THE ANALYSIS AND IMPROVEMENT OF FOCUSED SOURCE REPRODUCTION WITH WAVE FIELD SYNTHESIS

Robert Oldfield

Acoustics Research Centre, University of Salford,
Salford, UK

Submitted in Partial Fulfilment of the Requirements of the Degree of
Doctor of Philosophy, April 2013

For Susie

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

There are so many people that deserve thanks for the completion of this thesis. Particular thanks must go to Dr. Ian Drumm for his fantastic supervision during this project. I'm grateful for his time, encouragement, guidance and jokes all of which were never in short supply. Thanks are also due to in the Acoustic Research Centre at the University of Salford who have all been prepared to offer help and encouragement when needed.

It shouldn't be underestimated; the power of supportive family and friends and this I can say I have in abundance. Particular thanks to my fantastic parents for always believing in me, encouraging and inspiring me, your support and wisdom have been fantastic. Huge thanks go to my phenomenal wife, Susie who has had to walk through every step of this PhD with me and has been the best wife and friend anyone could ask for through it all. But the one to whom I am most thankful is Almighty God in whose strength and grace I can do all things!

ABSTRACT

This thesis presents a treatise on the rendering of focused sources using wave field synthesis (WFS). The thesis describes the fundamental theory of WFS and presents a thorough derivation of focused source driving functions including, monopoles, dipoles and pistonic sources. The principle characteristics of focused sources including, array truncation, spatial aliasing, pre-echo artefacts, colouration and amplitude errors are analysed in depth and a new spatial aliasing criterion is presented for focused sources. Additionally a new secondary source selection protocol is presented allowing for directed and symmetrically rendered sources. This thesis also describes how the low frequency rendering of focused sources is limited by the focusing ability of the loudspeaker array and thus derives a formula to predict the focusing limits and the corresponding focal shift that occurs at low frequencies and with short arrays. Subsequently a frequency dependent position correction is derived which increases the positional accuracy of the source. Other characteristics and issues with the rendering of focused sources are also described including the use of large arrays, rendering of moving focused sources, issues with multiple focused sources in the scene, the phase response, and the focal point size of focused sound field.

The perceptual characteristics are also covered, with a review of the literature and a series of subjective tests into the localisation of focused sources. It is shown that an improvement in the localisation can be achieved by including the virtual first order images as point sources into the WFS rendering.

Practical rendering of focused sources is generally done in compromised scenarios such as in non-anechoic, reverberant rooms which contain various scattering objects. These issues are also covered in this thesis with the aid of finite difference time domain models which allow the characterisation of room effects on the reproduced field, it is shown that room effects can actually even out spatial aliasing artefacts and therefore reduce the perception of colouration. Scattering objects can also be included in the model, thus the effects of scattering are also shown and a method of correcting for the scattering is suggested. Also covered is the rendering of focused sources using elevated arrays which can introduce position errors in the rendering.

GLOSSARY OF SYMBOLS

Throughout this thesis letters or symbols in bold type describe vectors, a hat above the vector symbolises that it is a unit vector. A tilde above a symbol or letter denotes that the quantity is in the wavenumber domain. Other commonly used letters and symbols used in this thesis are as follows.

c Speed of sound ($343 ms^{-1}$)

f Frequency, *(Hz)*

f_{alias} Spatial aliasing frequency, *(Hz)*

$H_v^{(n)}$ Hankel function order, n and of the v^{th} kind

$j_m(..)$ m^{th} order spherical Bessel function of the first kind

$J_m(..)$ m^{th} order Bessel function of the first kind

j $\sqrt{-1}$

k Wavenumber, $k = 2\pi f / c$, (m^{-1})

t Time (s)

V Volume (m^3)

$\mathbf{x}_0$ Secondary source position

$\mathbf{x}_S$ Virtual source position

$\mathbf{x}_{Ref}$ Reference point where amplitude errors are minimal

$y_m(..)$ The m^{th} order spherical Bessel function of the second kind

$Y_m(..)$ The m^{th} order Bessel function of the second kind,

λ Wavelength, (m)

Δ_{LS} Loudspeaker spacing, *(m)*

ρ_0 Density of air ($1.21 kgm^{-3}$)

ω Angular frequency, $\omega = 2\pi f$, (s^{-1})

LIST OF ACRONYMS

ASW	Apparent source width
C_{80}	Clarity Index
D_{50}	Deutlichkeit
DML	Distributed Mode Loudspeakers
EDT	Early Decay Time
FDTD	Finite Difference Time Domain
HINT	Hearing In Noise Test
HOA	Higher Order Ambisonics
HRTF	Head-Related Transfer Function
IACC	Interaural Cross-correlation Coefficient
ILD	Interaural Level Difference
ITD	Interaural Time Difference
LEV	Listener envelopment
MADI	Multi-channel Audio Digital Interface
NFC-HOA	Near Field Compensated Higher Order Ambisonics
PC	Personal Computer
PML	Perfectly Matched Layer absorbing boundary in FDTD
RMS	Root Mean Square
RT60	Reverberation time: The time taken for sound energy to decay *60dB*
WFS	Wave Field Synthesis
3D	Three-dimensional
2D	Two-dimensional
2.5D	2.5-dimensional representing a cross between 2D and 3D systems

LIST OF FIGURES

LIST OF TABLES

1 Introduction

The wave field synthesis (WFS) technique for spatial audio reproduction utilises many loudspeakers to synthesise arbitrary wave fronts throughout the listening space and consequently offers many advantages over standard methods as it is based upon physical principles rather than psychoacoustical approximations. A combination of many sources rendered using wave field synthesis results in the accurate reproduction or synthesis of an entire sound scene with each virtual source generating the same auditory cues as if a real source were actually at that location. Wave field synthesis therefore does not attempt to fool the brain using psychoacoustical tricks to synthesis auditory cues but rather recreates the exact cues that the listener would experience if they were in the original scene. WFS can thus be considered a volumetric approach to spatial audio and therefore has the advantage of not having a sweet spot, as the recreated wave fronts propagate throughout the listening space so the source will be localised to the same position irrespective of listener location.

Perhaps one of the most striking advantages of this technique however is the ability to create sources in front of the loudspeaker array. Such *focused source*s are in fact not real sources but rather they have the auditory appearance of a real source as all the acoustic energy is focused on to one point and thus re-radiates giving the impression of an acoustic source at that location. The rendering of a focused source allows listeners much more interaction with their environment than would otherwise be possible as they provide a physical audio event which listeners can localise and walk around. This is a remarkable feature of WFS and indeed any sound field synthesis technique and provides an increased realism and depth-of-scene for any synthesised sound environments, allowing sources to be positioned very close to listeners. When combined with 3D or holographic projection techniques this can therefore result in very realistic virtual scenes (Boone et al., 1996). Focused sources have also been shown to be useful for creating virtual loudspeaker signals in hybrid systems (Menzel et al., 2005) and virtual sound field synthesis systems (Spors and Ahrens, 2010a).

Despite the unique and striking nature of focused sources in WFS, there has been relatively little attention given specifically to this topic to date. This thesis presents a treatise on WFS focused sources; analysing and characterising the physical and perceptual attributes and highlighting some improvements in the rendering and analysis thereof.

1.1 Aims of research

The aim of this research was the full characterisation of focused source reproduction in wave field synthesis. Following on from this characterisation, it was aimed that improvements could be made to the rendering of focused sources such that they could be reproduced to a greater degree of accuracy and localised over a wider listening area.

1.2 Focus of thesis

The thesis therefore focuses specifically on the wave field synthesis reproduction of sources in front of the loudspeaker array, termed ‘focused sources’ and the subsequent improvements in the rendering thereof.

1.3 Original contributions

The following original contributions are made in this thesis:

- Derivation of a secondary source selection criteria specific to focused sources for directional and symmetrical propagating virtual sources.
- A definition of the un-aliased zone and derivation of a new aliasing condition for focused sources.
- The prediction of, and correction for, the position error in focused source rendering caused by diffraction errors and resulting focal shift in the rendered virtual source at low frequencies or with short arrays.
- The derivation of a direct-to-pre-echo ratio for focused sources which can be used as a rough measure of localisation performance.
- The characterising of the localisation properties of focused sources with respect to listener angle.
- A perceptual improvement in the localisation of focused sources based on the inclusion of first order image sources into the rendering for a focused source.
- An analysis of room effects on focused sources using the finite difference time domain technique.
- Analysis of the effect of scattering objects on the focused sound field.
- The reduction in the effects of scattering in the convergent part of the sound field using time reversal techniques.

1.4 Structure of thesis

The thesis begins with a brief overview of spatial audio techniques and acoustic localisation mechanisms which contextualise the wave field synthesis, sound field reconstruction system used for this work. Chapter 3 presents an in-depth theoretical background of wave field synthesis describing the relevant driving functions, source selection criteria and some of the errors in the reproduced wave field. This chapter also reviews some of the current major topics of research in the area of wave field synthesis and includes the description of four wave field synthesis systems that have been designed, built, calibrated and commissioned at the University of Salford as part of this project. Chapter 4 deals specifically with focused source reproduction in wave field synthesis; presenting the important driving functions and the errors

associated with their rendering. The chapter pays specific attention to the definition of the un-aliased region in the reproduced wave field and the correction for the source position errors resulting from the focal shift phenomena. The chapter also includes the derivation of a direct-to-pre-echo ratio which is used for the analysis of the rendering of focused sources using large arrays towards the end of the chapter. Chapter 5 reviews some of the perceptual attributes of focused sources and presents results from a localisation test. In this chapter the key findings in the literature are reviewed with a brief description of the experiments performed therein. The chapter also describes a localisation test to determine the acoustic view angle of focused sources and demonstrates the improvements that can be made in the reproduction of focused source with the addition of the first order image sources into the rendering. Chapter 6 then describes the implementation of focused sources in compromised/practical environments. The chapter introduces the finite difference time domain as a method of modelling sound reproduction in rooms with arbitrary geometry and frequency dependent absorbing boundaries. This model is used in the chapter to analyse the effects of the reproduction room on the focused source reproduction. Also analysed in this chapter is the effect of both simple and complex scattering objects. A scattering correction approach is presented based upon the finite difference model and time reversal techniques. The chapter also includes an analysis of the effects of using elevated loudspeakers on the reproduction of focused sources. Chapter 7 provides a summary of the work performed and outlines some further work that will be undertaken as a result of these findings.

2 Spatial Audio Perception and Techniques

2.1 Introduction

This chapter describes some of the fundamental theory of spatial audio techniques and perception. The chapter will both put into context and provide a theoretical basis for the work described in this thesis.

Since the beginning of audio reproduction technology in 1876 with Bell's telephone (Bell, 1876) and 1877 with Edison's phonograph (Edison, 1878), there has been an ever growing desire by audio developers and enthusiasts alike to recreate sound scenes to a high level of accuracy in order to either give the listener the impression that they are at the performance or conversely that the performers are with them. Commonly people have sought an increased timbral fidelity, attempting to recreate audio such that it sounds as similar as possible to the original performance or the

intended audio impression/experience that the producer wishes to create. Recently innovation has been more keenly focused on the spatial fidelity of the sound scene, whether it is in order to give the impression that the listener is in the space where the recording was made or whether it is making the synthesised scene as believable as possible. Human spatial audio perception and the recreation of the cues thereof is vitally important as it allows us to interpret, interact and make sense of our environment whether real or virtual. The goal of spatial audio reproduction is to improve the listener experience and generally to create a sense of realism. In the case of cinema and computer games this could even be an experience of 'hyper-reality' where the spatial audio scene is created artistically to evoke certain impressions even if this is not exactly how the scene would be experienced in reality. This audio scene fidelity encompasses many psychoacoustical properties such as localisation, spaciousness/envelopment (Morimoto and Maekawa, 1989), apparent source width (Griesinger, 1997) and various other depth of scene cues, some of which will be mentioned in the following. (Re)creating an audio scene therefore requires that both the spatially discrete sources and the ambient sound be played back in a realistic manner as the brain is very good at interpreting all of the information to compose a picture of what the scene is like, if there is an incongruity in the cues then this will hinder the sense of presence and immersion that the listener will experience of the scene.

Spatial audio systems find applications in many areas such as virtual reality and computer games but really the developments have been fuelled primarily by both the home and domestic cinema markets; as such in recent years there has been a constant development in the technology and systems used. Many of the systems that have been developed for this purpose to exploit various psychoacoustic mechanisms to evoke the impression of an acoustic source at a given location with the minimum number of sources used for the reproduction.

Section 2.2 describes some of the principle localisation mechanisms that are important for spatial audio rendering whilst section 2.3 presents an overview of the major developments in the area of spatial audio reproduction systems up to now.

2.2 Audio localisation

In order to gain a full understanding of spatial audio perception it is useful to define some terms from the outset that describe the relationship between the physical outputs from real or reproduced sources and the perception thereof. A *sound event* relates to the physical outputs from either a real or reproduced source while the term *auditory event* relates to the perception of the sound event. Multiple sound events form a *sound scene* which in turn evokes the perception of an *auditory scene* at the listener.

The interpretation of an auditory scene and in particular the localisation is based upon many *cues*. Cues describe the auditory information used by the brain to interpret the auditory event or scene. Although there are some useful monaural cues, much of human spatial audio perception and cognition relies upon binaural cues resulting from differences between the signals arriving at the two ears; whether, timing, level or spectral differences, the signals from both ears arrive at the brain down the auditory nerve where higher cognitive processes determine the evocation and perception of the auditory scene. Because of this binaural processing, it is apparent that the recreation of any audio scene cannot be properly done monaurally. The recreation of these binaural cues is therefore done using multiple loudspeakers. This has been known for many years but as yet there is not a single technique that has risen as being the *de facto* method of spatial audio reproduction. Principally however, spatial audio systems work in a similar manner, seeking to reproduce the necessary psychoacoustical cues to evoke the sense of an audio source coming from a specific location. The debate over which cues are the most important and the methods of which these cues are created is still active in the literature and is the reason for the many different spatial audio systems that are currently used, (as described in section 2.3).

This section discusses briefly the principle human localisation mechanisms that will be addressed and mentioned throughout this thesis.

2.2.1 Lateral localisation

Human auditory localisation is better in the lateral plane than elevation localisation in the median plane, this is the result of both ears being in the same lateral plane, which provide more information. In the early 1900's Lord Rayleigh formed his Duplex theory (Strutt, 1907) which states that *interaural time differences* (ITD) and *interaural level differences* (ILD) are the principle localisation cues in the lateral plane with ITD being a predominantly low frequency cue and ILD being a higher frequency cue (above 500*Hz*). Rayleigh's experiments were done with sinusoidal signals generated using tuning forks and as such the results (having since been proven with additional experiments by other researchers) constitute only a specific case. The reality is that natural sounds are more complex both in terms of spectral content and also amplitude envelope thus requiring a more complex interpretation of ILD and ITD cues.

2.2.1.1 Interaural time difference

The temporal displacement of the signals arriving at one ear with respect to the other is often referred to as the *interaural time difference* (ITD). This temporal displacement is generated when a sound is incident off the central axis as shown in Fig. 2.1. A source at 90˚ corresponds to the maximum possible ITD which has a value in the region of about *650-700µs*. As noted by Blauert (1997) the lateral displacement of the source increases linearly with time delays up to about *630µs* above this value the difference in ITD causes a smaller rate of change in lateral position until about *1ms* where there is no change at all.

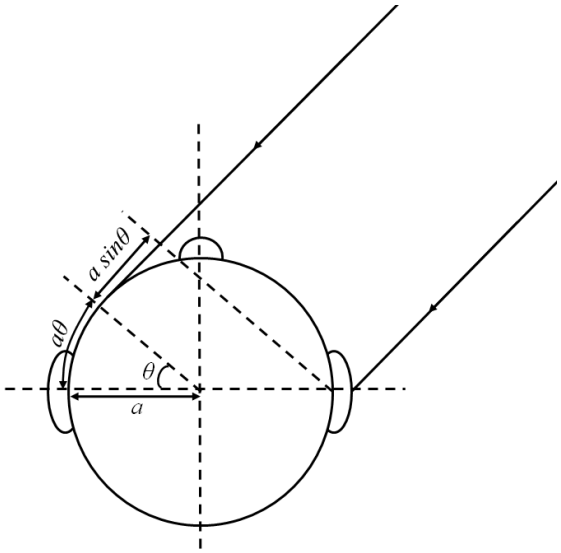


Fig. 2.1 Geometry used for the derivation of ITD calculations

$$ITD = \frac{a(\theta + \sin\theta)}{c} \qquad (2.1)$$

ITDs are particularly effective for signals with significant transient content like speech or other impulsive sounds where, the brain can easily determine the difference in arrival times. It should also be noted that even though ITD has been presented here as a frequency independent function it does actually vary slightly with frequency (Feddersen, 1957; Moore, 2003). Most real signals contain lots of frequency components which could each have a slightly different ITD therefore making the localisation a more complex process, however for most purposes *(2.1)* can be considered a good approximation.

Practically the ITD is often calculated using the *interaural cross-correlation coefficient* (IACC). Two signals are recorded at the ears of a human or dummy head and the cross-correlation between them is computed, with the peak in the function occurring at the delay between the two signals which corresponds to the ITD. An example ITD of a real source measured with a dummy head can be seen in Fig. 2.2 The glitches in the data are caused by the ambiguity of phase differences at higher frequencies, hence ITD is a predominantly low frequency cue. ITD provides predominantly low frequency cues and is generally considered the most important

localisation cue up to about *1.5kHz* for broadband signals (Wightman and Kistler, 1992). Wightman and Kistler also noted that if the low frequency energy is removed from the stimulus, the ILD becomes the most significant localisation cue.

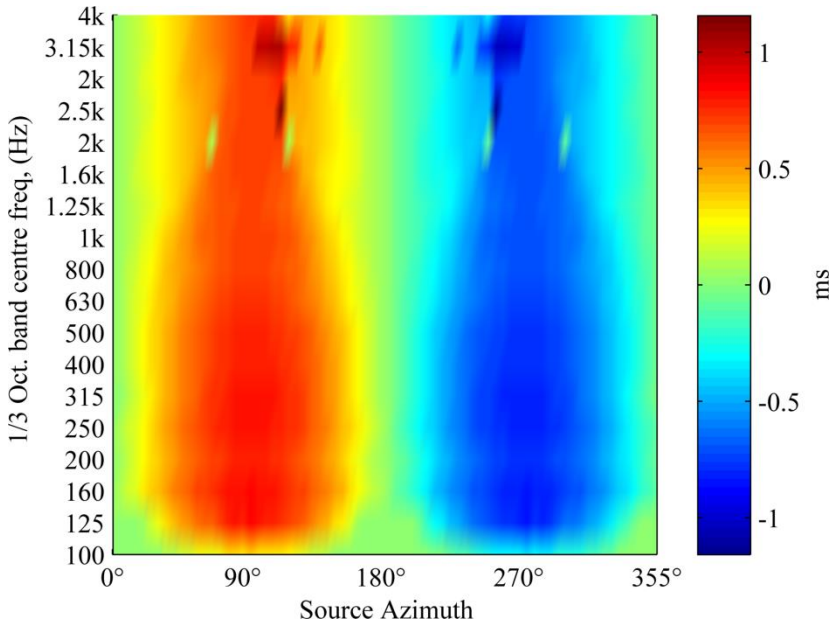
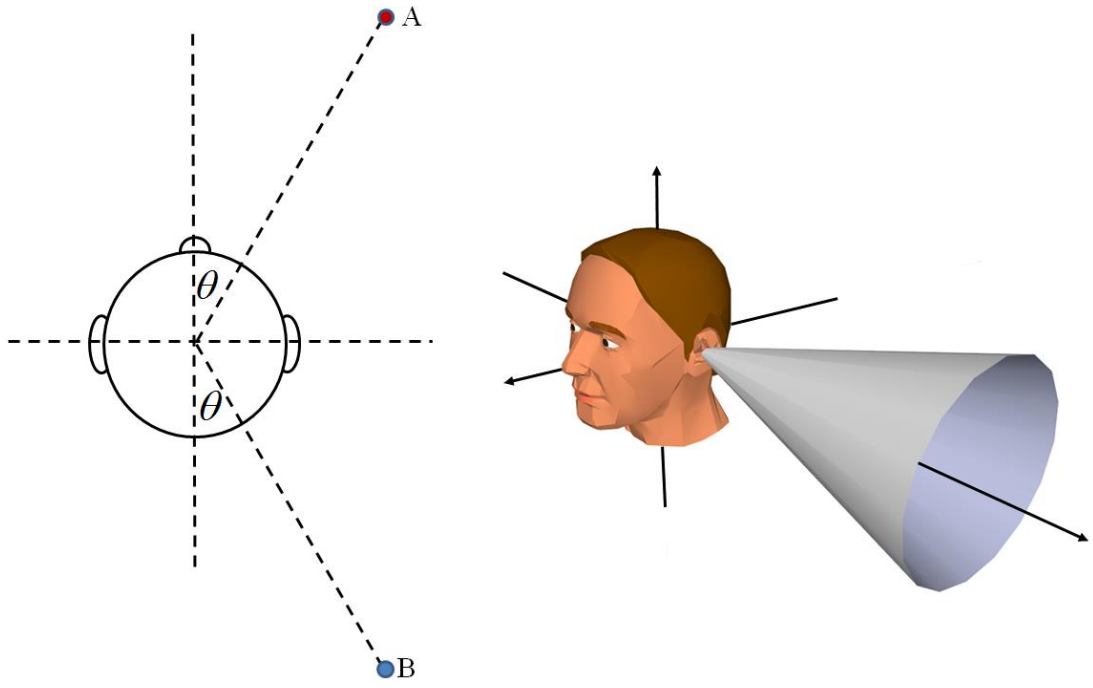


Fig. 2.2 Broadband ITD measurement with a KEMAR dummy head

ITD cues do not provide much information as to front and back localisation as there will be no difference in the relative arrival times of the signals at each ear for a sound at a given angle in front of the listener as at the same angle behind as shown in Fig. 2.3a where the ITD as a result of a source at position *A* will be the same as that from position *B*. This is a specific case of what is commonly referred to as the *cone of confusion* (von Hornbostel and Wertheimer, 1920; cited in Blauert, 1997, p179) as shown in Fig. 2.3b. As the time difference of arrival is only known at only two points (each ear), there is an ambiguity of possible source positions. In the two dimensional case the possible source position would be anywhere along a hyperbolic path between the ears where the difference in distance travelled from the source to each ear is constant. In the three dimensional case, the possible source positions are anywhere on the surface of a hyperbolic cone as illustrated in Fig. 2.3b. This localisation confusion/ambiguity is resolved through additional cues generated from the changes in ITD with movements of the head.



a) b)

Fig. 2.3 a) Example of front to back ambiguity of ITDs, b) Diagrammatic representation of the cone of confusion

The role of ITD in spatial audio perception and localisation has been well documented in the literature and an excellent overview can be found in Blauert (1997).

2.2.1.2 Interaural level differences

Rayleigh's duplex theory also outlined the importance of *interaural level differences* (ILD) for the localisation of sound. ILDs are caused primarily by the shadowing effect of the head. At higher frequencies, the head is of significant size with respect to the incident sound wavelength and as such acts as a barrier causing a difference in the levels received at each ear for a given source that varies with the angle of incidence. This effect is much less for lower frequencies (longer wavelengths) where the sound easily diffracts around the head resulting in very small or no difference in the levels between the signals received at the two ears. This effect can be seen from the plot of broadband ILDs as measured using a KEMAR dummy head. Below about 500Hz the difference in relative levels is very small; consequently this cue is not useful for low frequencies. For closer sources this ILD may even be caused by the

extra distance from the source to the ears and the corresponding distance attenuation thereof.

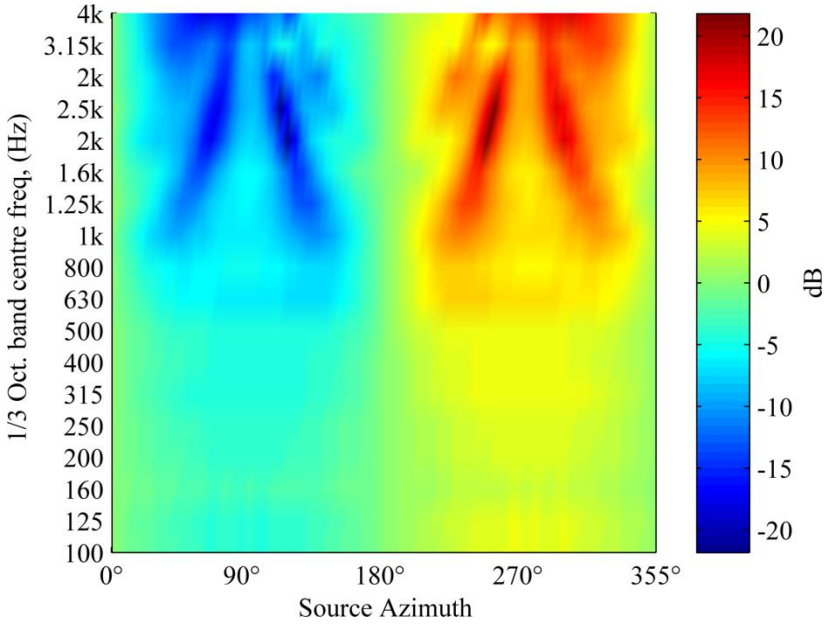


Fig. 2.4 Broadband ILD measurement with a KEMAR dummy head

The magnitude of the ILD is dependent upon the frequency of the incident wave as shown in Fig. 2.4, additionally it has a source position dependency as, when sources are closer to the head, the wave front curvature becomes significant and causes larger ILDs (Brungart et al., 1999), conversely the same authors found that ITD is frequency independent.

2.2.1.3 Spectral cues

As previously mentioned, neither ILD nor ITD provide much information for front to back localisation. One of the mechanisms for determining whether a sound is in front or behind the listener is the spectral differences resulting from the filtering due to reflections from the person's head and shoulders and more importantly the pinna. Everyone's pinnae are different but learned cognitive patterns mean the reflections caused by the outer ear provide information as to source direction. Due to the unusual shape of the pinna each source position will cause different resonances and reflections making it possible to infer the source position. These effects are characterised as an individual's *head-related transfer function* (HRTF). Every

person's HRTF will be different due to anatomical differences, although there are some obvious trends/common feature that are true for most people, giving rise to various generic HRTF models (Wenzel et al., 1993) that can be used for the binaural techniques mentioned in the following section. HRTFs can be problematic and time consuming to measure as there is a need for a large number of source angles needed for a complete description of the HRTF. Work has been done however using interpolation methods to reduce the number of measurements needed (Zotkin and Gumerov, 2004) additionally some researchers have used numerical methods to characterise the head and torso to provide the potential for a continuous set of HRTFs at different source angles (Katz, 2001a, 2001b)

Once again it is most likely the differences in the HRTF signals at each ear that is important for the localisation of sounds. Whilst it is possible for the brain to do some pattern matching with just monaural cues, with stationary sources it is difficult to separate the timbral characteristics of the source from the spectral content received at the ear. There is evidence that the brain learns this sort of directional encoding and that in fact it has an adaptive ability if a listener's HRTF were to change for any reason (Hofman, 1999; Wanrooij and Opstal, 2005).

2.2.2 Vertical localisation

Vertical localisation in the median plane is more difficult than horizontal localisation in the lateral plane due to the fact that the ears are on the same level, meaning there can be very little difference in the timing or level of the signals at each ear for different source elevations. Vertical localisation therefore is calculated by the brain based on the aforementioned spectral differences from sounds of differing elevations by way of the learnt HRTF. Generally humans have a far more limited capacity and localisation resolution in the vertical plane than in the lateral plane. The sound reproduction systems mentioned in this thesis are confined to two dimensions; as such a more in-depth analysis of vertical localisation mechanisms is outside the scope of this thesis.

2.2.3 Depth and distance perception

The perception of depth and distance are not equivalent measures. Depth in this context defines not just distance perception but the overall perspective of an auditory scene and the dimensions of the entire scene, thus depth perception depends upon the distance perception of all the sources in the sound scene. As human localisation is based on binaural cues there is no possibility of any cognitive triangulation to position a source accurately in terms of distance so other localisation cues such as environmental characteristics become important. The primary mechanisms for distance perception are outlined below as given by Wittek (2009):

i) Level (sound pressure)
ii) Direct-to-reverberant energy ratio
iii) Reflection pattern (timing, level and directions of early reflections)
iv) Frequency spectrum (for very near and for far sources)
v) Binaural differences: acoustic parallax as well as intensity/phase differences
vi) Motion parallax (changes of perspective with source/listener movements)
vii) Interaction with source familiarity and non-acoustical cues such as visual cues

Perhaps the most important distance cue is the level of the source which will decrease by approximately *6dB* for every doubling of distance due to spherical spreading and air absorption. This is a very useful cue if the original level of the source is known, i.e. if there is a reference. Where there is no reference, the level of the source provides very little localisation information unless the source moves in which case the changes in level provide the necessary localisation information. Spectral differences also provide useful information as to source distance as the atmosphere acts as a low-pass filter so distant sources will have less high frequency energy. Once again the value of this cue for localisation is dependent on prior knowledge of the source content, unless the source has a dynamic position.

It has been found in several experiments that distance localisation is very poor in anechoic environments or where there is no reference for comparison (Nielsen,

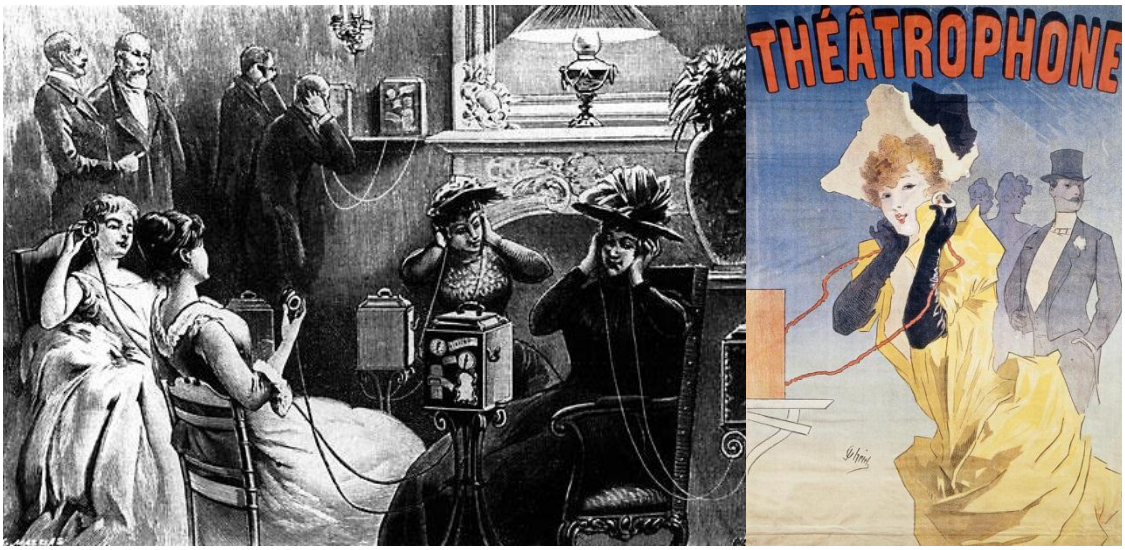
1991) in this case localisation is based primarily on level cues with louder sounds being localised nearer to the listener, which confirms the findings of Gardner (1969) that level is the most important cue where there are no room reflections present. Where there are reflections present it is possible to infer the source localisation from the relative strength and timing of the direct sound to its reflections, of particular importance is the first reflection from the floor in the reproduction environment (Bech, 1998) as this is generally the first reflection received at the listener position. The relative timing of this reflection (and others) with respect to the direct source hence helps inform the auditory system of the source distance.

2.3 An overview of spatial audio techniques

Spatial audio systems can broadly be split into three categories: intensity-based systems, binaural reproduction and sound field reconstruction systems. A brief overview and history of each of these categories here follows.

2.3.1 Intensity-based stereophonic systems

The earliest recorded spatial audio experiments including, recording, transmission and reproduction were carried out at the Electrical Exhibition (‘La Lumiere Electrique’) held at the Paris Grand Opera in 1881 (Moncel, 1881). The system was designed by Clement Ader, and utilised four microphone pairs connected to binaural telephone receivers in four remote locations in the opera house allowing visitors to the exhibition to listen stereoscopically to the live performance.



a) b)

Fig. 2.5 a) The first stereophonic transmission at 'La Lumiere Electrique' by Clement Ader. b) Advert showing the subsequent commercialisation of the Theatrophone for stereophonic transmission and

However it was not until 1931 when the theoretical background for stereophonic systems was developed, following the experiments by Blumlein and his subsequent patenting of the first intensity stereo system (Blumlein, 1931), this contribution was complemented by stereo recording methods using multiple microphones.

Around the same time the Bell labs were also looking into spatial audio reproduction techniques and the work by Steinberg and Snow (1934) proposed an early three-channel stereo system, replaying a recording made with 3 spatially disparate microphones (shown in Fig. 2.7). Remarkably this early work into stereophonic reproduction shows many similarities to the now state-of-the-art wave field synthesis systems as described later this thesis. They suggested the idea of the *Acoustic Curtain* as shown in Fig. 2.6.

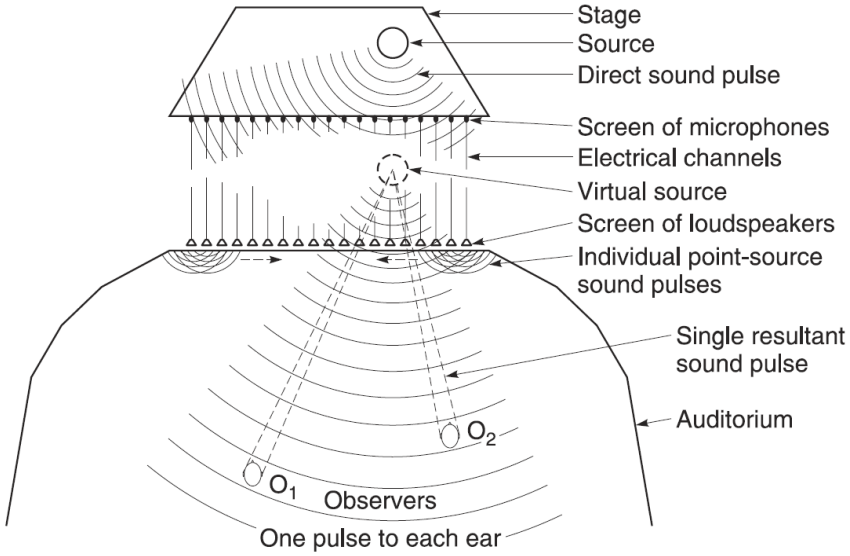


Fig. 2.6 Steinberg and Snow's acoustic curtain concept, recording and recreating the wave fronts from the stage (after Rumsey (2001))

The idea was to have a row of microphones across the stage that exactly matched the spatial configuration of loudspeakers at the reproduction end. Quite correctly, they pointed out that this would recreate the original wave fronts and therefore the localisation cues that would be present if the listener were at the location. However this was put forward as the ideal situation which due to technical constraints could not be practically realised. They therefore came up with a compromise solution which utilised just three microphones and three loudspeakers as shown in Fig. 2.7.

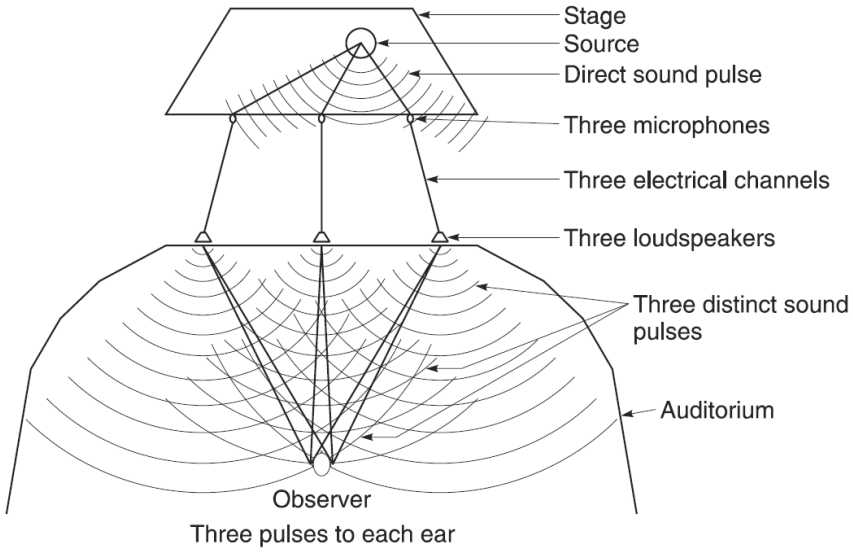


Fig. 2.7 Steinberg and Snow's compromise version of the acoustic curtain concept, recording and recreating the wave fronts from the stage with three channels (after Rumsey (2001))

Whilst the system in Fig. 2.7 was original devised as a compromised version of that shown in Fig. 2.6, Steinberg and Snow did point out that these two situations are in fact different in the localisation processes, as the latter compromised version actually relies more on the precedence effect rather than the information from the wave fronts reaching the ears, however they note that the performance of the compromised system was actually very good, which provides another example of the surprisingly good performance of stereophonic reproduction which has seen it so widely adopted.

Despite the works from Steinberg and Snow and Blumlein, it was not until the 1950's that commercial stereophonic systems became realised; the market was fuelled in the 1960's when two-channel vinyl records became popular. Stereophonic audio grew greatly in popularity and became ubiquitous for many years, to a great extent usurping monophonic reproduction.

Despite the success of two-channel stereophony however, there was still a desire in the cinematic world for a greater spatial impression and rendering to fit with the moving image; consequently quadraphony was introduced as a concept. Quadraphony provided an extra two channels at the rear of the listener, increasing spatial impression and the creative options available to the producer/sound engineer. Quadraphony was however rather unpopular for consumer audio systems primarily due to compatibility issues and consumers not seeing the need for extra loudspeakers in their living rooms thus it did not supersede two-channel stereophony.

With the invention of the CD in 1982 and the subsequent upsurge in digital audio technology, more audio channels have become easily available, driving the consumer market. Multi-channel sound reproduction is subsequently becoming more popular and is being utilised increasingly in an ever growing home cinema market. Since the consumer adoption of stereophony, people have increased the number of loudspeakers used within a given system such that it is now common to see 5.1, 7.1 and even 22.2 (Hamasaki et al., 2005) surround sound systems.

All of these systems use both level and phase panning in order to provide the necessary information for the listener to be able to localise the sound in either a two or three dimensional space, the greater the number of loudspeakers the greater the definition of the localisation. A problem with the aforementioned systems however is the so-called 'sweet spot'. The systems are generally optimised for just one position meaning that the listener will only receive correct spatial imaging if he or she is sat in the correct place in the room. This is because sources are positioned in the reproduced sound scene based on level and phase differences between each loudspeaker signal (generating the required ILDs and ITDs at the listener); if the listener moves even a small amount, these cues will be altered and the perceived location of the source will change thus it is said that the phantom source is unstable. These problems can be reduced if the sweet spot is set to be further away from the speakers; however this method still only allows true surround sound for one or two individuals.

2.3.2 Binaural Systems

Binaural audio can be considered as a special case of 2-channel stereophonic reproduction. Binaural audio is realised with recordings using a dummy head with a similar geometry to the proposed listener (as shown in Fig. 2.8) with the microphones placed in the ears such that the exact signals received at the microphone position can be played through headphones to the user giving them a potentially very realistic spatial impression of the original sound scene. Binaural stereo can also be realised by processing the sources in the scene with the HRTF (*Head Related Transfer Function* - see section 2.2.1.3) corresponding to the intended source location, thus an entire scene can be synthesised for binaural rendering if the source signals, and the HRTF of each source to listener, are combined. This will be very realistic if the HRTF that is used is correct (i.e. matches the source position and the individual), but as mentioned above, generic and interpolated HRTFs are often used which are an approximation and are therefore a compromise of the necessary localisation cues.

Fig. 2.8 Brüel & Kjær HATS dummy head used for binaural recording

It is intended that the use of this method retains all the original auditory cues that the listener would have had if actually present at the recording location; it does however have several shortcomings. Primarily everyone's ears are different and the geometry of the pinna is crucial in the spatial perception of an auditory stimulus. Results are also impaired if there is any additional processing on the signals such as compression and amplification or any noise in the system.

The playback should be done with headphones, or it is possible to inverse filter the signal with the HRTF to obtain a transauralisation (Cooper and Bauck, 1989) played via loudspeakers but this is limited to a listening area of only a few centimetres so is often considered impractical. The fact that the playback is done using headphones can lead to an unrealistic spatial impression with binaural recordings, not only because headphones provide an unusual listening experience but also in real life localisation is greatly helped by small movements of the head to resolve ITDs and ILDs especially for depth perception. This resolving cannot be done inherently with headphones as head movements will have no effect on the received signals, thus optimum realism is achieved if the user does not move their head during playback or more complex head tracking is implemented (Begault et al., 2001). There is also the problem that with the use of headphones many people have the sensation that the sound source is located inside their head (so-called *intracranial localisation*) and externalisation is therefore difficult to achieve. The use of headphones also prevents

natural communication with other listeners which can be a problem for many applications. A good review of binaural audio techniques can be found in Begault (1994).

2.3.3 Sound field synthesis/reproduction techniques

The concept of recording an entire sound field and transmitting it to a remote location for playback is not a new concept and was arguably first conceived of by Steinberg and Snow (1934), however practical limitations meant that the idea would remain merely theoretical for many years. With the vast improvements in computational power and the relatively cheap audio hardware now available, the possibility of not just using artificially created psychoacoustic cues to evoke an impression of a spatial scene but to accurately recreate the sound field such that all of the original localisation cues are present can now be a reality. Sound field synthesis techniques typically use a large number of loudspeakers and aim once again to recreate all of the localisation cues present at the original scene. As opposed to binaural techniques, sound field synthesis includes cues generated by the small movements of the head and don't suffer from the problems associated with reproduction over headphones mentioned above.

The goal of sound field synthesis/reproduction is the physical representation of the sound field in a given space. Whilst the aforementioned systems have their roots in physical principles, they all offer different compromises to reproduce a reduced set of auditory cues to provide a believable sound scene. Sound field reconstruction on the other hand aims at recreating all of the cues needed for the accurate perception of the intended sound scene at a given point in space or volume. Three predominant sound field reproduction methods will be mentioned in this section, namely A*mbisonics (including higher orders), Wave Field Synthesis* and the *Spectral Division Method*.

2.3.3.1 Ambisonics

Ambisonics uses spherical harmonics to describe the sound incident from all directions at one point in space using a spherical Fourier transform. This circular

harmonic expansion of the wave equation is written in circular coordinates such that the sound field can be written as a Fourier-Bessel series. Ambisonics thus offers a hierarchical representation such that the order of the spherical harmonic representation of the sound field defines the spatial resolution of the reproduced sound field.

Much of the early theoretical work into the Ambisonics technique for sound field recording, encoding and reproduction was done by Michael Gerzon (Gerzon, 1973) at Oxford University and has since been complemented by many others e.g. (Daniel et al., 2003; Lee and Sung, 2003; Poletti, 1996). Despite the strong technical roots of Ambisonics and its theoretical benefits of offering a scalable system it has not been widely adopted for commercial systems and is mainly used in research and bespoke audio applications. However as it offers a scalable representation of a sound field it has much potential to be used to spatially encode and transmit a recorded scene for a format agnostic rendering (Batke et al., 2011) and consequently may yet prove to be adopted on a wider scale.

Arguably, Ambisonics can be thought of as sound field synthesis as the technique does provide a physically accurate rendering of a sound field (even though remarkably the equations essentially become panning functions), however, generally this is optimised for a point rather than being a volumetric solution such as the other techniques mentioned in this section. However the Ambisonics technique is scalable and as such, using a higher order of coefficients for the sound field representation allows for a sound field reproduction over a wider area.

2.3.3.2 Higher Order Ambisonics (HOA)

The initial Ambisonics approach of the 1970s utilised only first order spherical harmonics for the sound field representation, more recently however, the research on Ambisonics techniques has been more keenly focused upon *Higher Order Ambisonics (HOA)* using higher order spherical harmonic representations for rendering (Daniel, 2001; Malham, 2003) and recording (Meyer and Elko, 2002; Teutsch, 2007). Generally the term *Higher Order Ambisonics* is reserved for Ambisonics above and including 2nd order.

An increasingly popular extension to HOA is the so called *Near-Field Compensated Higher Order Ambisonics (NFC-HOA)* (Daniel, 2003). In principle, the sources in Ambisonics are assumed to be plane wave sources, but when loudspeakers are used for the reproduction the sources are much better represented as point sources, thus NFC-HOA is a method of applying a suitable correction to the loudspeaker driving signals such that the consequences of this source mismatch are minimised. Furthermore, the use of NFC-HOA means that the playback system is able to reproduce sources closer to the listener.

2.3.3.3 Wave field synthesis (WFS)

Whilst higher order Ambisonics systems can recreate a sound field over a large area, in most cases it is not considered a volumetric system as the region of accurate reproduction is limited and generally defined for one location in space. Unlike Ambisonics however, the wave field synthesis (WFS) technique, is a volumetric solution aiming at recreating/synthesising a given sound field over the entire reproduction space. This requires a very large number of loudspeakers with some systems currently utilising several hundred loudspeakers (De Vries, 2009).

As will be discussed in this thesis, the WFS theory is based on the Huygens-Fresnel principle and the Kirchhoff-Helmholtz equation and can be considered to be a practical realisation of the *acoustic curtain* principle originally proposed by Steinberg and Snow (1934) (see Fig. 2.6). Practically a WFS system will consist of many loudspeakers surrounding the reproduction space whose output are individually defined and controlled. The superposition of all these loudspeaker signals results in reproduction of arbitrary wave fronts. The first formal implementation of WFS was proposed by (Berkhout et al., 1993; Berkhout, 1988) and since then there has been an active research community looking at many system improvements and analyses, some of which will be covered in this thesis.

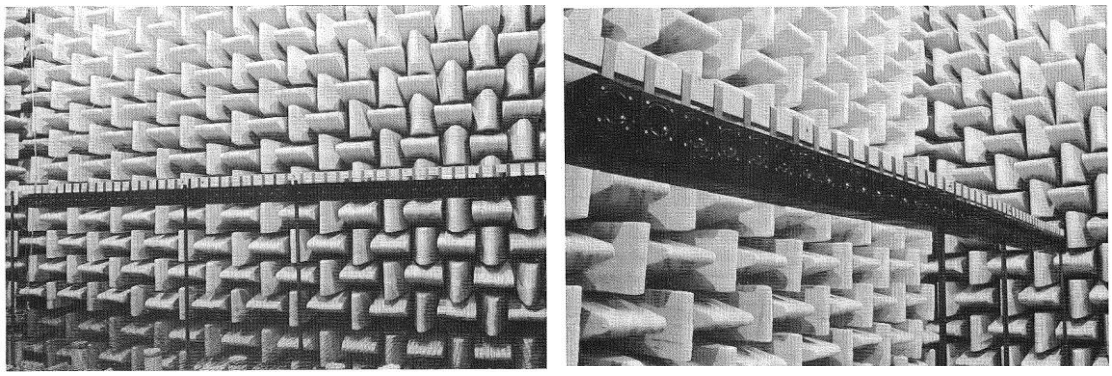


Fig. 2.9. The early prototype wave field synthesis system installed in an anechoic chamber at TU Delft by Peter Vogel (image after (Vogel, 1993))

Perhaps the main advantage of the physical sound field reconstruction using WFS is its ability to create arbitrary sound fields that are accurately reproduced throughout the entire listening space, these sound fields can even include so-called *focused sources* that appear to be radiating from in front of the loudspeaker array (as is the focus of this thesis). This provides very accurate depth of scene cues and increases the sense of presence in the reproduced acoustic environment. WFS provides a set of driving functions for various source types such that the composition of an acoustic scene or *soundscape* can be achieved very easily by combining many different audio objects which have a defined position in space and audio content, in this case WFS will reproduce all of the original cues needed for the localisation and perception of all the sources in the scene.

