \tag{2.24}$$

$$f_{l_n} = 2^{-\frac{1}{2\,b}}\ f_{c_n} \tag{2.25}$$

$$f_{u_n} = 2^{\frac{1}{2\,b}}\ f_{c_n} \tag{2.26}$$

(Ravetta, 2005).

2.3 Properties of Beamformed Maps

2.3.1 Point Spread Function and Sidelobes

The point spread function is the beamformed map that would result from a unit point source located at $\vec{\boldsymbol{X}}_s$. The shape of the point spread function depends on the frequency and position of the point source and the number and geometry of the microphones in the array. Figure 2.4 shows an example of a point spread function. Here it can be seen that the point spread function consists of a main lobe and side lobes. To enhance plotting, side lobes below -18dB have not been shown.

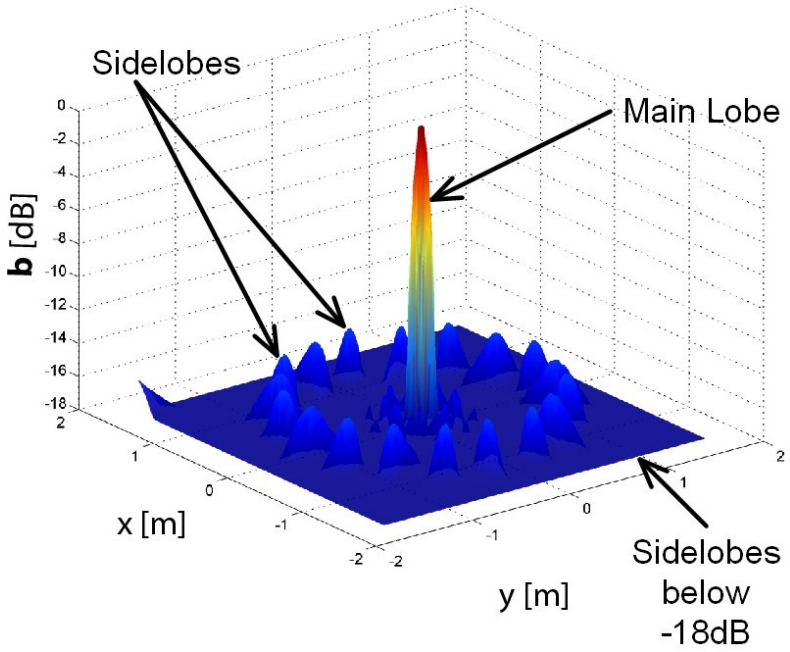


Figure 2.4: An example of a near-field point spread function showing the main lobe and sidelobes for an Underbrink multi-arm spiral array using frequency domain beamforming at 3.3kHz.

The dynamic range refers to the difference in dB from the peak intensity of the main lobe to the peak intensity of the largest sidelobe. In a beamformed acoustic map, it can be hard to identify secondary sound sources which have a main lobe peak below the sidelobe peaks of the primary sound source. The array resolution or beamwidth (BW) is the width, in metres, for near field imaging, of the main beam at 3dB down from the main peak. The array resolution can be described by

$$BW_R \propto \frac{R\,\lambda}{D} \tag{2.27}$$

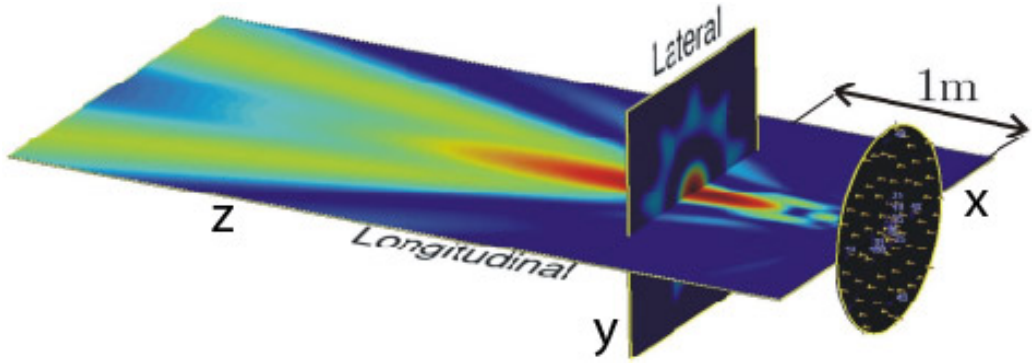
where R is the distance of the sound source from the center of the array, λ is the wavelength, and D is the diameter of the array. It is also related to the placement of microphones in the array (Underbrink, 2002).

The beamwidth affects the ability of acoustic imaging to resolve closely spaced sound sources. This can be expressed in terms of the Rayleigh criterion

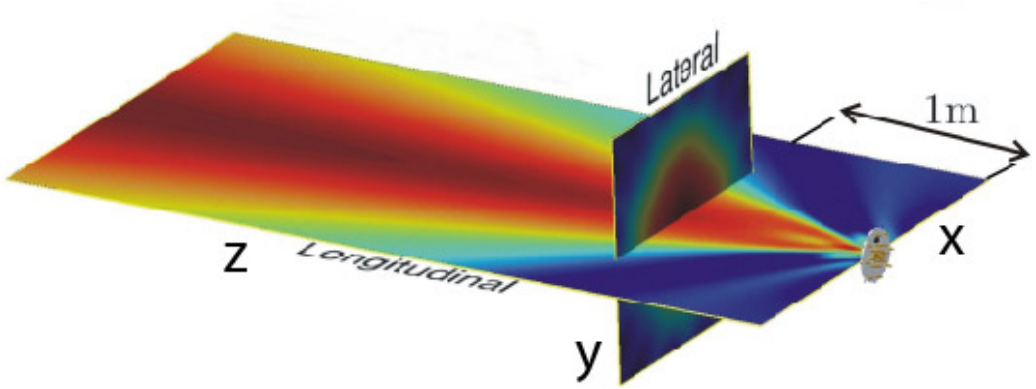
$$\mathcal{W} = \frac{d\,D}{\lambda\,R} \tag{2.28}$$

where d is the separation of the sound sources and $\mathcal{W} = 1.22$ is the Rayleigh limit. This is the distance between the peak and the first zero of an ideal circular diffraction pattern, below which multiple sources cannot be separated (Ramachandran et al., 2012).

The beamwidth in the plane parallel to the array (lateral beamwidth) is generally smaller



(a) Underbrink multi-arm spiral array beamformed maps for 3.3kHz.



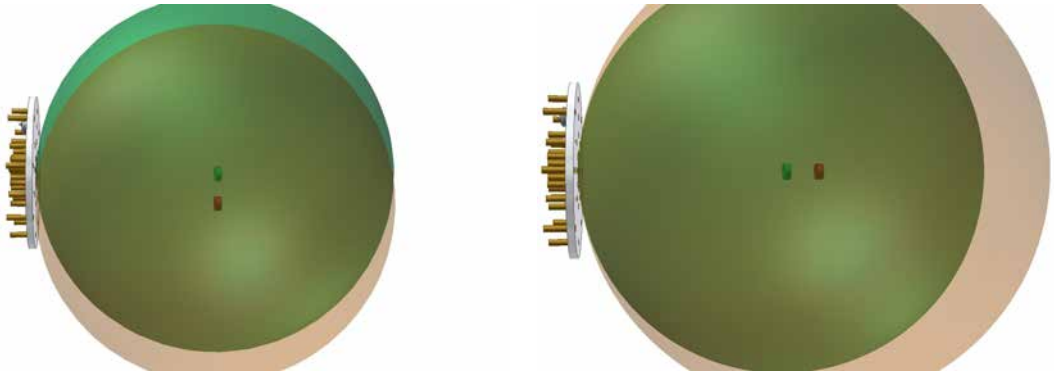
(b) $SADA$ array beamformed maps for 6.6kHz.

Figure 2.5: Simulated array lateral and longitudinal beamformed maps for a unit point source located at the centre of the lateral beamformed maps one metre from the array. These may also be referred to as lateral and longitudinal point spread functions. The dimensions of the longitudinal beamformed maps are 2×4m. Normalised steering vectors have been used for the simulation.

than the beamwidth is the direction along the array axis (so called longitudinal beamwidth) for most arrays commonly used. Figure 2.5 shows simulated longitudinal and lateral beamformed maps for an Underbrink multi-arm spiral array (Underbrink, 2002) and a Small Aperture Directional Array ($SADA$) array (Humphreys et al., 1998). This difference in beamwidth results from the fact that a greater difference in acoustic propagation distance between microphones generally occurs if the sound sources are moved in the X-Y plane (lateral beamwidth) compared to if the sound sources were moved the same distance in the Z axis direction (longitudinal direction), see Figure 2.6 as an example.

Depth of Field of an Array

The longitudinal resolution, has a high dependance on the geometry of the microphones in the array (Brooks and Humphreys, 2005). Consider an acoustic wave propagating from a sound source located on the Z axis of the array to microphones in a planar array. If all the microphones are in the X-Y plane in a circular pattern centered on the Z axis, the



(a) Wavefronts for two sound sources spatially separated 50mm in the lateral direction.

(b) Wavefronts for two sound sources spatially separated mm in the longitudinal direction.

Figure 2.6: Diagram illustrating the difference in wavefronts for two spatially separated sound sources.

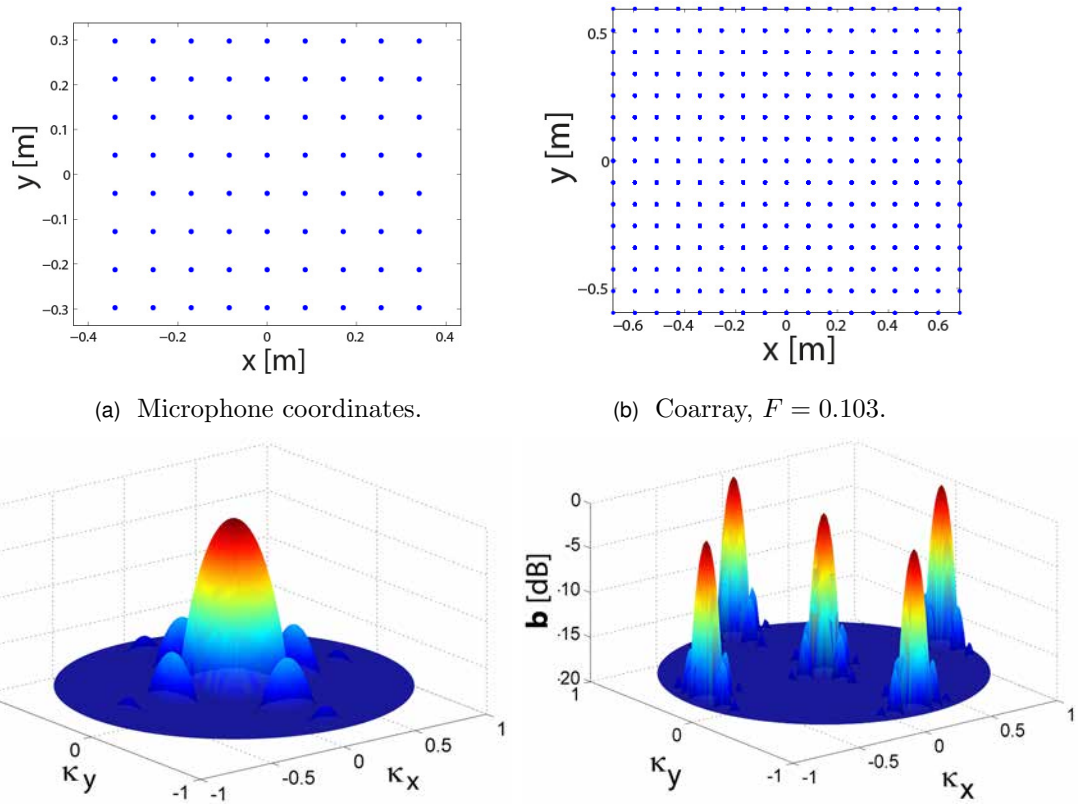
acoustic signal will arrive at all the microphones at the same time regardless how far from the array the sound source is located. This means that if one performs delay and sum beamforming, any scan point along the Z axis will result in constructive interference of the summed signals. The beamwidth along the Z direction will be infinite at all frequencies for this configuration of microphones. Consider now that the microphones are distributed in the X-Y at different distances from the Z axis, say in a spiral pattern. The acoustic signal from a sound source on the Z axis will, therefore, arrive at each microphone at a different time. The closer the sound source is to the array and the greater the difference in distances of the microphones from the Z axis, the more difference in travel path there will be between microphones. At low frequencies, this difference in travel path may be a small fraction of a wavelength. As the frequency increases, however, this difference in travel path may became close to a half a wavelength. Delay and sum beamforming will, therefore, give different degrees of destructive interference for scan points along the Z axis depending on the position of the sound source, the wavelength, the diameter array and placement of the microphones. The longitudinal resolution will, therefore, also depend on these factors.

The depth of field of an array describes the amount that the measured sound pressure level, obtained using beamforming, decreases as the scanning surface is offset from the true sound source location, in the direction perpendicular to the array. This reduction is due to incorrect time delays, or focus, being used. This results in some degree of destructive interference of the microphone signal during the delay and sum beamforming. Döbler et al. (2008) showed that this error increased as frequency increased and as the source was positioned closer to the array. It was also shown that this error varied significantly with the array type. For example, a spiral array was shown to have a measured SPL reduction of several dB for a scan surface offset of a metre from the sound source while a

ring array had relatively little reduction in measured SPL, see to Figure 1.4.

2.3.2 Spatial Aliasing

In order to achieve high spatial resolution, the microphone array should have as wide an aperture as possible. Since there will be a limited number of microphone channels available, increasing the array's diameter will result in increased spacing between microphones. However, to avoid spatial aliasing, for sound propagating from all directions, the minimum spacing between microphones must be less than half the wavelength of the highest frequency of the acoustic signal. If this condition is not met, a sound source may cause multiple main lobes in the beamformed map and it can be impossible to distinguish which main lobe peak corresponds to the location of the real acoustic source. This is similar to temporal aliasing where the Nyquist Theorem states that a time domain analog signal must be sampled at twice the sampling rate of the highest frequency in the signal to avoid frequency components being aliased to lower frequencies.



(a) Microphone coordinates.

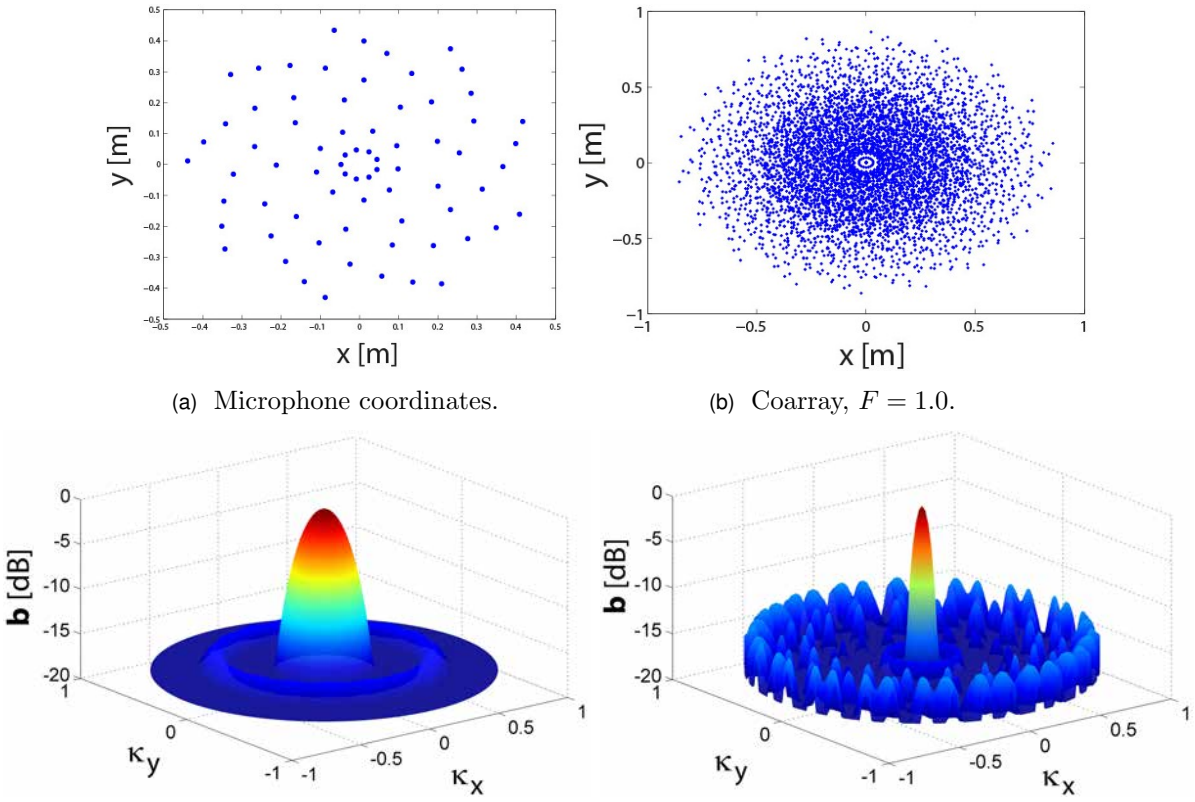
(b) Coarray, $F = 0.103$.

(c) Point spread function at 1.5kHz.

(d) Point spread function at 4.5kHz. Note spatial aliasing.

Figure 2.7: Rectangular regularly spaced microphone array.

The spacing between microphones may be visualised by the "coarray". This is a plot of vector spacings between microphone positions. Each vector spacing is represented as a



(a) Microphone coordinates.

(b) Coarray, $F = 1.0$.

(c) Point spread function at 1.5kHz.

(d) Point spread function at 4.5kHz.

Figure 2.8: Underbrink multi-arm spiral far field array pattern.

dot. The distance and direction of the dot relative to the origin give the magnitude and direction of the vector. These vectors can be calculated by

$$\vec{\boldsymbol{X}}_p = \vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_{m'}, \tag{2.29}$$

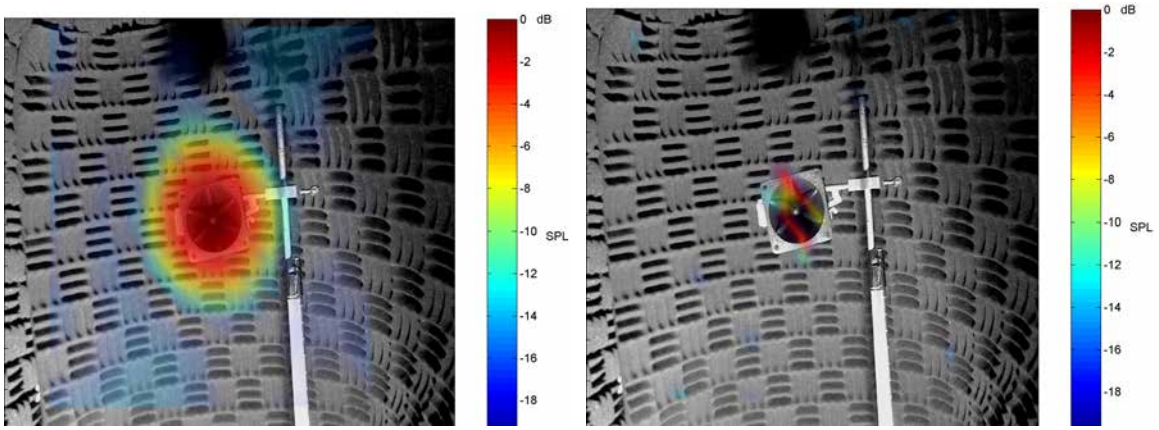
where $m = 1, 2, \cdots, M$ and $m' = 1, 2, \cdots, M$. Let μ be the number of unique vector spacings. Since there are M^2 vector spacings and M of these must be zero, the maximum possible number of unique vectors is given by

$$\mu_{max} = M^2 - M + 1, \tag{2.30}$$

if the zero vector is included. The ratio

$$F = \frac{\mu}{\mu_{max}} \leq 1 \tag{2.31}$$

is referred to as the figure of merit of an array. When designing an array for broadband frequency purposes, it is usually desirable to use an array design where $F = 1$. This is referred to as being a non-redundant array (Underbrink, 2002; Dougherty, 1998; Hoctor and Kassam, 1990). This can be seen by comparing point spread functions for an evenly



(a) Beamformed acoustic map (b) *DAMAS* acoustic map

Figure 2.9: Example of a beamformed map and corresponding *DAMAS* map. Note that the deconvolution resolution used was low due to the high processing time. The plotting method used, a mesh plot, was not ideal for this resolution and a more appropriate plotting method would have been to plot the deconvolution map as pixels.

spaced rectangular grid array, see Figure 2.7, and an Underbrink multi-arm spiral array, see Figure 2.8. While the rectangular array ($F = 0.103$) shows aliasing at 4.5kHz, the spiral array ($F = 1.000$) shows no secondary main lobes to frequencies above 50kHz, even though the same number of microphones and similar array diameter is used for both arrays.

2.4 Deconvolution

A point source located at $\vec{\boldsymbol{X}}_s$ may be described by the source distribution

$$q = q_o \, \delta(\vec{\boldsymbol{\xi}} - \vec{\boldsymbol{X}}_s) \tag{2.32}$$

where q_o is the amplitude of the point source and δ is the Dirac delta function. Ideally the resulting beamformed image would be given by $b(\vec{\boldsymbol{\xi}}) = q_o \, \delta(\vec{\boldsymbol{\xi}} - \vec{\boldsymbol{X}}_s)$. In reality, however, the beamformed map can be considered as the convolution of q with the point spread function $\mathbf{psf}(\vec{\boldsymbol{\xi}}, \vec{\boldsymbol{X}}_s)$. Deconvolution aims to take a beamformed acoustic map and use the known properties of the array to get the 'true' source distribution. Two deconvolution algorithms are outlined below, *DAMAS* and *CLEAN-SC*.

2.4.1 *DAMAS*

The *DAMAS* (Deconvolution Approach for the Mapping of Acoustic Sources) algorithm was introduced by Brooks and Humphreys in 2004. It assumes that the beamformed

map can be described by the combination of theoretical point spread functions due to statistically independent noise sources. These point spread functions are assumed to vary depending on the position of the noise sources. Errors can occur if the theoretical point spread functions are incorrect.

The assumption is made that the source distribution may be modelled as monopole sources located at the beamforming scan points. Consider a beamformed map containing N scan points. For the s^{th} scan point having coordinate $\vec{\boldsymbol{\xi}}_s$, a $1 \times N$ theoretical point spread function is calculated assuming a monopole sound source located at $\vec{\boldsymbol{\xi}}_s$ using

$$\mathbf{psf}(\vec{\boldsymbol{\xi}}_n, \vec{\boldsymbol{\xi}}_s) = \ \mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_n)\ [\boldsymbol{C}(\vec{\boldsymbol{\xi}}_s)\boldsymbol{C}^{\dagger}(\vec{\boldsymbol{\xi}}_s)]\ \ \mathbf{w}(\vec{\boldsymbol{\xi}}_n), \quad \vec{\boldsymbol{\xi}}_n = \vec{\boldsymbol{\xi}}_1 \ldots \vec{\boldsymbol{\xi}}_N \tag{2.33}$$

where the steering vectors $\mathbf{w}$ are of the form given in Equation (2.17). This point spread function becomes the s^{th} row of matrix $\mathbf{A}$. The beamformed map is assumed to be described by

$$\mathbf{Y} = \mathbf{A}\ \mathbf{Q} \tag{2.34}$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \ldots & A_{1N} \\ \vdots & A_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ A_{N1} & \ldots & \ldots & A_{NN} \end{bmatrix}, \ \mathbf{Q} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_N \end{bmatrix},$$

and $\mathbf{Q}$ is a column matrix of noise source levels. A Gauss-Seidel interactive method is then used to solve for $\mathbf{Q}$ using

$$\mathbf{Q}_s = \max\left(0, \quad \mathbf{Y}_s - \left[\sum_{n=1}^{s-1} \mathbf{A}_{sn}\mathbf{Q}_n + \sum_{n=s+1}^{N} \mathbf{A}_{sn}\mathbf{Q}_n\right]\right). \tag{2.35}$$

Iteration may be performed using $s = 1 \ldots N$ or may alternate between $s = 1 \ldots N$ and $s = N \ldots 1$ on consecutive iterations (Brooks and Humphreys, 2004). *DAMAS* was extended to allow for 3D scanning grids (Brooks and Humphreys, 2005) and coherent sources (Brooks and Humphreys, 2006b).

The *DAMAS* algorithm is computationally expensive and hence can be time-consuming to solve. For example, a 100×100 beaformed map will have a matrix $\mathbf{A}$ with $10^4 \times 10^4 = 10^8$ elements. A 3D grid of size $100 \times 100 \times 100$ will have a matrix $\mathbf{A}$ with $10^6 \times 10^6 = 10^{12}$ elements. To address this issue, *DAMAS2* and *DAMAS3* were developed by Dougherty (2005b). These are faster versions of the original *DAMAS* algorithm, which use iterative Fourier methods to solve for the source distribution, and assume a shift-invariant point spread function. To extend the region over which the shift invariant assumption holds, a

change of coordinates was introduced.

2.4.2 *CLEAN-SC*

The $CLEAN-PSF$ algorithm uses the beamformed map as a 'dirty' map. It iteratively removes peaks from this dirty map by subtracting scaled point spread functions and inserts clean beams at the peak locations in a 'clean' map. A problem with this method is that it uses theoretical point spread functions which may be inaccurate. To overcome this problem, the *CLEAN-SC* algorithm was developed by Sijtsma (2007b). *CLEAN-SC* is performed in the frequency domain. A time domain version, referred to as *TIDY* was developed by Dougherty, see (Dougherty and Podboy, 2009; Dougherty, 2009).

For iteration number l, the highest peak value $\boldsymbol{b}_{max}^{\{l-1\}}$ and the coordinate $\vec{\boldsymbol{\xi}}_{max}^{\{l\}}$ of that peak are obtained. Rather than using theoretical point spread functions to remove this peak, spatial coherence is used. Spatial coherence requires that for all $\mathbf{w}$

$$\mathbf{w}^{\dagger}\,\mathbf{G}^{\{l-1\}}\,\mathbf{w}_{max}^{\{l\}} = \mathbf{w}^{\dagger}\,\mathcal{G}^{\{l\}}\,\mathbf{w}_{max}^{\{l\}} \tag{2.36}$$

where $\mathcal{G}^{\{l\}}$ is the CSM that would result from a sound source located at $\vec{\boldsymbol{\xi}}_{max}^{\{l\}}$ and $\mathbf{w}_{max}^{\{l\}}$ is the steering vector corresponding to $\vec{\boldsymbol{\xi}}_{max}^{\{l\}}$. This simplifies to

$$\mathbf{G}^{\{l-1\}}\,\mathbf{w}_{max}^{\{l\}} = \mathcal{G}^{\{l\}}\,\mathbf{w}_{max}^{\{l\}}. \tag{2.37}$$

This does not have a unique solution. However, if one assumes a single coherent source component, one can find a solution

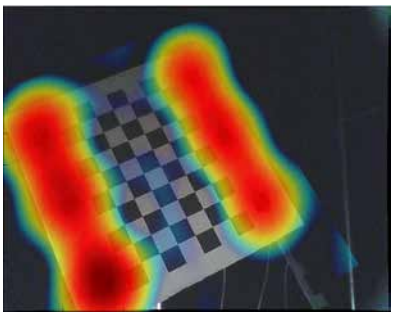
$$\mathcal{G}^{\{l\}} = \boldsymbol{b}_{max}^{\{l-1\}}\,\left(\boldsymbol{h}^{\{l\}}\,\boldsymbol{h}^{\dagger\{l\}} - \mathbf{H}^{\{l\}}\right) \tag{2.38}$$

where $\boldsymbol{h}^{\{l\}}$ is the coherent source component and

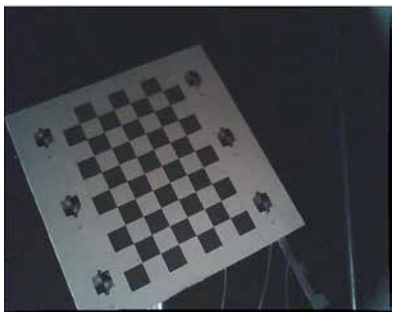
$$\mathbf{H}_{m,m'}^{\{l\}} = \begin{cases} 0, & m \neq m' \\ \boldsymbol{h}_{m}^{\{l\}}\,\boldsymbol{h}_{m'}^{\dagger\{l\}}, & m = m' \end{cases} \tag{2.39}$$

if CSM diagonal removal (DR) is used, see Section 2.2.2, or otherwise $\mathbf{H} = \mathbf{0}_{(M,M)}$, a $M \times M$ matrix of zeros. A summary of the *CLEAN-SC* algorithm is outlined below.

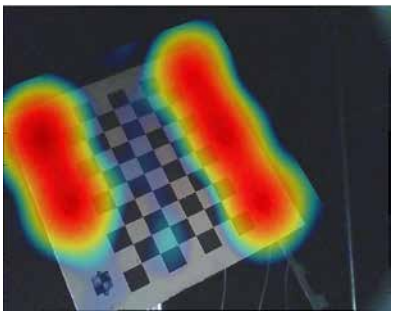
CLEAN-SC is initialised with iteration $l = 0$ by setting $\mathbf{G}^{\{0\}} = \mathbf{G}$ to be the cross spectral matrix obtained from the raw data using Equation (2.12) and setting $\boldsymbol{b}^{\{0\}} = \boldsymbol{b}$ the beamformed map obtained using Equation (2.11). For each iteration, a degraded map



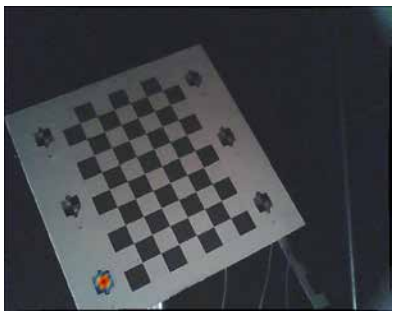
(a) Beamformed map used as the dirty map for *CLEAN-SC* iteration 0.



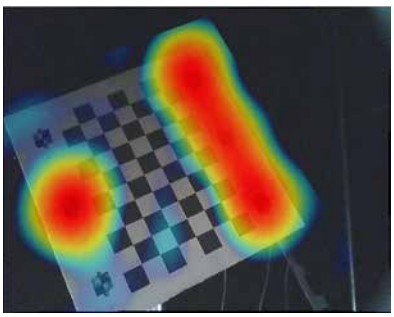
(b) Clean map, *CLEAN-SC* iteration 0. ($\mathcal{Q}^{\{0\}}$.)



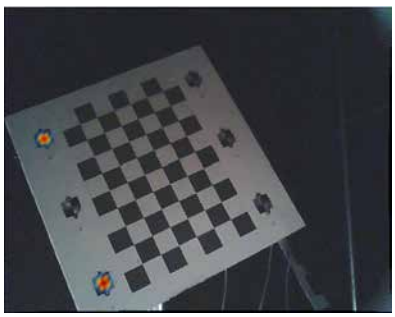
(c) Dirty map, *CLEAN-SC* iteration 1. (Primary sound source peak with sidelobes removed.)



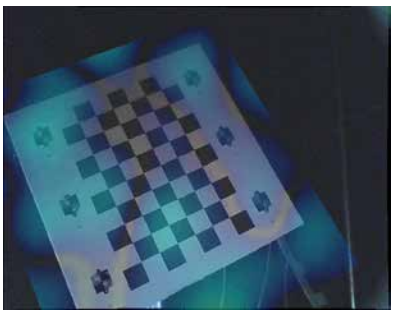
(d) Clean map, *CLEAN-SC* iteration 1. (Primary sound source $\mathcal{Q}^{\{1\}}$).)



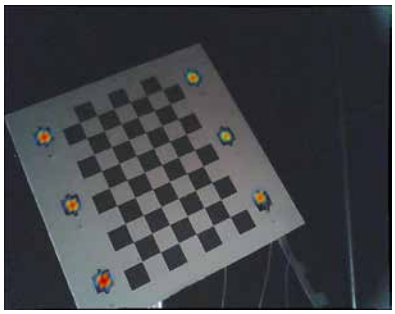
(e) Dirty map, *CLEAN-SC* iteration 2. (Secondary sound source peak with sidelobes removed.)



(f) Clean map, *CLEAN-SC* iteration 2. (Primary and secondary sources $\sum_{l=1}^{2} \mathcal{Q}_n^{\{l\}}$.)



(g) Dirty map, *CLEAN-SC* iteration 14. (Theoretically just noise remaining.)



(h) Clean map, *CLEAN-SC* iteration 14. (Final deconvolution map $\sum_{l=1}^{14} \mathcal{Q}_n^{\{l\}}$.)

Figure 2.10: Example plots of *CLEAN-SC* deconvolution iterations using a loop gain φ of 95%. A spherical array was used to image sound from six speakers set into holes in a piece of perspex, See Section 7.2 for more details.

is obtained for each scan point using

$$\boldsymbol{b}_n^{\{l-1\}} = \mathbf{w}(\vec{\boldsymbol{\xi}}_n)^\dagger \, \mathbf{G}^{\{l-1\}} \, \mathbf{w}(\vec{\boldsymbol{\xi}}_n). \tag{2.40}$$

The highest peak in the degraded beamformed map is then identified and its peak value $\boldsymbol{b}_{max}^{\{l-1\}}$ and corresponding scan point coordinate $\vec{\boldsymbol{\xi}}_{max}^{\{l\}}$ obtained. A clean map $\boldsymbol{\mathcal{Q}}^{\{l\}}$, having the same peak value $\boldsymbol{b}_{max}^{\{l-1\}}$ and location $\vec{\boldsymbol{\xi}}_{max}^{\{l\}}$, is obtained using

$$\boldsymbol{\mathcal{Q}}_n^{\{l\}} = \boldsymbol{\varphi} \, \boldsymbol{b}_{max}^{\{l-1\}} \, \boldsymbol{\Phi}(\vec{\boldsymbol{\xi}}_n - \vec{\boldsymbol{\xi}}_{max}^{\{l\}}), \tag{2.41}$$

where $\boldsymbol{\varphi}$ is a loop gain constant (≤ 1) and $\boldsymbol{\Phi}$ is a normalised clean beam. The normalised clean beam used in the thesis work has the form

$$\boldsymbol{\Phi}(\vec{\boldsymbol{\xi}}_n - \vec{\boldsymbol{\xi}}_{max}^{\{l\}}) = e^{-\text{const} \, \|\vec{\boldsymbol{\xi}}_n - \vec{\boldsymbol{\xi}}_{max}^{\{l\}}\|^2 / \|\vec{\boldsymbol{\xi}}_n\|}.$$

The contribution of the peak and its sidelobes may be removed from the degraded CSM using

$$\mathbf{G}^{\{l\}} = \mathbf{G}^{\{l-1\}} - \boldsymbol{\varphi} \, \boldsymbol{b}_{max}^{\{l-1\}} \left(\boldsymbol{h}^{\{l\}} \, \boldsymbol{h}^{\dagger\{l\}} - \mathbf{H}^{\{l\}} \right) \tag{2.42}$$

where

$$\boldsymbol{h}^{\{l\}} = \frac{1}{\left(1 + \mathbf{w}_{max}^{\dagger\{l\}} \, H^{\{l\}} \, \mathbf{w}_{max}^{\{l\}}\right)} \left[\frac{\mathbf{G}^{\{l-1\}} \, \mathbf{w}_{max}^{\{l\}}}{\boldsymbol{b}_{max}^{\{l-1\}}} + H^{\{l\}} \, \mathbf{w}_{max}^{\{l\}} \right]. \tag{2.43}$$

This is not an explicit expression for $\boldsymbol{h}^{\{l\}}$, since $\mathbf{H}^{\{l\}}$ is a function of $\boldsymbol{h}^{\{l\}}$. Therefore, $\boldsymbol{h}^{\{l\}}$ is initially set equal to the array propagation vector $\boldsymbol{C}(\vec{\boldsymbol{\xi}}_{max}^{\{l\}})$ and Equation (2.43) is iteratively repeated until convergence is reached. If diagonal removal is not used, however, $\mathbf{H}^{\{l\}}$ may be set to a matrix of zeros of size $(M \times M)$. In this case, Equation (2.43) simplifies to

$$\boldsymbol{h}^{\{l\}} = \frac{\mathbf{G}^{\{l-1\}} \, \mathbf{w}_{max}^{\{l\}}}{\boldsymbol{b}_{max}^{\{l-1\}}}. \tag{2.44}$$

Note that the primary sound source is contained in the clean map $\boldsymbol{\mathcal{Q}}_n^{\{1\}}$, second main sound source in $\boldsymbol{\mathcal{Q}}_n^{\{2\}}$, etc. The *CLEAN-SC* map for the n^{th} scan point due to all sources may be obtained by summing these

$$\boldsymbol{\mathcal{Q}}_n = \sum_{l=1}^{L} \boldsymbol{\mathcal{Q}}_n^{\{l\}}, \tag{2.45}$$

where there were L iterations used, see Figure 2.10. A stop criteria for iterations may be chosen as

$$\|\mathbf{G}^{\{l\}}\| > \|\mathbf{G}^{\{l-1\}}\|. \tag{2.46}$$

It is assumed, when this condition is reached, that all likely sound sources have been

found and that only noise remains in the degraded CSM.

.

3

Hardware

3.1 Phased Array Instrumentation

3.1.1 Array Hardware

Microphones, Preamplifiers, and Cabling

The array hardware contains 73 Panasonic *WM-61A* electret condenser microphones. They have a relatively flat frequency response, having a 5dB amplitude variation from 20Hz to 30kHz (Holm Acoustics, 2009), have a sensitivity of about 35 ± 4dB (0dB = 1V/Pa, 1kHz), and self-noise of about 32dB(A)(Panasonic). They are low cost and are a commonly used microphone in microphone phased arrays (Booth and Humphreys, 2005; Yardibi et al., 2009; Oerlemans, 2009).

(a) Front. (b) Side. (c) Back.

Figure 3.1: Photos showing microphone and pre-amp circuit board.

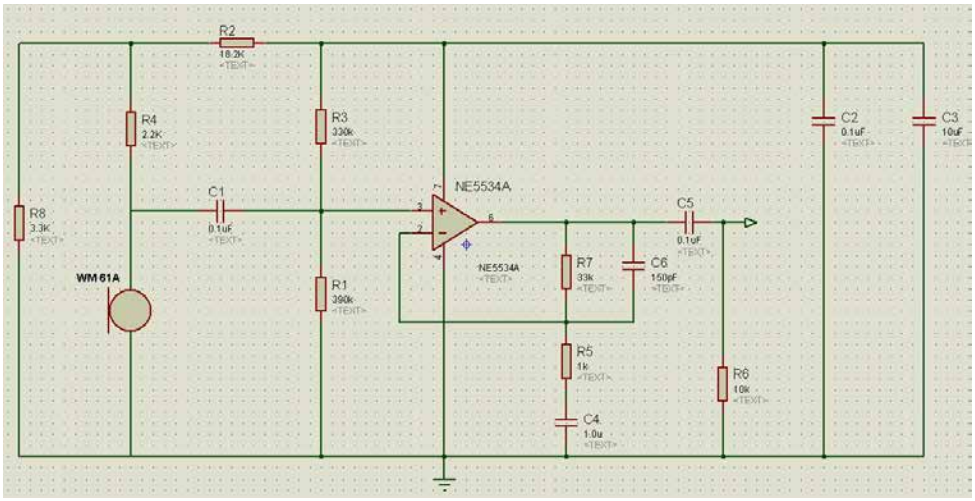
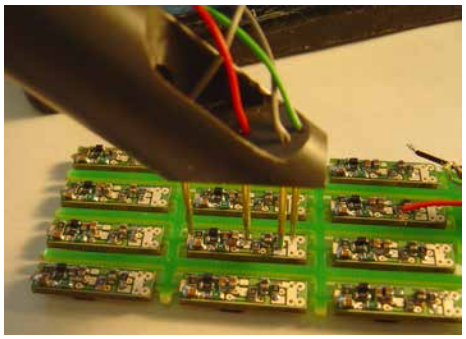
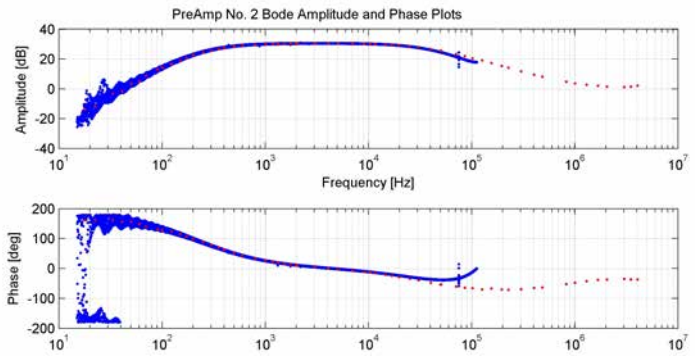


Figure 3.2: Pre-amp circuit diagram.

To minimise electrical noise, each microphone is soldered directly to a preamplifier circuit board. The profile of this circuit board was designed to be as close as possible to the diameter of the microphone to reduce scattering effects when used in a 3D lattice type array, see Figure 3.10b. A double sided circuit board was designed and built using hand soldered 0805 surface mount components, see Figure 3.1. A circuit diagram of the preamplifier is provided in Figure 3.2. The preamplifier, which has a gain of 30dB and includes some high and low pass filtering, see Figure 3.3, uses the low noise op-amp, NE5534A. The circuit boards are soldered to 2m long, 4mm outer-diameter cables, see Figure 3.4.



(a) Testing rig used to check amplitude and phase response.

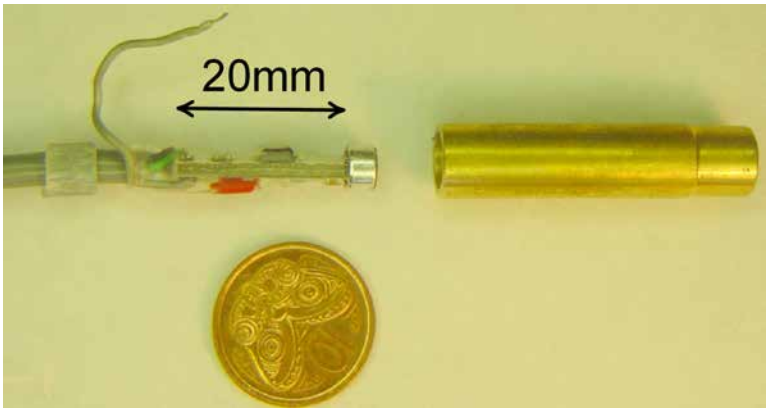


(b) Pre-amp Bode amplitude and phase plots. Blue dots are automatically generated using white noise.

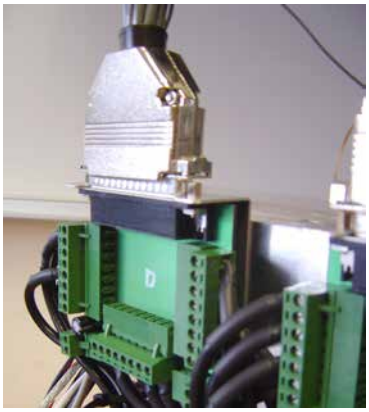
Figure 3.3: Pre-amp circuit boards testing.

These cables are composed of independently shielded, twin twisted pairs. One pair is used for the output signal and the other supplies 24V to the pre-amp.

The microphone and pre-amp circuits are inserted into microphone cases machined from brass rod as shown in Figure 3.4a. These provide both electrical shielding and physical protection for the pre-amp circuits. Heat shrink over the circuit prevents shorting. A



(a) Microphone pre-amp with wiring and case.



(b) Breakout circuit board.

(c) Microphone cables.

Figure 3.4: Cabling.

groove beside the microphone ensures equal air pressure on the front and back of the microphone. Two half rings, laser cut from perspex plastic, are pushed from the back of the microphone case to form a wedge between the cable and the inside surface of the case. While providing a firm fit, this ensures that the microphone pre-amp circuit can be accessed if desired.

To decrease the setup time of the array, the microphone cables are divided into six groups of twelve. The cables in each groups are soldered to a single 37-way D-Sub connector plug. These plug into a breakout circuit board which provides 12 BNC cables for connection to ADC boards and 24V from two batteries, see Figure 3.4b. A diode was included to provide voltage protection for the microphone pre-amps. Two stands were also made for mounting the breakout cables and ADC boards. Apart from the microphone cases, which were built by the *Faculty of Science Workshop*, the design and build described above was performed by the author.

ADC/DAC Hardware

The analog signals from the microphone pre-amps are converted to digital data using three Data Translation DT9836-12-2-BNC boards and six Data Translation DT9816-S boards, see Figure 3.5. Table 3.1 outlines the specifications of these two Data Translation boards. These provided a total of 72 simultaneous sampling analog input channels,

Table 3.1: Features of two Data Translation data acquisition boards.

	DT9836-12-2-BNC	DT9816-S
Sampling Type	Simultaneous	Simultaneous
Maximum Sampling Rate	225 kSPS per channel	750 kSPS per channel
Resolution	16-bit	16-bit
Analog Input Channels	12 (single ended)	6 (single ended)
Analog Output Channels	2	0
Digital I/O Lines	32	16
Counter/Timers	2	1
Quadrature Encoders	3	0
Input Voltage Range	$\pm 10V$	$\pm 10V$
Triggering	External or internal	External or internal
Clock	External or internal	External or internal
Powered	Mains	USB

six analog outputs, and 192 digital I/O lines. Synchronisation is achieved using a clock signal, generated by one of the DT9836 boards, as an external input clock and trigger for all boards. These boards are connected through two USB hubs to a laptop and are

(a) DT9836-12-2-BNC (b) DT9816-S

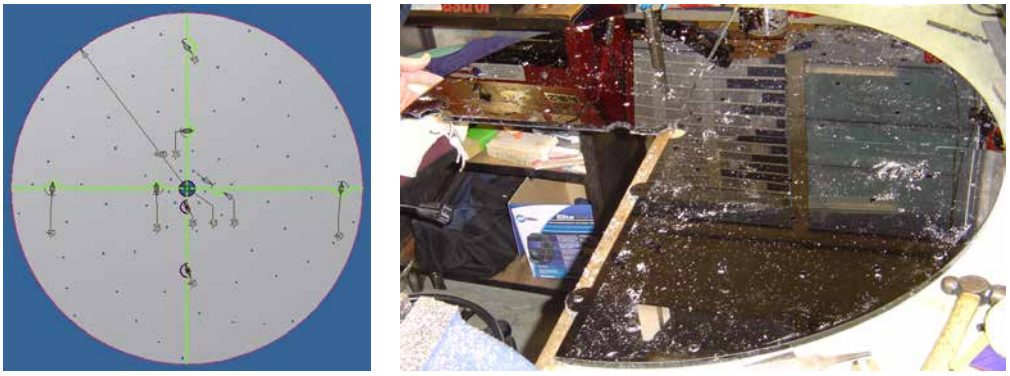
Figure 3.5: Data Translation data acquisition boards.

controlled using *MATLAB* code which uses the '*MATLAB* Data Acquisition Toolbox'.

Sound may be played on speakers by generating signals from the DT9836's six analog outputs. In addition to this, four circuits boards containing Atmel ATTINY13a microcontroller chips could be used to generate white noise. These white noise generators were

made by a colleague, Travis Wiens, in exchange for microphone/op-amp circuits. They output samples at 48 kSPS and are designed to be as uncorrelated as possible, using different sequences in a Kasami set (Wiens, 2009a).

3.1.2 Array Structures

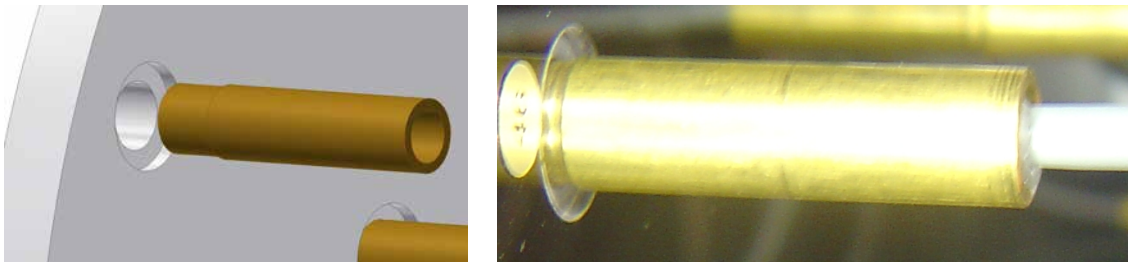


(a) Cad file used for the laser cutter.

(b) Laser cut perspex panels being prepared for gluing together in order to form a single spiral array panel.

Figure 3.6: Spiral array panel build.

The array hardware has been designed to be versatile so that a range of 2D or 3D array geometries or multiple arrays could be investigated. Planar array panels are constructed by making a CAD model of the array panel and using this to laser cut sections from 10mm thick perspex plastic, see Figure 3.6. These sections are then glued together and mounted onto a stand. The microphones are flush mounted by inserting the microphone cases into holes laser cut in the panels, refer to Figure 3.7. This technique is based on the method described by Underbrink (2002).

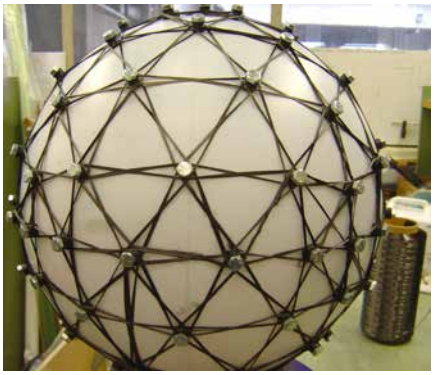


(a) CAD model illustration.

(b) Photo.

Figure 3.7: Planar array with a microphone case inserted through a hole laser cut into the perspex panel. A counter sunk hole ensures that the microphone is flush with the front surface of panel.

The 3D array was made by wrapping carbon fibre tow/string around grommet-like microphone holders which are bolted onto a mold and then set using epoxy resin, see Figure



(a) Microphone holders bolted onto a mould and carbon fiber tow wrapped between in readiness for setting with epoxy resin.

(b) Microphone case inserted through holder in 3D carbon fiber lattice structure.

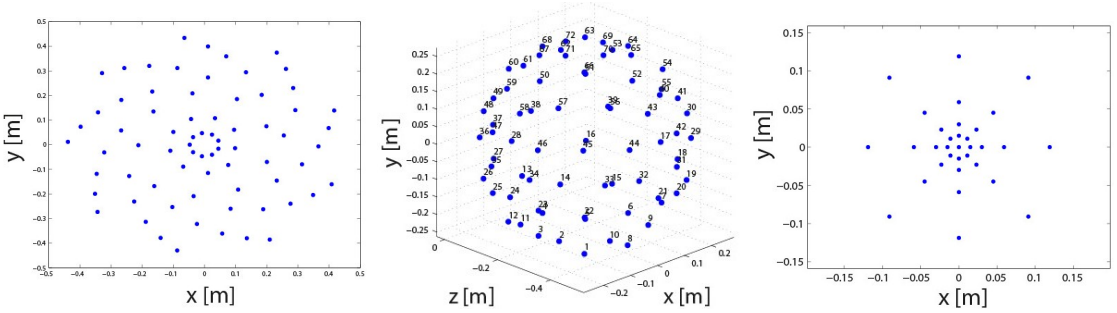
Figure 3.8: Lattice array construction.

3.8. The grommet-like microphone holders are made by laser cutting washers from perspex plastic and gluing them together on the same rod from which the microphone cases where made. This results in a lattice-like structure which is designed to be as acoustically transparent as possible. The microphones are mounted by inserting the microphone cases through the grommet-like holders. The same microphones/microphone cases are used in both the 3D and planar arrays.

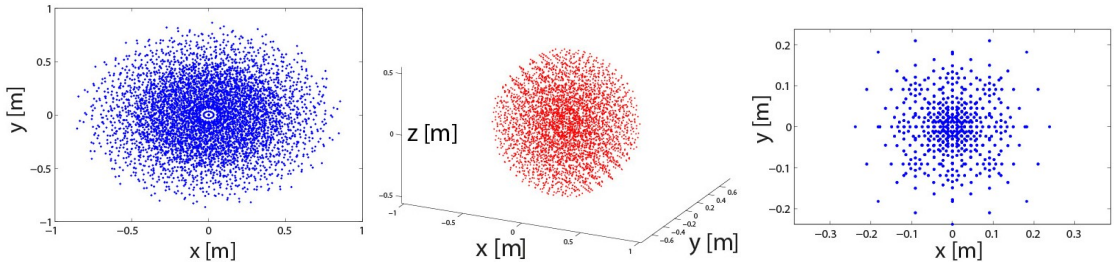
Array Design

Different array designs were investigated. Rather than trying to create totally new designs, the use of well established array geometries was preferred. Certain requirements were decided on, such as the maximum diameter of the array, the frequency range of interest, and the required resolution and dynamic range. The coordinates of microphones in a proposed array were then usually obtained using an iterative-like process. The dynamic range and resolution were obtained from far-field and near-field point spread functions for different microphone positions and frequencies until optimum coordinates were chosen.

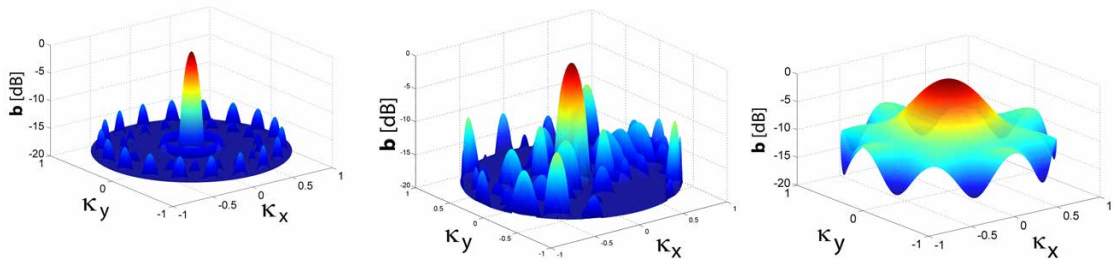
Three arrays were eventually decided on: an Underbrink multi-arm spiral, a spherical array, and a SADA, see Figure 3.10. The Underbrink spiral provides a well established 2D array type with a good lateral dynamic range, about $11-14$dB, over a wide frequency range and good longitudinal resolution for its size. Since it is a planar array, however, ignoring any shadowing effects due to its array structure, it cannot distinguish if sound sources are located in front or behind the array. The spherical array provides imaging in all directions, a typical lateral dynamic range of about 10dB, and low longitudinal resolution. This provides a comparison to the spiral array's higher longitudinal resolution. The SADA



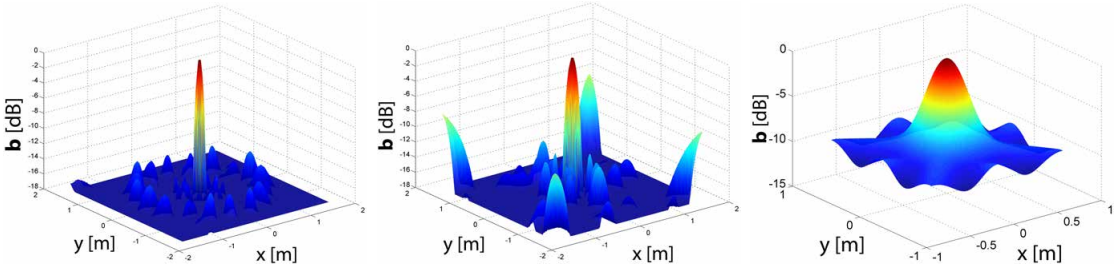
(a) Spiral array microphone coordinates. (b) Spherical microphone coordinates. (c) SADA microphone coordinates.



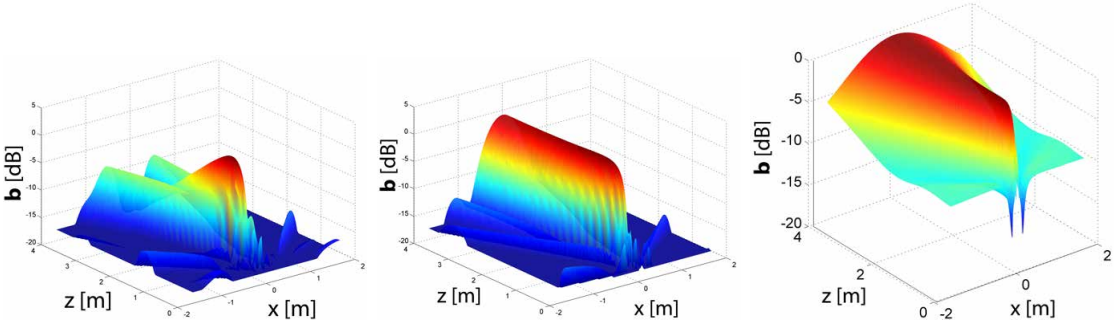
(d) Coarray. (e) Coarray. (f) Coarray.



(g) Far-field point spread function. (h) Far-field point spread function. (i) Far-field point spread function.



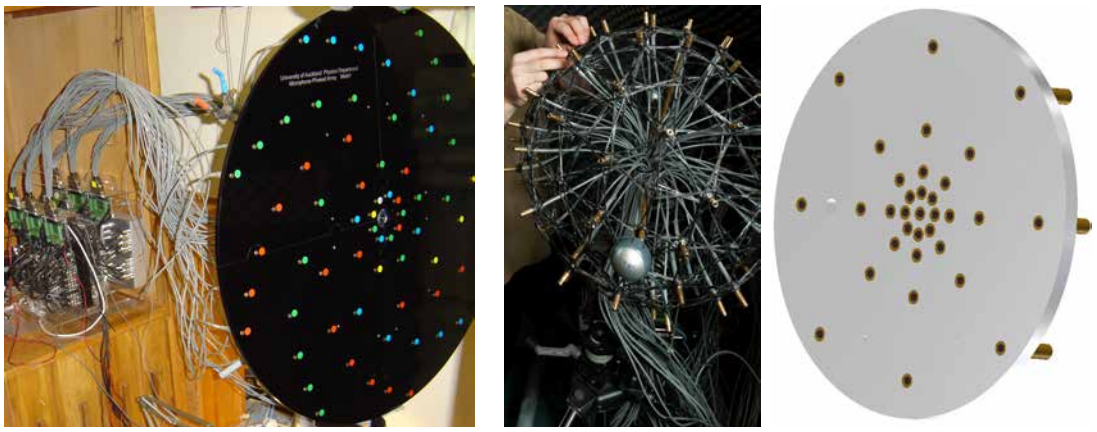
(j) Near-field lateral point spread function. (k) Near-field lateral point spread function. (l) Near-field lateral point spread function.



(m) Longitudinal point spread function. (n) Longitudinal point spread function. (o) Longitudinal point spread function.

Figure 3.9: Comparison of microphone array characteristics for 3.5kHz, where each column relates to a different array. See Appendix 10 for larger images. (Note that spherical spreading correction has been omitted.)

was also included due to its smaller diameter, compared to the other and arrays, and its poor longitudinal resolution. See Figure 3.9 and Appendix 10 for microphone coordinates and point spread functions.



(a) Underbrink multi-arm spiral array. (b) Spherical array. (c) SADA (scaled) array.

Figure 3.10: Photos of built arrays.

The Underbrink multi-arm spiral microphone array was designed using the equations and design process described by Underbrink (2002) and (Dougherty, 1998). Nine logarithmic spiral arms were equally spaced about a circle. The position of the microphones along the spirals were obtained using seven equal aperture-area intervals with an additional ring added near the centre of the array. The array panel and stand were built by the *Faculty of Science Workshop*.

A spherical array was designed and built using a geodesic structure based on vertices of a fifth frequency, second class, octahedron. (Frequency describes the number of segments into which a side of a triangle segment of the octahedron is subdivided in order to generate the vertices. Class describes the manner of the subdivision (Kenner, 1976).) The coordinates were generated using a trial version of '*CADRE Geo 6*'[1]. These vertices where marked onto a spherical mold (500mm diameter lamp shade) using a *Faro-Arm*[2], robotic arms measurement tool. Microphone holders were bolted through holes drilled at these positions and carbon fibre wrapped between and set with epoxy. Three sections were merged to form a full sphere. See Figure 10.4 for photos of the construction process. This array was almost entirely built by the author with a few components being built by the *Faculty of Science Workshop*.

The other array constructed was a Small Aperture Directional Array (*SADA*). This has microphone coordinates scaled by 130% from that described by (Humphreys et al., 1998).

[1] `http://www.cadreanalytic.com/cadregeo.htm`
[2] `www.faro.com/gage/`

This scaling was required because the original microphone spacing, near the centre of the array, was too close for the microphone cases (9.6mm diameter). The stand was made by the *Faculty of Science Workshop* but the panel was made by the author.

3.1.3 Cameras and Projector

The cameras used in the arrays were three inexpensive Logitech C500 webcams (two of which were able to be attached to the side of all arrays) and a 'no brand' mini webcam. These have maximum resolutions of 1280x1024 and 640x480 respectively. The main data

Table 3.2: Web cameras used in arrays.

Array	Logitech C500	Mini Web Camera
Underbrink Multi-arm Spiral	1	0
Spherical	2	1
SADA	0	1

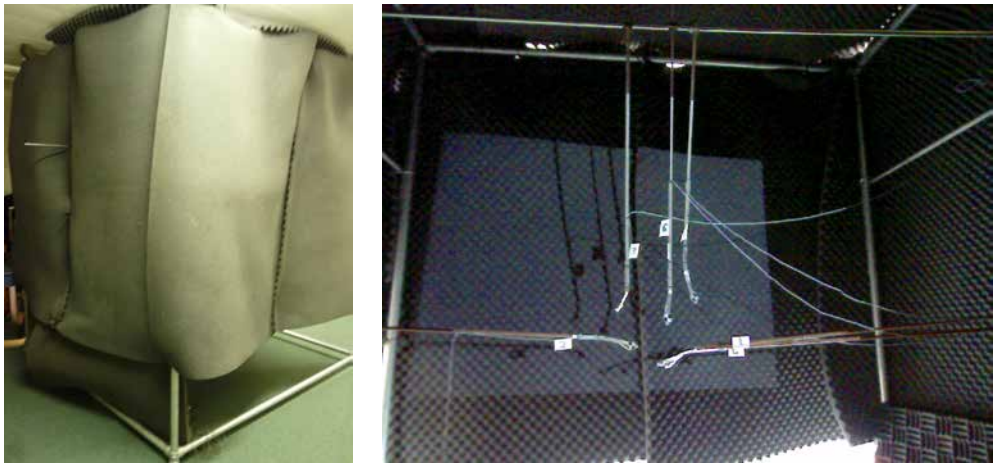
projector used is a Dell 2300MP. This was mounted onto a separate stand.

3.1.4 Acoustic Foam Enclosure

The main acoustic measurements were made in a semi-anechoic enclosure (2.2×2.2×2.4m), see Figure 3.11. This was constructed using acoustic foam attached to framework made using galvanised pipe and key clamp fittings. The acoustic foam did not cover the lower part of two of the four walls and also did not cover the floor. However, the floor was covered in carpet with a thick underlay and, therefore, was not highly reflective. The acoustic foam combined with the room being relatively large appeared to stop the main refections. Traffic noise was present. This may have been a problem for low frequency noise. The framework was also designed to allow speakers to be mounted against a 3D sound source object without them having to be physically attached to the object.

3.2 3D Models

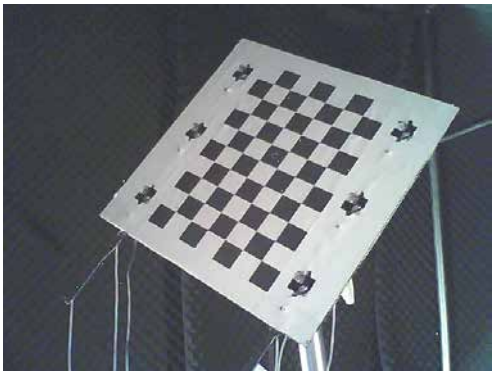
The 3D surfaces for acoustic imaging were also constructed. These include a plane, a small cardboard box (approx 220 × 260 × 290mm), a larger plywood wooden box (1.2 × 1.2 × 1.225m), and a 1.2m diameter plastic sphere, see Figure 3.12. The planar object, which consisted of a perspex (lucite) sheet with speakers inserted into its surface, was



(a) Exterior. (b) Interior with speakers mounted from framework.

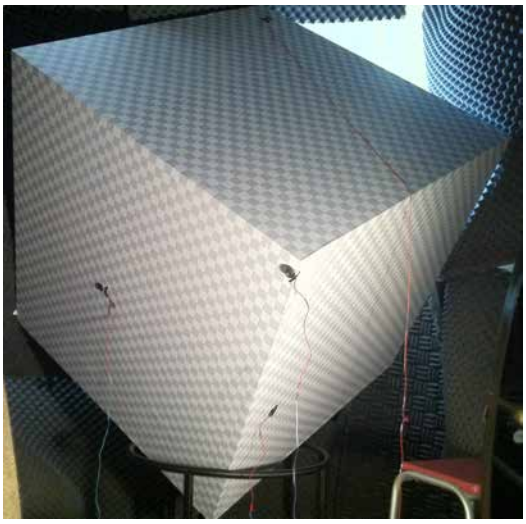
Figure 3.11: Acoustic foam enclosure.

laser cut from perspex plastic by the author. The large cube was constructed by the *Faculty of Science Workshop* from plywood. They also constructed the large sphere by modifying two hemispheres purchased from a playground equipment manufacturer. Acoustic imaging was intended to be performed both inside and outside the large cube and sphere. A small hole/slot was, therefore, cut into the base of these objects to enable the microphone arrays' stands to be passed through and set up independent of the 3D object. Legs were constructed to ensure that the objects were at an appropriate height relative to the arrays. The objects (with the exception of the sphere) had a checkerboard pattern glued onto their surface. This was done to enable the position and orientation of the object to be determined relative to the array, see Section 5.2.1. Also, by plotting a deconvolution acoustic map over a camera image of the object, the checkerboard pattern was intended to enable the position error of a deconvolution peak relative to a speaker to be visually identified.



(a) Plane with six speakers inserted into its surface.

(b) Small cardboard box.



(c) Large cube.

(d) Two halves of a large sphere.

Figure 3.12: Sound source objects.

4

Camera 3D Vision Theory

This chapter briefly outlines the computer vision theory used in this work. For more detailed information, refer to works by Hartley and Zisserman (2004), Ma et al. (2003), Faugeras (1993), Szeliski (2010), Tsai (1987), Zhang (1998), Bradski and Kaehler (2008), and Bouguet (1999). An extensive computer vision bibliography is also provided by Price[1].

4.1 Camera Model

The camera reference frame $\mathcal{F}_c = (O, X, Y, Z)$ is defined such that the X and Y axes are parallel to the image plane and the Z axis lies along the optic axis. These axes are right handed and mutually orthogonal, see Figures 4.2 and 1.3b.

The convention for describing a cartesian coordinate vector in microphone phased array literature is to use a row vector of the form

$$\vec{\boldsymbol{X}} = [X, Y, Z] \tag{4.1}$$

[1] http://www.visionbib.com/bibliography/contents.html

in units of metres. However, the convention in computer vision theory is to use a column vector of the form

$$\vec{\boldsymbol{X}} = [X, Y, Z]^{\intercal} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{4.2}$$

in units of millimetres. Rather then changing the equations from that presented in the literature, the convention used in this thesis is that microphone array and computer vision equations will use position vectors in the traditional forms given respectively in Equation (4.1) and Equation (4.2). Another conflict in convention between the two sets of literature is in the notation for a 3D Euclidean coordinate vector. In acoustic imaging literature, it is usual to use lower case letters, such as $\vec{\boldsymbol{x}}$. In computer vision literature, however, it is usually to use capital letters, such as $\vec{\boldsymbol{X}}$, for 3D coordinate vectors and lower case letters for 2D coordinates/projection vectors. The computer vision convention has be followed where possible to avoid confusion.

4.1.1 Pinhole Projection Model

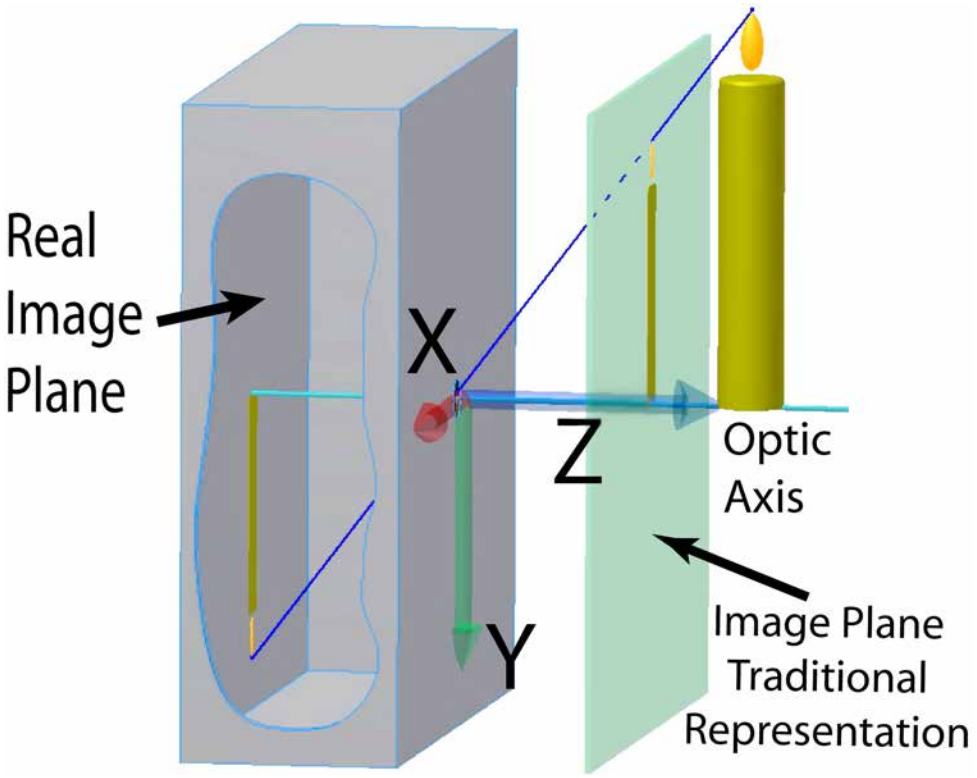


Figure 4.1: Diagram illustrating a pinhole camera.

Figure 4.1 shows a diagram of a pinhole camera which is the basic model for a camera. This is composed of two screens which are separated a distance f_o apart. The first screen

has a hole, referred to as the centre of projection, through which light shines. An inverted image is formed on the second screen which is referred to as the image plane. The convention, however, when drawing this model, is to place the image plane on the side of the hole from which the light shines a distance f_o in front of the hole. This gives a non-inverted image where the light rays pass through this virtual image plane to the hole, see Figures 4.1 and 4.2 (Lanman and Taubin, 2009b).

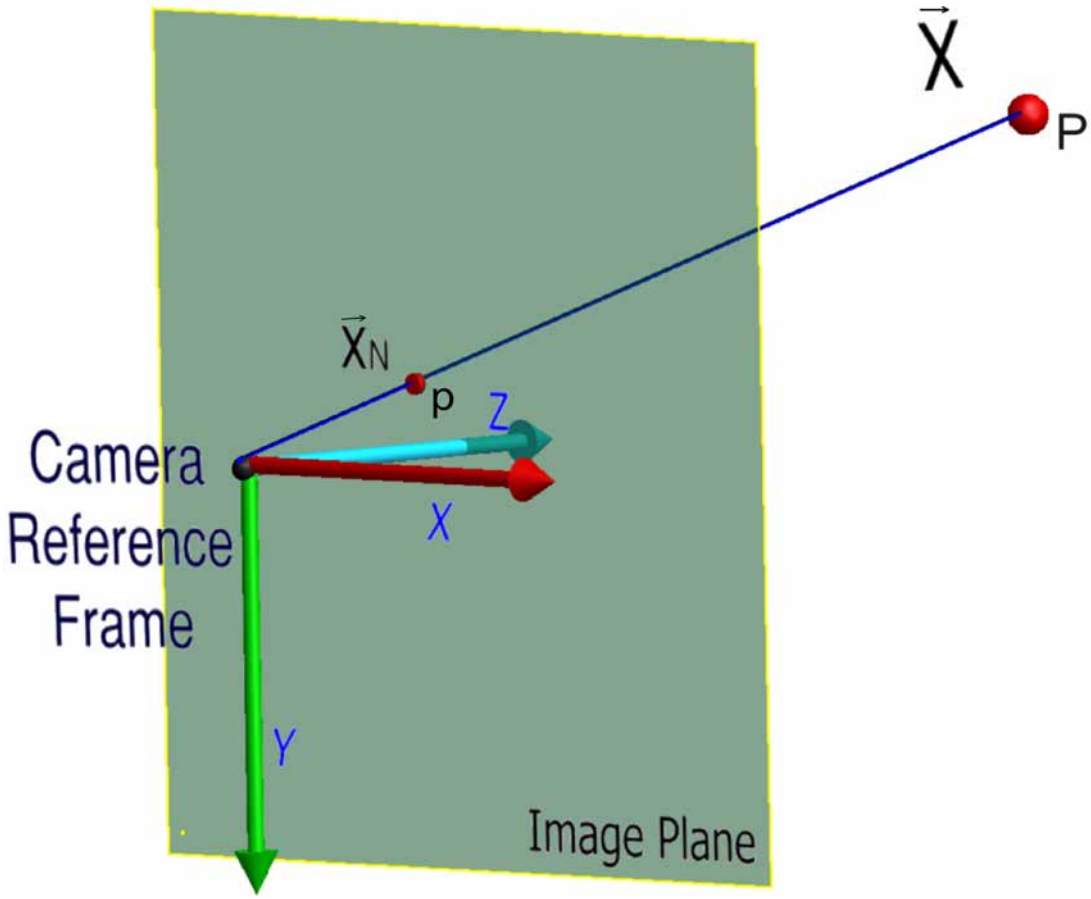


Figure 4.2: Diagram showing the standard pinhole camera model where a point is projected onto the image plane.

The pinhole camera uses perspective projection to project light from a 3D object onto the 2D image plane. Consider a point P with coordinates $\vec{\boldsymbol{X}} = [X, Y, Z]^\intercal$ in 3D space in the camera reference frame, where the pinhole is the origin and Z lies along the optic axis. Light from P forms an image on the 2D scanning surface with 3D coordinates

$$\vec{\boldsymbol{X}}_{(\text{Image Plane})} = \begin{bmatrix} X/Z \\ Y/Z \\ 1 \end{bmatrix} f_o. \tag{4.3}$$

Note that negative signs have been omitted from Equation (4.3), since the convention

of a virtual image plane in front of the camera has been used. Rather than using 3D coordinates, each point on the 2D image plane is associated with a normalised coordinate

$$\vec{\boldsymbol{x}}_N = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X/Z \\ Y/Z \end{bmatrix}. \tag{4.4}$$

Depth information, Z is lost in the projection of a 3D point onto the 2D image plane. To recover the 3D coordinate of P,

$$\vec{\boldsymbol{X}} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} Z = \bar{\mathbf{x}}\, Z \tag{4.5}$$

where $\bar{\mathbf{x}} = [x, y, 1]^\intercal$ is referred to as a homogeneous coordinate. However, any scaled version of $\bar{\mathbf{x}}$ represents the same point p on the image plane and lies on a line which passes from O through both p and P. The homogeneous vector $\bar{\mathbf{x}}$, therefore, describes an optical ray (Bouguet, 1999).

4.1.2 Camera Distortion Model

Instead of a pinhole, real cameras use lenses which cause distortion of the projected image. There are many distortion models in the literature, see Ricolfe-Viala and Sanchez-Salmeron (2010) for example. The distortion model used here was first developed by Brown (1966) and is the one used by the camera calibration software '*MATLAB* Camera Calibration Tolbox' and 'openCV'. This distortion may be modelled as being due to the sum of a radial and tangential distortion, see Figure 4.3. Distorted normalised coordinates may be calculated, using a sixth order distortion model, as

$$\vec{\boldsymbol{x}_{\boldsymbol{d}}} = \left(1 + \boldsymbol{\zeta}_1\, R^2 + \boldsymbol{\zeta}_2\, R^4 + \boldsymbol{\zeta}_5\, R^6\right) \vec{\boldsymbol{x}}_N + \delta_x \tag{4.6}$$

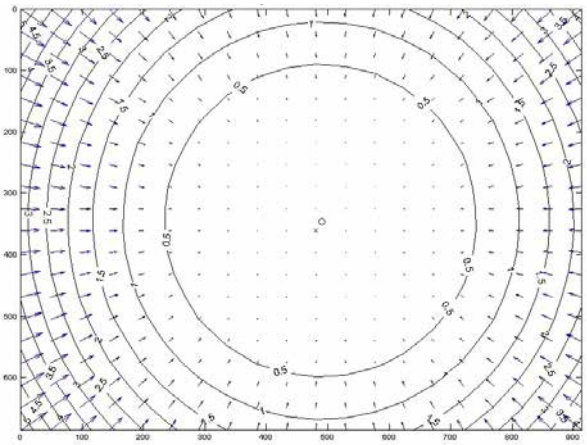
where

$$\vec{\delta}_x = \begin{bmatrix} 2\,\boldsymbol{\zeta}_3\, x\, y + \boldsymbol{\zeta}_4\,(R^2 + 2\,x^2) \\ \boldsymbol{\zeta}_3\,(R^2 + 2y^2) + 2\,\boldsymbol{\zeta}_4\, x\, y \end{bmatrix} \tag{4.7}$$

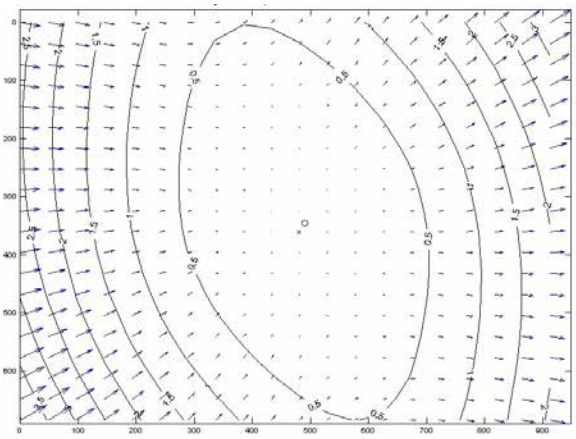
is the tangential distortion, $R = \|\vec{\boldsymbol{x}}_N\| = \sqrt{x^2 + y^2}$, and $\boldsymbol{\zeta}$ is a (5×1) array of distortion coefficients. For most cameras, first order distortion

$$\vec{\boldsymbol{x}_{\boldsymbol{d}}} = \left(1 + \boldsymbol{\zeta}_1\, R^2\right) \vec{\boldsymbol{x}}_N, \tag{4.8}$$

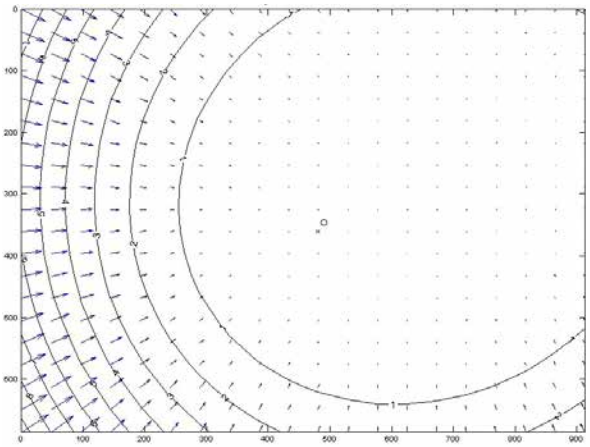
is usually sufficient.



(a) Radial distortion.



(b) Tangential distortion.



(c) Complete distortion.

Figure 4.3: Example of camera distortion.

4.1.3 Conversion from Normalised Projection Coordinates to Pixel Coordinates

In modern cameras, the image is projected onto optical sensors and converted into discrete pixels. The top left pixel is taken to be the origin of the pixel coordinate system and $\boldsymbol{c}_c$ are the coordinates of the centre point. The distorted coordinates $\vec{\boldsymbol{x_d}}$ may be converted to pixel coordinates by

$$\vec{\boldsymbol{p}} = \begin{bmatrix} \boldsymbol{f}_{c1}\ (\vec{\boldsymbol{x_{d1}}} + \alpha_c\ \vec{\boldsymbol{x_{d2}}}) + \boldsymbol{c}_{c1} \\ \boldsymbol{f}_{c2}\ \vec{\boldsymbol{x_{d2}}} + \boldsymbol{c}_{c2} \end{bmatrix} \tag{4.9}$$

where α_c is the angular skew between the x and y axes and $\boldsymbol{f}_c$ is the focal length in pixels. If dx and dy are the dimensions of each pixel sensor in mm, $\boldsymbol{f}_{c1} = f_o/dx$ and $\boldsymbol{f}_{c2} = f_o/dy$, where f_o is the focal length in mm. The angle α_c is usually zero for modern cameras. The variables $\boldsymbol{f}_c$, $\boldsymbol{c}_c$, α_c, and $\boldsymbol{\zeta}$ are referred to as the intrinsic parameters of a camera and are

obtained using camera calibration, see Section 4.2.

Equation (4.9) may also be written as

$$\bar{\mathsf{p}} \simeq \boldsymbol{\mathcal{K}}\, \bar{\mathsf{x}}_{\mathsf{d}} \tag{4.10}$$

where $\bar{\mathsf{p}} = [\vec{\boldsymbol{p}}_1, \vec{\boldsymbol{p}}_2, 1]^\intercal$ are the homogeneous pixel coordinates, $\bar{\mathsf{x}}_{\mathsf{d}} = [\vec{\boldsymbol{x_d}}_{\mathbf{1}}, \vec{\boldsymbol{x_d}}_{\mathbf{2}}, 1]^\intercal$ are the homogeneous distorted normalised coordinates in the image plane, and

$$\boldsymbol{\mathcal{K}} = \begin{bmatrix} \boldsymbol{f}_{c\,1} & \alpha_c\, \boldsymbol{f}_{c\,1} & \boldsymbol{c}_{c\,1} \\ 0 & \boldsymbol{f}_{c\,2} & \boldsymbol{c}_{c\,2} \\ 0 & 0 & 1 \end{bmatrix} \tag{4.11}$$

(Bouguet, 2008).

4.1.4 Reverse Projection

The camera parameters enable one to convert from pixel coordinates back to normalised coordinates. For a first order radial distortion, see Equation (4.8), Bouguet (1999) gives

$$\vec{\boldsymbol{x}}_N \approx \frac{\vec{\boldsymbol{x_d}}}{1 + \boldsymbol{\zeta} \left|\left| \frac{\vec{\boldsymbol{x_d}}}{1+\boldsymbol{\zeta}\|\vec{\boldsymbol{x_d}}\|^2} \right|\right|^2}. \tag{4.12}$$

Converting this into a homogeneous vector

$$\bar{\mathbf{x}} = \begin{bmatrix} X/Z \\ Y/Z \\ Z/Z \end{bmatrix}.$$

To obtain the correct 3D position, $\bar{\mathbf{x}}$ must be multiplied by the correct value Z, see Equation (4.5).

4.1.5 Rigid Body Motion Transformations

In the previous section, the 3D coordinates $\vec{\boldsymbol{X}}$ of a point were in the camera's reference frame $\mathcal{F}_C$. In many applications, however, it is more convenient to define this point by coordinates $\vec{\boldsymbol{X}}_{RW\mathcal{F}}$ in a 'real world' reference frame $\mathcal{F}_{RW}$, see Figure 4.4. To convert

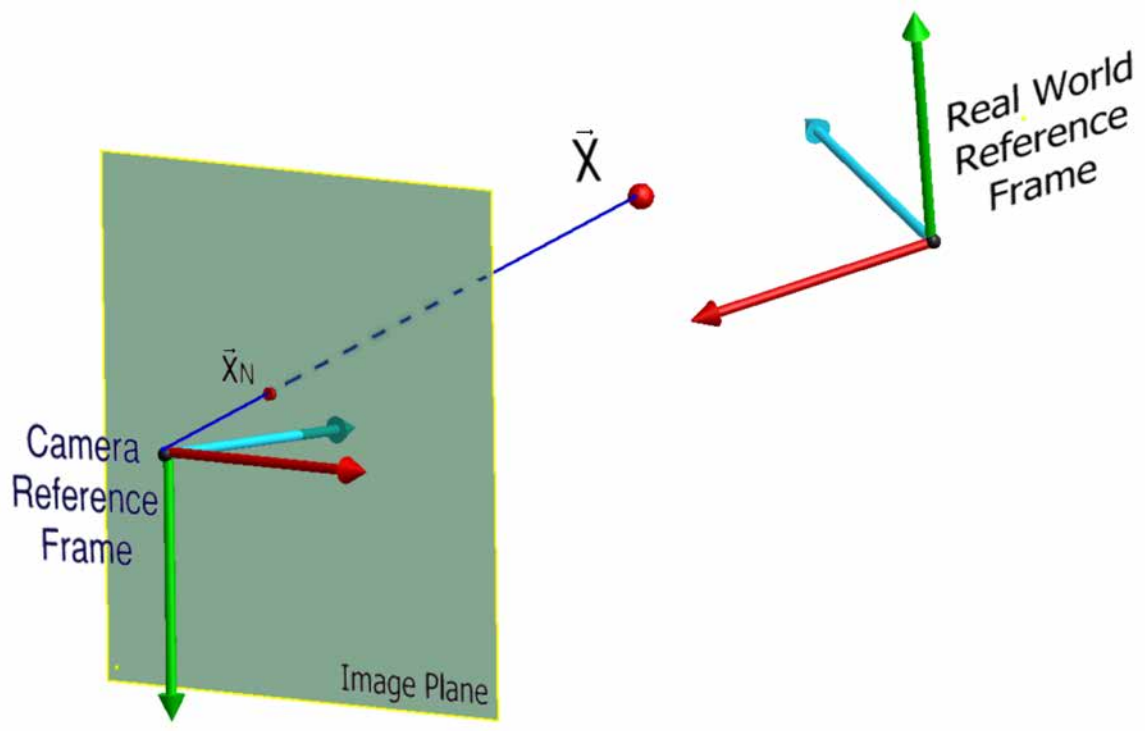


Figure 4.4: Illustration showing real world and camera reference frames.

these coordinates into camera reference frame coordinates, one may use a rigid body transformation

$$\vec{\boldsymbol{X}} = \mathbf{R}\,\vec{\boldsymbol{X}}_{RW\mathcal{F}} + \vec{\boldsymbol{T}} \tag{4.13}$$

where $\vec{\boldsymbol{T}}$ is a (3×1) translation vector and $\mathbf{R}$ is a (3×3) rotation matrix. The parameters $\vec{\boldsymbol{T}}$ and $\mathbf{R}$ are referred to as the extrinsic parameters of the camera and define the camera's position and orientation relative to a real world coordinate system. Note that if an array of coordinates is used

$$\vec{\boldsymbol{X}}_{RW\mathcal{F}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \\ Y_1 & Y_2 & \cdots & Y_N \\ Z_1 & Z_2 & \cdots & Z_N \end{bmatrix}, \tag{4.14}$$

then $\vec{\boldsymbol{T}}$ in Equation (4.13) is replaced by a $(3 \times N)$ matrix $\vec{\boldsymbol{T}}_{vect}$ where each column is the (3×1) vector $\vec{\boldsymbol{T}}$. An alternative form of Equation (4.13) is

$$\bar{\mathbf{X}} = \mathrm{g}\,\bar{\mathbf{X}}_{\boldsymbol{RW}} \tag{4.15}$$

where $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}_{\boldsymbol{RW}}$ are homogeneous vectors of the form $[X, Y, Z, 1]^{\intercal}$ and

$$\mathrm{g} = \begin{bmatrix} \mathbf{R} & \vec{\boldsymbol{T}} \\ 0_{1\times 3} & 1 \end{bmatrix}. \tag{4.16}$$

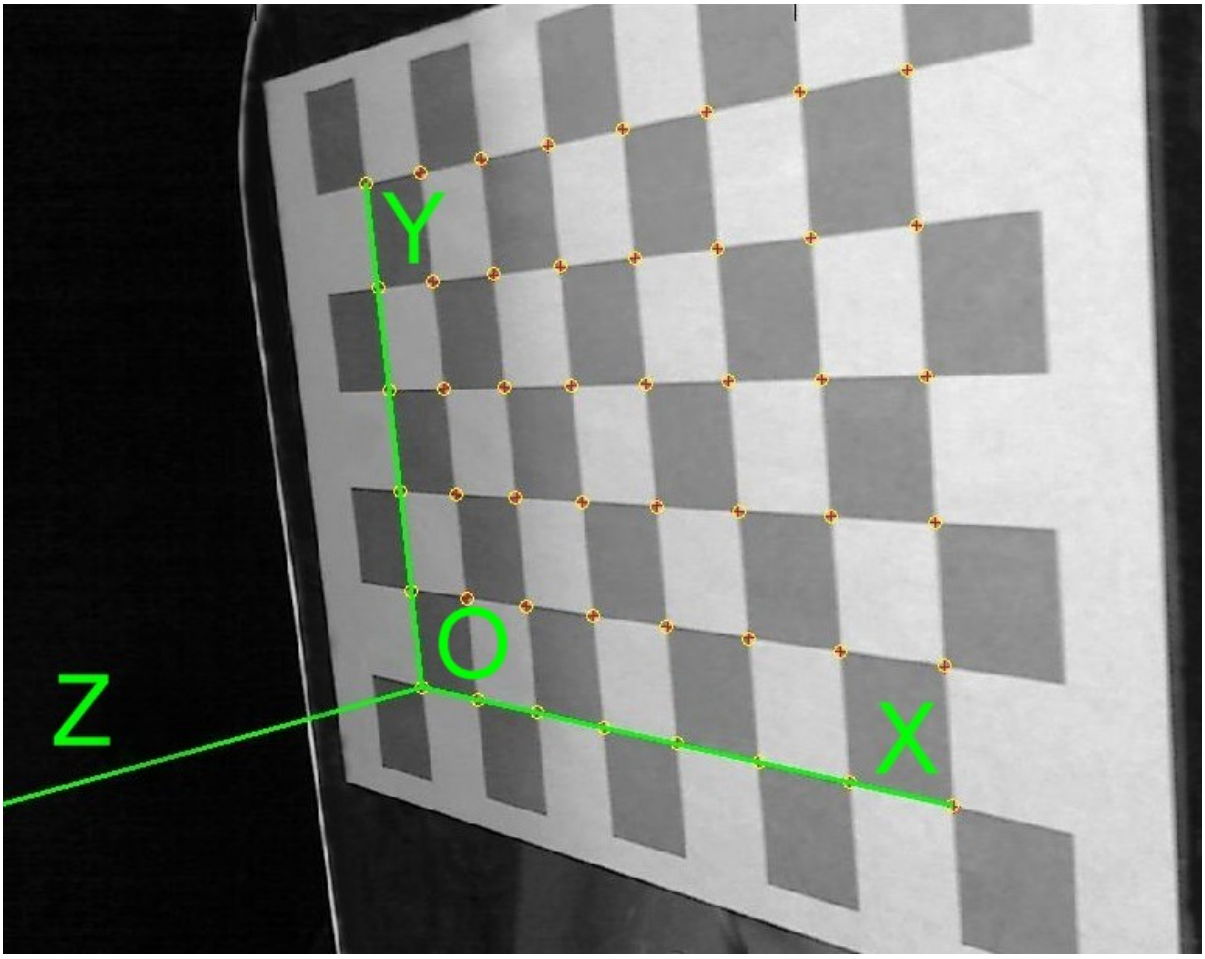


Figure 4.5: Example of a checkerboard image used in the '*MATLAB* Camera Calibration Tolbox'. Real world reference frame axes and detected corners have been plotted.

$\mathbf{R}$ has the properties that $\mathbf{R}^\intercal\ \mathbf{R} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and $det(\mathbf{R}) = 1$. Let $\vec{\mathbf{\Omega}} = [\Omega_x, \Omega_y, \Omega_z]^\intercal$ be a (3×1) rotation vector, where each component describes the amount of rotation in a given axis. The rotation matrix $\mathbf{R}$ may be calculated from the rotation vector using

$$\mathbf{R} = e^{i\Lambda} \tag{4.17}$$

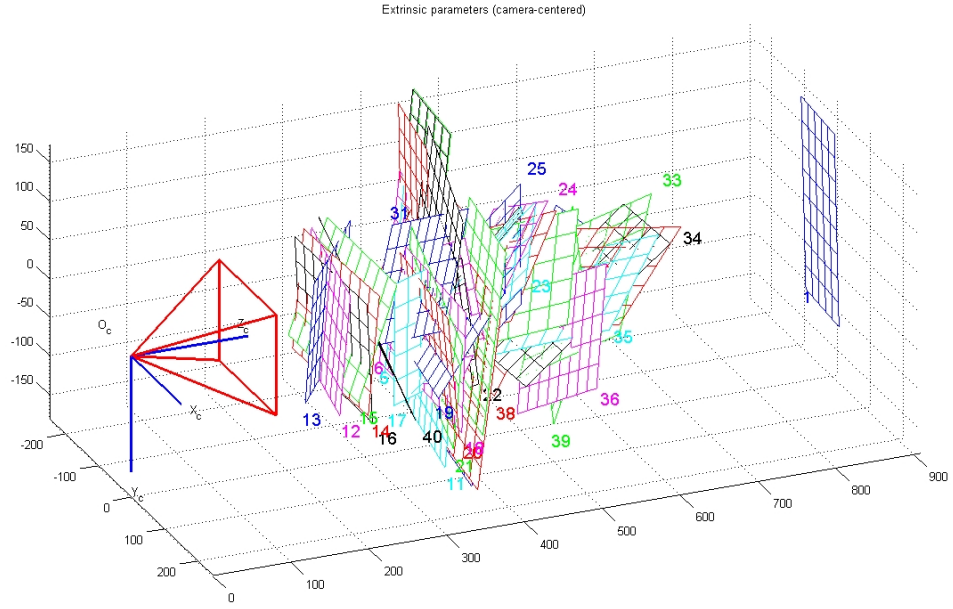
where

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & -\Omega_z & \Omega_y \\ \Omega_z & 0 & -\Omega_x \\ -\Omega_y & \Omega_x & 0 \end{bmatrix}. \tag{4.18}$$

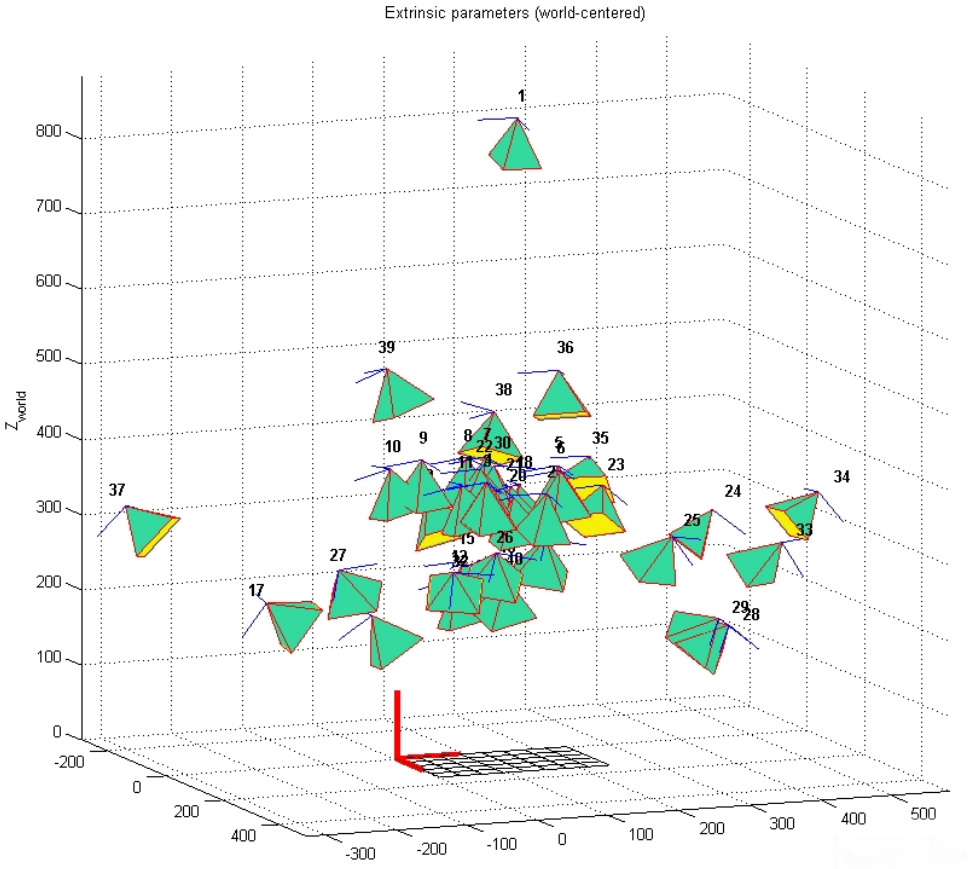
This may also be calculated using Rodrigues' formula

$$\mathbf{R} = \mathbf{I}_{3\times 3}\ \cos(\|\vec{\mathbf{\Omega}}\|) + [\mathbf{\Lambda}]\frac{\sin(\|\vec{\mathbf{\Omega}}\|)}{\|\vec{\mathbf{\Omega}}\|} + \vec{\mathbf{\Omega}}\vec{\mathbf{\Omega}}^\intercal\ \frac{1-\cos(\|\vec{\mathbf{\Omega}}\|)}{\|\vec{\mathbf{\Omega}}\|^2} \tag{4.19}$$

(Ma et al., 2003; Bouguet, 1999). To revert back to the original reference frame, one can



(a) Extrinsic parameters in camera-centred reference frame.



(b) Extrinsic parameters in world-centred reference frame.

Figure 4.6: Extrinsic parameters obtained during a camera calibration.

use

$$\vec{\boldsymbol{X}}_{RW\mathcal{F}} = \mathbf{R}^{\intercal} \, (\vec{\boldsymbol{X}} - \vec{\boldsymbol{T}}). \tag{4.20}$$

4.2 Camera Calibration

In this work, the '*MATLAB* Camera Calibration Toolbox' written by Bouguet (2008) is the main camera calibration software that has been used. This uses the Zhang calibration technique (Zhang, 1998) where a series of images of a checkerboard pattern are used to obtain the camera calibration parameters. The same method is used by the 'openCV' C++ calibration software (Bradski and Kaehler, 2008).

A checkerboard pattern is printed onto a sheet of paper and then glued onto a flat smooth surface such as perspex (or lucite) plastic. A series of images is then taken of this checkerboard pattern from a range of positions and orientations. For each image, Bouguet's software requires four extreme inner checker pattern corners to be manually selected by the user. All inner corners are then automatically detected using a corner detection algorithm and their coordinates $\vec{\boldsymbol{p}}$ in the pixel coordinate system are obtained. The 'openCV' software uses real time automatic corner detection which does not require manual selection by the user.

A real world reference frame $\mathcal{F}_{RW}$ is defined such that the origin O is the first corner selected, the X and Y axes are in the plane of the checkerboard and aligned with the checker pattern, and the Z axis points outwards perpendicular to the checkerboard plane, see Figure 4.5. The coordinates $\vec{\boldsymbol{X}}_{RW\mathcal{F}}$, of the checkerboard corners are then defined by the software in this reference frame using the dimensions $(dX \times dY)$ of the checkerboard squares.

The calibration software then uses both the pixel coordinates $\vec{\boldsymbol{p}}$ and the Euclidean coordinates $\vec{\boldsymbol{X}}_{RW\mathcal{F}}$ to obtain the intrinsic parameters $\boldsymbol{f}_c$, $\boldsymbol{c}_c$, α_c, and $\boldsymbol{\zeta}$ with errors. The extrinsic parameters $\vec{\boldsymbol{T}}$ and $\mathbf{R}$ with errors are also obtained for each calibration checkerboard pattern image, see Figure 4.6. Once the calibration has been performed, the intrinsic parameters may be used to obtain the extrinsic parameters for any image of a checkerboard pattern, where the dimensions of the checkerboard squares are known. Note that $\vec{\boldsymbol{T}}$ gives a point on the plane of the checkerboard while the third column of $\mathbf{R}$ is a vector which is perpendicular to the checkerboard plane. The combination of point $\vec{\boldsymbol{T}}$ with this vector gives the Z axis and describes the plane of the checkerboard.

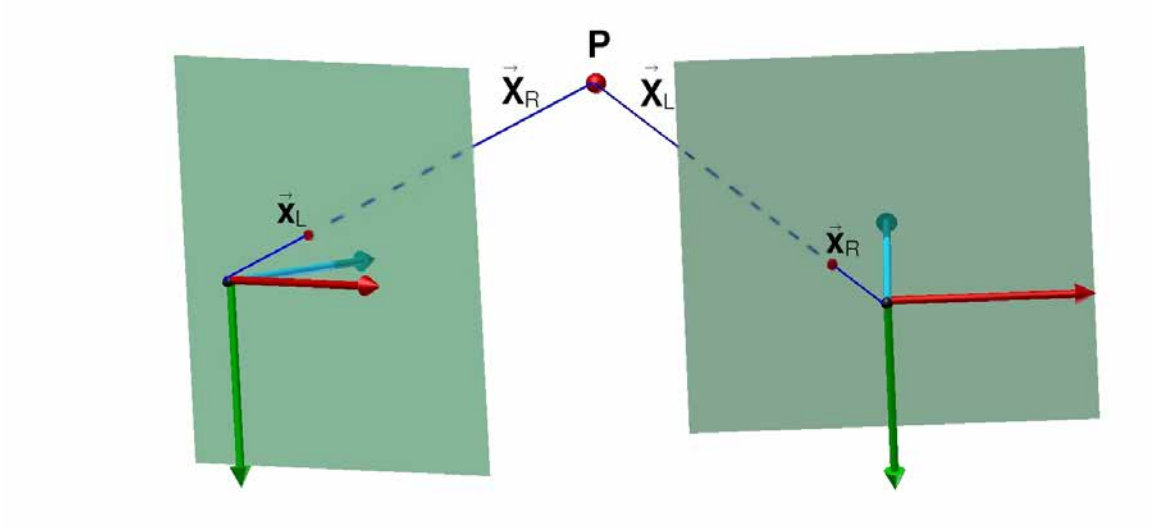


Figure 4.7: Stereo cameras setup with pinhole projection of a point P to normalise coordinates $\bar{\mathbf{x}}_L$ and $\bar{\mathbf{x}}_R$.

4.3 Stereo Triangulation

Consider a point P viewed by two cameras from two different positions and orientations. This point will correspond to normalised coordinates $\vec{\boldsymbol{x}}_L$ and $\vec{\boldsymbol{x}}_R$ in the left and right camera reference frames, see Figure 4.7. The aim of triangulation is to convert these normalised coordinates into 3D Euclidean coordinates $\vec{\boldsymbol{X}}_L$ and $\vec{\boldsymbol{X}}_R$. A summary of the method given in Bouguet (1998, 1999) is presented below.

If $\vec{\boldsymbol{T}}$ and $\mathbf{R}$ are the translation vector and rotation matrix that describe rigid body translation between the right and left camera reference frame, then

$$\vec{\boldsymbol{X}}_L = \mathbf{R}\,\vec{\boldsymbol{X}}_R + \vec{\boldsymbol{T}}. \tag{4.21}$$

This can be written in terms of homogeneous coordinates by

$$Z_L\,\bar{\mathbf{x}}_L = \mathbf{R}\,Z_R\,\bar{\mathbf{x}}_R + \vec{\boldsymbol{T}} \tag{4.22}$$

or in matrix form

$$\begin{bmatrix} -\mathbf{R}\,\bar{\mathbf{x}}_R & \bar{\mathbf{x}}_L \end{bmatrix} \begin{bmatrix} Z_R \\ Z_L \end{bmatrix} = \vec{\boldsymbol{T}}. \tag{4.23}$$

Substituting in $\mathcal{A} = \begin{bmatrix} -\mathbf{R}\,\bar{\mathbf{x}}_R & \bar{\mathbf{x}}_L \end{bmatrix}$ gives

$$\begin{bmatrix} Z_R \\ Z_L \end{bmatrix} = (\mathcal{A}^\intercal\,\mathcal{A})^{-1}\,\mathcal{A}^\intercal\,\vec{\boldsymbol{T}} \tag{4.24}$$

where

$$\begin{bmatrix} Z_R \\ Z_L \end{bmatrix} = \frac{\|\bar{\mathbf{x}}_L\|^2 \, (\alpha_R \cdot \vec{\boldsymbol{T}}) - (\alpha_R \cdot \bar{\mathbf{x}}_L) \, (\bar{\mathbf{x}}_L \cdot \vec{\boldsymbol{T}})}{\|\alpha_R\|^2 \, \|\bar{\mathbf{x}}_L\|^2 - (\alpha_R \cdot \bar{\mathbf{x}}_L)} \tag{4.25}$$

where $\alpha_R = -\mathbf{R}\,\bar{\mathbf{x}}_R$ and $(\cdot)$ is a dot product. The 3D Euclidean coordinates in the two cameras' reference frames are then given by $\vec{\boldsymbol{X}}_R = Z_R\,\bar{\mathbf{x}}_R$ and $\vec{\boldsymbol{X}}_L = Z_L\,\bar{\mathbf{x}}_L$. Different Z_L distanced from the camera result in different combinations of $\bar{\mathbf{x}}_L$ and $\bar{\mathbf{x}}_R$ values.

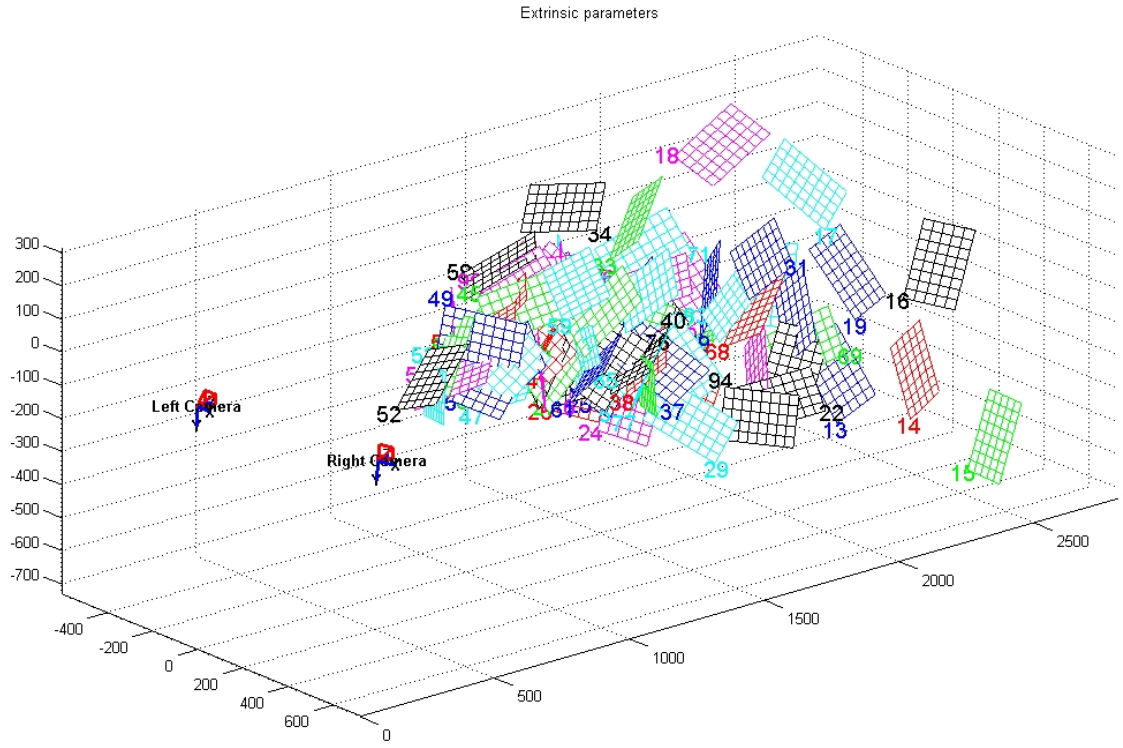


Figure 4.8: Stereo camera extrinsic parameters.

The *MATLAB* Calibration Toolbox includes stereo camera calibration. The two cameras capture images of a checkerboard pattern at the same time and two separate calibrations are obtained. These calibrations are then combined and the intrinsic parameters of each camera are refined and rigid body translation parameters $\vec{\boldsymbol{T}}$ and $\mathbf{R}$ are calculated in order to convert coordinates from the right camera's reference frame to the left camera's reference frame, see Figure 4.8 (Bouguet, 2008).

5

Surface Geometry Reconstruction Using Computer Vision

Acoustic imaging has traditionally used scan points located on a 2D plane with its normal parallel to the array axis. Recently, however, there has been work done on using a 3D scanning surface which corresponds to the real surface of an object emitting sound for beamforming. This chapter describes the computer vision theory and methodology used in this thesis to obtain the 3D surface geometry of objects for 3D acoustic imaging. While these methods are not in themselves novel, their application to acoustic imaging is.

5.1 Common Surface Reconstruction Techniques

A CAD model of an object may exist or be able to be created. If this is the case, the CAD model may be used to obtain the 3D scanning surface for acoustic imaging, see Section 5.2.1. Often, however, the CAD model may not exist or may be inaccurate or be incomplete. The object may also not be static and may have moving parts. This section outlines common methods used for obtaining the 3D geometry of an object.

5.1.1 Contact Surface Measurement Methods

Contact surface measurement techniques generally measure the coordinates of individual points on the surface of the object by placing a probe directly onto the object's surface. These data points are then used to generate a mesh corresponding to the surface of the object. This process can be time-consuming if the measurement area is large and/or complex, and such equipment can be bulky and expensive. Also contact with the surface of the object may not be desirable if, for example, the object is fragile or dynamic. Examples of contact measurement devices are robotic arms such as the *Faro-Arm*[1] or sonic devices such as *InterSense's IS-900*[2]. The latter obtains the 3D position of the probe by obtaining the time of flight of an ultrasound signal which is transmitted by the probe and received by multiple detectors.

5.1.2 Non-contact Surface Measurement Methods

Surface scanning using non-contact methods may be divided into two groups, passive and non-passive techniques. Passive techniques use only ambient lighting while active scanning uses light projected onto the object.

Passive Methods

Stereoscopic imaging is the most common passive method used for surface reconstruction. This method uses two or more calibrated cameras. Pixel coordinates in each camera's image are identified which correspond to the same points on the surface of the object. Triangulation is then used to reconstruct the 3D surface of the object. A problem with this method is that it requires a textured surface/key points which can be matched between multiple camera images. This can make 3D surface reconstruction unreliable. Other passive techniques developed include shape from shading, focus and defocus (Zhang et al., 1999; Jarvis, 1983; Krotkov, 1987; Subbarao et al., 1992).

Active Methods

Active methods use controlled lighting to overcome the problem of correspondence found in the stereoscopic passive method. However, active methods can be suspectable to errors if the object being scanned is reflective, transparent or in a brightly illuminated environment.

[1] `www.faro.com/gage/`
[2] `www.intersense.com/pages/20/14`

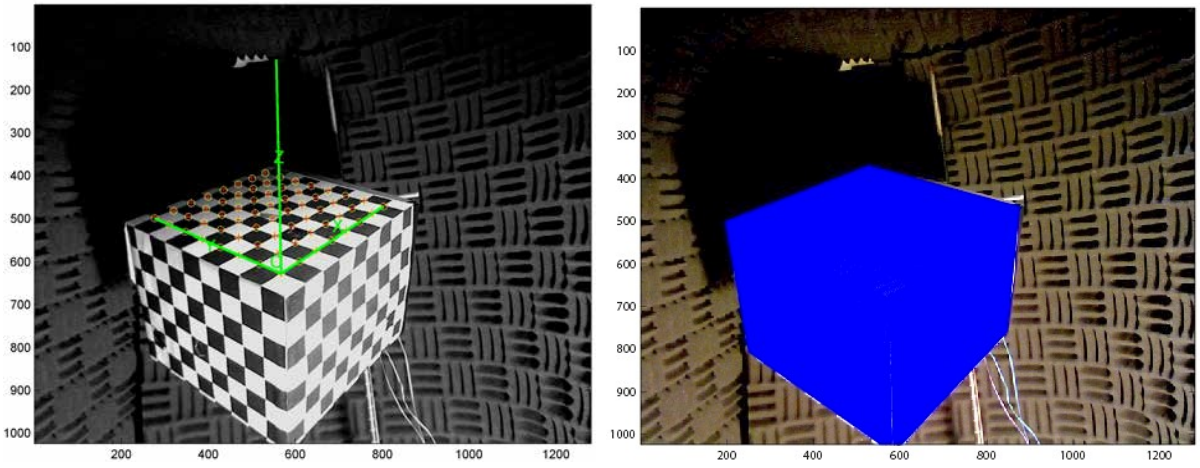
Laser Scanner: A laser scanner is commonly used as an active scanner. Mirrors are used to scan a laser beam or laser stripe across the surface of an object. A camera, which is offset from the laser, is used to capture images of this laser light on the surface of the object. Triangulation, which uses the offset of the camera and laser, allows the scanner to obtain high resolution images of the 3D surface of the object (Levoy, 1999). Since the 3D geometry can only be captured one point or line at a time, the scan process can be slow and not suited for dynamic scenes. Sections of the object that cannot be 'seen' by the laser or camera will result in 'holes'. Multiple scans from different positions might be required to obtain the full 3D surface of the object.