There has been some work done, looking at the similarities and differences between WFS and HOA (Daniel et al., 2003; Spors and Ahrens, 2008). One of the big differences is that for WFS there is an explicit solution for the loudspeaker driving functions for arbitrary geometries, whereas for HOA, there are only exact solutions for limited loudspeaker configurations. Another interesting difference is in the spatial aliasing artefacts (see section 3.5.2 for more details), for HOA there is a circular region in the centre of the reproduction space which doesn't exhibit these artefacts whereas for WFS spatially aliased contributions are present throughout the reproduction space (above the aliasing frequency). There is however still work to be done to determine the perceptual consequence of the differences between the two techniques.

2.3.3.4 *Spectral division method (SDM)*

A recent innovation in sound field synthesis has been proposed by Ahrens and Spors (2010) and is referred to as the Spectral division method (SDM). This method describes the desired sound field in the spatio-temporal frequency domain, or *wavenumber domain* (see section 3.5.2.3). From this mathematical description, the loudspeaker driving functions are calculated by taking the inverse spatial Fourier transform. In many respects the SDM shares many similarities with WFS, exhibiting some similar restrictions such as spatial aliasing and loudspeaker array truncation (refer to section 3.5.2 and section 3.5.3) but has been shown in some circumstances to produce more favourable results (Spors and Ahrens, 2010b). Work in this field is still in its infancy but the driving functions are slightly easier to define than the corresponding WFS functions and some of the reproduction errors can also be reduced such as it does not rely on a far field approximation as WFS does, so the method shows much promise.

2.3.4 Other methods

There are many other methods used for spatial audio reproduction including vector base amplitude panning (Pulkki, 1997), numerical approaches (Fazi et al., 2008; Kirkeby and Nelson, 1993), approaches based on optimisation (Hannemann and Donohue, 2008) and the solutions of inverse problems (Ise, 1991). These references and the references therein will provide the interested reader with more information on these techniques but they will not be covered in any more detail here as they fall outside the scope of this thesis.

2.4 Summary

This chapter has briefly outlined some of the primary localisation mechanisms relevant to spatial audio and hence the study of focused sources in wave field synthesis as is addressed in this thesis. Also described in this chapter is a historical overview of some of the major spatial audio rendering techniques in order to contextualise the work described in this thesis. The three primary categories of spatial audio rendering systems have been reviewed and many of the principles and

concepts mentioned in this chapter serve to provide a framework from which to interpret the findings in the rest of this thesis.

3 Wave Field Synthesis

3.1 Introduction

This chapter begins with an overview of the history of the wave field synthesis technique, briefly covering its historical and technical roots. This is followed by a detailed review of the theory of wave field synthesis (WFS) including a full derivation of the Kirchhoff-Helmholtz equation which forms the basis of the driving functions which are also derived in detail. The driving functions include the major source types, point source, plane wave, and dipoles and also briefly covers sources with more complex radiation characteristics. The subsequent section describes the design, installation, calibration and use of 4 WFS systems installed in the University of Salford as part of this project. This chapter then concludes with a description of some of the key subjects of research in the WFS literature and some open issues.

3.2 History

Arguably the first conception of WFS as it is now known was proposed by Steinberg and Snow (1934) as shown in the previous chapter (see Fig. 2.6). Due to the technological and practical constraints of the time, the system described here was never actually implemented and indeed the ideas mentioned laid dormant for many years until Berkhout (1988) formulated a more mathematically rigorous theory. Berkhout's work was originally applied to seismic vibrations (Berkhout, 1987) but was later applied to audio acoustics in what was at that time named acoustic holography (or 'holohphony'). Later the term 'wave field synthesis' was presented by the same author along with other academics at the Technical University of Delft (Berkhout et al., 1992, 1993), proposing practical uses of the system the principles of which had previously been outlined. Following these seminal papers many PhD theses were written at the same institution (Vogel, 1993; Start, 1997; Verheijen, 1998; de Bruijn, 2004; Hulsebos, 2004), adding more rigorous theory and practical considerations of the WFS technique.

Since this early conception of WFS there has been much interest in the area as many have realised its potential to (re)create sound fields that are spatially and physically accurate and that consequently provide listeners with the same localisation cues as if they were actually present in the recorded or synthesised audio scene.

3.3 Theory

Wave field synthesis is in essence an application of the well-known Huygens-Fresnel principle, the theory of which states that each point on a given wave front can be considered as the centre of a spherical wavelet radiating from that point. The superposition of an infinite number of these wavelets carries all the information of the original wave front. Thus, conceptually, if the pressure is recorded over an infinite number of points along the wave front, it can be completely reconstructed using secondary sources. A diagram of this principle is shown in Fig. 3.1.

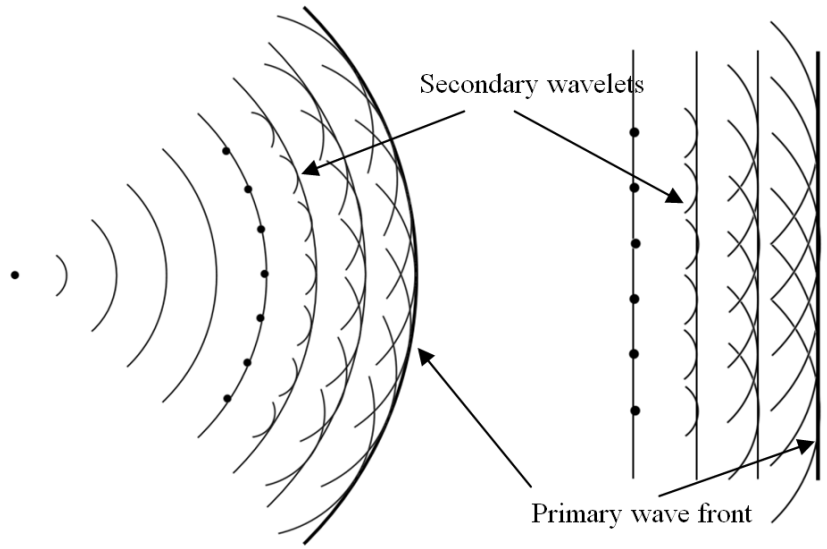


Fig. 3.1 Diagrammatic representations of the Huygens Principle

The Huygens principle originally dates back to the late 1600's and was later modified by Augustin-Jean Fresnel to include the full effects of the superposition of the secondary wavelets. A more mathematically rigorous formulation which can be considered as a generalised form of this principle was proposed by Kirchhoff in the early 1800's, for which we now have the Kirchhoff-Helmholtz equation *(3.8)* which forms the basis of WFS. The Kirchhoff-Helmholtz equation describes how the pressure field in a source-free volume can be completely described if the pressure and its gradient normal to the surface at a continuum of points on that surface are known.

3.3.1 Derivation of the Kirchhoff-Helmholtz equation

The starting point for the derivation for the Kirchhoff-Helmholtz equation is Green's second theorem which states that if the field on the surface of a closed source free volume is known then the field within that volume can be completely determined. This theorem is given mathematically by equation *(3.1)*:

$$\underbrace{\int_V f\nabla^2 g - g\nabla^2 f \, dV}_{\text{Field in the volume, } V \text{ enclosed by a surface, } S} = \underbrace{\oint_S \left(f\nabla g - g\nabla f \right).\mathbf{n}\, dS}_{\text{Field on the surface, } S \text{ of a volume, } V} \qquad (3.1)$$

Where *f* and *g* are functions of all three spatial dimensions *x*, *y* and *z* and **n** is the inward pointing unit vector normal to the surface, *S*. For acoustic scenarios these functions *f* and *g* will be acoustic pressures. The geometry used for *(3.1)* and the derivation of the Kirchhoff-Helmholtz equation is shown in Fig. 3.2.

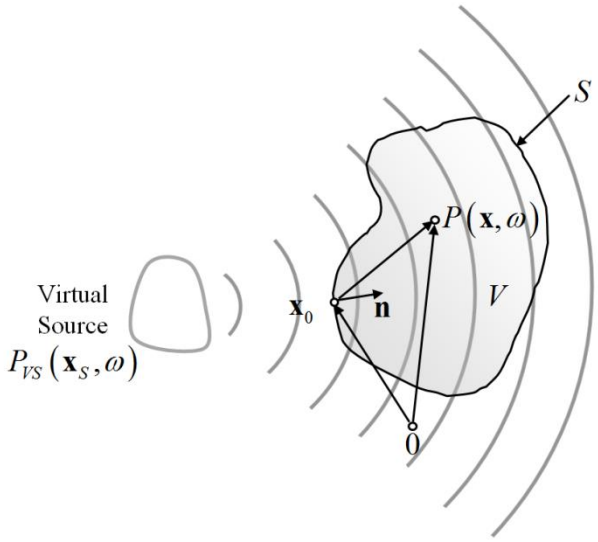


Fig. 3.2 Geometry used for Green's 2nd theorem and derivation of Kirchhoff-Helmholtz equation

To determine the pressure at a receiver point, **x** inside the volume according to Green's theorem we need to know the pressure on the surface, $P(\mathbf{x}_0,\omega)$ as a result of a distribution of acoustic sources outside of the volume, where *k* is the wave number in air given by $k=\omega/c$. The function $P(\mathbf{x}_0,\omega)$ will obey the homogeneous wave equation, such that:

$$\nabla^2 P(\mathbf{x}_0,\omega)+k^2 P(\mathbf{x}_0,\omega)=0 \qquad (3.2)$$

This gives the pressure on the surface as a result of a distribution of sources outside of the volume, $P_{VS}(\mathbf{x}_S,\omega)$. To determine the pressure at point, **x** within the volume we need to know the function, *g* of a source inside the volume that obeys the inhomogeneous wave equation:

$$\nabla^2 G(\mathbf{x}_0\,|\,\mathbf{x},\omega)+k^2 G(\mathbf{x}_0\,|\,\mathbf{x},\omega)=4\pi\delta(\mathbf{x}_0-\mathbf{x}) \qquad (3.3)$$

This function, $G(..)$ is called the *Green's function* and can be any function that obeys equation *(3.3)*. A function that obeys this is that of a monopole so we place a monopole at the receiver position and determine that the Green's function is:

$$G(\mathbf{x}_0 \mid \mathbf{x}, \omega) = \frac{e^{-jk|\mathbf{x}_0 - \mathbf{x}|}}{|\mathbf{x}_0 - \mathbf{x}|} \tag{3.4}$$

This is the same as for a monopole at a point, $\mathbf{x}_0$ on the surface propagating to the receiver, **x** due to the principle of acoustic reciprocity (that is to say, the transmission path between source and receiver is the same as from receiver to source). Thus

$$G(\mathbf{x} \mid \mathbf{x}_0, \omega) = G(\mathbf{x}_0 \mid \mathbf{x}, \omega) \tag{3.5}$$

So substituting $P(\mathbf{x}_0, \omega)$ for *f* and $G(\mathbf{x} \mid \mathbf{x}_0, \omega)$ for *g* in equation *(3.1)* gives:

$$\begin{aligned} \int_V P(\mathbf{x}_0, \omega)\nabla^2 G(\mathbf{x} \mid \mathbf{x}_0, \omega) - G(\mathbf{x} \mid \mathbf{x}_0, \omega)\nabla^2 P(\mathbf{x}_0, \omega)\, dV = \ldots \\ \oint_S \left(P(\mathbf{x}_0, \omega)\nabla G(\mathbf{x} \mid \mathbf{x}_0, \omega) - G(\mathbf{x} \mid \mathbf{x}_0, \omega)\nabla P(\mathbf{x}_0, \omega) \right).\mathbf{n}\, dS \end{aligned} \tag{3.6}$$

Now substituting $\nabla^2 P(\mathbf{x}_0, \omega)$ from *(3.2)* and $\nabla^2 G(\mathbf{x} \mid \mathbf{x}_0, \omega)$ from *(3.3)* and *(3.5)* into *(3.6)* gives:

$$\int_V P(\mathbf{x}_0, \omega) 4\pi\delta(\mathbf{x} - \mathbf{x}_o) dV = \oint_S \left(P(\mathbf{x}_0, \omega)\nabla G(\mathbf{x} \mid \mathbf{x}_0, \omega) - G(\mathbf{x} \mid \mathbf{x}_0, \omega)\nabla P(\mathbf{x}_0, \omega) \right).\mathbf{n}\, dS \tag{3.7}$$

Utilising the sifting property of delta functions then results in the Kirchhoff-Helmholtz equation:

$$P(\mathbf{x}, \omega) = -\frac{1}{4\pi} \oint_S \left(G(\mathbf{x} \mid \mathbf{x}_0, \omega) \frac{\partial}{\partial \mathbf{n}} P(\mathbf{x}_0, \omega) - P(\mathbf{x}_0, \omega) \frac{\partial}{\partial \mathbf{n}} G(\mathbf{x} \mid \mathbf{x}_0, \omega) \right) dS \tag{3.8}$$

With

$$\frac{\partial}{\partial \mathbf{n}} \equiv \nabla \cdot \mathbf{n} \tag{3.9}$$

The term $\frac{\partial}{\partial \mathbf{n}} P(\mathbf{x}_0, \omega)$ is the directional gradient of the pressure normal to the surface and can be given from *(3.9)* and the equation of motion as:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{n}} P(\mathbf{x}_0, \omega) &= \nabla P(\mathbf{x}_0, \omega).\mathbf{n}(\mathbf{x}_0) \\ &= -j\omega\rho_0 V_n(\mathbf{x}_0, \omega) \end{aligned} \tag{3.10}$$

And with

$$\frac{\partial}{\partial \mathbf{n}} G(\mathbf{x} \mid \mathbf{x}_0, \omega) = -\frac{\left(1 + jk|\mathbf{x}-\mathbf{x}_0|\right)}{|\mathbf{x}-\mathbf{x}_0|^2} e^{-jk|\mathbf{x}-\mathbf{x}_0|} \frac{\mathbf{x}-\mathbf{x}_0}{|\mathbf{x}-\mathbf{x}_0|} \tag{3.11}$$

(3.8) thus becomes:

$$P(\mathbf{x}, \omega) = \frac{1}{4\pi} \oint_S \left(\frac{e^{-jk|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|} j\omega\rho_0 V_n(\mathbf{x}_0, \omega) + P(\mathbf{x}_0, \omega) \frac{\left(1 + jk|\mathbf{x}-\mathbf{x}_0|\right)}{|\mathbf{x}-\mathbf{x}_0|^2} e^{-jk|\mathbf{x}-\mathbf{x}_0|} \frac{\mathbf{x}-\mathbf{x}_0}{|\mathbf{x}-\mathbf{x}_0|} \right) dS \tag{3.12}$$

The physical interpretation of this equation is that the pressure at any point can be determined in a source free volume if the pressure, $P(\mathbf{x}_0, \omega)$ and normal component of the particle velocity, $V_n(\mathbf{x}_0, \omega)$ on the surface are known. The first term of the integrand represents a continuous monopole distribution and the second term defines a continuous dipole distribution on the surface of the volume as shown diagrammatically in Fig. 3.3.

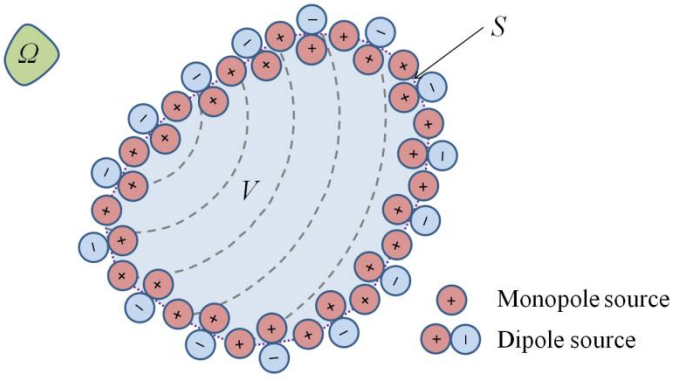


Fig. 3.3. Monopole and dipole distribution of Kirchhoff-Helmholtz equation

3.3.2 Derivation of the Rayleigh I equation

Thus *(3.8)* states that the pressure field $P(\mathbf{x},\omega)$ in the volume, V enclosed by the surface, S (where $\mathbf{x}\in V$) can be entirely determined by the pressure and its gradient on the boundary. It is important to note that outside of the volume, the pressure will be equal to zero. This can be intuitively seen as the result of the cancellation from the negative lobes of the dipole sources and the positive pressure from the monopoles. Thus the strict application of this equation will control only the pressure field within the volume. For practical reasons however it is often preferable to eliminate either the monopoles or the dipoles from the equation such that only one type of source is needed to reproduce the field in the volume. Commonly it is the dipoles that are eliminated as these are harder to realise using real loudspeakers, conversely monopoles can be approximated effectively using a small loudspeaker driver in a sealed enclosure. The dipoles can be removed from equation *(3.8)* by carefully choosing the Green's function such that the second term in the integrand (corresponding to the dipoles) goes to zero, so that:

$$\frac{\partial}{\partial \mathbf{n}} G(\mathbf{x}\,|\,\mathbf{x}_0,\omega) = 0\,,\ (\text{where } \mathbf{x}_0 \in S) \qquad (3.13)$$

The condition described in *(3.13)* is a *Neumann boundary condition* i.e. the boundary is assumed to be an acoustically rigid surface and therefore there is a mirror image source created the other side of the boundary from $\mathbf{x}$ at $\mathbf{x}_R$ (assuming a linear geometry) as shown in Fig. 3.4.

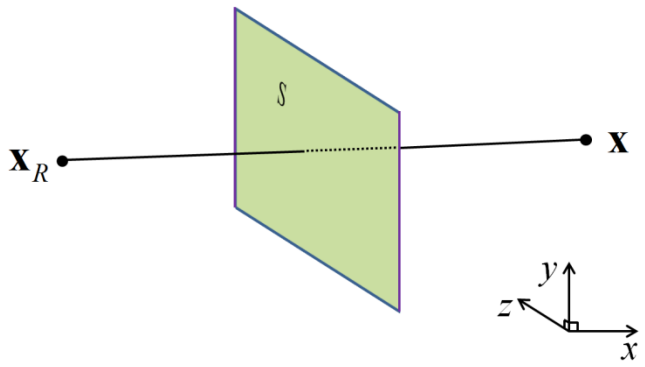


Fig. 3.4 The restricted geometry imposed by the use of the Neumann Green's function

So the Green's function, $G(..)$ in *(3.13)* can be described by a Neumann Green's function, G_N given by (Williams, 1999) as

$$G_N(\mathbf{x}\,|\,\mathbf{x}_0,\omega) = G(\mathbf{x}\,|\,\mathbf{x}_0,\omega) + G(\mathbf{x}_R\,|\,\mathbf{x}_0,\omega), \tag{3.14}$$

which is valid only for the linear planar boundary shown in Fig. 3.4. As this source is the mirror image of the secondary source, equation *(3.14)* becomes:

$$G_N(\mathbf{x}\,|\,\mathbf{x}_0,\omega) = 2G(\mathbf{x}\,|\,\mathbf{x}_0,\omega). \tag{3.15}$$

Substituting this into *(3.8)* successfully eliminates the dipole term and leaves just the monopoles driven at twice the pressure at the surface. This equation is often referred to as the *Rayleigh I integral*.

$$P(\mathbf{x},\omega) = -\frac{1}{4\pi}\oint_S \left(2G(\mathbf{x}\,|\,\mathbf{x}_0,\omega)\frac{\partial}{\partial\mathbf{n}}S_P(\mathbf{x}_0,\omega)\right)dS \tag{3.16}$$

It is also possible to remove the dipoles from *(3.12)* resulting in the so-called *Rayleigh II equation* (see e.g. Verheijen (1998)).

The term $\partial/\partial\mathbf{n}\,S_P(\mathbf{x}_0,\omega)$ represents the normal differential of the pressure field on the surface of the volume in question; it is determined using *(3.9)* and is referred to as the *driving function*, $D(\mathbf{x}_0,\omega)$ when scaled by a factor of $-1/2\pi$ for the

secondary sources in 3-dimensional WFS. The pressure field as a result of the source distribution on the surface, *S* can then be given as

$$P(\mathbf{x},\omega) = \oint_S \left(G(\mathbf{x} \mid \mathbf{x}_0, \omega) D(\mathbf{x}_0, \omega) \right) dS \qquad (3.17)$$

with

$$D(\mathbf{x}_0, \omega) = -\frac{1}{2\pi} \frac{\partial}{\partial \mathbf{n}} S_P(\mathbf{x}_0, \omega) \qquad (3.18)$$

where $S_P(\mathbf{x}_0, \omega)$ is the pressure on the boundary at $\mathbf{x}_0$ as a result of a distribution of external primary sources.

This equation is limited in that it can only be used for a specific geometry (see Fig. 3.4), namely that of a linear planar array of secondary sources radiating into hemispherical space (Verheijen, 1998) thus it is not the full practical solution for sound reproduction as planar arrays are difficult to realise, thus some further steps should be taken in the derivation of the driving function such that the reproduction can be done with linear arrays of discrete loudspeakers.

Removing the dipoles from *(3.12)* has the consequence that the field outside of the volume will no longer be zero. This is not generally a problem for sound field synthesis systems as loudspeakers are often mounted flush or at least close to the walls of the listening room and it is more important to control the sound field *inside* the room than outside it. Problems can occur however if the loudspeaker array is convex such that the radiation from the rear of one driver may affect the radiation from the front of another. For such loudspeaker geometries it is possible to select only the loudspeakers that don't cause the undesirable artefacts for the reproduction using, for example, the time-averaged acoustic intensity vector as proposed by (Spors, 2003) (see section 3.4).

3.3.3 Derivation of WFS driving functions

In order to have a realisable WFS system, a set of functions need to be derived that define the radiation of the secondary sources (loudspeakers), such *driving functions* can be derived for several major primary/virtual source types and can also be extended for sources with more complex radiation characteristics (Corteel, 2007; Baalman, 2008; Ahrens and Spors, 2007). This section presents the derivation for some of the more common source types including, plane waves, monopoles and dipoles.

The first step towards a WFS driving function is to determine the pressure on the boundary, normal to that boundary, as a result of some primary/outside source.

For whichever source type that is chosen, we need to solve *(3.19)*:

$$\frac{\partial}{\partial \mathbf{n}} S_P\left(\mathbf{x}_0, \omega\right) = \nabla S_P\left(\mathbf{x}_0, \omega\right).\mathbf{n}\left(\mathbf{x}_0\right) \qquad (3.19)$$

Where $S_P\left(\mathbf{x}_0, \omega\right)$ is the source function of the primary source at $\mathbf{x}_0$. So from *(3.18)* the 3-dimensional WFS driving function is given as:

$$D_{3D}\left(\mathbf{x}_0, \omega\right) = -\frac{1}{2\pi} \nabla S_P\left(\mathbf{x}_0, \omega\right).\mathbf{n}\left(\mathbf{x}_0\right) \qquad (3.20)$$

Using this driving function, the pressure field within *V* can be given from the Rayleigh I integral as:

$$P\left(\mathbf{x}_0, \omega\right) = \oint_S D_{3D}\left(\mathbf{x}_0, \omega\right) G_{3D}\left(\mathbf{x} \middle| \mathbf{x}_0, \omega\right) dS \qquad (3.21)$$

3.3.3.1 3D to 2D conversion

The Kirchhoff-Helmholtz equation defines the wave field in a 3-dimensional volume, however in practice it is more useful and practical to assume that the wave field in the *z*-dimension does not affect the overall impression such that the wave

field can be produced in two dimensions only. This however requires a 3D to 2D transformation which causes spectral errors in the driving function such that the driving functions for 2D WFS have a frequency dependant amplitude correction due to the changing of the secondary source Green's function.

In the traditional WFS literature this was done by the stationary phase approximation but here the simplification has been done by considering the 2D Green's function (of a line source) as per (Spors et al., 2008). Doing this means that the amplitude has to be optimised for a specific point or reference line, $\mathbf{x}_{Ref}$. As the signal is only correct at this reference point it has the effect of introducing amplitude errors in the reproduced wave field at all other locations. It essentially means that a plane wave source will actually be a cylindrical source which introduces an extra -3dB per doubling of distance; point sources will also have inherent amplitude errors. These small amplitude errors experienced elsewhere in the listening space are however not considered to be subjectively significant (Wittek et al., 2004).

When calculating the pressure from the secondary sources (3D planar array) a free-field 3D Green's function of a point source is considered i.e.

$$G_{3D}\left(\mathbf{x}\middle|\mathbf{x}_0,\omega\right)=\frac{1}{4\pi}\frac{e^{jk|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|} \tag{3.22}$$

However for 2D WFS a different Green's function is required, i.e. that of a 2D point source (a line source). This will consequently change the Rayleigh I equation. The concept of 2D WFS is really to use a series of line sources to recreate the wave field in 2-dimensional space. Each line source is theoretically infinitely long and intercepts the boundary at $\mathbf{x}_0$ as shown in Fig. 3.5.

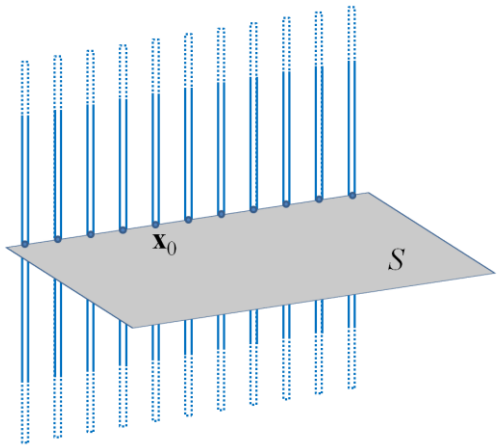


Fig. 3.5. Diagrammatic representation of line sources used for 2D WFS reproduction

3.3.3.2 The 2.5D operator

2-dimensional WFS systems are rarely, if ever, used in practice as line sources are difficult to realise. The derivation of 2-dimensional systems does however provide a useful intermediate step to the derivation of the more practical 2.5D WFS which uses point sources as the secondary source, as such 2D WFS is mentioned only briefly here. The Green's function for monopole radiation by a line source is given as:

$$G_{2D}\left(\mathbf{x}\middle|\mathbf{x}_0,\omega\right)=\frac{j}{4}H_0^{(2)}\left(k\left|\mathbf{x}-\mathbf{x}_0\right|\right) \qquad (3.23)$$

Where $H_0^{(2)}(...)$ is a Hankel function of the zero-th order and of the second kind. It is possible from here to derive a new driving function for the line sources that takes into account the differences in radiation characteristics of line sources and planar sources, this Green's function is substituted into *(3.21)* (where $D(\mathbf{x}_0,\omega)$ is the same driving function as for the 3D case, because this is independent of the reproduction system/sources). This will however result in amplitude errors because we are trying to radiate a 3D monopole (for example) using a 2D system. In this case a cylindrical wave should be used as the virtual source model rather than a spherical source. In any case line sources are not practical for WFS reproduction, plus they should really be infinitely long for correct reproduction which is clearly not possible. Consequently it is desirable to convert these line sources into monopole sources which can be realised using small loudspeakers in cabinets. It is not too difficult to

convert these line sources into point sources if one uses the asymptotic version of *(3.23)*. The large argument approximation of the asymptotic expansion of a Hankel function of the second kind and zeroth order is given as

$$H_0^{(2)}(Z) \simeq \sqrt{\frac{2}{\pi Z}}\, e^{-j(Z-\pi/4)} \tag{3.24}$$

Thus *(3.23)* can be approximated as:

$$\begin{aligned} G_{2D}(\mathbf{x}|\mathbf{x}_0,\omega) &\simeq \frac{j}{4}\sqrt{\frac{2}{\pi k|\mathbf{x}-\mathbf{x}_0|}}\, e^{-j\left(k|\mathbf{x}-\mathbf{x}_0|-\pi/4\right)} \\ &\simeq \frac{j}{4}\sqrt{\frac{2}{\pi k|\mathbf{x}-\mathbf{x}_0|}}\, e^{-jk|\mathbf{x}-\mathbf{x}_0|}\sqrt{j} \\ &\simeq -\sqrt{\frac{2\pi|\mathbf{x}-\mathbf{x}_0|}{jk}}\,\frac{1}{4\pi}\frac{e^{-jk|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|} \end{aligned} \tag{3.25}$$

The second term in this equation is the Green's function of a 3D monopole source as per *(3.22)*, thus the 2D Green's function is the same as the 3D Green's function but with a frequency dependant amplitude correction. This asymptotic approximation however only stands for large arguments i.e. in this case, when $k|\mathbf{x}-\mathbf{x}_0| \gg 1$. The error in *dB* between *(3.23)* and *(3.25)* is plotted in Fig. 3.6.

As can be seen from Fig. 3.6, only at low frequencies and small secondary source to receiver distances is the approximation inaccurate. So with this limitation in mind *(3.21)* can be rewritten as:

$$P(\mathbf{x}_0,\omega) = -\oint_S \sqrt{\frac{2\pi|\mathbf{x}-\mathbf{x}_0|}{jk}}\, D(\mathbf{x}_0,\omega)\, G_{3D}(\mathbf{x}|\mathbf{x}_0,\omega)\, dS_0 \tag{3.26}$$

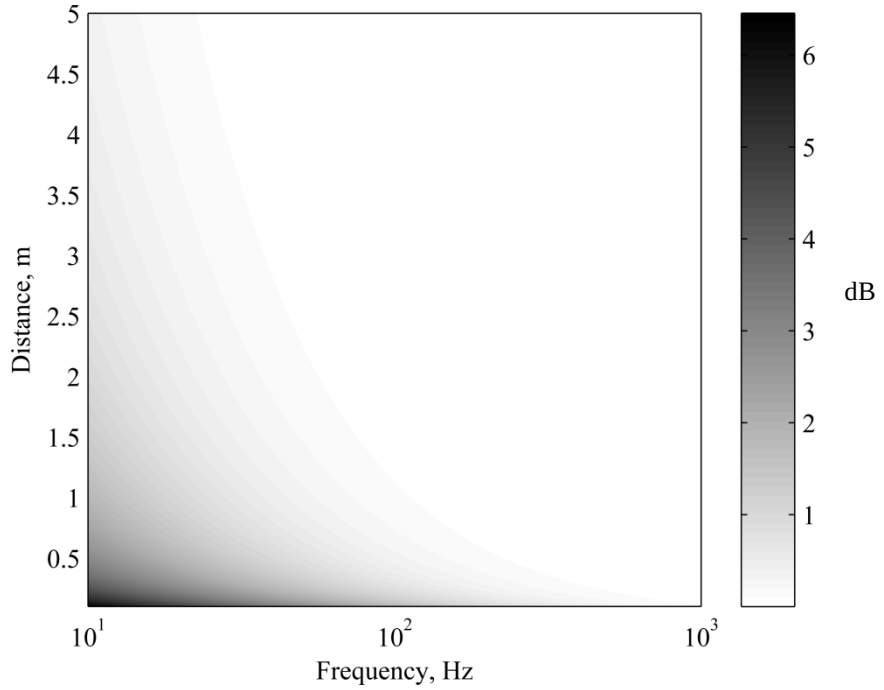


Fig. 3.6. Error in dB between 2D Green's function *(3.23)* and approximated version from asymptotic approximation *(3.24)*

It can be seen from *(3.26)* that the driving function for a 2D WFS system using monopole secondary sources can be given as:

$$D_{2.5D}\left(\mathbf{x}_0,\omega\right)=-\sqrt{\frac{2\pi\left|\mathbf{x}-\mathbf{x}_0\right|}{jk}}D_{2D}\left(\mathbf{x}_0,\omega\right) \qquad (3.27)$$

This equation is not ideal as it depends upon the listener position, $\mathbf{x}$ which is undesirable for WFS, therefore a position, $\mathbf{x}_{\text{Ref}}$ is chosen at which the amplitude will be correct; at the other positions there will be a position dependant error in the level with respect to a real point source although the phase will be correct over the entire reproduction plane (Verheijen, 1998). The amplitude errors with a 2.5D WFS reproduced plane wave compared to a real plane wave are shown in Fig. 3.7.

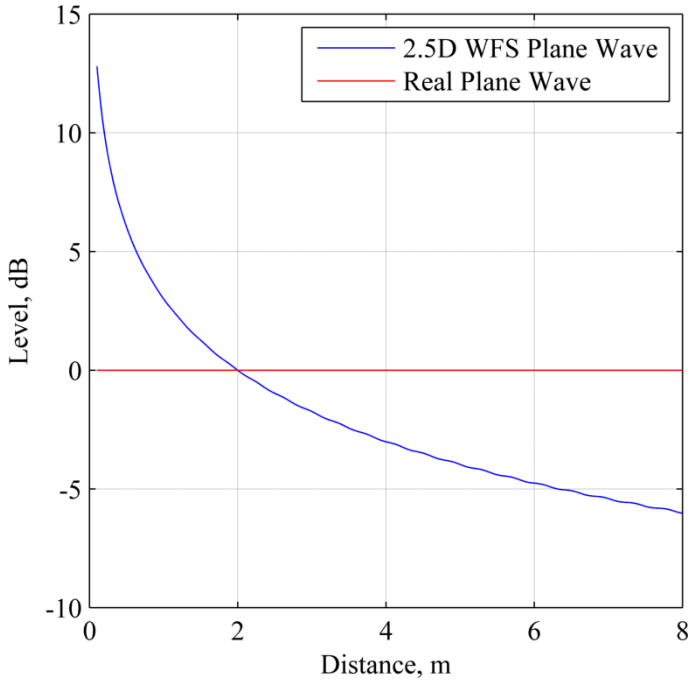


Fig. 3.7 Decay of a real and 2.5D WFS plane wave.

$\mathbf{x}_{\text{Ref}}$ is usually chosen to be the most likely listener location or in the middle of the reproduction space. Thus the 2.5D driving function becomes:

$$D_{2.5D}\left(\mathbf{x}_0,\omega\right)=-\sqrt{\frac{2\pi\left|\mathbf{x}_{\text{Ref}}-\mathbf{x}_0\right|}{jk}}D_{2D}\left(\mathbf{x}_0,\omega\right). \tag{3.28}$$

Now that the driving function has been derived it is necessary to determine the primary source characteristic to fully determine the driving function needed to reproduce a given source.

3.3.3.3 Plane wave sources

Perhaps the most basic source model is that of a plane wave as all sound fields can indeed be decomposed into plane wave components, thus they are the first source model dealt with here. A monochromatic plane wave source can be given by

$$S_{plane}\left(\mathbf{x},\omega\right)=A\left(\omega\right)e^{-jk\mathbf{n}_{PW}\mathbf{x}} \tag{3.29}$$

where $\mathbf{n}_{PW} = [\cos\theta, \sin\theta]$ is the unit vector of the plane wave direction with θ as the angle with respect to the *x*-axis and $A(\omega)$ is the spectral amplitude of the source. So taking the grad of this function and multiplying it by the normal unit vector on the surface at $\mathbf{x}_0$ gives the 3D WFS driving function for a plane wave, $D_{plane,3D}(\mathbf{x}_0, \omega)$ from *(3.20)*:

$$D_{plane,3D}(\mathbf{x}_0, \omega) = -\frac{1}{2\pi} \nabla S_P(\mathbf{x}_0, \omega).\mathbf{n}(\mathbf{x}_0) = \frac{1}{2\pi} \mathbf{n}_{plane}^T \mathbf{n}(\mathbf{x}_0)\, jkA(\omega) e^{-jk\mathbf{n}_{PW}^T \mathbf{x}_0} \qquad (3.30)$$

To reproduce a plane wave using monopole sources rather than planar arrays or line sources we need to add the spectral correction derived in *(3.28)*. As the driving function is for a plane wave there is no need to derive a 2-dimensional case as the spectral radiation characteristics are the same for the 2D and 3D in the reproduction plane. Thus the 2.5D driving function for a plane wave is given by substituting *(3.28)* into *(3.30)*.

$$\begin{aligned} D_{plane,2.5D}(\mathbf{x}_0, \omega) &= \frac{1}{2\pi} \sqrt{jk} \sqrt{2\pi |\mathbf{x}_{\mathbf{Ref}} - \mathbf{x}_0|}\, \mathbf{n}_{plane}^T \mathbf{n}(\mathbf{x}_0) A(\omega) e^{-jk\mathbf{n}_{PW}^T \mathbf{x}_0} \\ &= \sqrt{jk} \sqrt{\frac{|\mathbf{x}_{\mathbf{Ref}} - \mathbf{x}_0|}{2\pi}}\, \mathbf{n}_{plane}^T \mathbf{n}(\mathbf{x}_0) A(\omega) e^{-jk\mathbf{n}_{PW}^T \mathbf{x}_0} \end{aligned} \qquad (3.31)$$

Applying *(3.17)*, the wave field over the reproduction volume can be plotted as shown in Fig. 3.8.

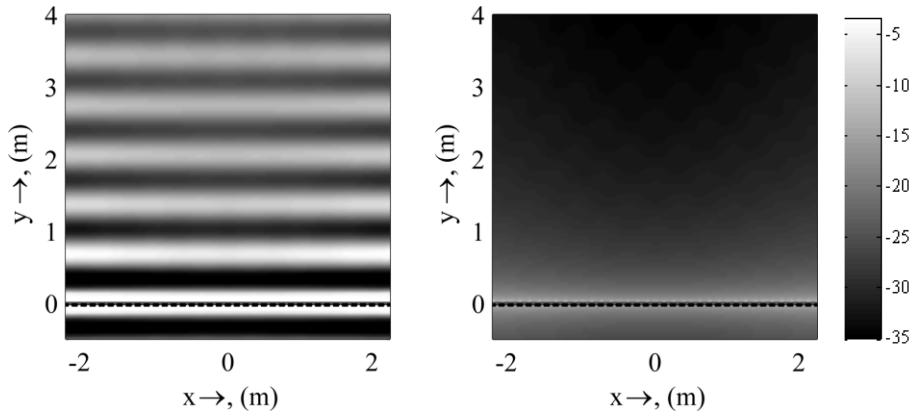


Fig. 3.8 2.5D WFS reproduction of a plane wave using an infinitely long array. Left: Real part of pressure field. Right: Pressure field in *dB*. Note the amplitude decay is not the same as a real plane.

The factor $\sqrt{jk}$ term in *(3.31)* is independent of both the primary and secondary source position, thus it can be computed in advance; consequently it is often referred to as the pre-filter. It is actually a difficult filter to implement recursively and there has been research done to approximate using fractional order IIR filters (Salvador, 2010) and more recently to perform the filtering of the entire WFS driving function recursively in real-time from a streamed input (Drumm and Oldfield, 2013). The frequency response of the pre-filter function is show in Fig. 3.9.

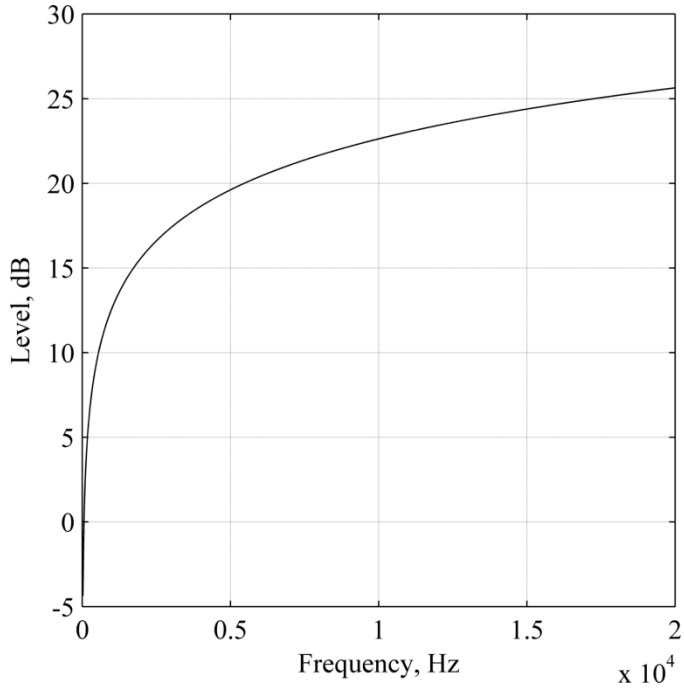


Fig. 3.9 Pre-filter function for plane wave and point sources in WFS

3.3.3.4 Monopole point sources

As the 2.5D driving function is based on the 2D driving function one should consider the 2D primary source function for the virtual source rather than a 3D source model. For a plane wave it is possible to use the 3D source function as the frequency characteristics are identical in both cases. For a spherical wave however, a line source should really be chosen as the primary source function, but line sources don't have the same spectral characteristics as point sources and as such need a frequency correction. In most cases a point source model is adopted for the derivation of the driving function (as shown below) as point sources don't have any spectral amplitude terms. However if a line source is used as the virtual source model, a

frequency correction term must be added into the equation. This frequency correction can be obtained using the far field approximation of the Hankel function and is given below.

For the purposes of comparison to other WFS methods and systems, the derivation of the 2.5D driving function for a point source is given here but will also be shown for a monopole line source in the following.

The source function at a point, $\mathbf{x}_0$ of a monopole point source at $\mathbf{x}_S$ is given by *(3.32)* as:

$$S_{point}(\mathbf{x}_0, \omega) = A(\omega)\frac{e^{-jk|\mathbf{x}_0 - \mathbf{x}_S|}}{|\mathbf{x}_0 - \mathbf{x}_S|} \qquad (3.32)$$

Where $\mathbf{x}_0$ and $\mathbf{x}_S$ are 3 dimensional vectors, $\mathbf{x}_0 = \{x_0, y_0, z_0\}$ and $\mathbf{x}_S = \{x_S, y_S, z_S\}$.

Defining $\mathbf{R} = (\mathbf{x}_0 - \mathbf{x}_S)$ such that $R = |\mathbf{x}_0 - \mathbf{x}_S|$ and $\hat{\mathbf{R}} = \frac{\mathbf{R}}{R}$, thus *(3.32)* becomes:

$$S_{point}(\mathbf{x}_0, \omega) = A(\omega)\frac{e^{-jkR}}{R} \qquad (3.33)$$

And the driving function becomes

$$D_{point,3D}(\mathbf{x}_0, \omega) = -\frac{1}{2\pi}A(\omega)\frac{\partial}{\partial \mathbf{n}}\frac{e^{-jkR}}{R} \qquad (3.34)$$

From *(3.9)* , this can be expressed as

$$D_{point,3D} = -\frac{1}{2\pi}A(\omega)\mathbf{n}(\mathbf{x}_0)\cdot\nabla\frac{e^{-jkR}}{R} \qquad (3.35)$$

where

$$\nabla \frac{e^{-jkR}}{R} = \hat{\mathbf{R}} \frac{\partial}{\partial R} \frac{e^{-jkR}}{R} = -\frac{\hat{\mathbf{R}}}{R} e^{-jkR} \left(jk + \frac{1}{R} \right) \tag{3.36}$$

So finally the 3D driving function for a monopole source in WFS can be expressed as:

$$D_{point,3D}(\mathbf{x}_0, \omega) = \frac{1}{2\pi} \frac{(\mathbf{x}_0 - \mathbf{x}_S) \cdot \mathbf{n}(\mathbf{x}_0)}{|\mathbf{x}_0 - \mathbf{x}_S|^2} A(\omega) e^{-jk|\mathbf{x}_0 - \mathbf{x}_S|} \left(\frac{1}{|\mathbf{x}_0 - \mathbf{x}_S|} + jk \right) \tag{3.37}$$

Thus it follows from *(3.27)* that the 2.5D driving function for a spherical wave can be given as:

$$D_{point,2.5D}(\mathbf{x}_0, \omega) = \sqrt{\frac{|\mathbf{x}_{\mathrm{Ref}} - \mathbf{x}_0|}{2\pi jk}} \frac{(\mathbf{x}_0 - \mathbf{x}_S) \cdot \mathbf{n}(\mathbf{x}_0)}{|\mathbf{x}_0 - \mathbf{x}_S|^2} \ldots \times A(\omega) e^{-jk|\mathbf{x}_0 - \mathbf{x}_S|} \left(\frac{1}{|\mathbf{x}_0 - \mathbf{x}_S|} + jk \right) \tag{3.38}$$

To write this equation in a more recognisable form that can be compared more easily with the traditional driving functions as originally formulated by Verheijen (1998), *(3.38)* can be rearranged as:

$$D_{point,2.5D}(\mathbf{x}_0, \omega) = \frac{(\mathbf{x}_0 - \mathbf{x}_S)^T \mathbf{n}(\mathbf{x}_0)}{|\mathbf{x}_0 - \mathbf{x}_S|^2} \sqrt{|\mathbf{x}_{Ref} - \mathbf{x}_0|} \ldots \times A(\omega) e^{-jk|\mathbf{x}_0 - \mathbf{x}_S|} \left(\frac{1}{\sqrt{jk2\pi}|\mathbf{x}_0 - \mathbf{x}_S|} + \sqrt{jk} \right) \tag{3.39}$$

This is similar to driving function as derived by (Spors et al., 2008) apart from an amplitude factor of $1/4\pi$ introduced as a result of Spors not including this factor in the Kirchhoff-Helmholtz equation. The frequency dependent terms have been grouped together for clarity.