Structured Light Scanner: Structured light scanners attempt to decrease the scan time compared to a laser scanner. A projector is used to project an encoded light sequence onto the object. For a static object, high resolution scans can be achieved by projecting a temporally encoded pattern of light onto the object. An example of this is the binary or gray code, see Section 5.2.2. If the object is dynamic, has moving parts for example, a single spatially encoded light sequence may be projected onto the object, though a lower resolution will be achieved compared to temporal encoding sequences. An example of this is to use a colour code sequence such as the *De Bruijn* sequence (Li et al., 2002; Salvi et al., 2004; Batlle et al., 1998; Curless, 1999; Davis et al., 2005). Hybrid versions exist to increase resolution while keeping scan speeds high, for example the three phase-shift method (Zhang, 2005). The *Xbox Kinect*, formally known as Project Natal, is a game controller that was released by Microsoft in November 2010 and costs about $150USD. This uses *PrimeSense*'s infrared LED laser and a micromirror array to project dots of infrared light onto a scene. An infrared camera, a colour camera, and chip are then use to form a realtime 640×480 structured light system. Open source SDK and other software has been developed to enable the *Kinect* to be used with a computer, see OpenKinect or Burrus (2011) for example. As with a laser scanner, 'holes' appear in structured light scans where either the projector or cameras cannot see the object.

Time of Flight Camera: A time of flight camera obtains depth information by transmitting a pulse of light and measuring the time that it takes for light to reflect off the surface of the object and travel back to the receiver. The light source and receiver are co-located. A time of flight camera is, therefore, a radar like system and does not use triangulation. This allows dynamic scenes to be imaged. The resolution, however, is typically less than that of a laser or structured scanner and time of flight cameras are relatively expensive. These have less 'holes' or shadow effects, however, than a laser or structured light scanners, which have the light source and receiver spatially separated.

5.2 Surface Reconstruction Methods Used in this Work

5.2.1 Using Known 3D Surface Geometry and Computer Vision



(a) Extrinsic parameters using checkerboard image. (b) Scan surface over 3 sides of cube.

Figure 5.1: Example of the use of extrinsic parameters to obtain the 3D surface geometry of an object in the camera reference frame.

Often the 3D surface geometry of an object being acoustically imaged may be known; a CAD model of the object may exist or the geometry may be easily defined. If this is the case, this geometry may be used to define acoustic imaging scan points on the surface of the object. The coordinates of these points will usually be defined in a 'real world' reference frame, which is generally different from the reference frame used for acoustic imaging.

In this work, acoustic imaging is performed in the reference frame of the camera in the array. The coordinates of the scan points may be converted into the camera reference frame using the rigid body translation

$$\vec{\boldsymbol{\xi}}_{Cam\mathcal{F}} = \mathbf{R}\,\vec{\boldsymbol{\xi}}_{RW\mathcal{F}} + \vec{\boldsymbol{T}},$$

where $\mathbf{R}$ and $\vec{\boldsymbol{T}}$ are respectively a rotation matrix and a translation vector and are the extrinsic parameters of the camera relative to the object. These extrinsic parameters are obtained using an image of the checkerboard pattern, attached to the object, and using computer vision software. Acoustic imaging scan points were then calculated in the checkerboard 'real world' reference frame using the known geometry of the object. The

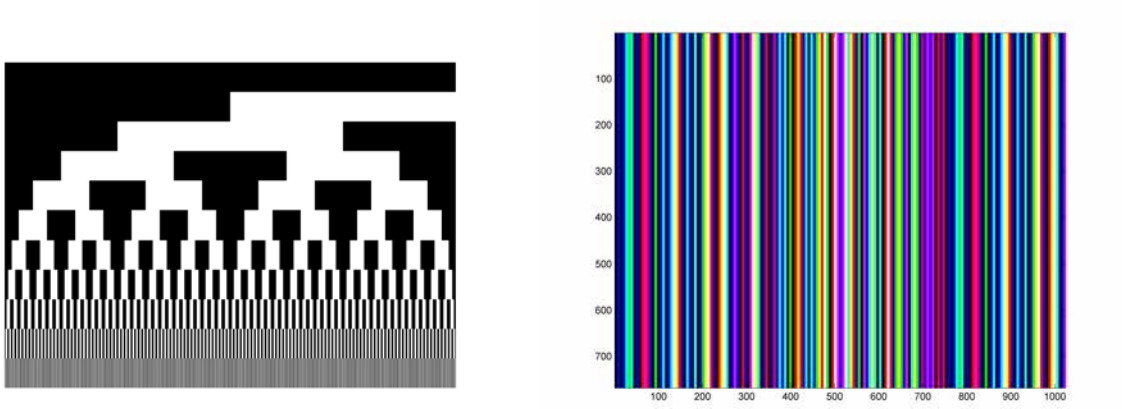
coordinates of these scan points were then converted to the camera/array reference frame using the extrinsic parameters and a rigid body rotation. Figure 5.1 shows an example where this method has been used for a cube. Acoustic imaging may then performed using these scan points in the camera reference frame.

The above method of using a checkerboard to define the reference frame of an object will be limited in application. For basic shapes such as boxes or planes, it is ideal. For more complex geometries, however, an alternative method would be required. Though not implemented in this work, the extrinsic parameters of the array could be determined using key points on the surface of the object. The pixel coordinates of these key points in the camera image and the corresponding 3D coordinates, in the object's 'real world' reference frame, could be used by computer vision software to obtain the extrinsic parameters. Beacons such as LEDs could be attached to these points for automatic detection in the camera images. Meyer and Döbler (2006) of *GFaI* developed an automatic method of aligning a CAD model with an array based on image recognition. An image is captured by the array's camera of the object and the CAD model projected onto this image. The CAD model is then iteratively translated and rotated until image recognition software obtains a match between the camera image and the projected CAD model image. It would appear that such a method would need distinctive features for image recognition and could be suspectable to discrepancies between the object and the CAD model or parts of the object not being included in the CAD model.

The beamforming algorithm assumes free field propagation from a scan point to the microphones. Scan points should, therefore, not be included in the beamforming if they are 'hidden' from view of the array. The scan points which do not have free field propagation paths to the centre of the array may be calculated using a ray tracing technique and these points excluded from the beamforming algorithm Schröder (2009).

5.2.2 Surface Reconstruction Using Structured Light

In the previous section, it was assumed that the object's geometry was known and could be defined relative to a checkerboard pattern or key points. Often, however, this will not be the case. A CAD model of the object may not exist or may be inaccurate. An automatic method of obtaining the scanning surface of an object is, therefore, needed. It was decided to use a structured light systemUnlike a laser scanning system, structured light systems are capable of real-time shape acquisition. A structured light scanner uses a data projector and one or more cameras to generate a scan of the surface of an object. This scan consist of a set of data points each having a 3D coordinate and a magnitude

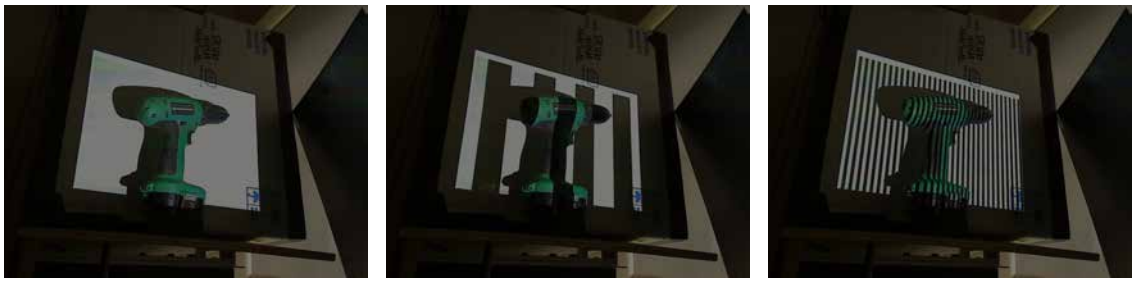


(a) Representation of the gray code structured light sequence where each row represents a projection image. The inverse (black to white, white to black) of each row is also displayed.

(b) Zhang's colour coded de Bruijn structured light sequence for real-time shape acquisition.

Figure 5.2: Examples of structured light sequences generated using code by Lanman and Taubin (2009a).

relating to the color of the object at that point. This is often referred to as a coloured 3D point cloud. These are often processed to form a mesh of the surface of an object.



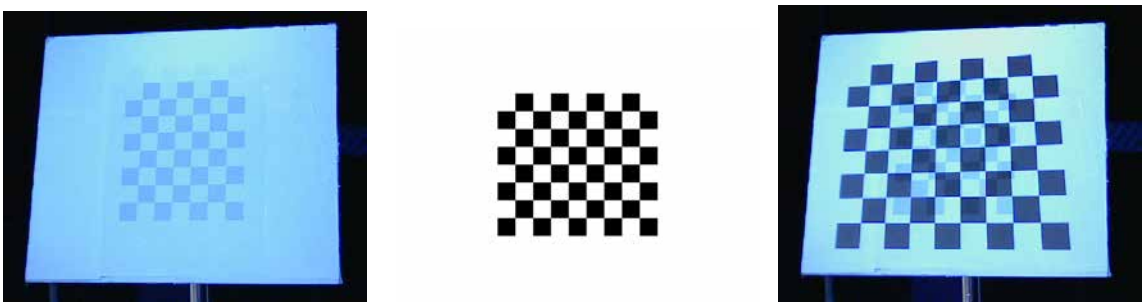
(a) (b) (c)

Figure 5.3: A selection of 42 gray code structured light sequence images for a scan of a electric drill.

Most microphone phased arrays contain a camera and hence only the addition of a data projector and software would be required to convert these microphone arrays into a structured light scanner. The point cloud generated will be in the reference frame of the main camera in the array. This point cloud may then be used as the acoustic imaging scan points and may also be used to plot the resulting 3D acoustic map. A disadvantage, however, of a standard single projector/camera structured light system is that the field of view is limited. This would be an issue for imaging the interior of cars, for example, where it might be required to view the full interior.

(a) (b)

Figure 5.4: A coloured point cloud generated by a structured light scan of an electric drill using Lanman and Taubin's 'cvStructuredLight' code. Two views of a point cloud from a single scan are shown.



(a) Screen with printed checkerboard.

(b) Checkerboard image sent to data projector.

(c) Projected checkerboard on the screen captured by camera.

Figure 5.5: *MATLAB* structured light calibration example images.

Structured Light Calibration

The projector can be modelled as a camera where the light travels in the opposite direction to a camera. To achieve accurate structured light scans, the intrinsic and extrinsic parameters of the projector and cameras need to be obtained using calibration. In this work, structured light calibration was mainly achieved using a combination of modified versions of 'mlStructuredLight' (Lanman and Taubin, 2009a) structured light *MATLAB* code, 'Camera-Projector Calibration Toolbox' (Falcao et al., 2009), and '*MATLAB* Camera Calibration Toolbox' (Bouguet, 2008). The 'cvStructuredLight' (Lanman and Taubin, 2009a) openCV C++ code was also used. This is more straightforward to use but does not include multiple camera calibration/scanning.

Calibration parameters were obtained by projecting a checkerboard pattern onto the screen which has a printed checkerboard pattern glued onto it. An image is captured and the projected checkerboard pattern is turned off. Without moving the screen, an image is then captured of the printed checkerboard pattern on the screen. This is repeated

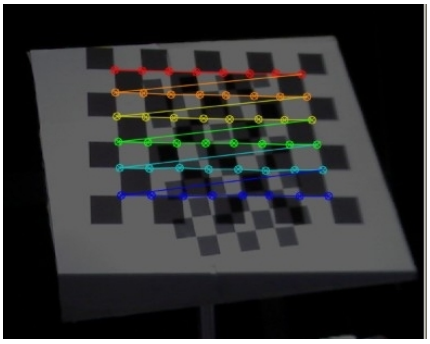


Figure 5.6: Example of the use of an openCV structured light calibration method written by Lanman and Taubin.

for a range of positions and orientations of the screen, see Figures 5.5 and 5.6. These data are used to obtain the camera's and projector's intrinsic and extrinsic parameters in the camera's reference frame, see Figure 5.7. The extrinsic parameters of multiple cameras may be obtained using images of a printed checkerboard pattern in one position.

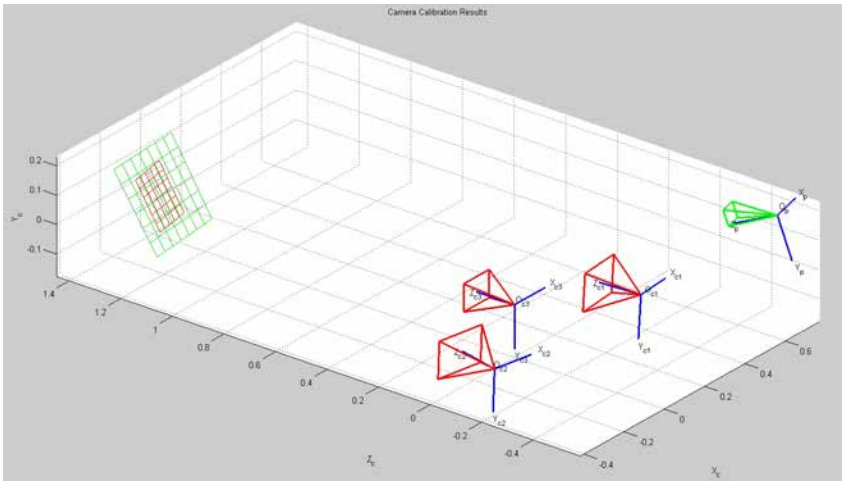


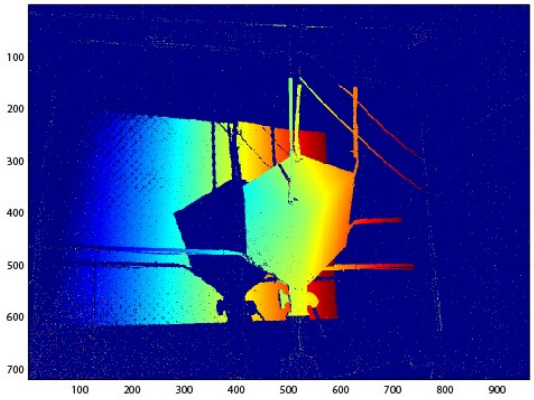
Figure 5.7: Extrinsic parameters for three cameras and a projector obtained using *MATLAB* structured light calibration.

Description of Structured Light Implementation

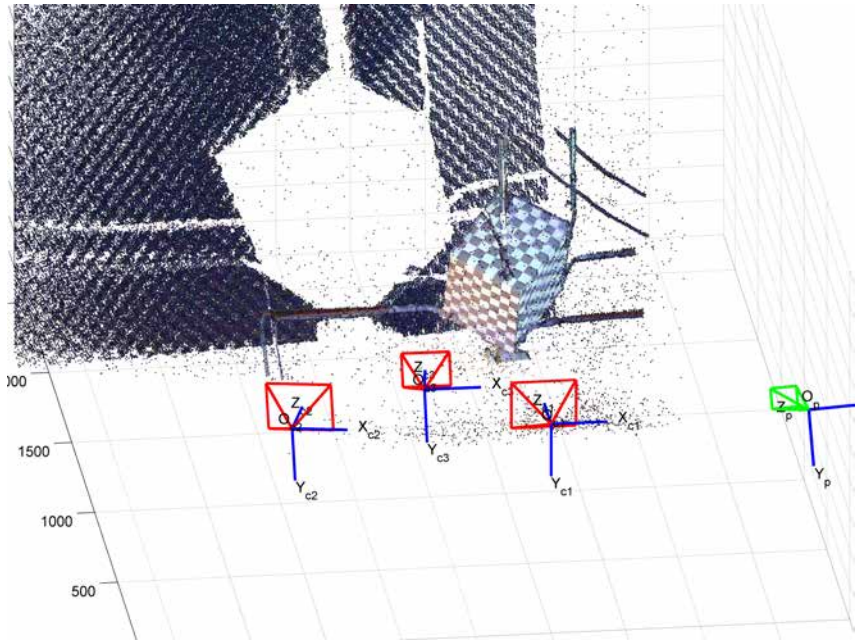
There is a range of structured light methods such as colour coded, three phase and binary or gray code structured light sequences, see Figure 5.2. Binary and gray code structured light scanning software, written by Lanman and Taubin, has been used in this work. A sequence of images, consisting of columns and rows of light and dark stripes are projected onto an object. One or more cameras are used to capture images of this sequence reflected from the surface of the object, see Figure 5.3 for several examples. The deflection of these stripes enables a coloured 3D point cloud to be generated using triangulation, see Figures

5.4 and 5.8. Problems in detecting the surface can occur if the object is shiny, transparent or in a brightly lit environment. An extension to this work would be to implement a structured light method, such as a colour coded structured light sequence, that would allow dynamic objects to be imaged.

(a) Image used for texture map.



(b) Depth map.



(c) Raw structured light point cloud.

Figure 5.8: Structured light scan.

Point Cloud Processing/Merging

The raw point cloud obtained by a structured light scan was processed to limit the spatial 3D range, remove noise, and reduce the number of scan points. First, scan points outside the area of interest were removed. To remove noise, isolated points were removed using an algorithm that calculated the distance of each point from its N nearest neighbours. If the

sum of these distances was greater than a cut off value, the point was removed from the point cloud, compare Figures 5.8 and 5.9. In this example, scan points further than 5mm from their nearest neighbours were considered to be noise an were removed. Structured light scans can generate very dense point clouds with a large number of scan points, which could be computationally expensive if used directly for acoustic imaging. To address this, the Euclidean distance of a scan point to its neighbour was calculated and removed if this distance was less than a cut off value. This cut off distance was iteratively increased until the desired resolution was achieved, see Figure 5.10. For this example, the cutoff value was initially 1.5mm. This value was interactively increased by a factor of 1.01 until the number of scan points was less than 10,000.

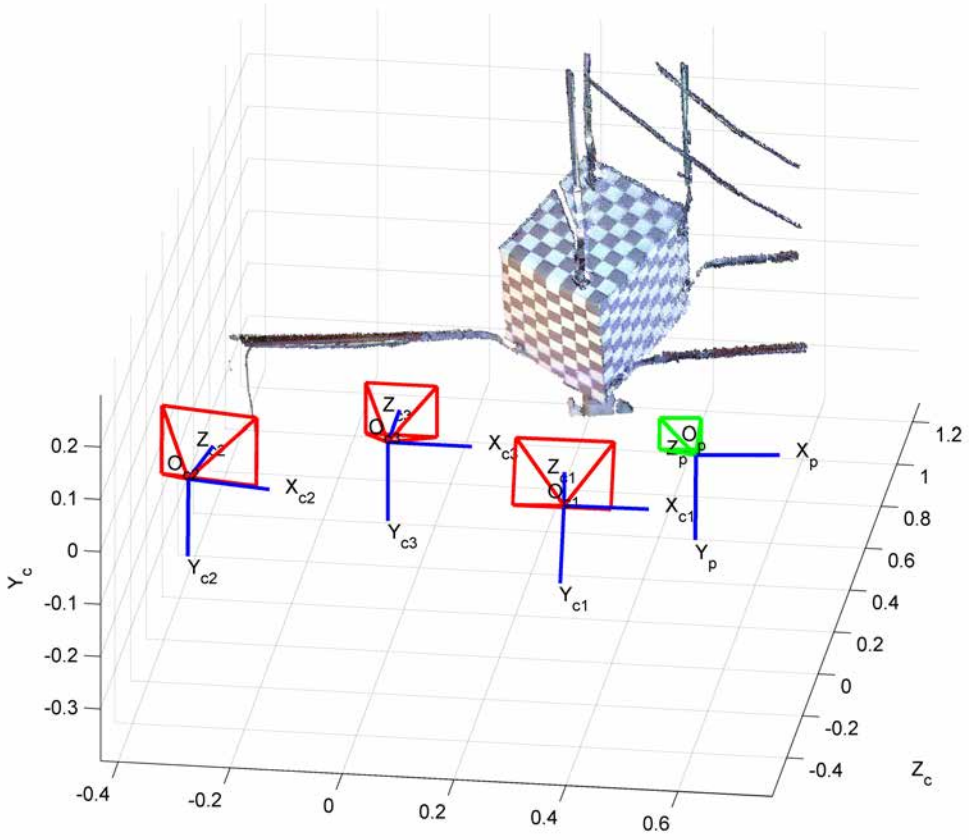
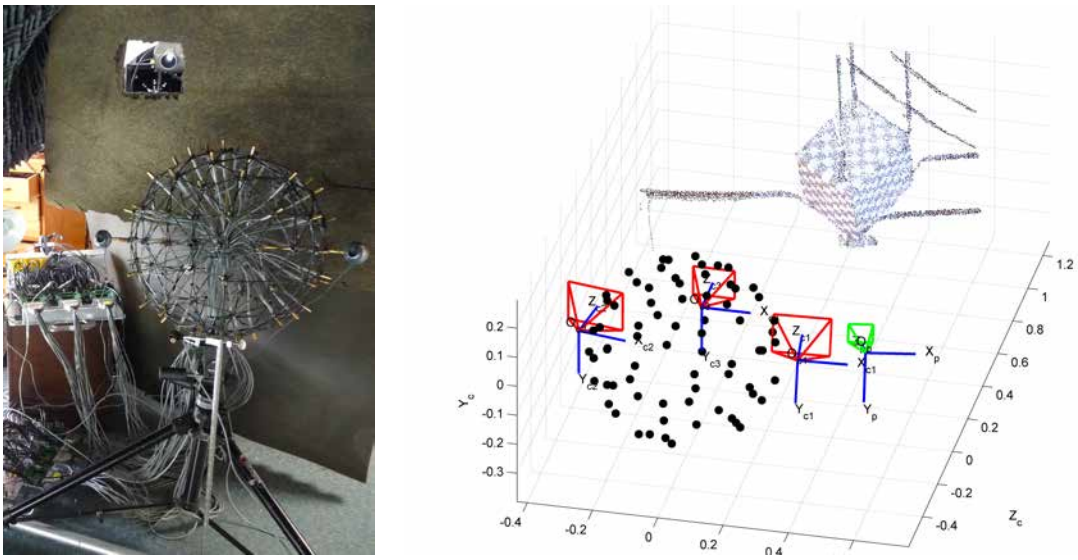


Figure 5.9: Structured light point clouds and filtering, background and isolated data points removed. The positions of the cameras (red icons) and projector (green icon) are also plotted.

The scanner can only detect the portion of the object's surface which is illuminated by the data projector and which can be seen by the camera/cameras. To generate a full model of an object, multiple scans may need to be made, with the object at different orientations, and the resulting point clouds merged. To merge two point clouds, a rigid body rotation is used to roughly align the point clouds. The parameters of this transformation may be obtained from knowledge of the extrinsic parameters of the scanner relative to the object's

reference frame. Alternatively, common features in the two point clouds may be selected, manually or automatically, and the rigid body transformation parameters calculated. An iterative closest point (ICP) algorithm of overlapping sections may then be used to fine tune the merging of the point clouds.



(a) Structured light setup. Three camera are attached to the spherical microphone array. A standard data projector is mounted to a stand behind and to the side of the array.

(b) Down-sampled point cloud for acoustic imaging. Microphone coordinates, obtained using calibration techniques described in Section 6.3, are also plotted.

Figure 5.10: Structured light scan processed for acoustic imaging.

5.2.3 Possible Extensions

A draw-back of using a standard structured light system is that the scan has a limited field of view. For imaging enclosures, such as the interior of a car, one would wish to image the full interior. A non-ideal option could be to use a structured light scan of a section of the object to align the CAD model with the array using an algorithm such as the iterative closest point algorithm (ICP). Alternatively, multiple, wider field of view, camera/projector pairs or a revolving single camera/projector pair could be used perhaps using the same IR structured light sensor components developed by *PrimeSense* and used in the *Xbox Kinect*. An omnidirectional system could also be investigated, see (Orghidan, 2005). Perhaps a method similar to that used by Béguet and Robin (2010) could be used where the array and or the the scanner system are rotated by software.

A more ideal scenario, however, might be to use a computer vision code that allows the array, and hence the structured light scanner, to be moved about an object or enclosure

and the computer automatically to know the position and orientation of the array. A 3D surface could then be built up as the array is moved. Such computer vision methods have been developed (Klein and Murray, 2007, 2009; Robbel). Acoustic imaging could be performed as this is happening or after the full surface has been built up.

6

Microphone Position Calibration in Web Camera Reference Frame

A microphone position calibration technique has been developed that uses computer vision and time of arrival data to obtain microphone coordinates in a reference frame of the camera in the microphone array. To avoid any confusion with the terms 'acoustic camera' and optical camera, the term 'web camera' will be used in this chapter.

6.1 Outline of Problem

The coordinates of microphones in an array are required for delay and sum beamforming. The required accuracy increases as the acoustic frequency increases. While the eigenvalue calibration technique developed by Dougherty (2002), see Section 2.2.2, includes some correction for some microphone position errors, it still requires relatively accurate microphone positions. Underbrink (2002) states that microphone position accuracy needs to be better than 2.5 mm for higher audio frequencies.

Acoustic maps are commonly overlayed over an image obtained by a web camera located at the centre of the array. Consider an acoustic map generated using scan points $\vec{\xi}$ and

microphone coordinates $\vec{\boldsymbol{X}}_m$ in a web camera's reference frame. This acoustic map may be plotted as a transparency over an undistorted web camera image using the web camera's intrinsic parameters and Equations (4.4) and (4.9)

$$\vec{\boldsymbol{x}}_N = \begin{bmatrix} \vec{\boldsymbol{\xi}}_1 / \vec{\boldsymbol{\xi}}_3 \\ \vec{\boldsymbol{\xi}}_2 / \vec{\boldsymbol{\xi}}_3 \end{bmatrix}$$

$$\vec{\boldsymbol{p}} = \begin{bmatrix} \boldsymbol{f}_{c1} \ (\vec{\boldsymbol{x}}_{N1} + \alpha_c \, \vec{\boldsymbol{x}}_{N2}) + \boldsymbol{c}_{c1} \\ \boldsymbol{f}_{c2} \, \vec{\boldsymbol{x}}_{N2} + \boldsymbol{c}_{c2} \end{bmatrix}.$$

This acoustic map may also be plotted over a distorted web camera image by distorting $\vec{\boldsymbol{x}}_N$ using Equation (4.6).

This assumes, however, that both the microphone and scan coordinates are in the web camera's reference frame. In reality, however, the microphone coordinates will be measured in a 'microphone reference frame'. For example, a planar array generally has the origin at the centre of the array, the X and Y axes in the plane of the array, and has the Z axis pointing outwards from the array, see Figures 1.3 and 6.1. Acoustic maps, generated using these coordinates, will also be in the 'microphone' reference frame. The web camera may be positioned and rotated until the web camera's reference frame is the same as the microphone reference frame. This process, however, can be time-consuming. A more general method is to convert the microphone coordinates into the web camera's reference frame using

$$\vec{\boldsymbol{X}}_{m_{Cam\mathcal{F}}} = \mathbf{R}_m \, \vec{\boldsymbol{X}}_{m_{Mic\mathcal{F}}} + \vec{\boldsymbol{T}}_m, \tag{6.1}$$

where $\mathbf{R}_m$ and $\vec{\boldsymbol{T}}_m$ are a rotation matrix and a translation vector. Note that a $[3 \times 1]$

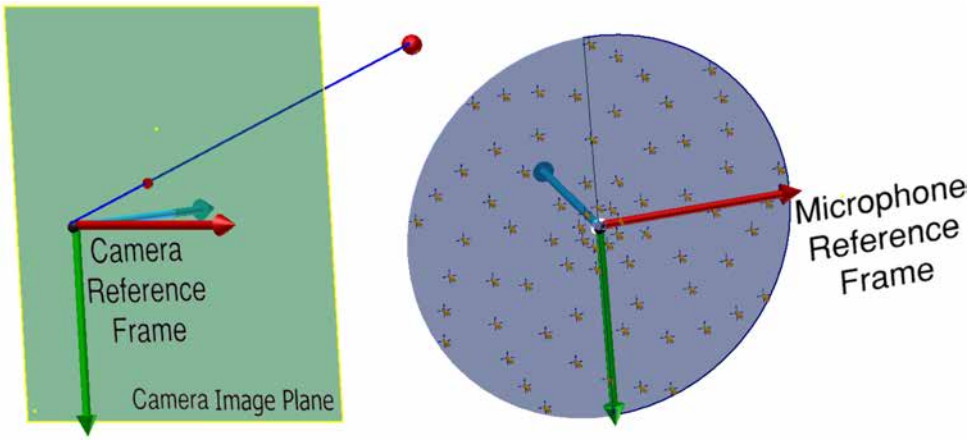
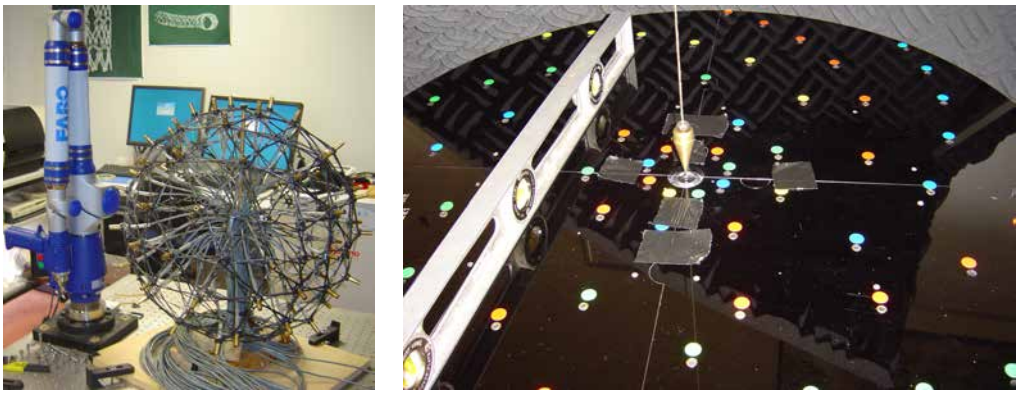


Figure 6.1: Web camera and microphone reference frames.

rotation vector $\vec{\boldsymbol{\Omega}}_m$ may be converted into the equivalent $[3 \times 3]$ rotation matrix $\mathbf{R}$ using Rodrigues' equation, see Equation (4.19). The parameters $\mathbf{R}_m$ and $\vec{\boldsymbol{T}}_m$ are the extrinsic parameters of the web camera relative to the microphone reference frame. If a web camera

is to be used in an array, calibration should either give microphone coordinates in the web camera reference frame or measure the extrinsic parameters between the web camera and the microphone reference frames.



(a) Measurement of microphone coordinates using a Faro-arm.

(b) Approximate physical alignment of web camera with array. The optical centre of the web camera image was aligned with a plumb bob when raised to near the ceiling. Cross wires gave orientation with panel coordinates.

Figure 6.2: Photos illustrating the initial methods used to roughly align the camera and measure microphone coordinates.

The initial calibration of the arrays, described in Chapter 3, was done using a manual process. Microphone coordinates for 2D arrays were obtained using the CAD models used to laser cut the array panels. The microphone coordinates for the spherical array were obtained using a Faro Arm, see Figure 6.2a. Careful labelling of each microphone and holder ensured that the correct microphone coordinates were given to each microphone. The main web cameras were initially aligned with the array, see Figure 6.2b. A speaker was setup in front of the array. The microphone coordinates were then incrementally rotated and translated, using Equation (6.1), until extrinsic parameters were found that resulted in the main peak in the beamformed/deconvolution map overlaying the speaker in the web camera image for a range of speaker positions.

This calibration process was time-consuming. An automatic calibration technique is desirable. This technique ideally would allow the user to arbitrarily position microphones and web cameras in 3D space, and be able to obtain the microphone coordinates in any of the web camera reference frames without any *a priori* knowledge of their positions.

A new calibration technique has been developed by the author that combines web camera calibration with microphone position calibration. The 3D coordinates of the sources are obtained using information obtained during web camera calibration. These locations are in the web camera's reference frame. Time of Arrival (TOF) of sound emitted from speakers to each microphone is used to obtain the microphone positions in the web camera's reference frame. No *a priori* information about microphone positions is required.

6.2 Previous Calibration Methods

6.2.1 Manual Microphone Position Calibration

If a planar array panel is built using a milling machine or laser cutter, for example, the coordinates may be known to very high accuracy. If this is not the case, however, the coordinates will need to be measured. Birchfield (2003); Birchfield and Subramanya (2005) use a point based multi-dimensional scaling technique which requires as an input measurements of the distances between the microphones using a tape measure. Measurements of microphone coordinates with a tape measure become very difficult as the number of microphones and geometric complexity increases. An alternative method is to use measurement tools such as a Faro-arm, a laser scanner, survey equipment (Manthe et al., 2008), or a sonic digitiser (Underbrink, 2002). Such equipment, however, is costly and often is not a practical option. An automatic method for obtaining microphone positions is, therefore, desirable.

6.2.2 Automatic Microphone Position Calibration

Calibration of the positions of microphones in an array, commonly known as array shape calibration, may be divided into two parts: self calibration, where source positions are unknown; and calibration using known source location.

Using Unknown Source Locations

A self calibration technique was developed by Rockah and Schultheiss (1987a,b) for the case where microphone coordinates in a 2D array were not known precisely. Three non-collinear sound sources are used. The orientation of a 2D array was obtained by defining the coordinates of one sensor and measuring the direction to a second sensor. Weiss and Friedlander (1989, 1991) proposed an eigenstructure based method which is similar to the Multiple Signal Classification (MUSIC) algorithm.

Raykar et al. (2003, 2005) proposed a method, using calibration signals (short chirps), for calibrating a network of general-purpose computers (GPC) such as laptops, PDAs, and tablets which contain audio/video sensors. No prior information of microphone positions or synchronisation of platforms was allowed for. This method uses the time difference of flight

$$TDOF_{mm's} = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{X}}_s\|}{c} \tag{6.2}$$

between multiple microphone pairs $\vec{\boldsymbol{X}}_m$ and $\vec{\boldsymbol{X}}_{m'}$ for sound sources located at $\vec{\boldsymbol{X}}_s$. If a non-synchronised system was used, additional time delay terms were added. The actual time of flight values were obtained using General Cross Correlation (GCC) with Phase Transform ($PHAT$) (Knapp and Carter, 1976) weighting between microphone signals. Metric multidimensional scaling (MDS) was used to get an approximate initial guess for the microphone and speaker locations and then an iterative nonlinear least square optimisation procedure was applied. A microphone position error of 3.8cm was obtained for a synchronised system. Hörster et al. (2005) propose a way to automatically calibrate multiple cameras in a distributed platform of GPCs. The position of any microphone relative to these web cameras is not addressed, however. A simplified method is presented in Raykar and Duraiswami (2004) and Zotkin et al. (2007) where at least five mini-speakers were attached to different microphones in an array and time of flight

$$TOF_{ms} = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\|}{c} \tag{6.3}$$

was used.

McCowan et al. (2008) proposed an unsupervised method for relatively small arrays that did not require any calibration signal but instead used a diffused noise field. No *a priori* knowledge of microphone positions is required. An average error of 0.5 to 1.7cm is obtained with this method. This method was used by Hennecke et al. (2009) for local shape calibration of individual groups of microphones followed by a network calibration performed using the Steered Response Power algorithm in combination with a Phase Transformation ($SPR\text{-}PHAT$). In Redondi et al. (2009) and Valente et al. (2010) a method is proposed which estimates the rigid body motion between the coordinate systems of two microphone arrays using least squares. Although no cameras were used, the two arrays are modelled as cameras and computer vision algorithms are used.

A problem with self calibration techniques was that the microphone position errors were too large for high frequency beamforming measurements. To address this, Döbler et al. (2010) present a technique which increases the number of sources to 50 - 100 using a taser (spark generator). Time difference of arrival measurements were obtained using a reference microphone and an iterative method was used to obtain both microphone and sound source locations. Conditions, such as the array being planar and the source only being in front of the array, were used in the algorithm. To help reach convergence, the microphone coordinates should be known prior to calibration to within 0.1m. After calibration, an accuracy of better than 2mm for a 2D array and 4mm for a 3D array was achieved. Other self calibration techniques include those by Moses and Patterson (2002),Thrun (2006), Chen et al. (2007), and Jager et al. (2009). Also see a recent paper

by Teal and Poletti (2010) on self calibration of the phase of microphones in an array for acoustic holography.

Using Sources with Known Locations

Seymour et al. (1987) proposed an algorithm for compensating for microphone position errors using acoustic signals from sensors from known directions. The algorithm is based on the MUSIC method developed by Schmidt (1986). Lo and Marple Jr (1987) used analysis of the eigenvectors and eigenvalues of the sample covariance matrix of signals from two sound sources. Cevher and McClellan (2001, 2006) used a moving source to calibrate the position of sensors. An approximate position of the source is obtained by the Global Positioning System (GPS).

Sachar et al. (2002, 2005) use a calibration rig with source points at known locations for shape calibration for the 'Huge Microphone-Array'. The calibration consisted of five tweeters on a tetrahedron shaped structure. It uses time-delay estimation of short chirp signals. The technique was also used for automatic gain calibration. A microphone position error of the order of 0.5cm was achieved. Lauterbach et al. (2009) (Patent DE 10 2008 017 001.1-09) use 8 tweeters at known locations. To increase accuracy of the measured source locations, the sound from the tweeters is funnelled through conduits inserted into a panel. Microphone position errors of the order of 1 mm could be achieved. The error in microphone positions was shown to decrease as the number of sources was increased. Other microphone position calibration techniques using sources with known locations include those by Ng and Nehorai (1991); Ng and Ser (1992); Ng and See (1996), Koerber and Fuhrmann (1993) Xu et al. (2010); Basten et al. (2010) and Moebus (2011).

6.2.3 Calibration of Web Camera Extrinsic Parameters Relative to Microphone Array Reference Frame.

Literature was found that obtained the extrinsic parameters of the web camera relative an array. This literature may be divided into three groups, acoustic imaging, conference room speech enhancement, and underwater sonar.

Acoustic Imaging Arrays

Most commercial microphone phased arrays used for acoustic imaging appear to use a single fixed web camera located at the centre of the array. Dougherty (2011) describes

in Optinav's Array 24 operation manual a method of aligning a web camera with the beamformed map. A beamformed map is generated for a distant speaker location and overlayed over a web camera image. In software, the web camera pan and tilt parameters are adjusted until the beamformed spot is on the centre of the speaker. A nearby speaker is then used to obtain the web camera offset parameter. The x_{camera} and y_{camera} offset parameters are adjusted, iteratively as necessary, while keeping other parameters fixed. It would appear that either the microphone coordinates or the coordinates of points on the 2D beamformed map are being rotated and translated before being projected onto the image plane. Camera distortion correction is also allowed for. The method used by other manufacturers is not known by the author.

Li et al. (2008); Yang et al. (2011) use two high speed cameras and a cross array for acoustic holography imaging of noise from a moving car. The cameras are positioned on either side of the centre of the array in the plane of the array. The two cameras are calibrated as a stereo pair using computer vision techniques. The stereo pairs are then used to determine the distance to a marker on the side of a car. A 2D scanning surface parallel to the array is then generated in a 'real world reference frame' at this distance which is used for acoustic imaging. The acoustic map is then projected onto the image of one of the cameras. The method used to obtain the rigid body rotation between the cameras and the microphone array is unclear. Conformation of alignment is achieved by using a speaker with one of the markers attached to it.

Conference Room Arrays

There is a significant amount of literature related to combining cameras and microphone arrays for conference room purposes. The location of a person's head might be detected on a camera image using methods such as face detection. This is then used to steer the microphone array for speech enhancement. Alternatively a pan and tilt camera may be steered using the location of a speaker obtained using microphone array data. In much of the literature, the alignment of the camera and array is assumed to be known. Below are several references that were found where camera alignment with microphone array calibration is included.

Zotkin et al. (2001) obtained microphone coordinates by using stereo camera images of the microphones and triangulation. These measurements were described as being inaccurate, however, due to the distance of the microphones from the cameras. O'Donovan et al. (2007b,a) models a solid spherical array and spherical beamforming as a 'central-projection' camera. A method is proposed which uses computer vision to obtain the extrinsic parameters of a camera relative to the microphone array. This camera was setup

so that it can see the microphone array. Beamformed acoustic maps and camera images were then obtained of a mini-speaker with a LED attached for multiple positions. The LED pixel coordinates for each camera image and the corresponding main peaks in the acoustic maps are then used to obtain the essential matrix and hence the extrinsic parameters of the camera relative to the array. The geometry of the camera and array was used to transfer a point from a camera image to a 1-D curve in the microphone array directional space. This technique was used with face detection for speech enhancement.

Ettinger and Freund (2008) present a calibration technique for a system that automatically steers a pan and tilt (PTZ) camera using an ad-hoc microphone array. A LED was attached to the centre of a speaker. This was used to keep the camera centred on the speaker as it was moved about a room. Time difference of arrivals for all microphones in the array were obtained for a range of positions of speaker in a room. These delay times were used to obtain the relationship between the time differences of arrivals and the pan and tilt angles of the PTZ camera. Schulz et al. (2008) present a patented method of obtaining microphone position as a function of the pan and tilt angles and distance from a PTZ camera based on detection of LEDs on an array and size of array in the image.

SONAR

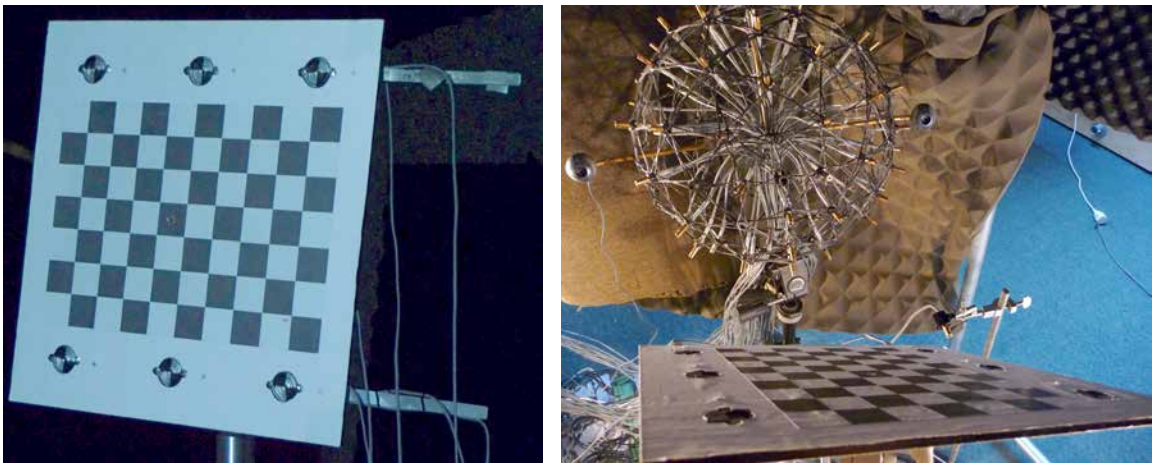
An "optio-acoustic sensor" refers to a merging of a sonar device with a camera for underwater object imaging. These sonic devices transmit ultrasound pulses and, by analysing the reflected signal, are able to generate a point cloud image of objects in front of them. A camera is incorporated for closeup higher resolution imaging. These devices differ from microphone phased arrays, which only listen to the sound emitted by objects and have very low resolution in a direction along the array axis. They have been included here for completeness. The extrinsic parameters of camera sensors relative to sonar devices was developed using computer vision techniques with registration of the acoustic point cloud scan of an object with camera images (Fusiello and Murino, 2000, 2004; Negahdaripour, 2005; Negahdaripour et al., 2007; Negahdaripour, 2010). A SONAR scan made over a checkerboard has also been used (Vilarnau, 2009; Hurtós et al., 2010).

6.3 Microphone and Camera Calibration Using Computer Vision

Array calibration for microphone coordinates and camera alignment requires three steps: obtain microphone coordinates, obtain camera's intrinsic calibration parameters, and then

extrinsic parameters/alignment of the camera relative to the microphone reference frame. A new calibration technique is developed here which performs all three steps in one process. The 3D locations of sound sources are obtained as part of the camera calibration process and time of flight measurements from these sources are obtained for each microphone. An iterative method is then used to obtain the microphone coordinates.

6.3.1 Calibration Rig



(a) Calibration rig.

(b) Setup at one position and orientation of calibration rig.

Figure 6.3: Photo of calibration rig used to calibrate three cameras and spherical array.

A calibration rig was laser cut from perspex plastic. Six holes were cut, through which 26mm diameter speakers were inserted. A checkerboard pattern was also engraved by the laser cutter into the surface of the perspex and used to align a printed checkerboard pattern which was glued over the top, see Figure 6.3. The coordinates of the speakers, relative to the checkerboard speaker pattern, are given in Table 6.1. Each speaker was

Speaker	X [mm]	Y [mm]	Z [mm]
1	-60	0	-2.5
2	210	0	-2.5
3	-60	240	-2.5
4	210	240	-2.5
5	-60	120	-2.5
6	210	120	-2.5

Table 6.1: Speaker coordinates in checkerboard reference frame.

connected through cables to one of the analog outputs of Data Translation DT9836 boards. These analog outputs were synchronised to the analog inputs using an external clock and

trigger. This ensured that a repeatable signal was able to be played and recorded. This calibration rig was attached to a tripod.

6.3.2 Data Acquired for Calibration

The first step in data acquisition was to obtain a reference recording for each speaker. This was obtained by placing a reference microphone about 2mm from the surface of the speaker and simultaneously playing and recording a short burst of maximum length sequence white noise. The reference signal was to be used later for time of flight calculations.

If automatic speed of sound measurements were to be made, the reference microphone was then placed on a stand in front of the array just outside the view of the camera. The corner of the checkerboard calibration rig was placed at the centre of the microphone, see Figure 6.4, and a camera image obtained of the checkerboard pattern. This image was later used to enable the 3D coordinates of a reference microphone to be obtained in the camera's reference frame.

(a) Front view. (b) Back view.

Figure 6.4: Setup for obtaining the 3D coordinates of the reference microphone in the camera's reference frame. This was used for automatic speed of sound measurements.

The checkerboard rig was then removed and setup in view of the camera. A camera image of the checkerboard was then taken. For each speaker in turn, a burst of white noise, identical to that used for the reference recording, was played and simultaneously recorded on all microphone channels. A separate recording file was created for each sound source. This process was repeated for a range of positions and orientations of the calibration rig. The file structure used for saving the microphone data were of the form '*./mic/ImageNumber/SpeakerNumber*' while for camera images it was of the form '*./cam/CameraNumber/ImageNumber*'.

6.3.3 Web Camera Calibration

The web camera calibration intrinsic parameters ($\boldsymbol{f}_c$, $\boldsymbol{c}_c$, α_c, and $\boldsymbol{\zeta}$) and extrinsic parameters ($\mathbf{R}$, $\vec{\boldsymbol{T}}$) for each image were obtained using the images of the printed checkerboard pattern and Bouguet's MATLAB *Camera Calibration Toolbox* (see Section 4.2 for more details). If more than one camera partially share the same field of view, stereo calibration was then performed. Other software such as *openCV* (Bradski and Kaehler, 2008) could have been used instead of Bouguet's toolbox.

The extrinsic parameters $\mathbf{R}$ and $\vec{\boldsymbol{T}}$ of a web camera describe its position and orientation relative to the checkerboard 'real world' reference frame. This reference frame is right handed and is defined such that one extreme inner corner is the origin, X and Y axes lie in the plane of the board and are aligned with the checker pattern such that Z pointing out of the board, see Figure 6.5.

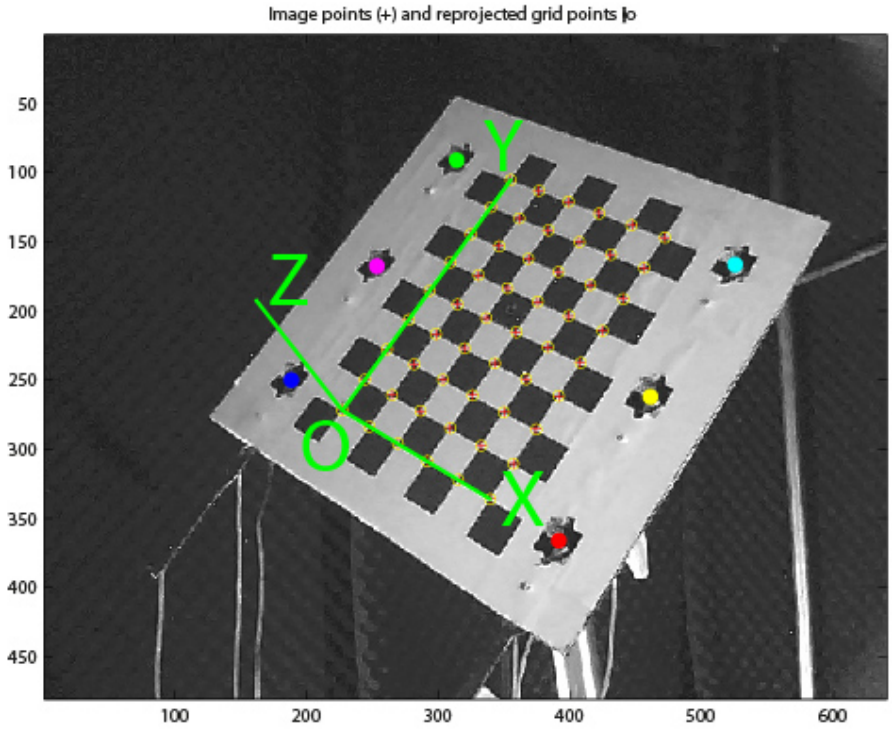
6.3.4 Obtaining 3D Coordinates of Sound Sources and Reference Microphone Using a Web Camera's Extrinsic Parameters

The 3D coordinates of the speakers in the calibration rig may be obtained in a web camera's reference frame using the extrinsic parameters obtained during web camera calibration. For each web camera image, the extrinsic parameters $\mathbf{R}$ and $\vec{\boldsymbol{T}}$ will be obtained. If the coordinates of a speaker in the checkerboard 'real world' reference frame are $\vec{\boldsymbol{X}}_{s_{\{\mathrm{RW}\mathcal{F}\}}}$, the coordinates of the speaker in the web camera reference frame will be

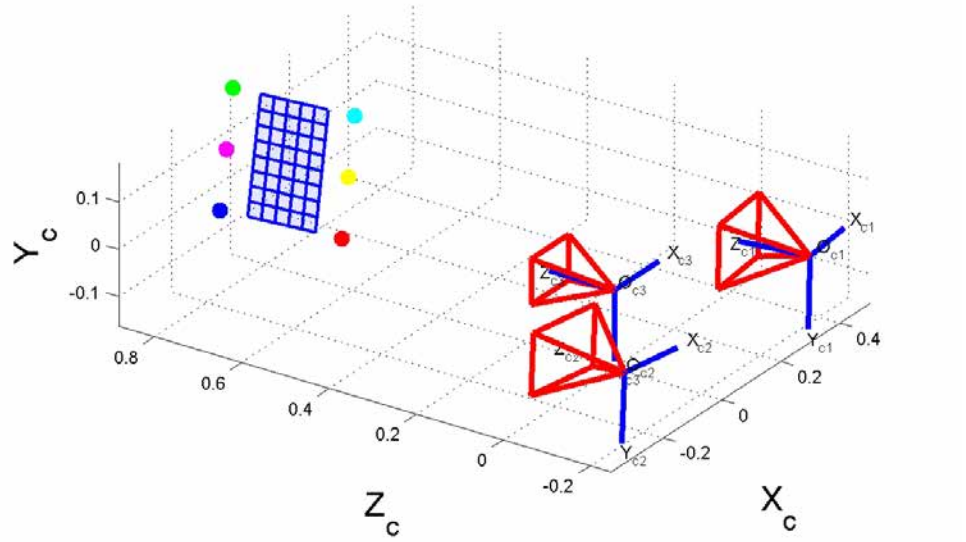
$$\vec{\boldsymbol{X}}_{s_{\{\mathrm{CAM}\mathcal{F}\}}} = \mathbf{R}\, \vec{\boldsymbol{X}}_{s_{\{\mathrm{RW}\mathcal{F}\}}} + \vec{\boldsymbol{T}}.$$

This is performed for each speaker for each web camera image to generate a point cloud of speaker coordinates, see Figure 6.6.

The same technique was used to obtain the 3D coordinates of a reference microphone. Consider that the reference microphone is setup just outside the web camera's field of view. A point, such as the corner, on the calibration rig may be placed next to the centre of the microphone. The coordinates of this point, and hence the coordinates of the microphone, are known in the checkerboard 'real world' reference frame. A web camera image of the checkerboard pattern is then captured and the extrinsic parameters for this image obtained. These parameters, combined with coordinates of the microphone in the 'real world' reference frame, enable the 3D coordinates of the microphone to be determined in the web camera's reference frame.



(a) Example photo of the calibration rig captured by the central camera in the spherical array during calibration. The extrinsic parameters obtained from the checkerboard image were used to obtain the 3D coordinates of the six speakers in the reference frame of this camera. These have been projected into the image and plotted using a different coloured dot for each speaker.



(b) Plot showing the 3D position and orientation of the checkerboard (blue grid) on the calibration rig and coordinates of the six speakers (coloured dots) relative to the three cameras in the spherical microphone array.

Figure 6.5: Detection of the 3D coordinates of speakers in calibration rig using web camera calibration data.

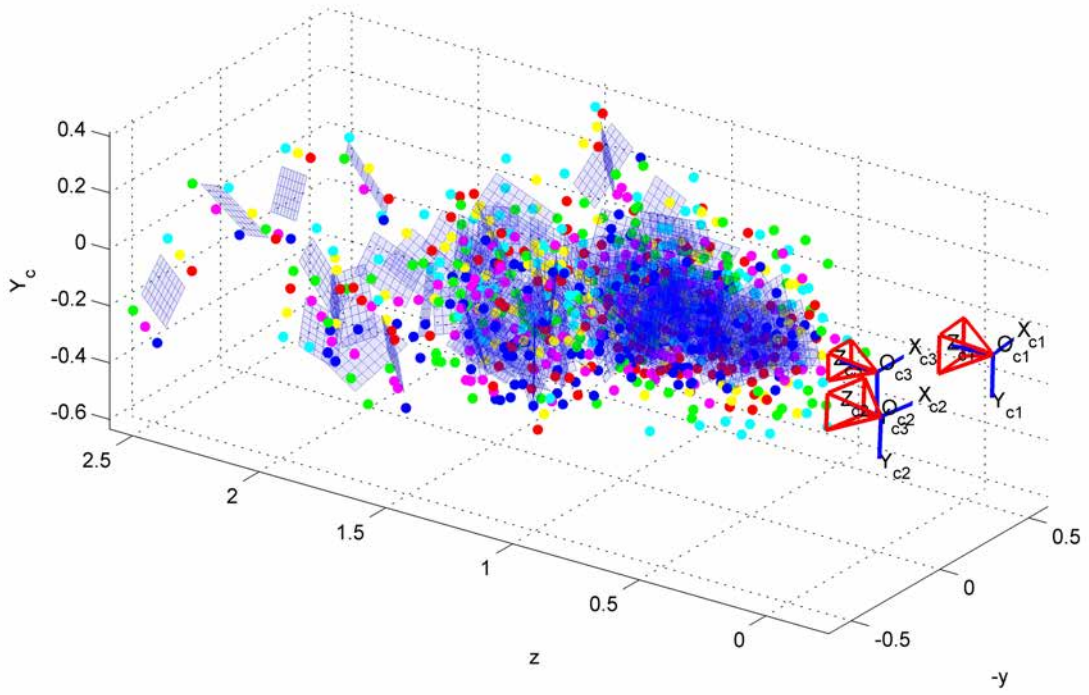


Figure 6.6: Plot showing the positions and orientations of the checkerboard patterns (blue grids) and 3D coordinates of the speakers (colour dots) in the central web camera's reference frame for multiple positions of the calibration rig. This is the same as Figure 6.5b except that data from multiple calibration rig positions and orientations have been shown.