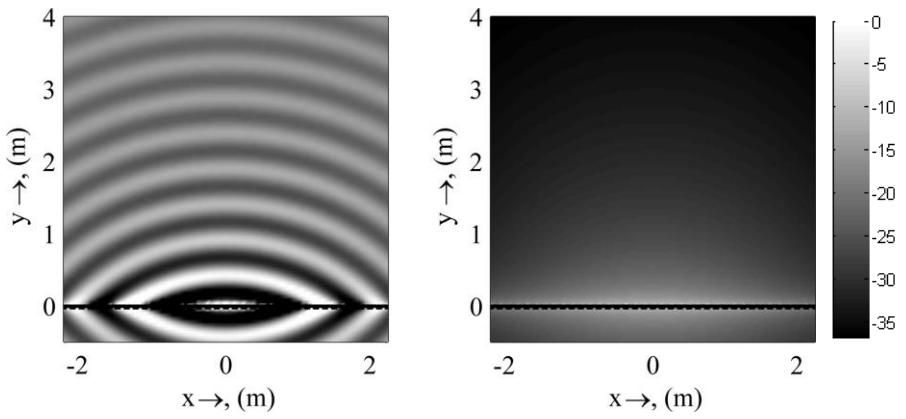


Fig. 3.10 2.5D WFS point source $f = 800Hz$ with an inter-loudspeaker spacing, $\Delta_{LS} = 0.125m$. Left: Real part of pressure field, Right: Pressure field in dB

3.3.3.5 Monopole line sources

As we are considering 2.5D WFS the virtual source model should ideally be a 2D source model such that the error as a result of the dimensionality miss-match is only caused by the secondary sources rather than the primary source model. Consequently a line source is used as the virtual source model in the following. The virtual source model of a line source is thus given as:

$$S_{line}(\mathbf{x}_0, \omega) = \frac{j}{4} H_0^{(2)}\left(k\left|\mathbf{x}_0 - \mathbf{x}_S\right|\right) \tag{3.40}$$

The problem with this source model is that it has a spectral characteristic that doesn't match that of a 3D point source, thus there will be errors in reproduction. These spectral characteristics can be eliminated from the driving function by once again examining the far field approximation of this source function.

$$S_{line}(\mathbf{x}_0, \omega) \approx -\sqrt{\frac{1}{jk}}\sqrt{2\pi\left|\mathbf{x}_0 - \mathbf{x}_S\right|}\,\frac{1}{4\pi}\frac{e^{-jk|\mathbf{x}-\mathbf{x}_0|}}{\left|\mathbf{x}_0 - \mathbf{x}_S\right|} \tag{3.41}$$

So to remove the spectral differences between this function and that of a 3D point source we need to multiply *(3.41)* by $\sqrt{jk}$ such that we now have

$$G_{2D}\left(\mathbf{x}_0,\omega\right)=\frac{j}{4}\sqrt{jk}\;H_0^{(2)}\left(k\left|\mathbf{x}_0-\mathbf{x}_S\right|\right) \tag{3.42}$$

So to calculate the driving function we need to determine from *(3.18)*

$$D\left(\mathbf{x}_0,\omega\right)=-\frac{1}{2\pi}\frac{\partial}{\partial\mathbf{n}}S_{line}\left(\omega\right)\equiv-\frac{1}{2\pi}\mathbf{n}.\nabla S_{line}\left(\omega\right) \tag{3.43}$$

So

$$\nabla S_{line}\left(\mathbf{x}_0,\omega\right)=\hat{\mathbf{R}}\frac{\partial}{\partial R}\frac{j}{4}\sqrt{jk}\;H_0^{(2)}\left(kR\right) \tag{3.44}$$

With $\hat{\mathbf{R}}=\mathbf{R}/R$ where $R=\left|\mathbf{x}-\mathbf{x}_0\right|$ and $\mathbf{R}=\left(\mathbf{x}-\mathbf{x}_0\right)$ such that from the chain rule, we have

$$\nabla S_{line}\left(\mathbf{x}_0,\omega\right)=-\hat{\mathbf{R}}\frac{j}{4}\sqrt{jk}\;k\;H_1^{(2)}\left(kR\right) \tag{3.45}$$

So from *(3.43)* the 2-Dimensional driving function for a line source can be formulated as:

$$D_{2D,line}\left(\mathbf{x}_0,\omega\right)=\frac{1}{2\pi}A\left(\omega\right)\hat{\mathbf{R}}\cdot\mathbf{n}\left(\mathbf{x}_0\right)\frac{jk}{4}\sqrt{jk}\;H_1^{(2)}\left(kR\right) \tag{3.46}$$

$$D_{2D,line}\left(\mathbf{x}_0,\omega\right)=-A\left(\omega\right)\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)\cdot\mathbf{n}\left(\mathbf{x}_0\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\frac{jk}{8\pi}\sqrt{jk}\;H_1^{(2)}\left(kR\right) \tag{3.47}$$

The generic large argument asymptotic expansion approximation is given as (Abramowitz and Stegun, 1965):

$$\begin{aligned}H_\nu^{(1)}\left(Z\right)&=\sqrt{\frac{2}{\pi Z}}e^{j\left(Z-\nu\pi/2-\pi/4\right)}\\H_\nu^{(2)}\left(Z\right)&=\sqrt{\frac{2}{\pi Z}}e^{-j\left(Z-\nu\pi/2-\pi/4\right)}\end{aligned} \tag{3.48}$$

Thus the driving function can be approximated as:

$$\begin{aligned} D_{2D,line}\left(\mathbf{x}_0,\omega\right) &\simeq -A(\omega)\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)\cdot\mathbf{n}\left(\mathbf{x}_0\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\frac{jk}{8\pi}\sqrt{jk}\ \sqrt{\frac{2}{\pi k\left|\mathbf{x}-\mathbf{x}_0\right|}}e^{-j\left(k\left|\mathbf{x}-\mathbf{x}_0\right|-3\pi/4\right)} \\ &\simeq A(\omega)\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)\cdot\mathbf{n}\left(\mathbf{x}_0\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\frac{jk}{8\pi}\ \sqrt{\frac{2}{\pi\left|\mathbf{x}-\mathbf{x}_0\right|}}e^{-j\left(k\left|\mathbf{x}-\mathbf{x}_0\right|\right)} \end{aligned} \qquad (3.49)$$

Remembering the relationship between the 2.5D and 2D driving functions:

$$D_{2.5D,line}\left(\mathbf{x}_0,\omega\right) = -D_{2D,line}\left(\mathbf{x},\omega\right)\sqrt{\frac{2\pi\left|\mathbf{x}-\mathbf{x}_0\right|}{jk}}e^{-j\left(k\left|\mathbf{x}-\mathbf{x}_0\right|\right)} \qquad (3.50)$$

So combining *(3.49)* and *(3.50)* results in the 2.5D driving function for a monopole line source as:

$$D_{2.5D,line}\left(\mathbf{x}_0,\omega\right) = \sqrt{\frac{2\pi\left|\mathbf{x}_{\mathrm{Ref}}-\mathbf{x}_0\right|}{jk}}A(\omega)\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)\cdot\mathbf{n}\left(\mathbf{x}_0\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\frac{jk}{8\pi}\sqrt{jk}\ H_1^{(2)}\left(kR\right) \qquad (3.51)$$

It is often useful to express driving functions by their large argument approximation as it separates out the pre-filter term which, as previously mentioned, can be applied to sources prior to the WFS rendering to reduce the computational overhead. Thus recalling the approximation made in *(3.49)*, *(3.51)* can be expressed as:

$$\begin{aligned} D_{2.5D,line}\left(\mathbf{x}_0,\omega\right) &\simeq \sqrt{\frac{2\pi\left|\mathbf{x}_{\mathrm{Re}\,f}-\mathbf{x}_0\right|}{jk}}\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)\cdot\mathbf{n}\left(\mathbf{x}_0\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\frac{jk}{8\pi}\sqrt{\frac{2}{\pi\left|\mathbf{x}-\mathbf{x}_0\right|}}A(\omega)e^{-j\left(k\left|\mathbf{x}-\mathbf{x}_0\right|\right)} \\ &\simeq \frac{1}{4\pi}\sqrt{jk}\sqrt{\frac{\left|\mathbf{x}_{\mathrm{Re}\,f}-\mathbf{x}_0\right|}{\left|\mathbf{x}-\mathbf{x}_0\right|}}\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)\cdot\mathbf{n}\left(\mathbf{x}_0\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}A(\omega)e^{-j\left(k\left|\mathbf{x}-\mathbf{x}_0\right|\right)} \end{aligned} \qquad (3.52)$$

This equation can be considered to be superior to *(3.39)* as the pre-filter of $\sqrt{jk}$ is more straightforward and is the same as that for plane waves as shown in Fig. 3.9 thus is easier to implement in a real system.

3.3.3.6 Dipole sources

The sound field of a dipole line source can be given as:

$$S_{Dipole,line}\left(\mathbf{x}_0,\omega\right)=-A\left(\omega\right)jk\pi\cos\theta H_1^{(2)}\left(k\left|\mathbf{x}-\mathbf{x}_0\right|\right) \qquad (3.53)$$

The derivation of the driving function for a focused dipole is given in full detail in chapter 4 and the working can easily be adapted for the non-focused case. For an non-focused dipole source therefore, the driving function can be expressed as:

$$D_{Dipole,2.5D}\left(\mathbf{x}_0,\omega\right)=A\left(\omega\right)\sqrt{k}\sqrt{\frac{\left|\mathbf{x}_{Ref}-\mathbf{x}_0\right|}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\,\pi x_S\,\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|^2}\cdot\mathbf{n}\left(\mathbf{x}_0\right)\left(1+jk\left|\mathbf{x}-\mathbf{x}_0\right|\right)\frac{e^{-jk\left|\mathbf{x}-\mathbf{x}_0\right|}}{\left|\mathbf{x}_0-\mathbf{x}_S\right|} \qquad (3.54)$$

This can basically be thought of as monopole radiation with an angular dependant term and an additional amplitude term. Fig. 3.11 shows the resultant wave field produced by the driving function, clearly showing the positive and negative lobes characteristic of dipole sources.

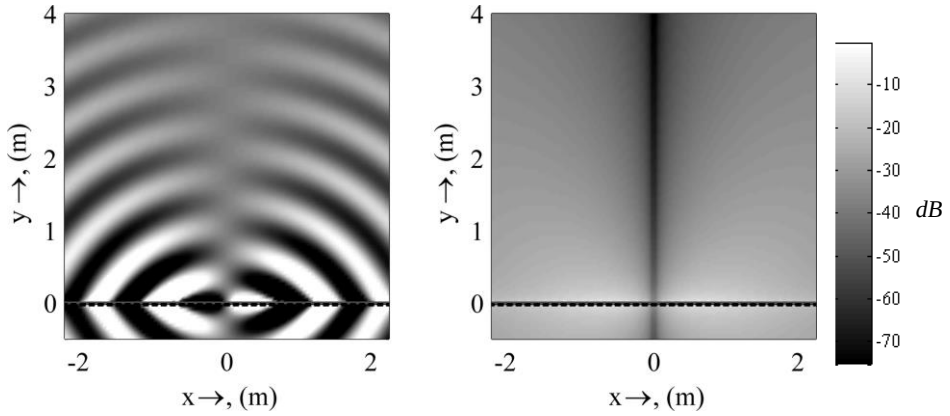


Fig. 3.11 2.5D WFS dipole source $f=800Hz$ with $\Delta_{LS}=0.125m$. Left: Real part of pressure field. Right: Pressure field in *dB*

3.3.3.7 Sources with more complex directivity patterns

It has been shown by several authors (Ahrens and Spors, 2007; Baalman, 2008; Corteel, 2007; Ahrens and Spors, 2008b), that WFS systems can reproduce primary sources with arbitrary directivity characteristics although the derivation of such driving functions is more complex.

A driving function for a source with arbitrary directivity characteristics defined by spherical harmonics can be adapted from Ahrens (2012):

$$D_{copmlex,2.5D}\left(\mathbf{x}_0,\omega\right)=A\left(\omega\right)\sqrt{\frac{\left|\mathbf{x}_{\mathrm{Ref}}-\mathbf{x}_0\right|}{2\pi jk}}\cos\left(\alpha_n-\alpha_0\right)\frac{e^{-jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\ldots \times\sum_{m=0}^{N-1}\sum_{m=-n}^{n}j^n S_n^m\left(\omega\right)Y_n^m\left(\frac{\pi}{2},\alpha_0\right) \qquad (3.55)$$

Where

$$S_n^m\left(\omega\right)=\begin{cases}\left(-1\right)^{(m+n)}\ j^{-n}\dfrac{\left(N-1\right)!N!}{\left(N+n\right)!\left(N-n-1\right)!}Y_n^{-m}\left(\beta_{or},\alpha_{or}\right), & \forall n\le N-1\\ 0\ , & \text{elsewhere}\end{cases} \qquad (3.56)$$

Here N is the order or spherical harmonics α_{or} and β_{or} are the angles of the source propagation in terms of azimuth and elevation respectively, α_0 is the reference angle and α_n is the angle of which is the driving function is calculated at i.e. the angle between the primary and secondary source. The resultant wave field produced by this driving function is shown in Fig. 3.12.

Sources with more complex radiation patterns are not used in this thesis and therefore no greater details on their rendering will be given here. However the next chapter does present the theory for the driving function of a virtual loudspeaker using the source model of a circular piston in an infinite baffle which has a frequency dependent directivity pattern.

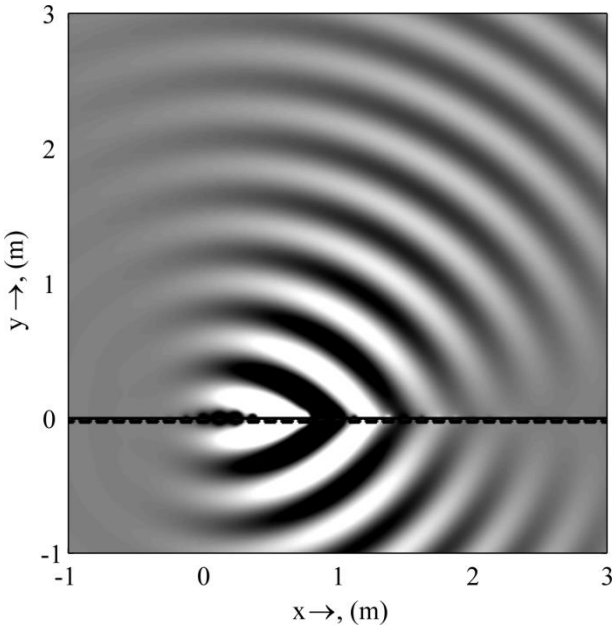


Fig. 3.12 2.5D WFS source with complex directivity $f = 800Hz$ with $\Delta_{LS} = 0.125m$. Based on a Matlab script from (Ahrens, 2012), $\alpha_{or} = \pi/3$, $\beta_{or} = \pi/2$ and $N = 31$.

3.4 Secondary source selection

The above driving functions can readily be applied to arbitrary shaped loudspeaker arrays e.g. circular (as shown in Fig. 3.13 and Fig. 3.14) if the correct secondary source selection is made. The source selection is most often made by comparison of the source direction vector and the direction unit vector of the loudspeakers. The criterion is slightly different depending on the source type. For plane waves the source selection criteria is given by Spors (2003) as:

$$a_{plane}\left(\mathbf{x}_0\right) = \begin{cases} 1, & \text{if } \left\{\mathbf{n}_{plane} \cdot \mathbf{n}\left(\mathbf{x}_0\right)\right\} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.57}$$

Where $\{\cdot\}$ is the dot product of the two vectors. Simply stated, a loudspeaker is set to be active when the angle between the plane wave and the angle normal to the loudspeaker array at loudspeaker position $\mathbf{x}_0$ is less than 90°.

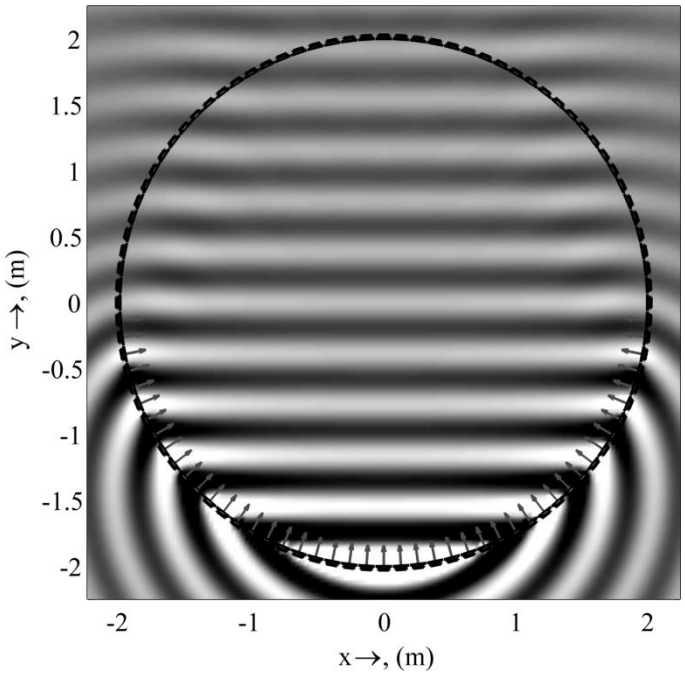


Fig. 3.13 2.5D WFS plane wave $f = 800Hz$ with $\Delta_{LS} = 0.125m$ with a circular array of loudspeakers. The arrows denote the direction vector of the active loudspeakers that satisfy *(3.57)*

For a source with a specific location such as a monopole, dipole or other directional sources the source selection criteria can be expressed as:

$$a_{point}\left(\mathbf{x}_0\right) = \begin{cases} 1, & \text{if } \left\{\left(\mathbf{x}_0 - \mathbf{x}_S\right)\cdot \mathbf{n}\left(\mathbf{x}_0\right)\right\} > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (3.58)$$

In this case it is the angle between the primary and secondary source which must be less than 90° for the loudspeakers to be active. The source selection criterion for focused sources is slightly more complicated and will be dealt with in the next chapter (see section 4.2).

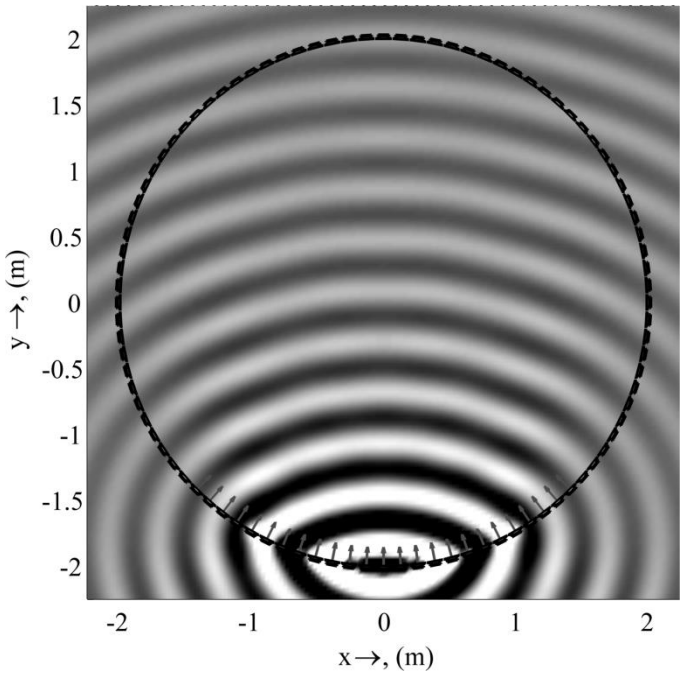


Fig. 3.14 Pressure field of a 2.5D WFS point source $f = 800Hz$ with $\Delta_{LS} = 0.125m$ with a circular array of loudspeakers. The arrows denote the direction vector of the active loudspeakers that satisfy *(3.58)*

3.5 Reproduction Errors

Wave field synthesis reproduction is based on a physical model of sound fields as described by the Kirchhoff-Helmholtz equation, but the simplifications needed to enable practical sound field reproduction mean that there are several sources of error in the reproduction. The principle errors that are found in real WFS systems are amplitude errors, resulting from the transformation from 3D to 2D using point sources rather than line source, truncation errors, due to the finite size of the secondary source array and spatial aliasing as a consequence of a discretised non-continuous secondary source array. This section briefly reviews these errors and the consequences thereof.

3.5.1 Amplitude errors

Most WFS systems concentrate on the accurate reproduction in one plane only mainly for practical reasons apart from exceptional systems (Komiyama et al., 1991) where a planar array of loudspeakers were employed for distance perception tests. The benefits of such a system however rarely out-ways the large amount of effort

and hardware needed for its implementation and consequently they are rarely if ever seen.

Amplitude errors are inherent in a non-3D implementation of WFS. As shown in Fig. 3.7, for a plane wave scenario the amplitude decay is proportional to $1/\sqrt{r}$ (where r is the distance from the virtual source to receiver) which equates to approximately 3dB per doubling of distance, which is in contrast to a real plane wave which exhibits no attenuation with respect to distance. This amplitude error is a consequence for using a 2D array of point sources rather than planar arrays as would be the case in a true 3D setup. This means that plane wave virtual sources are actually cylindrical waves and thus exhibit distance attenuation accordingly (Toole, 2008).

For virtual monopole sources, the error in distance attenuation between real and 2.5D WFS sources is dependent on the virtual source distance to the loudspeaker array and is plotted for reference in *dB* in Fig. 3.15.

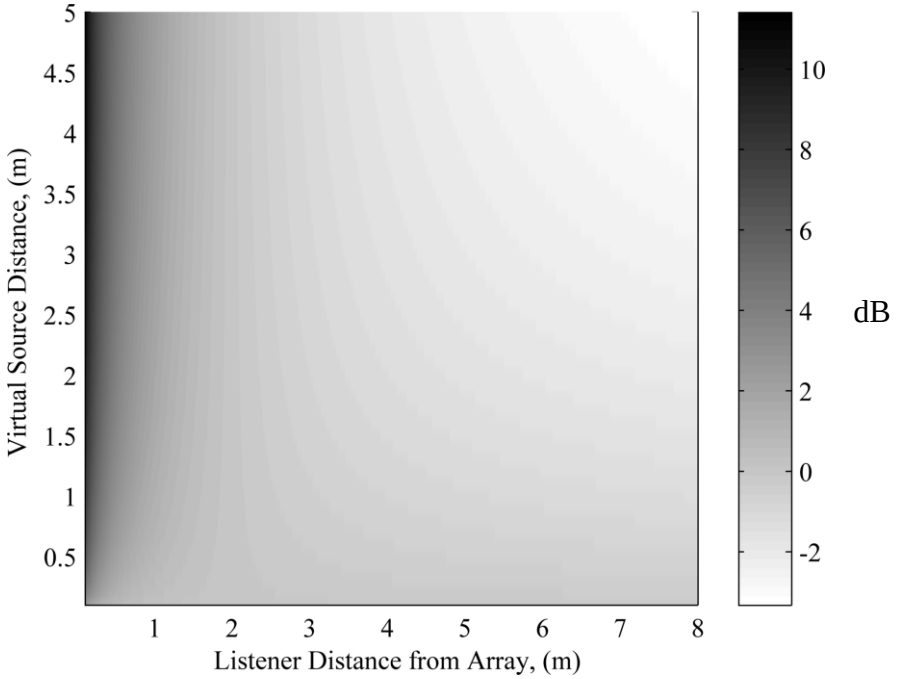


Fig. 3.15 Error in *dB* between the distance attenuation of a real point source and a 2.5D WFS source with varying virtual source distances. The reference point is chosen as *2m* from the array.

This plot shows that the amplitude mismatch between real and WFS sources is greatest for virtual sources far from the loudspeaker array and when the listener is close to the secondary sources as is intuitive because this is where the gradient of the decay from the secondary sources is greatest and when the primary source is a long way away from the listener in a real scenario the gradient of the amplitude decay is much less.

The amplitude errors associated with WFS reproduction have been well documented in the literature for non-focused sources and more details can be found in (e.g. Sonke et al., 1998 and Verheijen, 1998). The amplitude errors pertaining to focused source reproduction can be found in the next chapter.

3.5.2 Spatial sampling errors

3.5.2.1 Introduction

The driving functions presented in this chapter have been derived using the assumption that the reproduction array is continuous. Continuous arrays however are not currently realisable; hence real WFS systems introduce a compromise by sampling the secondary source array at discrete locations using loudspeakers. This spatial sampling is often dictated by driver dimensions, limiting how close they can be placed together. The physical artefacts in the reproduced wave field introduced by this sampling manifest in the form of additional wave contributions in a different direction to the intended source, reducing the spatial resolution of the reproduced wave front and thus introducing distortion in the spatial and timbral fidelity of the intended sound field as shown in Fig. 3.16 and Fig. 3.17.

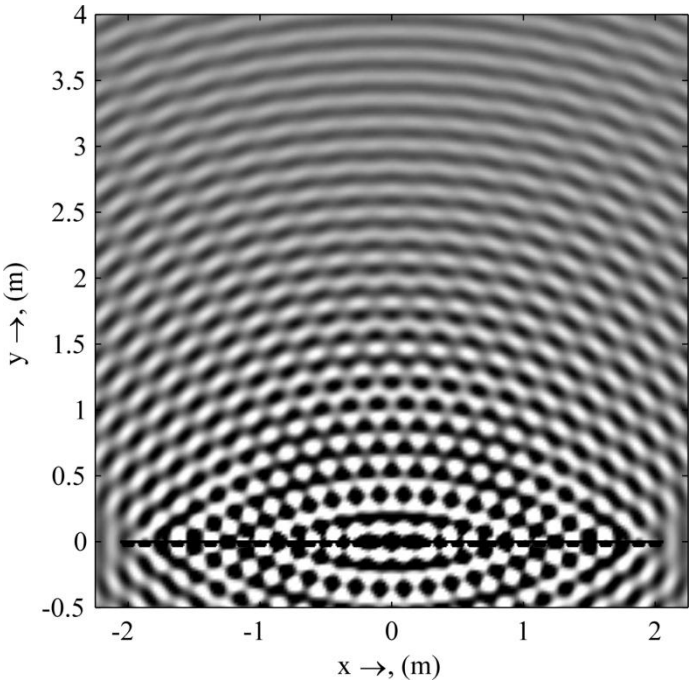


Fig. 3.16 WFS point source exhibiting spatial aliasing. $f = 2500Hz$ using an array of loudspeakers $2m$ long with $\Delta_{LS} = 0.125m$

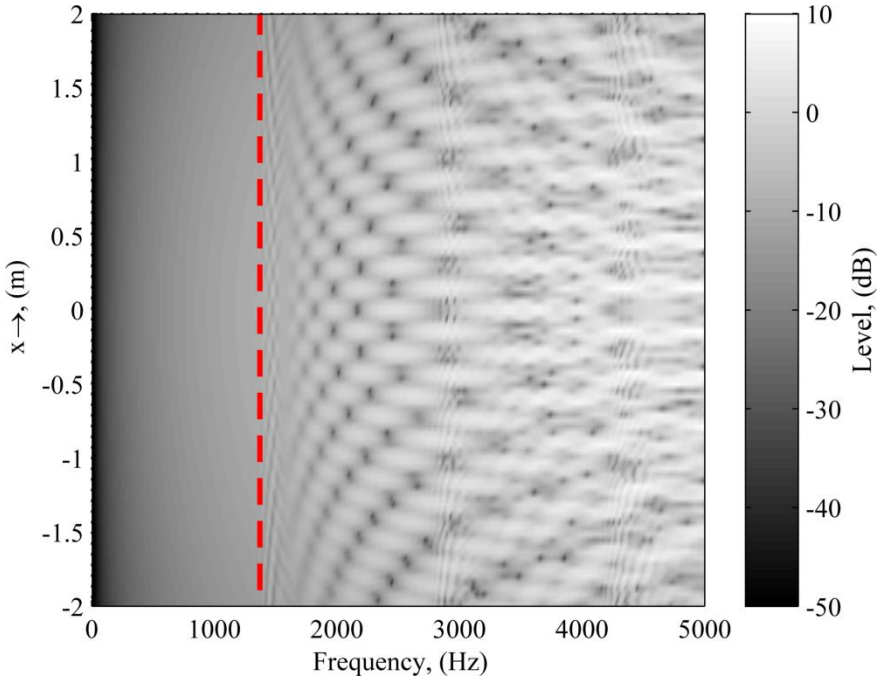


Fig. 3.17 Broadband WFS point source $1m$ behind a loudspeaker array with $\Delta_{LS} = 0.125m$ registered along a line of receivers $0.5m$ in front of the array, exhibiting spatial aliasing. The red line shows the aliasing frequency $f_{alias} = 1372Hz$ beyond which colouration is seen in the reproduced wave field.

Generally, spatial aliasing artefacts don't cause significant localisation problems and are often considered to be imperceptible in a non-anechoic environment (Spors and

Rabenstein, 2006), however spatial aliasing effects will manifest perceptually as a position dependent colouration in the reproduced sound field (as shown in Fig. 3.17), a fact that was further investigated by De Bruijn (2004) who looked specifically at the differences in colouration with different loudspeaker spacings, noting that as long as the loudspeaker spacing was below *~0.25m* the colouration introduced was not perceptually significant.

Spatial aliasing artefacts are thus caused when the spacing between secondary sources is too great to accurately reproduce the desired wave front; it is a sampling problem in space analogous to temporal sampling in digital audio conversion and as such can be analysed in a similar manner but where the analysis of temporal aliasing is done in the frequency domain, the analysis of spatial aliasing is done in the spatio-temporal frequency domain or *wavenumber domain* (as described in section 3.5.2.3). The most comprehensive analysis of spatial aliasing artefacts using this technique is given by (Spors and Rabenstein, 2006) and many of the same principles used in this paper have been applied in the analysis presented here.

3.5.2.2 History of the analysis of spatial aliasing

Spatial aliasing errors have been well documented in the literature, since the earliest conception of WFS e.g. (Start, 1997; Verheijen, 1998; Corteel, 2006; Spors and Rabenstien, 2006; Wittek, 2009) resulting in several different versions of the equation for the frequency above which, spatial aliasing artefacts can be expected in the reproduced wave field, f_{alias} , here follows a brief overview of some of the important developments in this area.

The traditional method draws upon an analogy of time domain sampling utilising a spatial application of the Shannon sampling criteria. Shannon sampling theory states that the sampling of a time domain signal needs to be made twice per wavelength to result in enough information for a correct reconstruction of the original signal, giving rise to a sampling criterion of

$$f_{\max} = \frac{f_S}{2} \tag{3.59}$$

where f_S is the sampling frequency of the signal. If the signal is sampled less than this, unwanted aliased frequencies will appear in the reconstructed waveform due to the ambiguity in level fluctuations between samples that arises in the discretised signal.

Applying this sampling criteria for a spatial application demands that the spatial sampling of the sound field contains at least two secondary sources per wavelength such that the spatial aliasing frequency is given as (Verheijen, 1998);

$$f_{alias} = \frac{c}{2\Delta_{LS}} \tag{3.60}$$

This definition is rather simplistic in that it fails to take into account the differing directions of propagation of the plane wave components of the virtual source, and thus this constitutes the strictest form of the spatial aliasing criteria. This criteria is useful in that it holds true for all source types and incidence angles, however it will predict aliasing for some situations where there will be no such artefacts. In reality f_{alias} is dependent upon the angle at which the source is incident on the array as is pointed out by (Start, 1997) and can be seen from Fig. 3.18b which demonstrates that the wavelength as registered along the axis of the loudspeaker array is greater than the signal wavelength, thus allowing a larger loudspeaker spacing whilst still avoiding aliasing.

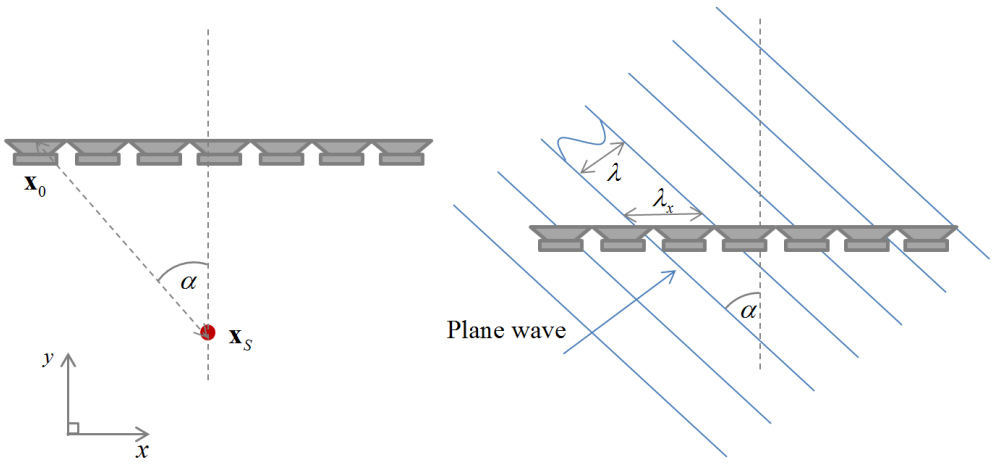


Fig. 3.18 Source geometry for the definition of the spatial aliasing criteria. a) point source propagation, b) plane wave propagation

This gives rise to another spatial aliasing criterion with the inclusion of the source angle.

$$f_{alias} = \frac{c}{2\Delta_{LS}\sin\alpha_{max}} \tag{3.61}$$

The problem with this aliasing criterion is that at grazing incidence when $\alpha_{\max}$ is 0°, the aliasing frequency becomes infinite which is unrealistic. Consequently a less optimistic equation has been presented by Spors and Rabenstein (2006), given by *(3.62)*.

$$f_{alias} = \frac{c}{\Delta_{LS}\left(1+\left|\sin\alpha_{max}\right|\right)} \tag{3.62}$$

This condition was formulated based on a plane wave virtual source but the authors note that all sources can be split into plane wave components therefore the model described can be used for the generic case as well. When looking at the extremes of possible source angles ($\alpha_{\max} = 0°$ and $\alpha_{\max} = \pm 90°$) it is clear that this formula predicts the following aliasing condition.

$$\frac{c}{\Delta_{LS}} > f_{alias} > \frac{c}{2\Delta_{LS}} \tag{3.63}$$

All of these conditions are derived based on a frequency or spatial frequency domain approach but it is also possible to perform the analysis in the time domain as per (Corteel, 2006a). Such a temporal description has the advantage of describing the aliasing effects with respect to listener location which can be useful in some circumstances.

$$f_{alias}(\mathbf{x},\Psi) = \frac{1}{\max_{l=1\to L-1}\left\{t_{l+1}(\mathbf{x},\Psi) - t_l(\mathbf{x},\Psi)\right\}} \tag{3.64}$$

Where $\max_{l=1\to L-1}\left\{t_{l+1}(\mathbf{x},\Psi) - t_l(\mathbf{x},\Psi)\right\}$ is the maximum time difference of arrival at the listener location, $\mathbf{x}$ from successive loudspeakers for a given virtual source, Ψ where L is the total number of loudspeakers in the reproduction array. This criterion can be shown to produce similar results to *(3.62)* but is more complicated to calculate and for many circumstances the spatial aliasing frequency formulation acts only as a guide and it is more useful to simply have a frequency above which aliasing will occur rather than knowing the position dependent aliasing effects unless e.g. room equalisation algorithms are being employed (Corteel, 2006b).

3.5.2.3 *Wavenumber domain*

In the same way that temporal aliasing can be viewed in the frequency domain, spatial sampling can be viewed nicely in the spatial frequency domain often referred to as the *wavenumber domain* or *'k-space'*

The wavenumber domain provides a very useful representation of reproduced sound fields and driving functions and is an excellent tool for analysing attributes of a given sound field such as spatial aliasing (Spors and Rabenstien, 2006), near field effects (Spors and Ahrens, 2007) and truncation (Start, 1997) etc.

The spatial Fourier Transform of a function, $S(\mathbf{x},\omega)$ can be thought of as the spatial equivalent to the temporal Fourier Transform and can be formulated as *(3.65)*.

$$\tilde{S}(k_x, y, z, \omega) = \int_{-\infty}^{\infty} S(\mathbf{x},\omega) e^{-jk_x x dx} \tag{3.65}$$

In this equation, k_x is the variable in the Fourier domain and $\omega = 2\pi f$, is the angular frequency. Although this operation can be applied independently along each of the three Cartesian directions it is given here only in the x-direction for convenience. The inverse of this operation can also be performed as *(3.66)*

$$S(\mathbf{x},\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{S}(k_x, y, z, \omega) e^{-jk_x x dx} dk_x \tag{3.66}$$

One of the great advantages of using the spatial frequency representation is that it allows, on inspection, the direction of the propagating wave(s) to be determined and indeed offers a technique for a plane wave decomposition of a function. A broadband plane wave plotted in the wavenumber domain is shown in Fig. 3.19 the equivalent, monochromatic plane wave would simply be a discrete point on this plot rather than a line.

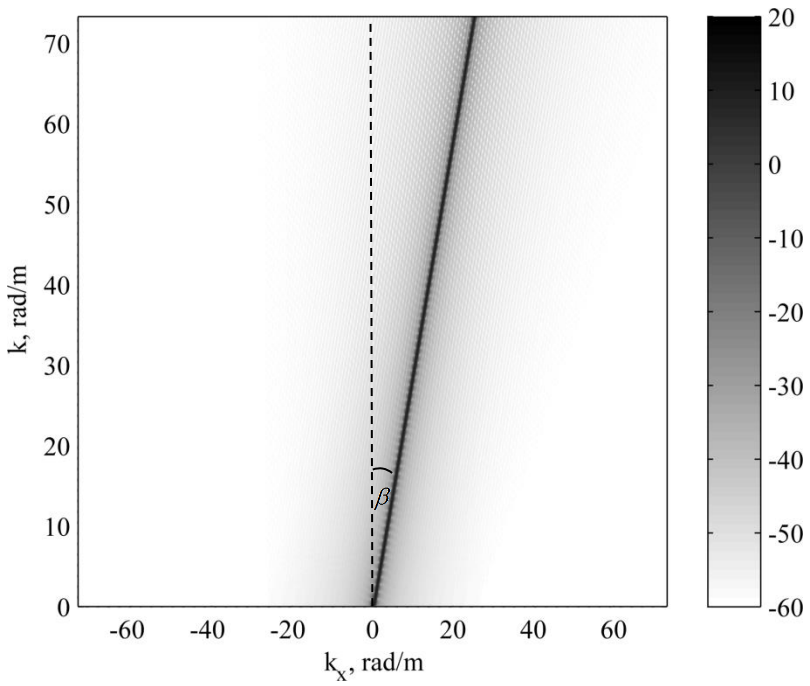


Fig. 3.19 Broadband plane wave at $\theta = \boldsymbol{\pi}/\mathbf{4}$ radians with respect to the *x*-direction in the wavenumber domain

The angle of the line with respect to the *x*-axis determines the direction of the propagating plane wave and is given by *(3.67)*.

$$\tan\beta = \frac{k_x}{k} = \sin\theta \qquad (3.67)$$

Where β is the angle with respect to the *y*-axis (shown in Fig. 3.19) and θ is the propagation angle of the plane wave component with respect to the *x*-axis.

Equation *(3.66)* can be thought of as a plane wave decomposition of the wave field into monochromatic plane waves whose direction is given by *(3.67)* hence for a spherical source where the propagating energy spreads out equally in all directions, the k-space representation takes the form of a triangular area depicting the propagating part of the sound field (as shown in Fig. 3.20). If the transform is done in three dimensions then the triangular section becomes a conical section.

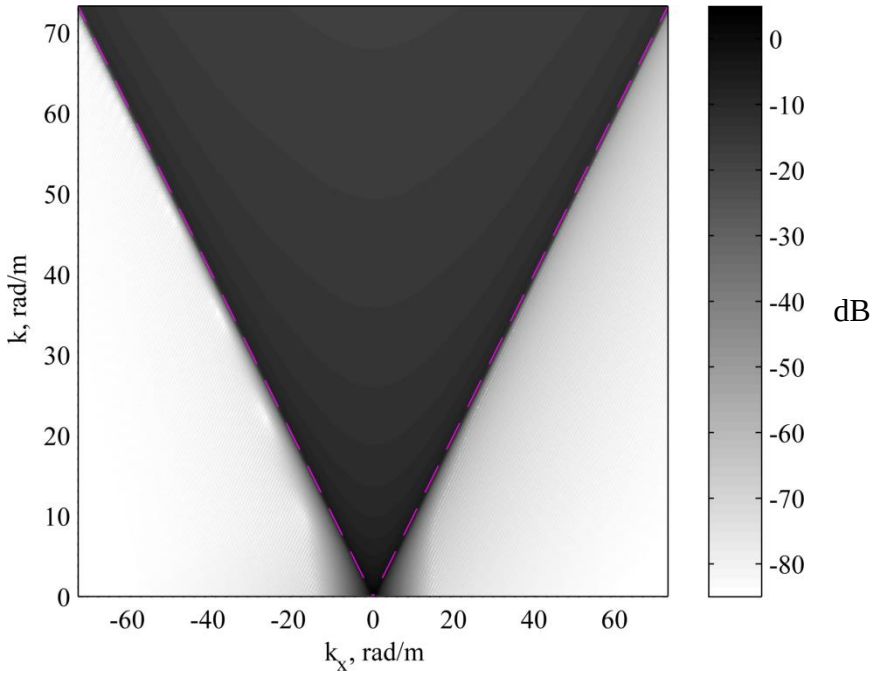


Fig. 3.20 Broadband point source in the wavenumber domain, the magenta lines demark the boundary between the propagating and evanescent contributions

The Fourier variable, k_x can take on any value between $-\infty$ and ∞ meaning that the wave field can be decomposed into both *propagating waves* (where $|k_x| < k$) and an *evanescent waves* (where $|k_x| > k$). The magenta dotted line in Fig. 3.20 depicts the boundary in spatial frequency between these propagating and evanescent contributions. Evanescent waves are caused at the source position and die away exponentially in the source direction in a frequency dependent manner such that the higher frequencies decay faster than the low frequency contributions. The perceptual consequence and audibility of these evanescent contributions is not well understood in the literature but an analysis of these contributions in the spatial frequency domain has been done by Spors and Ahrens (2007) to demonstrate the near field effects of WFS sources.

It is possible to use this technique to analyse sound fields, by placing a row of receivers and then calculating the spatial Fourier transform along the line. This can be used to show the frequency at which spatial aliasing occurs as shown by Fig. 3.21.

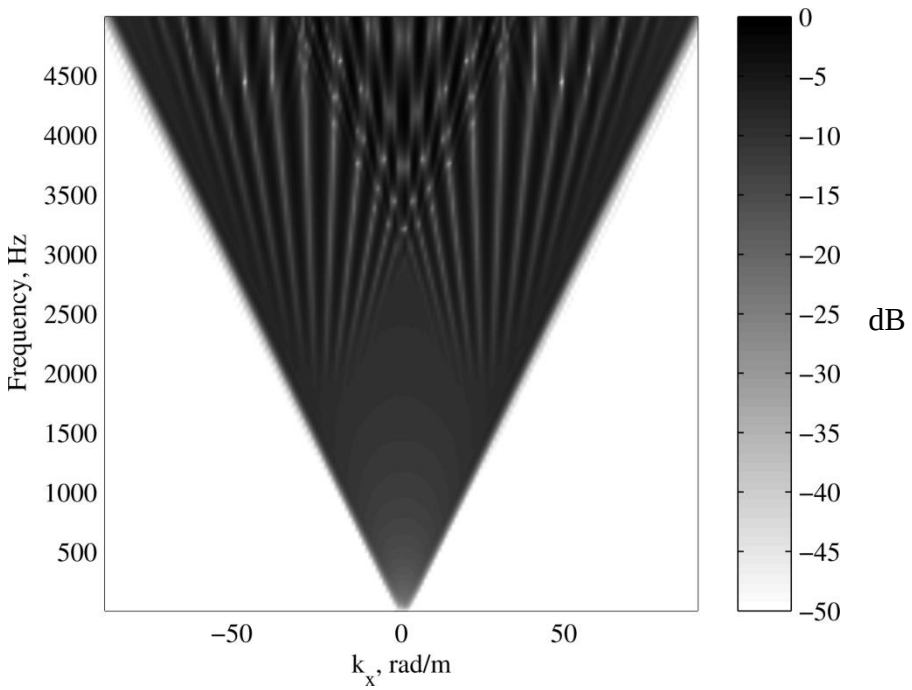


Fig. 3.21 Aliased broadband WFS point source, $\Delta_{LS} = 0.125m$ ($f_{alias} = 1372Hz$), placed *1m* in front of the centre of a *10m* loudspeaker array.

Fig. 3.21 shows that the aliased contributions form additional propagating components in a direction different from the intended virtual point source.

3.5.3 Truncation errors

Truncation errors are caused by the finite dimensions of the secondary source array. The effects of the array truncation are caused by diffraction waves radiating from the edges of the loudspeaker array. These diffraction effects can be seen in the broadband case shown in Fig. 3.24a and also from the curved wave fronts in Fig. 3.22. In the broadband case, the artefacts from array truncation result in echoes or pre-echoes for focused sources (as described in section 4.10). The consequences of secondary source array truncation are that the listening area is reduced and also the wave front curvature will be decreased. Spors states that truncation could also lead to an increases spatial aliasing frequency (Spors and Ahrens, 2010b).

The effects of truncation can be reduced to some extent as described in section 3.5.3.1.

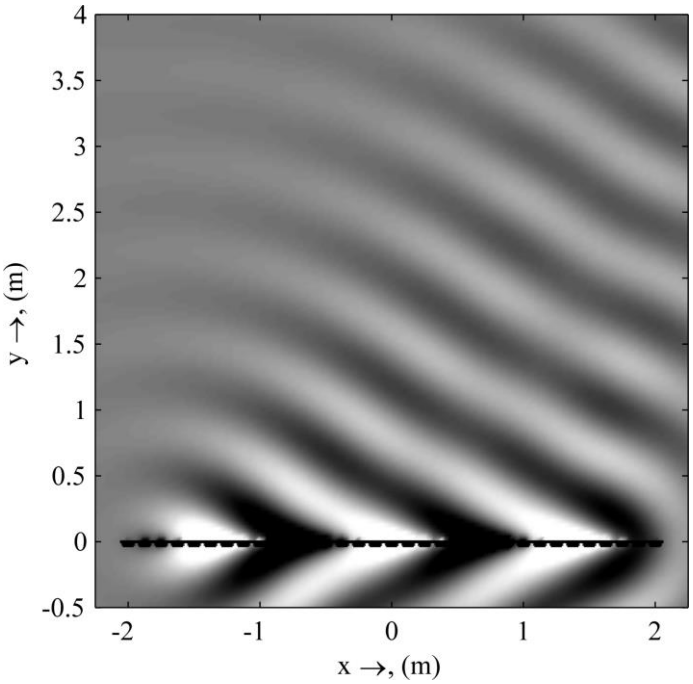


Fig. 3.22 WFS plane wave at *60°* to *x*-axis, showing wave font distortion at the edges due to truncation error. The loudspeaker array in this case was *4m* long with $\Delta_{LS} = 0.125m$

For a linear array, truncation effects can be modelled by multiplying the driving function by a windowing function $w(x_0)$ as shown by Start (1997). The reproduced wave field of the truncated array can then be given from *(3.21)* as,

$$P_{trunc}(\mathbf{x}_0, \omega) = \int_{-\infty}^{\infty} w(x_0) D(\mathbf{x}_0, \omega) G(\mathbf{x}|\mathbf{x}_0, \omega) dx_0 . \qquad (3.68)$$

The windowing function can be any windowing function but for the analysis of the errors of array truncation a sensible choice of windowing function is a rectangular window given by,

$$w(x_0) = \begin{cases} 1, & \text{if } |x_0| \le \frac{L}{2} \\ 0, & \text{otherwise} \end{cases} \qquad (3.69)$$

where L is the length of the loudspeaker array. It can be helpful to plot the effects of the array truncation in the wavenumber domain. As such the windowing function can be written in the wavenumber domain, which for the rectangular window given in *(3.69)* is expressed as:

$$w(x_0, k_x) = |L| \operatorname{sinc}\left(\frac{k_x}{2} L\right) \tag{3.70}$$

Thus for the rectangular case the truncation errors will manifest as spatial amplitude lobes in the reproduced wave field, in the same way that a time domain signal will result in peaks and troughs in the frequency response if multiplied by a rectangular window. In the temporal domain, these lobes are reduced by using different window shapes such as a Hanning window etc. and thus the frequency response can be smoothed. In the same way the effects of a rectangular truncation of the secondary array in spatial frequency domain can be reduced to some extend by using a tapering window rather than a rectangular window as described in the next section.

3.5.3.1 Reduction of truncation errors

The unwanted artefacts associated with the truncation of the secondary source array can be reduced to some extent by applying the well-known technique of tapering/windowing of the loudspeaker array (Vogel, 1993; Boone et al., 1995; Start, 1997; Verheijen, 1998). Tapering involves decreasing the level of the loudspeakers at the edges of the array. This is typically done using a $\cos^2$ window, half-Hanning or Tukey window as shown in Fig. 3.23. Practically, the length of this tapering function is usually about 20% of the array dimension array (Start, 1997) but a strict criteria for the window length cannot be found in the current literature, the first and last loudspeakers in the array are generally not set to zero amplitude but rather the array is considered to have an imaginary loudspeaker at each end with zero amplitude, to maximise the use of the available hardware.