6.3.5 Time of Flight

The time difference of arrivals of signals from two microphones may be measured from the cross correlation of the microphone signals. The General Cross Correlation (GCC) is given by

$$\mathcal{R}_{m,m'}(\tau) = \int_{-\infty}^{+\infty} \boldsymbol{\Psi}_{m,m'}(\omega)\, \boldsymbol{U}_m(\omega)\, \boldsymbol{U}^{\dagger}_{m'}(\omega)\, e^{i\omega\tau}\, d\omega \tag{6.4}$$

where $\boldsymbol{\Psi}$ is a function (Knapp and Carter, 1976) that allows filtering prior to peak detection. If $\boldsymbol{\Psi} = 1$, then the standard cross correlation function is obtained. For high reverberation, low noise environments, the Phase Transform ($PHAT$) weighting function

$$\boldsymbol{\Psi}_{m,m'}(\omega) = \frac{1}{|\boldsymbol{U}_m(\omega)\, \boldsymbol{U}^{\dagger}_{m'}(\omega)|} \tag{6.5}$$

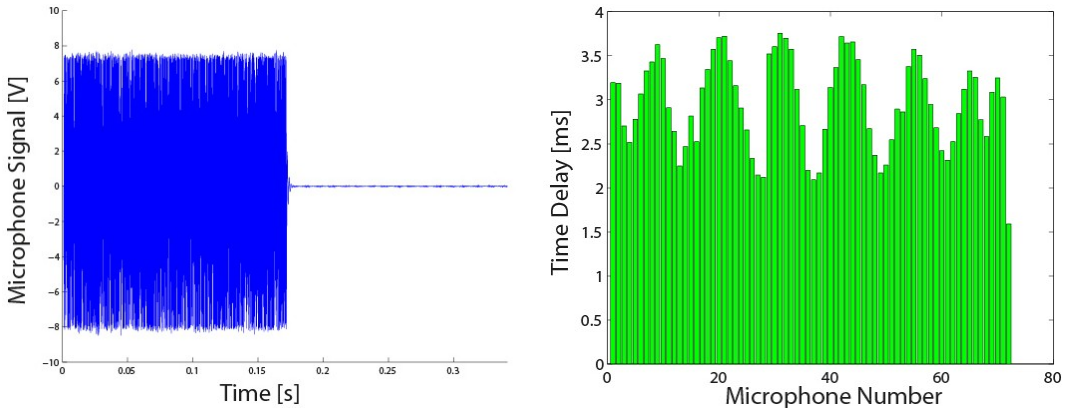
is commonly used (Knapp and Carter, 1976). Other weighting functions include ROTH (Roth, 1971), SCOT (Carter et al., 1973), and Maximum Likelihood (ML) (Brandstein

et al., 1995). The time difference of flight is obtained by obtaining the peak in the cross correlation

$$TDOF_{m,m'}^{Measured} = arg\ \max_{\tau}\ [\ \mathcal{R}_{m,m'}(\tau)\ ]\,. \tag{6.6}$$

The microphone signals will have been sampled at discrete times. This means that the best resolution able to be obtained, using Equation (6.6), will be equal to the sampling period. Methods have been developed to achieve sub-sample accuracy. For example, a function, such as a parabolic or Gaussian, can be fitted to the three points consisting of the peak and a value on each side (Jacovitti and Scarano, 1993; Zhang and Wu, 2006; Wiens and Bradley, 2009).

For the microphone position calibration, the delay between the transmitted and received signal, $TOF_{m,m\_ref}^{Measured}$, was measured using sub-sample delay estimation code written by a colleague (Wiens and Bradley, 2009; Wiens, 2009b). Wiens's iterative method fits the peak of a standard cross correlation to achieve sub-sample time delay measurements. The signal transmitted by the speaker was recording by placing a microphone about 2mm from the surface of the speaker and synchronously playing and recording a burst of white noise. This synchronisation of the analog output and input channels was achieved using external clock and trigger. This method was used, because it contained information on the amplitude and phase frequency response of the combined microphone, pre-amp, and speakers. Initial tests indicated that it gave more accurate results compared to that obtained using the signal sent to the analog output channels. An alternative approach would have been to record the voltage signal at the speaker terminals and use this as a reference signal for time delay estimation. This would have the advantage that synchronisation of the analog input and outputs would not be required. However, tests might need to be made to ensure that it gave accurate results. Figure 6.7b shows an example of time delays obtained using this method for a single sound source and 72 microphones.



(a) Reference microphone recording of Speaker 1 calibration signal.

(b) Example of time delays for microphone position calibration.

Figure 6.7: Example of a reference microphone recording and calculated time delays.

The time of flight from a sound source at $\vec{\boldsymbol{X}}_s$ to a microphone at $\vec{\boldsymbol{X}}_m$ may be given by

$$TOF_{m,s}^{\ Theory} = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\|}{c}. \tag{6.7}$$

In reality, an extra term $\Delta(\vec{\boldsymbol{X}}_m, \vec{\boldsymbol{X}}_s)$ should be added to allow for extra propagation time due to extra distance travelled by the sound wave due to the microphone facing away from the sound source or obstructions such as the array structure or cables. This effect has not been addressed in the calibration method used here.

6.3.6 Calculating the Speed of Sound Using TOF and Computer Vision

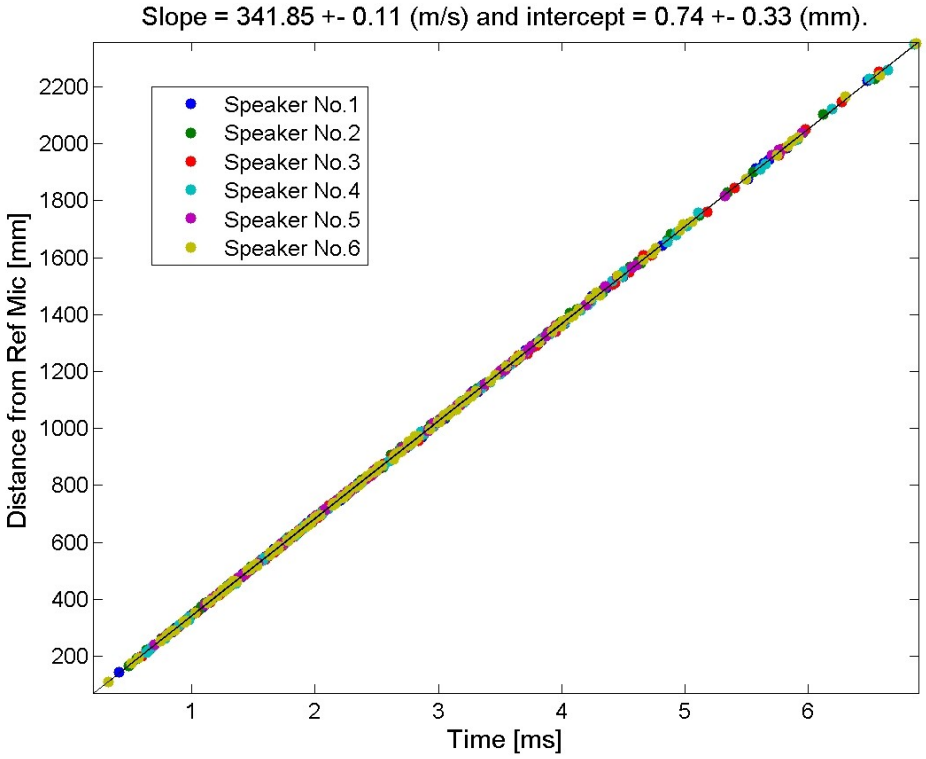


Figure 6.8: Speed of sound calculation.

The theoretical time of flight values given by Equation (6.7), rely on accurate speed of sound estimates. These may be obtained by measuring the temperature (Sachar et al., 2005). This, however, requires a thermometer with an accuracy of the order of $\pm 0.1°$C for accurate measurements. Döbler et al. (2010) suggest using measured distances between two or more microphones to enable solving for the speed of sound. Another factor that could lead to errors in Equation (6.7), is if the dimensions of the checker pattern used for

web camera calibrations were inaccurate. This could make the positions of the sources appear closer or further away than they really are.

To correct for this, an effective speed of sound is proposed. The position of a reference microphone is obtained in the web camera reference frame using the method described in Section 6.3.4. For each sound source, the Euclidean distance is calculated from the sound source coordinates to the reference microphone coordinates and the corresponding time of flight values obtained. A least squares fit is then performed to obtain the effective speed of sound. In Figure 6.8, an example of speed of sound measurements is shown. The measured effective speed of sound obtained using this method was 341.9m/s with standard deviation of 0.1m/s.

6.3.7 Maximum Likelihood Estimator

The microphone coordinates may be obtained using the theoretical and measured time of flight values. A maximum likelihood estimator may be defined as

$$\Theta_{ML} = arg\ \min_{\Theta} \left[\ J\left(\Theta, TOF^{\,Measured}\right)\ \right] \tag{6.8}$$

where

$$J\left(\Theta, TOF^{\,Measured}\right) = \sum_{m}\sum_{s} \frac{\left(TOF_{m,s}^{\,Theory}(\Theta) - TOF_{m,s}^{\,Measured}\right)^2}{\varphi_{m,s}^2} \tag{6.9}$$

is the cost function to be minimised, Θ are the suite of parameters to be estimated and $\varphi_{m,s}$ is a weighting value related to the noise (Raykar and Duraiswami, 2004). The parameters to be estimated are microphone coordinates and fine tuned sound source coordinates. If an initial estimate of the microphone coordinates is known, then the rotation and translation vector between the web camera and microphone reference frame also need to be obtained. To allow for this, the microphone coordinates in the web camera reference frame are defined as

$$\vec{\boldsymbol{X}}_{m_{\{\mathrm{Cam}\mathcal{F}\}}} = \mathbf{R}_m\ \vec{\boldsymbol{X}}_{m_{\{\mathrm{Mic}\mathcal{F}\}}} + \vec{\boldsymbol{T}}_m.$$

$\mathbf{R}_m$ is calculated from a $[3 \times 1]$ rotation vector $\vec{\boldsymbol{\Omega}}_m$, using Equation (4.19). The suite of parameters Θ to be solved for are $\vec{\boldsymbol{X}}_m$, $\vec{\boldsymbol{T}}_m$, $\vec{\boldsymbol{\Omega}}_m$, and refined values of $\vec{\boldsymbol{X}}_s$.

6.3.8 Algorithm Used to Find Microphone Coordinates

An iterative method is used to solve for the unknown parameters by the minimisation in Equation (6.8). This method is based on the method described by Döbler et al. (2010). The entire process may be described by

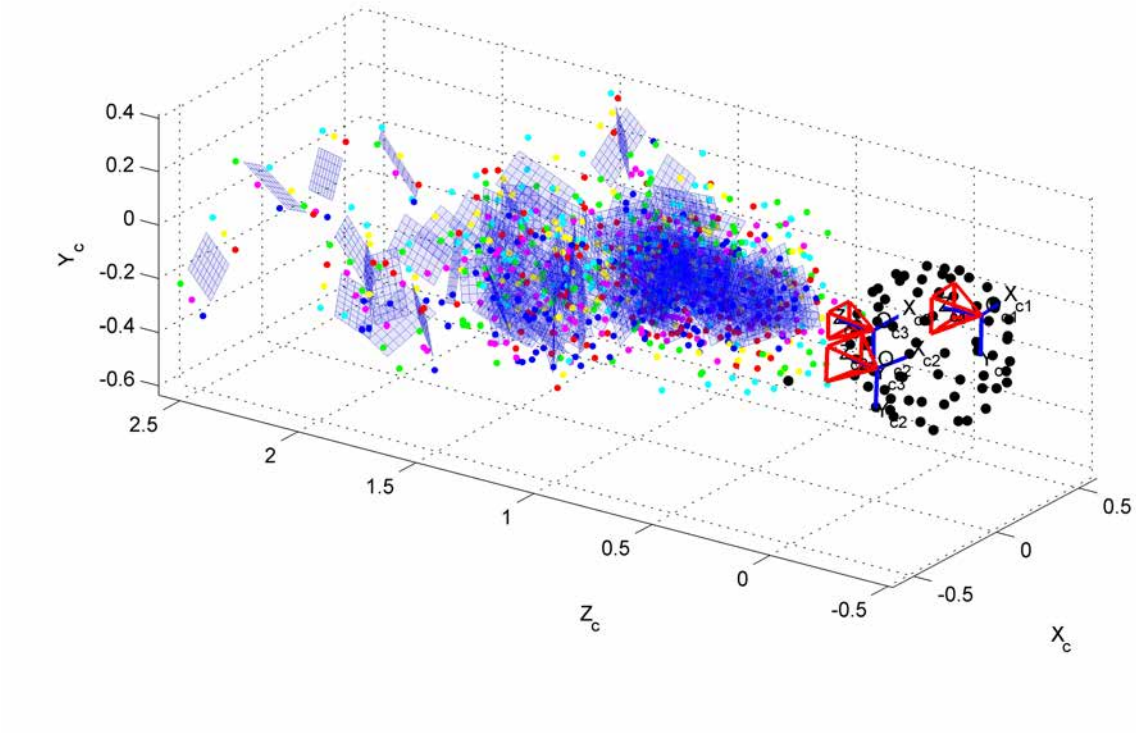
1. The sound source 3D coordinates are obtained in a web camera's reference frame as described in Section 6.3.4.
2. The effective speed of sound is obtained as described in Section 6.3.6.
3. Initialise $\mathbf{R}_m$ and $\vec{\mathbf{\Omega}}_m$ to [0;0;0];
4. Initialise microphone coordinates.
   (a) If approximate microphone coordinates are known, these are used.
      i. Adjust $\vec{\mathbf{\Omega}}_m$ until a minimum of $J(\Theta)$ is reached.
      ii. adjust $\vec{\boldsymbol{T}}_m$ until a minimum of $J(\Theta)$ is reached.
      iii. Repeat (i) to (ii) with decreasing step sizes until exit condition met.
   (b) Else, if no *a priori* knowledge of the microphone positions exist, initialise microphone coordinates using random numbers.
5. Adjust $\vec{\boldsymbol{X}}_m$ for each microphone in turn until a minimum of $J(\Theta)$ is reached.
6. Repeat 5 with smaller step sizes until exit condition is reached.
7. Adjust $\vec{\boldsymbol{X}}_s$ for each sound source until a minimum of $J(\Theta)$ is reached.
8. Adjust weighting of $\varphi_{m,s}$ to decrease effect of sound sources causing most contribution to $J(\Theta)$.
9. Repeat 5 to 8 until exit condition is met.

To find an optimal value of a parameter, its value is adjusted until a minimum value in Equation (6.9) is achieved. For example, the translation vector $\vec{\boldsymbol{T}}_m$ may be adjusted by:

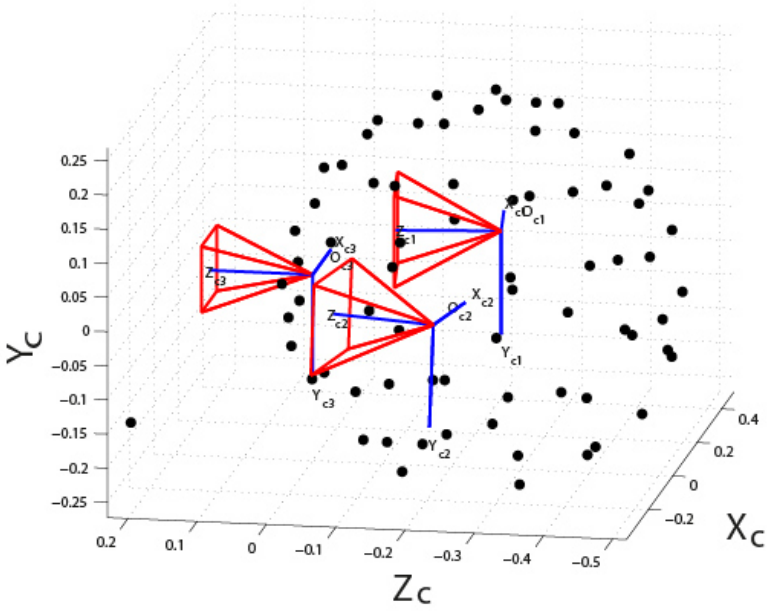
Define a step size δ, $\boldsymbol{sign} = 1$, and component $index = 1$.

1. Reduce step size δ.
2. Change component $index$.
3. Use
$$\vec{\boldsymbol{T}}_m(index) = \vec{\boldsymbol{T}}_m(index) + \boldsymbol{sign} \times \delta \tag{6.10}$$
until $J(\Theta)$ increases, in which case reject the latest value of $\vec{\boldsymbol{T}}_m$. If $J(\Theta)$ increased on the first iteration of 3, try $\boldsymbol{sign} = \boldsymbol{sign} \times -1$.
4. Repeat 2 to 3 (*minimised error in next component*) until $J(\Theta)$ increases on consecutive iterations for all components of $\vec{\boldsymbol{T}}$.
5. Repeat 1 to 4 until exit condition is reached.

The iterative method used here was based on that used by Döbler et al. (2010).



(a) Plot showing the microphone coordinates (black dots) relative to the three web cameras in the spherical array and their position and orientation relative to the calibration rig checkerboard patterns (blue grids) and the coordinates of speakers (coloured dots) used for the calibration.



(b) Microphone coordinates relative to the cameras in the array.

Figure 6.9: Microphone and web camera positions obtained during calibration.

6.3.9 Measured Error

Figure 6.9 shows the microphone coordinates obtained using this calibration method for the spherical array. A mean Euclidean distance error of 4mm was obtained relative to Faro-Arm measurements with a maximum value of 14mm at the back of the array behind the web camera. Differences could be partially due to extra time of flight related to microphone direction and array structure obstruction. There may also be a rotation between the Faro-Arm and camera reference frames not properly corrected for during the above error calculation. Also, there will be some error in the Faro-Arm measurements. There is some parallax error associated by the angle of the probe point (a 6mm sphere at the end of the robotic arm) contacting the surface of the microphone. For microphones at the front, one would expect to get little error since the probe pointed straight down onto the microphone. For the back, however, the robotic arm could only measure the microphones with the probe at a relatively high angle to the microphone. The parallax error, therefore, might be expected to be negligible at the front and to increase to several mm at the back of the array. In addition to this error, measurements of three points on the array were repeated a second time. These gave position errors of between 1 and 3.5mm compared to the previous measurements. It is possible that some movement of the array may have occurred during the measurements despite efforts to mount it securely to the measurement table. Additional movement of the microphones could also have occurred during transport and experimental setup. Although a maximum position error of 14mm was obtained for microphone coordinates by the calibration technique compared to the Faro-Arm measurements, this does not necessarily mean an error for beamforming since it may be correcting for the extra travel path caused by sound having to pass around the edge of the microphones plus the array structure. If this was the case, the Faro-Arm measurements would actually be causing errors for beamforming since they assume free field propagation. It Section 7.5.4, a mean position accuracy from 5 to 15kHz was 3.7mm when using 3D scanning surfaces. This is remarkably accurate considering that the speaker diameter was 26mm and the scanning surface spacing was 6mm. This indicates that the combined calibration methods were providing accurate results.

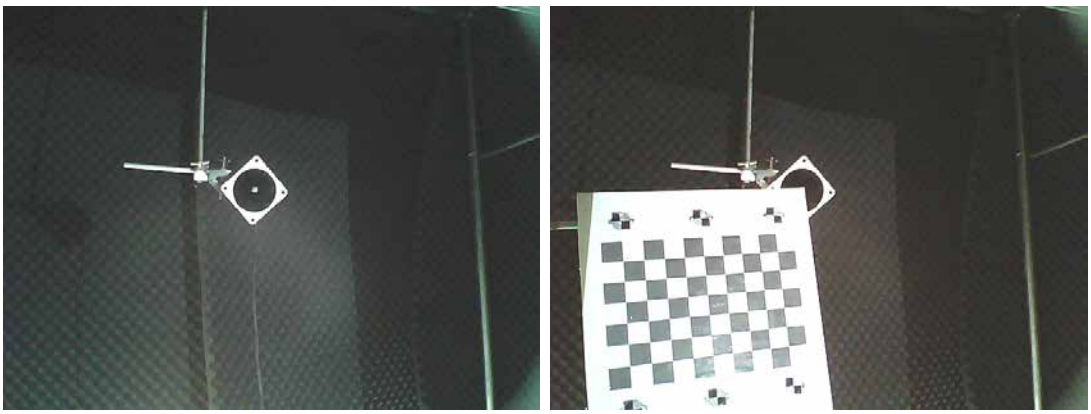
A better estimation of the error would be achieved using a planar or linear array, where more accurately determined microphones positions should be able to be a achieved since one can use a milling machine or a laser cutter. If the array is made of a rigid material, it should maintain its shape better than a lattice type structure. Also, since the microphones all point in the same direction there is no refraction distance to be accounted for in the travel times as is the case if microphones face in different directions. Future work should investigate the accuracy of microphone coordinates as a function of number of sound sources used and how many are needed to give accurate results.

6.3.10 Possible Variations on Method Used

A reference microphone could have been placed in the calibration rig if synchronisation of the analog input and output hardware was not an option. Its coordinates could then be obtained in the same way the speaker coordinates were obtained. $TDOF$ would then be used for this microphone instead of TOF using Equation (6.2). A least squares method could also be investigated instead of the iterative method.

An alternative method of finding the 3D position of a speaker would be to attach a LED (or covered laser pointer) to the speaker, detecting the LED in the web camera's images and using triangulation. By moving the speaker to various positions, the 3D coordinates of multiple sound sources could be determined. Multiple web camera calibration techniques have been developed which use similar methods. This would have the advantage of not requiring a checkerboard calibration rig, but would require more than one web camera to be used and synchronisation of the web cameras and audio acquisition.

6.3.11 Eigenvalue Calibration



(a) (b)

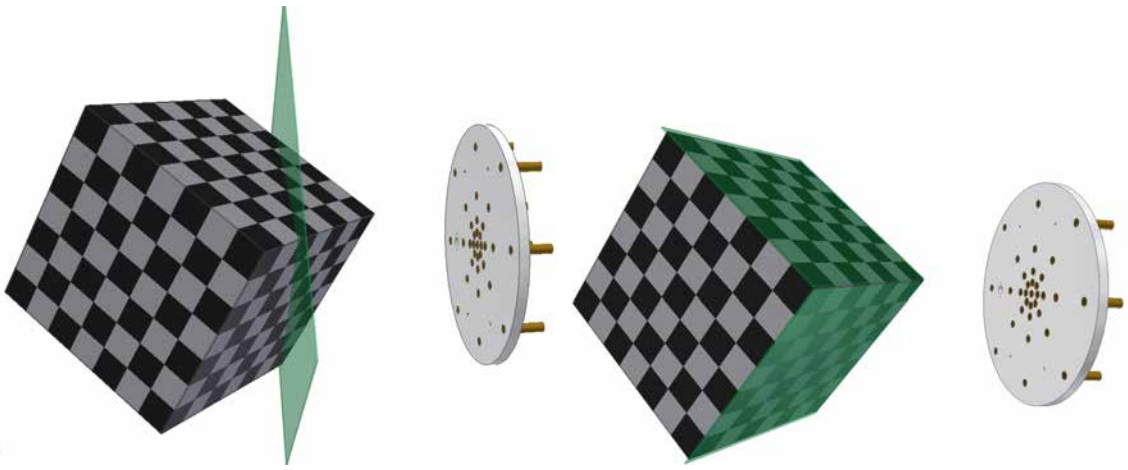
Figure 6.10: Eigenvalue microphone calibration speaker position images.

The microphone position calibration technique presented in this work was designed to obtain the position of the microphones in the camera reference frame and optionally the speed of sound. Since the acoustic signals used for time of arrival measurements were broadband white noise, the calibration method also includes broadband group phase corrections. However, in its current form, it does not include corrections for variation in phase between different frequencies. This was intended to be provided by using the eigenvalue calibration technique developed by Dougherty (2002), see Section 2.2.2. This eigenvalue calibration technique requires accurate knowledge of the position of the microphones and

the speaker location and does not include camera alignment. This information was provided by the calibration technique discussed in this chapter. The sound source location was obtained using an image of a checkerboard, whose corner was temporarily placed over the centre of the tweeter in free space in front of the array, see Figure 6.10. This was the same method used in Section 6.3.4 for obtaining the 3D position of the reference microphone. An alternative method might be to used the white noise recordings, used for time of flight calculations, to obtain an average calibration matrix. A problem with this method, however, could be any scattering effects due to the calibration rig structure.

7

3D Beamforming and Deconvolution



(a) Traditional 2D scanning surface.

(b) A 3D scanning surface corresponding to the surface of the object.

Figure 7.1: Illustration of 2D and 3D scanning surfaces and a SADA (Humphreys et al., 1998) microphone array.

7.1 Beamforming and Deconvolution Using 2D and 3D Scanning Surfaces

7.1.1 Potential Phase and Spherical Spreading Correction Errors Associated with the Use of 2D Scanning Surfaces

The position and magnitude of a sound source is usually obtained from an acoustic map by measuring the position and magnitude of a peak in the map. A position error will occur if the coordinate of the peak $\vec{\boldsymbol{\xi}}_{max}$ is different from the scan point coordinate $\vec{\boldsymbol{\xi}}_s$ corresponding to the sound source location $\vec{\boldsymbol{X}}_s$. A magnitude error will occur if the peak magnitude does not correspond to the sound pressure level of the source. Different factors such as channel noise or interference of the sound source's beam pattern with the beampattern of other sound sources can result in errors. Such effects can be particularly problematic at long wavelengths, where the beamwidths are large. Even in the absence of other noise sources, errors in the magnitude and location of sound sources can occur due to incorrect phase shifts and spherical spreading corrections being used during beamforming. This section looks at the potential errors that can be caused by using 2D scanning surface which is offset from the sound source location.

Delay and Sum Beamforming

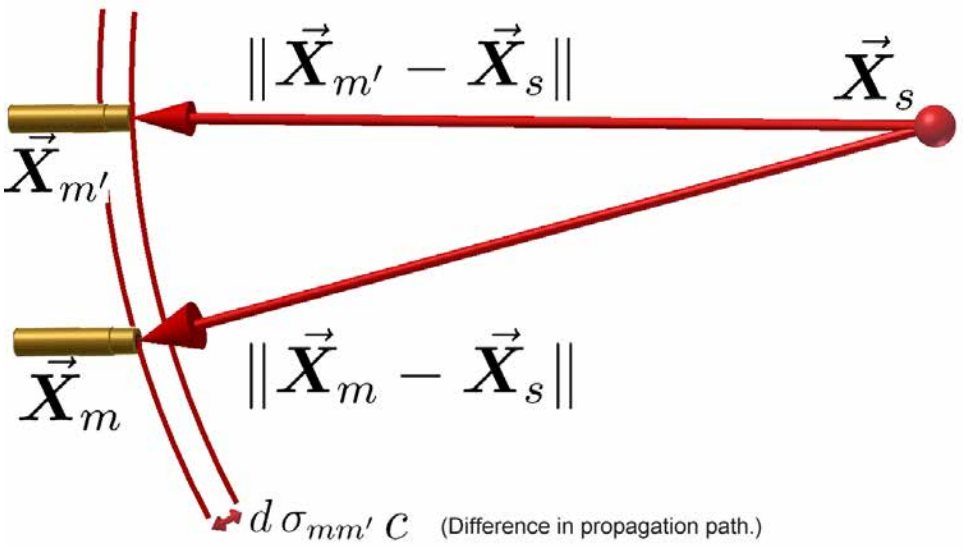


Figure 7.2: Diagram illustrating the difference in propagation time $d\,\sigma_{mm'}(\vec{\boldsymbol{X}}_s)$ for a sound wave propagating from a sound source at $\vec{\boldsymbol{X}}_s$ to two microphones located at $\vec{\boldsymbol{X}}_m$ and $\vec{\boldsymbol{X}}_{m'}$.

Consider sound waves propagating under free-field conditions from a sound source located at position $\vec{\boldsymbol{X}}_s$ to two microphones located respectively at positions $\vec{\boldsymbol{X}}_m$ and $\vec{\boldsymbol{X}}_{m'}$ in an array containing M microphones. The propagation time from the sound source to microphone m is

$$\sigma_m(\vec{\boldsymbol{X}}_s) = \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\|/c,$$

see Figure 7.2. The difference in propagation time between the two microphones will, therefore, be

$$d\,\sigma_{mm'}(\vec{\boldsymbol{X}}_s) = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{X}}_s\|}{c} \tag{7.1}$$

Now consider that delay and sum beamforming is performed using the signals from the

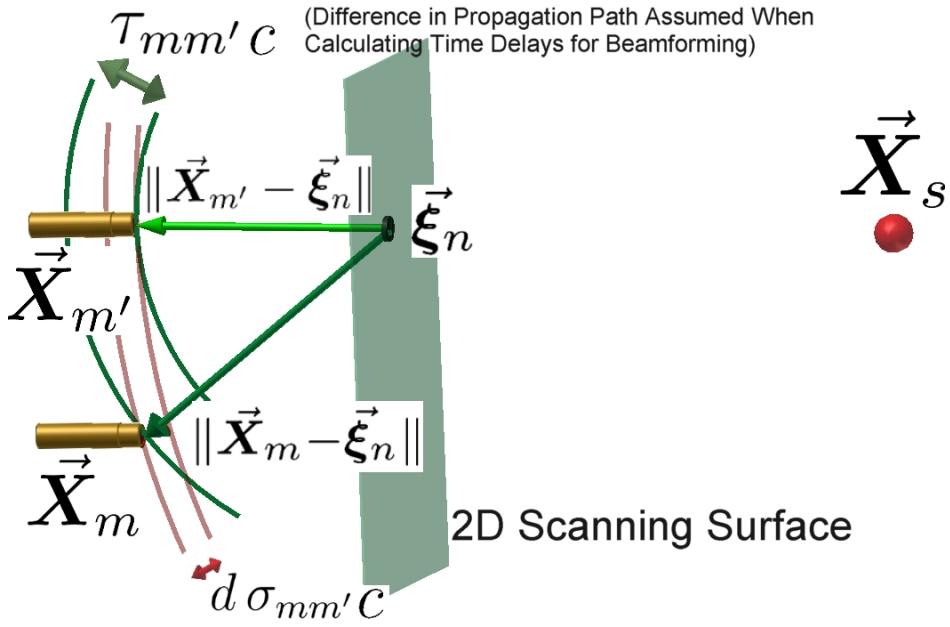


Figure 7.3: Diagram illustrating the time delay $\tau_{mm'}(\vec{\xi}_n)$ that might be used for beamforming for a scan point located at $\vec{\xi}_n$. A difference between the propagation time (red curves) and the beamforming time delay can occur if a 2D scanning surface is used for beamforming which does not pass through the sound source locations. This factor can lead to magnitude and position error of sound sources.

two microphones and a 2D scanning surface. The time delays used for beamforming may be calculated using

$$\tau_m(\vec{\boldsymbol{\xi}}_n) = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|}{c},$$

where $\vec{\boldsymbol{\xi}}_n$ is the coordinate of the n^{th} scan point, refer to Figure 7.3. The difference between the time delays used for each microphone signals would, therefore, be

$$d\,\tau_{mm'}(\vec{\boldsymbol{\xi}}_n) = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{\xi}}_n\|}{c}. \tag{7.2}$$

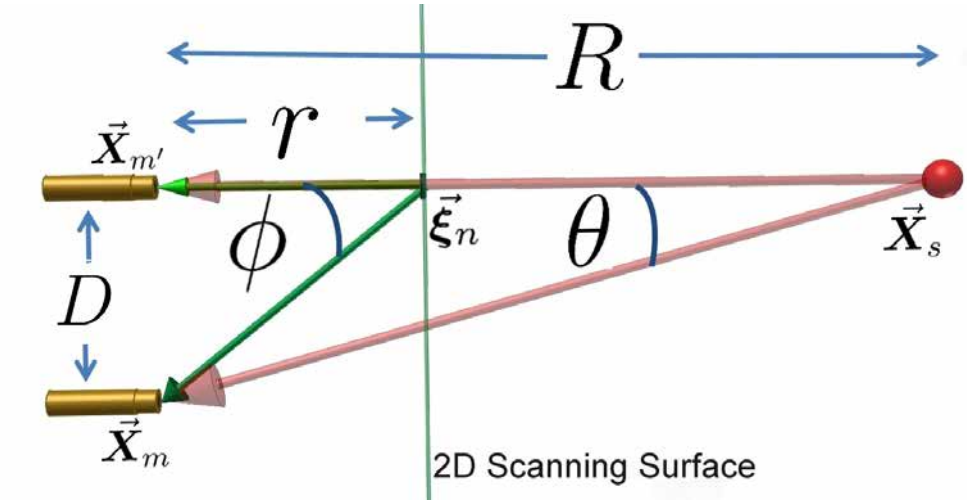


Figure 7.4: Simplified geometry for two microphones in a planar array to enable a general/simplified phase error equation to be obtained.

The relative phase of two delayed and summed signals may be described by

$$\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\boldsymbol{\xi}}_n) = \frac{2\pi}{\lambda} \left[\left(d\, \sigma_{mm'}(\vec{\boldsymbol{X}}_s) - d\, \tau_{mm'}(\vec{\boldsymbol{\xi}}_s) \right) c \right]. \tag{7.3}$$

Using Equations (7.1) and (7.2), this phase difference may be expressed as

$$\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\boldsymbol{\xi}}_n) \tag{7.4}$$

$$= \frac{2\pi}{\lambda} \left[\left(\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{X}}_s\| \right) - \left(\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{\xi}}_n\| \right) \right] \tag{7.5}$$

$$= \frac{2\pi}{\lambda} \left[\delta_m(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s) - \delta_{m'}(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s) \right], \tag{7.6}$$

where $\delta_m(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s) = \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|$. Constructive interference will occur during beamforming if

$$\delta_m(\vec{\boldsymbol{\xi}}_s,\, \vec{\boldsymbol{X}}_s) - \delta_{m'}(\vec{\boldsymbol{\xi}}_s,\, \vec{\boldsymbol{X}}_s) \ll \frac{\lambda}{2}, \tag{7.7}$$

for all combinations of microphones indices m and m'.

Figure 7.4 shows a simplified geometry used here to obtain a generalized equation relating the geometry of the microphones, sound source, and scanning surface to the phase error obtained during beamforming. The microphones are assumed here to lie in the plane of the X- Y axes. The microphone with coordinate $\vec{\boldsymbol{X}}_{m'}$ is located at the origin while the microphone with coordinate $\vec{\boldsymbol{X}}_m$ is located a distance D from $\vec{\boldsymbol{X}}_{m'}$. The sound source is located on the Z axis at a distance R from the array. A 2D scanning surface is located parallel to the array at distance z from the array with the scan point $\vec{\boldsymbol{\xi}}_n$ which is be

evaluated here for phase error is located on the Z axis.

$$\delta_m(\vec{\boldsymbol{\xi}}_s,\, \vec{\boldsymbol{X}}_s) = \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{\xi}}_s\| = R - r \tag{7.8}$$

$$\delta_{m'}(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s) = \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_s\| = R\,\cos\theta - z\,\cos\phi \quad \text{where} \tag{7.9}$$

$$\theta = \tan^{-1}\left(\frac{D}{R}\right) \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{D}{r}\right). \tag{7.10}$$

Therefore,

$$\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\boldsymbol{\xi}}_n) = \frac{2\pi}{\lambda}\left[\delta_m(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s) - \delta_{m'}(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s)\right] = \frac{2\pi}{\lambda}\left[(R-r) - (R\,\cos\theta - r\,\cos\phi)\right] \tag{7.11}$$

The Taylor expansions for $\tan^{-1}\theta$ is $\theta - \theta^3/3 + \theta^5/5\cdots$ while that for $\cos\theta$ is $1 - \theta^2/2 + \theta^4/4!\cdots$. If the separation of the microphones D is small compared to the distances R and r, then $\tan^{-1}(\theta) \approx \theta$ and $\tan^{-1}(\phi) \approx \phi$. Using this assumption,

$$\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\boldsymbol{\xi}}_n) \approx \frac{2\pi}{\lambda}\left[(R-r) - \left\{R\left(1 - \frac{D^2}{2\,R^2}\right) - r\left(1 - \frac{D^2}{2r^2}\right)\right\}\right] \tag{7.12}$$

$$= \frac{\pi D^2}{\lambda}\left[\frac{1}{R} - \frac{1}{r}\right]. \tag{7.13}$$

The phase difference between the two microphone signals, after the beamforming delay has been applied, increases as

- the wavelength λ reduces,
- the distances R and r reduces,
- the offset $dr = |R - r|$ of the 2D scanning surface from the sound source location increases, and
- the separation D of the microphones increases.

It will also depend on the relative positioning of the microphones, as will be discussed in more detail later. Note that Equation (7.13) has a similar form to the inverse of the beamwidth given in Equation (2.27)

$$BW_R \propto \frac{R\,\lambda}{D}.$$

Frequency Domain Beamforming

The same results may also be obtained starting from the frequency domain beamforming equation given in Equation (2.8)

$$\boldsymbol{b}_{FD}(\vec{\boldsymbol{\xi}}_n) = \left\langle \left| \ \mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_n)\ \boldsymbol{U}(j) \right|^2 \right\rangle, \quad \vec{\boldsymbol{\xi}}_n = \vec{\boldsymbol{\xi}}_1 \ldots \vec{\boldsymbol{\xi}}_N,$$

where $\boldsymbol{U}(j)$ is the j^{th} segment of the frequency domain signal arriving at the array microphones and $\langle\ \rangle$ is the average over J data blocks, see Equation (2.10). For a single sound source located at $\vec{\boldsymbol{X}}_s$, the frequency domain signal $\boldsymbol{U}(\vec{\boldsymbol{X}}_s, j)$ may be modelled, using Equation (2.2), as

$$\boldsymbol{U}_m(\vec{\boldsymbol{X}}_s, j) = C_m(\vec{\boldsymbol{X}}_s)\ Q(\vec{\boldsymbol{X}}_s, j),$$

where $Q(\vec{\boldsymbol{X}}_s, j)$ is the source strength and $C_m(\vec{\boldsymbol{X}}_s)$ is the array propagation vector. The frequency domain beamforming output for the n^{th} scan point due to a sound source located at $\vec{\boldsymbol{X}}_s$ may, therefore, be modelled as

$$\boldsymbol{b}_{FD}(\vec{\boldsymbol{\xi}}_n) = \left| \ \mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_n)\ \boldsymbol{C}(\vec{\boldsymbol{X}}_s)\ \left\langle Q(\vec{\boldsymbol{X}}_s, j) \right\rangle \right|^2, \quad \vec{\boldsymbol{\xi}}_n = \vec{\boldsymbol{\xi}}_1 \ldots \vec{\boldsymbol{\xi}}_N. \tag{7.14}$$

Lets assume that $\mathrm{w}_m^{\dagger}(\vec{\boldsymbol{\xi}}_n)\ C_m(\vec{\boldsymbol{X}}_s)$ can be written in the form

$$\mathrm{w}_m^{\dagger}(\vec{\boldsymbol{\xi}}_n)\ C_m(\vec{\boldsymbol{X}}_s) = \beta_m(\vec{\boldsymbol{\xi}}_n,\ \vec{\boldsymbol{X}}_s)\ e^{-i\,2\pi\ \delta_m(\vec{\boldsymbol{\xi}}_n,\ \vec{\boldsymbol{X}}_s)/\lambda} \tag{7.15}$$

where, as above, $\delta_m(\vec{\boldsymbol{\xi}}_n,\ \vec{\boldsymbol{X}}_s) = \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_n\|$. Beamforming aims to generate an acoustic map which is proportional to the sound source distribution by using steering vectors such that, for a scan point $\vec{\boldsymbol{\xi}}_s$ corresponding to a sound source location $\vec{\boldsymbol{X}}_s$, $\mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_s)\ \boldsymbol{C}(\vec{\boldsymbol{X}}_s)$ produces a matched output. To measure the sound pressure level relative to the source location, for scan point coordinate $\vec{\boldsymbol{\xi}}_s$ corresponding to $\vec{\boldsymbol{X}}_s$, the steering vector should be chosen such that

$$\mathbf{w}^{\dagger}(\vec{\boldsymbol{\xi}}_s)\ \boldsymbol{C}(\vec{\boldsymbol{X}}_s) = 1.$$

As stated above, the correct amount of phase shifting, leading to constructive interference, will occur if

$$\delta_m(\vec{\boldsymbol{\xi}}_s,\ \vec{\boldsymbol{X}}_s) - \delta_{m'}(\vec{\boldsymbol{\xi}}_s,\ \vec{\boldsymbol{X}}_s) \ll \frac{\lambda}{2}, \tag{7.16}$$

for all combinations of microphones indices m and m'. The right amount of spherical spreading correction, giving the sound pressure level relative to the source location, will

(a) (b)

Figure 7.5: Diagram illustrating the difference in wavefront curvature for two sound sources separated by 50mm when they were at different distances from the array. Near the array distinct differences exist between the wavefronts arriving at the array from the two sound sources. However, as the sound sources are moved away from the array, the wavefronts become very similar. It is only by going to a larger array that differences in wavefront arrivals between microphones could be detected for the two sound sources.

be used if

$$\sum_{m=1}^{M} \beta_m(\vec{\boldsymbol{\xi}}_s,\, \vec{\boldsymbol{X}}_s) \approx 1. \tag{7.17}$$

Thus the factor by which the measured intensity of a sound source is reduced due to an offset of a scanning surface from the true sound source location may be obtained by

$$\varepsilon(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s) = 1 - \left| \sum_{m=1}^{M} \beta_m(\vec{\boldsymbol{\xi}}_n,\, \vec{\boldsymbol{X}}_s)\; e^{i\,2\pi\,\delta_m(\vec{\boldsymbol{\xi}}_n, \vec{\boldsymbol{X}}_s)/\lambda} \right|^2, \quad \vec{\boldsymbol{\xi}}_n = \vec{\boldsymbol{\xi}}_1 \ldots \vec{\boldsymbol{\xi}}_N \tag{7.18}$$

7.1.2 Scanning Surface Corresponding to the Real 3D Surface Geometry of the Object

Consider that a scanning surface is used which corresponds to the surface geometry of the object emitting sound. Assuming that the sound source location is on the surface of the object, there should exist some scan point such that $\vec{\boldsymbol{\xi}}_s \approx \vec{\boldsymbol{X}}_s$. In this case, if one assumes free-field propagation vectors of the forms given in Equation (2.4) then constructive interference of the shifted microphone coordinates should occur for steering vectors of the form given in either Equations (2.15) and (2.17) because

$$\delta_m(\vec{\boldsymbol{\xi}}_s = \vec{\boldsymbol{X}}_s,\, \vec{\boldsymbol{X}}_s) = 0$$

for all microphone pairs. If a normalised steering vector such as given in Equation (2.15) is used for beamforming, the measured sound pressure level of a source will depend on the distance of the sound source from the object. To recover the correct sound pressure

level, relative to the source location, a distance weighted steering vector such as given in Equation (2.17) could be used since this allows

$$\sum_{m=1}^{M} \beta_m(\vec{\boldsymbol{\xi}}_s = \vec{\boldsymbol{X}}_s, \vec{\boldsymbol{X}}_s) \approx 1.$$

However, caution should used with such steering vectors since they can cause distortion of the beam pattern leading to position errors of sound sources. These errors will be most prominent if there are high variations in the distance of scan points from the array and at low frequencies where the beamwidth becomes large.

7.1.3 Traditional 2D Scanning Surface

Traditionally beamforming and deconvolution have used scan points $\vec{\boldsymbol{\xi}}_n$ located on a 2D scanning surface orientated perpendicular to the array axis, see Figure 7.1a. Few if any of these scan points may correspond to the surface of an object. To enable sound source locations to be identified, some form of projection is required to relate the 2D acoustic map to the 3D object.

If $\vec{\boldsymbol{\xi}}_n \neq \vec{\boldsymbol{X}}_s$ for all scan points, constructive interference during beamforming may still occur if the condition defined in Equation (7.7) holds. This feature is dependent on the geometry of the microphones in the array, the position of the sound source relative to the array, the displacement of the scanning surface from the true sound source location, and the wavelength. For example, consider that a circular array with is used to image a sound source located on the Z axis. The distance $\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\|$ will be the same for each microphone and hence the signal will arrive at all the microphones at the same time. A 2D scanning surface located at any distance from the array would result in constructive interference for scan points $\vec{\boldsymbol{\xi}}_s$ along the Z axis since $\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{\xi}}_s\| - \|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\|$ is the same for all microphones and hence $\delta_m(\vec{\boldsymbol{\xi}}_s, \vec{\boldsymbol{X}}_s) - \delta'_m(\vec{\boldsymbol{\xi}}_s, \vec{\boldsymbol{X}}_s) = 0$, for all microphone pairs. On the other hand, if a spiral array is used to image a sound source along the Z axis, the distance $\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\|$ will be different for each microphone and hence the signal will arrive at each microphone as a slightly different time. This means that unless a plane is used which passes through the sound source location $\delta_m(\vec{\boldsymbol{\xi}}_s, \vec{\boldsymbol{X}}_s) - \delta'_m(\vec{\boldsymbol{\xi}}_s, \vec{\boldsymbol{X}}_s) \neq 0$, for $\vec{\boldsymbol{\xi}}_s$ on the Z axis. However, the smaller the offset dr of the 2D scanning surface from the sound source location, the further distance R the sound source is from the array, the smaller the diameter D of the array, and the larger the wavelength λ the more likely that $\delta_m(\vec{\boldsymbol{\xi}}_s, \vec{\boldsymbol{X}}_s) - \delta'_m(\vec{\boldsymbol{\xi}}_s, \vec{\boldsymbol{X}}_s) << \lambda/2$, and hence more likely that constructive interference will occur, see Figure 7.5 and Equation (7.13). If $\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\boldsymbol{\xi}}_n) \approx 0$, constructive interference will occur. If this is not the case, errors in the magnitude and position of a sound

source, obtained during beamforming will occur. These errors in the beamformed map will be passed onto the deconvolution map. This chapter presents experimental work which investigates the error in the magnitude and position of sound sources obtained using beamforming and *CLEAN-SC* when using 2D and 3D scanning surfaces.

7.2 Beamforming and *CLEAN-SC* Methodology

Software was written, using *MATLAB*, to generate white noise from speakers, record microphone data and capture camera images. A modified version of the *MATLAB* code, written by Lanman and Taubin (2009a), was also incorporated into the code for structured light calibration and scanning. All raw data were saved to files and later processed. Software was also written in *MATLAB* to perform octave frequency band calibration, beamforming, and *CLEAN-SC* for both 2D and 3D scanning surfaces.

The sound sources used were 25mm diameter speakers. These speakers could be driven using four separate microcontroller circuits which had separate clocks and on/off buttons. These repeatedly played uncorrelated white noise, at 48kSPS, using different sequences in a Kasami set. Alternatively the speakers could be unplugged from the microcontrollers and plugged into the six analog outputs of the Data Translation DT9836 boards. These were used to generate uncorrelated bursts of white noise, also using different sequences in a Kasami set (Wiens, 2009a), and using a sampling rate of 48kSPS.

The acoustic signals arriving at the 72 microphones in an array were simultaneously sampled using the analog inputs of the three DT9836 and six DT9816 Data Translation boards using a sampling rate of 96 kSPS per channel, with a resolution of 16 bits. Synchronisation between the boards was achieved using a clock output from one of the DT9836 boards as an external clock and hardware trigger for all boards. This clock and trigger was also used to synchronise the analog outputs of the DT9836 boards. The microphone data were saved to file.

The sensitivity η of each microphone channel was different. To correct for this, the data for each microphone channel were divided by the inverse of the microphone sensitivity factor in units of V/Pa. This microphone sensitivity factor had been obtained by placing a speaker near to the surface of the microphone, using a calibration rig, playing white noise on the speaker, and obtaining the rms (root mean square) value of the microphone voltage. The sound pressure level of the transmitted sound was obtained using a sound pressure level meter. No frequency domain calibration was performed. The *WM-61A* microphones have a relatively flat frequency response. Future work, however, would ideally include

the use of a calibrated microphone to obtain the frequency response of the speakers and microphones.

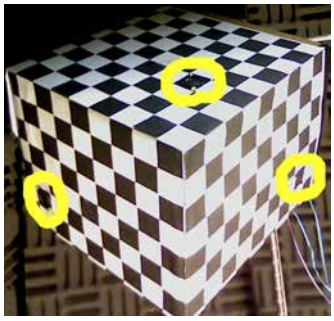
These microphone data were then used to obtain cross spectral matrices using Equation (2.12). A data block length K of 4096 samples, an overlap of data blocks of fifty percent and a Hann window were used. The cross spectral matrices were then summed into third or twelfth octave frequency bands, calculated using Equation (2.24). In later experiments, just before the main recording was made, a recording had been made of background noise. This was used to generate background noise CSM which was then subtracted from the main recording CSM. This was done to try to reduce the effect of low frequency traffic noise which came in through the windows. The effectiveness of this was not investigated.

These octave band cross spectral matrices were then calibrated using the eigenvalue calibration method described in Section 2.2.2. In later experiments, the position of the calibration speaker was obtained using an image of a checkerboard, captured by the camera in the array, see Figure 6.10, and using computer vision software. The white noise was generated by one of the microcontrollers, amplified by a Crown power amplifier and played on a *KSN* tweeter speaker or on a *Visaton SC 10N 8*Ω speaker.

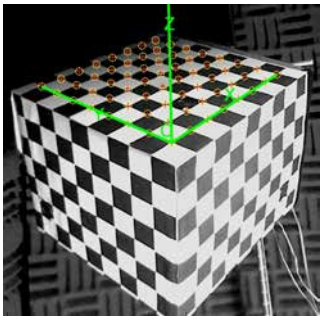
These cross spectral matrices were used to generate beamformed acoustic maps in each octave frequency band using Equation (2.11). The array steering vectors were initially calculated using Equation (2.15). However, since they are normalised, they apply the same gain to each scan point and do not correct for spherical spreading. This resulted in the measured sound source magnitudes obtained using these steering vectors being strongly dependant on distance of sound source from the array. Therefore, Equation (2.17) was also used. This is a modified version the the steering vector given by Humphreys et al. (1998) and Brooks and Humphreys (2004) which attempts to correct for spherical spreading and is, therefore, more appropriate for 3D imaging then Equation (2.15). Deconvolution of these beamformed maps was performed using *CLEAN-SC*, see Section 2.4.2. For each beamformed plot, 100 iterations of *CLEAN-SC* were used with an loop gain $\boldsymbol{\varphi} = 0.95$. The *DAMAS* algorithm code was also written but not used extensively due to the slow processing speed.

7.3 Initial Experiments

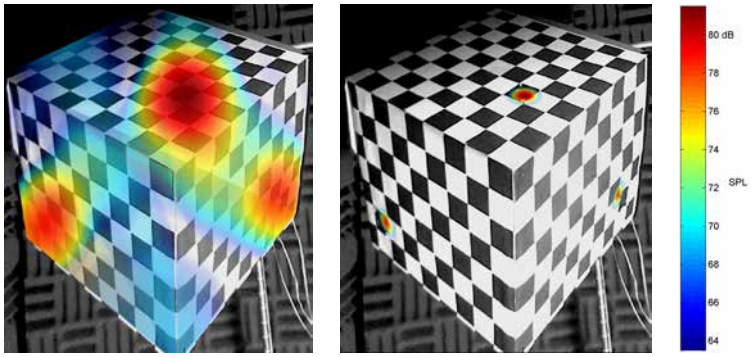
This section describes initial experiments that were carried out to investigate 3D beamforming and deconvolution. They were carried out before the calibration technique described in Section 6.3 was developed. Therefore, there was some degree of doubt as to



(a) Speaker locations.

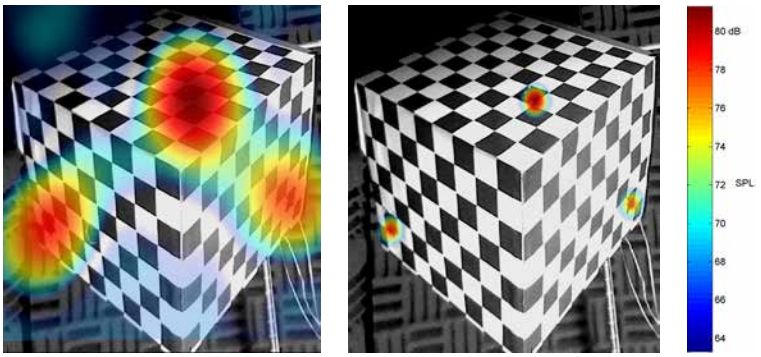


(b) Object reference frame.



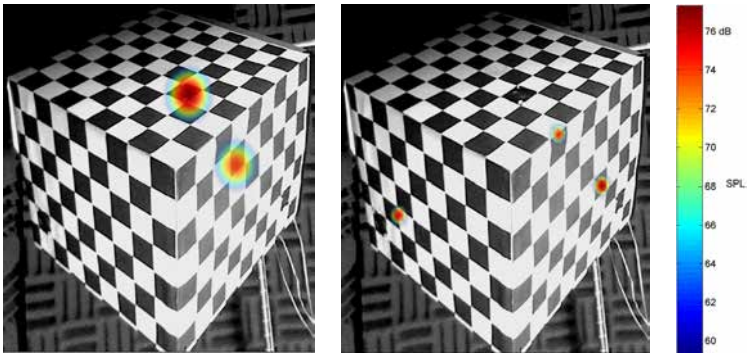
(c) Beamformed 3D map.

(d) *CLEAN-SC* 3D map.



(e) Beamformed 2D map, Z = 0.60m, plane through centre of cube.

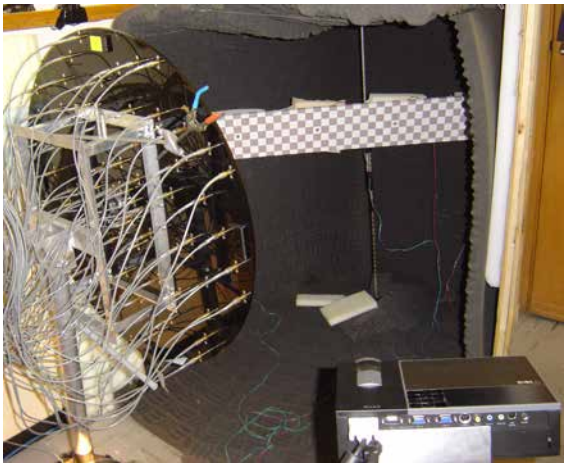
(f) *CLEAN-SC* 2D map, Z = 0.60m, plane through centre of cube.



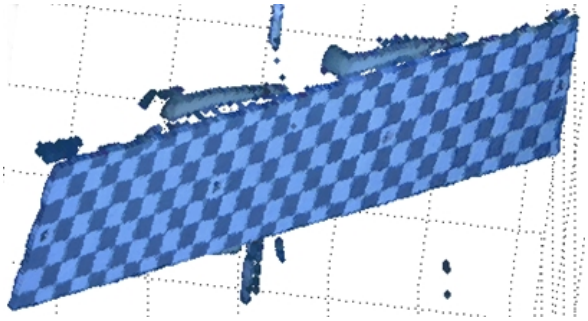
(g) *CLEAN-SC* 2D map, Z = 0.40m, plane at front of cube.

(h) *CLEAN-SC* 2D map, Z = 1.00m, plane behind cube.

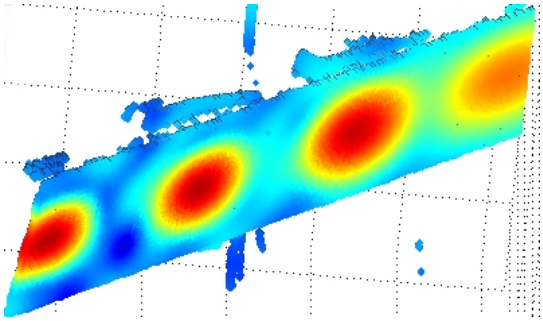
Figure 7.6: Initial comparison of 2D and 3D *CLEAN-SC* deconvolution maps using three speakers mounted flush into the surface of a cardboard box and the third octave frequency band between 5.8 and 7.3kHz. The Z component distances of the speakers from the array were 571, 681 and 652mm.



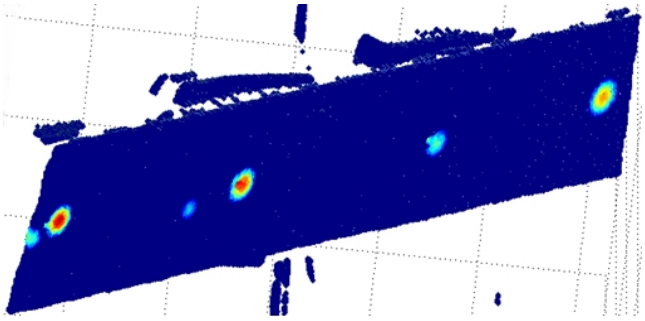
(a) Experimental setup.



(b) Structured light 3D scan.



(c) Beamforming 3D map obtained using the structured light scan.



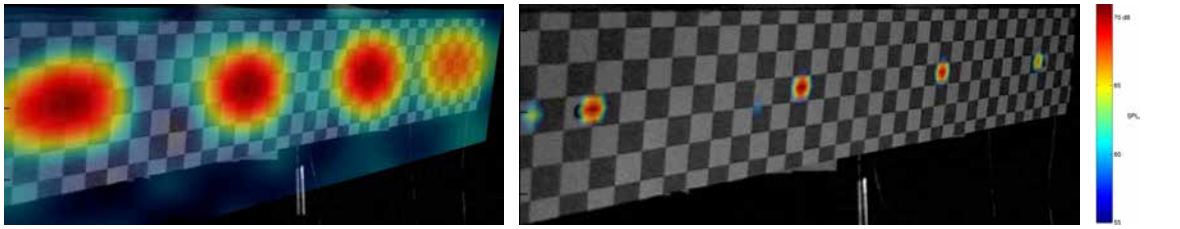
(d) *CLEAN-SC* 3D map obtained using the structured light scan.

Figure 7.7: Structured light 3D scan beamforming and *CLEAN-SC*. The experiment was carried out using an Underbrink multi-arm spiral array, data projector, and a board with four speakers mounted behind holes spaced 300mm apart.

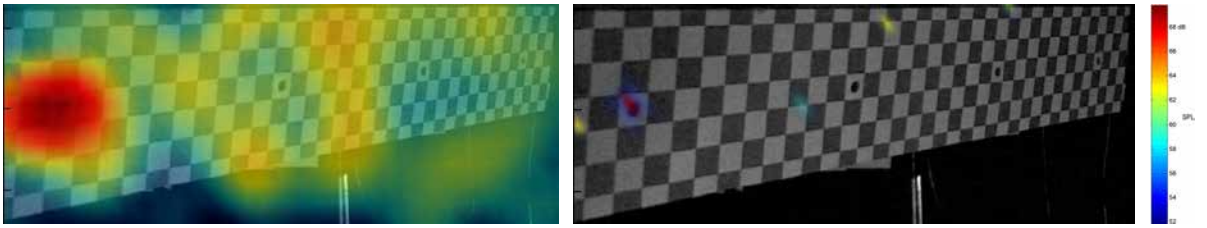
whether an error was due to inaccuracies in the coordinates of microphones, alignment of the camera, speed of sound, or the use of a 2D or 3D scanning surface. The results obtained were very subjective and only provided data for a few sound source locations and limited frequency range. They are presented here to illustrate the motivation for developing the error analysis technique presented in Section 7.4.

Initial comparisons of the use of 2D and 3D scanning surfaces for beamforming and *CLEAN-SC* were carried out using an Underbrink multi-arm spiral array. The 3D scanning surfaces were generated using the computer vision techniques described in Section 5.2. The 2D planes were generated using a 2D grid at a set distance Z from the array. Three 3D objects were used in these initial experiments: a small cardboard box, a rectangular plane, and a large cube.

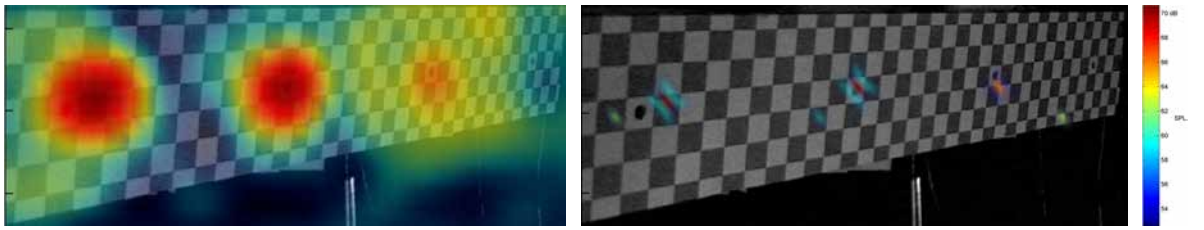
The first experiment used a small cardboard box ($220 \times 260 \times 290$mm) which contained three speakers which were mounted flush with the surface of the box. The box was filled with acoustic foam. The Z axis coordinates of the speakers were 571, 681 and



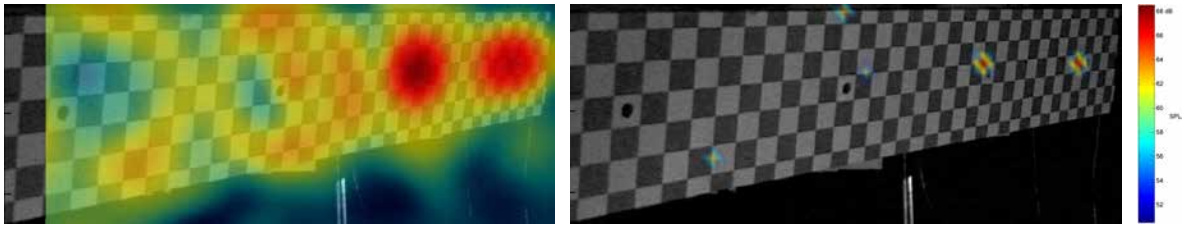
(a) Beamformed 3D map. (b) *CLEAN-SC* 3D map.



(c) Beamformed 2D map, Z = 0.68m. (d) *CLEAN-SC* 2D map, Z = 0.68m.



(e) Beamformed 2D map, Z = 0.89m. (f) *CLEAN-SC* 2D map, Z = 0.89m.



(g) Beamformed 2D map, Z = 1.58m. (h) *CLEAN-SC* 2D map, Z = 1.58m.

Figure 7.8: Initial comparison of 3D and 2D beamforming and *CLEAN-SC* maps for third octave frequency band 5.79 to 7.28 kHz for speakers mounted behind holes, spaced 300mm apart, in a board positioned at an angle to the Underbrink multi-arm spiral array.

652mm. The 3D surface geometry was obtained using an image of a checkerboard pattern attached to the cube captured by the array's camera and using computer vision software, see Section 5.2.1. Figure 7.6 gives example figures for 3D and 2D scanning surfaces for the 1/3 octave band from 5.8 to 7.3kHz. The use of a 3D scanning surface with *CLEAN-SC* gave relatively accurate results, (about 10-15mm position error). A 2D scanning surface at $Z = 600$mm, through the centre of the cube, gave similar results for *CLEAN-SC* compared to the 3D case (about 15mm position error). Significant error was obtained for *CLEAN-SC*, however, for 2D scanning surfaces located at $Z = 400$mm, in front of the cube (about 10 and 120mm position error and one peak missing), and for $Z = 1000$mm, behind the cube (about 40, 60, and 100mm).

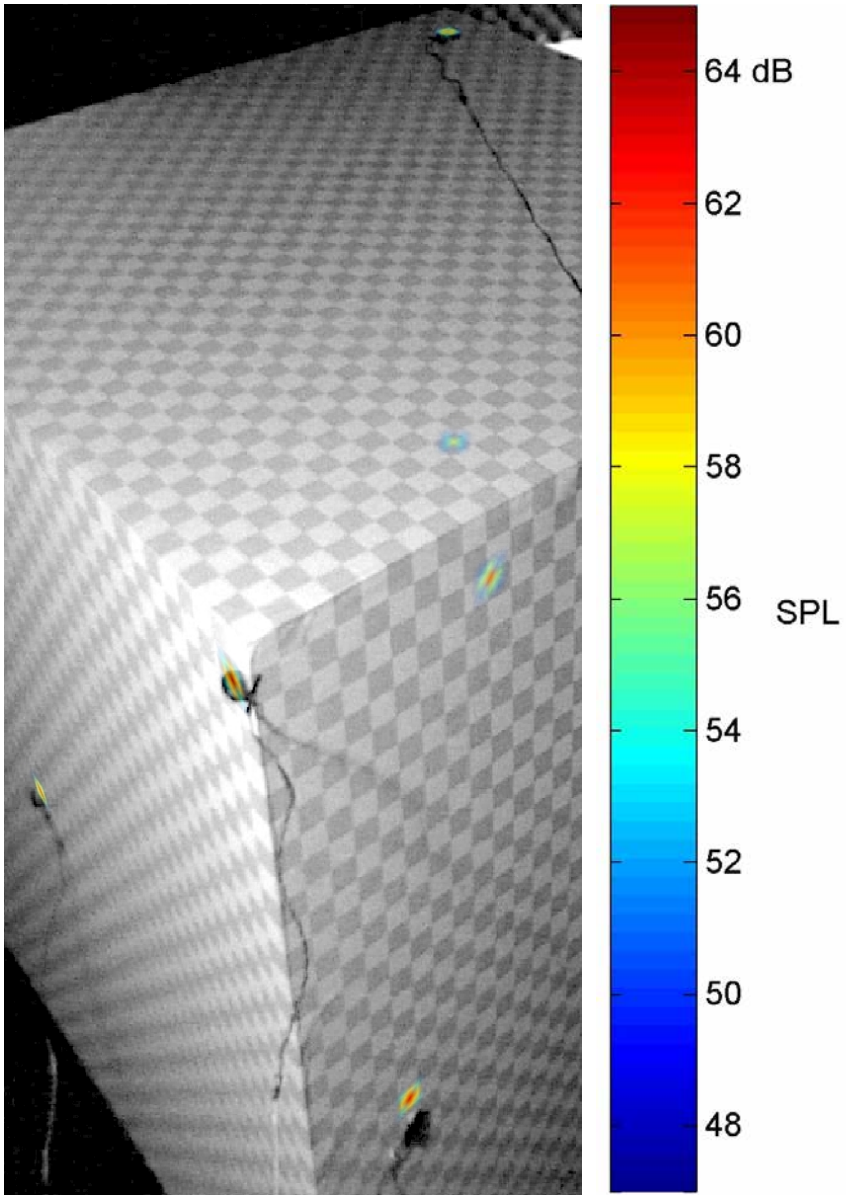


Figure 7.9: Initial *CLEAN-SC* test using speakers on the surface of a large cube.

The next experiment used a board with speakers mounted behind holes, which were located in a line, spaced 300mm apart, see Figure 7.7. The board was positioned at an angle to the Underbrink multi-arm spiral array. This experiment provides a greater range of distances of sound sources from the array than was able to be achieved for the small cube. The 3D surface geometry of the object was obtained using both the checkerboard pattern method and using a structured light scan, see Section 5.2.2. Again the 3D *CLEAN-SC* gives good results, see Figures 7.8 and 7.7. For the 2D scanning surfaces, however, only two adjacent speakers were accurately imaged at a time. The other two speakers had large errors in their positions and measured sound pressure levels, see (Figure 7.8). This was thought to be mainly due to the poor depth of field of the array.

A large cube ($1.2 \times 1.2 \times 1.225$m) was then used as a 3D object. The 3D surface geometry was obtained using an image of a checkerboard pattern and computer vision software, see Figure 7.9. The sound sources were relatively well located (about 10-60mm position error) using 3D *CLEAN-SC*, see Figure 7.9. No attempt was made at using a 2D scanning surface since the previous experiment had shown that the depth of field of the Underbrink array would only allow the array to accurately detect one speaker at a time.

These experiments indicated that the use of a 3D scanning surface for *CLEAN-SC* gives good results. Accurate results from 2D scanning surfaces are limited to situations where the distance of the 2D scanning surface from the sound source was small. Approximate position errors were obtained by covering the object with a checkerboard pattern and overlaying *CLEAN-SC* maps over camera images of the object. The checkerboard pattern was then used to identify the position error of a deconvolution peak relative to a speaker location. There were many factors that could effect this error and results were very qualitative. A more quantitative method was desired.

7.4 Methodology for Error Analysis of 2D and 3D *CLEAN-SC* Maps

A new technique was developed for error analysis of *CLEAN-SC* acoustic maps. It was designed to be an extension to the new calibration technique presented in Section 6.3. It uses the raw microphone data, captured as part of the calibration technique, to generate beamforming and *CLEAN-SC* acoustic maps. These maps, combined with the measured position of the speakers, were then used for error analysis of the *CLEAN-SC* acoustic maps.

7.4.1 Raw Data

The sound source object used was a planar rig, composed of six speakers inserted into a piece of perspex plastic surrounding a checkerboard pattern, attached to a tripod, see Figure 7.10. The rig was setup in front of the array and, for a range of positions and orientations of the calibration rig, camera images, microphone recordings, and optionally structured light scans made. Each position or orientation of the calibration rig corresponded to a single 'image number'. Nine microphone recordings were then made for each image number. Table 7.1 summarises the speakers numbers and the DAC hardware used for each record number. For 'record number' 1 to 6, a burst of white noise

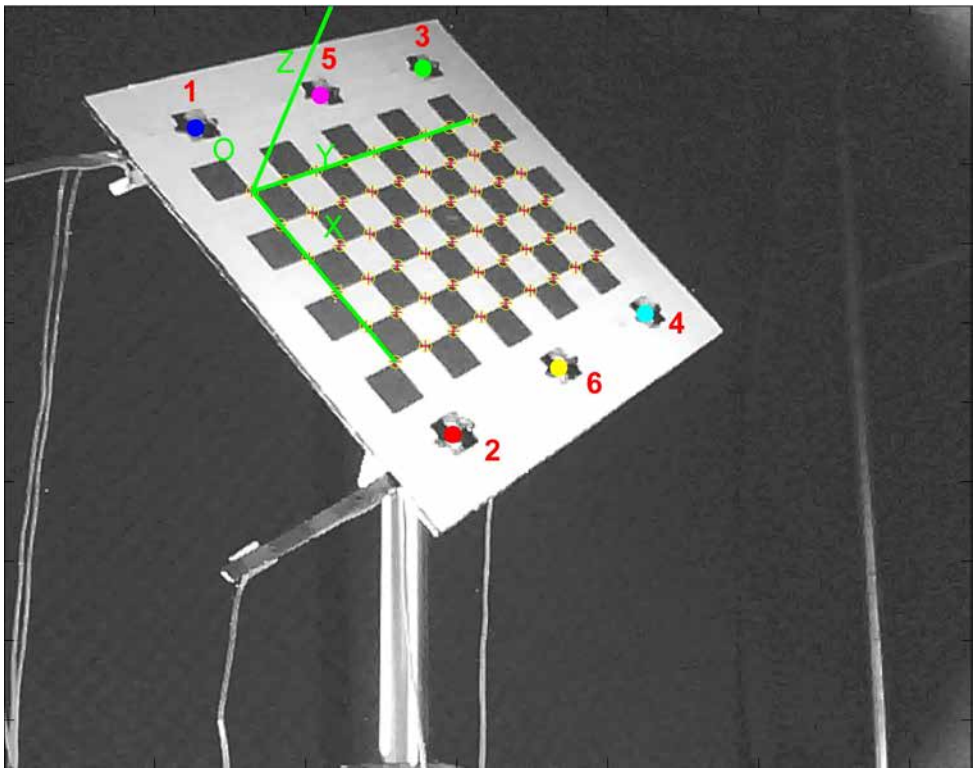


Figure 7.10: Sound source setup.

was played on only one speaker at a time using the speaker number corresponding to the record number. These six record numbers were used to generate time of arrival data which was used for microphone position calibration in the camera reference frame, see Section 6.3. An additional three recordings were made using uncorrelated white noise played on multiple speakers at once. These were designed to test *CLEAN-SC*'s ability to locate sound sources as the number of sound sources increased and the spacing between them decreased. Record No.9 used the four microcontrollers to generate continuous uncorrelated white noise (separate clocks and start buttons and different sequences), see Section 3.1.1. These were used because there was concern that there might be correlation between the outputs of the DT9836 boards which shared the same clock and hardware trigger. Also, while the microcontrollers used continuous white noise signals, the Data Translation DT9836 analog outputs had been used to generate 0.175s bursts of white noise, see Figure 6.7a. The microphone data were recorded for 0.34s at 98kSPS per channel.

Speaker Coordinates $\vec{X}_s$ in the Camera Reference Frame

The coordinates of the speakers, in the array's camera reference frame, were obtained by capturing an image of the checkerboard and using computer vision software to obtain the

Record Number	Speaker Number	White Noise Generator
1	1	DT9836
2	2	DT9836
3	3	DT9836
4	4	DT9836
5	5	DT9836
6	6	DT9836
7	1 to 4	DT9836
8	1 to 6	DT9836
9	1 to 4	μcontrollers

Table 7.1: Play method used for each record number.

extrinsic parameters of the checkerboard reference frame. The coordinates of the speakers were defined in this reference frame, see Table 6.1, and a rigid body rotation was then used to convert them into the camera reference frame. Figure 7.12 shows the errors in the translation and rotation vectors provided by the computer vision software and shows that the 3D position accuracy will decrease as the sound source is moved way from the array. A higher quality camera might be expected to achieve more accuracy.

For an arbitrarily shaped object where speakers could not be defined in a checkerboard reference frame, different methods could be used. An alternative method for obtaining the speaker coordinates could be to making a structured light scan and selecting the centre of the speaker from the resulting point cloud. Alternatively, the 3D coordinates of the speaker could be obtained by identifying the pixel coordinates $\vec{\boldsymbol{p}}$ of the speaker in images captured by two or more cameras in the array and using stereo triangulation, see Section 4.3. Colour coding or LEDs might be used to automate the speaker selection process from the raw data. These methods should give similar results to the checkerboard method.

7.4.2 Beamforming and *CLEAN-SC* Methodology

Beamforming and *CLEAN-SC* were performed in 12^{th} octave frequency bands from 300Hz to 23kHz using the methodology described in Section 7.2. Software was written that took the raw data and automatically generated the beamformed and *CLEAN-SC* acoustic maps for each image number, record number, and octave band. For each *CLEAN-SC* map, the peak magnitudes $\boldsymbol{b}_{max}$ and their coordinates $\vec{\boldsymbol{\xi}}_{max}$ in the *CLEAN-SC* maps are obtained and used for error analysis.

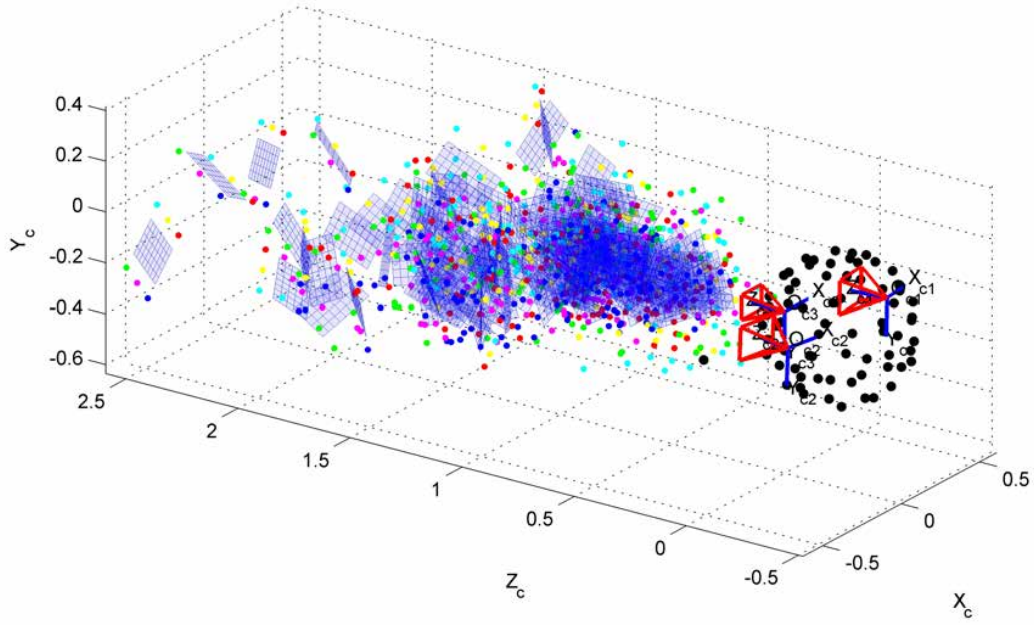


Figure 7.11: Plot showing the microphone coordinates (black dots), camera positions and orientations, speaker coordinates (coloured dots), and the position and orientation of the checkerboard patterns. These data was obtained during calibration and were used for beamforming and *CLEAN-SC*and for position error analysis.

Scanning Surface Corresponding to the 3D Surface Geometry of the Object

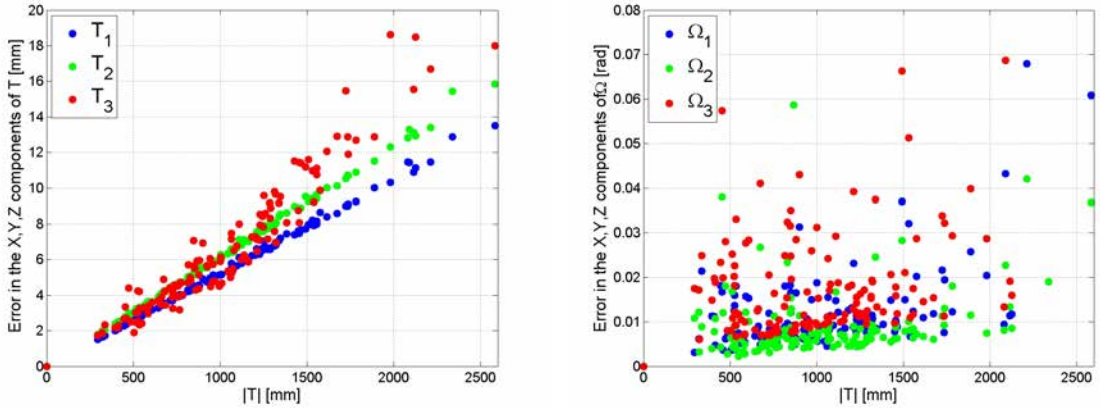
Real 3D scanning point coordinates were defined, in the checkerboard reference frame, using a X and Y axes range of -150 to 350mm with a spacing between scan points of 6mm. A rigid body transformation

$$\vec{\boldsymbol{\xi}}_R = \mathbf{R}_R \, \vec{\boldsymbol{\xi}}_R{}' + \vec{\boldsymbol{T}}_R$$

was then applied to these coordinates using the extrinsic parameters of the checkerboard, see Figure 7.14.

Traditional 2D Scanning Surface which is Not Related to the Object's Surface Geometry

The simplest method of generating the traditional planar scanning surface would have been to define a 2D grid at a distance Z from the array. However, it was suggested by



(a) Translation vector $\vec{\boldsymbol{T}}$ errors.

(b) Rotation vector $\vec{\boldsymbol{\Omega}}$ errors.

Figure 7.12: Translation and rotation vector errors provided by the MatLab Camera Calibration Toolbox for the images used, in combination with time of arrival data, to obtain the 3D coordinates of the speakers. These were also used to obtain the position and orientation of the 3D scanning surfaces used for beamforming and *CLEAN-SC* error analysis.

Dougherty (2010b) that the traditional 2D scanning surface should be generated using the same techniques as the 3D scanning surfaces so that any errors in the 3D scanning surface would be the same as in the 2D scanning surface. The traditional 2D scanning surfaces were, therefore, generated using images, captured by the cameras in the array, of the checkerboard pattern at a range of distances from the array, see Figure 7.16. Computer vision software was used to obtain the extrinsic parameters $\mathbf{R}_F$ and $\vec{\boldsymbol{T}}_F$ of the checkerboard. These were then used to generate the traditional 2D scanning surfaces using

$$\vec{\boldsymbol{\xi}}_F = \mathbf{R}_F \, \vec{\boldsymbol{\xi}}_F{}' + \vec{\boldsymbol{T}}_F.$$

Scan points where composed of 85×85 equally spaced grid points. The X and Y axes limits of this grid were obtained by projecting the real surface geometry of the object onto the 2D scanning surface with $Z = \vec{\boldsymbol{T}}_3$ using the perspective projection

$$\vec{\boldsymbol{\xi}}_N = \vec{\boldsymbol{\xi}}_R \, \frac{\vec{\boldsymbol{T}}_{F3}}{\vec{\boldsymbol{\xi}}_{R3}}. \tag{7.19}$$

7.4.3 Method Used to Define *CLEAN-SC* Position Error

A comparison between the position error of 2D and 3D *CLEAN-SC* maps was problematic. Consider a speaker located at $\vec{\boldsymbol{X}}_s$, which results in a peak in a *CLEAN-SC* map at location $\vec{\boldsymbol{\xi}}_{max}$. For a 3D scanning surface corresponding to the surface of the object, one could define the position error to be

$$\varepsilon = \|\vec{\boldsymbol{X}}_s - \vec{\boldsymbol{\xi}}_{max}\|. \tag{7.20}$$

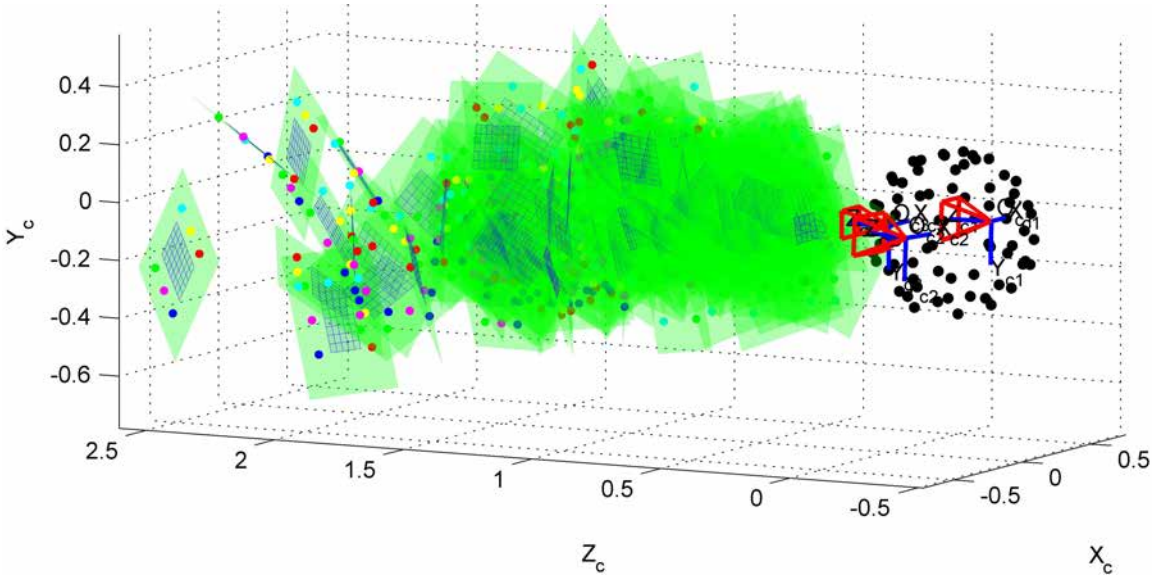


Figure 7.13: Multiple 3D scanning surfaces corresponding to surface geometry of the object being imaged. (Image numbers 1 to 140.)

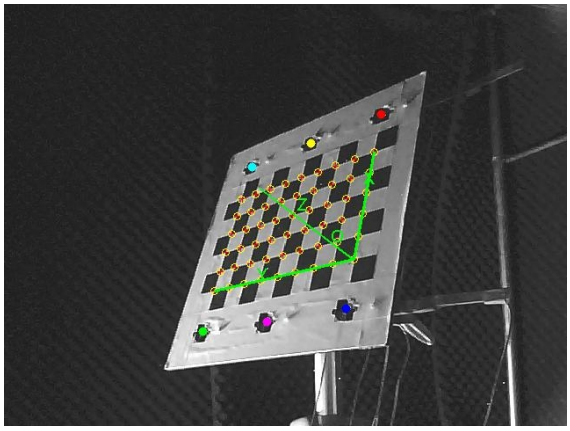
However, if a 2D scanning surface is used, where the scanning surface does not pass through a speaker, it would not seem logical to use this method since this would mainly be measuring the displacement of the scanning surface from the speaker. The deconvolution peak, therefore, could be projected from $\vec{\boldsymbol{\xi}}_{max}$ to a point $\vec{\boldsymbol{\xi}}_p$ in a plane perpendicular to the array axis at the same distance $Z = \vec{\boldsymbol{X}}_{s\,3}$ as the speaker. The error may, therefore, be defined as

$$\boldsymbol{\varepsilon} = \|\vec{\boldsymbol{X}}_s - \vec{\boldsymbol{\xi}}_p\|. \tag{7.21}$$

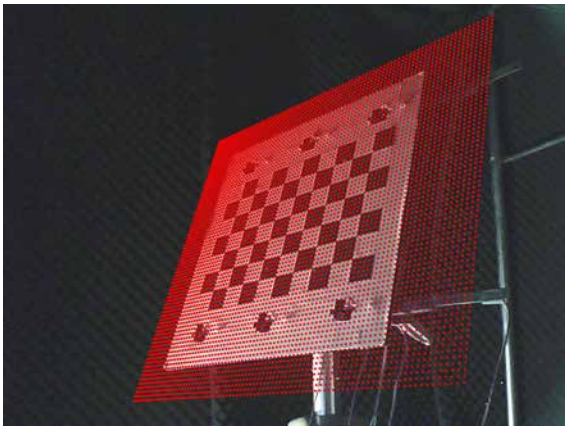
This is the position error mainly used in this work. The rest of this section describes the method used to obtain the projection coordinate $\vec{\boldsymbol{\xi}}_p$ and the corresponding speaker coordinate $\vec{\boldsymbol{X}}_s$ so that $\boldsymbol{\varepsilon}$ in Equation (7.21) may be calculated.

Calculation of the Projected Coordinate $\vec{\boldsymbol{\xi}}_p$ of *CLEAN-SC* Peaks

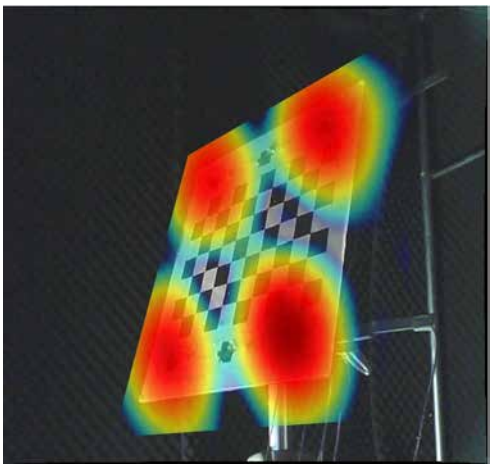
The only literature found relating to projection of acoustic maps was O'Donovan et al. (2007b,a). They modelled a solid spherical array using spherical beamforming as a perspective projection camera, with the centre of projection being the centre of the array. However, commonly the 2D acoustic maps are overlayed as transparencies over photos of the 3D object, captured by a camera in the array, using perspective projection. This method appears to make the assumption that an array can be modelled as a perspective projection camera with the centre of the projection being the location of the camera in the array. Some literature plots acoustic maps over 2D CAD-like drawings of 3D objects using what appears to be orthogonal projection.



(a) Image used to obtain the extrinsic parameters of 3D the object. These were used to generate the 3D scanning surface and obtain the coordinates of the speakers in the array's camera reference frame.



(b) Scan points corresponding to the real 3D geometry of the object.



(c) Beamformed map.

(d) *CLEAN-SC* map.

Figure 7.14: An example of a scanning surface corresponding to the real 3D geometry of the object and the 3D beamformed and *CLEAN-SC* maps generated using this scanning surfaces (Image No.74, Record No.7, and 12^{th} octave frequency range 5.014 to 5.312 kHz).

In this work, the coordinates of a deconvolution peak $\vec{\boldsymbol{\xi}}_{max}$ were projected to a plane passing through the speaker location, perpendicular to the array. Projection methods investigated were perspective projection and orthogonal projection.

Orthogonal projection projects the deconvolution peak position, along the Z axis, using

$$\vec{\boldsymbol{\xi}}_p = [\vec{\boldsymbol{\xi}}_{max\,1}, \vec{\boldsymbol{\xi}}_{max\,2}, \vec{\boldsymbol{X}}_{s\,3}]^{\intercal}. \tag{7.22}$$

Perspective projection was achieved by first using a change of reference frame

$$\vec{\boldsymbol{\xi}}'_{max} = \mathbf{R}\,\vec{\boldsymbol{\xi}}_{max} + \vec{\boldsymbol{T}} \quad \text{and} \quad \vec{\boldsymbol{X}}'_s = \mathbf{R}\,\vec{\boldsymbol{X}}_s + \vec{\boldsymbol{T}} \tag{7.23}$$

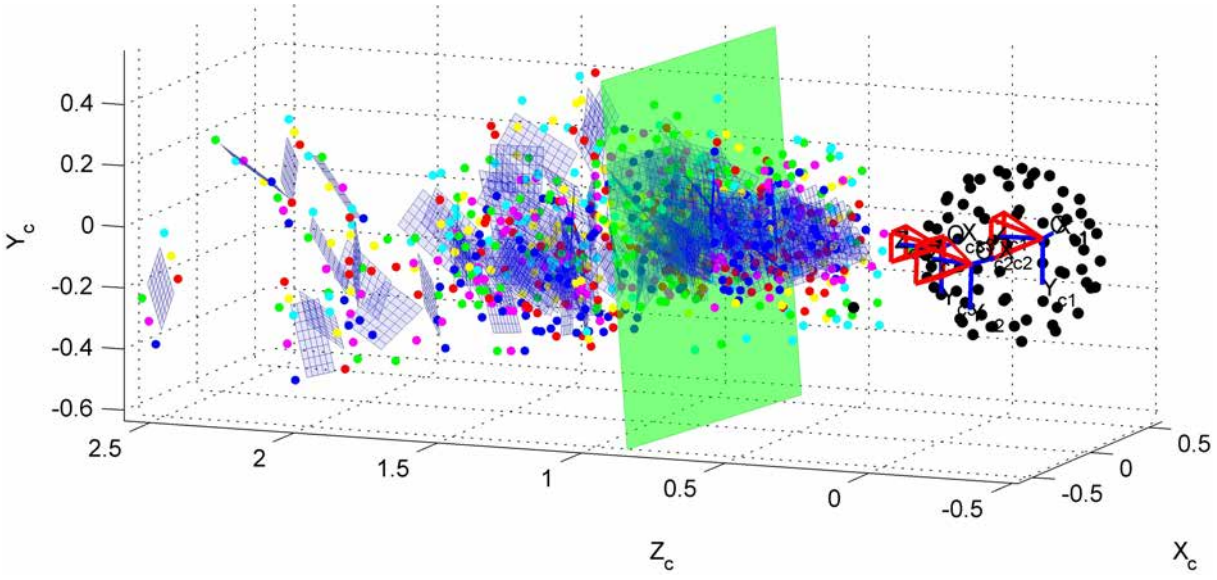


Figure 7.15: One traditional 2D scanning surface used to acoustically image speakers at multiple locations. (Image numbers 1 to 140 and traditional 2D scanning surface No.11)

and then using perspective projection of the form

$$\vec{\boldsymbol{\xi}}'_p = \frac{\vec{\boldsymbol{X}}'_{s\,3}}{\vec{\boldsymbol{\xi}}'_{max\,3}}\,\vec{\boldsymbol{\xi}}'_{max}. \tag{7.24}$$

A change of reference frame

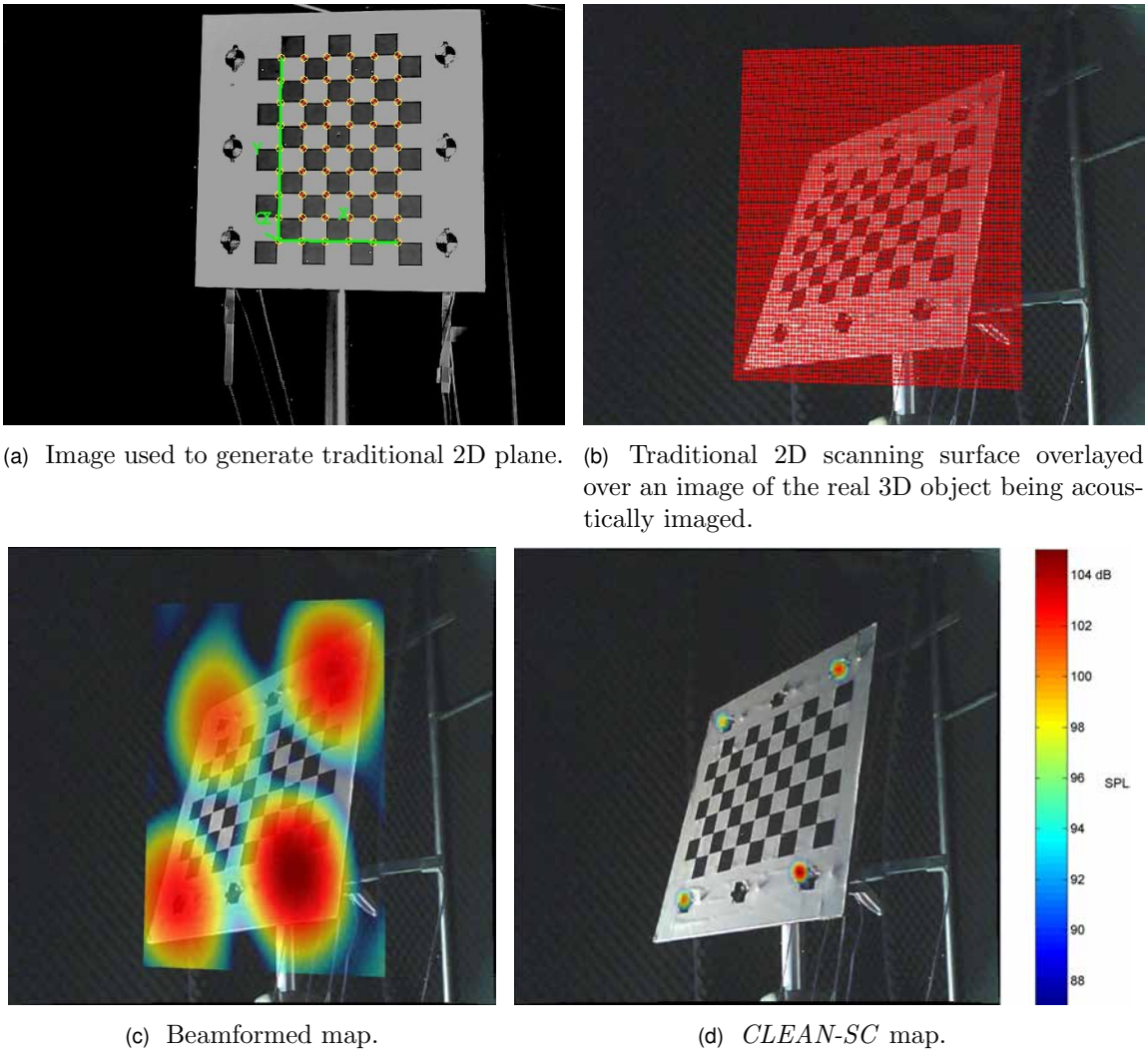
$$\vec{\boldsymbol{\xi}}_p = \mathbf{R}^{\intercal}\,(\vec{\boldsymbol{\xi}}'_p - \vec{\boldsymbol{T}}) \tag{7.25}$$

was then used to convert back to the original reference frame. For projection from the centre of the array, the mean microphone coordinates

$$\vec{\boldsymbol{T}} = -\frac{1}{M}\sum_{m=1}^{M}\vec{\boldsymbol{X}}_m, \tag{7.26}$$

were used as the centre of projection and no rotation, $\mathbf{R} = \boldsymbol{I}_{(3,3)}$, were used. For perspective projection from the main camera in the array, no change of reference frame was required since $\vec{\boldsymbol{\xi}}_{max}$ and $\vec{\boldsymbol{X}}_s$ were in this camera reference frame. For perspective projection from one of the other cameras, however, $\mathbf{R}$ and $\vec{\boldsymbol{T}}$ are the stereo extrinsic parameters describing the rigid body rotation between the main camera and the camera being used for projection.

Figure 7.17 gives an example of perspective projection of deconvolution peaks using the centre of the microphone coordinates as the centre of projection. This method of projection was found to give the best results, compared to orthogonal projection or using perspective projection with camera positions, which are offset from the centre of the array, as the centre of projection, see Figure 9.5. Instead of using the mean microphone



(a) Image used to generate traditional 2D plane.

(b) Traditional 2D scanning surface overlayed over an image of the real 3D object being acoustically imaged.

(c) Beamformed map.

(d) *CLEAN-SC* map.

Figure 7.16: An example of a traditional 2D scanning surfaces, passing through the centre of the object and the beamformed and *CLEAN-SC* maps generated using this scanning surface. The range of the scanning grid was limited using the extrinsic parameters of the 3D object and perspective projection. ('Traditional plane' Image No.11, 'True' Image No.74, Record No.7, and 12^{th} octave frequency range 5.014 to 5.312 kHz)

coordinates, perhaps the centre of projection could have been obtained using perspective projection to project back from the speaker locations $\vec{\boldsymbol{X}}_s$ through the peak coordinates $\vec{\boldsymbol{\xi}}_{max}$ and then use triangulation to identify a centre of projection.

Relating *CLEAN-SC* Peak Projection Coordinates $\vec{\boldsymbol{\xi}}_p$ to Speaker Coordinates $\vec{\boldsymbol{X}}_s$

The measurement of the position error of a projected *CLEAN-SC* coordinate set requires that the speaker generating the *CLEAN-SC* peak is correctly identified. For recordings where one speaker played at a time (Record No.1-6), it is assumed that the speaker used to generate the acoustic signal corresponds to the *CLEAN-SC* first iteration peak coordinate

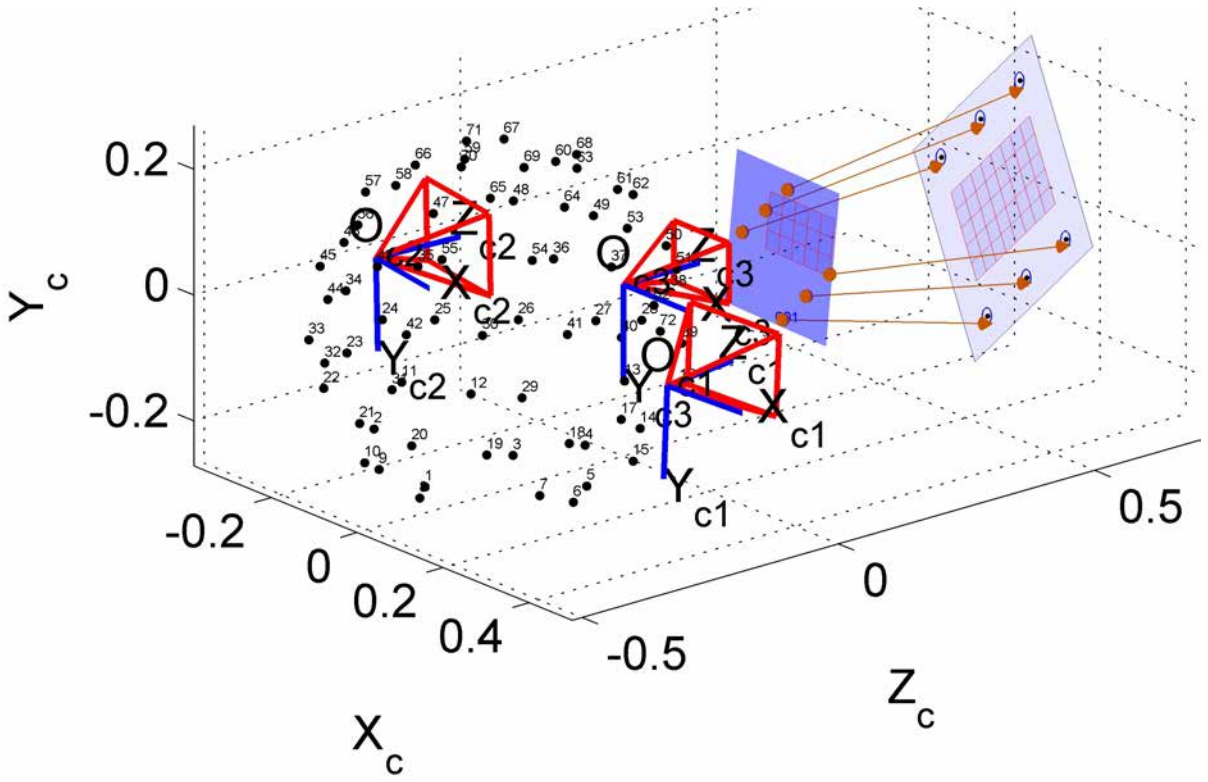
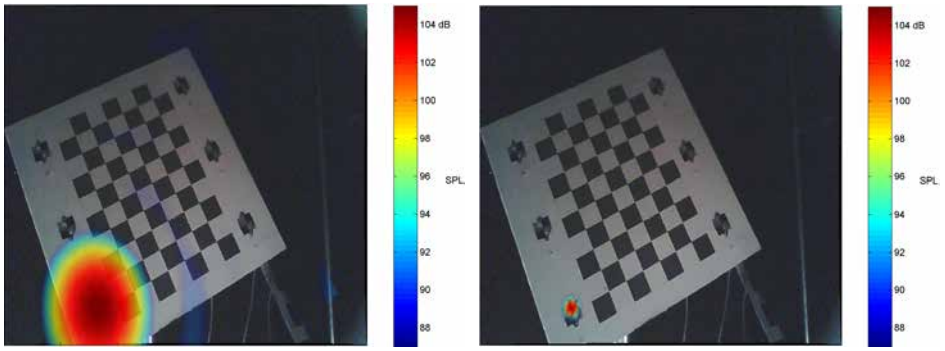
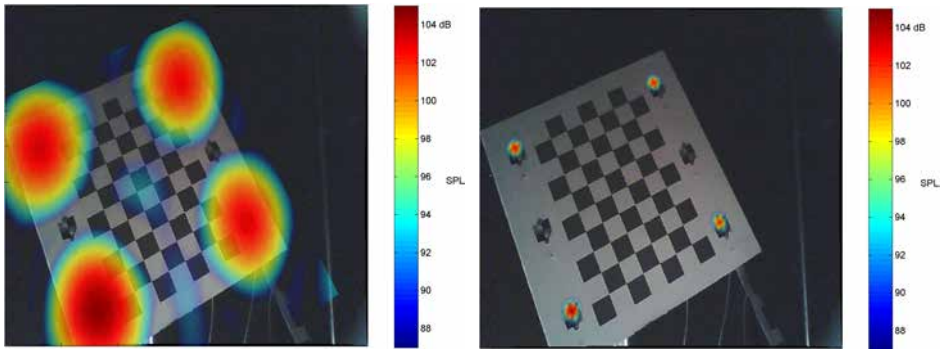


Figure 7.17: Perspective projection of the *CLEAN-SC* intensity peak coordinates $\vec{\xi}_{max}$, on a traditional planar scanning surface, to the corresponding speaker distances Z from the array. The centre of projection used was the centre of the microphone coordinates.

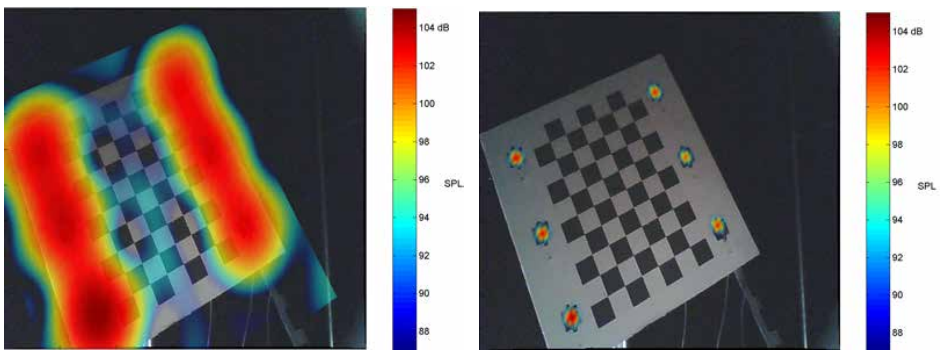
$\vec{\xi}_{max}^{\{1\}}$. This was the peak with the highest SPL value, see Figure 7.18b. However, for recordings where the acoustic signal was generated on multiple speakers simultaneously, such as was the case for Record No.7-9, see Figures 7.18d and 7.18f, for example, it is more difficult to identify which speaker a *CLEAN-SC* coordinate may correspond to. Consider that white noise was played on S speakers simultaneously. In this case, only the coordinates corresponding to the first S *CLEAN-SC* iterations $\vec{\xi}_{max}^{\{1\}} \cdots \vec{\xi}_{max}^{\{S\}}$ were used. Each peak coordinate was projected to all S speakers and position errors calculated using Equation (7.21). The speaker that gave the least position error was assumed to be the speaker that corresponded to that *CLEAN-SC* peak. One problem with this method is that a *CLEAN-SC* peak position error greater than half the spacing between speakers may appear to correspond to the neighbouring speaker, resulting in an underestimation of the true position error. Thus, the maximum measured position error for Record No.7-9, where multiple speakers were played simultaneously, may be lower than for Record No.1-6, where only one speaker was played at a time.



(a) Beamformed map, Recording No 1. (b) *CLEAN-SC* map, Recording No 1.



(c) Beamformed map, Recording No 7. (d) *CLEAN-SC* map, Recording No 7.



(e) Beamformed map, Recording No 8. (f) *CLEAN-SC* map, Recording No 8.

Figure 7.18: Example of beamformed and corresponding *CLEAN-SC* acoustic maps, where the real 3D geometry of the object was used as a scanning surface, for Image No.137, twelfth octave frequency band 5.014 to 5.312kHz. These maps illustrate the relationship between the Record Number and the speaker play method used for each image number.

7.5 Measured *CLEAN-SC* Depth of Field and Position Error

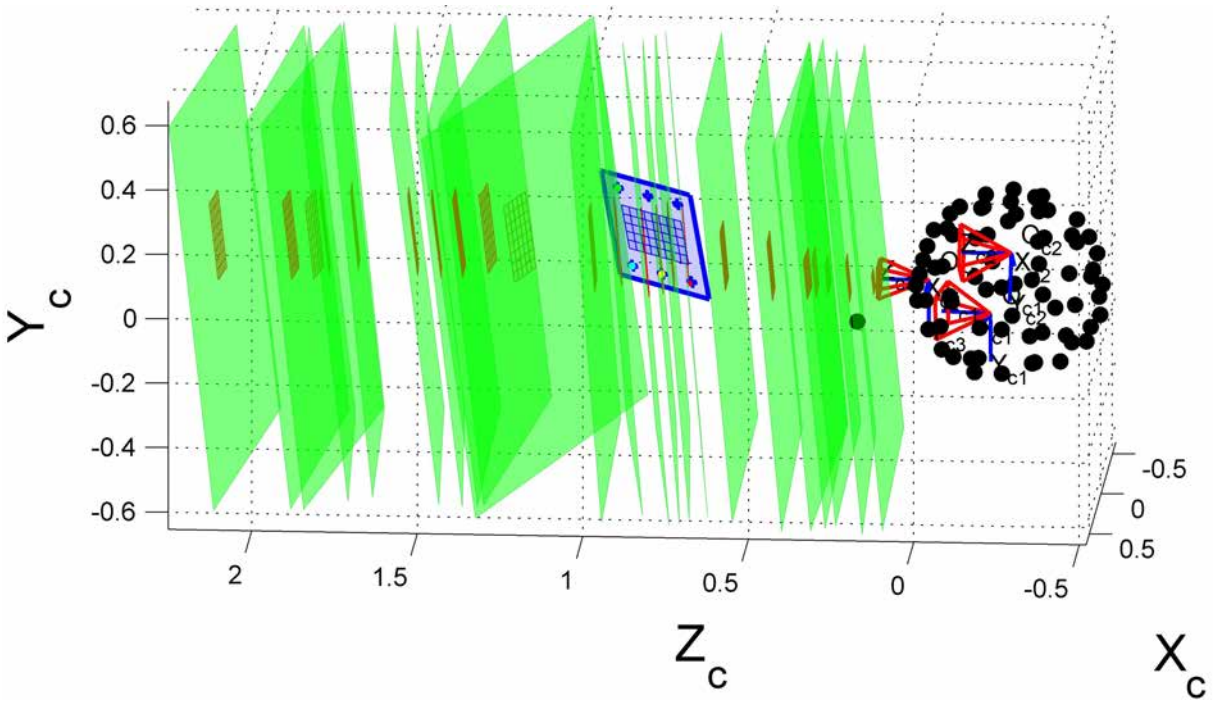
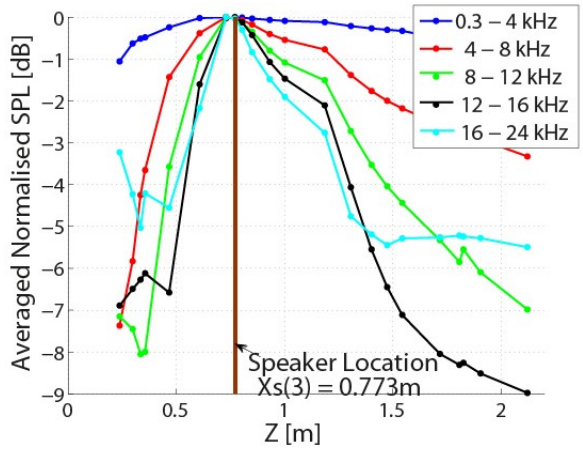


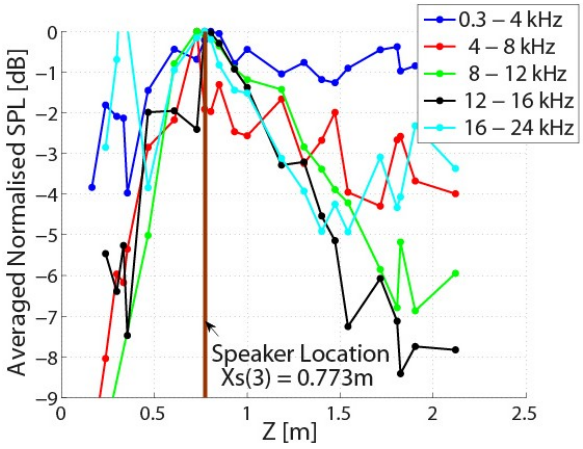
Figure 7.19: Multiple traditional 2D scanning surfaces used to image a speaker at one location.

The speaker rig shown in Figure 7.10, which contained six speakers, was setup at 140 different positions and orientations relative to the array, giving 840 different speaker positions, see Figure 7.11. For each position/orientation of the speaker rig, a total of nine microphone recordings were obtained. White noise was played first on only one speaker at a time and then simultaneously on four, six, and then again on four speakers, see Section 7.4.1 and Table 7.1 for more details. A *MATLAB* code was written to automatically cycle through these raw data, performing beamforming and *CLEAN-SC*, and obtaining the *CLEAN-SC* peak intensity value $\boldsymbol{b}_{max}$ and coordinate $\vec{\boldsymbol{\xi}}_{max}$ for each speaker in 12^{th} octave frequency bands ranging from 300Hz to 23kHz. This was performed using scanning surfaces corresponding to the 3D surface geometry of the object, see Figure 7.13, and using traditional planar scanning surfaces which had no relationship to the surface geometry of the object, see Figure 7.15. The *CLEAN-SC* peak coordinates $\vec{\boldsymbol{\xi}}_{max}$ were then projected to the speaker locations using perspective projection and the centre of the microphone coordinates as the centre of projection, see Equations (7.23) to (7.26). The position error was then calculated using Equation (7.21) for each speaker and each 12^{th} octave frequency band.

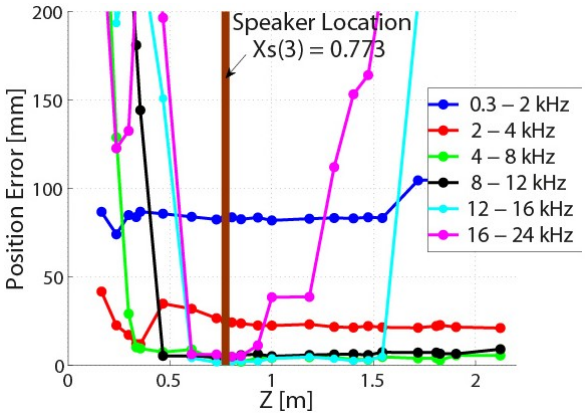
7.5.1 Measured Depth of Field of the Array for *CLEAN-SC* and Beamforming Using Multiple Traditional 2D Scanning Surfaces



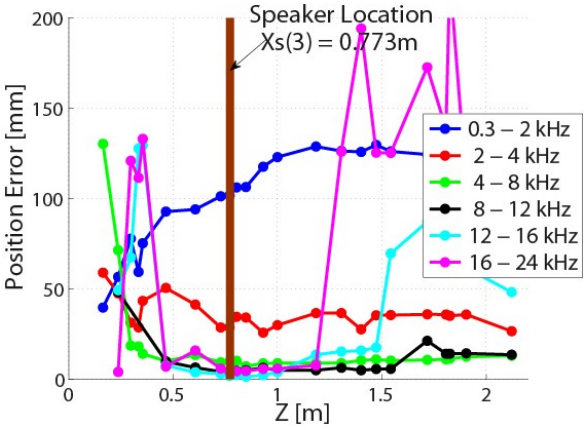
(a) Depth of field for Speaker No.1 where sound was only played on Speaker No.1 (Record No.1).