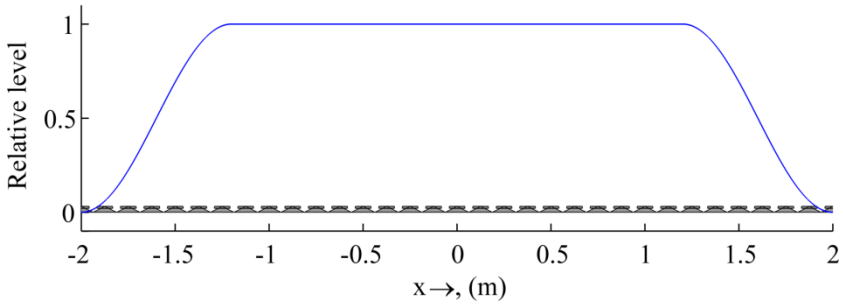
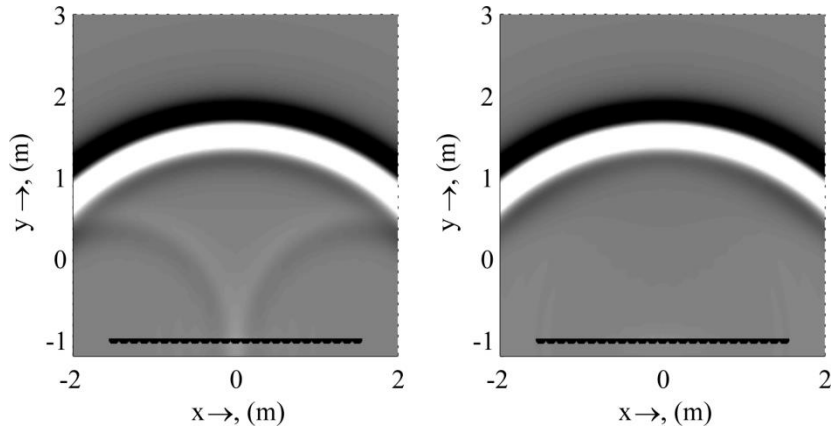


Fig. 3.23 Graphical representation of loudspeaker tapering

De Vries et al. (1994) suggested an alternative approach of modelling the diffraction waves from the edges of the array as additional sources with a specific directivity (although the implementation used point sources), and showed good results in decreasing the level of the truncation errors by placing additional sources at the edge of the array to reduce these contributions significantly at a desired listening position. Although the results of this technique are good, the method doesn't offer any advantages over the standard tapering technique and as such it is not often used.



a) b)

Fig. 3.24 Narrowband ($fx = 1000Hz$) 2.5D WFS point source, at $\mathbf{x}_S = [0m, -1.5m]$ with $\Delta_{LS} = 0.125m$. a) Un-windowed and b) Windowed array

Tapering can also be applied to the corners of arrays where, if untreated the discontinuity can cause errors in the reproduced sound field, a more thorough treatment of corner effects can be found in (Verheijen, 1998).

3.6 Practical WFS implementations

With the growing popularity of wave field synthesis as a sound reproduction technique, there has been an ever-growing number of systems being installed throughout the world. Popular commercial solutions include the IOSONO® and SonicEmotion™ systems but the resulting expense of these systems has led many institutions to seek their own custom built WFS installations. A review of many of the more prominent WFS systems has been published as an AES monograph (De

Vries, 2009), this document describes in detail the hardware and software of these systems and also gives an overview of the theory of WFS. Whilst all WFS systems are different in purpose and construction, several key aspects pervade all of the designs; that is they all require a large number of discrete audio channels and a powerful workstation on which to calculate the loudspeaker signals.

This section describes the design, installation and commissioning of four wave field synthesis systems at the University of Salford installed as part of this project. Whilst all of the systems have much in common in terms of their architecture and hardware, each system is configured in a specific way to fit the requirements of the installation environment and intended uses of the system. The first system described is a 128-channel semi-portable system which can be erected in any suitably sized room. This system is used predominantly for the testing and optimisation of WFS algorithms and for much of the work described here has been set up in the anechoic chamber; this introduces several technical and practical challenges as will be described in the following. The second system is a smaller 64-channel portable system which can be transported easily for various demonstrations and trade shows etc. The third system is a 112-channel permanent installation in a ITU spec listening room (Oldfield et al., 2011). The fourth system is a 128-channel permanent system installed in the OCTAVE 3D virtual reality laboratory at the University of Salford.

3.6.1 General Considerations

Regardless of the size and location of a WFS installation, there are a number of general issues that must be addressed such as the hardware choice and mounting.

3.6.1.1 Loudspeakers

For this project it was decided that passive drivers were more suitable than active drivers as they are lighter weight, generally have a smaller rim (allowing the drivers to be closer together increasing the spatial aliasing frequency) and to provide greater flexibility for future applications. It is also easier to distribute the power to passive

loudspeakers, as multichannel amplifiers can be used, requiring one mains connection per 8 channels (16 in total as opposed to 128 if active drivers are used).

All of the systems described here consist of arrays of the same type of drivers. The loudspeakers that were chosen were KEF CR100QR, which comprise a *100mm* woofer and a coaxial *19mm* tweeter in a sealer *0.7 litre* enclosure which also contains the passive crossover (at *3kHz*). Originally designed for in-ceiling installation, they are ideally suited to creating WFS arrays since they include mounting hardware and their individual enclosures eliminate coupling through the enclosure volume. Individually they reproduce *110Hz* to *27kHz* ±*6dB* and can be fitted with separation of just *12.95cm*, giving an aliasing frequency of approximately *1.33kHz*.

These loudspeakers were specifically chosen for their coaxial design i.e. the tweeter is in the centre of the woofer. This design means that the radiation centre remains the same for all frequencies thus the phase response is very stable and the wave front coherency is also very good, as such they are ideally suited to sound field reconstruction techniques.

For all of the systems described in this section, the nominal spacing between the drivers was kept the same across the array, however this spacing increases slightly at the point where the arrays are joined (as can be seen from Fig. 3.27a) and from the corners (as seen in Fig. 3.26). This has only a minor effect on the reproduced wave field in terms of the spatial aliasing artefacts but not on the reproduced waveform below the aliasing frequency as the loudspeaker positions are explicitly input into the driving function. The effects of this slight irregularity on the spatial aliasing can be however be quantified by using a time domain approach as shown by Corteel (2006a) (see section 3.5.2.2).

3.6.1.2 Audio Hardware

When designing a WFS system there are a number of issues that first need to be addressed. For example it is not a trivial matter to have a system consisting of 128 individual audio output channels as this is obviously reflected right down the signal

chain, requiring large numbers of converters, amplifier and loudspeakers accordingly.

The control PCs in this case are each fitted with 1 or 2 RME HDSPe MADI soundcards each of which is capable of outputting 64 channels of audio in MADI (*Multi-channel Audio Digital Interface*) format. Each MADI signal is then fed via coax into two daisy-chained RME M-32DA converters which each convert 32 channels of the MADI signal into 32 audio outputs (as shown in Fig. 3.25). These signals in-turn feed Cloud CX A850 8-channel power amplifiers. The loudspeaker cables (each one carrying 8 channels) are fed through to the loudspeakers. The overall system level is set by calibrating the amplifier gain using a dummy load resistance of 10Ω on each channel and setting the RMS voltage with a calibration signal.

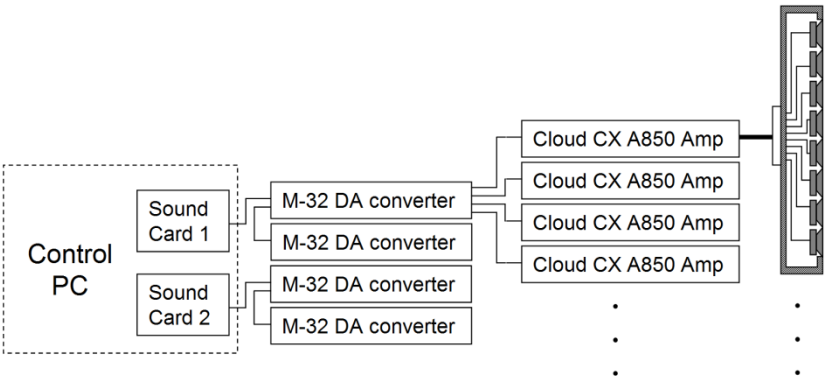


Fig. 3.25 Hardware diagram for WFS installation

3.6.2 Semi-portable system

As mentioned previously, the semi-portable system is used primarily in the fully anechoic chamber which constitutes a challenging installation environment for a WFS system, not least for the mounting of the loudspeakers, as there is no floor on which to place any stands etc. There are however several mount points distributed throughout the room allowing for a framework of poles to be erected, on which the loudspeakers could be mounted. It was also important to design this framework mounting for the loudspeaker arrays carefully such that it offers sufficient rigidity, flexibility and allows for the loudspeakers to be mounted as closely as possible to

each other to increase the spatial aliasing frequency. This design allows the speaker separation and height to be changed easily; also, several bars can be mounted on each pair of stands to allow for planar WFS testing if required.

Fig. 3.26 Semi-portable WFS installed in the anechoic chamber

Although the anechoic chamber offers the theoretically optimal acoustical environment for WFS, such rooms rarely offer the cabling for the required number of channels of audio input, thus for the system implemented here, the cables were fed around the outer and inner doors such that there was a small opening. For most test scenarios this is not too problematic as it only affects listeners who are in line of sight with the opening. The motivation for a semi-permanent system in the anechoic chamber was to enable testing of a system in a truly reflection free environment rather than to exploit the extremely low background noise levels offered by the chamber, thus the small gap made by the door is not problematic and indeed can be bridged with additional absorbent material.

3.6.3 Portable system

For the portable system, the loudspeaker system was designed to be modular in nature. Lightweight 6mm plywood enclosures were constructed that each contained eight loudspeakers. The enclosures were designed to be as small as possible for obvious practical reasons, as each driver has its own sealed enclosure this had no bearing on the overall frequency response or radiation characteristics of the system. This system utilises the same hardware architecture as the semi-portable system,

however this system only supports 64 channels. The portable system is shown in Fig. 3.27a.

a) b)

Fig. 3.27 a) Portable WFS installed in a typical classroom environment, b) Audio amplifiers and converters in the listening room void.

This WFS system is generally used for tradeshows, demonstrations and public engagement activities and can be setup in an hour, it is also useful for determining the effects of WFS propagation in rooms with different acoustic properties.

3.6.4 Listening Room Installation

As with the portable system, for the permanent installation in the listening room, the loudspeakers were placed in boxes of eight. For the installation in the listening room these were clamped directly onto the supporting frame, which was constructed from *48.3mm* steel pipe. Fourteen enclosures were arranged in a rectangle four enclosures deep and three wide, using 112 loudspeakers in total (as shown in Fig. 3.28). The control PC, audio digital-to-analogue conversion hardware and amplification are all located in the void around the side of the room as shown in Fig. 3.27b.

The design brief required that the system could be used at a range of different listening heights, from seated to standing, and should not compromise the acoustics of the room for other uses. In response to this the innovative solution of suspending the loudspeaker on a winch system was adopted. This was straightforward because

the room had been constructed with a concrete slab ceiling for acoustic isolation purposes and a flying system of the type most commonly seen in theatre lighting grid systems was designed and installed by a certified contractor according to LOLER regulations. The steel pipe was supported at four hanging points which were diverted by a system of pulleys back to a single ratcheted hand winch. The loudspeakers can be positioned at any height from $0.7m$ upwards and winched to within a few inches of the ceiling so that they are roughly flush with the existing acoustic treatment and both visually and acoustically unobtrusive. Cabling was routed around the steel ring and slung across to a strain relief bar (located so as to avoid the cable's weight pulling the flying frame askew) and finally through glands on a plate on the wall furthest from the door. Photos and a plan view of the system are shown in Fig. 3.28 and Fig. 3.29 respectively.

Fig. 3.28 Loudspeakers suspended on winch system in listening room

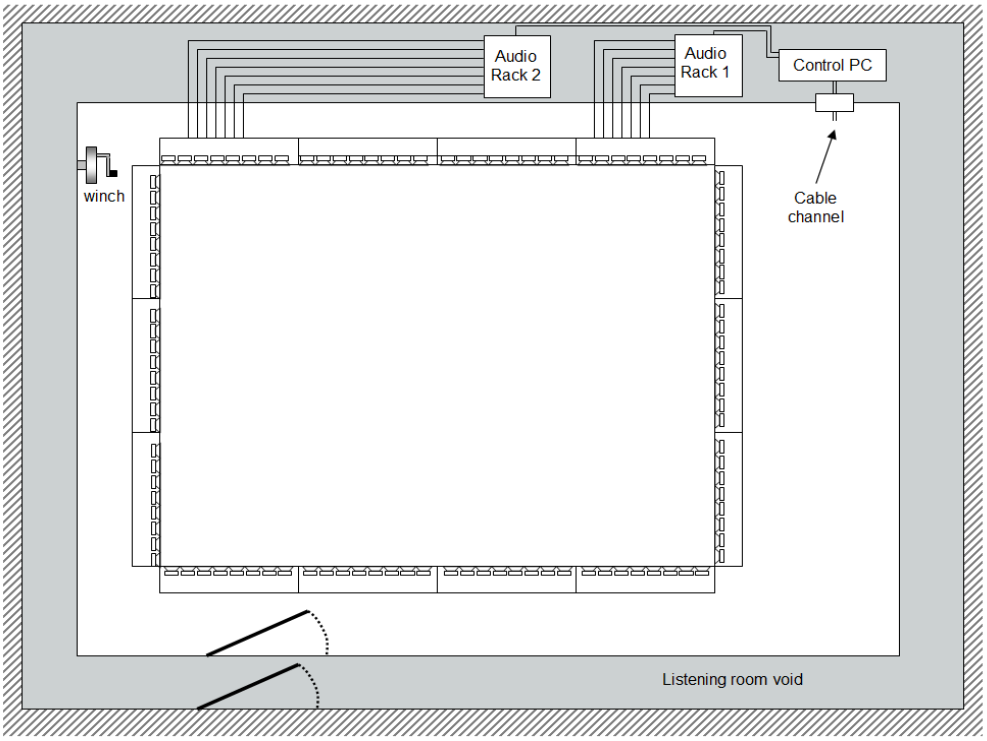


Fig. 3.29 Plan view of listening room with WFS install

The room was tested before and after the installation of the system and in both cases it conforms to the specifications set out in the standards ITU-R BS 1116-11, BS 6840-13 and IEC 268-132.

3.6.5 Octave Installation

The installation in the Octave is very similar in construction to that of the listening room. The arrays have exactly the same construction but the positioning of the arrays is at a fixed height. Due to the restrictions imposed by the rear projectors on the Octave it was neither possible to have the loudspeakers in font or behind the screens, so the solution was to place the arrays above the screens pointing at an angle to the listening area. Although this will reduce the vertical localisation accuracy, the azimuth localisation should remain stable. The problems associated with elevated loudspeaker arrays is an open research question and is briefly covered in section 6.5, more information on elevated sources in WFS can be found in (Lopez et al., 2010). The system installed in the Octave is shown in Fig. 3.30.

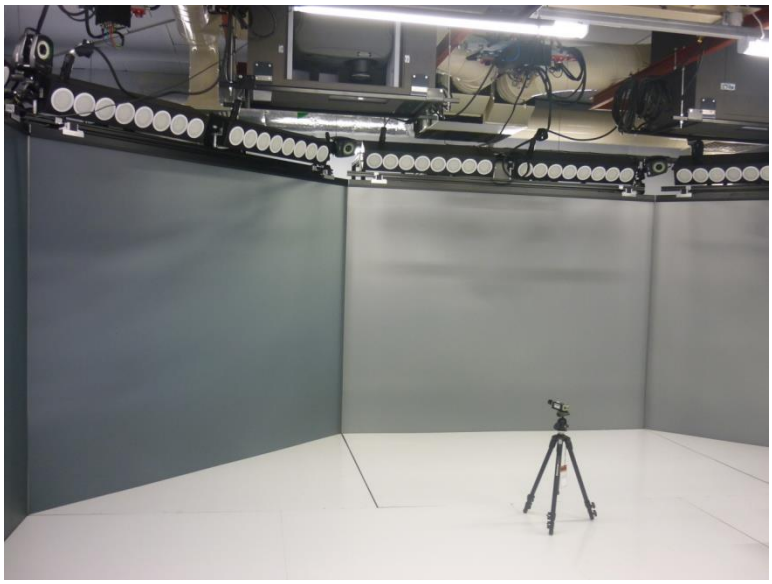


Fig. 3.30 128 channel WFS system installed in the Octave virtual reality lab at the University of Salford

The bespoke WFS rendering software has a UDP connection from the 3D visual rendering engine such that the positions of the audio sources can be adjusted dynamically to match those of the visual objects. This is particularly important for 3D rendering with audio so there is no incongruity in the audio and visual localisation cues and to prevent any problems associated with audio visual parallax error (Melchior et al., 2003; Springer et al., 2006), the UDP connection also enables the use of the tracking system with the audio system for optimised WFS rendering as suggested by (Melchior et al., 2008).

3.6.6 Software Development

Different versions of the control software have been developed that serve different practical purposes. The first implementation is a .dll library written in C++ that can be accessed in real-time by other pieces of software allowing the positioning and rendering of sources in real-time with UDP and TCP/IP links to allow input streams from other applications. Additionally a MATLAB® implementation has been written allowing for the testing and characterising of new driving functions and rendering methods as described in this thesis. The MATLAB® implementation is not real-time but does facilitate multiple sources and also accommodates moving sources. Each sound scene needs to be rendered offline in advance and rendered later.

3.6.6.1 Real-time C++ implementation

The controlling software for the WFS system has been written and visual C++ and has been complied to .dll format such that it can be accessed via a C# program that allows the arbitrary positioning of primary and secondary sources based on the XNA framework. The C++ part of the software takes advantage of the Windows Audio Session API (WASAPI), first introduced in Windows Vista which allows for simpler communication directly to the buffers on audio devices such that a desired output can be written straight to any audio channel of a soundcard. The C# side of the software works by using the positional data from the source and receiver, it sends a link object containing the corresponding spectral amplitude and delay from each source to each loudspeaker which are changed every time the virtual source is moved.

3.6.6.2 MATLAB® Implementation

For the ease of subjective tests an implementation of the software was written in MATLAB® such that the files could be rendered beforehand and played back on the request of the participant. The advantage of this technique is that the source files can be verified analytically insuring a correct auralisation for the test itself, the downside is that it does not work in real time. It has however been programmed in the frequency domain and time domain so it allows for moving virtual sources and there is no limit to the number of virtual sources that can be included in the sound scene.

3.6.6.3 Verification of the software

Before the setup could be used for research it had to be verified that it is correctly recreating acoustic wave fronts. For this to be confirmed a simple test has been done on the software to check the syncing between each output channel from the soundcard, if there is a syncing problem between the audio outputs, it will not be possible to recreate an acoustic wave front, as the small differences in phase between each loudspeaker are the key to successful sound field reconstruction.

A Gaussian pulse was put through the system with equal amplitude and phase on each channel, the ADAT outputs from the soundcard were recorded directly to

eliminate any other sources of delay. Results demonstrated that there is a delay of several samples between output channels. To check that it was in fact the output from the soundcard that was producing the error rather than the latency of the measurement system, the outputs from the soundcard were fed into an oscilloscope and the waveforms studied, this revealed that there was a delay with respect to the channels such that it was a problem at the output side. The measurement software was also checked by splitting the sinusoidal output from a signal generator into 8 identical signals and recording all 8 channels. The sine waves matched up perfectly meaning that the WFS software was producing the error.

The delay was caused by errors in WASAPI which later received a Windows® service pack update that fixed the problem a year later. The software sync between channels was again checked and revealed that a maximum of just 4 samples delay could be seen between channels which was due to the time it takes for the software to loop through all of the channels and release the audio buffers to the soundcard one at a time, the error was therefore corrected by determining the time it takes for the loop to complete the buffer release. This was done using clock function in the WASAPI used in the C++ part of the software, resulting in zero delay between channels.

3.6.7 System calibration

The system has been calibrated using a multi-meter on the output terminals of the power amplifiers. Given a sinusoidal output from each channel with the same amplitude, the voltage drop across a 10Ω resistor for each channel of the amplifier was adjusted to be the same. This technique assumes that the sensitivity of each of the loudspeakers is equal such that the output sound pressure from each one is the same. In order to verify this, the entire WFS system would have to be moved into the anechoic chamber and the acoustic output measured from each driver. It was decided therefore that the error in the relative loudspeaker sensitivities will be smaller than those introduced by the errors introduced by measuring the SPL from each driver separately such as the positioning of the microphone so the technique of calibrating the output level of the amplifiers was adopted.

3.6.8 System verification

The WFS system had been verified in several ways. Analytically using the MATLAB® implementation of the software, and also using multi-trace impulse response measurements with a linear microphone array.

3.6.8.1 MATLAB® implementation verification

Various scenarios can be rendered in MATLAB® and the corresponding wave field analysed before rendering to verify that the system, or at least the software, is working correctly.

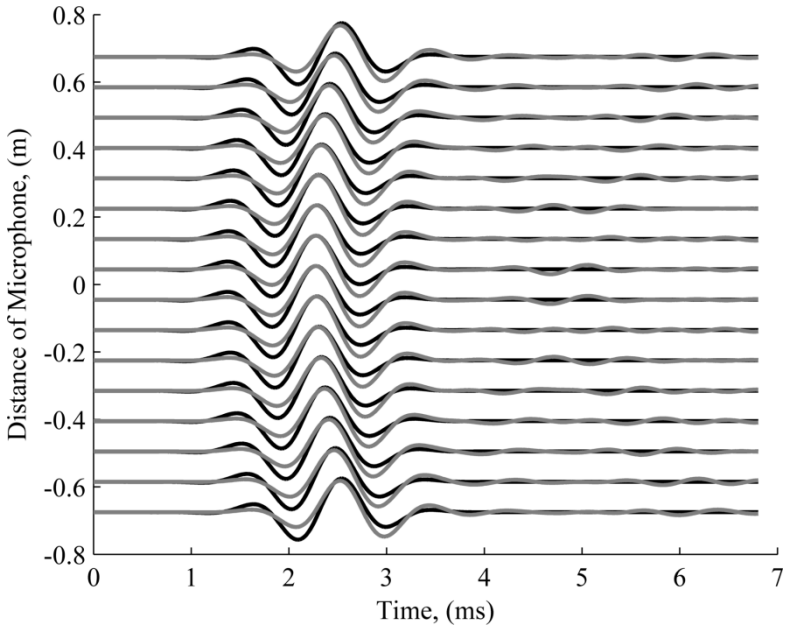


Fig. 3.31 16-channel multi-trace impulse response for a real source (black trace) and a WFS source (grey trace) with a Gaussian pulse 2.5m back from the centre of a linear loudspeaker array

Fig. 3.31 shows the impulse response at 16 locations recorded for both a real source and a WFS point source. The simulation demonstrates that the MATLAB® implementation aligns well with the analytical solution of a point source. There is some ripple evident in the impulse response after the main pulse which is due to echo artefacts from the WFS array as a consequence of truncation errors. Truncation errors are also responsible for the small delay error in the impulse response for the edge receivers causing some small errors in wave front curvature.

3.6.8.1 Multi-trace impulse testing

To objectively verify that the full system was working, a linear array of 16 microphones was set up in the centre of the listening room and various source types were output from the WFS system with a number of test signals to see if the rendered wave field was as intended. The method of checking this is the multi-impulse test method which lines up the time histories from each microphone in the array to reveal how the wave field changes with time over the dimensions of microphone array.

Fig. 3.32 shows the results from these experiments showing both the measured and analytical data. The results not only illustrate that the system successfully recreates the correct wave fronts but also demonstrates how the WFS driving functions match up well with the analytical solutions.

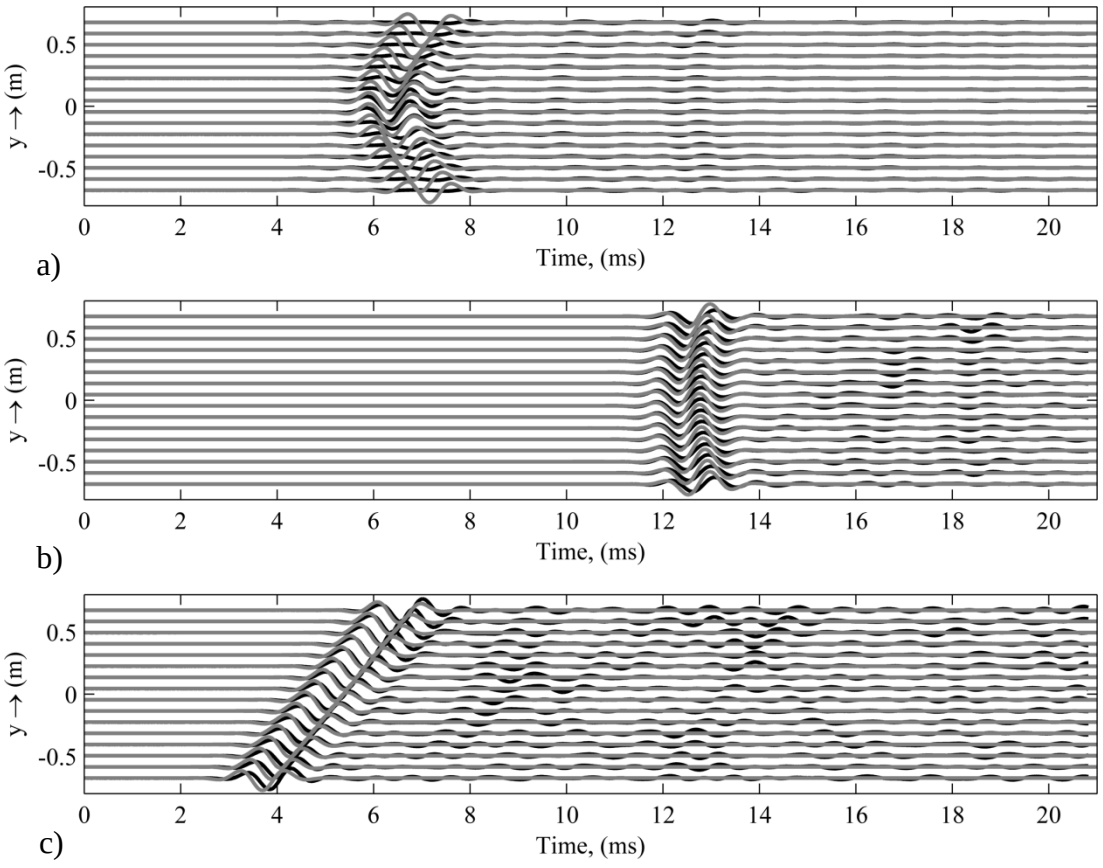


Fig. 3.32 Multi-trace impulse response measurements in the listening room, with a) a focused point source, b) a normal point source and c) a plane wave at 45˚. The black trace shows the measured data and the grey trace the analytical solution.

Results from the multi-trace impulse tests show that the wave fronts are being rendered accurately for the three main source types (plane wave, point source and focused point source). There are some minor differences in the wave front curvature for scenario a) and b) which are a consequence of array truncation. The room effects are also seen as a rippling in the impulse responses after the main pulse, these effects will be dealt with in detail in chapter 6. For the focused source case in a) there are some pre-echo artefacts present (i.e. contributions arriving before the intended wave front), these are specific to focused sources and will be described in more detail in chapter 4.

3.7 Areas of research in WFS literature

Research in WFS has taken many directions to date and the literature covers a broad range of applications. The following is a brief description of some of the major research areas in WFS.

3.7.1 Room compensation

The early seminal papers provided a strong theoretical foundation for WFS and demonstrated in particular the possibility of using the idea for acoustic control in rooms (Berkhout et al., 1993; Berkhout, 1988). This has remained one of the strongest areas of WFS research and much attention is given to the use of WFS for room compensation (Spors et al., 2007) with a number of techniques currently being used (e.g. Corteel and Nicol, 2003; López et al., 2005; Petrausch et al., 2005; Spors et al., 2004), similar principles can also be employed to use WFS for variable acoustics applications (Sonke et al., 1998) and for active noise control (Kuntz and Rabenstein, 2004; Peretti et al., 2007; Spors and Buchner, 2004). Wave field synthesis is ideal for such applications as it can recreate acoustic wave fronts of almost any shape and arriving from any direction, even sources within the room can be produced/reproduced effectively. This means that if the field in the room can be measured accurately enough and to a high enough resolution, the WFS system can radiate the opposite wave field and can cancel it out. This raises a number of issues and is certainly not an easy task, especially as the sound field within a room is 3-

dimensional whereas most WFS systems are just 2-dimensional. There are also problems with the frequency limitations of WFS. Spatial aliasing at higher frequencies means that an accurate spatial wave field can no longer be recreated thus it will not cancel effectively with the original field, possibly compounding the original problem, there is also the problem of time variance in a room especially when it is in use which is hard to correct for. Research in this area has however lead to what has now been termed Adaptive Wave Field Synthesis (Gauthier and Berry, 2006) which describes the combination of a standard wave field synthesis system with room compensation built in to the driving functions to correct for the unwanted reflections (Gauthier and Berry, 2007) in the room which damage the reproduced wave field, this is an important area of research because the definition of the WFS driving functions assumes free field source free conditions which is certainly not the case in a real room as each reflection essentially acts as a new source. Ideally therefore WFS reproduction should be done in an anechoic space or the reflections corrected for using active techniques.

3.7.2 Sound reinforcement

Another of the main uses of WFS is for sound reinforcement and there has correspondingly been much written on this topic (e.g. Verheijen, 1998; de Vries, 1996; de Vries et al., 1994). WFS provides accurate rendering in the spatial, temporal and frequency domain therefore to use it for sound reinforcement is ideal as it eliminates the separation that can occur with the audience seeing the performer in one position but hearing the audio coming from loudspeakers in various discrete locations. WFS can help alleviate this problem by amplifying the sound and also retaining the original spatial information. For this to be successful, careful attention has to be given to the microphone arrangement recording the performers so the signals can be accurately reproduced spatially. One can adopt an array based approach as (Snow, 1953) or else use close miking (Theile, 2004) or other more complicated schemes (Hulsebos et al., 2002).

3.7.3 Virtual reality applications

Possibly the most ubiquitous application of WFS is in virtual reality, cinema and sound scene creation (Sporer, 2004; Strauss et al., 2004). WFS allows very accurate spatial reproduction of a sound field and thus retains all of the spatial hearing cues that humans use to localise sound, it therefore provides excellent source localisation and a realistic impression, immersion and presence. There is consequently great scope to use WFS for soundscape rendering (Geier et al., 2008) and analysis as a recorded sound scene can be recreated almost perfectly so when combined with multiple visual displays (Springer et al., 2006) a very realistic impression of a given place/scene can be realised. Coupled with the field of virtual reality there is also a widening body of research into videoconferencing which attempt to increase the realism in the communication between two geographically separated groups. Wave field synthesis is seen as a good solution to the problem of matching up the audio signals with the visual information as conflicting auditory and visual cues reduces the realism and can cause confusion in the communication (De Bruijn, 2004). There is however research suggesting that WFS is only better for these applications when combined with a 3-dimensional projection (Melchior et al., 2003). The 2-dimensional visual display with a 3-dimensional audio signal (rendered using WFS) actually provides conflicting cues and with a parallax error between the two source positions, in one study it was found by (De Bruijn and Boone, 2003) that the subjects localised better with multi-channel stereophony rather than WFS with a 2-dimensional display; a 3-dimensional display however yielded better results with WFS.

3.7.4 Using flat panel loudspeakers

The concept of using WFS with multiple visual displays has driven research into the use of flat panel (distributed mode/multi actuator) loudspeakers for the rendering of WFS (Boone and De Bruijn, 2000). This proves to be a troublesome area as there are many corrections that one has to do for both the response of the individual panels (Spors et al., 2005) and also for reflections caused by the different panels (Corteel et al., 2004) as they are significantly larger than moving coil loudspeakers where this is not a problem, consequently there is a growing body of research in this area. This is

one of the key areas for improvement in WFS as it will open up the scope for many more applications and places where systems could be installed. At the moment, WFS systems consist of a lot of hardware and the loudspeakers themselves can be heavy, making installation not only problematic but also unsightly. The use of flat panel loudspeakers would reduce these problems and provide greater ease of installation, the loudspeakers could also have dual functionality as projection screens. The use of flat panel loudspeakers does have a subjective effect on the recreated sound field however and work has been done on the perceptual evaluation of this (Corteel et al., 2007) to ensure that the benefits outweigh the shortcomings.

3.7.5 Perceptual aspects

There are many different aspects to perception of WFS systems as the sound recreation takes place in the temporal, frequency and spatial domain, providing many variables for analysis, consequently there has been much work done to analyse these perceptual characteristics. Within this area there is a need to compare the perception of WFS with other existing spatial audio systems such as stereophony (Wittek, 2009), Ambisonics (Daniel et al., 2003) and VBAP (Pulkki, 1997) etc. Work on the comparison between these methods has led to the conclusion that a combination of different rendering schemes can be profitable, combining the merits of stereophony, Ambisonics and VBAP to result in various hybrid approaches (Nicol and Emerit, 1999; Wittek et al., 2007a) which minimise processing and maximise localisation accuracy.

3.7.6 Different source types

A key area in the perception of WFS rendering is that of moving virtual sources (Franck et al., 2007). The theory of wave field synthesis inherently takes into account a moving source and a good subjective impression can be achieved by rendering a series of stationary sources at advancing locations using an iterative approach. Further research by (Ahrens and Spors, 2008a) however revealed that rendering a series of stationary sources does not accurately represent the sound field produced by a real moving source. They used an analytical approach and demonstrated a reproduced sound field of a moving sources with greater realism

using an analytical driving function which explicitly models the Doppler Effect. It seems however that whilst the wave field looks more accurate with Ahrens and Spors' technique, the perception of the traditional method of rendering moving sources is perfectly adequate providing the sources are not moving at speeds greater than $40ms^{-1}$.

The traditional WFS literature allows for sources of a plane wave or spherical wave type but not much attention was paid to the topic of rendering directional sources, however this is becoming a more prominent area of research as there is a greater need for increased accuracy in WFS rendering. It is possible to derive driving functions from first principles that take into account a desired directivity (Ahrens and Spors, 2007; Baalman, 2008; Corteel, 2007), Ahrens also presented an approach to deriving a focused source with a given directivity (Ahrens and Spors, 2008b) which would allow a user to walk around the focussed source and experience a more realistic directivity pattern rather than that of a point source.

3.7.7 Analysis and reduction of errors in WFS reproduction

All of the above uses of WFS are subject to the inherent errors of WFS such as discretisation, spatial aliasing and truncation effects/edge effects. Truncation errors are caused by the finite length of the arrays used for the reproduction, causing diffraction errors at the end of the arrays. Edge errors are also caused in the corners where two loudspeaker arrays meet. Whilst these can be minimised by adopting various filtering schemes, there are consequent limits on the available applications of WFS. Analyses of the errors inherent in WFS reproduction due to the size of the arrays being used to reproduce the signals and the discretisation of the array can be found in (Spors and Rabenstien, 2006; Start, 1997; Verheijen, 1998)

Much research is still to be done in the area of WFS, looking at the optimisation of the WFS reproduction and coding, subjective effects resulting from simplifications that are made, evaluation and prediction of wave fields, novel applications of WFS, perceptual effects and dual modality investigation when using WFS with 2-dimensional or 3-dimensional displays. Consequently there is an ever increasing body of research into this exciting subject of wave field synthesis. This thesis aims at

the characterisation and improvement of focused sources in WFS, which is an under-researched area so far despite the many novel applications and interesting properties of such sources.

3.8 Summary

This chapter has presented the background theory of the wave field synthesis sound field reconstruction technique. A comprehensive theoretical background has been presented, including the foundational mathematic constructs and a detailed derivation of the important driving functions used throughout this thesis. A brief analysis of some of the reproduction artefacts of WFS has also been described in this chapter including, amplitude errors, spatial sampling artefacts and truncation errors. Many of these errors are common to both focused and non-focused sources although the manifestation of the artefacts will be different for focused sources. This chapter has described only non-focused sources, the next chapter deals specifically with the objective analysis of WFS focused sources and the errors there inherent.

Having laid the theoretical background of WFS, the chapter then described the design, installation and calibration of four WFS systems that have been installed at the University of Salford that have been used in the work described in this thesis. Also covered is a generic literature review, presenting an overview of the main areas of research in WFS since its conception in the late 1980's.

4 Objective Analysis of WFS Focused Sources

4.1 Introduction

The ability of sound reproduction techniques based on physical principles and a sound field description such as wave field synthesis to recreate acoustic wave fronts of an arbitrary shape allows for the reproduction of so-called focused sources. Focused sources are in fact not really sources but rather they describe the focusing of acoustic energy to a point in space. As the point does not behave like an acoustic sink, the acoustic energy peaks at the 'focal point' and then re-radiates from there, giving the impression of a source at that point. This is a remarkable attribute of WFS systems and allows for several interesting applications. For example, focused sources can be used for 3D virtual reality systems such that the depth of the acoustic source matches more perfectly the depth of the visual source being generated/projected (Boone et al., 1996). Focused sources have also been showed to be useful for creating virtual loudspeaker signals in hybrid systems (Menzel et al., 2005).

In the strictest sense, focused sources in WFS cannot be fully described by the Kirchhoff-Helmholtz equation demonstrated in the previous chapter. The Kirchhoff-Helmholtz equation states that the volume in question should be source free whereas a focused source expressly describes a volume with a source present. However this is not a real source but rather the appearance of a source thus the method of derivation used for normal WFS sources also stands in this case.

This chapter begins with a brief overview of sound focusing techniques and describes some of the properties of the produced focused wave fields. The chapter then goes on to describe the driving functions for the main focused source models used for practical WFS implementations including monopoles, dipoles and pistonic sources. Following this is a description of the secondary source selection criteria specific to the reproduction of focused sources and the necessary correction needed for symmetrical rendering and directed sources. The primary attributes of focused sources in WFS are then described including amplitude errors, truncation and spatial aliasing errors, causality and pre-echo artefacts, focal spot size and phase response. A model is also derived to predict the position error in focused source as a result of the focal shift phenomena caused by diffraction effects at low frequencies. A full description of spatial aliasing artefacts, including an analytical expression for the position dependant spatial aliasing condition for focused sources and the definition of the aliasing-free zone in the near vicinity of the focal point is also presented. The chapter concludes with some discussions on the rendering of moving sources and sources using large arrays.

4.2 An overview of acoustic focusing techniques

This section briefly overviews some of the major techniques for acoustic focusing that this chapter will make reference to and covers some of the main applications for the techniques mentioned.

4.2.1 Time reversal acoustics

Time reversal acoustics essentially exploits the theory of acoustic reciprocity, that is to say the characteristic of the acoustic wave equation that states that the transfer

function between source and receiver is the same as from receiver to source. A time reversal mirror (Fink, 1992, 1999), often used in ultrasonic applications, exploits this phenomenon, by placing a line of transducers that can act in both receive and emit mode in front of an acoustic sources, the field is recorded by the transducers in receive mode. These recorded signals are time reversed (i.e. played backwards) and resulting in the sound energy focusing on the original source position (or at least at a point the equivalent distance from the array if the array is moved).

4.2.2 Phase conjugation

Phase conjugation is a technique often applied in the field of underwater acoustics. It shares many similarities with time reversal acoustics, but the underlying method is slightly different. Phase conjugation exploits the fact that in the frequency domain a conjugated wave propagates in the opposite direction to the original i.e. if $p = e^{-jkR}$ is a propagating point source, then $p = e^{jkR}$ describes an acoustic sink. Thus if the conjugate form of a recorded signal is transmitted, the result will be a time reversed version of the original signal (i.e. one that converges on to the original source location)

4.2.3 Focused transducers

Much of the work in focused acoustic fields has been done using focused transducers and much of the same theory can be applied to WFS focused sources as described in this chapter. A focused transducer is often a curved piezoelectric transducer that due to its curved geometry will radiate focused acoustic energy on to a point at a position defined by the surface geometry. Most often these devices will have a hemi-spherical or semi-circular geometry and will therefore focus energy on to a point at the centre of the sphere or circle. This is probably the simplest method of acoustic focusing and is used to increase the radiated power at a given receiver location. As the radius of curvature decreases the solution for the on-axis pressure of these devices tends towards that of a piston source.

4.2.4 Physical acoustic lens'

Physical acoustic lens' (Hurwitz, 1977; Miller, 1983) have many different designs and can be made out of a variety of acoustic meta-materials, often made up of arrays of acoustic scatterers known collectively as a sonic crystal. Whatever the material or design of these devices, the aim is to selectively delay some of the acoustic energy such that it all arrives at the target location (focal point) at the same point in time. This delay means that the device has to exhibit different wave propagation speeds or the acoustic paths have to further to achieve the necessary delays, this can be thought of as a variable acoustic refractive index (like an optical lens).

4.2.5 Applications of acoustic focusing

There are many applications of acoustic focusing (Fink, 2006; Fink et al., 2004) utilising many different methods of focusing acoustic energy. Many of the major applications are in ultrasonic frequencies and are found in the field of medicine where focusing has been used for the destruction of tumours and kidney stones (Thomas et al., 1996), brain therapy (Clement et al., 2000). Also predominantly in ultrasonic frequencies, acoustic focusing has been used for non-destructive testing for defects in metals (Zumpano and Meo, 2007). Ocean acoustics is another area where acoustic focusing is often employed in this case it is predominantly for communication and transmission purposes (Kuperman et al., 1998). Mine detection (Sutin and Johnson, 2005), acoustic microscopes (Lemons and Quate, 1974), and even ink printing (Elrod et al., 1988) are yet more examples of acoustic focusing applications.

Most of the aforementioned applications use acoustic energy at ultrasonic frequencies, however, recently there has been more attention paid to acoustic focusing of audio frequencies, specifically in sound reproduction which is possible using sound field synthesis systems such as wave field synthesis (Jensen, 1997; Spors et al., 2009), Higher Order Ambisonics (Ahrens and Spors, 2008c; Jensen, 1997; Spors et al., 2009) and the Spectral Division Method (Spors and Ahrens, 2010c). Methods of acoustic time reversal have also been applied in room acoustics e.g. (Yon et al., 2003a, 2003b).

The following section presents the theory of acoustic focusing using the WFS method and will derive the important driving function used.

4.1 WFS driving functions for focused sources

Driving functions for focused sources in WFS can be derived in a manner similar to that of an non-focused sources but using a time reversal/phase conjugation approach such that the virtual source models become acoustic sinks rather than acoustic sources. An acoustic sink describes a point in space which absorbs all acoustic energy incident on it. For WFS rendering, the energy is not absorbed by the sink, but rather the acoustic energy is focused on to this point and then re-radiates giving the impression of a source located at the sink's position. As shown in Fig. 4.1.

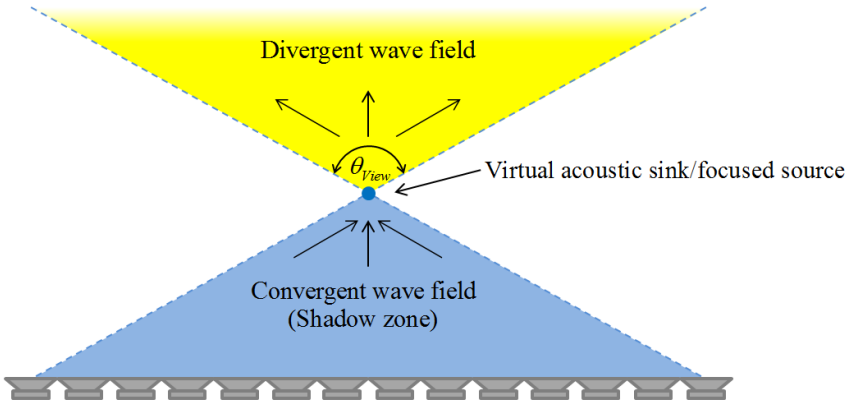


Fig. 4.1 Diagram of the theoretical view angle for focused source reproduction

This section briefly outlines the derivation of the driving functions for the major focused source types commonly used in WFS rendering. The geometry used for the derivation of the WFS driving function is shown in Fig. 4.2.

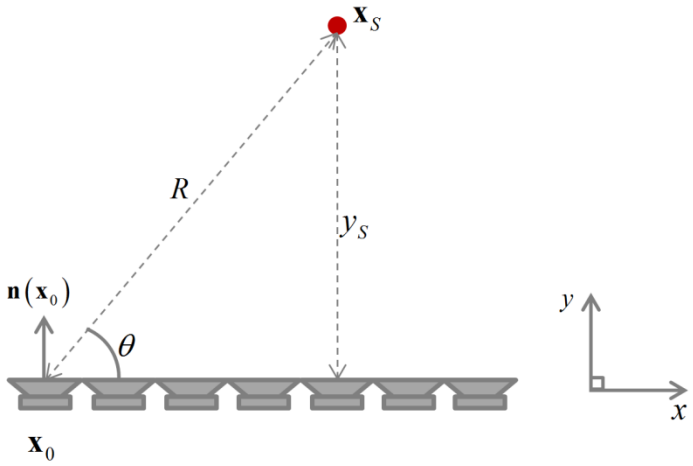


Fig. 4.2. Geometry used for the derivation of WFS focused sources driving functions

4.1.1 Monopole sources

For 3D WFS the time reversed equation of a point source would be sensible choice for the virtual source model. This describes the pressure at $\mathbf{x}_0$ as a result of a point sink at $\mathbf{x}_S$ and can be given as:

$$S_{Sink,3D}\left(\mathbf{x}_0,\omega\right)=\frac{e^{j\left(k\left|\mathbf{x}_0-\mathbf{x}_S\right|\right)}}{4\pi\left|\mathbf{x}_0-\mathbf{x}_S\right|} \tag{4.1}$$

For 2D WFS however, a more suitable source model for a monopole virtual source would be an acoustic line sink, the source function of which is given by (Williams, 1999) as *(4.2)*.

$$S_{Sink,2D}\left(\mathbf{x}_0,\omega\right)=\frac{j}{4}H_0^{(1)}\left(k\left|\mathbf{x}_0-\mathbf{x}_S\right|\right) \tag{4.2}$$

where $H_0^{(1)}(..)$ is the zeroth order Hankel function of the first kind. The line sink is used as the source model therefore for 2.5D WFS a spectral correction must be applied as was introduced in the previous chapter. This equation can be approximated using the large argument/asymptotic expansion of the Hankel function *(3.48)* (valid for $k\left|\mathbf{x}_0-\mathbf{x}_S\right|\gg 1$) as:

$$
\begin{aligned}
S_{Sink,2D}\left(\mathbf{x}_0,\omega\right) &\simeq \frac{j}{4}\sqrt{\frac{2}{\pi k\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\; e^{j\left(k\left|\mathbf{x}_0-\mathbf{x}_S\right|-\pi/4\right)} \\
&\simeq \frac{j}{4}\sqrt{\frac{2}{\pi k\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\; e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}\sqrt{\frac{1}{j}} \\
&\simeq \sqrt{\frac{j}{k}}\sqrt{2\pi\left|\mathbf{x}_0-\mathbf{x}_S\right|}\;\frac{e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}}{4\pi\left|\mathbf{x}_0-\mathbf{x}_S\right|}
\end{aligned}
\qquad (4.3)
$$

This source model, however is not suitable for 2.5D synthesis due to the spectral differences introduced when radiating 2D sound fields using point sources. From *(4.3)* and *(4.1)* it can be seen that the free field Green's function of line sink is the same as the 3D Green's function of a point sink but with a spectral correction of $\sqrt{j/k}$ and an additional amplitude weighting. The inverse of this spectral correction can be applied to the driving function for a time-reversed monopole line sink such that the source model behaves like a line sink but with the spectral characteristics of a point sink, and thus is a desirable source model for 2.5D synthesis. The spectral correction therefore takes the form of $\sqrt{k/j}$. The virtual source model with frequency correction applied can then be written as:

$$
S\left(\mathbf{x},\omega\right)=\frac{1}{4}\sqrt{jk}\;H_0^{(1)}\left(k\left|\mathbf{x}\;-\mathbf{x}_S\right|\right) \qquad (4.4)
$$

Where the position of the virtual source is given by the vector $\mathbf{x}_S$. Once again the 2D driving function can be expressed as:

$$
D_{2D}\left(\mathbf{x}_0,\omega\right)=-\frac{1}{2\pi}\nabla S\left(\mathbf{x}_0,\omega\right).\mathbf{n}\left(\mathbf{x}_0\right) \qquad (4.5)
$$

Inserting *(4.4)* into *(4.5)* gives the driving function as:

$$
D_{2D}\left(\mathbf{x}_0,\omega\right)=-\frac{1}{2\pi}\frac{\partial}{\partial R}\left(\frac{1}{4}\sqrt{jk}H_0^{(1)}\left(kR\right)\right)\hat{\mathbf{R}}\cdot\mathbf{n}\left(\mathbf{x}_0\right) \qquad (4.6)
$$

Which, noting that

$$\frac{\partial}{\partial Z} H_n^{(1)}(Z) = \frac{nH_n^{(1)}(Z)}{Z} - H_{n+1}^{(1)}(Z) \tag{4.7}$$

gives

$$D_{2D}(\mathbf{x}_0, \omega) = \frac{k}{8\pi} \frac{(\mathbf{x}_0 - \mathbf{x}_S)}{|\mathbf{x}_0 - \mathbf{x}_S|} \cdot \mathbf{n}(\mathbf{x}_0) H_1^{(1)}(kR) \tag{4.8}$$

So the 2.5D driving function is given by applying the amplitude correction from *(3.28)* (the spectral correction has already been applied) as:

$$D_{2.5D}(\mathbf{x}_0, \omega) = -\frac{k}{8\pi} \sqrt{2\pi |\mathbf{x}_{\text{Ref}} - \mathbf{x}_0|} \frac{(\mathbf{x}_0 - \mathbf{x}_S)}{|\mathbf{x}_0 - \mathbf{x}_S|} \cdot \mathbf{n}(\mathbf{x}_0) H_1^{(1)}(kR) \tag{4.9}$$

Where once again $\widehat{\mathbf{R}} = \mathbf{R} / R = (\mathbf{x}_0 - \mathbf{x}_S) / |\mathbf{x}_0 - \mathbf{x}_S|$, and $\mathbf{n}(\mathbf{x}_0)$ is the unit vector normal to the direction of the loudspeaker.