(b) Depth of field for Speaker No.1 where sound was played on all six speakers simultaneously (Record No.8).



(c) Position error for Speaker No.1 where sound was only played on Speaker No.1 (Record No.1).



(d) Position error for Speaker No.1 where sound was played on all six speakers simultaneously (Record No.8).

Figure 7.20: Plot showing the measured depth of field and position error in the projected *CLEAN-SC* peak coordinates for Speaker No.1, which was located at a distance of 0.773m from the spherical array. Multiple traditional 2D scanning surfaces were used at different distances from the array. *(Speaker No.1, Image No.45, Traditional Image No.1-23.)*

The depth of field of an array describes the reduction of the measured sound pressure level of a source as a traditional planar scanning surface is moved away from the true source location. This was investigated for the spherical array using a single sound source and 23 traditional 2D scanning surfaces at different distances from the array, see Figure 7.19. These scanning surfaces were generated using the extrinsic parameters of images captured by the camera in the array. For each scanning surface, a beamformed and *CLEAN-SC* acoustic map were generated using 12^{th} octave frequency bands and normalised steering

vectors, calculated using Equation (2.15). The *CLEAN-SC* peak intensity values $\boldsymbol{b}_{max}$ and coordinate $\vec{\boldsymbol{\xi}}_{max}$ corresponding to Speaker No.1 were obtained for each traditional 2D scanning surface and each octave band. The normalised peak intensity 12^{th} octave values $\boldsymbol{b}_{max}$, were then averaged into wider frequency bands and plotted as a function of the Z component of $\vec{\boldsymbol{\xi}}_{max}$.

Figure 7.20a shows the depth of field where only Speaker No.1 was played, see Section 7.4.1. This shows the *CLEAN-SC* depth of field for a single speaker. Since only one sound source is used, this is also equivalent to the beamforming depth of field. This plot shows a decrease in measured SPL as the traditional planar scanning surface was moved away from the true sound source location. This is consistent with the results obtained by Döbler et al. (2008), Maffei and Bianco (2008) and Irimia et al. (2009) for depth of field for beamforming. Figure 7.20b shows the *CLEAN-SC* depth of field for Speaker No.1 where all six speakers were played simultaneously. This plot shows similar results compared to the single speaker case but appears to have more noise. *CLEAN-SC* appears to be having difficulty separating the sound intensity due to each speaker. This could be due to some correlation between the sound sources. Alternatively, a beamforming peak from one speaker could be superimposing with a sidelobe from another speaker. This would cause CLEAN-SC to give inaccurate sound pressure levels.

7.5.2 Measured *CLEAN-SC* Position Error for Multiple Traditional 2D Scanning Surfaces and a Single Sound Source Location

The position error was determined using the methodology described in Section 7.4.3. The same *CLEAN-SC* peaks intensity coordinates $\vec{\boldsymbol{\xi}}_{max}$, used in the depth of field measurements in Section 7.5.1 were used for this experiment. These coordinates were projected to the speaker locations using perspective projection, where the centre of projection was the centre of the microphone coordinates, see Equations (7.23) to (7.26). The position error was then calculated using Equation (7.21) for each 12^{th} octave frequency band and each traditional planar scanning surface. This position error was then averaged into frequency bands and plotted as a function of the Z component of $\vec{\boldsymbol{\xi}}_{max}$. For Figure 7.20c only Speaker No.1 was played, while in Figure 7.20d all six speakers were played simultaneously. These plots show significant error in position for averaged frequency bands 0.3–2kHz and 2–4kHz for all traditional 2D scanning surfaces. However, above 4kHz, for traditional 2D scanning surfaces near the true sound source location, the position errors were on average 4.5mm. This was comparable to the scanning surface grid resolutions, once it had been projected to the speaker location. As distance from the true source

location increased, the error slowly increased and then jumped significantly. As the frequency increased, sound sources were generally only located accurately for 2D planes with increasingly smaller offsets of 2D planes from the true sound source locations.

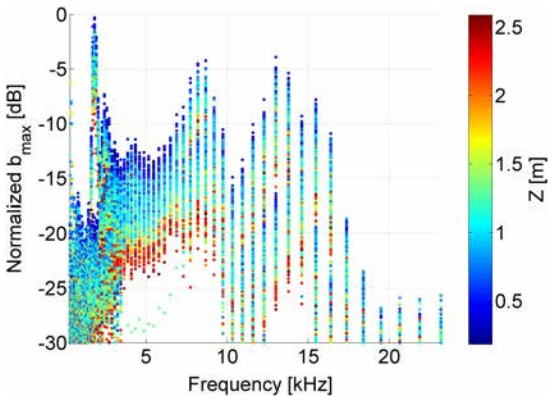
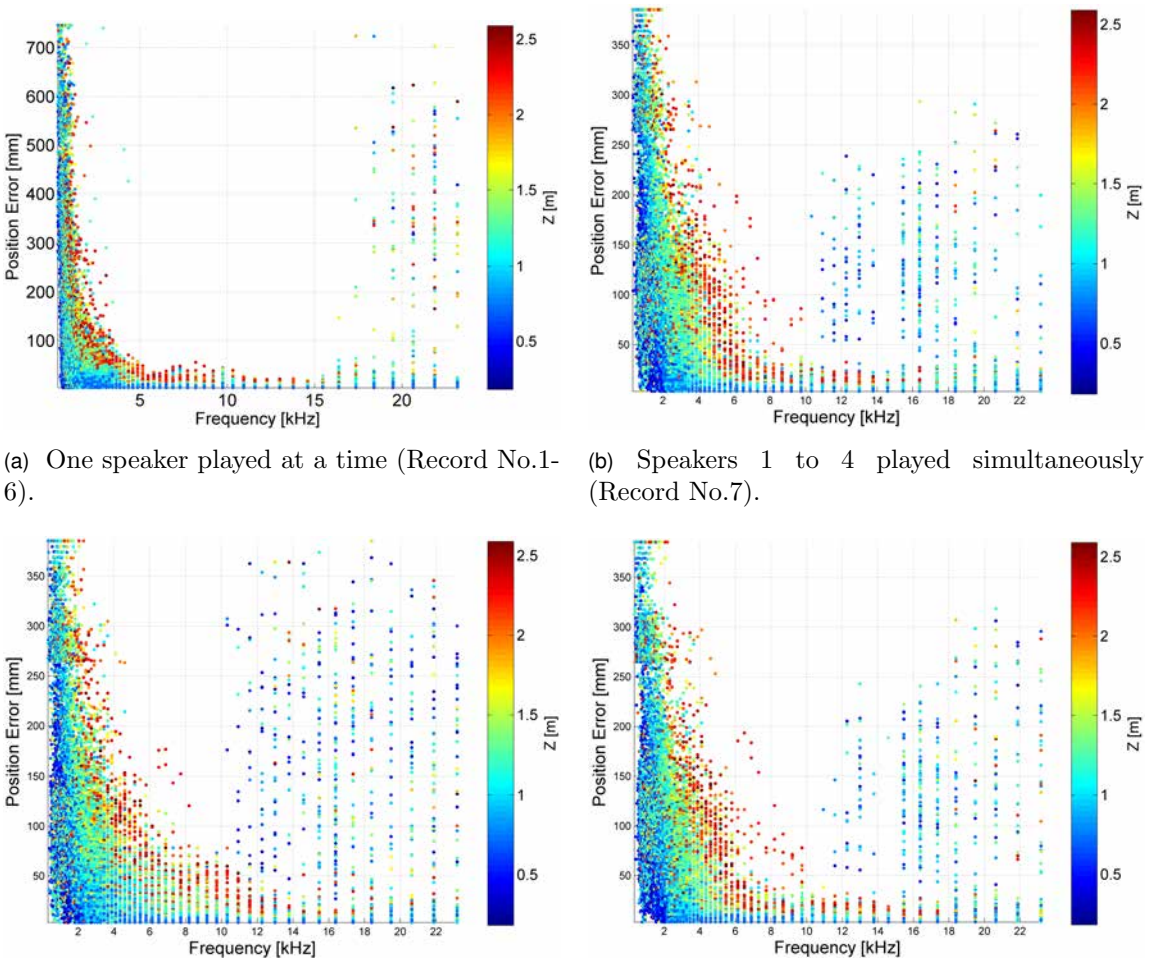


Figure 7.21: Measured normalised *CLEAN-SC* peak intensities $\boldsymbol{b}_{max}$ for scanning surfaces corresponding to the real 3D geometry of the object (Record No.1-6). The data for each sound location have been normalised by obtaining the peak intensity value from this sound source location for all frequencies and subtracting this value, in dB, from the peak values at other frequencies.

7.5.3 Scanning Surfaces Corresponding to the 3D Geometry of the Object

CLEAN-SC was then investigated for scanning surfaces corresponding to the 3D geometry of the object. Figure 7.21 shows the *CLEAN-SC* peak intensity values calculated using distance weighted steering vectors described in Equation (2.17). The intensity values reduce by about 11dB/m as the distance of the sound source from the array increases. This shows that insufficient spherical spreading correction is being used in the steering vector. Similar data, combined with the time of arrival information obtained during calibration, could be used to derive an experimentally determined propagation vector and steering vector equations.

The *CLEAN-SC* position errors obtained are plotted in Figure 7.22 for the different record numbers. These position errors were obtained using distance weighted steering vectors calculated using Figure 2.17. In Figure 7.23a, the position errors, where sound was played on only one speaker at a time, have been averaged into $1/12^{th}$ octave frequency bands. This graph shows the position errors obtained using both normalised and distance weighted steering vectors given respectively in Equations (2.15) and (2.17). Also shown in Figure 7.23a are simulation for both steering vectors. These simulations were performed for the first eight positions of the calibration rig giving 48 sound source locations. The



(a) One speaker played at a time (Record No.1-6).

(b) Speakers 1 to 4 played simultaneously (Record No.7).

(c) Speakers 1 to 6 played simultaneously (Record No.8).

(d) Speakers 1 to 4 played simultaneously using microcontrollers (Record No.9).

Figure 7.22: *CLEAN-SC* position errors for scanning surfaces corresponding to the real 3D geometry of the object.

Table 7.2: 3D Scanning Surface: Averaged *CLEAN-SC* position errors in mm for white noise played on one speaker at a time using the same perspective projection in the error calculation as used for the 2D scanning surface. The averaging has been performed in frequency ranges (rows) and into distances ranges from the center of the array (columns).

(a) Distance weighted steering vectors.

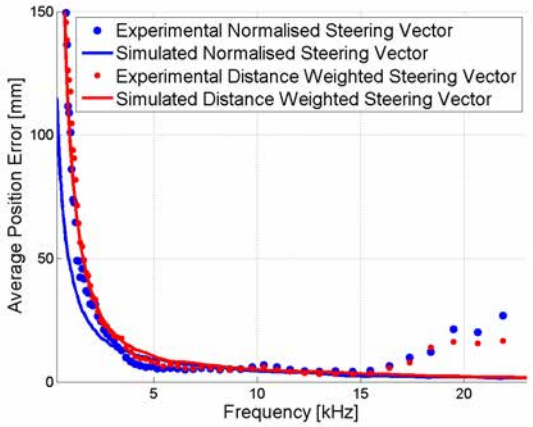
	0.5-1.0m	**1.0-1.5**m	**1.5-2.0**m	**2.0-2.5**m	**2.5-3.0**m
0.3-1kHz	200	220	240	260	260
1-2kHz	44	60	81	100	97
2-5kHz	11	17	23	36	39
5-15kHzx	3.3	4.6	6.2	8.4	11
15-20kHz	7.1	7.4	12	17	13
20-24kHz	6.5	14	17	30	50

(b) Normalised Steering Vectors.

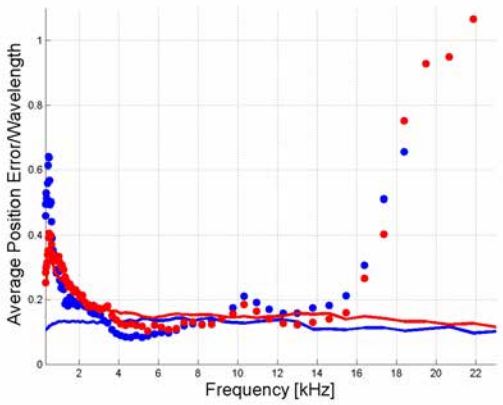
	0.5-1.0m	**1.0-1.5**m	**1.5-2.0**m	**2.0-2.5**m	**2.5-3.0**m
0.3-1kHz	260	320	360	380	440
1-2kHz	30	47	60	98	120
2-5kHz	12	14	18	28	34
5-15kHz	4.3	4.7	5.8	7.2	11
15-20kHz	8.7	8.3	15	16	21
20-24kHz	11	15	28	46	73

microphone data were simulated by calculating uncorrelated white noise signals for each speaker. For a given sound source location, these signals were then delayed and attenuated in the frequency domain for each microphone using Equation (2.2). The propagation vectors were calculated using Equation (2.4). A different frequency independant dependant gain was applied to each microphone channel. This was the only channel noise added to the microphone data. The microphone data were then converted into the time domain. The simulated microphone data were processed in the same manner as the experimental data. To investigate the high position error at low frequencies, additional simulations were performed. For example, summations were made with no variation in gain of the microphones, using $1/3^{rd}$ octave bands, using linear spaced frequency bands, using summation of sine waves with a single frequency components at the center frequencies each octave band, and and a smaller *CLEAN-SC* $\boldsymbol{\varphi}$. These simulations gave similar results to those plotted.

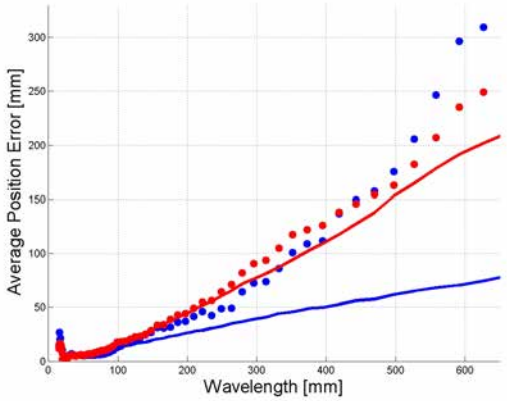
The mean Euclidean experimental position errors are summarised in Table 7.2 for six frequency ranges and five ranges of distances of sound sources from the center of the array. From 5-15kHz the average experimental position error increased with distance from the array from 3.3 to 11 mm with an average over all distances of 6.6mm. This is about $1/4^{th}$ of the diameter of the speaker and is close to the 6mm scanning grid spacing. Outside this frequency range higher position errors were obtained. The simulated average position errors indicate that the higher error for the 15 to 23kHz frequency range results from a



(a) Average position error as function of frequency.



(b) Average position error divided by wavelength as function of frequency.



(c) Average position error as function of wavelength.

Figure 7.23: Experimental and simulated *CLEAN-SC* position error for 3D scanning surfaces and one speaker played at a time (Record No.1-6). The mean *CLEAN-SC* Euclidean position error using 3D scanning surface was 3.7mm for frequencies between 5 and 15kHz. This is 1/7 of the 26 mm diameter of the speaker and only just larger than half the scanning grid spacing of 6mm.

low signal to noise ratio due to the low gain of the speaker at these frequencies, see Figures 7.21 and 9.2. The wavelength dependance of the position errors is investigated in Figures 7.23b and 7.23c. These indicate that the position errors is proportional to wavelength. In Figure 7.24, the position error for all sound source locations is plotted as a function of distance from the array. This shows a position error which is also proportional to wavelength λ and to the distance R of the sound source from the array. These results are related to the beamwidth of the array which is described in Equation (2.27) by

$$BW_R \propto \frac{R\,\lambda}{D},$$

where D is the array diameter. In addition to this, however, the increase in *CLEAN-SC*

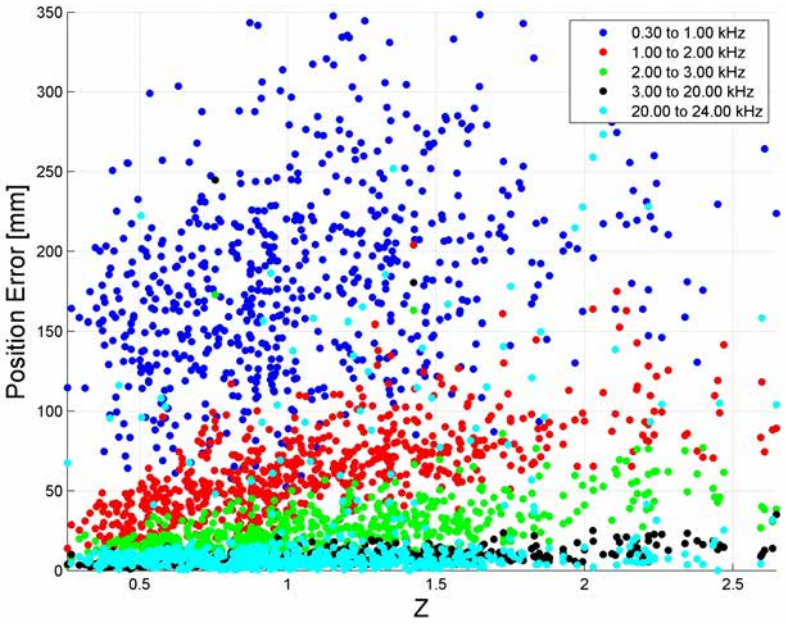


Figure 7.24: Position error as a function of distance Z from array for several frequency ranges for scanning surfaces corresponding to the real 3D geometry of the object (Record No.1-6).

position error with distance will be partially associated with decreased accuracy with distance of the camera extrinsic parameters, which were used to obtain the coordinates of the speakers and the position and orientation of the 3D scanning surfaces, see Figure 7.12.

The relationship between the magnitude of position errors and the beamwidth is illustrated by the comparison between normalised steering vectors and distance weighted steering vectors shown in Figure 7.23a. The distance weighted steering vectors place more weighting on scan points further from the array in an attempt to correct for spherical spreading. This can have the effect or biassing the position of the beamformed map peaks towards scan points further from the array, see Figure 9.4. At small wavelengths the beamwidth is small, meaning that the beamforming peak falls off rapidly with distance from the sound source location. Unless there is a high variation in distance scan points from of the array, relatively little offset of the peak will occur. However, at long wavelengths the beamwidth is large, meaning that the beamforming peak falls off by a small amount with distance from the sound source location. In this case, even if there is only a small variation in 3D distance of the scan points from the array, the distance weighting may dominate. If one cycles through beamforming maps starting from long wavelengths to short one will see the beamforming peak start at the scan point furthest from the array and move slowly toward sound source location as the beamwidth reduces. Normalised steering vectors do not have this same biasing. For this steering vector, assuming there is some form of unwanted noise in the microphone data, the same sequence of plots might show the beamforming peak jumping around apparently randomly and slowly converge

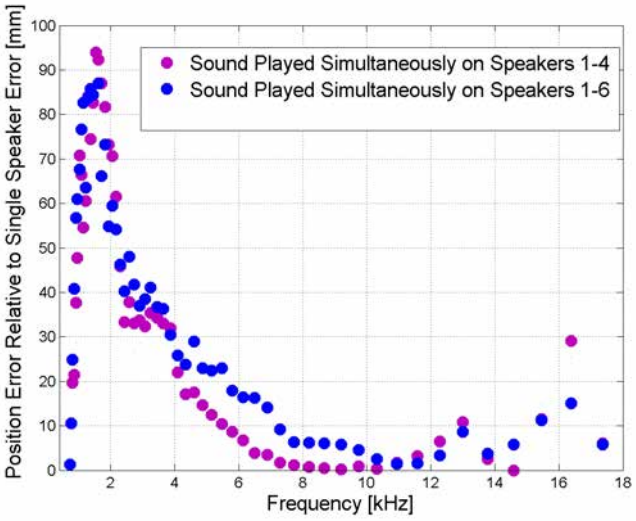


Figure 7.25: Difference in position error for sound played on multiple speakers simultaneously (Record No.7 and 8) relative to the position error obtained for a single speaker played at a time (Record No.1-6).

on the sound source location as the beamwidth reduces. Because *CLEAN-SC* uses the beamforming peak location, these errors in the beamforming peak location is passed onto *CLEAN-SC*.

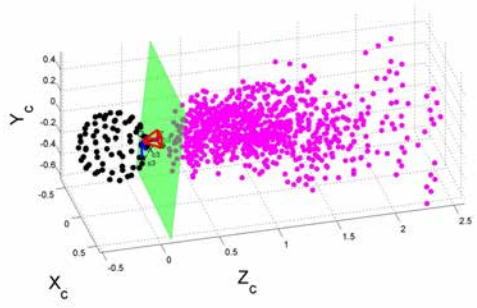
The beamwidth dependency on wavelength also effects the ability of a beamforming to separate closely spaced sound sources. This may be expressed in terms of the Rayleigh criterion described in Equation (2.28)

$$\mathcal{W} = \frac{d\ D}{\lambda\ R}$$

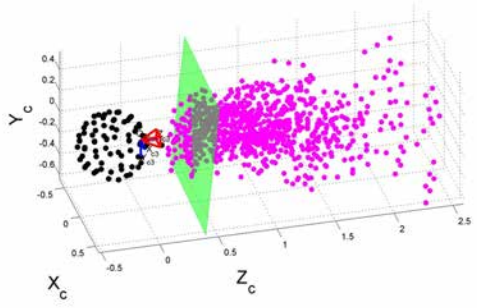
where d is the separation of the sound sources. Figure 7.25 shows the difference in position error where multiple speakers were played simultaneously compared to when only one speaker was played at a time.

7.5.4 Comparison of the Position Error for 3D Scanning Surfaces and Traditional 2D Scanning Surfaces

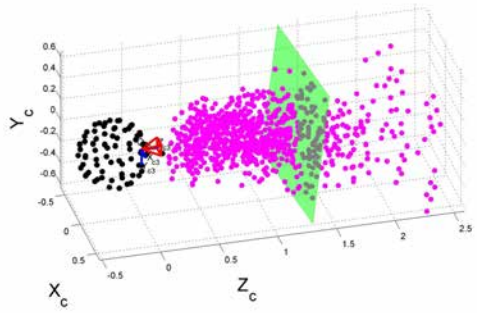
The *CLEAN-SC* position error for traditional 2D scanning surfaces, described in Section 7.5.2, used a single source and multiple traditional 2D scanning surfaces. In this section, about 840 sound sources were images using a single traditional 2D scanning surface at a time, see Figure 7.15. Although 23 of these traditional 2D scanning surfaces were used, the results from only five of these are shown in the plots below, see Table 7.3. The traditional 2D scanning surfaces were generated using the extrinsic parameters obtained



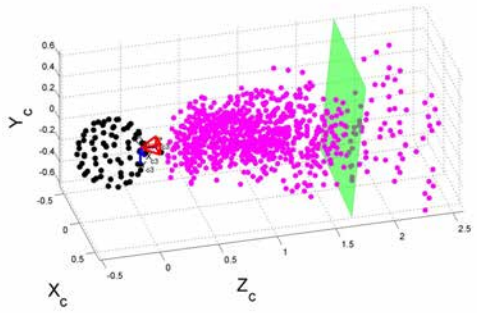
(a) 2D Scanning Surface No.1, $Z = 0.164$.



(b) 2D Scanning Surface No.6, $Z = 0.466$.



(c) 2D Scanning Surface No.15, $Z = 1.305$.



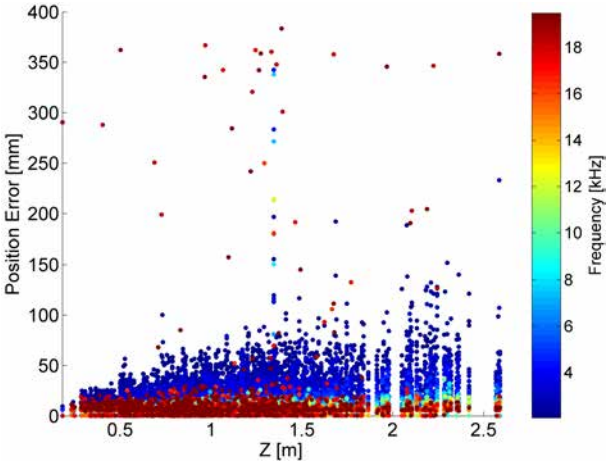
(d) 2D Scanning Surface No.19, $Z = 1.717$.

Figure 7.26: Example 2D scanning surface locations and orientations (green planes) and sound source locations (magenta coloured dots) relative to the microphone coordinates (black dots).

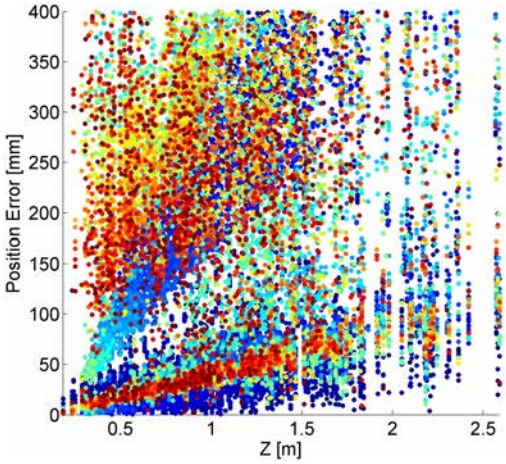
from camera images if the checkerboard pattern on the speaker rig, see Section 7.4.2. Therefore, the scanning surfaces are not necessarily perfectly perpendicular to the array axis. The position error obtained using these traditional 2D scanning surface are compared with that obtained in Section 7.5.3 where the 3D geometry of the object was used as a scanning surface.

The position error as a function of the Z component of the speaker positions was further investigated in Figure 7.27. In Figure 7.28 the same errors are plotted as a function of offset dR of the speaker from the traditional 2D scanning surface. To simplify the plots only frequencies from 2 to 20kHz have been plotted. Here it can be seen that poor performance was achieved for 2D scanning surface No.1 and No.6 ($R = 0.40$ and 0.70 m), for all speaker distances except for speaker coordinates near the 2D scanning plane. For 2D scanning surface No.15, No.19, and No.23 ($R = 1.54$, 1.94, and 2.36m), the position error was comparable to the 3D case for speakers located further than about 1 to 1.2m from the center of the array.

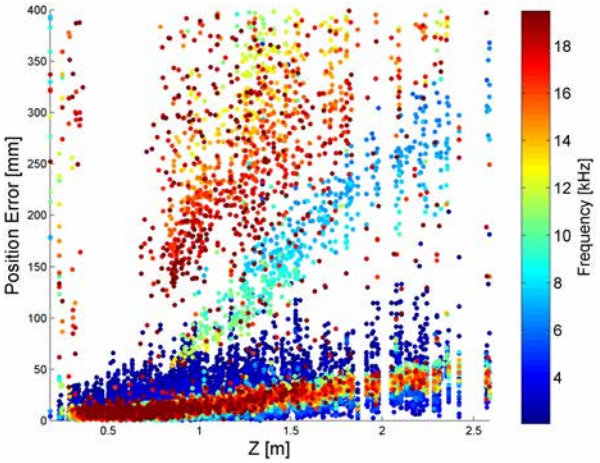
The *CLEAN-SC* position error was investigated as function of frequency for both the



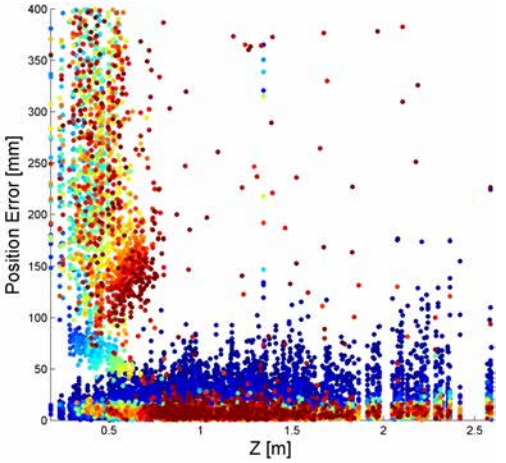
(a) True 3D scanning surface.



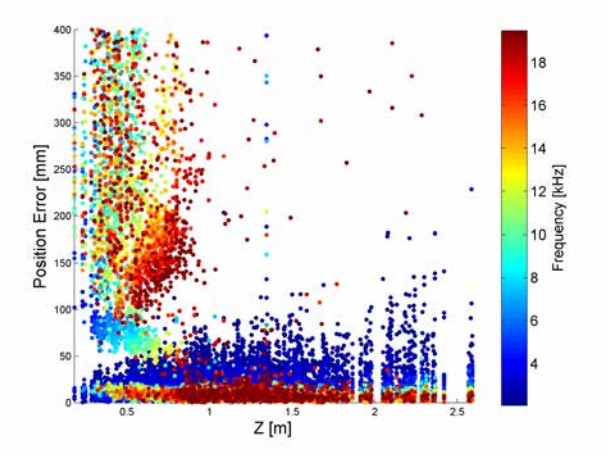
(b) Traditional 2D scanning surface No.1, $\vec{T}_3 = 0.164$ $(R = 0.40)$ m.



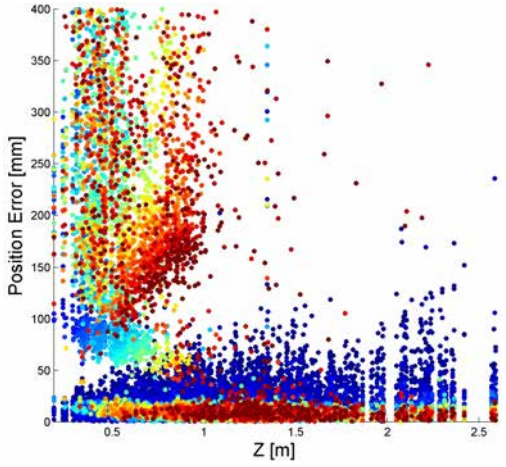
(c) Traditional 2D scanning surface No.6, $\vec{T}_3 = 0.466$ $(R = 0.70)$ m.



(d) Traditional 2D scanning surface No.15, $\vec{T}_3 = 1.305$ $(R = 1.54)$ m.

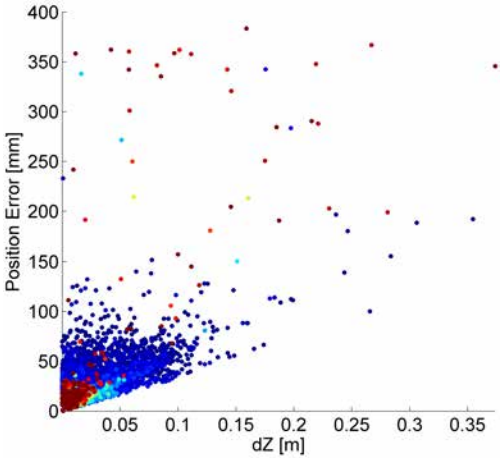


(e) Traditional 2D scanning surface No.19, $\vec{T}_3 = 1.717$ $(R = 1.94)$ m.

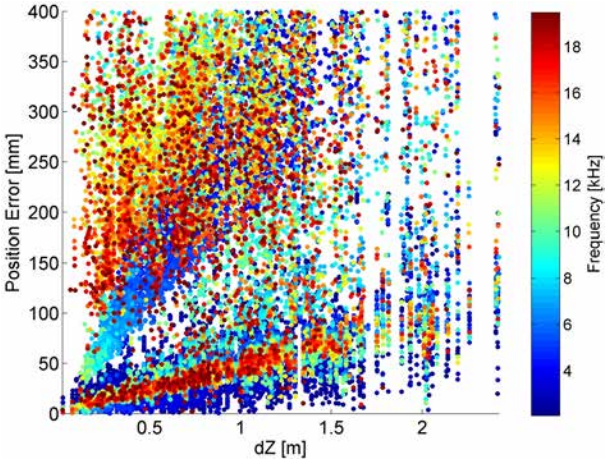


(f) Traditional 2D scanning surface No.23, $\vec{T}_3 = 2.120$ $(R = 2.36)$ m.

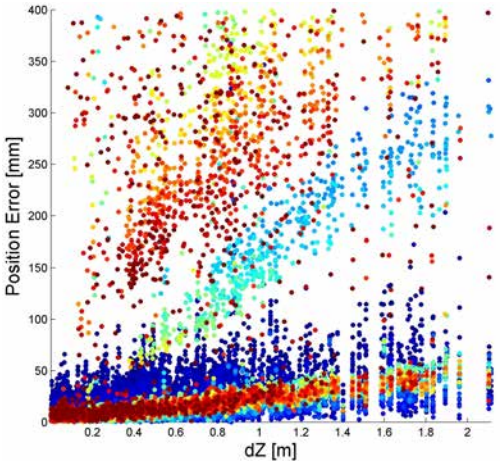
Figure 7.27: Comparison of position error as a function of distance Z in the reference frame of the camera at the front of the array) for the true 3D scanning surface and five traditional 2D scanning surface. The frequency range used was between 2 and 18kHz. (Record No.1-6)



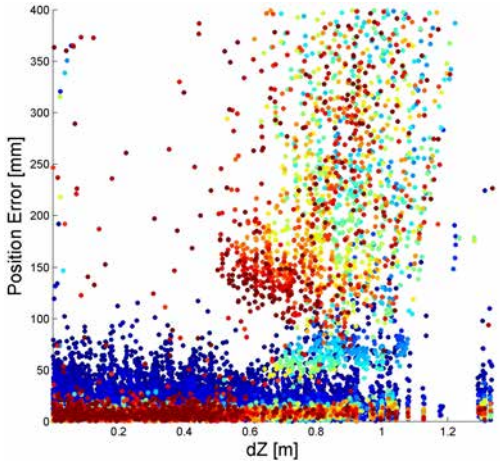
(a) True 3D scanning surface.



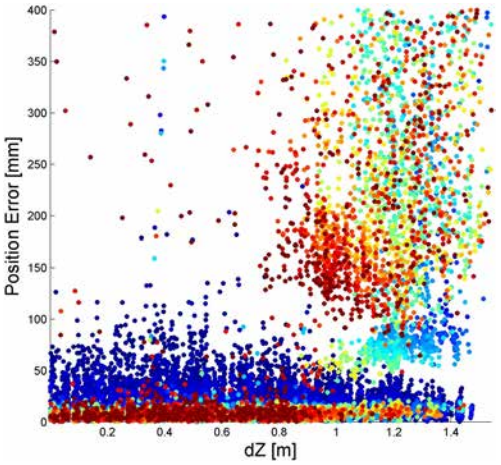
(b) Traditional 2D scanning surface No.1, $\vec{\boldsymbol{T}}_3 = 0.164$ $(R = 0.40)$ m.



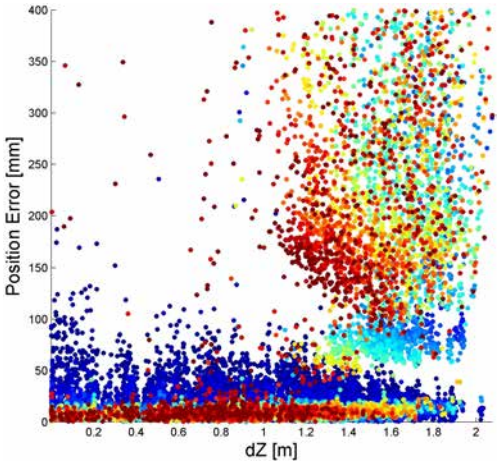
(c) Traditional 2D scanning surface No.6, $\vec{\boldsymbol{T}}_3 = 0.466$ $(R = 0.70)$ m.



(d) Traditional 2D scanning surface No.15, $\vec{\boldsymbol{T}}_3 = 1.305$ $(R = 1.54)$ m.



(e) Traditional 2D scanning surface No.19, $\vec{\boldsymbol{T}}_3 = 1.717$ $(R = 1.94)$ m.



(f) Traditional 2D scanning surface No.23, $\vec{\boldsymbol{T}}_3 = 2.120$ $(R = 2.36)$ m.

Figure 7.28: Comparison of position error as a function of offset $d\,R$ of a sound source from the traditional 2D scanning surface. The frequency range used was between 2 and 18kHz. (Record No.1-6)

2D Scanning Surface Number	R (Distance from the center of the array.) [m]
1	0.40
6	0.70
15	1.54
19	1.94
23	2.36

Table 7.3: Traditional 2D scanning surfaces numbers and their approximated Z component coordinates in the reference frame of the main camera located front of the array. The center of the array was located at 235mm behind the camera.

2D and 3D case. An average of the position error was made for all speaker positions for each 12^{th} octave frequency band. This averaged error was then plotted as a function of frequency for a white noise played on only a single speaker (Record No.1-6), see Figure 7.29a, and for white noise played simultaneously on six speakers (Record No.8), see Figure 7.29b. At low frequencies, the position error for the 2D case was the same as that for the 3D case. However, the position error for the traditional 2D scanning surfaces increased significantly above that of the 3D case for frequencies above 4.78kHz. An exception to this was for 2D scanning surface No.1 which departed from the 3D case at 1.4kHz. An interesting feature observed was that the position error for Record No.1-6, at lower frequencies was roughly linear if plotted on a loglog plot (X and Y axes are logarithmic), see Figure 7.29c. However, the position error for Record No.7-9, was approximately linear on a 'semilogy' plot (Y axis is logarithmic and X axis is linear), see Figure 7.29d. Up to about 5kHz for the 2D case and 10kHz for the 3D case, the position error for a single speaker could be modelled as

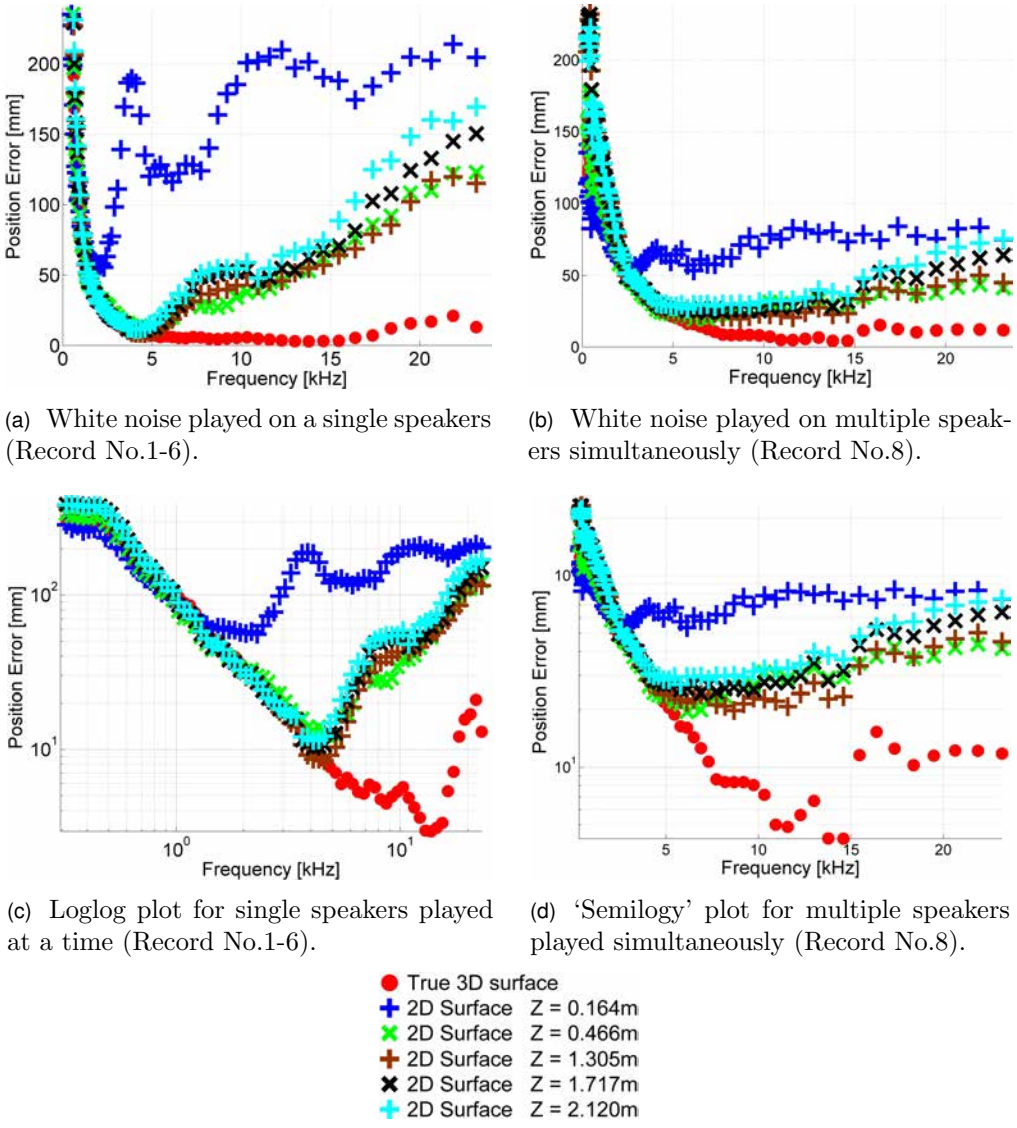
$$\varepsilon = e^b \, f^a \tag{7.27}$$

while the position error for multiple speakers could be modelled as

$$\varepsilon = e^{a\, f+b} \tag{7.28}$$

where a is the slope and b is the intercept.

Figure 7.30 also shows averaged position errors as a function of frequency. However, instead of making the average over all sound source location, it performs the frequency averaging by divides the errors into groups based on the distance of the sound source from the array. This averaging is performed for both the 3D and five of the 2D scanning surfaces for recordings where sound was played on only one speaker at a time. The results shown in this figure are summarised in Table 7.4. This table gives the percentage difference in the error for the 2D scanning surface compared to the corresponding position errors obtained for the 3D case given in Table 7.2(a). This shows that the position errors,



(a) White noise played on a single speakers (Record No.1-6).

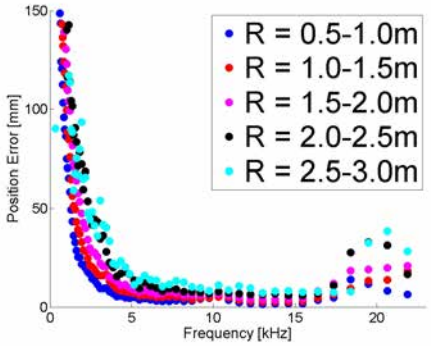
(b) White noise played on multiple speakers simultaneously (Record No.8).

(c) Loglog plot for single speakers played at a time (Record No.1-6).

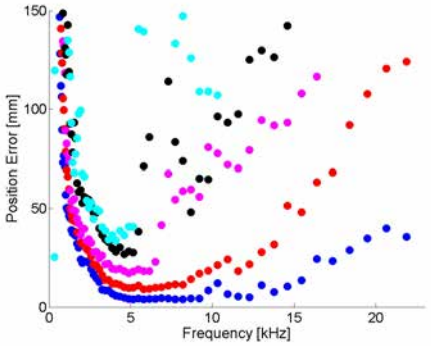
(d) 'Semilogy' plot for multiple speakers played simultaneously (Record No.8).

Figure 7.29: Position error averaged averaged over all sound source locations into 12^{th} octave frequency bands.

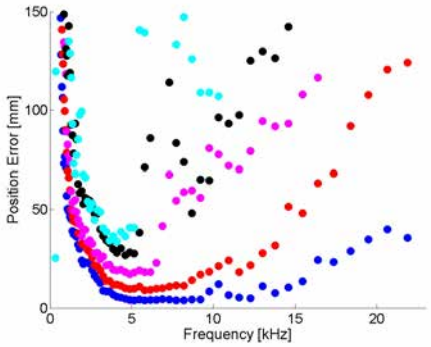
obtained using 2D scanning surface, were significantly larger than that obtained using 3D scanning surfaces except when the wavelength and the distance of the sound source from the array were large. As the distance of the scanning surface from the array increases, comparable results to the 3D case were obtained at increasingly smaller wavelengths. A position error dependance on the offset of the 2D scanning surface from the sound source location can also be seen. An interesting result seen is that in some cases, better results were obtained using the 2D scanning surface than that obtained using a 3D scanning surface. The cause of this has not been investigated. It is possible it may be be related to noise in the system or differences in the locations of the scan points used for the two types of scanning surfaces.



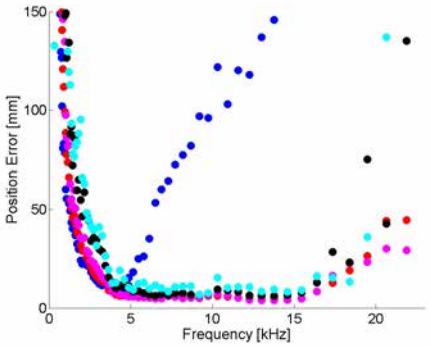
(a) 3D Scanning Surface.



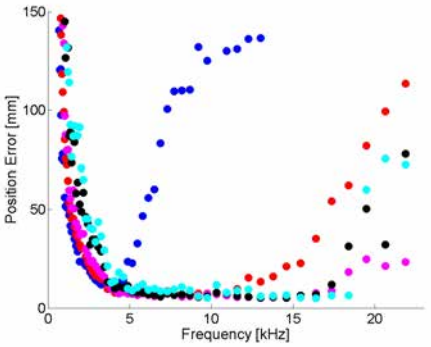
(b) 2D Scanning Surface No.1: $R = 0.40$m.



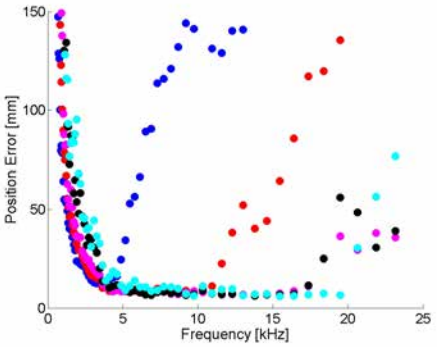
(c) 2D Scanning Surface No.6: $R = 0.70$m.



(d) 2D Scanning Surface No.15: $R = 1.54$m.



(e) 2D Scanning Surface No.19: $R = 1.94$m.



(f) 2D Scanning Surface No.23: $R = 2.36$m.

Figure 7.30: *CLEAN-SC* averaged position errors for recordings where sound was played on a single speaker at a time for 3D and 2D scanning surfaces. The position errors were divided into five groups based on the distance of the sound source from the center of the array. Each group of position errors were then averaged into $1/12^{th}$ octave frequency bands.

7.5.5 Summary

The deconvolution algorithm *CLEAN-SC* was investigated using 3D scanning surfaces corresponding to the surface geometry of an object, and compared to results obtained for traditional 2D scanning surfaces. It was found that using 3D scanning surfaces corresponding to the surface geometry of an object gave under certain conditions large improvements in the accuracy of measured sound source magnitudes and positions compared to those obtained using the traditional 2D scanning surface. The greatest benefit of using the 3D scanning surface was achieved for smaller wavelengths and sound sources located near the array. These experimental results are consistent with the theory presented in Section 7.1.1. Equation (7.13) gave the approximation for the beamforming phase error

$$\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\boldsymbol{\xi}}_n) \approx \frac{\pi D^2}{\lambda}\left[\frac{1}{R} - \frac{1}{r}\right]$$

for a sound source at located at $\vec{\boldsymbol{X}}_s$ and a 2D scanning surface at r. Like the experimentally measured position errors, the phase error becomes small as λ and distances R and r increase and as the separation between the sound source location and scanning surface $\|R - r\|$ decreases. An interesting feature observed when using a 3D scanning surface was that increased accuracy for lower frequencies was achieved as the sound location was moved closer to the array. This was in contrast to the 2D case where high errors was observed for sound sources located near the array.

Future work would repeat the experiment described above for the Underbrink multi-arm spiral array and the *SADA* array. The improvement in accuracy achieved using the 3D method, for the spherical array, is expected to be significantly greater for the Underbrink multi-arm spiral array, which has shorter depth of field. Other deconvolution algorithms could also be investigated using the same methodology.

Table 7.4: Percentage difference in averaged *CLEAN-SC* position error for 2D scanning surfaces compared to the 3D scanning surface average errors given in Table 7.2(a). The averaging was performed by using different ranges of frequency and distance R of sound sources from of the center of the array for. These were calculated using $\varepsilon(\%) = (\varepsilon_{2D} - \varepsilon_{3D})/\varepsilon_{3D} \times 100$. A positive value means the 2D position error was higher than for the 3D case.

(a) 2D Scanning Surface No.1: $R = 0.40$m.

	0.5-1.0m	**1.0-1.5**m	**1.5-2.0**m	**2.0-2.5**m	**2.5-3.0**m
0.3-1kHz	-20	-10	-3.3	-3.9	2.8
1-2kHz	-11	1.9	7.6	28	68
2-5kHz	280	590	650	540	550
5-15kHz	3400	3200	2800	2600	2500
15-20kHz	2000	2200	1600	1400	2300
20-24kHz	2700	1300	1300	840	360

(b) 2D Scanning Surface No.6: $R = 0.70$m.

	0.5-1.0m	**1.0-1.5**m	**1.5-2.0**m	**2.0-2.5**m	**2.5-3.0**m
0.3-1kHz	5.8	4.0	0.2	-0.7	-8.0
1-2kHz	-8.7	-15	-27	-6.6	12
2-5kHz	16	16	16	15	18
5-15kHz	81	290	830	1000	1400
15-20kHz	250	930	1000	980	1700
20-24kHz	510	810	980	640	370

(c) 2D Scanning Surface No.15: $R = 1.54$m.

	0.5-1.0m	**1.0-1.5**m	**1.5-2.0**m	**2.0-2.5**m	**2.5-3.0**m
0.3-1kHz	26	22	14	11	2.3
1-2kHz	-2.2	-9.5	-23	-2.4	11
2-5kHz	29	-6.8	-20	-20	-15
5-15kHz	2500	23	-15	-12	-6.6
15-20kHz	2700	95	22	75	42
20-24kHz	3700	230	53	180	220

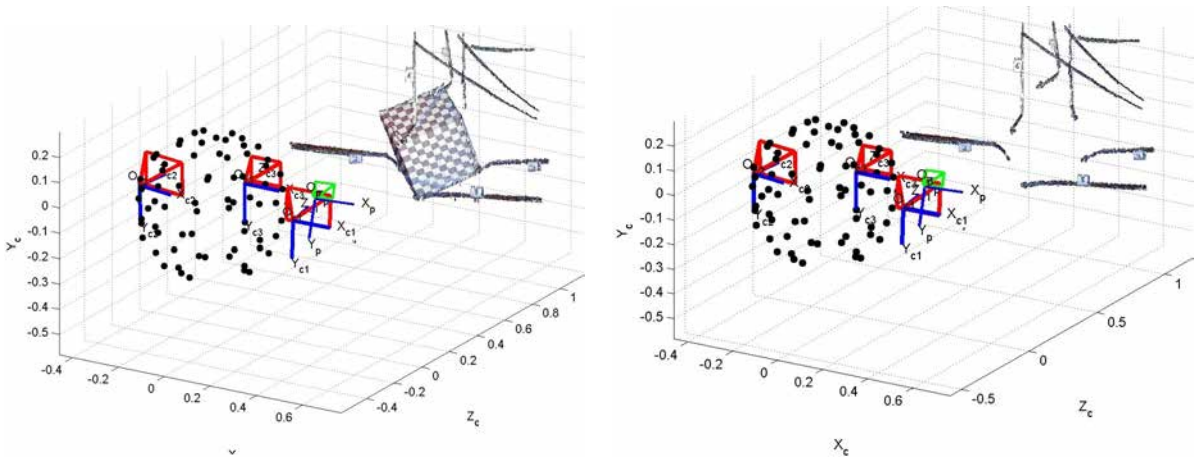
(d) 2D Scanning Surface No.19: $R = 1.94$m.

	0.5-1.0m	**1.0-1.5**m	**1.5-2.0**m	**2.0-2.5**m	**2.5-3.0**m
0.3-1kHz	27	24	17	12	6.7
1-2kHz	-6.9	-12	-24	-3.7	14
2-5kHz	46	-0.8	-16	-21	-13
5-15kHz	3100	106	1.9	-21	-21
15-20kHz	3000	590	5.2	27	31
20-24kHz	3800	730	29	69	20

(e) 2D Scanning Surface No.23: $R = 2.36$m.

	0.5-1.0m	**1.0-1.5**m	**1.5-2.0**m	**2.0-2.5**m	**2.5-3.0**m
0.3-1kHz	26	26	19	12	4.6
1-2kHz	-2.0	-8.6	-22	-1.5	13
2-5kHz	52	1.2	-15	-21	-12
5-15kHz	340	268	24	-11	-15
15-20kHz	3001	1300	43	27	-48
20-24kHz	3800	1100	110	29	8.5

7.6 Investigation of Scattering Effects



(a) Cube and speakers.

(b) Cube removed leaving speakers in place.

Figure 7.31: Structured light scans for investigating the effect scattering of the sound by a 3D object, the cube, for beamforming and deconvolution.

Current beamforming and deconvolution techniques assume free field propagation. Scattering effects are not included in the Green's functions used for beamforming and deconvolution (Dougherty, 2010b). Although not yet analysed, initial raw data measurements have been made for investigating scattering effects relating to the 3D surface of a geometry of an object, using methodology suggested to the author by Dougherty.

The small cardboard cube was setup in front of the array. Speakers were positioned against the surface of the cube. These speakers were not attached to the cube but instead mounted on rods attached to the acoustic foam enclosure framework. The cube could be removed without disturbing the speakers or the array. Microphone recordings of white noise played on the speakers were then obtained with the cube in place and a structured light scan made, see Figure 7.31. The cube was then removed and another set of microphone recordings obtained.

Beamforming and deconvolution could then be performed, using the 3D structured light scan as a scanning surface corresponding to the surface geometry of the cube. The same scanning surface would be used for both recordings where the cube was in place and recordings where the cube had been removed. The two deconvolution maps would then be compared to identify any scattering effect related to the surface geometry of the cube. Also, the acoustic propagation time from the speaker to the microphones would be measured using the time of flight technique described in Section 6.3.5 and be used as a comparison to any theoretical time delays which included scattering. These measurements were to be performed with the spherical array, Underbrink multi-arm spiral array, and

SADA array. This could also be investigated with the array outside and inside the 1.2m diameter sphere and cube, refer to Figure 3.12.

8

Summary, Conclusions, and Future Work

8.1 Summary

8.1.1 Methodology

The objective of this work was to investigate deconvolution of beamforming maps generated using a scanning surface corresponding to the 3D surface geometry of an object being acoustically imaged and contrast the results with those obtained using the traditional 2D method. The tasks initially identified were:

- Design and build microphone arrays and 3D objects.
- Write software to perform beamforming and deconvolution.
- Obtain a methodology for:
  - Calibrating a camera in the array, aligning the camera with the microphones array structure, and plotting acoustic maps, as transparencies, over camera images.

- Measuring the surface geometry of an object in the array's reference frame.
- Calculating the deconvolution position and magnitude error.

A 72 element microphone array and 3D sound source objects were designed and built. The array was designed to be versatile allowing the microphones to be mounted into different array structures. Three array structures were constructed; an Underbrink multi-arm spiral, a spherical array, and a SADA. These array types were chosen because they were well established designs and had respectively a high, medium, and poor longitudinal beamforming resolution. Software was written in *MATLAB* to capture the microphone and camera raw data, save it to file, and perform beamforming and deconvolution using the *CLEAN-SC* and *DAMAS* algorithms in octave frequency bands.

The initial camera calibration technique, that was used by the author with prototype arrays, had been a pinhole model. A search for a better method was made and the computer vision camera calibration codes *Camera Calibration Toolbox for Matlab* (Bouguet, 2008) and *OpenCV* (Bradski and Kaehler, 2008) were found. These were used for calibration of the cameras in the array and for projection/plotting of the 3D acoustic maps onto a camera image. They also provided a methodology for obtaining the surface geometry of an object. If the surface geometry of the object was known in an object reference frame, a checkerboard pattern was attached to the object. Beamforming and deconvolution scan points, located on the surface of the object, were defined in this checkerboard reference frame. An image was captured of the checkerboard pattern by a camera in the array. This image was then used by computer vision software to obtain the extrinsic parameters of the checkerboard pattern. These parameters were used to convert beamforming and deconvolution scan point coordinates from the checkerboard reference frame into the camera reference frame using a rigid body transformation.

A second method used to obtain the 3D surface geometry of the object in the camera reference frame was structured light scanning. This was implemented by adding a data projector to the microphone array setup and using code written by Lanman and Taubin (2009b). This provided a 3D point cloud on the surface of the object in the reference frame of the main camera in the array. This point cloud was used for beamforming and deconvolution scan points after removing isolated points, which were assumed to be noise, and reducing the spatial resolution of the scan points. The current structured light system only has a limited field of view. Also, holes in the scan occur for parts of the object where a camera in the array cannot see light projected by the data projector. An advantage of the structured light system, however, over a laser or contact measurement system is that it is capable of imaging dynamic objects if a colour coded structured light scan or Xbox

Kinect sensor is used. It is also more robust than a stereoscopic systems which generally have problems with pixel matching between camera pairs.