It is often considered desirable to separate the frequency dependent terms in WFS driving functions such that they can applied efficiently as a pre-filter, hence the far field approximation of the Hankel function in *(3.48)* can be applied to *(4.9)* resulting in the driving function of a focused monopole line source:

$$D_{2.5D}(\mathbf{x}_0, \omega) = -\frac{1}{4\pi} \sqrt{\frac{k}{j}} \sqrt{\frac{|\mathbf{x}_{\text{Ref}} - \mathbf{x}_0|}{|\mathbf{x}_0 - \mathbf{x}_S|}} \frac{(\mathbf{x}_0 - \mathbf{x}_S)}{|\mathbf{x}_0 - \mathbf{x}_S|} \cdot \mathbf{n}(\mathbf{x}_0) e^{\,j(k|\mathbf{x}_0 - \mathbf{x}_S|)} \tag{4.10}$$

This is the same as that given by (Spors et al., 2009) barring a normalisation factor. The resulting focused wave field produced by this driving function is shown in Fig. 4.3.

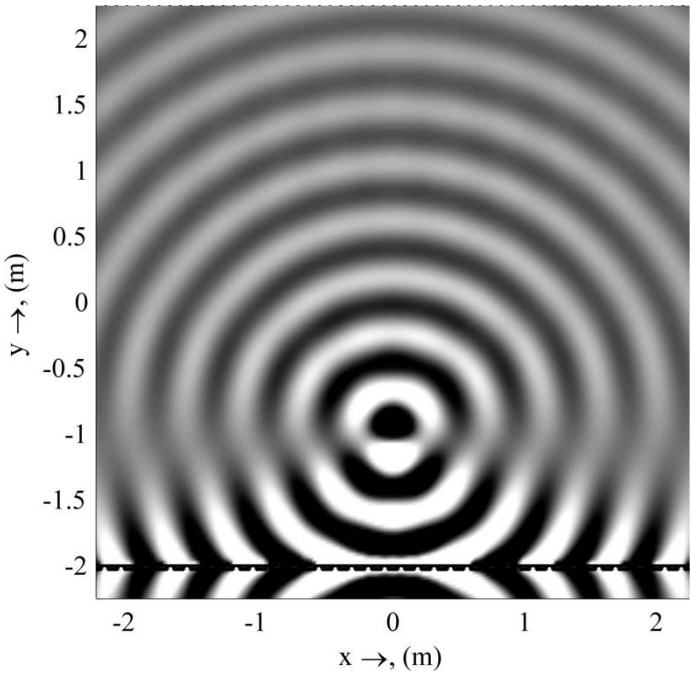


Fig. 4.3 Wave field of a focused line source in 2.5D WFS, $f = 800Hz$ at $\mathbf{x}_S = [0m, -1m]$ with $\Delta_{LS} = 0.125m$

4.1.2 Dipole focused sources

The source function for a focused acoustic dipole can be given by *(3.53)*, From this, the Green's function of a dipole line sink can be given as *(4.11)*, with the source geometry given by Fig. 4.2.

$$S(\mathbf{x}, \omega) = jk\pi \cos\theta H_0^{(1)}(kR) \tag{4.11}$$

Where

$$\cos\theta = \frac{-x_S}{|\mathbf{x}_0 - \mathbf{x}_S|} = \frac{-x_S}{R} \tag{4.12}$$

(4.11) becomes

$$S(\mathbf{x}, \omega) = -jk\pi \frac{x_S}{R} H_0^{(1)}(kR) \tag{4.13}$$

Thus, from *(4.5)* and *(4.13)*, the 2D driving function for a dipole line sink can be expressed as *(4.14)*

$$D_{2D}\left(\mathbf{x}_0,\omega\right)=-\frac{1}{2\pi}\widehat{\mathbf{R}}.\mathbf{n}\left(\mathbf{x}_0\right)\frac{\partial}{\partial R}\left(-jk\pi\frac{x_S}{R}H_0^{(1)}\left(kR\right)\right) \tag{4.14}$$

with

$$\frac{\partial}{\partial R}\left(-A(\omega)\,jk\pi\frac{x_S}{R}H_0^{(1)}\left(kR\right)\right)=\frac{j\pi kx_S\left(H_0^{(1)}\left(kR\right)+kRH_1^{(1)}\left(kR\right)\right)}{R^2} \tag{4.15}$$

which can be approximated with the far field approximation in *(3.48)* as:

$$\frac{\partial}{\partial R}S\left(\mathbf{x},\omega\right)\simeq\ \frac{1}{R^2}\,j\pi kx_S\sqrt{\frac{2}{\pi jkR}}\left(1-kR\right)e^{jkR} \tag{4.16}$$

Thus the 2D driving function for a line dipole sink is given is given as:

$$\begin{aligned}D_{2D}\left(\mathbf{x}_0,\omega\right)&=-\frac{1}{2\pi}\widehat{\mathbf{R}}\cdot\mathbf{n}\left(\mathbf{x}_0\right)\frac{1}{R^2}\,j\pi kx_S\sqrt{\frac{2}{\pi jkR}}\left(1-kR\right)e^{jkR}\\&=-\frac{1}{2\pi}\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\cdot\mathbf{n}\left(\mathbf{x}_0\right)\frac{1}{\left|\mathbf{x}_0-\mathbf{x}_S\right|^2}\,j\pi kx_S\sqrt{\frac{2}{\pi jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\times\\&\qquad\ldots\left(1-k\left|\mathbf{x}_0-\mathbf{x}_S\right|\right)e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}\\&=-\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|^2}\cdot\mathbf{n}\left(\mathbf{x}_0\right)\frac{x_S}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\sqrt{\frac{jk}{2\pi\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\times\\&\qquad\ldots\left(1-k\left|\mathbf{x}_0-\mathbf{x}_S\right|\right)e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}\\&=-\sqrt{jk}\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|^2}\cdot\mathbf{n}\left(\mathbf{x}_0\right)\frac{x_S}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\left(\frac{1-k\left|\mathbf{x}_0-\mathbf{x}_S\right|}{\sqrt{2\pi\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\right)e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}\end{aligned} \tag{4.17}$$

The necessary correction can also be applied to this equation resulting in the 2.5D driving function for a focused dipole line source from *(3.28)*.

$$D_{2.5D}\left(\mathbf{x}_0,\omega\right)=-\sqrt{\frac{\left|\mathbf{x}_{Ref}-\mathbf{x}_0\right|}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\frac{\left(\mathbf{x}_0-\mathbf{x}_S\right)}{\left|\mathbf{x}_0-\mathbf{x}_S\right|^2}\cdot\mathbf{n}\left(\mathbf{x}_0\right)\frac{x_S}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\left(1-k\left|\mathbf{x}_0-\mathbf{x}_S\right|\right)e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|} \tag{4.18}$$

The spectral correction for a dipole line source can be calculated from the large argument approximation of $H_0^{(1)}(kR)$ function and is $\sqrt{jk}$. So the spectrally corrected 2.5D driving function is therefore expressed as:

$$D_{2.5D}(\mathbf{x}_0,\omega) = -\sqrt{jk}\sqrt{\frac{|\mathbf{x}_{Ref}-\mathbf{x}_0|}{|\mathbf{x}_0-\mathbf{x}_S|}}\frac{(\mathbf{x}_0-\mathbf{x}_S)}{|\mathbf{x}_0-\mathbf{x}_S|^2}\cdot\mathbf{n}(\mathbf{x}_0)\frac{x_S}{|\mathbf{x}_0-\mathbf{x}_S|}\times \\ \ldots\left(1-k|\mathbf{x}_0-\mathbf{x}_S|\right)e^{jk|\mathbf{x}_0-\mathbf{x}_S|} \tag{4.19}$$

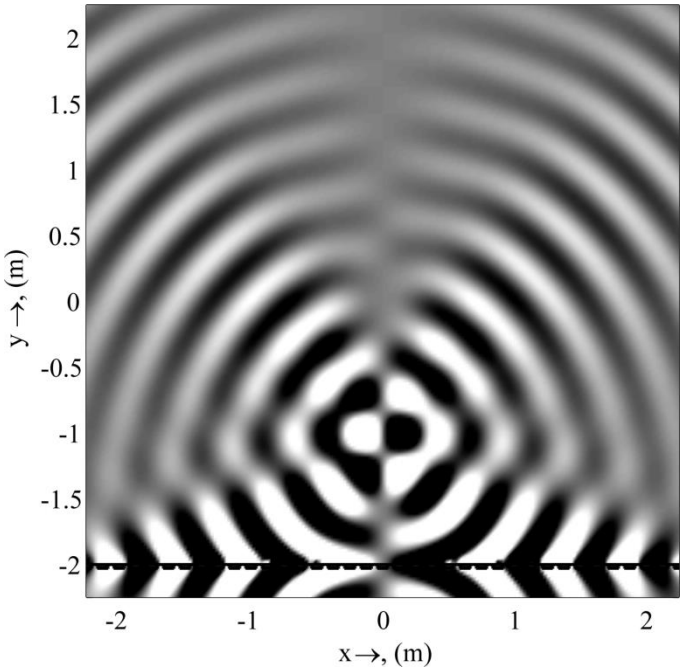


Fig. 4.4 Wave field of a focused dipole line source in 2.5D WFS, $f = 800Hz$ at $\mathbf{x}_S = [0m, -1m]$ with $\Delta_{LS} = 0.125m$

Note that due to the fact that focused sources are not real sources but rather the appearance of radiation as the sound passes through the focal point, the dipole source will be out of phase by π radians with respect to the desired pressure field, thus the driving function should be multiplied by minus one to correct for this phase shift.

The rendering of dipoles as WFS virtual sources offers the potential of generating sources with arbitrary directivity patterns and finite extent as any source directivity pattern can be created by the addition of multipole sources with different polarities and respective amplitudes (Williams, 1999). Of course this results in the need for

many different virtual sources which significantly increases the computational overhead of WFS systems, thus the commonly used techniques for creating more complex models is to use a spherical harmonic approach (Corteel, 2007).

4.1.3 Pistonic focused sources

Pistons are a useful source model for focused sources especially because they allow the rendering of a virtual loudspeaker in the listening space and therefore create the possibility of virtual surround loudspeakers (Boone and Horbach, 1999). A circular piston in an infinite baffle is often used as a mathematical model of a loudspeaker as it radiates much like a point source but with a frequency dependant directivity such that as $ka > 1$ the piston becomes increasingly directional as per a real loudspeaker in a sealed enclosure.

The far-field radiation of a circular piston sink in an infinite baffle can be given by the time reversed version of the equation of a pistonic source as (Kinsler et al., 2000a).

$$S_{piston}\left(\mathbf{x}_0,\omega\right) = A\left(\omega\right)\frac{jka^2 e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\left(\frac{2J_1\left(ka\sin\theta\right)}{ka\sin\theta}\right) \tag{4.20}$$

Where

$$\sin\theta = \frac{y_S}{\left|\mathbf{x}_0-\mathbf{x}_S\right|} \tag{4.21}$$

$$S_{piston}\left(\mathbf{x}_0,\omega\right) = A\left(\omega\right)\frac{jka^2 e^{jk\left|\mathbf{x}_0-\mathbf{x}_S\right|}}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\left(\frac{2J_1\left(\frac{ka\,y_S}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}\right)}{\frac{ka\,y_S}{\left|\mathbf{x}_0-\mathbf{x}_S\right|}}\right) \tag{4.22}$$

Where *a* is the loudspeaker radius and $J_n\left(..\right)$ is the n^{th} order Bessel function of the first kind. Assuming a coordinate system with the virtual source placed at the origin and substituting *(4.22)* with respect to R gives.

$$\frac{\partial}{\partial R} S_{piston}(\mathbf{x},\omega) = \frac{A(\omega) jka e^{jkR}\left(2jR^2 J_1\left(\frac{kay_S}{R}\right) - ax_S J_0\left(\frac{kay_S}{R}\right) + ax_S J_2\left(\frac{kay_S}{R}\right)\right)}{R^2 x_S} \qquad (4.23)$$

Thus the 2.5D driving function of a circular piston in an infinite baffle can be expressed as:

$$D_{2.5,piston}(\mathbf{x}_0,\omega) = -\sqrt{\frac{|\mathbf{x}_{Ref} - \mathbf{x}_S|}{2\pi jk}} \frac{(\mathbf{x}_0 - \mathbf{x}_S)}{|\mathbf{x}_0 - \mathbf{x}_S|} \cdot \mathbf{n}(\mathbf{x}_0) . \frac{A(\omega) jka e^{jkR}}{|\mathbf{x}_0 - \mathbf{x}_S|} .. \frac{\left(2j|\mathbf{x}_0 - \mathbf{x}_S|^2 J_1\left(\frac{kay_S}{|\mathbf{x}_0 - \mathbf{x}_S|}\right) - ax_S J_0\left(\frac{kay_S}{|\mathbf{x}_0 - \mathbf{x}_S|}\right) + ax_S J_2\left(\frac{kay_S}{|\mathbf{x}_0 - \mathbf{x}_S|}\right)\right)}{|\mathbf{x}_0 - \mathbf{x}_S| y_S} \qquad (4.24)$$

An example of the focused wave field produced with this driving function is given in Fig. 4.5. The virtual source has a similar appearance to a point source but with a frequency dependent directivity pattern as a function of the piston radius. A piston therefore has a broader radiation spot as seen in Fig. 4.5.

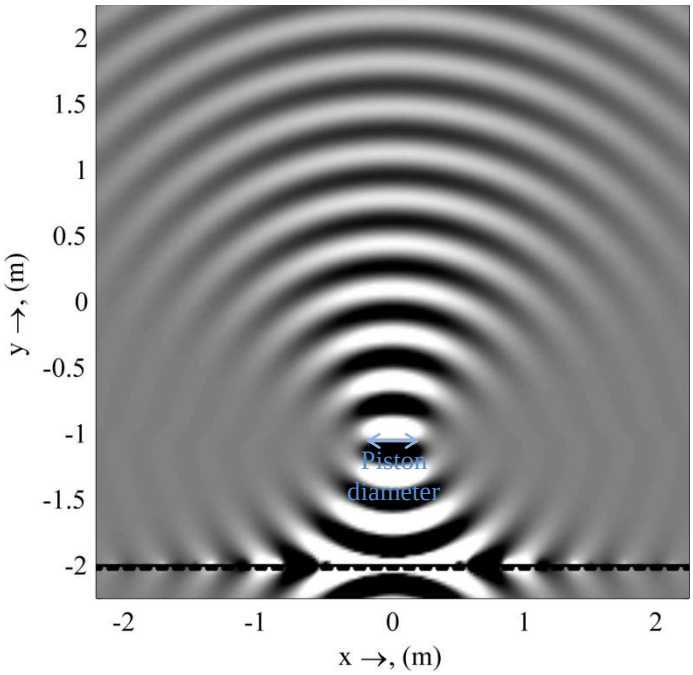


Fig. 4.5 Wave field of a focused loudspeaker modelled as a piston with a radius of 0.2m in 2.5D WFS, $f = 1000Hz$ at $\mathbf{x}_S = [0m, -1m]$ with $\Delta_{LS} = 0.125m$

The modelling of loudspeakers as focused piston sources in WFS is similar to the work done by Ahrens in modelling sources with extended radiation characteristics such as plates vibrating with different number of modes. A more detailed analysis can therefore be found in (Ahrens, 2012).

4.2 Secondary source selection

When practically rendering a focused source it is possible to choose the direction of propagation, this essentially means carefully selecting and weighting the loudspeakers such that the energy of the source is radiating in the required direction. This is not important for non-focused sources as they can be considered as omnidirectional point sources thus localisation is independent of listener position. For a focused source however the direction of propagation is important because a listener has to be within the source's view angle for correct localisation (see chapter 5). In order to choose a specific direction for a focused source, a source direction is chosen relative to the *x*-axis as a unit vector. This direction vector can then be compared with the unit direction vector of each loudspeaker in the normal way to check which ones should be active. Making sure that the dot product of the direction vectors of the secondary source $\mathbf{n}_{x_0}$ and the primary source $\mathbf{n}_{PS}$ is greater than zero:

$$\mathbf{n}_{x_0} \cdot \mathbf{n}_{PS} > 0 \qquad (4.25)$$

If this condition is not satisfied the loudspeakers are switched off. The second phase is to insure that the loudspeakers are behind the point of propagation. For a directed focused source this means drawing a line perpendicular to the intended source direction and making sure that only loudspeakers in the region behind this line are active. This can be tested using the set of conditions in *(4.26)* where x_{PS}, y_{PS}, x_0 and y_0 are the *x* and *y* coordinates of the focused source and loudspeaker respectively and θ_{focus} is the angle of the focused source with respect to the *x*-axis.

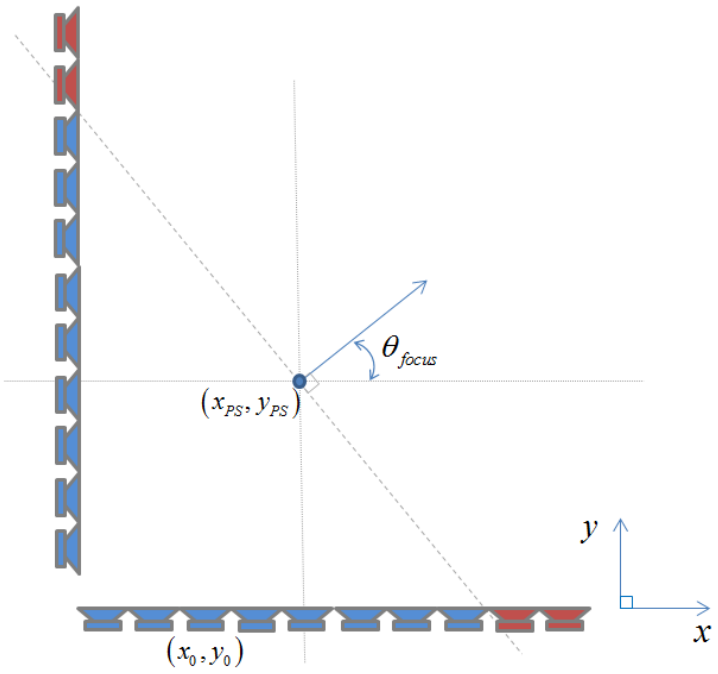


Fig. 4.6 Source selection geometry for symmetrical secondary source situations

For focused sources the secondary source selection criteria differs depending on the angle, θ_{focus}.

$$\text{if } \sin\left(\theta_{focus}\right) < 0$$

$$a\left(\mathbf{x}_0\right) = \begin{cases} 1, & \dfrac{x_{PS} - x_0}{\tan\left(\theta_{focus}\right)} + y_{PS} < y_0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{else if } \sin\left(\theta_{focus}\right) \geq 0$$

$$a\left(\mathbf{x}_0\right) = \begin{cases} 1, & \dfrac{x_{PS} - x_0}{\tan\left(\theta_{focus}\right)} + y_{PS} > y_0 \\ 0, & \text{otherwise} \end{cases} \qquad (4.26)$$

Applying these source selection criteria results in a focused source that can be positioned to propagate in any chosen direction with respect to the *x*-axis as shown in Fig. 4.7, where the arrows show the direction vectors of the active sources and the primary source.

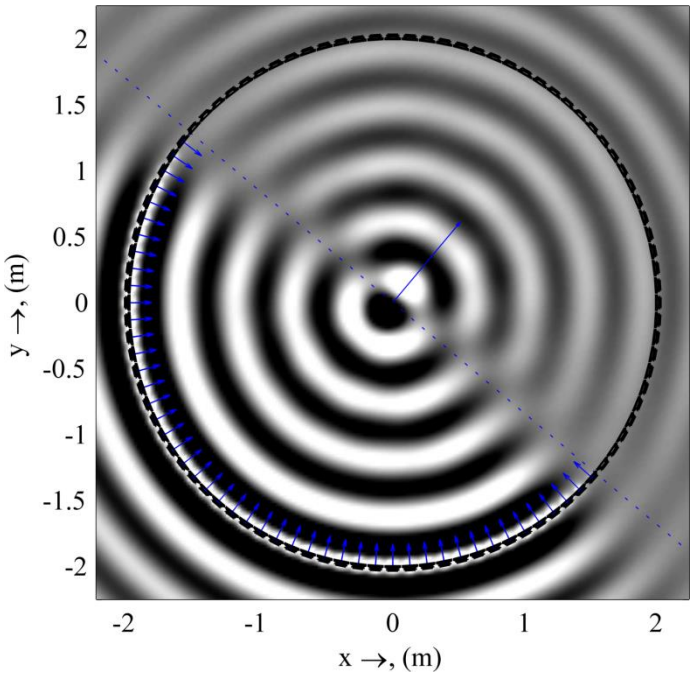


Fig. 4.7 A 2.5D WFS focused monopole virtual source with a position of [0m, 0m] travelling at 50° to the x-axis

It is very important, when defining an audio scene using WFS that the direction of any focused sources be considered carefully as the direction of the source dictates the areas where listeners will localise correctly. It has been suggested by (Melchior et al., 2008) that listener tracking can be used so update the WFS driving functions such that the listening space can be maximised. In the absence of such tracking or when there are multiple listeners it is important to consider the most probable location of listeners and then render the focused source directions accordingly. Rendering a focused source at an angle can also be useful for increasing the view angle of the focused source if using a square/rectangular array as it enables the source to be reproduced by all the loudspeakers in the corner of the array. This is not such a problem for circular arrays where the view angle is approximately 180˚ anyway (as is shown by the blue dotted line in Fig. 4.7) , but for rectangular arrays the view angle can be increase to nearly 180° which can be hugely advantageous. Obviously the source will then have a direction that may not be desired but for point sources this direction will not always be perceptible. For a source with a specific directivity pattern (e.g. Corteel, 2007) the angle of the source again becomes important.

A further consideration needed for the source selection of directed focused sources is to check whether the selected sources are symmetrically distributed, that is to say it is important that the angle between the source direction and the final source at each end of the array is the same. When this is not the case, the weighting of the source from one side of the array is greater than that from the other side of the array and the source radiates asymmetrically as seen in Fig. 4.8.

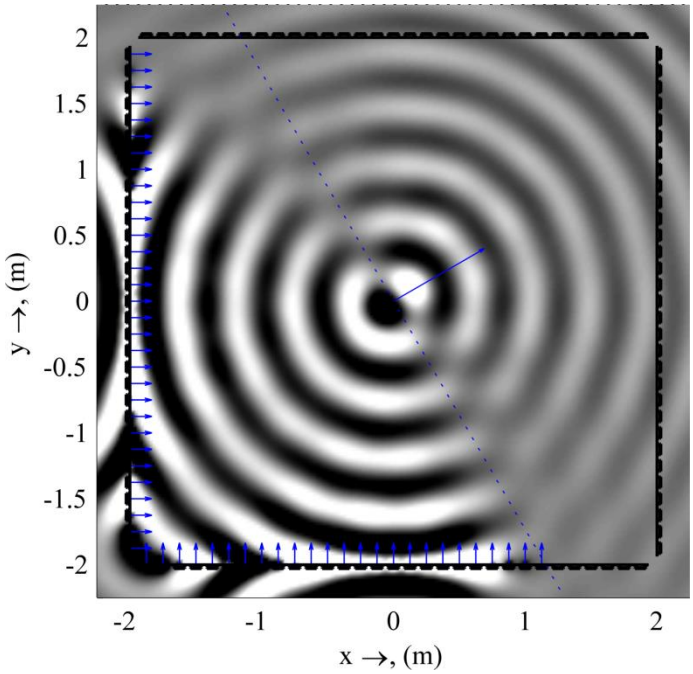


Fig. 4.8 Directed focused source with an asymmetric secondary source selection

On close inspection of Fig. 4.8 it can be noticed that the principle energy from the source is slightly off-axis from the intended direction as shown by the blue arrow. This is less problematic for monopole sources as in Fig. 4.8 for sources with a complex directivity pattern the angle of propagation is important and should not be affected by the secondary source selection. These problems with the asymmetric source distribution can be alleviated by employing an additional source selection requirement based on the geometry in Fig. 4.9.

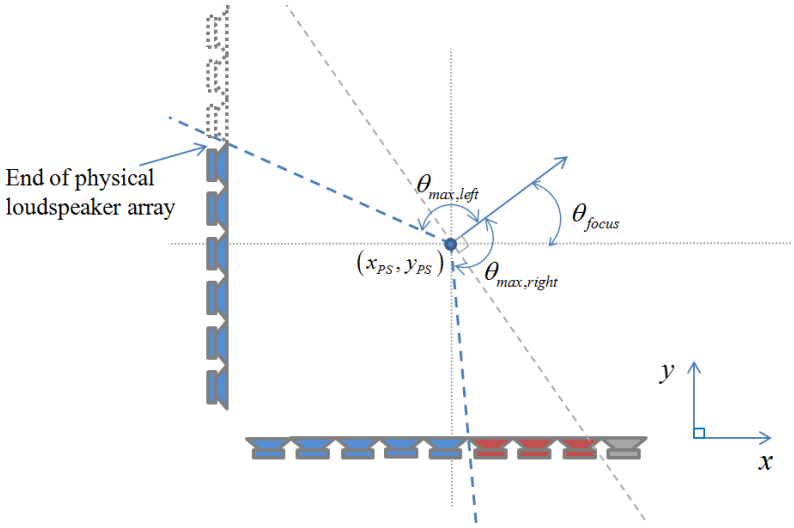


Fig. 4.9 Source selection geometry for non-symmetrical situations. The blue sources are the active sources, red sources are real sources that have been switched off, the grey source falls outside of the source selection criterion and the dotted loudspeakers are imaginary.

From Fig. 4.9 it becomes apparent that in order to have a symmetrical secondary source distribution, the source selection criteria must satisfy the relation $\theta_{max,left} = \theta_{max,right}$. As such a further source selection condition is therefore given that:

$$a(\mathbf{x}_0) = \begin{cases} 1, & \theta_{max,left} \geq \theta_{max,right} \\ 0, & \text{otherwise} \end{cases} \tag{4.27}$$

Equation *(4.25)* is formulated for the geometry Fig. 4.9. If this condition is adhered to the wave field in Fig. 4.8 becomes as shown in Fig. 4.10 which exhibits better symmetry around the focus point when compared with the asymmetric source selection case.

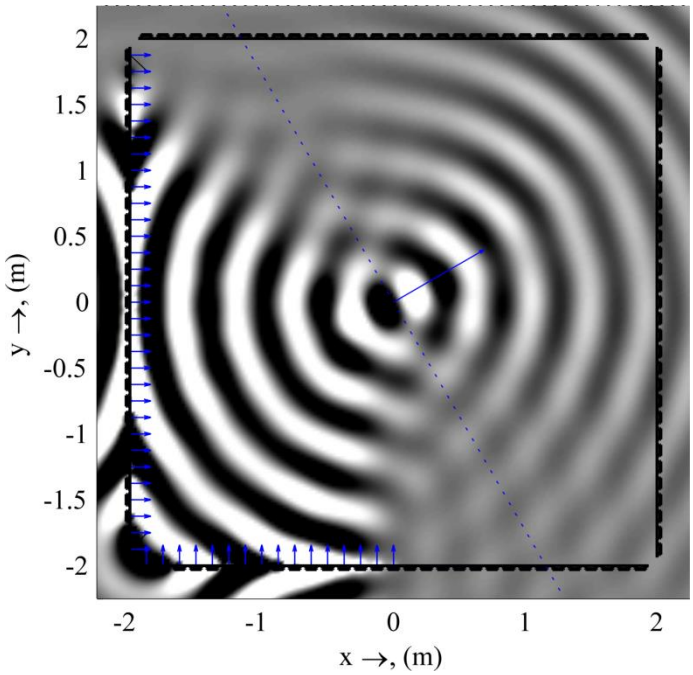


Fig. 4.10 Directed focused source with a symmetric secondary source selection

4.3 Amplitude errors

As mentioned in chapter 2, the primary localisation mechanism for distance perception of acoustic sources is amplitude cues. As such it is important for the creation of focused sources that the amplitude cues are recreated as faithfully as possible. It is more important that this is the case for focused sources as they are nearer to the listener than standard WFS sources thus movement of the listener with respect to the source has a greater effect. The amplitude decay of a focused source using the driving function of *(4.10)* is plotted in Fig. 4.11 compared with the amplitude decay of an equivalent real source.

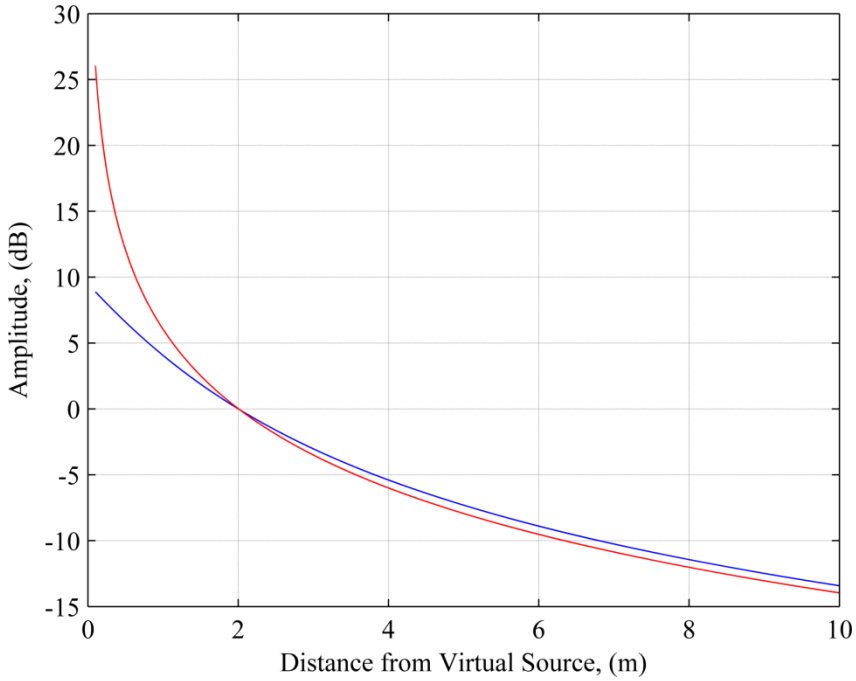


Fig. 4.11 Distance attenuation of a $f = 200Hz$ focused source with the reference point, $\mathbf{x}_{Ref}$ $2m$ in front of the source position (blue trace) compared with distance attenuation of an equivalent real source (red trace).

4.3.1 Position of the reference point

The reference point or line defines the position in space where the amplitude errors resulting from the conversion from 3D to 2.5D reproduction are equal to zero. This position can be arbitrarily set such that the reproduction is optimised for a given listener position. For focused sources however it is better that the reference point be situated in front of the virtual source position as to have zero amplitude error in the region behind the source position makes little sense as this is the convergent region of the focused source and the amplitude distribution of a focused source on the axis perpendicular to the loudspeaker array is not symmetrical about the focus point in the same way a real source is symmetrical about the point of propagation. Consequently it is better to define the reference point to be set to a given distance in front of the specific focused source or as close to the reference point used for the non-focused sources but still being in front of the focused source.

If the reference point is chosen to be too close to the source, the amplitude will be correct only in the very near vicinity of the virtual source position therefore a more

appropriate location for the reference line would be about 1-2m in front of the source, such that it is only very close to the source that the distance cues will be inaccurate.

Fig. 4.11 shows the amplitude error for one position, the amplitude error as a function of source distance from the loudspeaker array and listener distance is shown in Fig. 4.12. For this calculation of the distance errors for WFS focused monopole sources, the reference point is set to always be the same distance in front of the source position so the distance error can be compared for different source distances from the secondary source array.

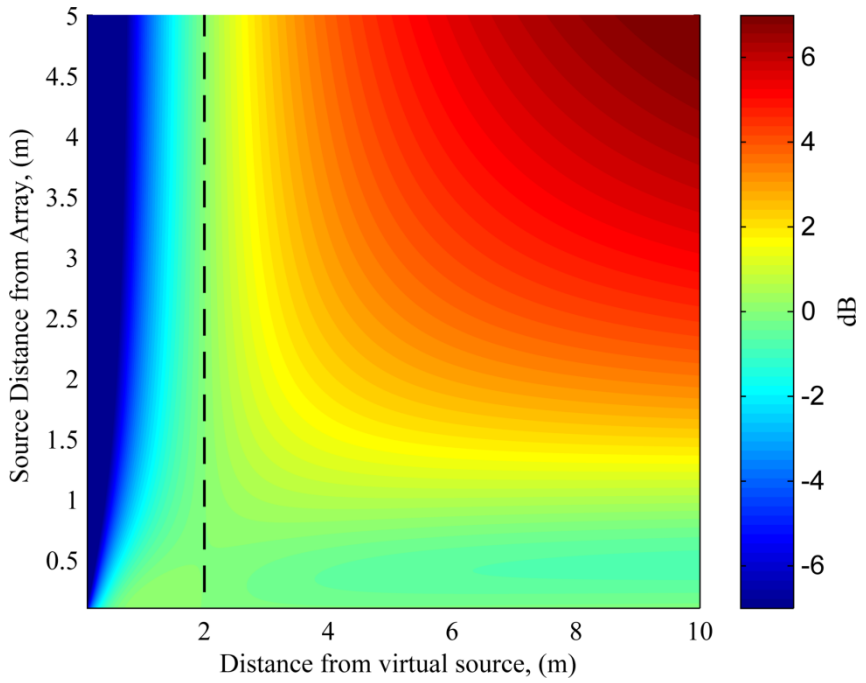


Fig. 4.12 Distance attenuation error between a real and WFS focused source, $f = 200Hz$ with the reference point, $\mathbf{x}_{Ref}$ $2m$ in front of the source position. The dashed black line shows the position of the reference line 0dB error

Fig. 4.12 shows that the errors in amplitude cues with respect to distance are not equal for different source positions. The errors are greatest when the source is positioned further from the array. When the source is very close to the array the amplitude errors with distance tend to zero, which is an expected result since, the closer the source is to the array the fewer active secondary sources there will be in

the rendering and thus the radiation tends towards that of a single loudspeaker where there will be no amplitude errors.

One of the interesting features of focused sources that can be seen in Fig. 4.11 is that as the distance from the source increases the gradient of the attenuation tends towards that of a real source. This characteristic is in contrast to non-focused sources where the difference in distance attenuation between a real and a WFS source is proportional to $1/\sqrt{r}$. The reason for this is that the recreated source, although based on a 2D source model is actually a 3D source because all of the energy is focused on to one point rather than a line which is theoretically infinitesimal, thus, from this source position, the radiation is in all directions and approximates that of a spherical source. The consequence of this is that if the listener is close to the virtual source and moves away from it, the distance cues will change in an unexpected manner and the listener may therefore localise the source incorrectly with respect to distance, however as the listener moves still further away from the virtual source the change in cues will more closely match those of a real source and it is therefore more likely that the listener will localise the source distance correctly.

4.3.2 Lateral amplitude errors

The amplitude in the lateral direction i.e. taking a line of receivers parallel to the *x-axis* can also be computed for a focused source and compared to that of a real point source as shown in Fig. 4.13.

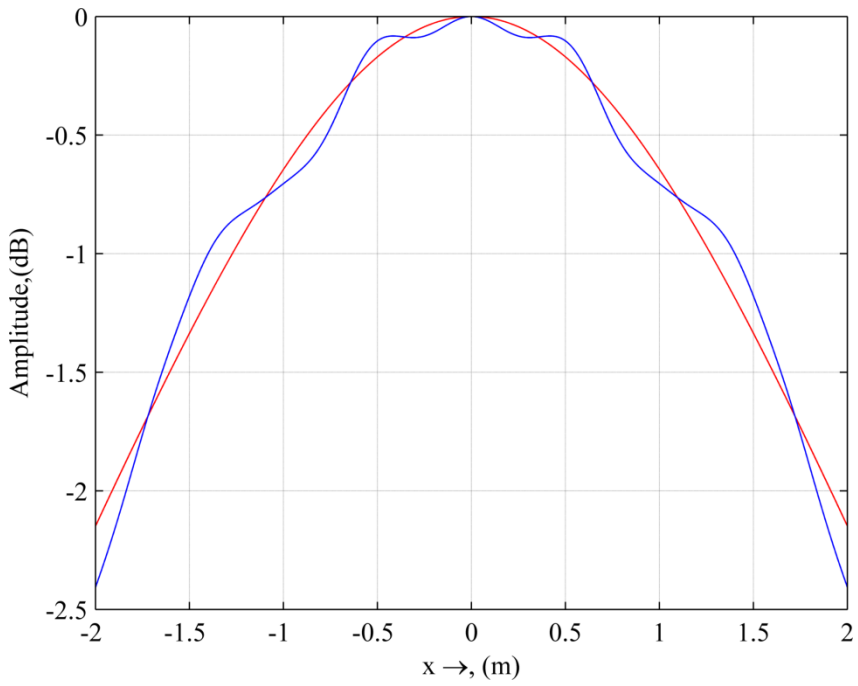


Fig. 4.13 Amplitude distribution in the lateral direction for a real source (red trace) and WFS focused source (blue trace) for a $800Hz$ virtual source *0.5m* in front of an *8m* loudspeaker array with $\Delta_{LS} = 0.125m$ and $\mathbf{x}_{Ref}$ *2.5m* in front of the virtual source. The analysis line is parallel with the *x-axis* at a distance of *2.5m* in front of the virtual source

The lateral amplitude in Fig. 4.13 is shown to be quite accurately reproduced by the WFS focused source, however there is a clear ripple in the amplitude distribution of a WFS focused source with respect to a real point source which is a consequence of array truncation and the limits of acoustic focusing such that a true point source cannot be fully recreated. Interestingly, the ripple still remains for very large arrays where pre-echoes cause aberrations in the rendered wave field as described in more detail in section 4.11. If the array is truncated too heavily, the focusing effect of the array is too weak and the Q-factor of this curve is too narrow and exhibits strong ripples.

The ripples are caused by the fact that the source is being reproduced by a line array of secondary sources and therefore an analytical solution for the ripples can be obtained from the theory of line array loudspeakers if desired.

4.4 Spatial aliasing

One of the interesting physical attributes of WFS focused sources is the spatial aliasing artefacts resulting from the discrete sampling of the secondary source array (see section 3.5.2). For a normal WFS source the spatial aliasing frequency is constant throughout the reproduction space and is often considered simply to be a function of the loudspeaker distance and the wave velocity. There are small changes in this frequency with respect to space as pointed out by (Corteel, 2006a) but most authors consider the aliasing frequency to be constant throughout the space. Contrary to this, the spatial aliasing frequency for focused sources is heavily dependent on the position in the listening space as can be seen from Fig. 4.17.

The effects of aliasing can also be seen in the frequency domain as shown in Fig. 4.14, where there is a well-defined frequency at which strong colouration artefacts occur in the reproduced wave field. For non-focused sources this frequency is independent of space as can be seen from the plot in Fig. 3.17, however for the focused source case colouration introduced in the received spectrum happens at a different frequency with respect to distance.

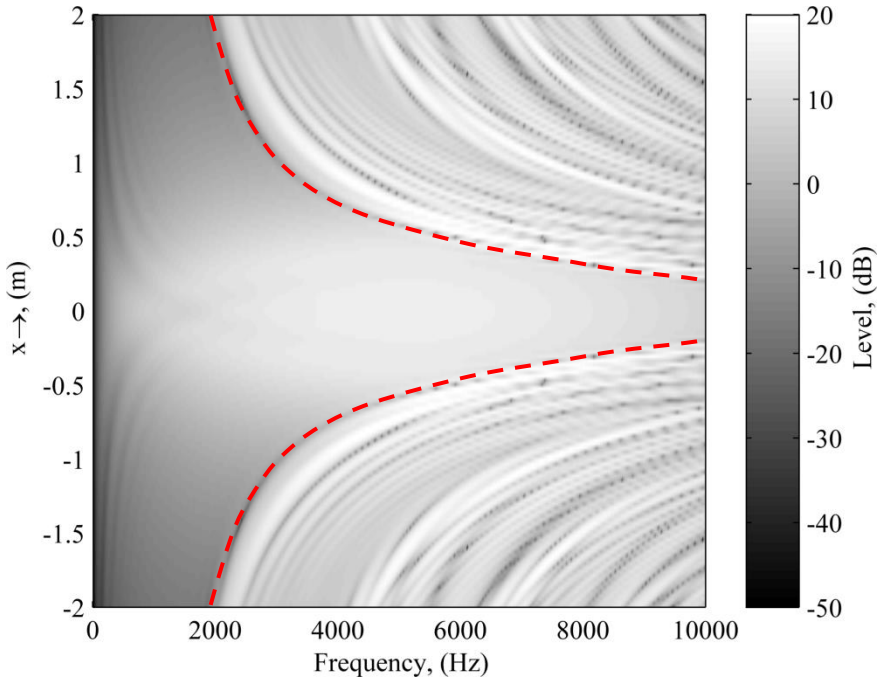


Fig. 4.14 Frequency spectrum of a focused source, with a 4m long receiver line *0.5m* in front of a broadband focused source *1m* from the secondary source array. The dashed red line shows the approximate boundary between the aliased and un-aliased region

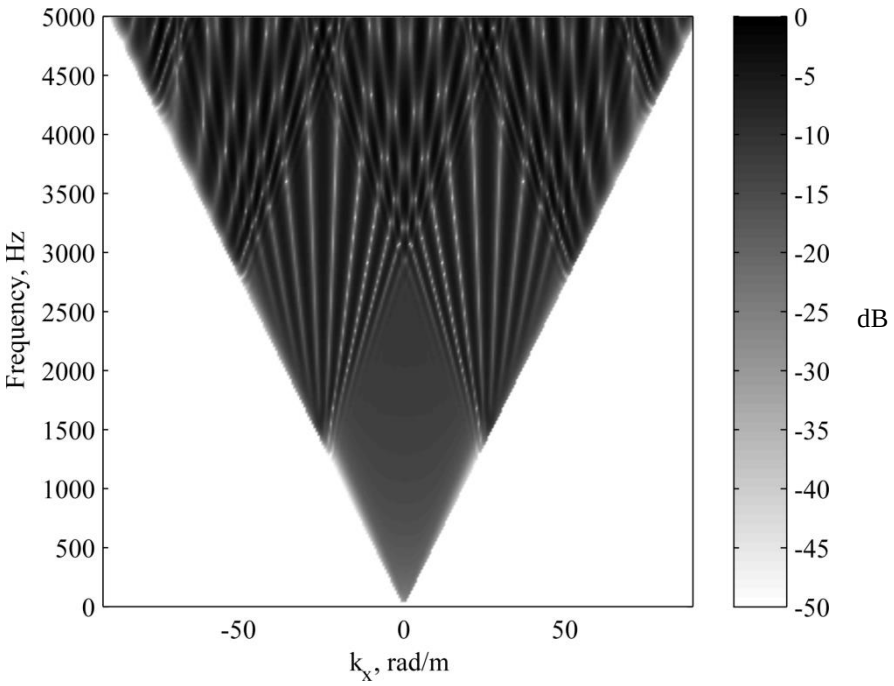


Fig. 4.15 Wavenumber domain plot of a focused source 1m in front of a 6m long loudspeaker array with $\Delta_{LS} = 0.125m$. The line of receiver points is 0.2m in front of the virtual source position

This position-dependent aliasing frequency means that as a listener moves throughout the reproduction space they will move from aliased regions to un-aliased regions. As described in chapter 3, spatial aliasing results in spatially varying coloration in the reproduced wave field as can be seen in the wavenumber representation in Fig. 4.15. Moving from a region of strong colouration artefacts to an un-aliased, uncoloured region will be perceptually disturbing and as such it is important to characterise the un-aliased region for a given focused source so as to help with the design of a sound scene or to inform user-dependent WFS rendering techniques.

4.4.1 Defining the un-aliased region

The fact that focused sources exhibit this extended area of un-aliased contributions has important ramifications for sound scene design and for the use of so-called *local sound field synthesis* techniques (Spors and Ahrens, 2010a) where focused sources are used as secondary sources in a virtual sound field synthesis array. The characterisation of this area is also important for the techniques of removing the

effects of scattering objects in the convergent region of focused sources as described in section 6.6.4.

4.4.1.1 Geometric method

The definition of the un-aliased area for focused sources can be done using a simple geometrical approach such that a radius around the source can be determined where there will be no spatial aliasing artefacts in the reproduced wave field. The area of this un-aliased region area will be quantified in the following for both the linear and circular array cases. The theory uses a geometrical sound propagation method to define a region in the wave field that will remain un-aliased for the frequency in question.

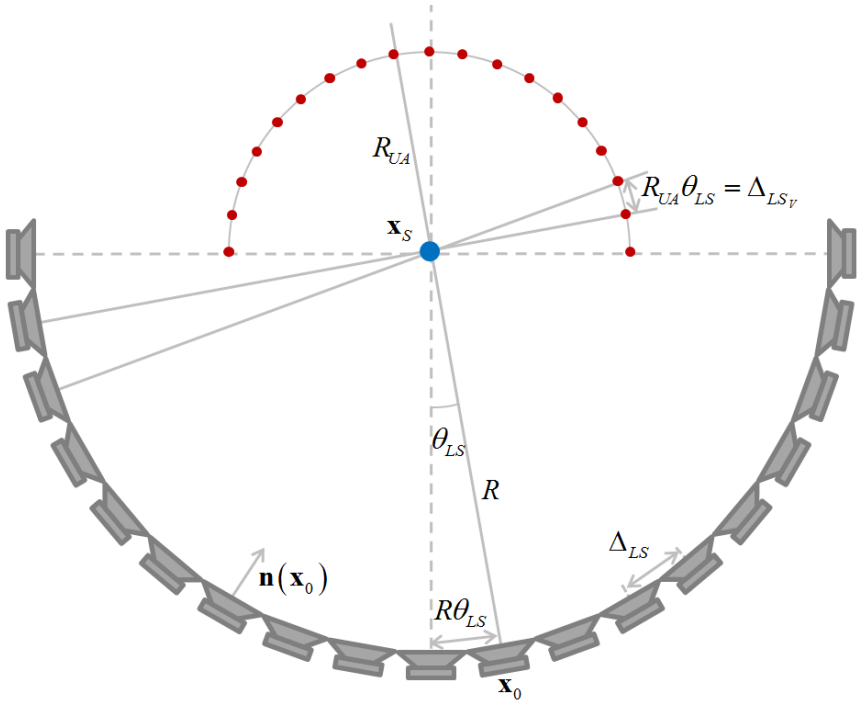


Fig. 4.16 Geometry for defining the un-aliased region for circular loudspeaker arrays

Using Huygens-Fresnel Principle the distance from the virtual source at which there will no aliasing effects can be calculated by assuming a ray-based propagation model as shown in Fig. 4.16. The Huygens-Fresnel principle states that the reproduced wave front can be considered to be the superposition of a distribution of secondary sources, here we assume that there is a finite number of secondary wavelets radiating with a position on the reproduced wave front determined by the intersection of a sound ray from each secondary source with the virtual source position as shown in

Fig. 4.16. Thus a set of virtual secondary sources can be assumed at a distance apart, Δ_{LS_V} which would result in aliasing for the given frequency of interest. As the radius of the wave front increases, there is still the same number of wavelets defined by the ray-based propagation model and thus there will be a radius, R_{UA} above which, the secondary wavelets will be too far apart to reproduce the wave field accurately. Thus a virtual loudspeaker separation Δ_{LS_V} can be defined, a value greater than which will result in aliasing in the reproduced wave field.