The 3D scan points corresponding to the surface geometry of the object were obtained in the reference frame of the main camera in the array. In order to use these scan points for beamforming and deconvolution, the microphone coordinates also needed to be able to be obtained in the reference frame of the main camera in the array. A new method was developed which combined camera calibration with microphone position calibration. A calibration rig was used which consisted of a checkerboard pattern glued onto a piece of perspex and speakers inserted into holes, laser cut in the perspex, at set locations surrounding the checkerboard pattern. Camera images of a checkerboard pattern were obtained for a range of positions and orientations of the calibration rig. For each of these images, a microphone recording was obtained of white noise played on one speaker at a time. These recordings were used to obtain time of flight data using a sub-sample cross correlation method. Speed of sound measurements were also able to be obtained. These microphone coordinates, time of flight, and speed of sound data were used to obtain the microphone coordinates in the camera reference frame. This method allows microphones and cameras to be set up in arbitrary positions to obtain the information required for generating beamformed and deconvolution acoustic maps and plotting these as transparencies over camera images.

Experiments comparing the *CLEAN-SC* acoustic maps for 2D and 3D scanning surfaces have been performed using the Underbrink multi-arm spiral array. The results obtained give accurate results for the 3D case. For 2D scanning surfaces located near the true sound source location, similar results to the 3D case are achieved. However, as the 2D scanning surface is removed away from the true sound source location the measured sound pressure level reduces and the position error increases. The results obtained were interpretive, however, and it was decided that a more systematic method of obtaining the *CLEAN-SC* errors was required.

A technique was developed to compare the accuracy of the *CLEAN-SC* deconvolution algorithm obtained using traditional 2D scanning surfaces with that obtained using scanning surfaces corresponding to the 3D surface geometry of an object. The microphone recordings, obtained during the calibration procedure, were used for beamforming and deconvolution. For each position of the calibration rig, nine microphone recordings were obtained where white noise was played on only one speaker at a time or on four or six speakers simultaneously. The position of the sound sources in the camera reference frame, calculated during the calibration, were used to measure the accuracy of *CLEAN-SC* peak intensity coordinates. For traditional 2D scanning surfaces, these peak intensity coordinates were projected to the same distance, in the Z axis direction, from the array as the

speaker. The most accurate projection method evaluated was perspective projection with the centre of projection being the centre of the microphone coordinates. By automatically combining deconvolution data from 1260 recordings and using 76 frequencies in 12^{th} octave bands from 0.3 to 24kHz, a dense point cloud of *CLEAN-SC* errors was generated. Analysis of the *CLEAN-SC* data were divided into four groups: recordings where white noise, which was generated using the DT9836 analog outputs, was played on one speaker, four, or six speakers simultaneously or where white noise, which was generated using independent microcontrollers, was played on four speakers simultaneously. The *CLEAN-SC* error analysis using a single speaker is considered to be equivalent to beamforming error analysis for a single speaker.

8.1.2 Results

The *CLEAN-SC* error analysis was performed using the above method for a spherical array. This array was used because it had a long depth of field and, therefore, was represented a near worst case scenario for the 3D method, compared with the Underbrink multi-arm spiral. The spherical array's depth of field, which is the reduction in measured sound pressure level as the 2D scanning is offset from the array, was measured for a single speaker located 0.773m from the array and using 23 traditional 2D scanning surfaces at distances from 0.164 to 2.121m from the array. The measured *CLEAN-SC* peak intensity was found to reduce with displacement of the traditional 2D scanning surface from the sound source location. The same data were also used for obtaining the *CLEAN-SC* position error as a function of the displacement of the traditional 2D scanning surface from the true sound source location.

The *CLEAN-SC* position error for about 840 sound sources was then obtained using scanning surfaces corresponding to the 3D geometry of the object. For recording where sound was played on only one speaker at a time, the position errors showed a roughly linear increase with wavelength λ and distance R of a sound source from the array. This is related to the beamwidth

$$BW_R \propto \frac{R\,\lambda}{D}.$$

Part of the increase in position error with distance could also have been related to the increase in error with distance of the camera extrinsic parameters used to identify the location of sound sources and obtain the 3D scanning surface. However, based on simulations performed it is not thought to be a dominant source of error. For frequencies from 5-15kHz, the average position error was about 6.7mm when averaged over all sound source locations. This was less than the 26mm diameter of the speaker used and comparable to the scanning surface grid spacing of 6mm. At higher frequencies, the position error

increased. Since this increased error was not seen in the simulations, it was believed that this increased error was due to poor gain of the speaker at higher frequencies. For recordings where multiple speakers were played simultaneously, the position error was higher at lower frequencies. Closer separation of sound sources required smaller wavelengths to achieve the same position error. This can be related to the Rayleigh criterion

$$\mathcal{W} = \frac{d\ D}{\lambda\ R}$$

where the ability to separate two sound sources is proportional to the the diameter D of the array and inversely proportional to the separation d of the sound sources and the wavelength.

The same 840 sound sources were then acoustically imaged using traditional 2D scanning surfaces at distances of 0.164, 0.466, 1.305, 1.717, and 2.120m from the front of the array. These showed similar results to the 3D scanning surface for frequencies up to about 4.78kHz. Above this frequency, the increased position error could be roughly related to the theoretical approximation for the phase error given in Equation (7.12)

$$\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\boldsymbol{\xi}}_n) \approx \frac{\pi D^2}{\lambda}\left[\frac{1}{R} - \frac{1}{r}\right]$$

where r is the distance from the the center of the array to the 2D scanning surface. The position error increased as the wavelength and the distances R and r increase and as the offset between the scanning surface and sound source increased. Scanning surfaces close to the array only gave accurate results for sound sources located very close to scanning surface. As the 2D scanning surface was moved away from the array, sound sources located at increasing distances from the scanning surface also gave accurate positioning of *CLEAN-SC* peaks, see Table 7.4.

8.2 Conclusions

The results of our 3D work, discussed in Chapter 7, clearly show improved accuracy of measured positions and magnitudes of sound sources under a range of circumstances. It is useful to condense these findings into quantitative terms, albeit for the particular spherical array used here. We will concentrate on position error, since other workers have not developed the methodology for investigating this. Some position error results shown in Figure 7.29a are summarised in Table 8.1 for a single point source. In using 3D methods, compared with 2D methods, there is not much improvement in position

	Z	$f = 5$kHz	$f = 10$kHz
2D	0.5m	15mm	40mm
2D	2m	15mm	50mm
3D		10mm	5mm

Table 8.1: Example *CLEAN-SC* position errors for a single source and 2D and 3D scanning surfaces. The Z values are the distances of each 2D scanning surface from the array.

estimation at lower frequencies, but up to two orders of magnitude smaller position errors for frequencies above 5kHz.

Similarly, some position error results from Figure 7.29b are summarised for multiple sources in Table 8.2. the improvements arising from use of 3D methods are not as large

	Z	$f = 5$kHz	$f = 10$kHz
2D	0.5m	30mm	30mm
2D	2m	35mm	35mm
3D		20mm	10mm

Table 8.2: Example *CLEAN-SC* position errors for multiple sources played simultaneously and 2D and 3D scanning surfaces. The Z values are the distances of each 2D scanning surface from the array.

as for single sources. Again, improvements are larger at higher frequencies.

Figure 7.27, which has results across a wide bandwidth, shows that 2D and 3D methods have comparable position errors for sources further than about 1m from the array, if a 2D scanning surface is used which is also more the a metre from the array.

The principal conclusion from the results summarised in Tables 8.1 and 8.2 is that there are large gains in position estimation using our new 3D methods for frequencies above about 5kHz, and with sound sources closer than about a metre from the array. Improvement at higher frequencies is to be expected, since the 2D phase errors are larger at higher frequencies. The improvement of the 2D case with increased range occurs because the depth of field is larger for larger Z.

The end result is that our new methods will likely find application when (1) the emitting object has significant depth in the longitudinal direction, in comparison with the array-to-object distance, and (2) when the wavelength of emitted sound is smaller than about 70mm for this particular array design. This array type was used since it had a relatively large depth of field. For arrays having a shorter depth of field, such as a spiral array, the improvement of the 3D case might be expected to be significantly greater. This was strongly indicated in the initial experiments using an Underbrink multi-arm spiral, see Figures 7.6 to 7.9 Examples of applications where this method might be used are in laboratory or workshop studies of machines, or in sound-mapping the interior of vehicles. Both of these are typical uses of acoustic arrays.

8.3 Future Work

The error analysis measurements were performed using the spherical array. This should be repeated with the Underbrink multi-arm spiral array and the *SADA* array.

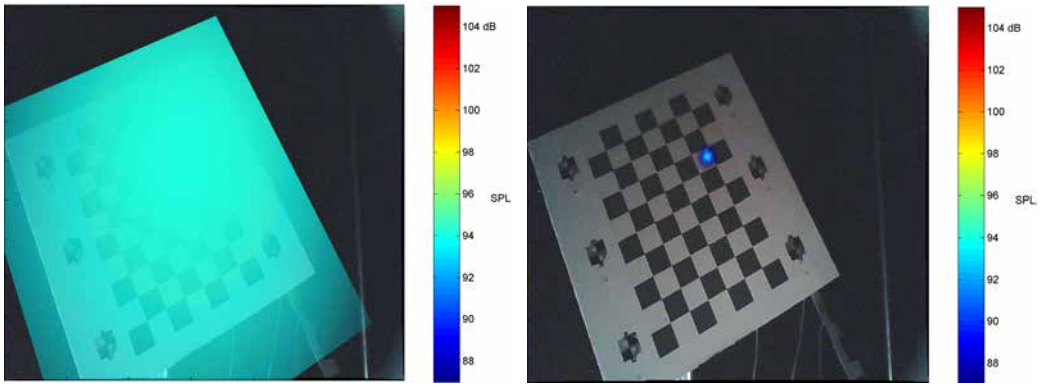
The deconvolution algorithm used in this work was *CLEAN-SC*. Future work would use the same error analysis method and data that was used for *CLEAN-SC* on other deconvolution algorithms such as *DAMAS2*,*DAMAS2*, or *TIDY*. Algorithms such as *DAMAS*, which are more computationally expensive, could also be used but will take significantly more computational time unless only a few sound sources or frequencies are used. Analysis of scattering effects would also be investigated, see Section 7.6 for more details.

The microphone position calibration software could be adapted to use *openCV* camera calibration software instead of the MATLAB *Camera Calibration Toolbox*. This would enable the calibration to be totally automatic and not require selection of checkerboard coordinates in camera images. To obtain a better measure of the microphone position error obtained, the calibration technique could be used with a linear or planar array.

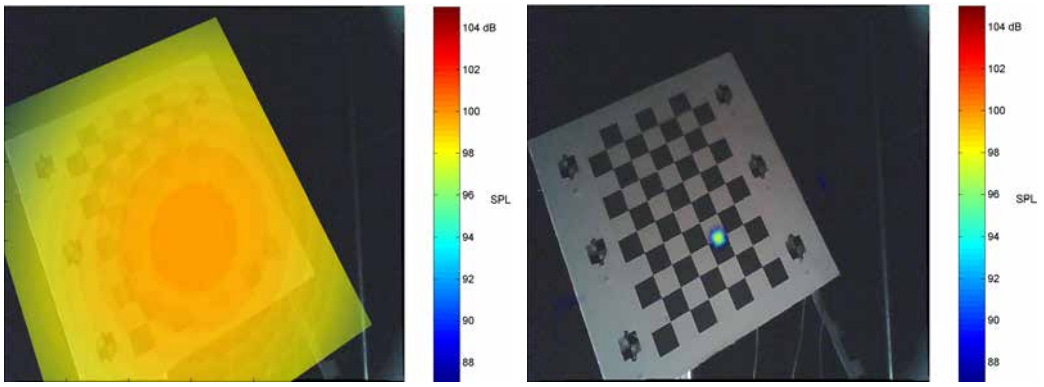
For structured light scanning of dynamic scenes, a colour coded structure light system or the use of Xbox Kinect sensors could be implemented. The current method uses grey code structured light scans which are only suitable for static objects. Methods of obtaining a wider field of view could also be investigated. More testing using structured light scans for beamforming and deconvolution could be performed.

9

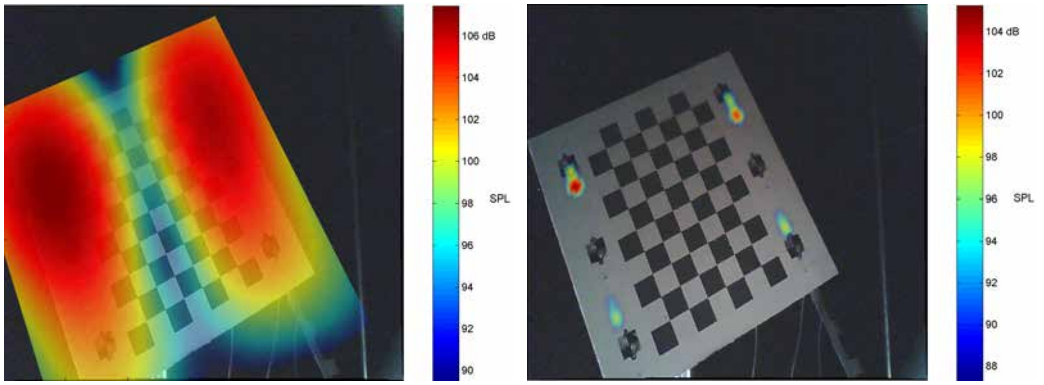
Beamforming and Deconvolution Plots



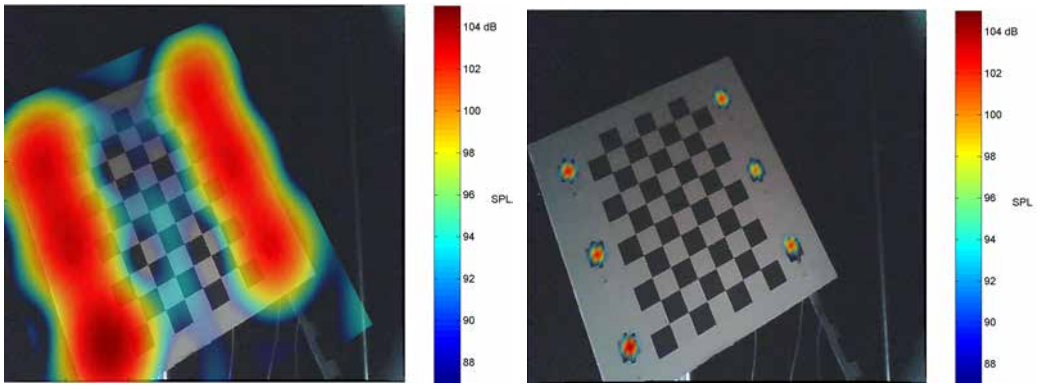
(a) Beamformed map, 0.527 to 0.558kHz (b) *CLEAN-SC* map, 0.527 to 0.558kHz



(c) Beamformed map, 1.054 to 1.117kHz (d) *CLEAN-SC* map, 1.054 to 1.117kHz

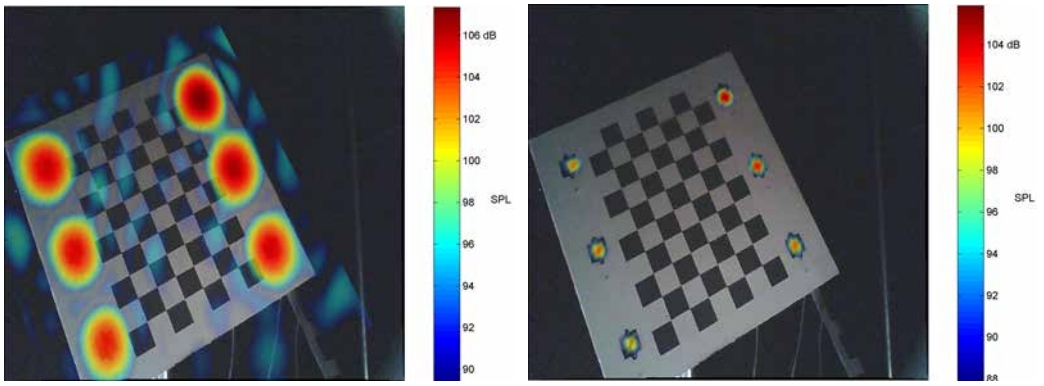


(e) Beamformed map, 2.981 to 3.158kHz (f) *CLEAN-SC* map, 2.981 to 3.158kHz

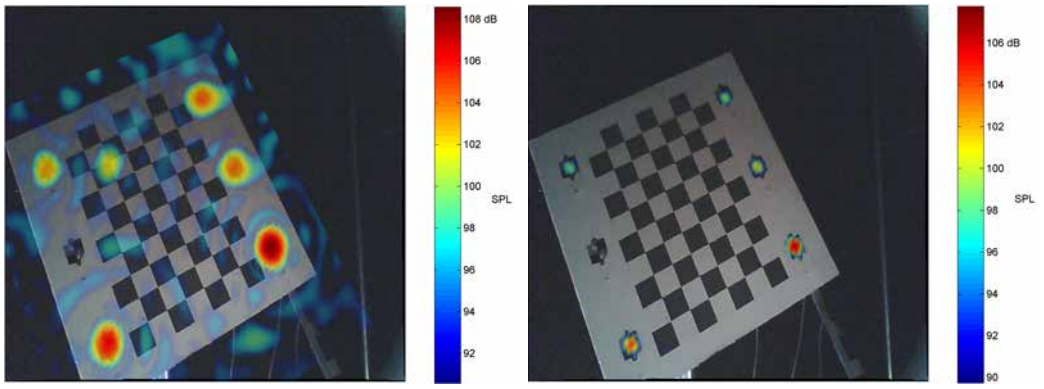


(g) Beamformed map, 5.014 to 5.312kHz (h) *CLEAN-SC* map, 5.014 to 5.312kHz

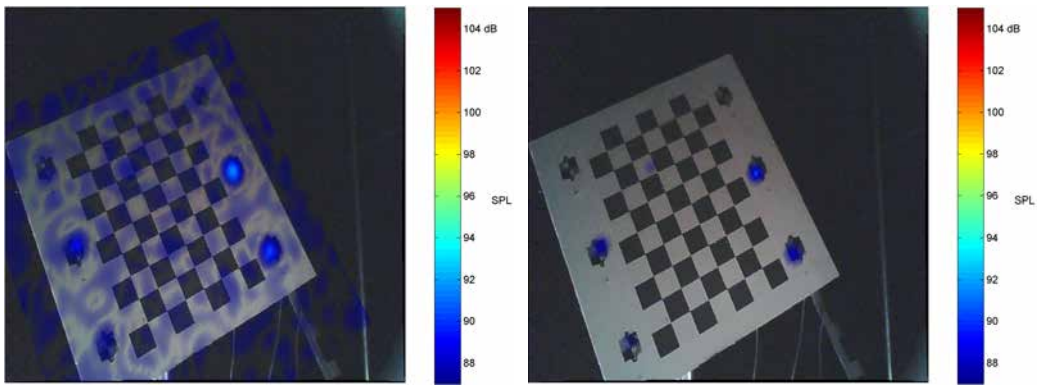
Figure 9.1: Examples of beamformed and *CLEAN-SC* maps for a ranged of octave frequency bands using a 3D scanning surface.



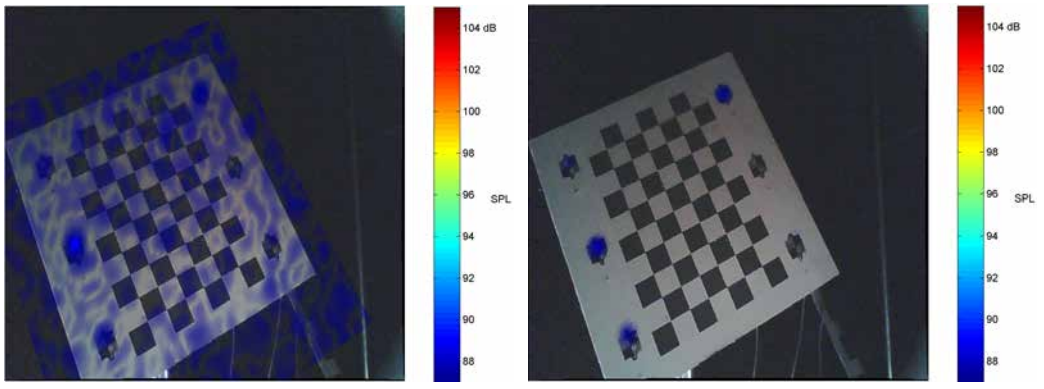
(a) Beamformed map, 9.465 to 10.027kHz (b) *CLEAN-SC* map, 9.465 to 10.027kHz



(c) Beamformed map, 14.181 to 15.024kHz (d) *CLEAN-SC* map, 14.181 to 15.024kHz

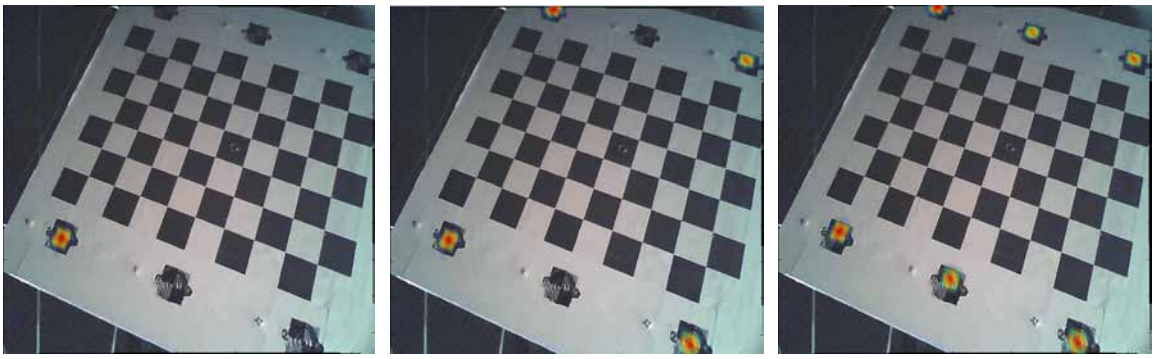


(e) Beamformed map, 18.929 to 20.055kHz (f) *CLEAN-SC* map, 18.929 to 20.055kHz

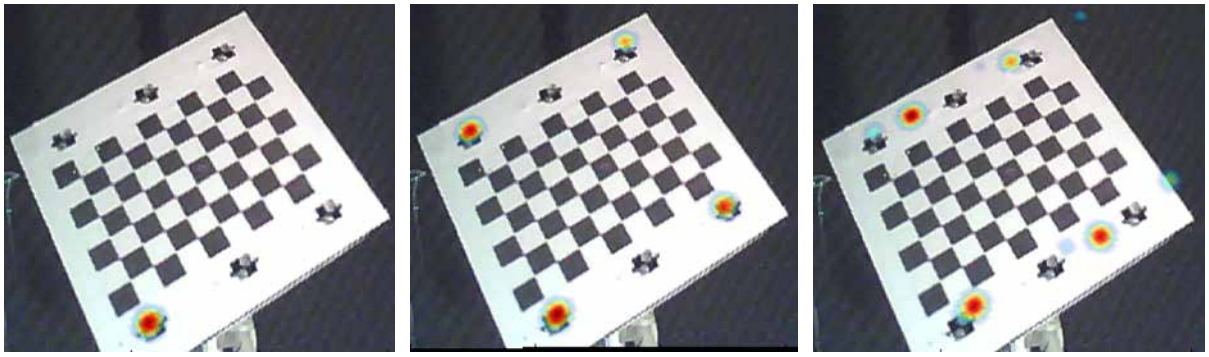


(g) Beamformed map, 22.511 to 23.849kHz (h) *CLEAN-SC* map, 22.511 to 23.849kHz

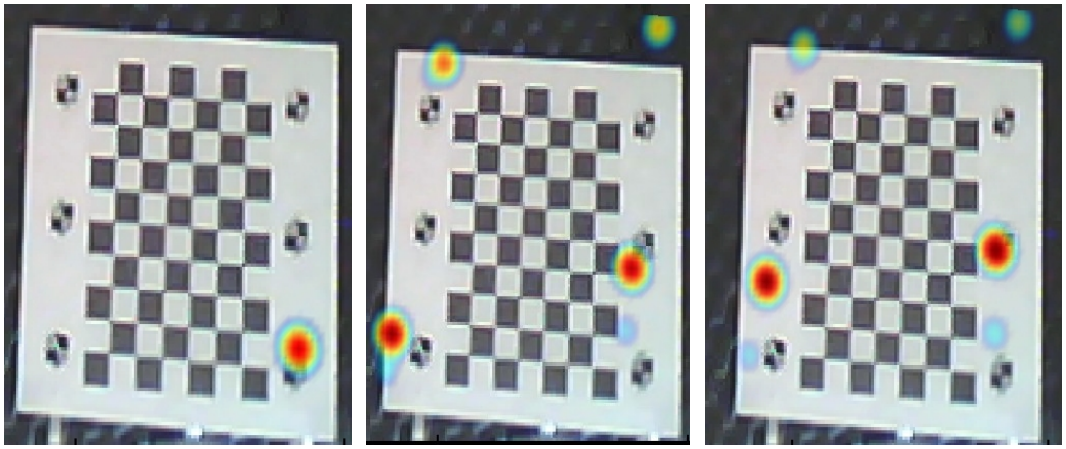
Figure 9.2: Examples of beamformed and *CLEAN-SC* maps for a ranged of octave frequency bands using a 3D scanning surface.



(a) Image No. 50,Record No. 2. (b) Image No. 50, Record No. 7. (c) Image No. 50, Record No. 8.

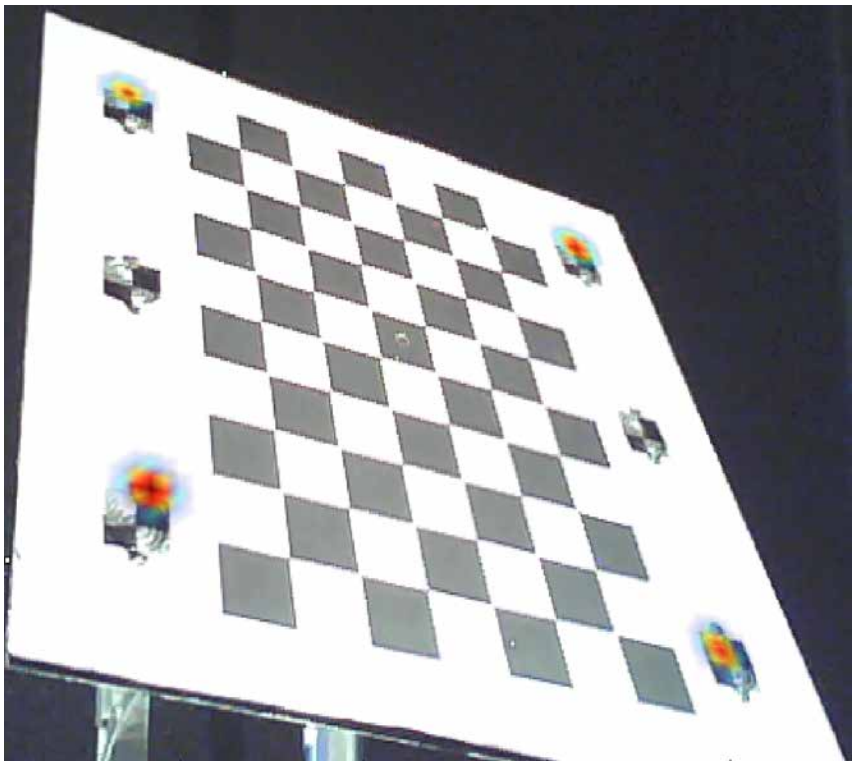


(d) Image No. 28, Record No. 2. (e) Image No. 28, Record No. 7. (f) Image No. 28, Record No. 8.

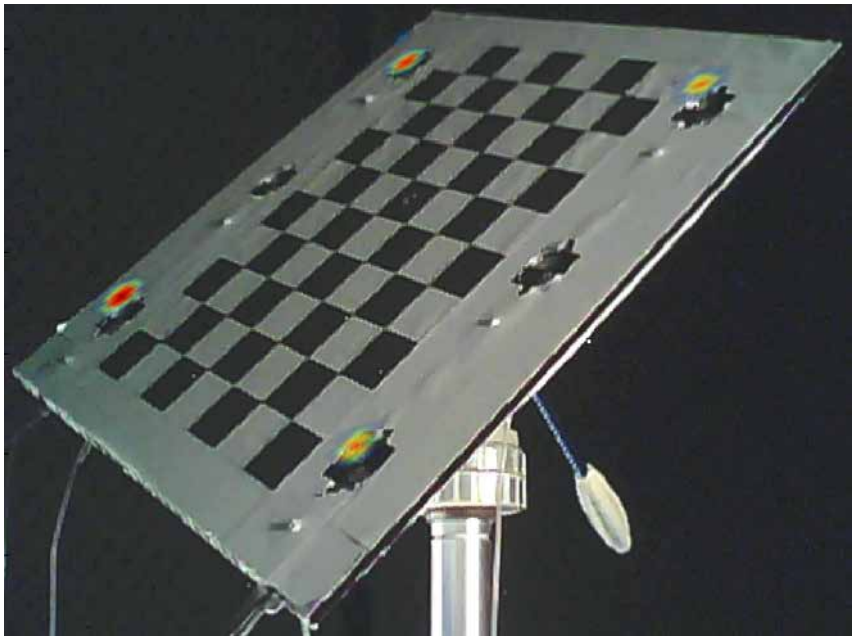


(g) Image No. 15, Record No. 2. (h) Image No. 15, Record No. 7. (i) Image No. 15, Record No. 8.

Figure 9.3: *CLEAN-SC* maps for three different distances of the checkerboard rig from the array for octave frequency band 5.014 to 5.312kHz using 3D scanning surfaces. These plots illustrate that as the distance of the sound source from the array increases, the spacing between sound sources needed to increase for *CLEAN-SC* to accurately locate the sound sources. A similar effect can be seen as the frequency decreases. The mean distance of the speakers from the array's main camera for Image No. 50, 28, and 15 are respectively 0.761, 1.510, and 2.451 m. Note that for Record No. 2 white noise is played on only one speaker, for Record No 7 uncorrelated white noise is simultaneously played on the four outer speakers, while for Record No. 8 uncorrelated white is simultaneously played on all six speakers simultaneously. (Note that the maps have been cropped to just show the checkerboard region.)

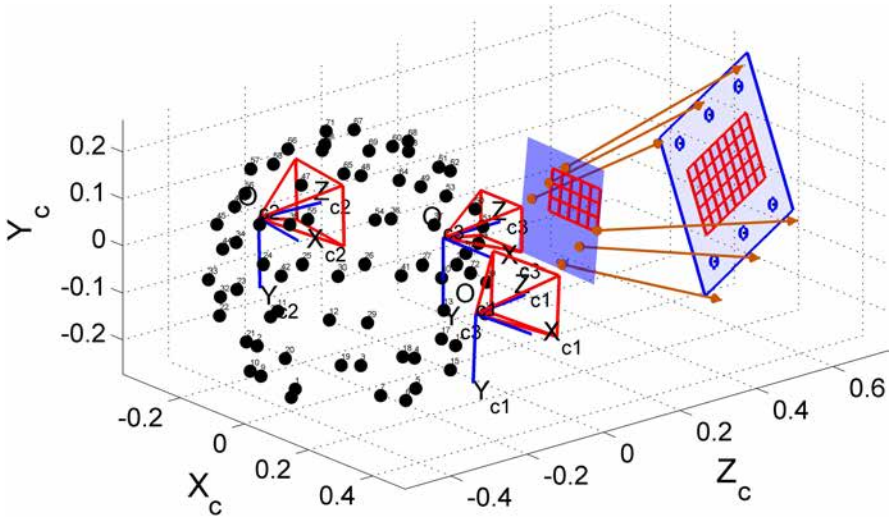


(a) Image Number 37, Record Number 7.

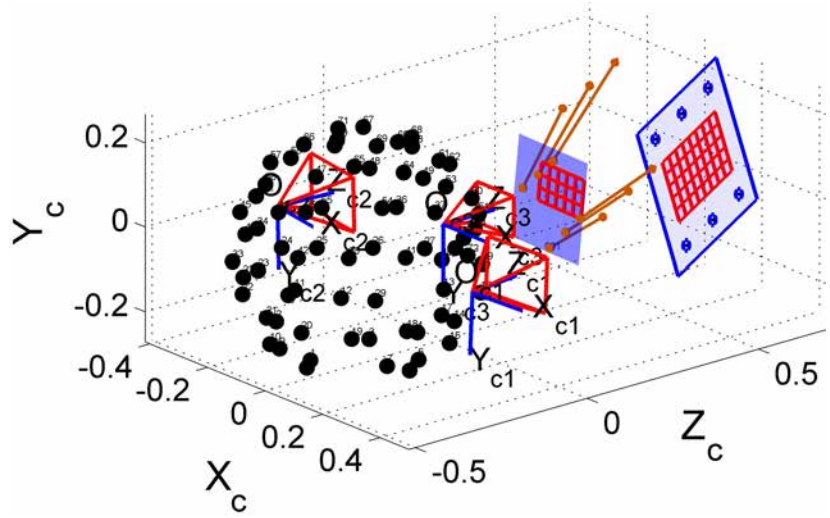


(b) Image Number 37, Record Number 7.

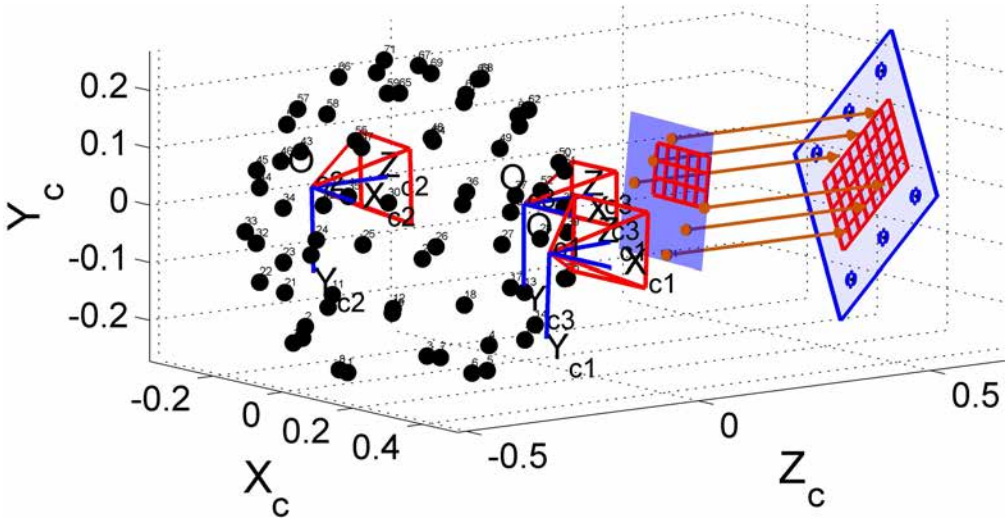
Figure 9.4: Example of apparent position errors seen in plots that occurred if the speaker rig was turned at high angle to the array, where 3D scanning surfaces are used. This is due to the use of the distance weighted steering vector given in Equation (2.17)



(a) Perspective projection from Camera 3 (main array camera located at the front of the array.



(b) Perspective projection from Camera 1 (secondary camera located to the side of the array.



(c) Orthogonal projection along Z axis.

Figure 9.5: Diagram illustrating projection methods investigated but not used due to high errors. The two red grids shows the position and orientation of the checkerboard used to obtain the traditional planar scanning surface (blue plane) and the coordinates of the speakers (blue circles).

10

Microphone Coordinates

Table 10.1: Underbrink multi-arm spiral array microphone coordinates.

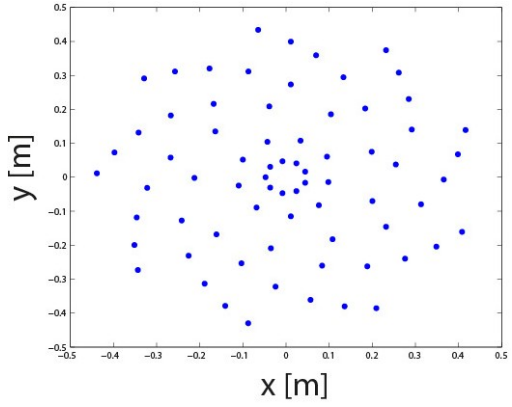
Mic. No.	X	Y	Z	Mic No	X	Y	Z
	[m]	[m]	[m]		[m]	[m]	[m]
1	-0.048	0.000	0.000	37	0.291	0.140	0.000
2	-0.036	-0.031	0.000	38	0.133	0.294	0.000
3	-0.008	-0.047	0.000	39	-0.087	0.311	0.000
4	0.024	-0.041	0.000	40	-0.267	0.182	0.000
5	0.045	-0.016	0.000	41	-0.322	-0.032	0.000
6	0.045	0.016	0.000	42	-0.226	-0.231	0.000
7	0.024	0.041	0.000	43	-0.025	-0.322	0.000
8	-0.008	0.047	0.000	44	0.188	-0.263	0.000
9	-0.036	0.031	0.000	45	0.313	-0.080	0.000
10	0.011	-0.115	0.000	46	0.285	0.230	0.000
11	0.076	-0.083	0.000	47	0.070	0.359	0.000
12	0.098	-0.015	0.000	48	-0.177	0.320	0.000
13	0.095	0.061	0.000	49	-0.342	0.131	0.000
14	0.034	0.108	0.000	50	-0.346	-0.119	0.000
15	-0.043	0.104	0.000	51	-0.189	-0.314	0.000
16	-0.100	0.052	0.000	52	0.057	-0.361	0.000
17	-0.110	-0.025	0.000	53	0.276	-0.240	0.000
18	-0.068	-0.090	0.000	54	0.366	-0.007	0.000
19	0.200	-0.071	0.000	55	0.262	0.308	0.000
20	0.199	0.075	0.000	56	0.011	0.399	0.000
21	0.104	0.185	0.000	57	-0.258	0.311	0.000
22	-0.039	0.209	0.000	58	-0.397	0.073	0.000
23	-0.164	0.135	0.000	59	-0.351	-0.200	0.000
24	-0.212	-0.002	0.000	60	-0.141	-0.379	0.000
25	-0.161	-0.168	0.000	61	0.136	-0.381	0.000
26	-0.035	-0.209	0.000	62	0.349	-0.204	0.000
27	0.108	-0.183	0.000	63	0.398	0.068	0.000
28	0.255	0.037	0.000	64	0.232	0.374	0.000
29	0.184	0.203	0.000	65	-0.065	0.434	0.000
30	0.011	0.273	0.000	66	-0.329	0.291	0.000
31	-0.168	0.216	0.000	67	-0.439	0.012	0.000
32	-0.268	0.058	0.000	68	-0.344	-0.273	0.000
33	-0.242	-0.128	0.000	69	-0.088	-0.430	0.000
34	-0.103	-0.253	0.000	70	0.209	-0.386	0.000
35	0.084	-0.261	0.000	71	0.408	-0.161	0.000
36	0.232	-0.146	0.000	72	0.416	0.139	0.000

Table 10.2: Spherical array microphone coordinates.

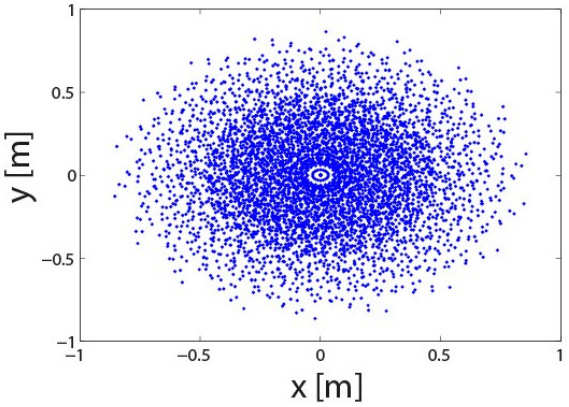
Mic. No.	X	Y	Z	Mic. No.	X	Y	Z
	[m]	[m]	[m]		[m]	[m]	[m]
1	-0.082	0.263	-0.332	37	-0.067	0.002	0.028
2	-0.187	0.194	-0.343	38	0.074	0.003	0.028
3	-0.083	0.261	-0.160	39	0.242	0.004	-0.092
4	0.002	0.240	-0.089	40	0.282	-0.063	-0.246
5	0.090	0.261	-0.161	41	0.262	-0.084	-0.334
6	0.162	0.240	-0.249	42	0.089	-0.083	-0.507
7	0.195	0.193	-0.341	43	0.003	-0.156	-0.485
8	0.003	0.240	-0.408	44	-0.084	-0.084	-0.505
9	0.002	0.160	-0.488	45	-0.190	-0.092	-0.437
10	-0.093	0.192	-0.439	46	-0.256	-0.083	-0.334
11	-0.235	0.161	-0.247	47	-0.255	-0.084	-0.162
12	-0.188	0.193	-0.153	48	-0.188	-0.095	-0.060
13	0.004	0.158	-0.010	49	-0.083	-0.082	0.010
14	0.098	0.194	-0.058	50	0.089	-0.083	0.010
15	0.196	0.193	-0.154	51	0.193	-0.093	-0.057
16	0.193	0.097	-0.058	52	0.263	-0.084	-0.162
17	0.282	0.071	-0.250	53	0.196	-0.189	-0.155
18	0.261	0.089	-0.334	54	0.196	-0.188	-0.344
19	0.192	0.099	-0.440	55	0.098	-0.187	-0.441
20	0.089	0.088	-0.506	56	-0.092	-0.186	-0.441
21	0.004	0.073	-0.524	57	-0.190	-0.187	-0.344
22	-0.186	0.099	-0.439	58	-0.237	-0.156	-0.247
23	-0.255	0.090	-0.335	59	-0.190	-0.186	-0.150
24	-0.274	0.072	-0.247	60	-0.090	-0.187	-0.055
25	-0.255	0.089	-0.163	61	0.005	-0.157	-0.014
26	-0.186	0.099	-0.057	62	0.098	-0.189	-0.060
27	-0.083	0.090	0.009	63	0.088	-0.257	-0.161
28	0.003	0.069	0.029	64	0.161	-0.238	-0.250
29	0.241	0.003	-0.406	65	0.089	-0.257	-0.335
30	0.192	-0.092	-0.441	66	-0.083	-0.255	-0.333
31	0.071	0.002	-0.525	67	-0.082	-0.255	-0.163
32	-0.067	0.004	-0.524	68	0.002	-0.235	-0.089
33	-0.156	0.003	-0.485	69	0.069	-0.273	-0.249
34	-0.271	0.003	-0.317	70	0.003	-0.275	-0.317
35	-0.274	0.004	-0.177	71	-0.068	-0.274	-0.248
36	-0.154	0.003	-0.011	72	0.001	-0.275	-0.177

Table 10.3: SADA array coordinates scaled by 130% from those given by Humphreys et al. (1998).

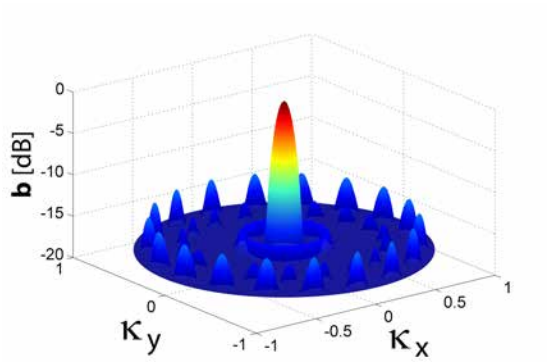
Mic. No.	X	Y	Z	Mic No	X	Y	Z
	[m]	[m]	[m]		[m]	[m]	[m]
1	0.000	0.000	0.000	18	0.000	-0.059	0.000
2	0.000	-0.015	0.000	19	0.045	-0.045	0.000
3	0.011	-0.011	0.000	20	0.059	0.000	0.000
4	0.015	0.000	0.000	21	0.045	0.045	0.000
5	0.011	0.011	0.000	22	0.000	0.059	0.000
6	0.000	0.015	0.000	23	-0.045	0.045	0.000
7	-0.011	0.011	0.000	24	-0.059	0.000	0.000
8	-0.015	0.000	0.000	25	-0.045	-0.045	0.000
9	-0.011	-0.011	0.000	26	0.000	-0.119	0.000
10	0.000	-0.030	0.000	27	0.091	-0.091	0.000
11	0.023	-0.023	0.000	28	0.119	0.000	0.000
12	0.030	0.000	0.000	29	0.091	0.091	0.000
13	0.023	0.023	0.000	30	0.000	0.119	0.000
14	0.000	0.030	0.000	31	-0.091	0.091	0.000
15	-0.023	0.023	0.000	32	-0.119	0.000	0.000
16	-0.030	0.000	0.000	33	-0.091	-0.091	0.000
17	-0.023	-0.023	0.000				



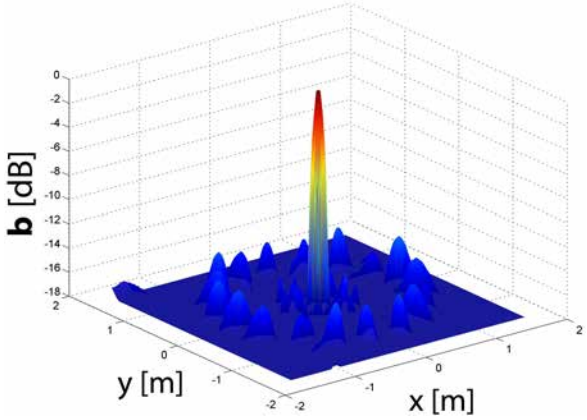
(a) Microphone coordinates.



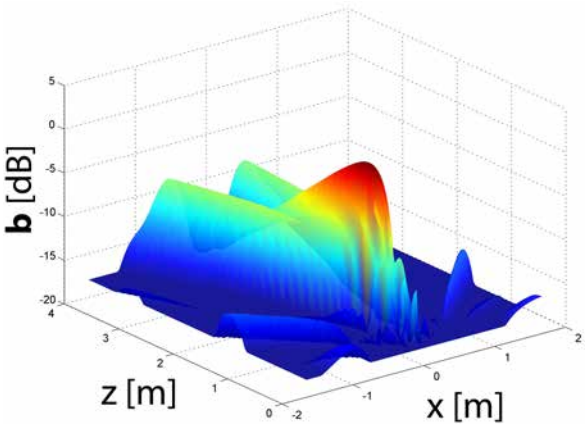
(b) Microphone coarray.



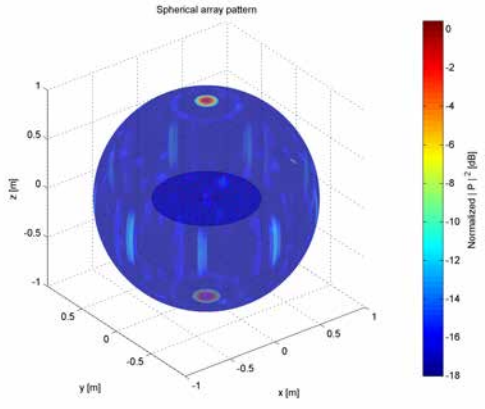
(c) Far-field beam pattern or point spread function.



(d) Near-field lateral point spread function.

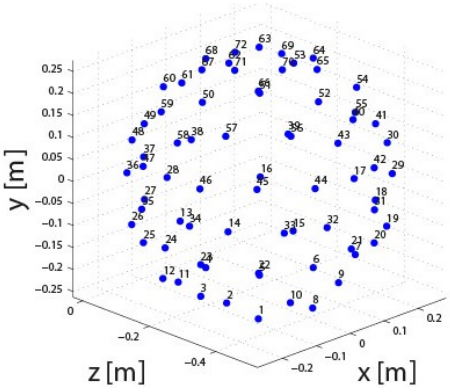


(e) Near-field longitudinal point spread function.

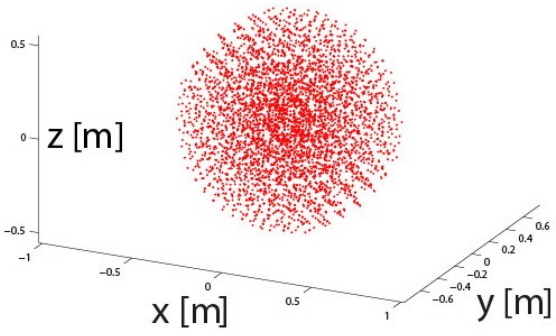


(f) Near-field spherical point spread function.

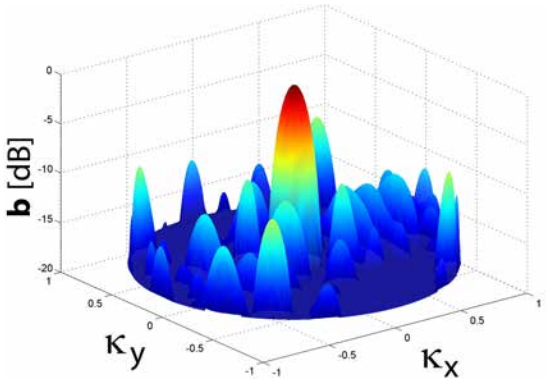
Figure 10.1: Near-field spiral point spread functions for 3.5kHz. (Note that spherical spreading correction has been omitted.)



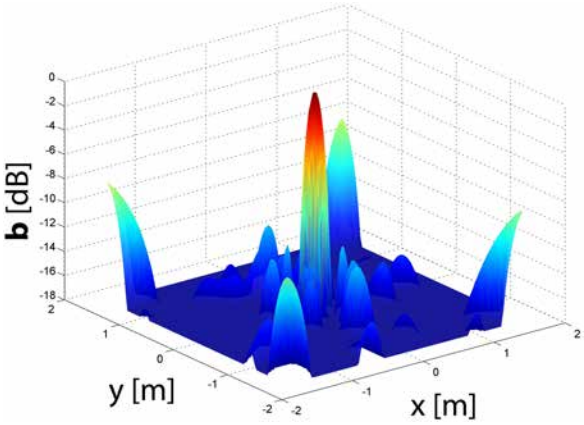
(a) Microphone coordinates.



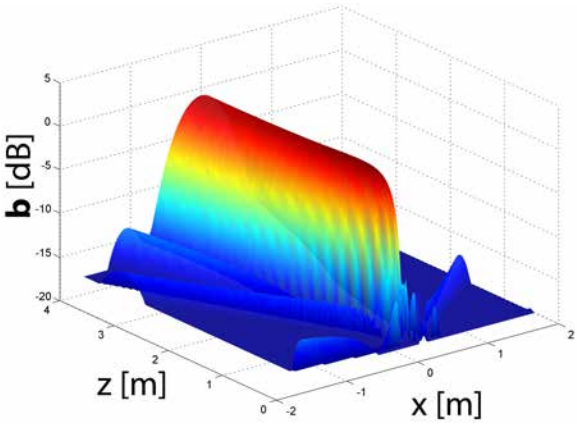
(b) Microphone coarray.



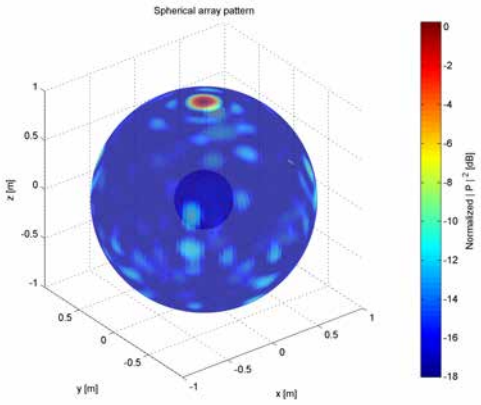
(c) Far-field beam pattern or point spread function.



(d) Near-field lateral point spread function.

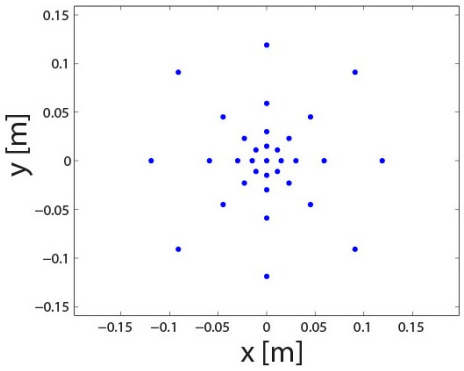


(e) Near-field longitudinal point spread function.

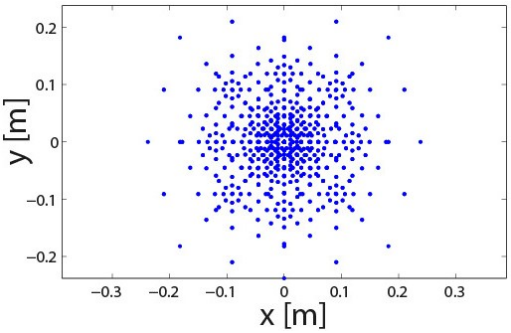


(f) Near-field spherical point spread function.

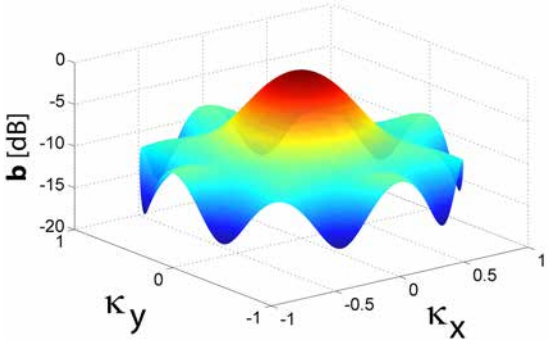
Figure 10.2: Spherical array point spread functions for 3.5kHz. (Note that spherical spreading correction has been omitted.)



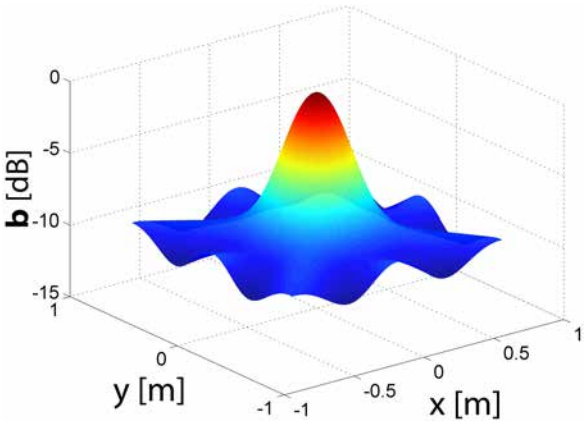
(a) Microphone coordinates.



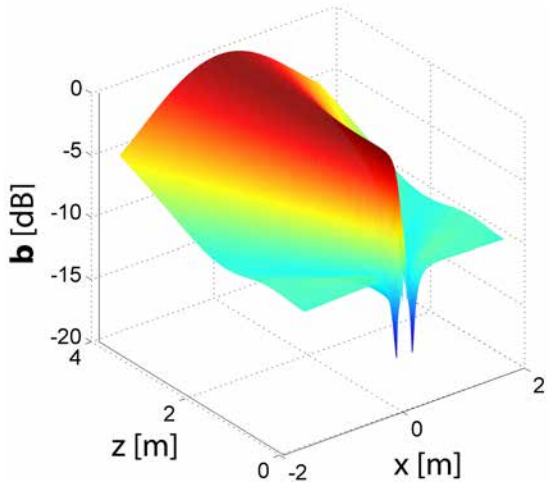
(b) Microphone coarray.



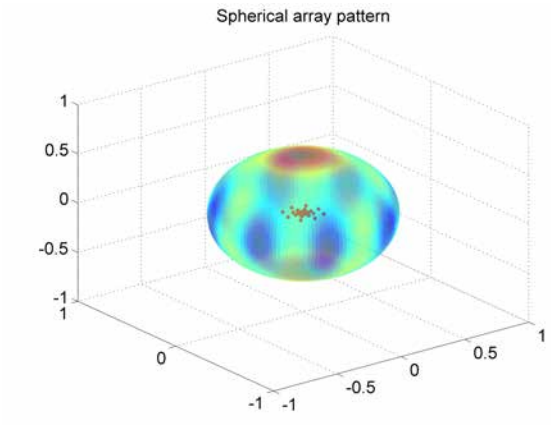
(c) Far-field beam pattern or point spread function.



(d) Near-field lateral point spread function.