The traditional spatial aliasing frequency as presented by Spors and Rabenstein (2006) (see section 3.5.2) can be given by *(4.28)*.

$$f_{alias} = \frac{c}{\Delta_{LS}\left(1+\left|\sin\theta\right|\right)} \tag{4.28}$$

Thus the loudspeaker separation at which aliasing contributions would begin to be seen in the reproduced wave field is expressed as:

$$\Delta_{LS} = \frac{c}{f_{alias}\left(1+\left|\sin\theta\right|\right)} \tag{4.29}$$

So the distance between the virtual sources, Δ_{LSv} (given by the red dots in Fig. 4.16) at which aliased artefacts will begin to appear in the reproduced wave field can be given as:

$$\Delta_{LS_V} = R_{UA}\,\theta_{LS} = \frac{c}{f_{alias}\left(1+\left|\sin\theta_{LS}\right|\right)} \tag{4.30}$$

$$R_{UA} = \frac{c}{f_{alias}\,\theta_{LS}\left(1+\left|\sin\theta_{LS}\right|\right)} \tag{4.31}$$

Where θ_{LS} is the angle in radians between the loudspeakers as shown in Fig. 4.16. R_{UA} describes the radius around the virtual source within which there will be no spatially aliased contributions in the reproduced wave field.

Applying *(4.31)* to the scenario shown in Fig. 4.17 predicts an un-aliased region surrounding the virtual source with a radius of *1.027m* which can be seen to be an accurate prediction.

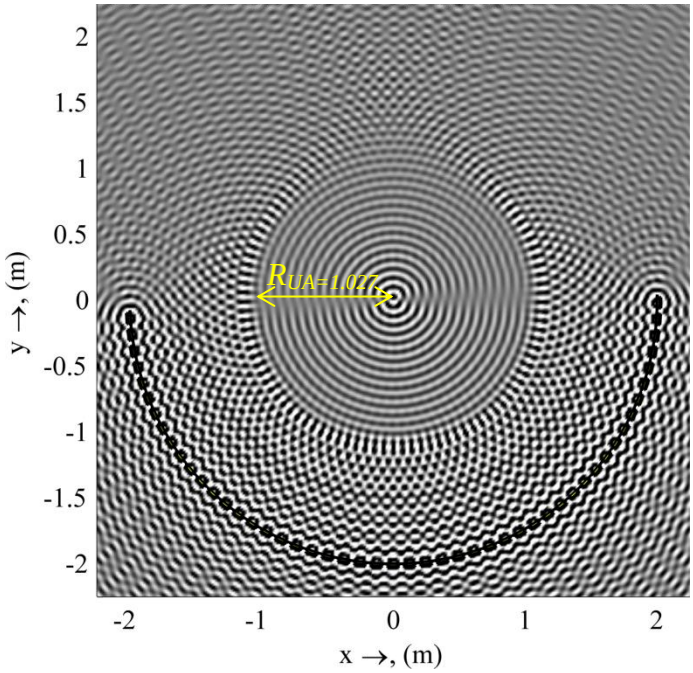


Fig. 4.17 Focused source, $f = 5000Hz$ at [0m, 0m] with $\Delta_{LS} = 0.125m$, exhibiting a clear un-aliased zone of radius, R_{UA} demarked by the yellow arrow

The radius of this un-aliased region is a function of the source frequency and the inter-loudspeaker angle, and so is a function of virtual source position as can be seen from Fig. 4.18. One of the major problems with focused sources therefore is that the spatial aliasing frequency changes constantly not only with virtual source position but also with listener position. This is problematic for colouration perception for moving virtual sources and also for pre-filtering of the source signal (which should only be applied up to the spatial aliasing frequency (Spors and Ahrens, 2010b)) as the cut-off frequency of the filtering function changes with listener position, as will be seen in the subsequent section.

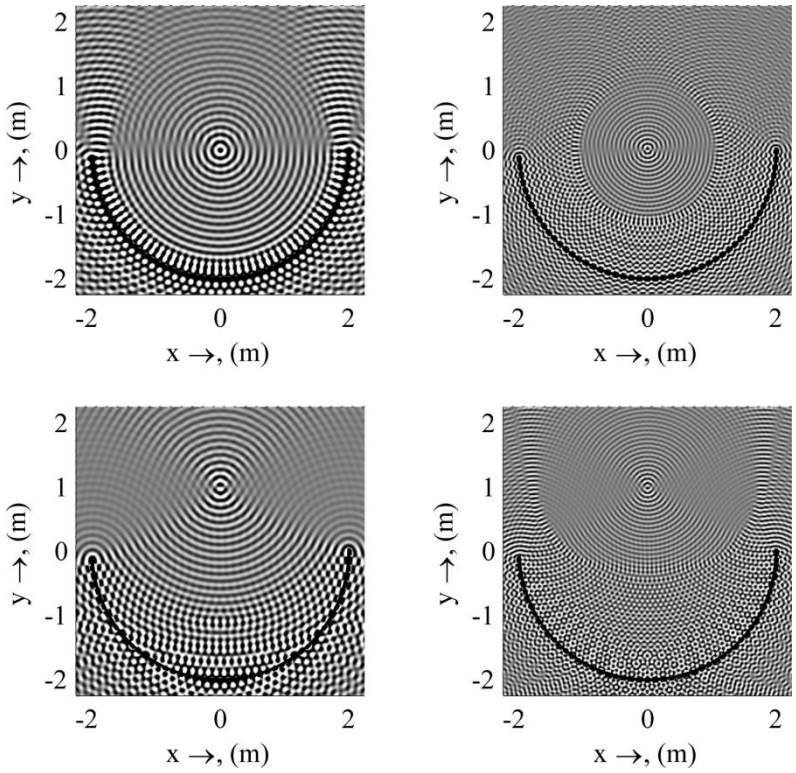


Fig. 4.18 Focused sources, with $\Delta_{LS} = 0.125m$, exhibiting a clear un-aliased zone varying in width with respect to frequency and source distance. Top left: focused source at $[0m, 0m]$, $f = 3000Hz$. Top right: focused source at $[0m, 0m]$, $f = 5000Hz$, Bottom left: focused source at $[0m, 1m]$, $f = 3000Hz$, bottom right: focused source at $[0m, 1m]$ $f = 5000Hz$

For the case of a linear array of loudspeakers or where the virtual source is not centrally located in a circular array, the shape of the un-aliased region becomes more complex and may extend beyond the limit defined by *(4.31)*. The aliased superposition of all of the secondary sources produces a complex wave field which cannot easily be defined by geometrical methods, however the wave field will still exhibit a circle around the virtual source with a radius defined by *(4.31)* at which aliasing contributions will not be present. This circle defines the minimum area around the source in which there will be no aliased contributions although the region may extend further than this in the propagation direction of the source. As shown in Fig. 4.19, the inter-loudspeaker angle, θ_{LS} changes across the array so in this case the aliasing condition is controlled by the maximum angle.

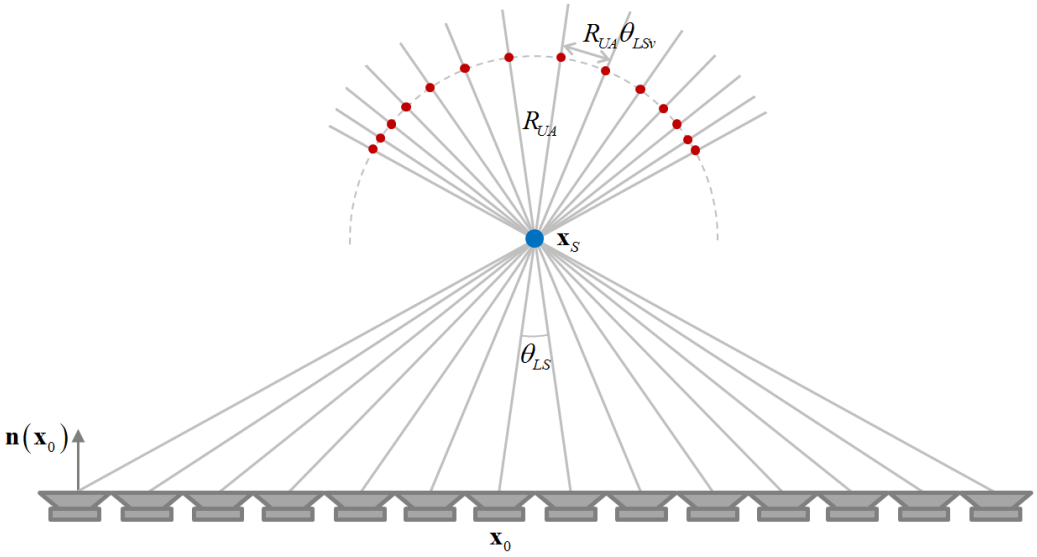


Fig. 4.19 Geometry for defining the un-aliased region for linear loudspeaker arrays

Fig. 4.20 demonstrates how for sources produced by linear loudspeaker arrays and those produced off-centre by circular arrays still exhibit a circular aliasing-free zone defined by *(4.31)* as shown by the yellow trace on the plot.

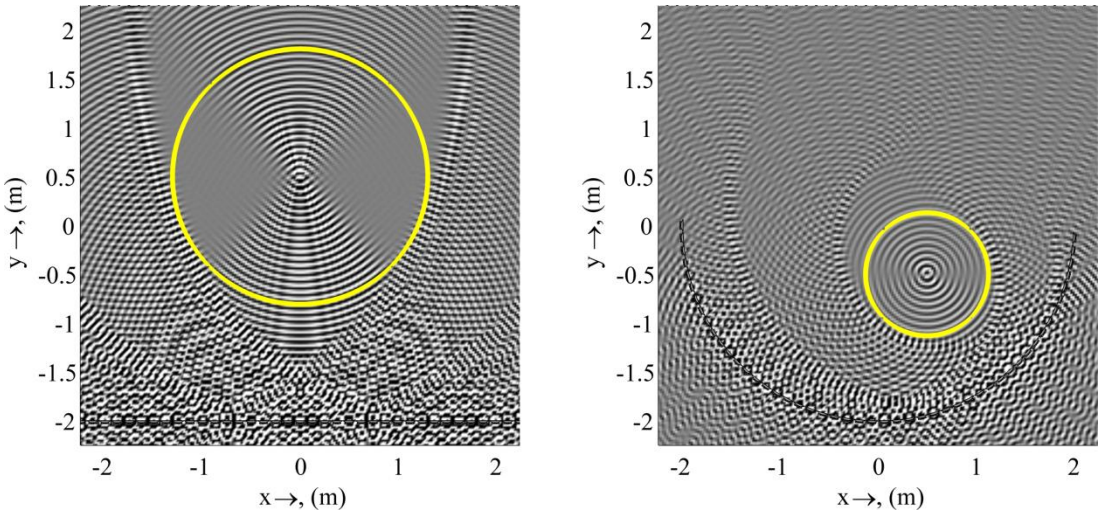


Fig. 4.20 Reproduced wave field of a *5kHz* focused source with $\Delta_{LS} = 0.125m$ including an un-aliased region (demarked by the yellow line) for a source at $[0m, 0.5m]$ produced by a linear array (left) and a source produced by a circular array at $[0.5m, -0.5m]$ (right)

If a listener is anywhere outside this un-aliased region it can be expected that they will experience colouration in the reproduced wave field, these colouration effects are spatially variant and are thus affected by both source and listener movements.

Further analysis on the colouration of focused sources in WFS is the topic of the subsequent section.

Spors and Ahrens (2011) also derive a formula for predicting this un-aliased region based on the work by Dickins et al. (2005) analysing the spatial dimensionality that can be synthesised using a limited number of loudspeakers.

$$R_{UA} = \frac{y_S c}{\Delta_{LS}} \tag{4.32}$$

Applying *(4.32)* to the example above and assuming that the distance, y_S is the largest distance in the *y*-direction, gives a value of the un-aliased region of *1.098m* which for this example slightly over-predicts the radius of this un-aliased region but is within a similar region to the condition as given here by *(4.31)*. A further limitation of this formula and indeed *(4.31)* is that they both fail to accurately predict the un-aliased region for off centre sources rendered using circular arrays as shown in Fig. 4.23. Thus there is a clear need for a more rigorous formula that can deal with arbitrary loudspeaker setups and source positions. Such a formula is described below.

4.4.1.2 Time domain method

For the case of a linear array of loudspeakers or where the virtual source is not centrally located in a circular array, the shape of the un-aliased region becomes more complex and may extend beyond the limit defined by *(4.31)* as shown in Fig. 4.20. The aliased superposition of all of the secondary sources produces a complex wave field which cannot easily be defined by geometrical methods, however the wave field will still exhibit a circle around the virtual source with a radius defined by *(4.31)* at which aliasing contributions will not be present. This circle defines only the minimum area around the source in which there will be no aliased contributions and can thus be considered a worst case scenario. In reality the region may extend further than this, particularly in the propagation direction of the source.

Due to the limitations of the geometrical approach to describe this more complex region, the un-aliased region can be better classified using a time domain approach similar to that used for unfocused sources by (Corteel, 2006a) by looking at the difference in arrival times between the direct energy from the secondary sources and the superposed energy via the virtual source position for consecutive loudspeakers.

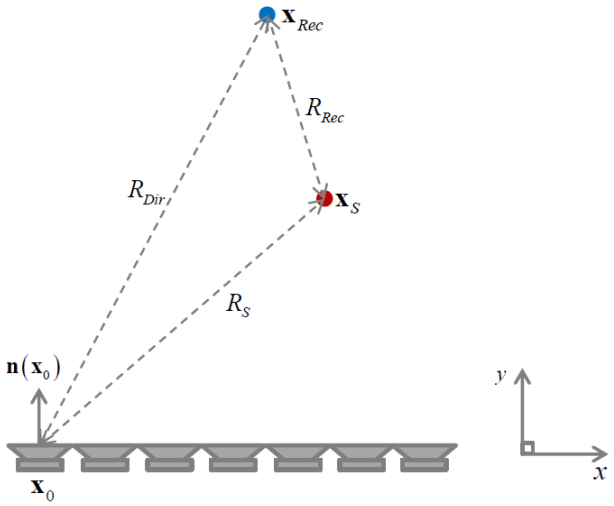


Fig. 4.21 Geometry used for the time domain derivation of the un-aliased region for focused sources

The geometry used for this derivation is given in Fig. 4.21 with the total distance, $R_{Indirect}$ between the source and receiver through the virtual source expressed as:

$$R_{Indirect} = R_S + R_{Rec} \tag{4.33}$$

The path difference, ΔR between the direct path from loudspeaker to receiver and the path through the virtual source therefore given as:

$$\Delta R = R_{Indirect} - R_{Direct} \tag{4.34}$$

To compute the anti-aliasing condition, the relative values of ΔR between consecutive loudspeakers is compared across the array, if the maximum difference across all loudspeakers exceeds the source wavelength, λ then aliasing contributions can be expected in the reproduced wave field. When this is computed for a given point in space it is possible to determine whether or not aliasing contributions can be expected. Performing this for each point in the room allows the plotting of a map of

the un-aliased region as shown in Fig. 4.22. This method also allows the determination of the un-aliased region for irregular loudspeaker spacings and complex array geometries as shown in Fig. 4.25.

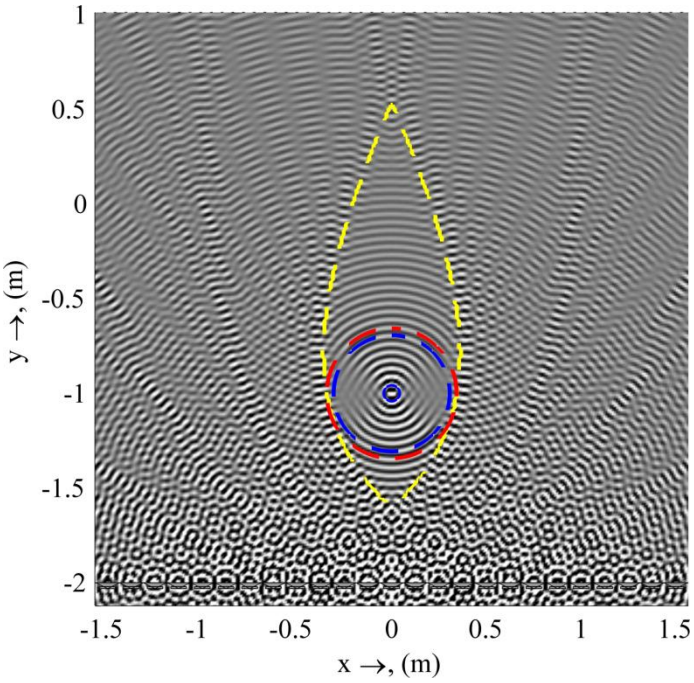


Fig. 4.22 Un-aliased Region for a *8kHz* source at *[0m, -1m]*. The Red trace marks the un-aliased region as given by *(4.32)* the blue trace given by the geometrical method from *(4.31)* and the yellow trace shows the time domain method given by *(4.35)*

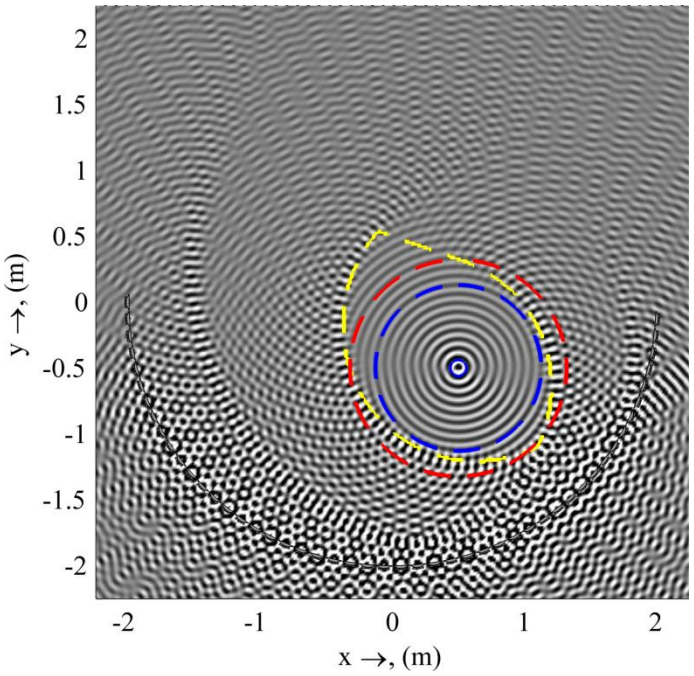


Fig. 4.23 Un-aliased Region for a *5kHz* source at *[0.5m,-0.5m]*. The Red trace marks the un-aliased region as given by *(4.32)*, the blue trace given by the geometrical method from *(4.31)* and the yellow trace shows the time domain method given by *(4.39)*

The spatial aliasing frequency for a given position in the listening space for a given virtual source, ψ can therefore be given as:

$$f_{alias}(\mathbf{x},\psi) = \frac{c}{\max \Delta R} \tag{4.35}$$

This equation allows the plotting of the variation in the spatial aliasing frequency across the listening space for a given source position as shown in Fig. 4.24. This is a useful tool as it allows, on inspection, the region of validity of a given focused source at and below a given frequency of interest.

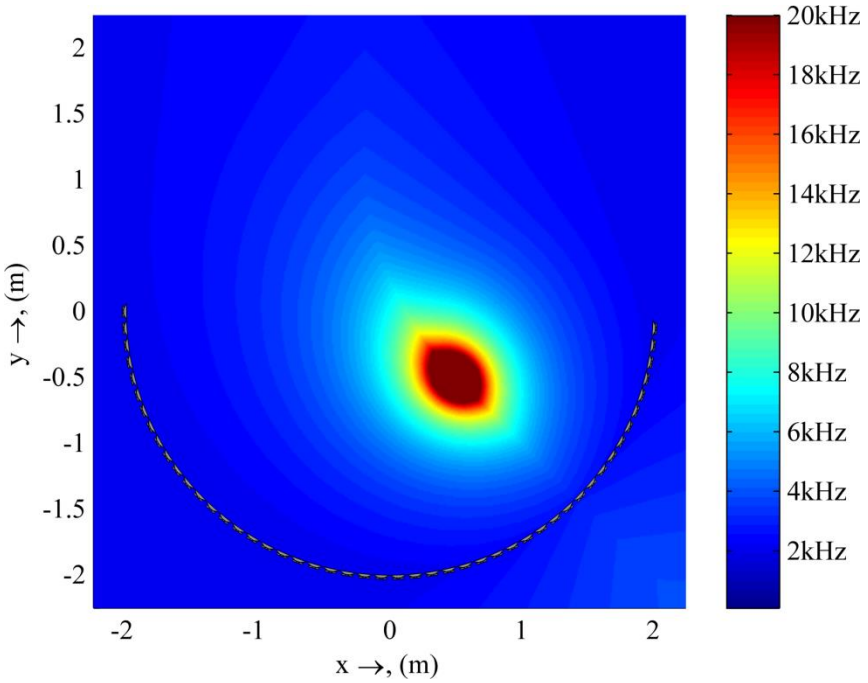


Fig. 4.24 A map of the variation in spatial aliasing frequency across the listening space for a source at *[0.5m, -0.5m]*, the colour bar is clipped at *20kHz* for illustrative purposes

If a listener is anywhere outside this un-aliased region it can be expected that they will experience colouration in the reproduced wave field, these colouration effects are spatially variant and are thus affected by both source and listener movements.

4.5 Colouration

The above mentioned spatial aliasing artefacts result in unwanted colouration in the reproduced wave field. As can be seen from Fig. 4.14, these colouration artefacts are strong and are highly position dependent, as such when a listener moves around the listening space or if the focused source is moving, changes in colouration become more obvious and undesirable to listen to. To some extent these effects are reduced when the reproduction system is placed in a non-anechoic environment as room effects smudge out the spatially aliased components as will be discussed in the following chapter.

Informal listening tests in an anechoic chamber suggest that these colouration effects are particularly disturbing and many listeners complained of 'phasing' effects as they moved about the chamber, which suggests that moving focused sources in an anechoic environment would be particularly problematic. The colouration in the reproduced wave field is primarily caused by the high frequency spatial aliased pre-echoes which cause comb filtering as described in more detail in section 4.10.3. These pre-echoes and the resulting colouration can be seen clearly in impulse response and frequency response shown in Fig. 4.25.

As there is a frequency and position dependent un-aliased zone for focused sources (see section 4.4.1), there are particular problems as listeners move around the reproduction area as they move from an aliased region to an un-aliased region etc.

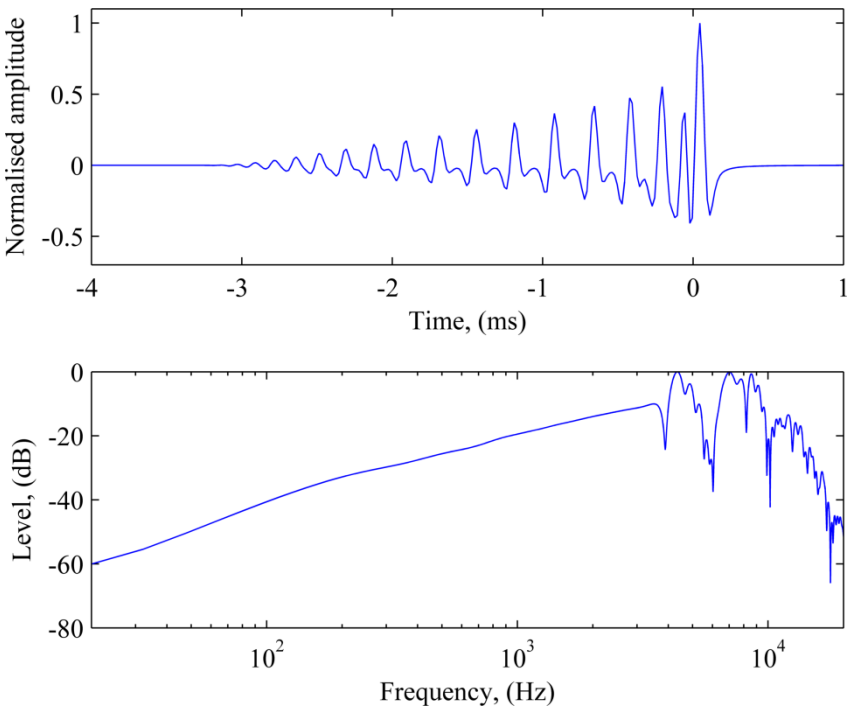


Fig. 4.25 Typical broadband impulse response and corresponding frequency response at *[0.5m, 0.5m]* for a WFS focused source at *[0m, 0m]* produced by a *4m* long array at $y = -1m$ with $\Delta_{LS} = 0.125m$. Note the repetitive aliased pre-echoes artefacts cause comb filtering in the frequency response above the aliasing frequency.

Further colouration can be caused by anomalies in the pre-filtering of WFS source. For non-focused sources the source functions are typically pre-filtered before rendering as the filtering function is not source position dependant, this is a useful technique as it greatly reduces the computational overhead. The pre-filtering function for a focused source is shown in Fig. 4.26, corresponding to a high pass filter function. However the pre-filtering is typically only applied up to the spatial aliasing frequency (Spors and Ahrens, 2010b) as the spatial sampling of the secondary source array already adds an increase in energy of roughly 3dB/octave. This poses a problem for focused sources as the spatial aliasing frequency is heavily position dependent, meaning that the cut-off frequency for the pre-filter should really be position dependent. It is undesirable to have such a position dependent filter thus typically only one pre-filter is applied which can be chosen to match the spatial aliasing frequency at a point in space e.g. the reference point, $\mathbf{x}_{Ref}$. The consequence of this is that some areas of the listening space will not have the correct equalisation applied and as such a colouration below the spatial aliasing frequency may be discernible at some listener positions. This problem can of course be alleviated by

applying the filtering operation at the point of rendering rather than as a pre-filter but the additional computational load needed for this can reduce the number of sources that a WFS can render and may therefore be undesirable. In reality the perceptual effect of these colouration artefacts is small as they only change gradually with listener position and will simply result in an increase in low frequency energy.

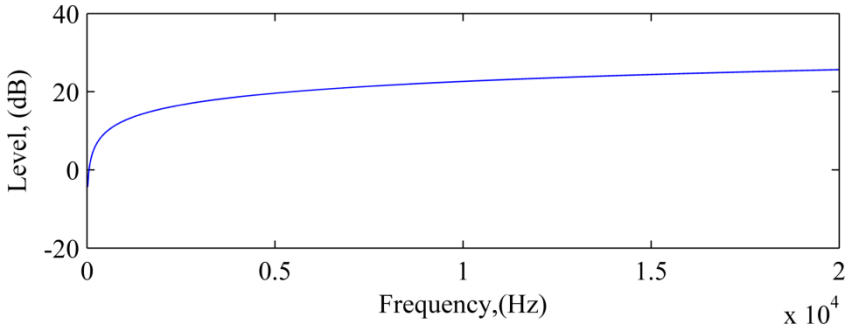


Fig. 4.26 Pre-filter for focused sources given by the $\sqrt{k/j}$ term in *(4.10)*

4.6 Truncation effects

The effects of array truncation have been covered in chapter 3 for non-focused sources. For the case of focused source the truncation of the secondary array introduces additional wave fronts from the edges of the array that arrive at all listener positions before the intended virtual source and thus form pre-echoes which are perceptibly disturbing. Truncation of the loudspeaker array can also cause rippling in the reproduced wave field as seen in Fig. 4.27.

One of the most prominent effects of the truncation of the secondary source array however is the resulting reduction in wave front curvature of the intended virtual source. As the array length decreases the requirement of strong focusing can no longer be met and the radiation from the extremities of the array reduces the wave front curvature at the edges of wave. These effects cause what is known as the focal shift phenomena which is described in more detail in section 4.9 and if uncorrected for will result in errors in the depth of the reproduced source. The focal shift phenomenon is linked with the limit of acoustic focusing, i.e. there has to be a sufficient number of secondary sources in the array in order to be able to achieve a given strength of focusing.

Truncation effects can also cause colouration in the reproduced wave field (Start, 1997) and impose limits on the size of the listening space. Truncation effects are especially prominent for focused sources when the array is not just one linear array i.e. using a corner or circular array. In these cases, the additional diffraction waves from the edges of the array radiate with greater amplitude into the listening area causing ripples in the reproduced wave field as shown in Fig. 4.27.

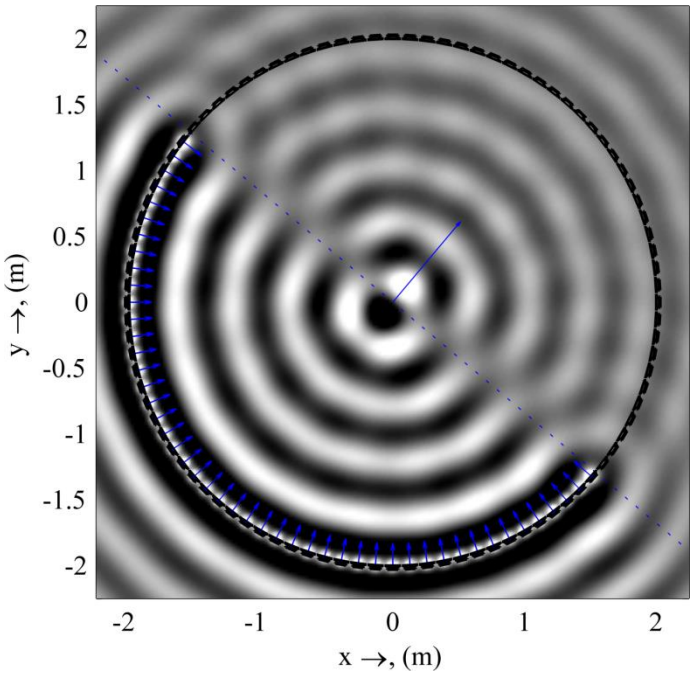


Fig. 4.27 Truncation artefacts present in the reproduced wave field of an *800Hz* focused source $[0m, 0m]$ at an angle of *50°* to the *x*-axis

One of the effects of truncation is that the reproduced focused source becomes broader in dimension at the focal point as described in the following section. The position of the focal point also is affected by the truncation as will be described in section 4.9.

4.7 Focal point size

It is well known in diffraction theory that the minimum size of a focal spot is limited by diffraction. This physical limit means that the dimension of the focal in the transverse direction is λ/2 whereas the depth resolution is λ. This can be seen from

the RMS amplitude distribution taken through the focal spot in both the *x*-direction and *y*-direction as shown in Fig. 4.28.

An analytical expression for the pressure in the focal plane is given by (Lucas and Muir, 1982) as:

$$P_{Lucas} = \left| \frac{2J_1(kax/R)}{kax/R} \right| \tag{4.36}$$

Which can be shown to match up well with the pressure distribution in the focal plane of a WFS focused source as shown in Fig. 4.28. The pressures zeros are then given by the zeros of the Bessel function.

The 3dB down width the of the focal spot in the focal plane is given by Chen et al. (1993) as:

$$Width_{dB_{-3}} \approx \frac{3.232R}{ka} \tag{4.37}$$

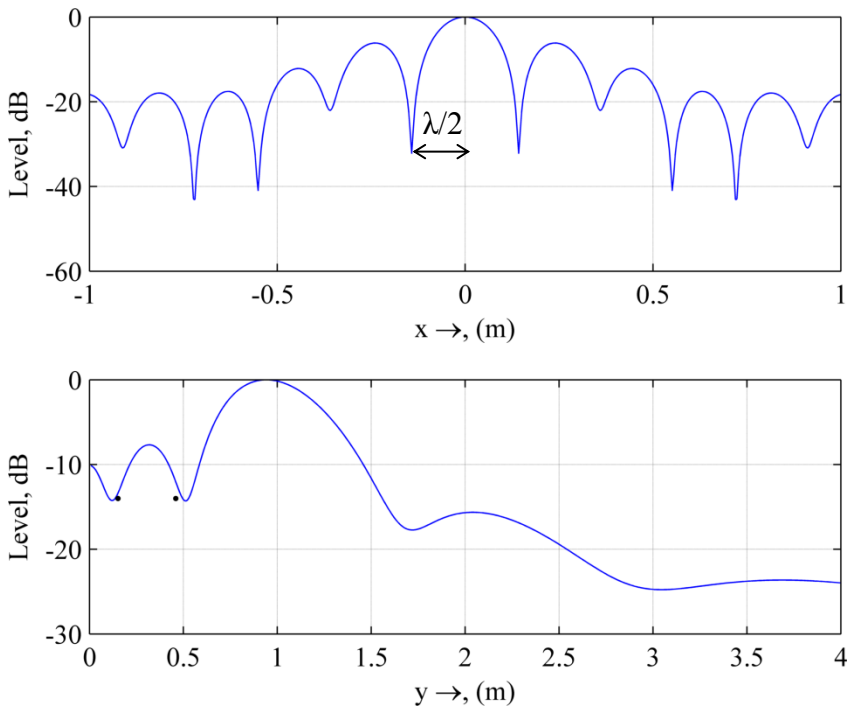


Fig. 4.28 RMS pressure distribution in dB through the focal point in the transverse direction (top) and range (bottom) for a monochromatic focused source *f = 1000Hz* at 1m with a 4m array. The black dots on the bottom plot shown the 1st two minima as predicted by *(4.38)*

An analytical expression for the on-axis pressure of a focusing source is given by O'Neil (1949) and also by Lucas and Muir (1982) both of whom provide analytical expressions for the locations of the pressure minima in the focal plane. The solution by Lucas and Muir can be considered as an approximation of that given by O'Neil who formulates the locations of the minima as:

$$z_n = \frac{\chi R}{2n\pi R + \chi}\left(1 - \left(\frac{2n\pi}{ka}\right)^2\right) \qquad (4.38)$$

Where the Rayleigh distance $\chi = ka^2/2$ and n is any positive integer. R is the geometrical focus position and a is the apparent array aperture, given by b in Fig. 4.33. These predictions match up well with the RMS pressure distribution for a WFS focused source shown by the black dots in Fig. 4.28.

4.8 Phase response

An interesting aspect of focused sources is that the phase response exhibits a shift of π radians with respect to a real source at the same position as the source crosses the focus point. This phase shift can be seen in Fig. 4.29.

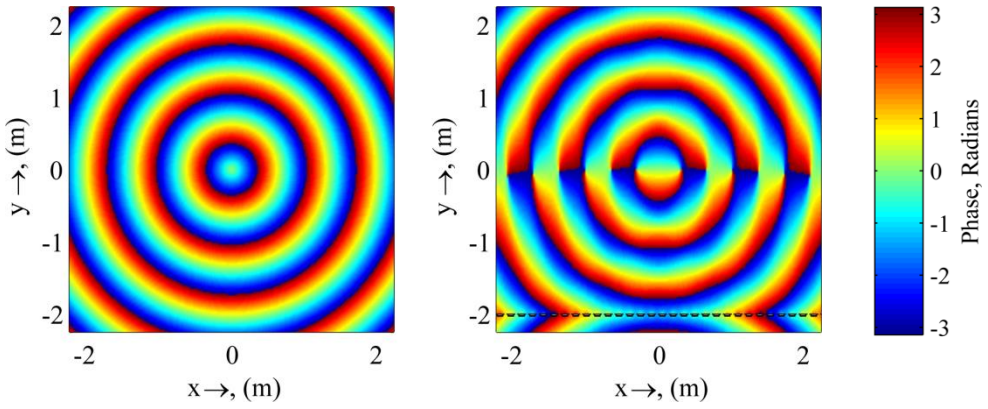


Fig. 4.29 Phase responses of a real source (Left) and a WFS focused source (Right) for $f = 500Hz$ $\Delta_{LS} = 0.125m$

In optical literature it is commonly referred to as the phase anomaly (Born and Wolf, 1999) but has also been applied for the assessment of acoustic focusing applications (Lucas and Muir, 1982) and can be seen in any wave-based focusing scenarios. The phase anomaly is the difference in phase between a real source at the source position and the source created by focusing waves to that same point. The 'real' source used for the focus anomaly is actually more like that of a plane wave but radiating as a point source (not a physically realisable source model) such that the phase is given as:

$$\phi = \begin{cases} -kR, & \text{when } R \leq 0 \\ +kR, & \text{when } R \geq 0 \end{cases} \qquad (4.39)$$

Where R is the distance from the focal point. This condition describes plane wave type propagation but with spherical wave fronts.

The phase anomaly is useful for depicting the position of the focused wave field however it cannot be used for the determination of the focal point as it requires a priori information of the source position for its computation. As such the focused source position can be determined by the location of the RMS pressure maximum as will be discussed more in the following section.

4.9 Position errors

At high frequencies and with large array apertures, the acoustic focusing follows geometric laws, thus the focal position can be chosen and rendered accurately using a simple ray-based geometry and scaling the delays of the secondary sources respectively. For low frequencies and small array apertures however, significant diffraction effects occur causing a shift in the expected focal point. Such a phenomena has been demonstrated, and is well known in both the field of optics (Born and Wolf, 1999) and ultrasonics (O'Neil, 1949; Lucas and Muir, 1982; Makov et al., 1982) but has not been applied in the field of audio acoustics and the rendering of focused sources in wave field synthesis. This section presents the basic theory of diffraction-based focal point position errors and presents an analytical model based

on (Makov et al., 1982, 2008) for predicting the location of the pressure maximum on axis to the loudspeaker array. Fig. 4.30 demonstrates the difference in focal point position as a result of different focusing effects of the diffraction and geometric zones of focusing.

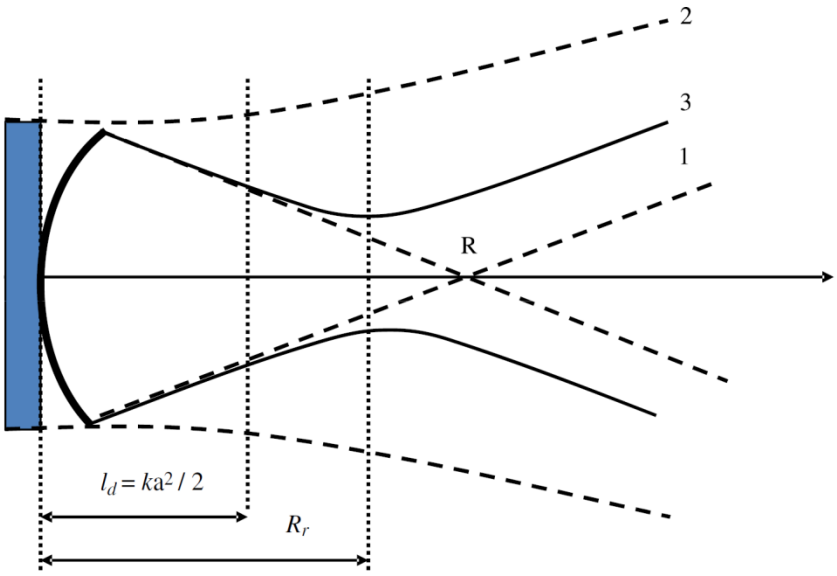


Fig. 4.30 Diagrammatic interpretation of the focal shift phenomena (after Makov et al. (2008)). Line 1 shows the geometrical focus, Line 2 shows the diffraction (without focusing) and Line3 shows the combined effect of Line 1 and Line 2 (diffractional divergence and geometrical focusing).

4.9.1 Makov's model

The theoretical background for the so called *focal shift* is well documented in the optical literature and was first discussed for acoustic scenarios by (O'Neil, 1949) and was subsequently improved upon by (Lucas and Muir, 1982). These papers provide models that aim to predict the local pressure maximum corresponding to the position of the acoustic focus point, and formulate how this focus position differs from the intended geometric focal point. Both of these models describe and quantify the phenomenon using diffraction theory, however they both utilise a simplification of the system such that the resulting equations are only valid for cases where the aperture is large with respect to the wavelength. Makov et al. (1982) also present a model for predicting the position of the pressure maximum for a spherical focusing transducer with the aim of demonstrating how the focal shift phenomena varies due to non-linear affects, as a result of their work in the non-linear region, they also present a more accurate model for the linear case which was later revisited and expanded upon in (Makov et al., 2008).

Makov's model for the focal shift of ultrasonic transducers was formulated for spherical focused transducers which consist of a plate that is curved such that the energy focuses onto a point, the position of which is dictated by the radius of curvature in a similar way to an optical lens. If the focusing lens is strong, i.e. there is a large initial wave front curvature, the lens behaves geometrically and the focus point can be predicted accurately, however when the initial wave front curvature is small, the lens tends towards the behaviour of a flat piston and thus there is little or no focusing present. These phenomena can be described well by using the Fresnel Number, well known in Optics (Born and Wolf, 1999).

The Fresnel number of the spherical focusing transducer can be given as:

$$N_F = \frac{a^2}{\lambda R} \tag{4.40}$$

with the geometry given in Fig. 4.31.

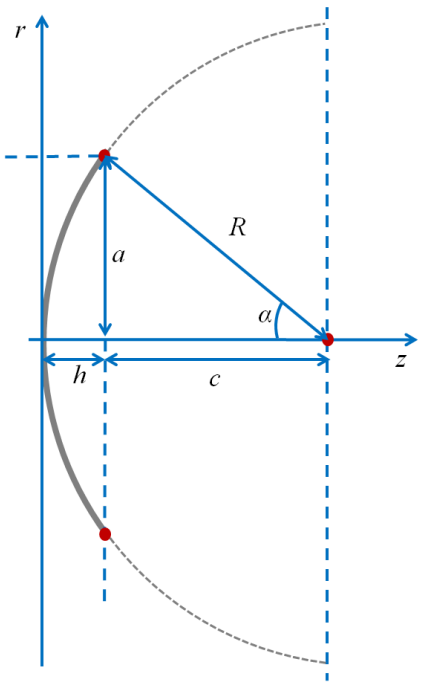


Fig. 4.31 Geometry for deriving Makov analytical model focal shift in focusing transducers

The Fresnel number can alternatively be thought of as the number of half wavelengths that are contained within *h* such that it is given as:

$$N_F = \frac{h}{\lambda/2} \tag{4.41}$$

Where *h* can be expressed in terms of the aperture, *a* of the transducer and the geometric focal distance, *R* as:

$$h = R - c = R - \sqrt{R^2 - a^2} \simeq \frac{a^2}{2R} \tag{4.42}$$

The focusing transducer can loosely be split into two regions corresponding to $N_F \gg 1$ where the lens will be a strong wave front curvature and the lens behaves approximately geometrically. Alternatively for $N_F \ll 1$ the initial wave front curvature is weaker and the effect of focusing is weaker such that transducer tends towards the solution of a plane circular driver and the lens will not follow geometric laws, being affected more by the diffractional divergence as shown in Fig. 4.30, resulting in a focal shift and a broadening of the Q-factor of the rms pressure distribution. As can be seen from *(4.40)*, the Fresnel number is controlled by both the radiated wavelength and the aperture, thus for low frequencies or small apertures the focusing effect is weaker and the focal shift phenomena becomes more prominent, conversely large apertures and higher frequencies exhibit focusing behaviour that can be easily predicted using geometric approaches.

As (Lucas and Muir, 1982), the axial pressure can be given as:

$$p(r,z,t) = A(r,z)e^{j(kz-\omega t)} \tag{4.43}$$

Where the complex pressure amplitude, $A(r,z)$ along the lens axis can be determined by the parabolic approximation of the ordinary wave equation:

$$A(r,z) = -\frac{jk}{z} e^{j\frac{kr^2}{2z}} \int_0^a e^{j\frac{kr'^2}{2z}} J_0\left(\frac{k}{z} rr'\right) A(r',0)\, dr' \qquad (4.44)$$

The boundary condition is then assumed to be that of a truncated Gaussian pressure distribution. From here Makov formulates the on axis pressure ratio where A_0 is the constant pressure amplitude across the transducer (assuming a uniform pressure distribution) as:

$$\left|\frac{A(0,\tilde{z})}{A_0}\right| = 2\left|\frac{1}{1-\tilde{z}} \sin\left(\frac{\pi N_F}{2}\frac{1-\tilde{z}}{\tilde{z}}\right)\right| \qquad (4.45)$$

The focal point is defined as the value of $\tilde{z}$ at the peak of *(4.45)*. There is no direct solution to this equation however. Taking the derivative of this function and equating it to zero allows for the formulation of the pressure maxima as:

$$\tilde{z}\cot(X) = 1 \qquad (4.46)$$

Where $X = \frac{\pi N_F}{2}\frac{1-\tilde{z}}{\tilde{z}}$

An exact solution of this transcendental equation is difficult to obtain therefore, in order to obtain an explicit formula for the focal shift of a given transducer, *(4.46)* can be re-written as:

$$\frac{\sin(X)}{X} = \left(\frac{2X}{\pi N_F}\right)\cos(X) \qquad (4.47)$$

This can be expressed as a 3rd order power series and taking only the significant terms can be expressed as:

$$24\tilde{z}^2 = \pi^2 {N_F}^2 \left(3 - 4\tilde{z} + \tilde{z}^2\right) \qquad (4.48)$$

The solution to *(4.48)* is given by Makov *et al* as:

$$\tilde{z}_{Max} = \frac{3\pi N_F}{2\pi N_F + \sqrt{\pi^2 N_F{}^2 + 72}} \tag{4.49}$$

This result is useful as it allows the analytical prediction of the focal shift for a given transducer as shown in Fig. 4.32. When compared with previous formulae from other authors (O'Neil, 1949; Lucas and Muir, 1982) it proves to give better predictions for low Fresnel number lenses as it is based on the wave equation and can thus be considered an exact solution (when calculated numerically from *(4.44)*). The analytical approximation given in *(4.49)* is consequently very useful in the case described below for the determination of the low frequency errors in the position of WFS rendered focused sources.