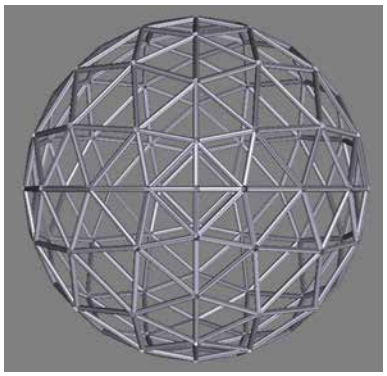


(e) Near-field longitudinal point spread function.

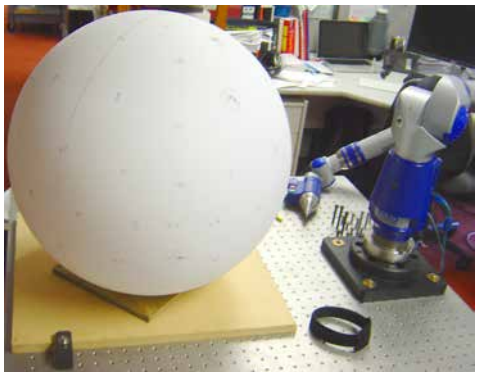


(f) Near-field spherical point spread function.

Figure 10.3: $SADA$ array point spread functions for 3.5kHz. (Note that spherical spreading correction has been omitted.))

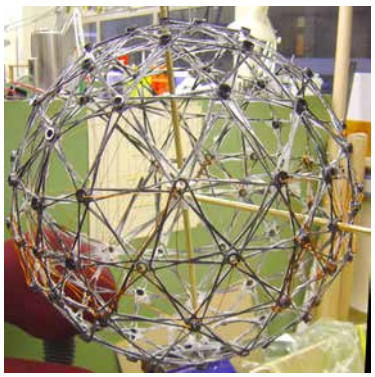


(a) Octahedron.



(b) Faro-Arm measurement of microphone location on mold.

(c) Microphone holders bolted to mold and carbon fiber tow wrapped between in readiness for setting with epoxy resin.



(d) Joining two parts of lattice together.

(e) Measurements of final microphone positions with Faro-Arm.

Figure 10.4: Spherical array build process.

11

Deconvolution Literature

Dougherty and Stoker (1998)

The first application of the CLEAR algorithm in acoustic array measurements is presented.

Brühl and Röder (2000)

A procedure is proposed to solve the complete inverse problem for microphone arrays. A method was presented which allows a construction of an acoustic source model based on the analysis of microphone array measurements during a train pass-by.

Wang et al. (2004a)

Wideband **RELAX** (*WB-RELAX*) and the wideband *CLEAN* (*WB-CLEAN*) algorithms are presented for aeroacoustic imaging using an acoustic array. *WB-RELAX* is a parametric approach that can be used efficiently for point source imaging without the sidelobe problems suffered by the delay-and-sum beamforming approaches. *WB-CLEAN*

does not have sidelobe problems either, but it behaves more like a nonparametric approach and can be used for both point source and distributed source imaging.

Point sources in 3D space were investigated. Although it was found that the 2D *SADA* array used had a limited resolution in the direction perpendicular to the array, it was believed that under certain test situations, the 3D location of sources could be estimated.

Brooks and Humphreys (2004) & Brooks and Humphreys (2006a)

A Deconvolution Approach for the Mapping of Acoustic Sources (*DAMAS*) algorithm represents a breakthrough in phased array imaging for aeroacoustics, potentially eliminating sidelobles and array resolution effects from beamform maps. *DAMAS* is an iterative non-negative least squares solver using the Gauss-Seidel iterative method. It gave better results than *CLEAN* algorithm in typical aeroacoustic applications.

Brooks and Humphreys (2005)

This paper examines the technique of using *DAMAS* for three-dimensional (3D) applications. First, the beamforming ability for arrays, of different size and design, to focus longitudinally and laterally is examined for a range of source positions and frequency. Advantage is found for larger array designs with higher density microphone distributions towards the centre. After defining a 3D grid generalised with respect to the arrays beamforming characteristics, *DAMAS* is employed in simulated and experimental noise test cases.

Dougherty (2005b)

A disadvantage of both *DAMAS* and the *CLEAN* algorithms is the relatively high computational effort that is required compared to conventional beamforming. The original *DAMAS* algorithm also lacks an explicit regularisation method to prevent noise amplification. *DAMAS2* and *DAMAS3* was developed by Dougherty to solve these issues. Similar to the Richardson-Lucy method these are iterative Fourier-based algorithms, where the point-spread function is considered to be shift-invariant. In contrast to the applications in astronomy, where the distance from the observed objects is huge compared to the object or array dimension, in aeroacoustic array measurements the point-spread function may change significantly within the source region. However, Dougherty obtained promising results for cases where the source region is limited to small angular sectors.

DAMAS2 provides a dramatic speedup of each iteration and adds regularisation by a low pass filter. *DAMAS3* also provides fast iterations, and additionally, reduces the required number of iterations. It uses a different regularisation technique from *DAMAS2*, and is partially based on the Wiener filter.

Also see Dougherty's page `http://www.optinav.com/Iterative-Deconvolve-3D.htm`

Ravetta (2005) & Ravetta et al. (2006)

A new approach to the deconvolution of acoustic sources is developed. The beamforming output is modelled as a superposition of 'complex' point spread functions and a nonlinear system of equations in terms of the sources amplitude is posted. Such system is solved using a new 2- step procedure. In the first step, an approximated linear problem is solved. In the second step, an optimisation is performed over the nonzero values obtained in the previous step. The solution to this system of equations renders the sources position and amplitude. In the case of coherent sources, their relative phase can also be recovered. The technique is referred as: noise source Localisation and Optimisation of Phased Array Results (*LORE*). A detailed analytical formulation, numerical simulations as well as sample experimental results are shown for the proposed post-processing.

Other issues present in the *DAMAS* approach, such as the position of the actual source relative to the scanning grid are not an issue in the new approach (so long as the actual source is inside the scanning grid). Unlike *DAMAS*2 and 3, the theoretical formulation is not restricted to small regions in space and can be applied to **2D or 3D space**.

Brooks and Humphreys (2006a)

An underlying premise of *DAMAS*, as well as that of classical array beamforming methodology, is that the noise regions under study are distributions of statistically independent sources. The present development, called *DAMAS-C*, extends the basic approach to include coherence definition between noise sources. The solutions incorporate cross-beamforming array measurements over the survey region. While the resulting inverse problem can be large and the iteration solution computationally demanding, it solves problems no other technique can approach. *DAMAS-C* is validated using noise source simulations and is applied to airframe flap noise test results.

Ehrenfried and Koop (2006) & Ehrenfried and Koop (2007)

The *DAMAS2* algorithm is compared with several other Fourier-based deconvolution approaches. One is the Richardson-Lucy method, which is widely used for the deconvolution of astronomical images. The second is a modified gradient-type NNLS approach, where spectral procedures are implemented to accelerate the computations. Both methods require a computational effort similar to the *DAMAS2* algorithm. All three algorithms use an approximate shift-invariant point-spread function. Furthermore it is described how the *DAMAS2* and the Fourier-based NNLS algorithms can be embedded in an outer iteration loop to take the variation of the point-spread function into account. The resulting methods have two nested iterations, which require much more numerical effort than a single *DAMAS2* iteration loop.

All methods are tested with synthetic data. At first an example with a simple linear array and a small opening angle is considered, where the variation of the point-spread function in the source region is negligible. In this test case the results of the *DAMAS2*, the modified gradient-type NNLS, and the Richardson-Lucy algorithm are compared. It is shown that these algorithms more or less introduce oscillations in the reconstructed source distribution. A second test case with a planar array and a large opening angle is presented, to demonstrate the influence of a strong variation of the point spread function. It is shown that the approximate methods lead to distorted results, while the methods with nested iterations give a significantly better reconstruction of the source distribution.

Guérin et al. (2006)

A deconvolution method of beamforming maps for moving broadband sources was formulated and applied to synthesised data. A modified Gauss-Seidel method presented in Brooks and Humphreys (2004) is used. This method contains no regularisation technique to prevent noise amplification. However it converges quickly towards a very acceptable solution. The method presented here is a hybrid one: the beamforming results are calculated in the time domain and the point spread function is approximated in the frequency domain taking into account an average Doppler frequency shift.

Papamoschou and Dadvar (2006)

Deconvolution was investigated in relation to jet noise sources. Deconvolution based on the Richardson-Lucy inversion method shows promise as a means to extract the true source distribution from the array power spectrum.

Recent approaches to deconvolving the array output include the *DAMAS* algorithm of Brooks and Humphreys and the *DAMAS2* and *DAMAS3* algorithms of Dougherty. Doughertys algorithms assume that the point spread function is or can be rendered in convolutional form, i.e., $V(x, \xi, \omega) = V(x - \xi, \omega)$, which is not the case for the present array. The *DAMAS* algorithm is basically a Gauss-Seidel (GS) iterative method, with non-negative constraint imposed at each iteration step. The performance of GS can depend on the ordering of the equations. The *DAMAS* approach was tried and, while the results seemed reasonable, their dependence on the order of the equations raised concern. For example, the location of peak noise changed by as much as three jet diameters depending on whether the rows of the linear system of Eq. 31 were ordered forward or backward in x.

Another method tested was the Richardson-Lucy (RL) iteration algorithm which has found wide application in image restoration. The advantage of RL over GS is that its output is inherently non-negative and does not depend on the sequence of the equations. Several tests indicated that the output was consistently smoother and more robust than that of the GS method. For these reasons, the RL method was selected for the deconvolution.

Quayle et al. (2006)

Source power estimates have been computed using a power integration method (POWINT) and verified using the *CLEAN* source identification method. By removing a proportion of the highest sources and re-evaluating the source map, the *CLEAN* algorithm allows successive sources to be identified and interpreted using knowledge of the model position. Using a source power integration method without carefully selecting the integration area for each frequency band concerned can lead to errors in source estimation, particularly where background tunnel noise is high. The *CLEAN* method is also excellent at rejecting sidelobes at higher frequencies, extending the useful frequency range of a given array.

Brooks and Humphreys (2006b)

The *DAMAS* microphone phased array processing methodology that allows the determination and separation of coherent and incoherent noise source distributions.

Pascal and Li (2007)

DAMAS (Gauss-Seidel) algorithm, Richardson-Lucy's algorithm, and Jansson-Van Cittert's algorithm are compared. The *DAMAS* method (Gauss-Seidel) presents a good resolution whereas the algorithm of Richardson-Lucy converges towards to a smooth but also more stable solution. Some difficulties may occur in the estimation process if information concerning the acoustic sources of interest is masked by a too wide main lobe at low frequencies. In this case, SOAP method gives advantages.

Sijtsma (2007a), Sijtsma (2007b) and Sijtsma (2008)

To overcome the disadvantages of a PSF-based method, a new deconvolution method is described here. This new method takes advantage of the fact that sources in source plots are spatially coherent with their side lobes. Beam patterns of individual noise sources are determined by analysand the measured spatial coherence, thus avoiding the use of synthetical PSFs. The new method is called '*CLEAN* based on Source Coherence' or, briefly, *CLEAN-SC*. It is an alternative version of the classical *CLEAN* method used in Astronomy, which iteratively removes PSFs of peak sources from a "dirty map". Essentially, *CLEAN-SC* iteratively removes the part of the source plot which is spatially coherent with the peak source. A feature of *CLEAN-SC* is its ability to extract absolute sound power levels from the source plots. The merits of *CLEAN-SC* were demonstrated using array measurements of airframe noise on a scale model of the Airbus A340 in the 86 m2 closed test section of DNW-LLF.

Böhning and Siller (2007)

In this study, the specific problems with the analysis of moving sources are addressed. A point spread function for moving sources is formulated and applied to the simple deconvolution method *CLEAN*. In the analysis of flyover measurement data of an Airbus A320, this method significantly improves the quality of the source distributions. The modified point spread functions are also successfully tested with the *DAMAS* algorithm for simulated moving point sources.

Brusniak (2008)

This paper examines the effectiveness of the *DAMAS2* algorithm for accurately determining flight test aircraft noise subcomponent spectral levels. First, the *DAMAS2* subcomponent spectral levels are shown to be qualitatively correct through comparison with the

expected behaviour as seen in the phased array beamform maps. The *DAMAS2* spectra are then compared to sample output from the original *DAMAS* algorithm and are seen to be generally within 0.1 dB of *DAMAS* algorithm spectral levels. Lastly, the shape of the cumulative *DAMAS2* spectral levels are compared with single microphone spectra and are seen to be in very good agreement. The *DAMAS2* algorithm is seen to be an effective tool for flight test aircraft noise subcomponent spectral measurement

Guérin and Weckmüller (2008) and Guérin and Siller (2008)

The use of frequency-domain reconstruction of the point-spread function for moving sources is investigated to be used by the *DAMAS* deconvolution method.

In order to determine the amplitude of the tonal source when it moves, the *DAMAS* method was adapted so that now the frequency shift on the sidelobes is taken into account in the formulation of the least-square problem. The modification of the beamforming pattern related to motion was also accounted for. At the end, the problem could be solved at the expense of an increase of the computing time and need in memory. Further testing with real data is necessary to validate the method.

Michel and Funke (2008a) and Michel and Funke (2008b)

It is demonstrated that the source strengths and directivities of all sound sources of a high bypass ratio aeroengine can be determined from measurements in an open air test bed with a line array of microphones, which is laid out parallel to the engine axis in the geometric near field of the engine.

An estimation of the source strengths from these maps is possible only in rare cases, e.g., when the source positions are sufficiently separated spatially. Sources along a line or distributed over an area or over a source volume yield results that depend on the beam width of the point-spread function and on sidelobes and aliases. Additional problems arise when the source directivities are non-uniform. The consequence is that amplitudes of sound sources are very difficult to derive from beamforming maps. The problem can be solved with deconvolution techniques. It is assumed that the point sources have a uniform directivity. The point spread functions of the array (or beamformed maps) are calculated for every possible source position and for each narrow-band frequency of interest. The source levels of the unknown sources have to be determined with a least square fit for the difference between the measured beamformed map and the estimate for the sum of point sources.This deconvolution procedure yields huge and often badly conditioned matrices.

Special iterative procedures are required to solve them with the side condition that only positive source levels are permitted.

Blacodon and Elias [1, 2] propose a different method for the determination of the source strengths of assumed sources. The method is based on modelling the cross-spectral matrix of the microphone signals. Instead of determining the sources such that the beamform maps are best approximated, Blacodon and Elias determine the sources for a best approximation of the cross spectral matrix. They generate a cross-spectral matrix for each possible point source position assuming a uniform directivity of the sources. The amplitudes of the sources are then determined with a least square fit between the modelled and the measured cross-spectral matrix. The beamforming map is not required for their method.

It is the purpose of the present paper to develop a method that is based on a line array of microphones, which permits the determination not only of the source strengths but also of the directivities of the sources. The method of Blacodon and Elias [2] is used as a starting point but extended to include the directivities of the sources.

Quayle et al. (2008)

In this paper we present a discussion of measurements from the tunnel and mechanisms for interference from wind tunnel sources. We confirm the existence of the effect using sources in an anechoic chamber. We also present a method for mitigating interference using the '*CLEAN-SC*' algorithm with appropriately reshaped scanning grids for beamforming. Results show that more consistent measurements are obtained from the two acoustic arrays, providing some validation of the method. This analysis has the potential to be significant in a variety of closed wind tunnel testing.

Romenskiy and Jaeckel (2008)

The paper presents a new method for the improvement of microphone array measurement results using an unconventional, relatively simple and very fast iterative post processing algorithm suitable for 2D-beamforming acoustic maps. This procedure is based on the mathematical model of a time reversed diffusion process which is numerically stabilised by means of a spatial low pass filter as well as by an additional amplitude renormalisation step. The method allows the reduction of the mainlobe width of acoustic point sources in the map and can thus be applied for the improvement of the usually very bad image contrast in the lower frequency range and also for a better separation of closely lying

sources which can not be clearly resolved by simple delay-and-sum beamforming. An outstanding property of the proposed algorithm is that it does not rely on the knowledge of the point spread function which is needed in complete deconvolution approaches. The iteration procedure merely utilises the local curvature information that is already hidden within the acoustic beamforming map to sharpen the individual sources while simultaneously smoothing out disturbing high frequency components that otherwise would lead to numerical instability and unlimited amplification of unwanted small false signal and noise components.

Sarradj (2008)

It is not necessary to perform the beamforming algorithm (3) to get quantitative information about the individual source amplitudes. To assign these amplitudes to sources at a location of interest, a reduced, so called orthogonal beamformer may be applied, where G is replaced by the reconstructed estimation of the cross spectral matrix due to the i-th source $V_i L_{M,ii} V_i^H$. A map of sound pressure contributions estimated using this beamformer will show only contributions from one source. Thus, the map has one distinct maximum that corresponds to the location of the source.

Yardibi et al. (2008b) and Yardibi et al. (2008a)

The classical delay-and-sum approach suffers from low resolution and high sidelobes and the resulting beamforming maps are difficult to interpret. The deconvolution approach for the mapping of acoustic sources (*DAMAS*) deconvolution algorithm recovers the actual source levels from the contaminated delay-and-sum results by defining an inverse problem that can be represented as a linear system of equations. In this paper, the deconvolution problem is carried onto the sparse signal representation area and a sparsity constrained deconvolution approach (*SC-DAMAS*) is presented for solving the *DAMAS* inverse problem. A sparsity preserving covariance matrix fitting approach (CMF) is also presented to overcome the drawbacks of the *DAMAS* inverse problem. The proposed algorithms are convex optimisation problems. Our simulations show that CMF and *SC-DAMAS* outperform *DAMAS* and as the noise in the measurements increases, CMF works better than both *DAMAS* and *SC-DAMAS*. It is observed that the proposed algorithms converge faster than *DAMAS*. A modification to *SC-DAMAS* is also provided which makes it significantly faster than *DAMAS* and CMF. For the correlated source case, the CMF-C algorithm is proposed and compared with *DAMAS-C*. Improvements in performance are obtained similar to the uncorrelated case. Also see Yardibi et al. (2010).

Fleury et al. (2008)

To take into account directivity effects, one may introduce an acoustic source model that allows coherent monopoles. The problem is then to restore the source Cross-Spectral Matrix (CSM). A method has already been defined to tackle this issue, *DAMAS-C*. The proposed algorithm however requires heavy computational costs. In this paper, two extensions are provided to overcome this difficulty. The first one allows to solve the full problem, and the other one is an approximated formulation, similar to *DAMAS2 3* but extended to correlated sources. These methods are applied to simulated data and experimental data on co-flow jet noise. These examples show the convergence difficulties when many source variables are to be restored. The introduction of *a priori* information and the application of a regularisation scheme allow to overcome the issue. The application of the full methodology then allows to restore smoothed source CSM and extrapolate the acoustic radiation pattern into the far-field in good agreement with the expectation. Conversely, the application of the approximated methodology should be restricted to problems where the point spread function is shift-invariant.

Suzuki (2008)

To resolve coherent (or incoherent), distributed (or compact), and multipole aerodynamic sound sources with phased-array pressure data, a new source-detection algorithm is developed based on generalised inverse techniques (not deconvolution). To extract each coherent signal, a cross spectral matrix is decomposed into eigenmodes. Subsequently, the complex source-amplitude distribution that recovers an eigenmode is solved using generalised inverse techniques with reference solutions which include multipoles as well as a monopole. The source distribution consisting of pre-defined source types is first solved with a least square approach and then iteratively narrowed from an underdetermined to an overdetermined problem. The capabilities of the proposed algorithm are demonstrated using various model problems by comparing with several existing beam-forming algorithms, and it is found that distributed sources as well as dipoles with arbitrary orientation can be identified regardless of coherency with another source. The resolution is nearly comparable to *DAMAS*2, and the computational cost is only few times more than that of existing iterative approaches. The proposed algorithm is also examined using previous model-scale test data taken in an openjet wind-tunnel from a study on jet-flap interaction, and some indication of dipole radiation is discerned near the flap edge.

Papamoschou (2008)

This study relates to the imaging of noise sources that are distributed and strongly directional, such as in turbulent jets. For jet noise, which is not only directional but whose nature changes with emission angle, it is necessary to conduct the measurements with a narrow-aperture array. The resulting images need to be processed to remove as much as possible the spatial distortions introduced by the array response. Two such methods are addressed here, a deconvolution technique based on the Richardson-Lucy inversion and a direct spectral estimation method that obviates delay-and-sum beamforming. Both approaches enhanced significantly the spatial resolution of the noise source maps and provided similar results. It is shown that the deconvolution integral amounts to the tensor inner product of the direct spectral estimation relation with the array response matrix.

Dougherty (2008a)

Beamforming with a standard 5° polar array in a static engine test is possible. The results will be strongly contaminated by sidelobes if conventional beamforming alone is used to process the data. Deconvolution processing with *DAMAS* and a standard model of the point spread function improves the resolution but not the sidelobes. The problem with *DAMAS* in this case is believed to be the fact that the simple point spread function model does not account for extended source coherence or turbulent decorrelation. Deconvolution with *CLEAN-SC* greatly reduces the sidelobes. This algorithm appears to be a major breakthrough in phased array technology. Even with *CLEAN-SC*, the 5 polar array has some sidelobes. A well-designed 1D phased array can be created adding microphones to a sector of the polar arc. This provides better high frequency performance than the 5 array, and gives results that characterise a particular range of directivity angle. The combination of a purposefully-designed phased array and *CLEAN-SC* can give detailed source maps over a wide frequency range with almost no sidelobes or difficulties with source coherence.

Dougherty (2008b)

A mode measurement array of 119 Kulite transducers was installed in an inlet duct section and tested with a Honeywell Tech977 turbofan engine. 2D mode analysis was performed using conventional beamforming and *CLEAN-SC*. *CLEAN-SC* is found to be a powerful enhancement for mode measurements, completely removing several types of array artifacts and facilitating reliable sound power estimates. An interesting feature of *CLEAN-SC* is

that the resulting mode map potentially combines the powers of several modes into a single spot. Applying *CLEAN-SC* to the spatial imaging results in a confusing distribution of apparent point sources. A need to improve the way *CLEAN-SC* represents extended sources has been identified.

CLEAN-SC may be an enabling technology for mode power estimates from simple in-duct arrays because of its ability to remove array effects of mode spreading and sidelobes while preserving incoherent modes. Some competing convolution methods, such as *DAMAS*, would not be able to separate coherent and incoherent modes.

The usefulness of *CLEAN-SC* for in-duct spatial imaging in this setting has not been demonstrated convincingly as *CLEAN-SC* source maps are confusing arrangements of spots in this case.

Fleury and Bulté (2009)

Only first page viewed.

Dougherty and Podboy (2009)

(Only first page viewed)

An advanced phased array system, OptiNav Array 48, and a new deconvolution algorithm, TIDY, have been used to make octave band images of supersonic and subsonic jet noise produced by the NASA-Glenn SHJAR rig. The results are much more detailed than previous jet noise images. Shock cell structures and the production of screech in an underexpanded supersonic jet are observed directly. The beamforming approach has been validated by agreement between the integrated image results and the conventional microphone data.

Dougherty et al. (2010)

The deconvolution algorithms *CLEAN-SC*, *DAMAS*, *TIDY*, and eigenvalue cancellation for are used for 3D imaging of turbofan engine jet noise using a cage array and a 3D grid composed of transverse planes.

Xenaki:2010

Dougherty's change of coordinates for *DAMAS2* was investigated to identify its effect on extending the 3D range over which the shift-invariant point spread function assumption was valid.

Sarradj:2010

Orthogonal beamforming is used to generate 3D maps of sound sources using a 3D grid of scan points. This method uses eigenvalue decomposition to generate acoustic maps. Sound sources were able to be distinguished in 3D space but it appears from the plots provided that a relatively poor resolution was achieved in a direction parallel to the array axis. While this is not deconvolution of beamformed maps, it is included here since it gives similar results and is faster than deconvolution methods. Also see related work by Sarradj (2010a).

Schmitt et al. (2010)

Two inversion methods (which are not deconvolution) are used to acoustically image the interior of cabin cabin.

Sijtsma (2010)

The first benchmark problem of the 2006 AARC Phased Array Workshop was reconsidered. The spectra emitted by two acoustic sources from synthetic time data of a small array of microphones was investigated using CSEM (Cross-Spectral Estimation Method) and *CLEAN-SC* (CLEAN based on spatial Source Coherence).

Fleury and Bulté (2011)

Several deconvolution approaches are proposed to estimate the narrow-band spectra of low-Mach number uncorrelated moving sources. DAMAS, DAMAS2, CLEAN, and CLEAN-SC are extended to moving sources.

References

Basten, T., J. Wind, B. Xu, H.-E. D. Bree, and E. Druyvesteyn. Amplitude, phase, location and orientation calibration of an acoustic vector sensor array, part ii: Experiments. In *Proceedings of Meetings on Acoustics*, volume 9, Baltimore, Maryland, USA, 19-23 April, 2010.

Batlle, J., E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern recognition*, 31(7): 963–982,, 1998.

Béguet, B. and M. Robin. Device for measuring and representing noise sources inside a space. Patent No. EP2182334 (A1), 5 May, 2010.

Birchfield, S. T. Geometric microphone array calibration by multidimensional scaling. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 5, pages 157–160, Hong Kong, China, 6-10 April, 2003. IEEE.

Birchfield, S. T. and A. Subramanya. Microphone array position calibration by basis-point classical multidimensional scaling. In *IEEE Transactions on Speech and Audio Processing*, volume 13, pages 1025–1034, September, 2005.

Brüel & Kjær. URL: `www.bksv.com`. Last accessed March 2012.

Böhning, P. and H. Siller. Study of a de-convolution method for aircraft flyover measurements. In *13th AIAA/CEAS Aeroacoustics Conference (28th AIAA Aeroacoustics Conference)*, Rome, Italy, 2007.

Booth, E. R. J. and W. M. J. Humphreys. Wake acoustic microphone array results - a preliminary sneak-peak. In *WakeNet-USA Meeting*, Boca-Raton, Florida, USA, 16-17 March, 2005.

Bouguet, J.-Y. Stereo triangulation in MATLAB, 1998. URL: `http://www.multires.caltech.edu/teaching/courses/3DP/ftp/98/hw/1/triangulation.ps`. Last accessed March 2012.

Bouguet, J.-Y. *Visual methods for three-dimensional modeling*. Ph.d, California Institute of Technology, Pasadena, California, USA, May 25, 1999.

Bouguet, J.-Y. Camera calibration toolbox for MATLAB, 2008. URL: `http://www.vision.caltech.edu/bouguetj/calib_doc/index.html`. Last accessed March 2012.

Bradski, G. and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 2008.

Brandstein, M. S., J. E. Adcock, and H. F. Silverman. A practical time-delay estimator for localizing speech sources with a microphone array. *Computer Speech and Language*, 9(2):153–170, 1995.

Brooks, T. F. and W. M. Humphreys. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294(4-5):856–879, July, 2006a.

Brooks, T. F. and W. M. J. Humphreys. Effect of directional array size on the measurement of airframe noise components. In *5th AIAA/CEAS Aeroacoustics Conference*, Bellevue, Washington, 10-12 May, 1999.

Brooks, T. F. and W. M. J. Humphreys. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. In *10th AIAA/CEAS Aeroacoustics Conference*, Manchester, UK, May 10-12, 2004.

Brooks, T. F. and W. M. J. Humphreys. Three-dimensional application of DAMAS methodology for aeroacoustic noise source definition. In *11th AIAA/CEAS Aeroacoustics Conference*, Monterey, California, USA, 23-25 May, 2005.

Brooks, T. F. and W. M. J. Humphreys. Extension of DAMAS phased array processing for spatial coherence determination (DAMAS-C). In *12th AIAA/CEAS Aeroacoustics Conference*, Cambridge, Massachusetts, USA, May 8-10, 2006b.

Brown, D. C. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3): 444–462, 1966.

Brühl, S. and A. Röder. Acoustic noise source modelling based on microphone array measurements. *Journal of Sound and Vibration*, 231(3):611–617, 2000.

Brusniak, L. DAMAS2 validation for flight test airframe noise measurements. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008.

Burrus, N. Information on kinect, 2011. URL: `http://nicolas.burrus.name/index.php/Research/Kinect`. Last accessed March 2012.

Carter, G. C., A. H. Nuttall, and P. G. Cable. The smoothed coherence transform. *Proceedings of the IEEE*, 61(10):1497–1498, 1973.

Cevher, V. and J. H. McClellan. Sensor array calibration via tracking with the extended kalman filter. volume 5, pages 2817–2820, Salt Lake City, USA, 7-11 May, 2001. Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on.

Cevher, V. and J. H. McClellan. Acoustic node calibration using a moving source. *Aerospace and Electronic Systems, IEEE Transactions on*, 42(2):585–600, 2006.

Chen, M., Z. Liu, L. W. He, P. Chou, and Z. Zhang. Energy-based position estimation of microphones and speakers for ad hoc microphone arrays. pages 22–25, New Paltz, New York, USA, 21-24 October, 2007. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.

Christensen, J. J. and J. Hald. Beamforming: Bruel & kjaer technical review no.1. Technical report, Bruel & Kjaer, 2004.

Curless, B. Overview of active vision techniques. Los Angeles, UAS, 1999.

Davis, J., D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, 27(2):296–302, 2005.

Deblauwe, F. and M. Robin. Capturing a noise source in an interior enclosure. In *NAG/DAGA International Conference on Acoustics*, Rotterdam, The Netherlands, 23-26 March, 2009.

Deblauwe, F., K. Janssens, and B. Béguet. A focalization technique to extend the usability of near-field acoustc holography and beamforming. In *INTER-NOISE 2007*, Istanbul, Turkey, 28-31 August, 2007.

Döbler, D., G. Heilmann, and R. Schröder. Investigation of the depth of field in acoustic maps and its relation between focal distance and array design. In *Inter Noise*, Shanghai, China, 26-29 October, 2008.

Döbler, D., G. Heilmann, and M. Ohm. Automatic detection of microphone coordinates. In *3rd Berlin Beamforming Conference*, Berlin, Germany, 24 - 25 February, 2010.

Dougherty, B. Iterative deconvolve 3D, 2005a. URL: `http://www.optinav.com/Iterative-Deconvolve-3D.htm`. Last accessed March 2012.

Dougherty, B. Phased array technology applied to aeroacoustics. In *Keynote Lecture, 15th AIAA/CEAS Aeroacoustics Conference*, Miami, Florida, USA, 11 - 13 May, 2009.

Dougherty, B. Optinav array 24 operation, 2011. URL: `www.optinav.com/Array24Operation.pdf`. Last accessed March 2012.

Dougherty, R. P. , *Aeroacoustic Testing*, pages 63–97. Springer-Verlag, Berlin, 2002.

Dougherty, R. P. Jet noise beamforming with several techniques. In *3rd Berlin Beamforming Conference*, Berlin, Germany, 24-25 February, 2010a.

Dougherty, R. P. Personal communication, 2010b.

Dougherty, R. P. and R. W. Stoker. Sidelobe suppression for phased array aeroacoustic measurements. In *4th AIAA/CEAS Aeroacoustics Conference, (19th AIAA Aeroacoustics Conference), Toulouse, France*, Toulouse, France, 2-4 June, 1998.

Dougherty, R. P., B. E. Walker, and D. L. Sutliff. Locating and quantifying broadband fan sources using in-duct microphones. Technical report, Cleveland, Ohio, USA, 2010.

Dougherty, R. P. Spiral-shaped array for broadband imaging. Patent No. 5,838,284, 1998.

Dougherty, R. P. Advanced time-domain beamforming techniques. In *10th AIAA/CEAS Aeroacoustics Conference*, Manchester, UK, 10 - 12 May, 2004.

Dougherty, R. P. Extensions of DAMAS and benefits and limitations of deconvolution in beamforming. In *11th AIAA/CEAS Aeroacoustics Conference (26th AIAA Aeroacoustics Conference)*, Monterey, California, USA, 23 - 25 May, 2005b.

Dougherty, R. P. Phased array beamforming with 100-foot polararc microphones in a static engine noise test. In *46th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, USA, 7 - 10 January, 2008a.

Dougherty, R. P. Nacelle in-duct beamforming using modal steering vectors. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, Vancouver, British Columbia, Canada, 5-7 May, 2008b.

Dougherty, R. P. and G. G. Podboy. Improved phased array imaging of a model jet. In *15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference)*, Miami, Florida, USA, 11 - 13 May, 2009.

Ehrenfried, K. and L. Koop. Comparison of iterative deconvolution algorithms for the mapping of acoustic sources. *AIAA Journal*, 45(7):1584–1595, 2007.

Ehrenfried, K. and L. Koop. A comparison of iterative deconvolution algorithms for the mapping of acoustic sources. In *12th AIAA/CEAS Aeroacoustics Conference (27th AIAA Aeroacoustics Conference)*, Cambridge, Massachusetts, USA, 8 - 10 May, 2006.

Ettinger, E. and Y. Freund. Coordinate-free calibration of an acoustically driven camera pointing system. In *2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, Stanford, California, USA, 7-11 September, 2008.

Falcao, G., N. Hurtos, J. Massich, and D. Fofi. Projector-camera calibration toolbox, 2009. URL: `http://code.google.com/p/procamcalib`. Last accessed March 2012.

Faugeras, O. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, Cambridge, MA, USA, 1993. ISBN 0-262-06158-9.

Fleury, V. and J. Bulté. Extension of deconvolution algorithms for the mapping of moving acoustic sources. *The Journal of the Acoustical Society of America*, 129:1417–1428, 2011.

Fleury, V. and J. Bulté. Extension of beamforming algorithms to analysis moving sources. In *15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference)*, Miami, Florida, USA, 11 - 13 May, 2009.

Fleury, V., J. Bulté, and R. Davy. Determination of acoustic directivity from microphone array measurements using correlated monopoles. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, Vancouver, British Columbia, Canada, 5 - 7 May, 2008. AIAA 2008-2855.

Fusiello, A. and V. Murino. Augmented scene modeling and visualization by optical and acoustic sensor integration. *IEEE Transactions on Visualization and Computer Graphics*, 10(6):625–636, 2004.

Fusiello, A. and V. Murino. Calibration of an optical-acoustic sensor. *Machine Graphics & Vision*, 9(1/2):207–214, 2000.

GFaI. Acoustic movie of ballon pop - 3D beamforming result .. in recording studio, a. URL: `http://www.youtube.com/watch?v=yTaiL8RB4YU`. Last accessed March 2012.

GFaI. Leipziger konzerthaus - 3D scan and acoustic impulse as a 3D acoustic movie, b. URL: `http://www.youtube.com/watch?v=4YCXJB8SL2M`. Last accessed March 2012.

Guérin, S. and H. Siller. A hybrid time-frequency approach for the noise localization analysis of aircraft fly-overs. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, Vancouver, British Columbia, Canada, 5 - 7 May, 2008.

Guérin, S. and C. Weckmüller. Frequency-domain reconstruction of the point-spread function for moving sources. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008.

Guérin, S., C. Weckmüller, and U. Michel. Beamforming and deconvolution for aerodynamic sound sources in motion. In *1st Berlin Beamforming Conference*, Berlin, Germany, 21 - 22 November, 2006.

Hald, J. Combined NAH and beamforming using the same array. *Sound and Vibration*, pages 18–25, December, 2004.

Hald, J., J. Mrkholt, J. Gomes, and S. Gade. Efficient interior noise source identification based on conformal mapping using sonah holography for details on selected panels. In *15th International Congress on Sound and Vibration (ICSV15)*, pages 2627–2634, Daejeon, Korea, 6-10 July, 2008.

Hartley, R. and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, West Nyack, New York, 2004.

Heilmann, G., A. Meyer, and D. Döbler. Time-domain beamforming using 3D-microphone arrays. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008a.

Heilmann, G., A. Meyer, and D. Döbler. Beamforming in the time-domain using 3D-microphone arrays. In *XIXth Biennial Conference of the New Zealand Acoustical Society*, Auckland, New Zealand, 27-28 November, 2008b.

Hennecke, M., T. Plotz, G. A. Fink, J. Schmalenstroer, and R. Hab-Umbach. A hierarchical approach to unsupervised shape calibration of microphone array networks. In *15th Workshop on Statistical Signal Processing*, pages 257–260, Cardiff, Wales, UK, 31 August - 3 September, 2009. IEEE.

Hoctor, R. T. and S. A. Kassam. The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging. *Proceedings of the IEEE*, 78(4):735–752, 1990.

Holm Acoustics. WM-61A calibration file, 2009. URL: `http://www.holmacoustics.com/downloads/MicrophoneCalibration/Mic_WM61A.cal`. Last accessed March 2012.

Hörster, E., R. Lienhart, W. Kellermann, and J. Y. Bouguet. Calibration of visual sensors and actuators in distributed computing platforms. In *Proceedings of the Computer Vision for Interactive and Intelligent Environment (CVIIE05)*, pages 39 – 50, Lexington, Kentucky, USA, 17-18 November, 2005.

Humphreys, W. M., T. F. Brooks, W. W. Hunter, and K. R. Meadows. Design and use of microphones directional arrays for aeroacoustic measurements. In *36st AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, USA, 12-15 January, 1998. NASA Langley Research Center.

Hurtós, N., X. Cufi, and J. Salvi. Calibration of optical camera coupled to acoustic multibeam for underwater 3D scene reconstruction. In *OCEANS 2010 IEEE*, pages 1–7, Sydney, Australia, 24-27 May, 2010.

Irimia, C., F. Deblauwe, K. Jannens, Z. Juhos, and S. Ignat. Improving the localization of noise sources inside a vehicle. *Scientific Bulletin: Automotive series, year XV*, B (19), 2009.

Jacovitti, G. and G. Scarano. Discrete time techniques for time delay estimation. *IEEE Transactions on Signal Processing*, 41(2):525–533, December, 1993.

Jaeckel, O. Strengths and weaknesses of calculating beamforming in the time domain. In *1st Berlin Beamforming Conference*, Berlin, Germany, 21 - 22 November, 2006.

Jager, P. D., M. Trinkle, and A. Hashemi-Sakhtsari. Automatic microphone array position calibration using an acoustic sounding source. In *4th IEEE Conference on Industrial Electronics and Applications (ICIEA 2009)*, pages 2110–2113, Xi'an, China, 25-27 May, 2009. doi: 10.1109/ICIEA.2009.5138521.

Jarvis, R. A. A perspective on range finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2):122–139, 1983.

Jordan, P., J. Fitzpatrick, and J. Valiere. Measurement of an aeroacoustic dipole using a linear microphone array. *Journal of the Acoustis Society of America*, 111(3):1267–1273, March, 2002.

Kenner, H. *Geodesic math and how to use it.* Univ of California Pr., 1976.

Klein, G. and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07)*, Nara, Japan, 13-16 November, 2007.

Klein, G. and D. Murray. Parallel tracking and mapping on a camera phone. In *Proc. Eigth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'09)*, Orlando, Florida, USA, 19 - 23 October, 2009.

Knapp, C. and G. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4):320–327, 1976.

Koerber, M. A. and D. R. Fuhrmann. Array calibration by Fourier series parameterization: Scaled principal components method. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-93)*, volume 4, pages 340–343, Minneapolis, Minnesota, USA, 27-30 April, 1993.

Krotkov, E. Focusing. *International Journal of Computer Vision*, 1(3):223–237, 1987.

Lamotte, L. and F. Deblauwe. Caractérisation de sources acoustiques en 3D par méthode inverse avec une sphére rigide. In *10éme Congrés Franais d'Acoustique*, Lyon, France, 2010.

Lamotte, L., M. Robin, and F. Deblauwe. Noise mapping and sound quantification in the space using spherical array. In *EURONOISE 2009*, Edinburgh, Scotland, UK, October 26-28, 2009.

Lanman, D. and G. Taubin. Build your own 3D scanner: optical triangulation for beginners, 2009a. URL: `http://mesh.brown.edu/byo3d/`. Last accessed March 2012.

Lanman, D. and G. Taubin. Build your own 3D scanner: 3D photography for beginners. In *SIGGRAPH 2009 Course Notes*, New Orleans, Louisiana, USA, 5 August, 2009b.

Lanslots, J., F. Deblauwe, and K. Janssens. Selecting sound source localization techniques for industrial applications. *Sound and Vibration*, 44:6–9, June, 2010.

Lauterbach, A., K. Ehrenfried, L. Koop, and S. Loose. Procedure for the accurate phase calibration of a microphone array. In *15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference)*, Miami, Florida, USA, 11 - 13 May, 2009.

Levoy, M. The digital michelangelo project. In *Proc. Second International Conference on 3-D Digital Imaging and Modeling*, pages 2–11, Ottawa, Ontario, Canada, 4-8 October, 1999.

Li, B., D. Yang, L. Shao, and X. Lian. Development of acoustic video camera system based on binocular vision and short-time beamforming. *INTER-NOISE 2008*, 26 29 October, 2008.

Li, Z., B. Curless, and S. M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission*, pages 24–36, Padova, Italy, 19-21 June, 2002.

Liu, Y., A. P. Dowling, and A. R. Quayle. Numerical simulation of beamforming correction for dipole source identification. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008.

Lo, J. and S. Marple Jr. Eigenstructure methods for array sensor localization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '87)*, volume 12, pages 2260–2263, Dallas, Texas, USA, 6-9 April, 1987.

Ma, Y., S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3D Vision: From Images to Geometric Models*. Springer Verlag, New York, 2003.

Maffei, M. and A. Bianco. Improvements of the beamforming technique in Pininfarina full scale wind tunnel by using a 3D scanning system. *SAE International Journal of Materials and Manufacturing*, 1(1):154–168, 2008.

Manthe, C., A. Meyer, and F. Gielsdorf. Geometric calibration of acoustic camera star48 array. In *Integrating Generations: FIG Working Week*, Stockholm, Sweden, 14-19 June, 2008.

Maynard, J. D. and E. G. Williams. Nearfield acoustic holography: theory of the generalized holography and the development of NAH. *Journal of the Acoustical Society of America*, 78(4):1395–1413, 1985.

McCowan, I., M. Lincoln, and I. Himawan. Microphone array shape calibration in diffuse noise fields. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3): 666–670, 2008.

Meyer, A. and D. Döbler. Noise source localization within a car interior using 3D-microphone arrays. In *1st Berlin Beamforming Conference*, Berlin, Germany, 21-22 November, 2006.

Michel, U. and S. Funke. Inverse method for the acoustic source analysis of an aeroengine. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19-20 February, 2008a.

Michel, U. and S. Funke. Noise source analysis of an aeroengine with a new inverse method sodix. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, Vancouver, British Columbia, Canada, 5 - 7 May, 2008b.

Moebus, M. *Object Detection and Classification for Mobile Platforms Using 3D Acoustic Imaging*. Ph.D. thesis, TU Darmstadt, Institut für Nachrichtentechnik, FG Signalverarbeitung, 2011.

Moses, R. and R. Patterson. Self-calibration of sensor networks. *Unattended Ground Sensor Technologies and Applications IV (Proceeding of SPIE)*, 4743:491, 2-5 April, 2002.

Negahdaripour, S. Calibration of DIDSON forward-scan acoustic video camera. In *Proceedings of OCEANS 2005 MTS/IEEE*, volume 2, pages 1287–1294, Washington D.C, USA, 18-23 September, 2005.

Negahdaripour, S. A new method for calibration of an opti-acoustic stereo imaging system. In *OCEANS 2010 MTS/IEEE Opti-acoustic stereo imaging, system calibration*, pages 1–7, Seattle, Washington, USA, 20-23 September, 2010.

Negahdaripour, S., H. Sekkati, and H. Pirsiavash. Opti-acoustic stereo imaging, system calibration and 3-D reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pages 1–8, Minneapolis, Minnesota, USA, 17-22 June, 2007.

Ng, B. C. and A. Nehorai. Optimum active array shape calibration. In *Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 893–897, Pacific Grove, California, USA, 4-6 November, 1991.

Ng, B. C. and C. M. S. See. Sensor-array calibration using a maximum-likelihood approach. *IEEE Transactions on Antennas and Propagation*, 44(6):827–835, June, 1996.

Ng, B. C. and W. Ser. Array shape calibration using sources in known locations. In *ICCS/ISITA '92. 'Communications on the Move'*, volume 2, pages 836–840, Singapore, 16-20 November, 1992.

O'Donovan, A., D. Ramani, and J. Neumann. Sensing the world with arrays of microphones and cameras. 2-7 September, 2007a.

O'Donovan, A., R. Duraiswami, and J. Neumann. Microphone arrays as generalized cameras for integrated audio visual processing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07),*, pages 1–8, Minneapolis, Minnesota, USA, 17-22 June, 2007b.

Oerlemans, S. *Detection of Aeroacoustic Sound Sources on Aircraft and Wind Turbines.* PhD thesis, University of Twente, Enschede, Twente, Netherlands, September, 2009.

OptiNav. URL: `www.optinav.com`. Last accessed March 2012.

Orghidan, R. *Catadioptric Stereo based on Structured Light Projection.* PhD thesis, Universitat de Girona, December, 2005.

Panasonic. Omnidirectional back electret condenser microphone cartridge WM-61A. URL: `http://www.panasonic.com/industrial/components/pdf/em06_wm61_a_b_dne.pdf`. Last accessed March 2012.

Papamoschou, D. Imaging of directional distributed noise sources. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, Vancouver, British Columbia, Canada, May 5-7, 2008.

Papamoschou, D. and A. Dadvar. Localization of multiple types of jet noise sources. In *12th AIAA/CEAS Aeroacoustics Conference (27th AIAA Aeroacoustics Conference)*, Cambridge, Massachusetts, USA, 8-10 May, 2006.

Pascal, J.-C. and J.-F. Li. Resolution improvement of data-independent beamformers. In *INTER-NOISE 2007*, Istanbul, Turkey, 28-31 August, 2007.

Price, K. Annotated computer vision bibliography. URL: `http://www.visionbib.com/bibliography/contents.html`. Last accessed March 2012.

Quayle, A. R., A. Dowling, W. R. Graham, and H. Babinsky. Comparison of source estimation algorithms / methods in closed tunnel noise measurements. In *1st Berlin Beamforming Conference*, Berlin, Germany, 21-22 November, 2006.

Quayle, A., W. Graham, A. Dowling, H. Babinsky, and Y. Liu. Mitigation of beamforming interference from closed wind tunnels using CLEAN-SC. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008.

Ramachandran, R., H. Patel, G. Raman, and R. Dougherty. Localization of wind turbine noise sources using a compact microphone array with advanced beamforming algorithms. In *4th Berlin Beamforming Conference*, 22-23 February, 2012.

Ravetta, P. A. *LORE Approach for Phased Array Measurements and Noise Control of Landing Gears*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA, 28 November, 2005.

Ravetta, P. A., R. A. Burdisso, and W. F. Ng. Noise source localization and optimization of phased array results. In *12th AIAA/CEAS Aeroacoustics Conference*, Cambridge, Massachusetts, USA, May 2006, 2006.

Raykar, V. C., I. Kozintsev, and R. Lienhart. Position calibration of audio sensors and actuators in a distributed computing platform. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, pages 572–581, Berkeley, Califoria, USA, 28 November, 2003.

Raykar, V. C. and R. Duraiswami. Automatic position calibration of multiple microphones. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004)*, volume 4, pages 69–72, Montreal, Canada, 17-21 May, 2004.

Raykar, V. C., I. V. Kozintsev, and R. Lienhart. Position calibration of microphones and loudspeakers in distributed computing platforms. *IEEE Transactions on Speech and Audio Processing*, 13(1):70–83, 2005.

Redondi, A., M. Tagliasacchi, F. Antonacci, and A. Sarti. Geometric calibration of distributed microphone arrays. In *IEEE International Workshop on Multimedia Signal Processing (MMSP'09)*, pages 1–5, Rio De Janeiro, Brazil, 5-7 October, 2009.

Ricolfe-Viala, C. and A. Sanchez-Salmeron. Lens distortion models evaluation. *Applied optics*, 49(30):5914–5928, 2010.

Robbel, P. The kinect sensor in mobile robotics: Initial experiments. URL: `http://www.youtube.com/watch?v=dRPEns8MS2o&feature=player_embedded#at=71Personal`. Last accessed March 2012.

Robin, M. and B. Béguet. Acoustic source localization inside cabins. In *Automotive and Railway Comfort Congress*, Le Mans, France, 19-20 November, 2008.

Rockah, Y. and P. Schultheiss. Array shape calibration using sources in unknown locations–part i: Far-field sources. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3):286–299, 1987a.

Rockah, Y. and P. Schultheiss. Array shape calibration using sources in unknown locations–part ii: Near-field sources and estimator implementation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(6):724–735, 1987b.

Romenskiy, I. and O. Jaeckel. Improvement of source separation for phased microphone array measurements. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008.

Roth, P. R. Effective measurements using digital signal analysis. *Spectrum, IEEE*, 8(4): 62–70, 1971.

Sachar, J. M., H. F. Silverman, and W. R. Patterson. Position calibration of large-aperture microphone arrays. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'02).*, volume 2, pages 1797–1800, Orlando, Florida, USA, 2002.

Sachar, J. M., H. F. Silverman, and W. R. Patterson. Microphone position and gain calibration for a large-aperture microphone array. *IEEE Transactions on Speech and Audio Processing*, 13(1):42–52, 2005.

Salvi, J., J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004.

Sarradj, E. A fast signal subspace approach for the determination of absolute levels from phased microphone array measurements. *Journal of Sound and Vibration*, 329(9): 1553–1569, 2010a.

Sarradj, E. Quantitative source spectra from acoustic array measurements. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008.

Sarradj, E. Improving speed with orthogonal beamforming. In *3rd Berlin Beamforming Conference*, Berlin, Germany, 24 - 25 Feb, 2010b.

Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, March, 1986.

Schmitt, A., L. Lamotte, and F. Deblauwe. Source identification inside cabin using inverse methods. In *3rd Berlin Beamforming Conference*, Berlin, Germany, 24 - 25 Feb, 2010.

Schröder, R. Threedimensional beamforming: Considering the effects of covered microphones. Acoustic Camera - Newsletter: Edition 3, December, 2009.

Schulz, D., G. Thompson, C. L. D. Lo, R. Goubran, and M. Nasr. System and method of self-discovery and self-calibration in a video conferencing system. US Patent No. 7,403,217, 22 July, 2008.

Seymour, L., C. Cowan, and P. Grant. Bearing estimation in the presence of sensor positioning errors. volume 12, pages 2264–2267, Dallas, Texas, USA, April, 1987. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '87).

Sijtsma, P. CLEAN based on spatial source coherence. In *13th AIAA/CEAS Aeroacoustics Conference*, Rome, Italy, 21-23 May, 2007a.

Sijtsma, P. CLEAN based on spatial source coherence. *International Journal of Aeroacoustics*, 6(4):357 374, 2007b.

Sijtsma, P. CLEAN based on spatial source coherence. In *2nd Berlin Beamforming Conference*, Berlin, Germany, 19 - 20 February, 2008.

Sijtsma, P. AARC benchmark 1 revisited. In *3rd Berlin Beamforming Conference*, Berlin, Germany, 24 - 25 Feb, 2010.

Subbarao, M., N. B. Agarwal, and G. Surya. Application of spatial-domain convolution/deconvolution transform for determining distance from image defocus. Technical report, Computer Vision Laboratory, Dept, 1992.

Suzuki, T. Generalized inverse beam-forming algorithm resolving coherent/incoherent, distributed and multipole sources. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, Vancouver, British Columbia, Canada, 5 - 7 May, 2008.

Szeliski, R. *Computer vision: Algorithms and applications*. Springer-Verlag, New York Inc, 2010.

Teal, P. D. and M. A. Poletti. Adaptive phase calibration of a microphone array for acoustic holography. *Journal of the Acoustical Society of America*, 127(4):2368–2376, 2010.

Thrun, S. Affine structure from sound. *Advances in Neural Information Processing Systems*, 18:13531360, 2006.

Tsai, R. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, August, 1987.

Underbrink, J. R. , *Aeroacoustic Testing*, pages 98–217. Springer-Verlag, Berlin, Germany, 2002.

Upton, R., K. Haddad, and J. Sorensen. Conformal mapping techniques for consumer products. *Sound and Vibration*, 42(7):8–11, 2008.

Valente, S., F. Antonacci, M. Tagliasacchi, A. Sarti, and S. Tubaro. Self-calibration of two microphone arrays from volumetric acoustic maps in non-reverberant rooms. In *4th International Symposium on Communications, Control and Signal Processing (ISCCSP2010)*, Limassol, Cyprus, 3-5 March, 2010.

Vilarnau, N. H. *Integration of optical and acoustic sensor data for 3D underwater scene reconstruction*. PhD thesis, University of Girona, Girona, Spain, 2009.

Wang, F. Y., H. Wassaf, R. P. Dougherty, K. Clark, A. Gulsrud, N. Fenichel, and W. H. Bryant. Passive wake acoustics measurements at denver international airport. In *4th NASA Integrated CNS Conference and Workshop*, Fairfax, Virginia, USA, April 26-30, 2004a.

Wang, Y., J. Li, P. Stoica, M. Sheplak, and T. Nishida. Wideband RELAX and wideband CLEAN for aeroacoustic imaging. *Journal of the Acoustical Society of America*, 115 (2):757–767, 2004b.

Weiss, A. J. and B. Friedlander. Array shape calibration using eigenstructure methods. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):1958–1966, December, 1989.

Weiss, A. J. and B. Friedlander. Array shape calibration using eigenstructure methods. *Signal Processing*, 22(3):251–258, 1991.

Wiens, T. and S. Bradley. A comparison of time delay estimation methods for periodic signals, 2009. URL: `http://www.nutaksas.com/papers/wiens_dsp_delay.pdf`. Last accessed March 2012.

Wiens, T. Kasami sequences, m-sequences, linear feedback shift registers, 2009a. URL: `http://www.mathworks.com/matlabcentral/fileexchange/22716-kasami-sequences-m-sequences-linear-feedback-shift-registers`. Last accessed March 2012.

Wiens, T. Subsample delay estimation, 2009b. URL: `http://www.mathworks.com/matlabcentral/fileexchange/25210-subsample-delay-estimation`. Last accessed March 2012.

Xenaki, A., F. Jacobsen, E. Tiana-Roig, and E. F. Grande. Improving the resolution of beamforming measurements on wind turbines. In *Proceedings of 20th International Congress on Acoustics (ICA)*, Sydney, Australia, 2327 August, 2010.

Xu, B., J. Wind, H.-E. D. Bree, T. Basten, and E. Druyvesteyn. Amplitude, phase, location and orientation calibration of an acoustic vector sensor array, part i: Theory. *Proceedings of Meetings on Acoustics*, 9(1), 2010.

Yang, D., Z. Wang, B. Li, and X. Lian. Development and calibration of acoustic video camera system for moving vehicles. *Journal of Sound and Vibration*, 330(11):2457–2469, May, 2011.

Yardibi, T., C. Bahr, N. S. Zawodny, F. Liu, L. N. Cattafesta, and J. Li. Uncertainty analysis of the standard delay-and-sum beamformer and array calibration. In *15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference)*, Miami, Florida, USA, 11 - 13 May, 2009.

Yardibi, T., N. S. Zawodny, C. Bahr, F. Liu, L. N. Cattafesta, and J. Li. Comparison of microphone array processing techniques for aeroacoustic measurements. *International Journal of Aeroacoustics*, 9(6):733–762, 2010.

Yardibi, T. *Source Localization and Power Estimation in Aeroacoustic Noise Measurements*. PhD thesis, University of Florida, Miami, USA, 2009.

Yardibi, T., J. Li, and L. N. Cattafesta. Sparsity constrained deconvolution approaches for acoustic source mapping. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, Vancouver, British Columbia, Canada, 5 - 7 May, 2008a.

Yardibi, T., J. Li, P. Stoica, and L. N. Cattafesta. Sparsity constrained deconvolution approaches for acoustic source mapping. *The Journal of the Acoustical Society of America*, 123(5):2631–2642, 2008b.

Zhang, L. and X. Wu. On the application of cross correlation function to subsample discrete time delay estimation. *Digital Signal Processing*, 16(6):682–694, 2006.

Zhang, R., P. S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.

Zhang, S. *High-resolution, Real-time 3-D Shape Measurement*. Ph.D. thesis, Stony Brook University, New York, USA, May, 2005.

Zhang, Z. A flexible new technique for camera calibration. Technical report, Microsoft Research, Microsoft Corporation, December, 1998.

Zotkin, D., R. Duraiswami, H. Nanda, and L. S. Davis. Multimodal tracking for smart videoconferencing. In *IEEE International Conference on Multimedia and Expo (ICME 2001)*, pages 36–39, Tokyo, Japan, 22-25 August, 2001.

Zotkin, D. N., V. C. Raykar, R. Duraiswami, and L. S. Davis. chapter 7: Multimodal tracking for smart videoconferencing and video surveillance. Artech House, Norwood, Massachusetts, USA, 2007.