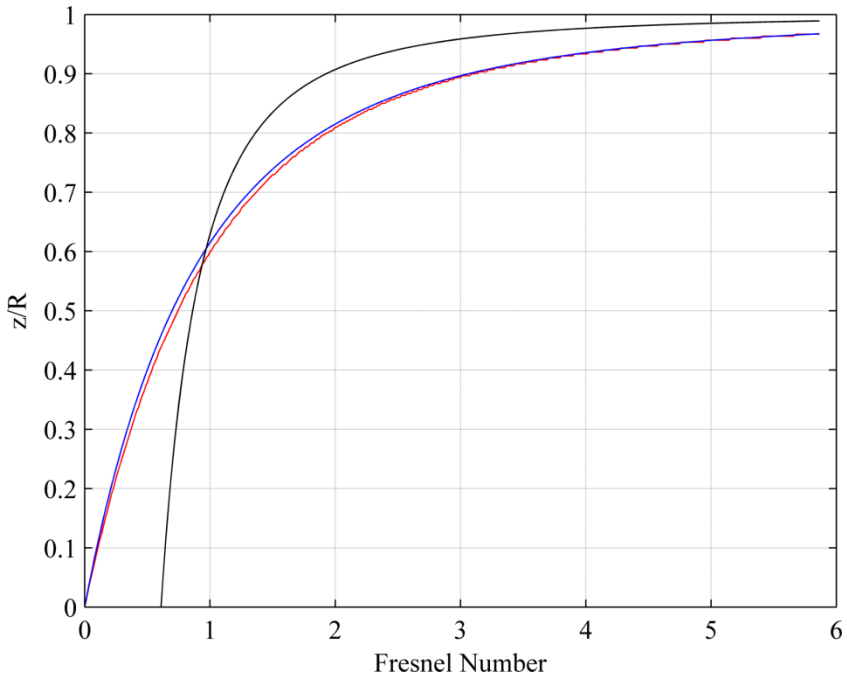


Fig. 4.32 Comparison of equations predicting the focal shift of focused transducers with a geometric focus of 2m. The blue trace is the exact solution given by *(4.44)*, the red trace is the approximation thereof given by *(4.49)*, and the black line is the prediction given by (Lucas and Muir, 1982; O'Neil, 1949)

4.9.2 Focal shift prediction for WFS focused sources

Fig. 4.36 shows the pressure on axis to the loudspeaker array at different distances and Fresnel numbers. The peak of this RMS pressure distribution denotes the focal

point of the system, which in this case is at 3m. The black dashed line shows the maximum RMS pressure for each Fresnel number. Fig. 4.36 demonstrates how the focal point deviates from the geometrical focal point for Fresnel numbers below approximately 4. This can be corrected to some extent as described in the following. For now we seek to quantify this deviation from the geometric focus based on the above described formula.

The model described in section 4.9.1 can be adapted to predict the case of a linear array of loudspeakers rendering a focused source as a result of the relative delays between the loudspeakers as is the case in WFS. The model is based on an amended geometry given in Fig. 4.33.

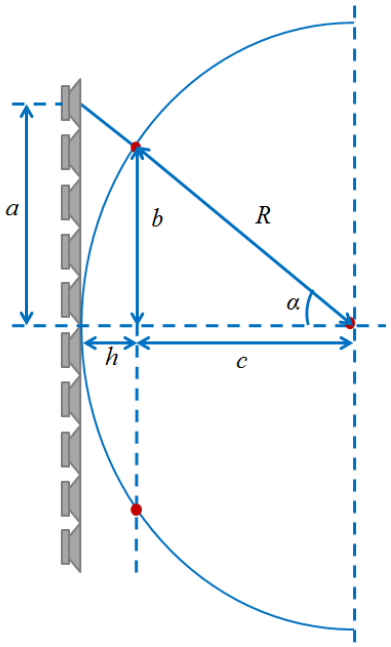


Fig. 4.33 Geometry for deriving analytical model for focused source position errors

Once again, the Fresnel Number can be given as the number of half wavelengths that can propagate within the distance, *h* between the lens and the aperture plane and is given by *(4.41)*. For focusing applications using linear arrays of loudspeakers, *h* can be inferred from Fig. 4.33 as:

$$h = R - c = R\left(1 - \cos\alpha\right) \tag{4.50}$$

Where

$$\alpha = \tan^{-1}\left(\frac{a}{R}\right) \tag{4.51}$$

Where a is the array aperture which is half of the length of the active loudspeaker array and R is the distance between the loudspeaker array and the intended focused source (also known as the geometric focal point). Combining *(4.50)* and *(4.51)*, h can be expressed as:

$$h = R\left(1 - \frac{1}{\sqrt{1+(a/R)^2}}\right) \tag{4.52}$$

From this the Fresnel-Number, N_F can be calculated from *(4.41)* for a given wavelength, λ. Then using *(4.49)* the focal shift of a WFS reproduced source can be predicted.

The sound pressure in dB on axis for a typical WFS focused source at different frequencies is shown in Fig. 4.34 exhibiting a strong peak at the focus point.

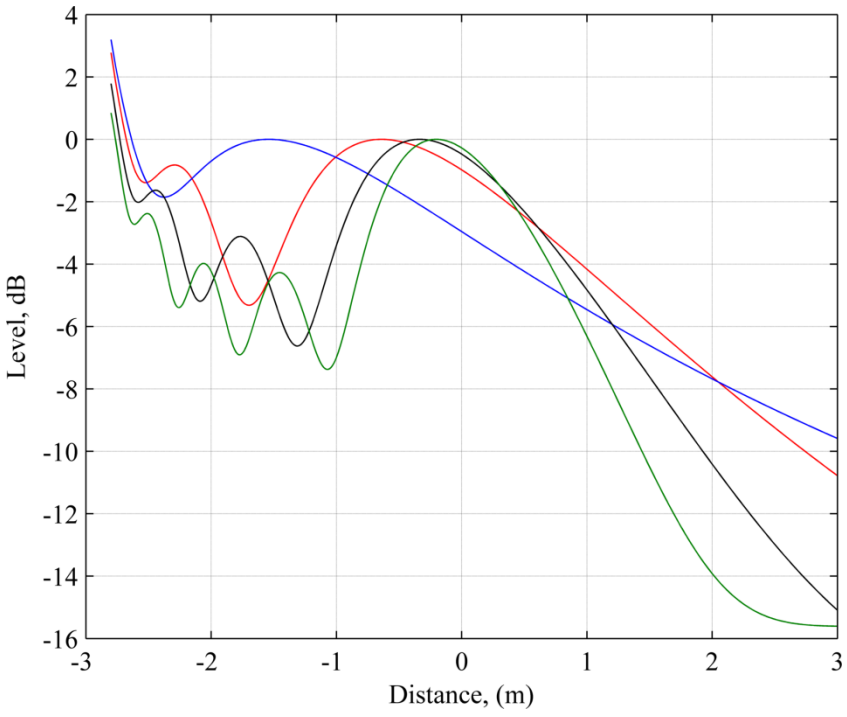


Fig. 4.34 Level in dB normalised to focal point for different frequency focused sources placed at 0m with the 4m loudspeaker array at -3m. Green: 1200Hz, Black: 900Hz, Red: 600Hz and Blue: 300Hz.

As the frequency is decreased, the position of this peak moves away from the geometrical focus and towards the loudspeaker array as expected. This behaviour is predicted accurately by *(4.49)* as shown in Fig. 4.35.

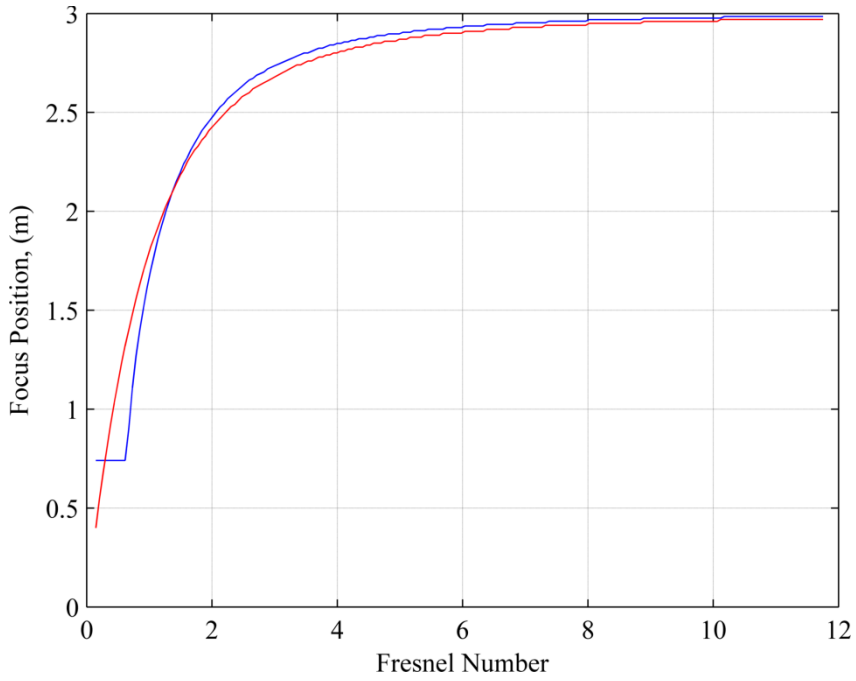


Fig. 4.35 Makov prediction of focus position (Red) and WFS focused source position (Blue)

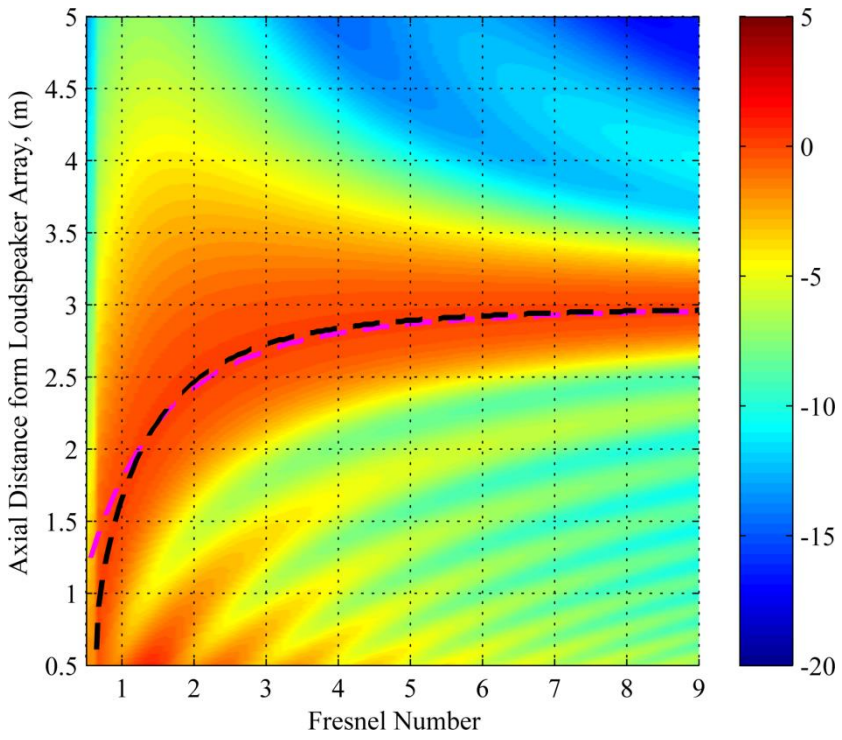


Fig. 4.36 RMS pressure in dB on axis, normal to the loudspeaker array. The dashed magenta line shows the pressure maximum corresponding to the focal point for each Fresnel number as predicted by *(4.49)*, the black dashed line shows the actual position of the pressure maxima in the WFS reproduced wave field. The pressure is normalised to the focus point. The array is *4m* long and the source is placed centrally, *3m* in front of the array

4.9.3 Correcting for the focal shift

Using the formula described here it is possible to provide a correction for the position error of the rendered focused source such that the position of pressure maxima is the same as the intended geometric focus position. This is done by rearranging *(4.49)* and recalling that $\tilde{z}_{Max}$ is the ratio of the pressure maximum to the geometric focus ($\tilde{z}_{Max} = z_{Max}/R$).

$$z_{Max} = \frac{R3\pi N_F}{2\pi N_F + \sqrt{\pi^2 N_F{}^2 + 72}} \qquad (4.53)$$

Where N_F is the Fresnel number given by *(4.41)*. Thus the correction in R_{corr} that is required to produce a pressure maximum at position defined by z_{Max} can be given as:

$$R_{corr} = \frac{z_{Max}\left(2\pi N_F + \sqrt{\pi^2 N_F{}^2 + 72}\right)}{3\pi N_F} \qquad (4.54)$$

So where, $R_{Intended}$ is the intended location of the source, the position to render the source to, R_{Render} is given as:

$$R_{Render} = R_{Intended} - R_{Corr} \qquad (4.55)$$

Applying *(4.54)* and *(4.55)*, the wave field produced by a *4m* long array with a virtual focused source positioned *3m* in front of it was modelled. The focus position was measured as the RMS pressure maximum on axis perpendicular to the loudspeaker array through the virtual source position. Fig. 4.37 shows the results of the simulation, with the focus point as predicted by the adapted Makov model from *(4.49)*, the measured focus point from the WFS array and also the corrected focus point. The focus point data is plotted with respect to frequency rather than Fresnel Number for clarity. The data shows that the source correction function is successful in correcting for the focal shift phenomena up to approximately *500Hz*. Below this frequency the improvements in the rendering begin to subside and errors in source position become large. It is apparent that the focal shift can only be corrected for to some extent as at very low frequencies, diffraction limits the possibility of focusing energy at all and therefore the focal point cannot be measured by a local pressure maxima. Correcting for the focal shift in this way has reduced position error at *1kHz* of nearly *0.3m*, suggesting the improvements are significant.

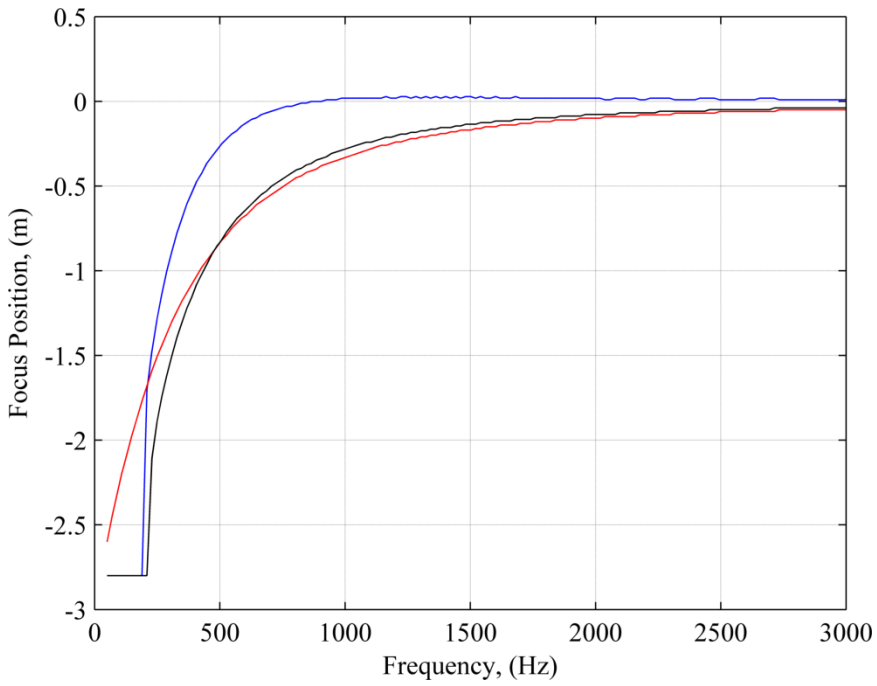


Fig. 4.37 Focal points as predicted by *(4.49)* (red trace), measured by RMS pressure maximum in reproduced wave field (black), corrected WFS focal points (blue)

As the Fresnel number is frequency dependent, the focal shift will vary with frequency which means that for a broadband signal, the pressure maximum will be in a frequency dependent position, causing a smearing in the amplitude distribution around the focus region. Applying this correction to the position of the intended virtual source helps alleviate this focal shift effect for a given frequency and thus increases the Q-factor of the amplitude peak at the source position (i.e. the RMS amplitude peak will be in the same place across all frequencies).

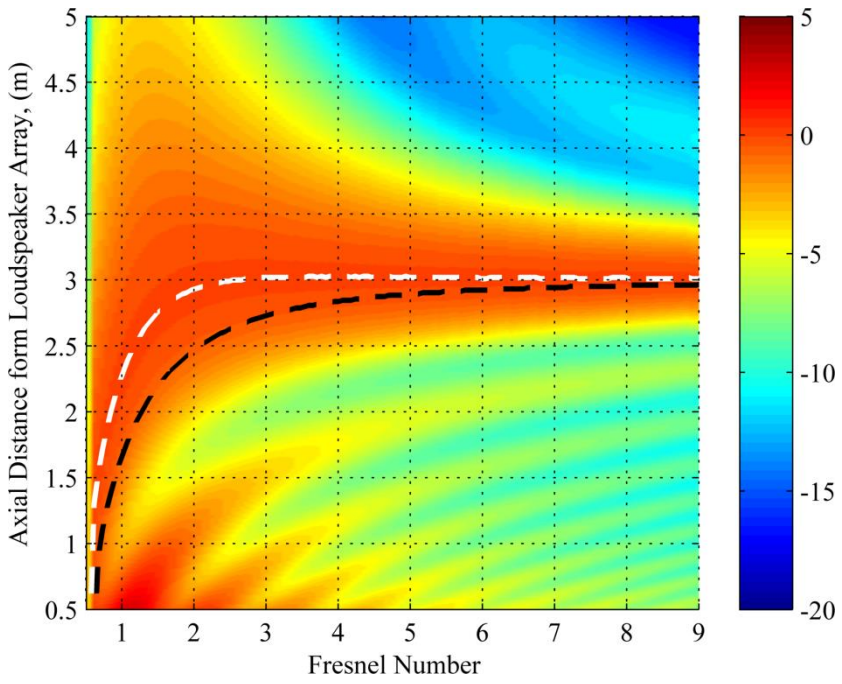


Fig. 4.38 RMS pressure in dB on axis, normal to the loudspeaker array, the black dashed line shows the position of the pressure maxima in the WFS reproduced wave field exhibiting focal shift. The white dashed line shows the max RMS pressure of the corrected wave field. The pressure is normalised to the focus point. The source is placed centrally, *3m* in front of a *4m* long array

Fig. 4.38 shows how the focal position of a focused source can be corrected for especially at low frequencies, although at very low Fresnel numbers (low frequency or short arrays) there are still limits on the ability of the loudspeaker array to focus the acoustic energy.

4.10 Pre-echoes

Inherent in the rendering of focused sources in WFS are the unwanted pre-echo artefacts that can cause spectral and temporal problems in the reproduced wave field. A pre-echo is defined as any energy that reaches the listener before the direct energy from the virtual source. Perceptibly these artefacts have been known to cause colouration or, 'phasiness' (Oldfield et al., 2010) of the virtual source, chirping (Geier et al., 2010) and temporal smearing; they can also introduce localisation errors as will be seen in the next chapter.

Pre-echoes are caused by spatial aliasing, truncation and causality artefacts. Each of these spatial aliasing and truncation artefacts are also seen in non-focused WFS

sources but in this case they arrive after the direct energy and as such are often masked. Due to the time reversal nature of focused sources however the artefacts in this case appear before the direct energy and therefore cannot be masked. This makes them perceptibly relevant and consequently they should be characterised as demonstrated in this section.

4.10.1 Causality effects

Using a WFS focused source to create a real point source is physically impossible to achieve using an array of loudspeakers at a remote location as a real system cannot be non-causal as such there will always be causality artefacts in the reproduced wave field. These artefacts are a result of the requirement that the loudspeakers at the edges of the array are activated before those in the middle of the array such that the sound is focused. The physical outworking of these causality effects is that at any point in the room, the listener will receive sound from the edge loudspeakers before, or potentially at the same time, as the sound from the virtual source. The impulse response at a given location will contain energy from the pre-echo artefacts and also the direct energy from the source as can be seen in Fig. 4.39.

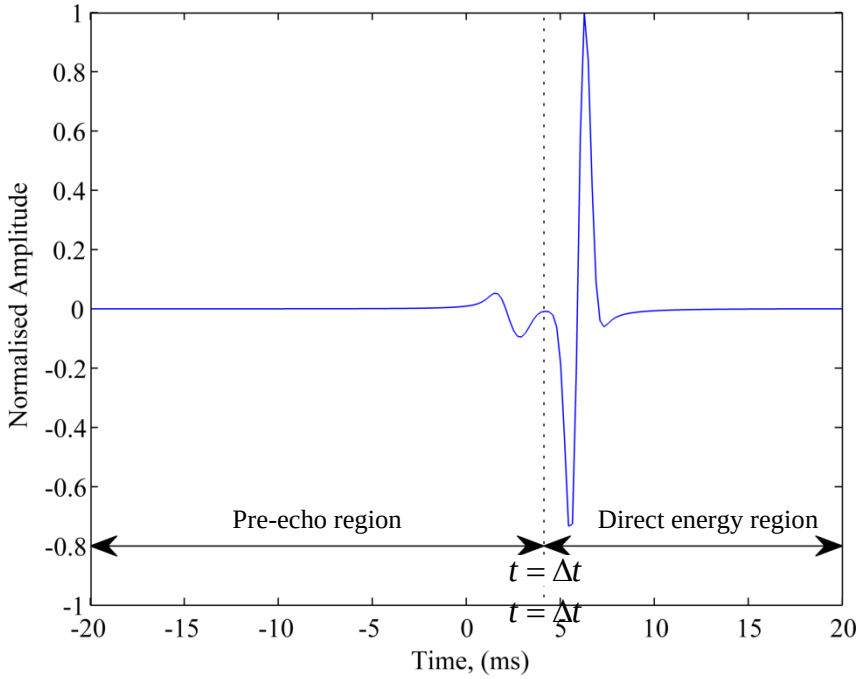


Fig. 4.39 Narrowband impulse response at position $[1m, 1m]$ with a source at $[0m, 0m]$, showing pre-echo energy before $t = \Delta t$

Causality artefacts like this are highly dependent upon the listener position both in terms of level and also in terms of the duration over which they occur as seen in section 4.10.4. Another important parameter is the array shape and dimensions. If the array is very long and contains many secondary sources, the temporal difference between the pre-echo artefacts and the direct sound will increase and their effect will be magnified as described in the following section.

4.10.2 Truncation

As previously stated, the truncation of the secondary source array results in additional sources from the edges of the array which distort the reproduced wave field and reduce the area over which correct localisation can be expected. For non-focused point sources, these artefacts appear after the initial wave front as can be seen in Fig. 3.24, but for a focused source the opposite is the case and the additional contributions from the edges of the array appear, before the direct energy, thus they constitute pre-echo contributions as shown in Fig. 4.40.

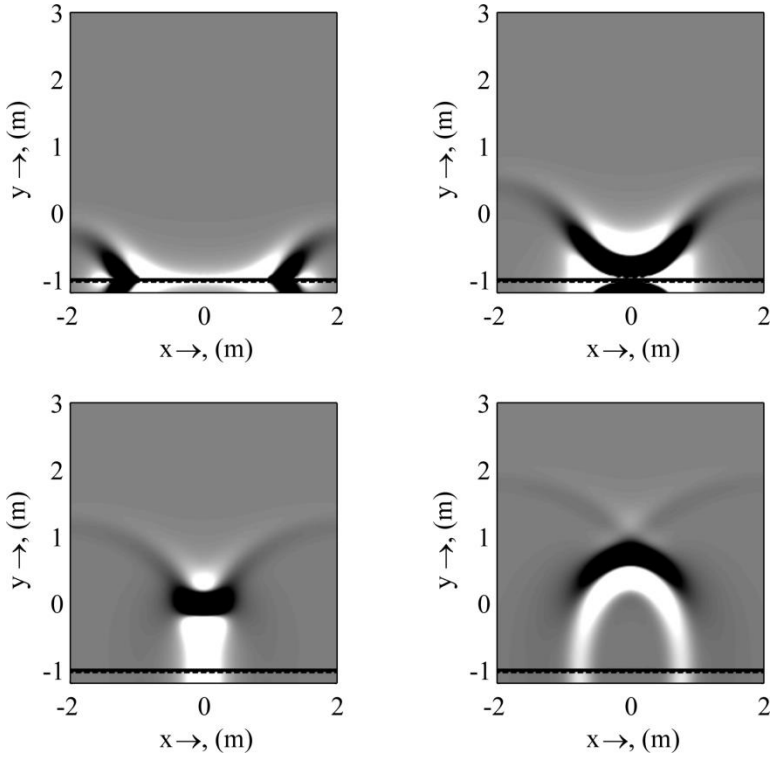


Fig. 4.40 Time evolution of a narrow band un-windowed focused source at $[0m, 0m]$, $\Delta_{LS} = 0.125m$, with a 2m array exhibiting pre-echo/truncation artefacts. Top left: $t = -3ms$, Top right: $t = -1.5ms$, Bottom left: $t = 0ms$, Bottom right: $t = 1.5ms$

These artefacts can once again be reduced to some extent with the use of windowing as shown in 3.5.3.1.

4.10.3 Spatial aliasing

The abovementioned causality effects are more problematic for a broadband scenario containing significant energy at frequencies above the spatial aliasing frequency as shown in Fig. 4.41, where the pre-echo contributions from each loudspeaker arrive at discrete time intervals. This results in a kind of comb filtering, causing colouration in the wave field and as a result they are particularly problematic both physically and perceptually.

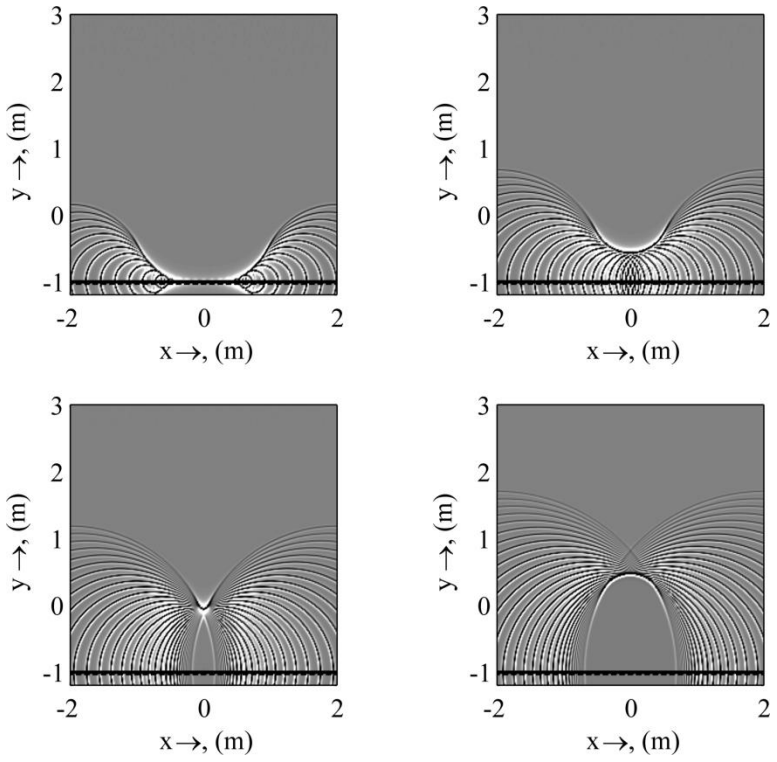


Fig. 4.41 Time evolution of a broadband focused source at $[0m, 0m]$ with $\Delta_{LS} = 0.125m$. The spatial aliased pre-echo artefacts are visible and arrive at the listener position before radiation from the primary source. Top left: *-3ms,* Top right: *-1.5ms,* Bottom left: *0ms,* Bottom right: *1.5ms*

Plotting the impulse response recorded at a single location in the reproduction space as shown in Fig. 4.42 demonstrate that spatial aliasing pre-echoes constitute a special case of causality artefacts where each loudspeaker signal is received at the listener

position at a discrete point in time, resulting in potentially unpleasant comb filtering and colouration as was seen in Fig. 4.25.

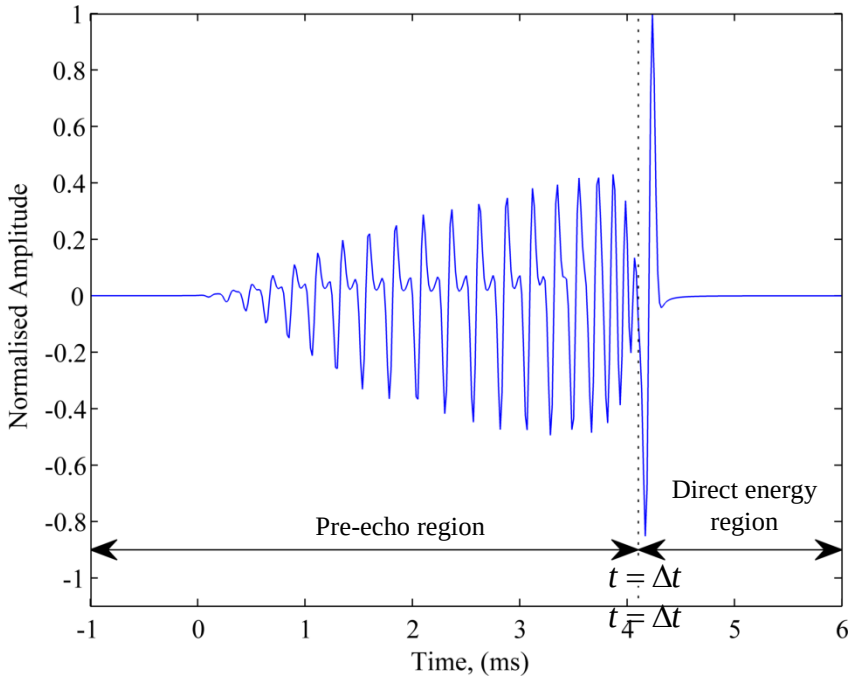


Fig. 4.42 Broadband impulse response at position $[1m,1m]$ with a source at $[0m,0m]$ produced with a *4m* array with a *12.5%* window, showing pre-echo energy before $t = \Delta t$

4.10.4 Direct to pre-echo ratio

To determine the severity of pre-echo artefacts it possible to calculate the ratio of the direct-to-pre-echo energy for a given listener position using a geometrical approach. The energy that arrives at or after $t = 0$ is a result of the primary source and provides only the correct localisation cues for the focused source. If the listener is in line with the virtual source and the edge loudspeaker, it is possible that the loudspeaker radiation reaches the listener at $t = 0$ as well but this is not considered to be a pre-echo as it will reinforce the focused source.

It can be seen from Fig. 4.43 that regardless of the listener position $\mathbf{x}_L$, the contribution from the primary source at position, $\mathbf{x}_S$ will always arrive after the contribution from the secondary source.

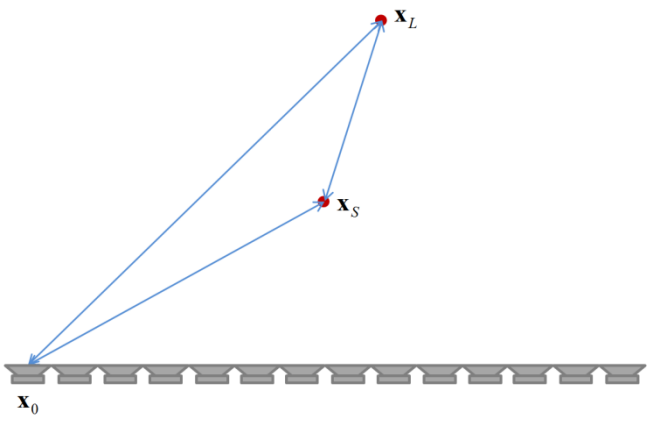


Fig. 4.43 Pre-echo source geometry, showing the primary source to listener distance is always greater than the secondary source to listener distance

From this, it is possible to separate the pre-echo contributions from the focused source contributions. If the focused source is generated at $t = 0$ then the contributions from the focused source will reach the listener at $t = \Delta t$ where:

$$\Delta t = \frac{\left|\mathbf{x}_L - \mathbf{x}_S\right|}{c} \tag{4.56}$$

Therefore any contributions at the listener location that arrive before $t = \Delta t$ can be attributed to pre-echoes or *causality* artefacts and any energy after Δt can be considered to be direct energy from the focused source. The ratio of the energy arriving before and after Δt results in the direct to pre-echo ratio which can be given as:

$$R = 10\log_{10}\frac{\int_{t\geq\Delta t} p_L(t)^2\,dt}{\int_{t<\Delta t} p_L(t)^2\,dt} \tag{4.57}$$

Applying this across the listening array results in a map of the pre-echo artefacts for a given source geometry as shown in Fig. 4.44.

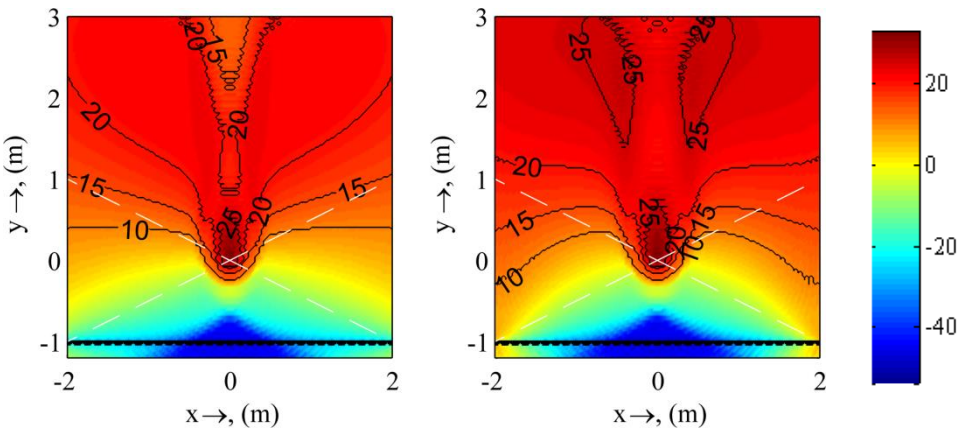


Fig. 4.44 Direct to pre-echo energy ratio in dB for a narrowband (un-aliased) focused source at $[0m, 0m]$ with a *4m* array. Left: Un-windowed, Right: Windowed (half-Hanning, *25%* of the array). The White dashed line shows the theoretical focused source view angle. The contours show the areas where the ratio is above *10*, *15*, *20* and *25dB*.

Fig. 4.44 shows the difference between the windowed and un-windowed array producing the focused source. The windowed array in the right hand plot exhibits a wider zone where the direct to pre-echo energy ratio is higher. So according to the precedence effect (see section 6.3.1.8), the localisation in this case, should be better.

A similar work has also been performed recently by (Song and Kim, 2011; Song et al., 2012). They claim that the levels of the pre-echo artefacts can be minimised for a specific listener location, by carefully selecting which loudspeakers should be active, a similar proposal was also suggested by (Wierstorf et al., 2010). The problem with this method is that WFS is inherently designed to be a multi-listener volumetric reproduction system and thus optimising it for a specific location may not be preferable in all cases. As was seen in section 4.2, the choice of active loudspeakers for the rendering of a focused source can affect the direction in which the source points and as described in section 4.9 could cause errors in the rendered source position, thus introducing this method of pre-echo suppression may result in directional sources pointing in the incorrect direction.

Another potential method for reducing the negative effects of pre-echoes could be to apply a frequency dependant windowing to the loudspeaker array such that the higher frequencies have a longer window than the lower frequencies. The longer

window results in a shorter active array so the temporal difference between the pre-echo and the direct energy would be reduced. The advantage of this method is that the colouration resulting from the spatially aliased pre-echoes can be reduced which is subjectively the most significant attribute of pre-echo artefacts. Using a shorter array at higher frequencies would result in a decrease in localisation accuracy as the wave front curvature would be incorrect and would introduce a spatial variation in frequency of the reproduced wave field, but the effects resulting in the changing in colouration as the listener moves around the listening space or if the source itself was moving would be reduced.

4.10.5 Localisation effects

The problem with the creation of focused sources is that, regardless of the listening position, the radiation from the secondary sources will reach the listener at the same time (if the primary source is on a direct line between the secondary source and the listener) or more generally, before the contribution from the intended primary source as can be seen by the wave field produced by a typical broadband focused source shown in Fig. 4.46.

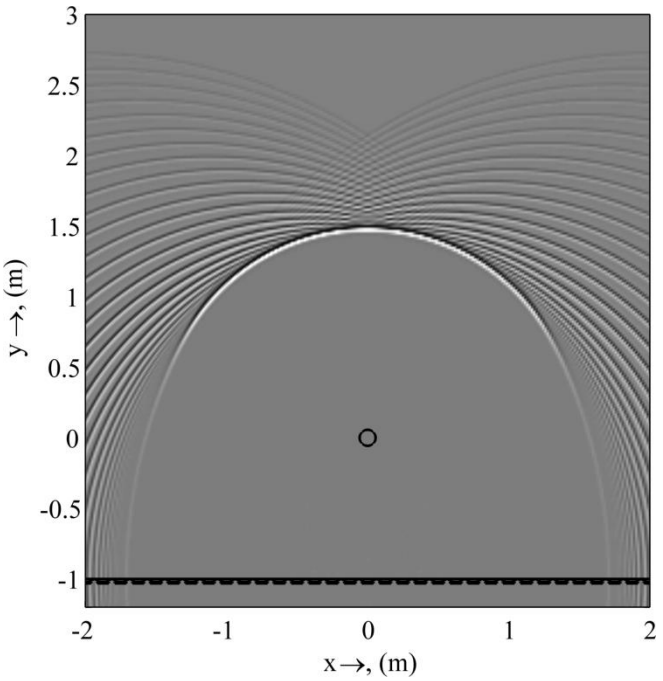


Fig. 4.45 Full band WFS demonstrating pre-echoes as a result of spatial aliasing for a focused source at $[0m, 0m]$

For normal (non-focused) WFS sources the opposite is true, with reproduction errors reaching the listener after the arrival of the direct sound (the virtual source) as shown in Fig. 4.46. Thus this case is not problematic in terms of confusing localisation cues but only in terms of colouration.

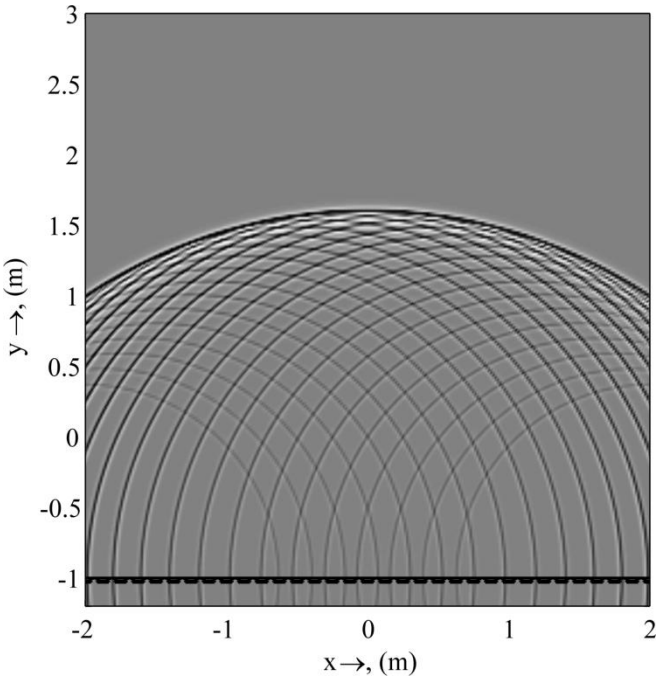


Fig. 4.46 Full band WFS demonstrating echoes as a result of spatial aliasing for an non-focused source at $[0m, -2m]$

Using the direct-to-pre-echo ratio it possible to make a rough prediction of the localisation quality of a focused source at a given listener position using the precedence effect. The precedence effect (Wallach et al., 1949) provides some explanation into the localisation of sounds in rooms or in environments where there are reflections present and states that humans localise sound based on the first arriving wave front providing that subsequent reflections arrive before *40ms* after the direct sound and that the subsequent reflections are not more than *~10-15dB* louder than the preceding sound. Thus the direct to pre-echo ratio map can help show regions of the listening space where these criteria are met.

This theory is of particular importance for the analysis of pre-echoes in WFS reproduction as the intention is that the listener localises the audio based not on the

first wave fronts which correspond to pre-echoes from the individual loudspeakers but rather on the later arriving energy from the virtual source. Thus the level of the pre-echoes needs to be significantly lower than that of the virtual source such that the localisation will be correct and not blurred. From Fig. 4.44 it can be seen that at different positions in the reproduction space the relative level of the direct sound and the pre-echo contributions is different. As the time difference between the direct sound and pre-echoes is less than *40ms,* according to the precedence effect it can be expected that the localisation performance of the system will be good for the regions where this ratio is greater than *15dB* and less good outside of these regions. The map of the direct-to-pre-echo ratio demonstrates therefore that the view angle of localisation approximates the geometrical view angle shown by the white lines on Fig. 4.44. This can also be seen from the results of the localisation tests described in chapter 5. It is also apparent from Fig. 4.44 that with the windowed array the localisation angle of the source increases as the plot exhibits a wider range of angles in which the direct-to-pre-echo ratio is higher.

Pre-echoes are most visibly obvious at frequencies above the spatial aliasing frequency as can be seen when comparing the narrowband case in Fig. 4.40 with the broadband case in Fig. 4.41. Pre-echoes are also perceptually more problematic above the aliasing frequency, for transient sounds and for sources reproduced with large arrays, as will be discussed in more detail in section 4.11. It has been shown (Geier et al., 2010) that aliased pre-echo contributions can result in chirping, smearing of transients, the perception of multiple sources and localisation confusion. The effects of spatial aliasing artefacts on the localisation of WFS virtual sources is still an open area of research, however when the direct-to-pre-echo ratio is plotted for the same source geometry as Fig. 4.44 but using a broadband source, the effects of the spatial aliasing can be seen clearly.

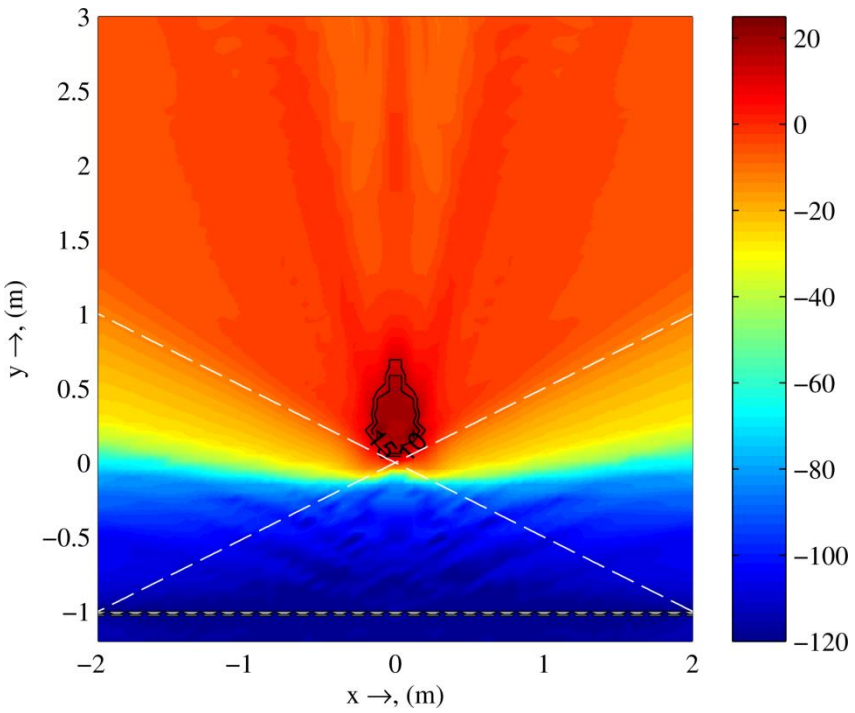


Fig. 4.47 Direct-to-pre-echo energy ratio in dB for a broadband (aliased) focused source at $[0m, 0m]$ with a *4m* array.

Fig. 4.47 demonstrates that the direct-to-pre-echo energy ratio is significantly lower than for the un-aliased case, suggesting that the precedence effect would cause errors in source localisation unless the listener was very close to the source or off axis from the source. This is congruous with the qualitative findings from Geier et al. The particularly low ratio of direct and pre-echo contributions may be due to the very large number of low amplitude pre-echoes rather than fewer pre-echoes with a very large amplitude, therefore localisation may not be as impaired as could be expected, which informal listening test suggest. The increase in spatial aliasing however will result in more localisation blur as pointed out by (Wittek, 2009).

4.11 Focused sources using large arrays

Unlike with non-focused sources, focused sources are not best produced by infinitely long arrays. In the case that the array is extremely long, the radiation from the pre-echoes not only cause temporal problems but also result in distortion in the wave front, manifesting as faint ripples as can be seen in the left hand plot in Fig. 4.48 even when a suitable windowing function is applied to the array. The effect of the ripples is relatively small and occurs only with arrays over about *15m* long. Longer

arrays do however result in a larger view angle as is evident from Fig. 4.48 and thus are able to reproduce focused sources with a more accurate wave front curvature.

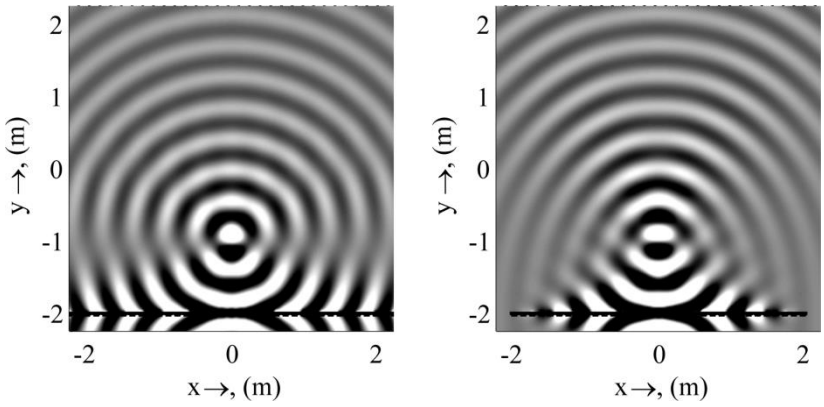


Fig. 4.48 $f = 800Hz$ focused sources at $[0m, -1m]$ with $\Delta_{LS} = 0.125m$. Left: *100m* array, Right *4m* array

The temporal artefacts produced by focused sources rendered with large arrays are particularly problematic due to the longer time discrepancy between the sound from the virtual source and the reproduction artefacts from the ends of the array (pre-echoes/truncation effects).

The reproduction of focused sources on a larger scale in large rooms or with long loudspeaker arrays can exhibit more severe errors due to the longer time discrepancy between the sound from the virtual source and the reproduction artefacts from the ends of the array (pre-echoes/truncation effects). To alleviate these problems it has been suggested by some authors (Wierstorf et al., 2010; Song et al., 2012) to use only a subset of loudspeakers for the reproduction. The problem with doing this for focused sources using large arrays, is that the application is usually for large listening spaces where there will be many people present and thus reducing the number of loudspeakers used for the reproduction, reduces the area over which correct localisation can be expected as can be inferred from Fig. 4.1 and seen from Fig. 4.48. It can also be expected that reducing the number of loudspeaker used for the reproduction of the focused source will result in a decrease in the positional accuracy of the source at low frequencies due to the focal shift phenomena described in section 4.9.

As the array gets larger the temporal difference between the direct energy and the pre-echo energy increases and as such eventually the precedence effect will result in people localising the pre-echoes rather than the direct source itself. The limiting factor of this can be determined using the direct-to-pre-echo ratio as described in section 4.10.4. The limiting temporal difference for the precedence effect is ~40ms which means that the difference in distance between the listener at $\mathbf{x}_L$ and focused source, $\mathbf{x}_S$ and the listener and the end of the loudspeaker array $\mathbf{x}_0$ must obey *(4.58)*.

$$\Delta t_{\max} = \frac{|\mathbf{x}_L - \mathbf{x}_S| - |\mathbf{x}_L - \mathbf{x}_0|}{c} < 40ms \qquad (4.58)$$

This means that as long as $|\mathbf{x}_L - \mathbf{x}_S| - |\mathbf{x}_L - \mathbf{x}_0| < 13.71m$ the precedence effect will prevent the pre-echoes from being localised instead of the intended source (assuming that the direct energy from the focused source is *10-15dB* higher that the energy from the pre-echoes.

Fig. 4.49 shows that for most of the listening space, the relative delay between the pre-echoes and the direct sound is within the required *40ms*. However, towards the edges of the array the delays are large and thus listeners may localise the individual pre-echoes and they may consequently be perceived as separate sources. It can be seen however that the in the region between the white lines on Fig. 4.49, the relative level between the direct and pre-echoes is high enough not to trigger the precedence effect and the delay between the respective contributions is lower than the *40ms*, this helps explain why (Geier et al., 2010) found that the performance of large arrays for producing WFS focused sources was not as bad as had previously been expected.

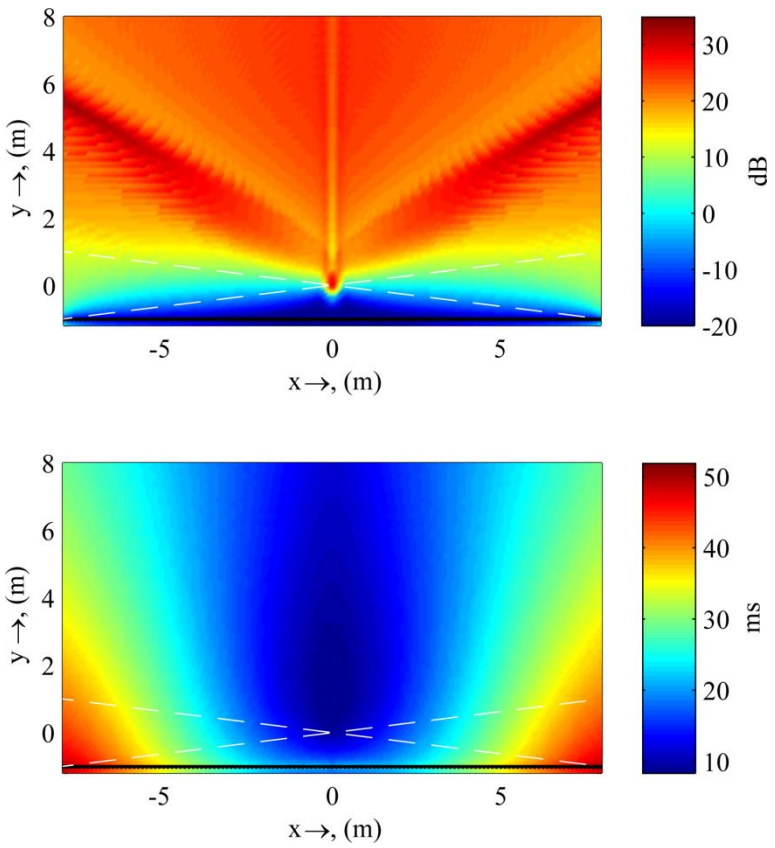


Fig. 4.49 Narrowband (un-aliased) focused source at $[0m, 0m]$ produced with a *16m* array with $\Delta_{LS} = 0.125m$. Top: Direct-to-pre-echo-ratio in *dB*. Bottom: Delay in *ms* between direct sound and pre-echo contributions

4.12 Multiple focused sources in the scene

Due to limitations in the view angle of focused sources, care has to be taken in the design of a reproduced audio scene/soundscape. There could arise the possibility where the listener could be in the divergent region of one focused source yet be in the convergent region of another focused source. Where this is the case, problems arise in terms of the loudness ratio between the two sources. The source of which the listener is in the convergent region of will be louder than it was designed to be and the localisation cues will be incorrect. This problem is difficult to negate without knowledge of the listener position (Melchior et al., 2008). However the scene can be designed to suit the most likely listener position and if necessary focused sources can be removed and instead rendered as non-focused point sources at a very small distance behind the loudspeaker array such that localisation is not impeded by the listener being in the convergent/shadow region. The direction in which the focused

source points can also be changed using the source selection criteria presented in section 4.2 such that the area in which it radiates into will best match the optimum listener position.

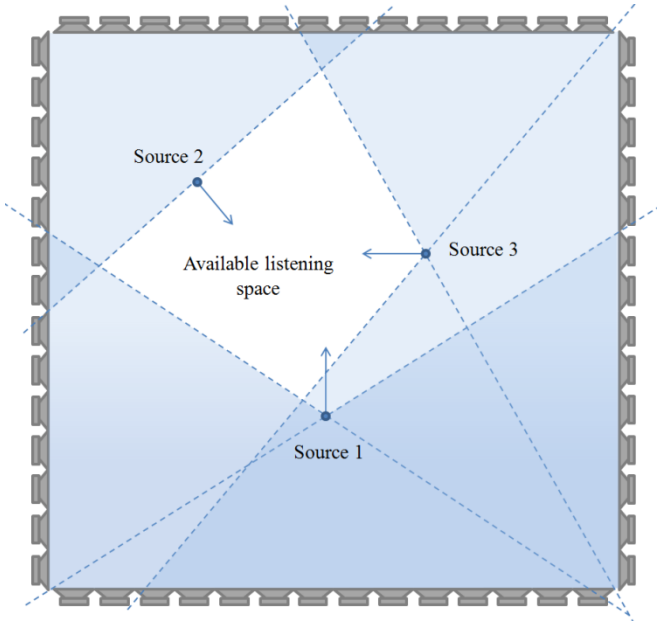


Fig. 4.50 Limited listening area as a result of multiple focused sources each with an individually defined view angle based on Fig. 4.1 and the source selection criteria in section 4.2

Fig. 4.1 earlier in this chapter demonstrates diagrammatically how the region in which all sources can be perceived accurately can be very small for sound scenes where there are multiple focused sources. If a greater number of sources are positioned in the scene there is a higher chance of the listener being in the convergent region of at least one of the sources. This situation is difficult to prevent if many sources are needed by the scene designer. If the region over which the source can be accurately localised can be increased, this can be less of a problem. One such technique of improving the focused source localisation view angle has been shown by (Oldfield et al., 2010) and will be demonstrated in chapter 5 of this thesis.

4.13 Moving focused sources

For moving focused sources, the source selection criteria as described in section 4.2 is of particular importance, as the source position is changing constantly with respect to the listeners, as such the listeners effectively move into the shadow region of the source where the wave field is convergent, thus as the source moves past the listener, the location of the sound source will suddenly flip providing an unusual and unnatural listening experience.

As shown in Fig. 4.51 there are four principle methods that can be adopted for the rendering of moving sources. The secondary sources can be selected such that:

1) The source points in the direction of travel
2) The source points at the centre of the room
3) The source points perpendicular to the direction of travel
4) The source points towards a tracked listener

Depending on the source selection protocol used here the listener will experience a different sensation of the focused source and may experience the source coming from the incorrect location not to mention errors in the Doppler Effect.

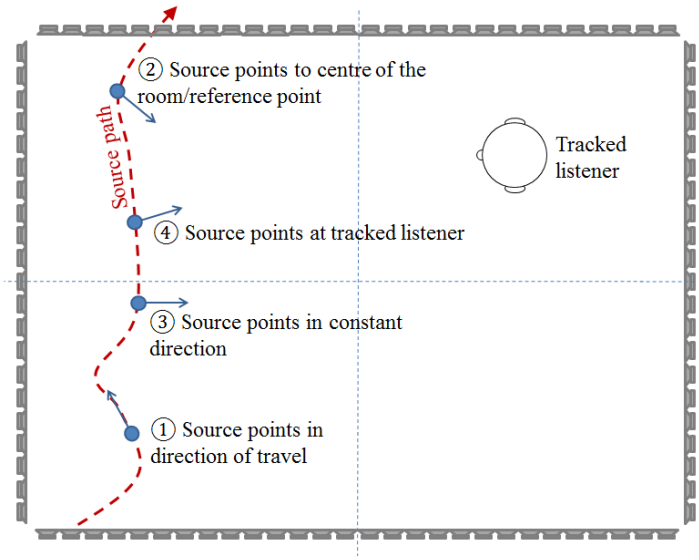


Fig. 4.51 Diagrammatic representation of the four primary methods of moving source rendering

For most practical WFS systems focused source are set to always point at the middle of the room (or a predefined area for optimal listening/sound field reproduction), in this case, as the source crosses the centre of the room or optimisation point the direction of the source will flip 180˚ and the sources on the opposite side of the array will become active. This secondary source flip usually takes the form of a fading/panning function between the sources such that the perception of the 180˚ flip is minimised for the listeners, but can be perceptually problematic in some cases.

The advantage of 1) is that there is no need to fade between loudspeakers on the other side of the array because it is always the same loudspeakers reproducing the source with additional sources becoming active gradually as the source moves. The disadvantage of this method is that the level normalisation is more complex and unwanted artefacts resulting from the switching on and off of the loudspeakers as the source moves in the form of clicks and pops can be undesirable and must be dealt with properly by using a gradual ramping function. A further disadvantage of this method is that the probability of a listener being in the shadow zone of the focused source is very high unless they are at either extremities of the listening space. Being in the shadow zone causes errors in the localisation of the source for static sources as described in chapter 5 and for moving sources there will also be amplitude errors introduced as the listener is in the shadow zone. These disadvantages mean that option 2) is often the best method of source selection for moving focused sources.

In some situations it may however be beneficial to employ option 3) for the source selection if for example the listener(s) location remains constant and the source moves in front of them like modelling a sound source on a stage as shown in Fig. 4.51.

Option 4) of pointing the source at a tracked listener (Melchior et al., 2008) will generally produce the best listening experience for the listener but as WFS is specifically aimed at scenarios where a volumetric approach to spatial sound reproduction is needed, it is less desirable to track listeners as this optimises the sound field for one point/person only.

It is not just important from a localisation perspective that moving focused sources are rendered correctly; a large problem with moving focused sources is also that the pre-echo artefacts that cause colouration at the listening position. These can vary greatly with listener location and as the source will be moving, the timbre of the source will constantly change for a given listener. It is difficult to alleviate these problems except by employing a low-pass filter cut off at the aliasing frequency (which clearly results in decreased timbral fidelity). As will be seen in chapter 6 however, the problems associated with these colouration artefacts are actually less disturbing when the reproduction takes place in a real (non-anechoic) room and therefore the anti-aliasing filters are generally not employed.

In summary, the secondary source selection for moving focused source is dependent on the reproduction scenario, most commonly the sources are optimised for the centre of the room but for the case of sources moving in front of a stationary audience it could be better to employ a source selection protocol such that the source always points perpendicular (or at a constant angle) to the direction of travel.

4.14 Summary

In this chapter, a set of driving functions has been derived from first principles demonstrating, monopoles, dipoles and pistonic focused sources. Where possible the sources were derived using a 2D representation of the primary source sound field such that the radiation characteristic match best the dimensionality of 2.5D synthesis. For this to be possible a frequency correction term is applied to the source driving functions to remove the spectral differences between 2D and 3D sources from the rendering.

The rendering of focused sources in WFS has been analysed with special attention paid to the spatial aliasing artefacts which near to the focal point are almost non-existent. This area of un-focused energy in the wave field has been quantified for the first time allowing the constraints of techniques such as local sound field synthesis to be quantified. Also analysed is the error in the position of the focused source with respect to the geometric focus. This error is greater at lower frequencies and smaller

loudspeaker arrays where diffraction becomes significant and prevents the system from focusing the energy at the geometric focal point. The effect has been modelled and it has been shown how to some extent this error can be corrected for by adding the required positional offset in the rendering.

This chapter has also re-visited and demonstrated the effects of truncation on the reproduced wave field for focused sources, and thus showed how these effects, combined with causality artefacts and spatial aliasing result in pre-echoes in the listening space.

Also briefly discussed are moving focused sources and the issues with the rendering thereof, the reproduction of focused source using long arrays and rendering multiple focused sources in a single sound scene.

5 Perceptual Aspects of Focused Sources

5.1 Introduction

This chapter describes a series of subjective tests used to evaluate some of the perceptual qualities of focused sources in wave field synthesis. There have been many subjective tests described in the literature, studying various perceptual attributes of WFS but as yet not much has been done specifically on the topic of focused source reproduction yet this is perhaps one of the most subjectively striking attributes of WFS and indeed one of the attributes of sound field synthesis techniques that constitute a significant advantage over traditional spatial audio techniques mentioned in Chapter 2 of this thesis.

The tests covered in this chapter include studies of distance and angular localisation in the horizontal plane for focused sources at different distances from a fixed listener position. A second test demonstrates how it is possible to increase the range of listener positions for which a focused source can be accurately localised by including the first order image source into the WFS rendering. For completeness, this chapter

also reviews the key published work in this area, summarising the significant results found therein.

5.2 Literature review

The first published subjective tests involving focused sources in WFS was done by Verheijen (1998). Verheijen used just two subjects for his localisation tests and adopted a simlar test as used by Hartman in his experiments on sound localisation in rooms (Hartman, 1983), processing the results based on the run standard deviation of the subjects' responses. The test stimuli were *2s* bursts of unfiltered pink noise. The stimuli were deliberately unfiltered so as to include the effects of spatial aliasing in the results. For his tests the subjects were allowed to freely move their heads as they localised each of the sources at the 25 test locations. Verheijen's test included both focused and non-focused sources and his conclusion was that there was no appreciable difference in the accuracy of the subjects' responses for focused sources as compared to non-focused sources. Due to hardware constraints, Verheijen's experiments did not employ the pre-filtering term $\sqrt{k/j}$, hence the sources would have had a low frequency emphasis. Verheijen repeated his experiments with a low pass filter to remove the effects of spatial aliasing and found that there was no appreciable difference in the results, suggesting that localisation is not affected much by pre-echo artefacts as is apparent from the investigation described in chapter 4 of this thesis. Verheijen noted that the results for focused sources exhibited a higher standard deviation (uncertainty) and therefore localisation is more reliable for normal WFS point sources. He does however warn against drawing rigid conclusions from these results due to the limited number of participants involved. Consequently there is scope for further investigation into the localisation of focused sources which is the motivation for the experiments described in this thesis.

Subsequent to the tests by Verheijen other experiments were done by Wittek (2009). Wittek's tests were designed to determine the effect of wave front curvature upon distance localisation and formed part of the work of developing his technique for enhancing the perceptual aspects of WFS by optimised phantom source imaging (OPSI) (Wittek et al., 2007b). Once again, for Wittek's tests, the subjects were able

to move their heads during the localisation tests. The subjects were asked to localise several test sources rendered using WFS and also real sources using discrete loudspeakers. All of the sources were close to the listeners as at larger distances the effects of wave front curvature are negligible. The tests were carried out in an anechoic chamber with stationary listeners with the sources placed perpendicular to the listener so to maximise the difference in the binaural cues.

Wittek's localisation experiments were not specifically analysing focused source preproduction although he did find that spatial aliasing causes a reduction in the locatedness (Rumsey, 2001) of WFS sources (including focused sources) and that WFS sources offer much better localisation-ability than stereo phantom sources but were not better than real sources.

Preliminary work by Spors commented briefly on some of the perceptual attributes of focused sources (Spors et al., 2009). This paper specifically mentions the effect of pre-echoes and presents some initial qualitative findings from a small set of subjective tests. Listeners were asked to describe the perceptual differences between WFS focused sources and real sources with specific attention to the colouration artefacts and changes in spatial impression. It was noted that listeners experienced audible artefacts as soon as the listening position was moved away from the central location. The artefacts that were noted were comb filtering, smearing of transients and chirping, it was noted that head movements had a large effect on the artefacts which was most likely due to the lack of a room used for the dynamic binaural resynthesis techniques (Lindau et al., 2007) they used for the tests. This is an approach whereby the secondary source signals are convolved with the HRTF corresponding to their position in space. The listener is tracked and the reproduction is done over headphones. The advantage of this technique is that it is possible to seamlessly switch between different rendering systems and very large systems can be tested without the need to set up hundreds of loudspeakers. Significantly they also mention the particular problems encountered when rendering focused sources using very large arrays as described in section 4.11.

Subsequent work from the same research group (Geier et al., 2010) again used a dynamic binaural resynthesis technique (Lindau et al., 2007) for the tests. These experiments provided a more detailed perceptual analysis and more concrete conclusions although much of the experimental work was qualitative. They used a panel of 25 expert listeners who noted that colouration, chirping, the perception of multiple sources and incorrect localisation when not in the central listening position. It was noted from these experiments that there is a region in the reproduction space that is perceptually artefact free but that this area was in fact very small. They also noted that the magnitude of these perceptual artefacts was greater for larger arrays although this effect was not as pronounced as expected.

As a result of these findings there has since been some attempts to improve the perceptual aspects of focused sources with the work by Wierstorf et al. (2010) and Oldfield et al. (2010). Wierstorf proposed some perceptual enhancements specifically aimed at the reproduction of focused sources using large scale arrays, pointing out that it could be perceptually beneficial to only perform the rendering using a subset of the available loudspeakers. As has been described in chapter 4 of this thesis, this has the undesirable effect of reducing the definition of the rendered source such that the focal spot is wider which also means that the distance at which the source is rendered to will be inaccurate. Oldfield's work is described in this chapter and aimed at improving the range of angles at which a focused source would be accurately localised by introducing the first order virtual reflections into the rendering and showed that indeed the so-called *view angle* of a focused source could thenceforth be increased.

5.3 Source colouration

One of the major problems with focused source reproduction is the colouration caused in particular by spatial aliasing artefacts. As shown in Chapter 4, there can be large discrepancies in the frequency response of the source at different listening positions. These colouration artefacts can become particularly perceptible when listeners move throughout the listening space and if the reproduction environment is acoustically dead. In the extreme case of an anechoic chamber many of the test

subjects who took part in the tests described below commented upon an apparent ‘phasiness’ as they walked around the listening space and even noticed the timbral character of the source stimulus change when it was moved to a different position in the room. These colouration artefacts are averaged out to some extent when the reproduction takes place in a more reverberant environment and pilot test studies revealed that in an environment such as the listening room system described in section 3.6.4, people no longer notice these colouration effects even with slow and fast moving sources. These findings are confirmed by the finite difference time domain models described in the next chapter where the effects of the reproduction room are analysed in the temporal, spatial and spectral domain.

Colouration causes problems as listeners move around the room as it becomes more apparent that the source is not behaving like a real source behaves in terms of the distance attenuation both in the high and low frequency region.

Some work has been done to try and alleviate some of the colouration introduced, for example, when using multi actuator panel (MAP) loudspeakers. These flat panel loudspeakers need additional frequency correction because of the complex interaction between the drivers at different points on the panel and consequently, complex multi-channel filtering algorithms have been presented for this purpose (Corteel et al., 2004; Spors et al., 2005)

For focused source reproduction however it is difficult to perform a spectral correction as the coloration is listener position dependent and therefore any improvement in one location will have a negative effect in another region of the listening space. However work done by Drumm and Oldfield (2013) shows how the frequency response can be corrected at a given receiver position by adopting a recursive filtering process. This approach could be applied to a small wireless microphone attached to a listener and the frequency response then corrected for that location.

5.4 Pre-echoes

The perception of pre-echo artefacts has been dealt with briefly in Chapter 4. There have been some recent attempts to categorise the perceptual effects of pre-echoes, based on the temporal separation of the pre-echoes from the direct energy and the difference in level between these artefacts and direct energy. Approaches have been presented in (Bianchi et al., 2012), (Song et al., 2012) and also in this thesis (see section 4.10.4). The approach adopted by Bianchi et al is interesting because it utilises the masking curves used as part of spatial audio coding algorithms.

The perceptual tests by Verheijen (1998), described above suggest that the effect of pre-echoes in the localisation of focused sources is not as great as would be expected, on the contrary the effects of pre-echoes, specifically from the spatially aliased contributions from the loudspeaker array seemed not to have a significant effect on the localisation performance of the subjects. To some extent, this has also been shown by Geier et al. (2010) who noted that when using the very large arrays for the rendering of focused sources, the effects artefacts perceived (which would be caused by pre-echoes) were not as severe as were expected.

The experiments described below include the effects of pre-echo artefacts to varying degrees as the source is moved around the listening position. When the source-receiver angle is greater the pre-echo component will be more significant and therefore affect the localisation performance.

5.5 Localisation test

This section describes a localisation test that was performed using the WFS system described in section 3.6.2 where the angle of the focused source with respect to the listener was changed. The purpose of this experiment was to test what range of angles a focused source could be accurately localised in and to act as a benchmark for the testing of subsequent improvements to the rendering techniques.

5.5.1 Introduction

First formally proposed by Verheijen (1998) it is often suggested that there is a certain angle for sources produced by WFS over which correct localisation can be expected, outside of this angle, errors in the wave front and consequently errors in localisation can be expected. This angle can be illustrated for both focused and non-focused sources by the diagram in Fig. 5.1. The extent of this window or *view angle* is dictated by the dimensions of the array and the position of the virtual source and consequently decreases if windowing/tapering is applied to the edges of the arrays as described in section 3.5.3.1.

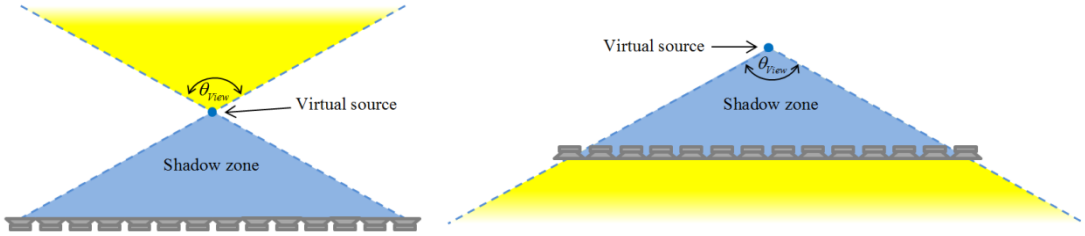


Fig. 5.1 Objective view angle for focused and non-focused WFS sources

The view angle defines the area over which the wave field will be accurately reproduced. The question that these subjective tests attempt to answer is: At what listener angle the localisation-ability of a focused source breaks down, i.e. what is the perceptual view angle of a focused source? A subjective test was performed in the anechoic chamber at the University of Salford with 10 participants. 25 sources were placed around the listeners and at 4 different radii. All of the sources were pointing directly straight ahead in the *+ve x*-direction, with the listener facing in the *–ve x*-direction. Thus the results provide data as to the polar response of localisation of such focused sources as the angle between the source and the listener changes with each source location.

The knowledge of the localisation view angle of focused sources helps in the definition of a sound scene as it enables the creator to know in which areas of the listening space the localisation will be correct. The results from the test also show how sources are localised when the listener is in the shadow region of the source i.e.

in between the source and the loudspeakers. A listener placed the 'wrong side' of a focused source like this will receive confusing and conflicting auditory cues with the wave field having the shape of a source radiating from within the reproduction space but a reversed propagation direction (a situation similar to the scattered sound from a convex reflector), this makes correct localisation difficult. Results show that as the angle of the focused source to the listener increases, the localisation accuracy decreases in a manner similar to the geometric case shown in Fig. 5.1.

The implication that focused sources can only be perceived accurately from in front of the source limits the use of such sources as it reduces the listening area for correct localisation and thus limits the number of listeners that will experience an accurate auditory scene. If, for example, two focused sources were placed propagating in different directions the only place where both sources could be accurately perceived would be between the two virtual sources such that the listener will be in front of both sources (as shown in Fig. 5.2), thus the possibilities of using multiple focused sources is greatly reduced. It is therefore important that the effects of focused sources as 'viewed' from all angles be determined.

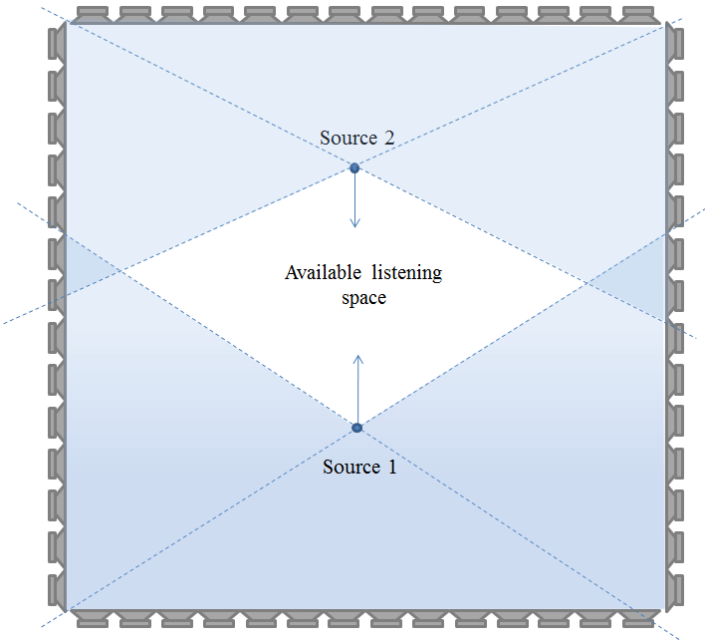


Fig. 5.2 Limited listening area when rendering multiple focused sources. The blue area depicts the shadow zone of the virtual sources

From the results it is possible to plot a 360° polar response graph of the localisation accuracy of focused sources. This can act as the benchmark from which improvements to the rendering of focused sources can be made (as described in section 5.6), it also provides a useful tool for sound scene creation as it shows where focused sources will be valid within the listening space. Anecdotal evidence from the test suggests that some subjects could discern an 'intimate' source but could not accurately localise it when the 'wrong side' of the source; the perception of such sources changes with listener orientation with some subjects localising image sources from behind the active loudspeaker arrays perpendicular to the direction they are facing due to the lateral energy from this direction. Also of interest is the area in the reproduction space where the listener transitions from being in front to being behind the virtual source, categorising this enables an analysis of the effective listening area of a focused source and the determination of the focused source *view angle*. The results provide interesting grounds for further research into the reproduction of focused sources to improve the evoked perception from all sides of the reproduction space. The possibility of using different directivity patterns of the source and employing windowing functions to the active loudspeaker arrays is suggested; this may decrease the sharpness of the virtual source but it is hoped will increase the listening area and consequently allow a greater number of users to experience the focused source(s).

5.5.2 Test Description

The listening tests were performed in an anechoic chamber with a cut off frequency of *100Hz*. Listeners were sat in the middle of the chamber surrounded by a $3m \times 4m$ rectangular array of 104 loudspeakers as shown in Fig. 5.3. An acoustically transparent sheet was set up in front of the loudspeakers to make a square 'room' measuring $2.75m \times 2.75m$ surrounding the listener, the purpose of the sheet was to ensure a natural response from the subjects rather than being affected by the visual cues introduced by loudspeakers. The acoustically transparent sheet also helps reduce the effects of the visually unnatural environment of the anechoic chamber. Additionally, the creation of a virtual room in this manner provided the correct

environment to test the effects of adding the first order image sources as described in section 5.6.1.

The loudspeakers were *1.45m* from the floor such that they were ear height for the participants when sat on the stool at the listener position.

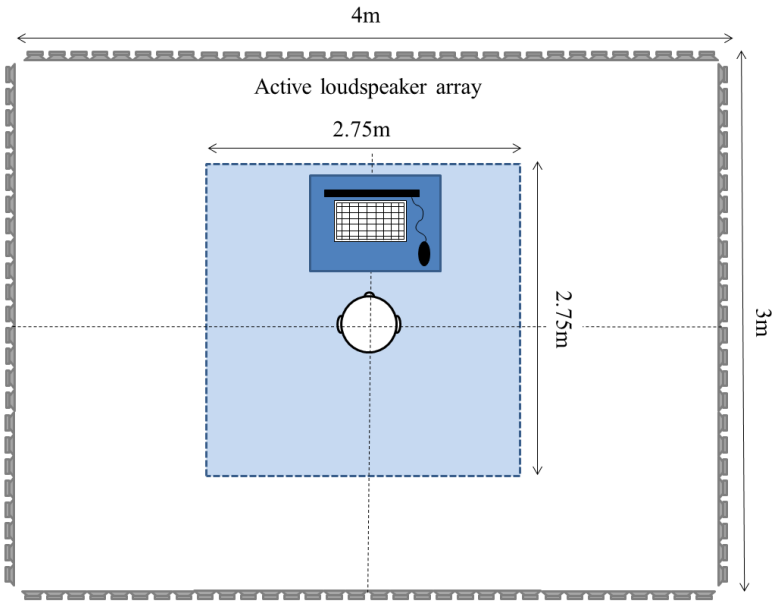


Fig. 5.3 Test set up for initial localisation test

5.5.2.1 Audio System

The audio system used for the WFS reproduction consisted of 2 RME HDSPe MADI 64 channel sound cards controlled by a single dual core PC. Each of the two MADI signals (64 channels encoded in a digital format down a BNC cable) from the sound cards were converted into 8 ADAT signals using two RME ADI-648 modules; each of these 8 ADAT signals were then converted into audio signals using RME ADI-8DS modules. These were amplified by Cloud CX A850 8-channel power amplifiers. The loudspeakers were Kef UniQ Ci100QR normally using as ceiling loudspeakers but chosen for this application because of the resulting stable phase and spatial responses due to the dual concentric design. More details of the hardware setup can be found in section 3.6.

The audio signals for each channel were controlled from a multichannel API .dll wrapped in MATLAB® function. Each of the test signals were pre-rendered as a

multichannel .wav file with each audio channel being sent to its corresponding channel on the soundcard to eliminate the need of real time processing.

5.5.2.2 Test procedure

Each test participant sat in the anechoic chamber with a laptop linked to the main control PC via a network cable so all the amps and control PC were outside the anechoic chamber to remove additional sound sources. The test participant controlled when the next source would be played and then clicked the location on the localisation map on the screen of the laptop where they perceived the source to be coming from and recorded with a slider the perceived distance of the source.

5.5.2.3 Reporting the Results

In order to retrieve the results of the test a user interface was constructed in MATLAB®. The user was asked to position the sound source on a circle on a 2-dimensional Cartesian plane and was then asked to report the distance of the source using a slider. The user interface is shown in Fig. 5.4.

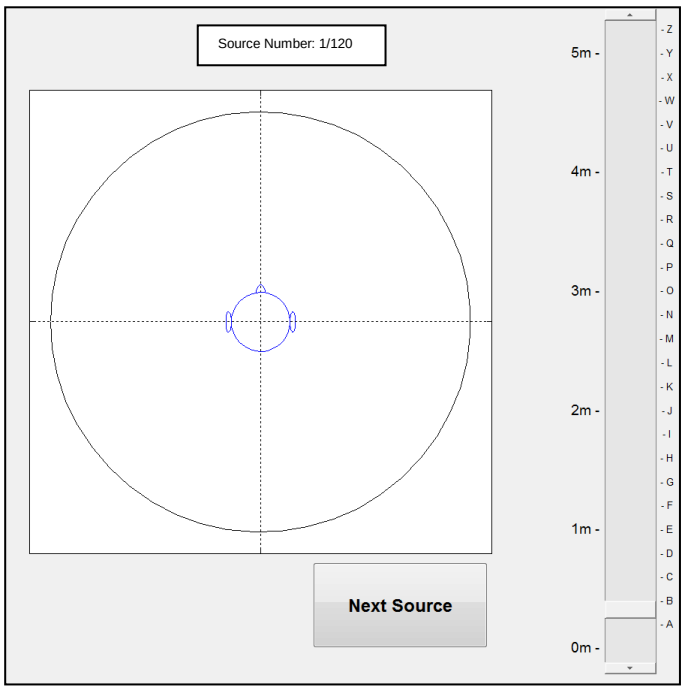


Fig. 5.4 User interface for subjective tests

By clicking on the circumference of the circle, the angular error in each participant's response was kept the same. For the distance perception, the letters *A-Z* were

included on the slider to help the participants scale their results. A line was put up in the listening space with the distances *A-F* marked on to provide the participant with an anchor to aid their distance evaluation. Labeling the line with letters rather than with numbers was decided to prevent people from recording the source distance based on their pre-conceived judgment of distances and so they felt at liberty to choose letters up to *Z* even though it was only *A-Z* that were visible in the room.

5.5.2.4 Test Panel

The initial test panel consisted of 10 participants, half of which were experienced in listening to surround sound and had taken part in sound localisation tests previously. The test took on average 14 minutes to complete, this was kept to a minimum to prevent listener fatigue affecting the results. Participants were told to face forward but were at liberty to move their heads during the test and were given a short training session of 5 sample sources to familiarise themselves with the test procedure.

5.5.2.5 Test Sources

The test was done using both noise bursts and also a speech stimulus. Broadband noise is a good test signal as it achieves optimal localisation performance as it offers a full range of cues to the listener especially with a short attack and decay ramp applied such that the transients provide additional information. Noise, however is only an approximation of a real signal and would not likely be the source type listened to must often by such a system. Additionally a speech source was also used for the tests as this constitutes a more ‘real world’ source.

The noise source consisted of two bursts of unfiltered pink noise of 1 second duration with a *0.4s* silence in between, the onset and offset transients of the signal were *0.1s* long. The envelope of the noise source is shown in Fig. 5.5.

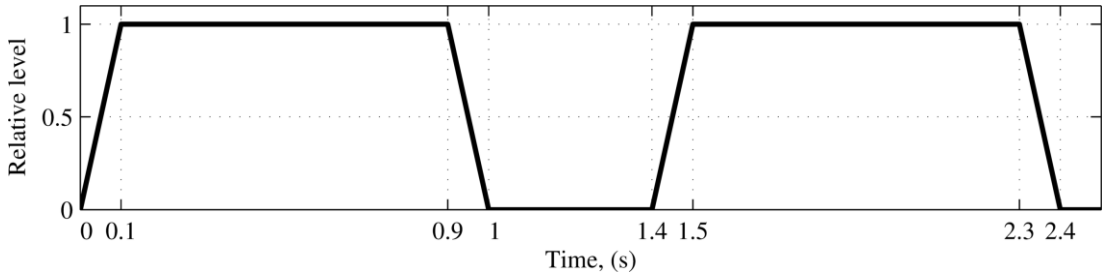


Fig. 5.5 Envelope of the pink noise test signal

The speech signal was a short sentence taken from the Hearing In Noise Test (HINT) (Nilsson et al., 1994):

"The yellow pears taste good"

In order to achieve useful results each source was reproduced identically such that the objective view angle was kept the same despite the distance from the loudspeaker array changing, this meant selecting a different set of loudspeakers for each source. By doing this each source could be considered identical and only the listener position with respect to the source was changed each time despite the listener remaining in the same location.

Small changes in the source position were introduced such that each source would be in the middle of the chosen loudspeakers to prevent the asymmetrical effects described in section 4.2. The source selection criteria insured that the geometric view angle was 70° for each source. Nominally the chosen angles were every 7.5^o but spread over both sides of the circle such that the angular separation was 15^o with a 7.5^o shift at 180^o.

The test consisted of 100 virtual sources positioned in circles around the listener at 4 different radii (*0.45m, 0.6m, 0.75m* and *0.9m*). The geometry of the test source and the loudspeakers is shown in Fig. 5.6.

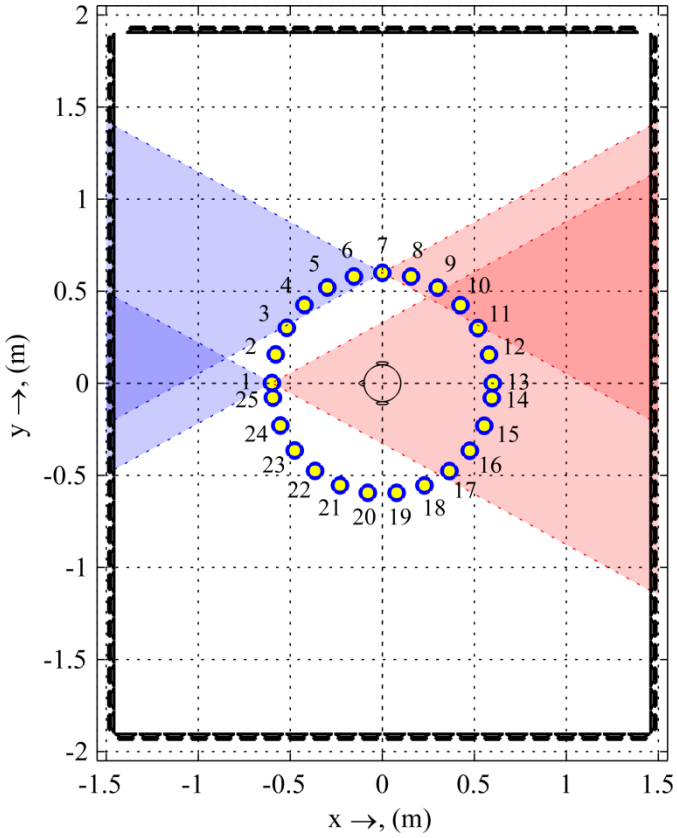


Fig. 5.6 Virtual source geometry and secondary source selection for constant view angle of 70˚

The direction that the focused sources point in was chosen to be in the *+ve x*-direction for all angles, thus only the array in front of the listener was active for each source. This meant that for each source position, it was the angle with respect to the listener that was changing but without the test participants having to move. Despite only the front array being active, the other loudspeakers however remained in place and connected up so participants would not show a bias in their responses to the front of the room.

The source order was randomised for each listener to prevent any bias to specific results and from listener fatigue having an overall effect on the results. For each of the virtual sources the number of active loudspeakers was chosen such that the view angle of was *70*°. A cos^2 window is applied to the last *20%* loudspeakers in the array as suggested by Start (1997).

5.5.3 Results

5.5.3.1 Noise

The first test source signal was the pink noise burst, for which the results are shown below for the 4 different source radii. The results are presented as reported angle versus the true virtual source angle for each source radius.

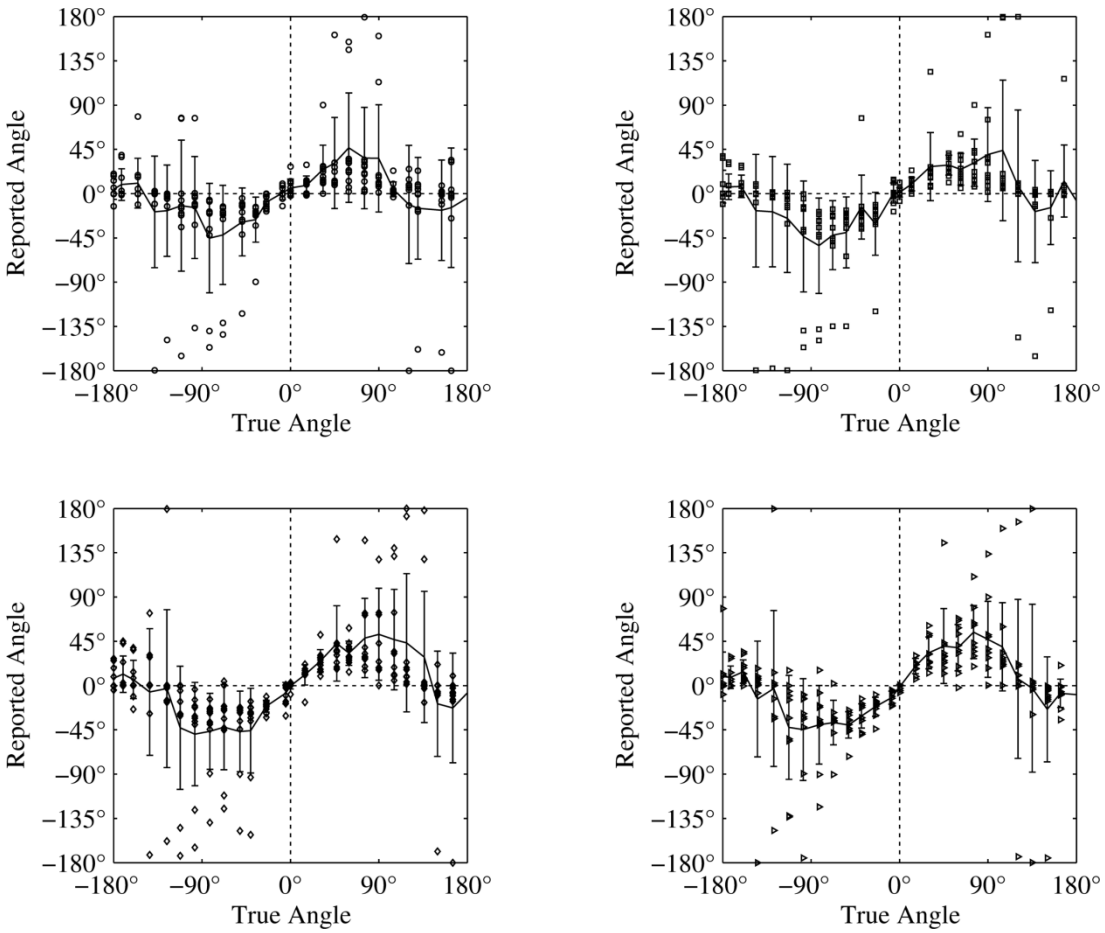


Fig. 5.7 Reported angle versus true angle of all 10 subjects for *noise* sources at a radius of *0.45m* (top left), *0.6m* (top right), *0.75m* (bottom left) and *0.9m* (bottom right). The error bars show the standard deviation of the reported angles.

The results show that for small virtual source angles most subjects localise the source correctly and the standard deviation of the responses is small. There is however a large standard deviation, especially at larger source-listener angles, this is due to a number of outliers resulting from the conflicting cues present when attempting to localise focused sources outside of the geometric view angle. It can be seen from Fig. 5.7 for example that some people managed to localise the sound

correctly even when they were substantially outside of the correct view angle, whilst others localised a source coming from the front array. Because of this high standard deviation the median was chosen to make an average as this reduces the effects of the outliers and can give a more faithful representation of the trend of the results. Plotting the median lines of all distances on the same set of axes demonstrates how the view angle of localisation changes with distance. Fig. 5.9 shows the data points of the median results per angle with the circles, triangles, diamonds and stars corresponding to the median scores per angle for source distances *0.45m, 0.6m, 0.75m* and *0.9m* respectively. A 5th order polynomial curve was fitted to the data points in a least squares sense and is plotted as the lines on the graph for clarity. Fig. 5.9 shows how the view angle of localisation accuracy is not constant with source distance but increases with listener distance.

It can also be helpful to view the data in the form of angular error for each source position as shown in Fig. 5.8.

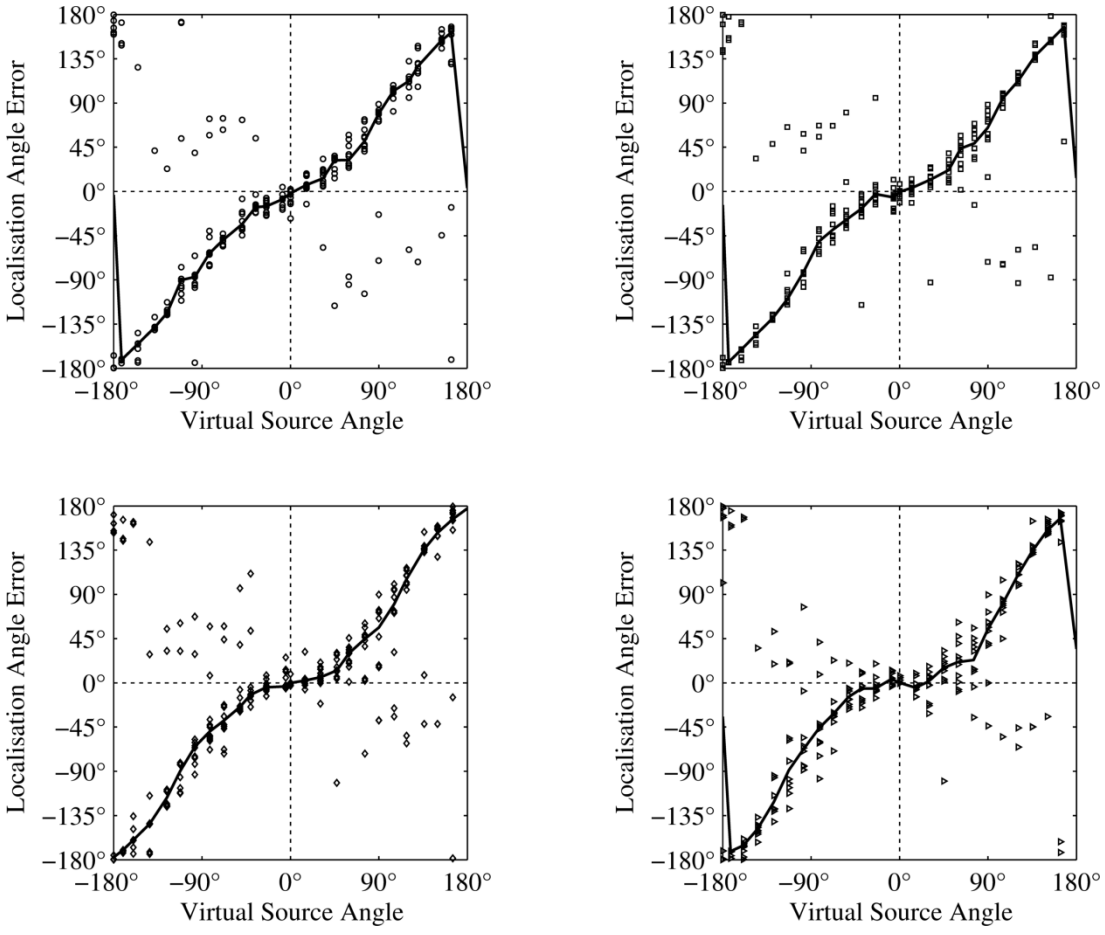


Fig. 5.8 Localisation angle error versus true angle of all 10 subjects for *noise* sources at a radius of *0.45m* (top left), *0.6m* (top right), *0.75m* (bottom left) and *0.9m* (bottom right)

Fig. 5.8 exhibits a flat region for smaller virtual source angles corresponding to the region over which the localisation error is smallest. As the virtual source distance is increased, the range of angles over which this flat region extends increases, showing that the subjective view angle increases with increasing source distance. Plotting the median results data with a 5th order polynomial curve fitting show the relationship between the results of each source distance as shown in Fig. 5.9.

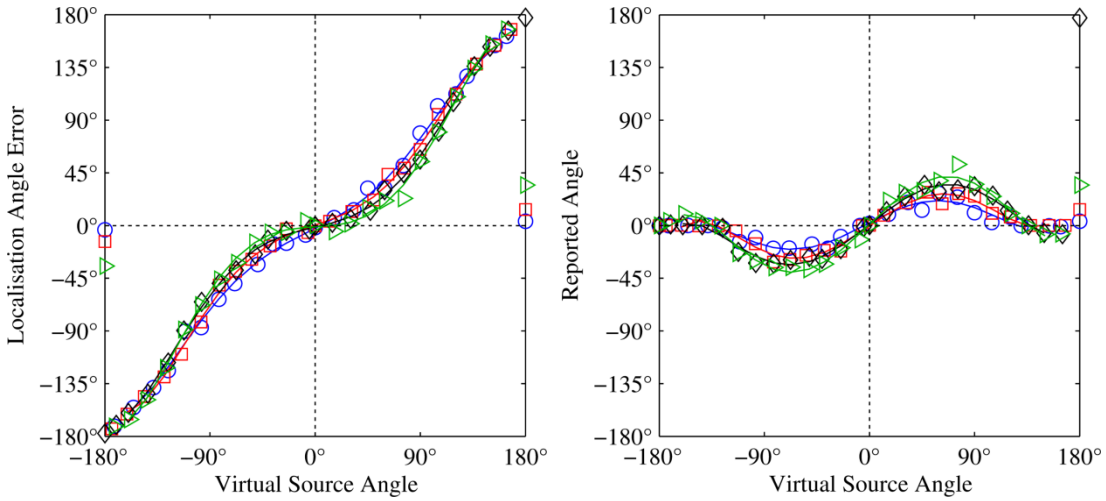


Fig. 5.9 Left: Median localisation error versus true angle, Right: Reported angle versus true angle. Both with 5th order polynomial curve fitting for the *noise* source at different distances. Blue 0.45m, Red: 0.6m, Black: 0.75m, Green: 0.9m. In each case, the line is the best fit line and the markers denote the median values

5.5.3.2 Speech

The same test was performed using the speech source, for which the same set of curves can be plotted for comparison between the two source signals.

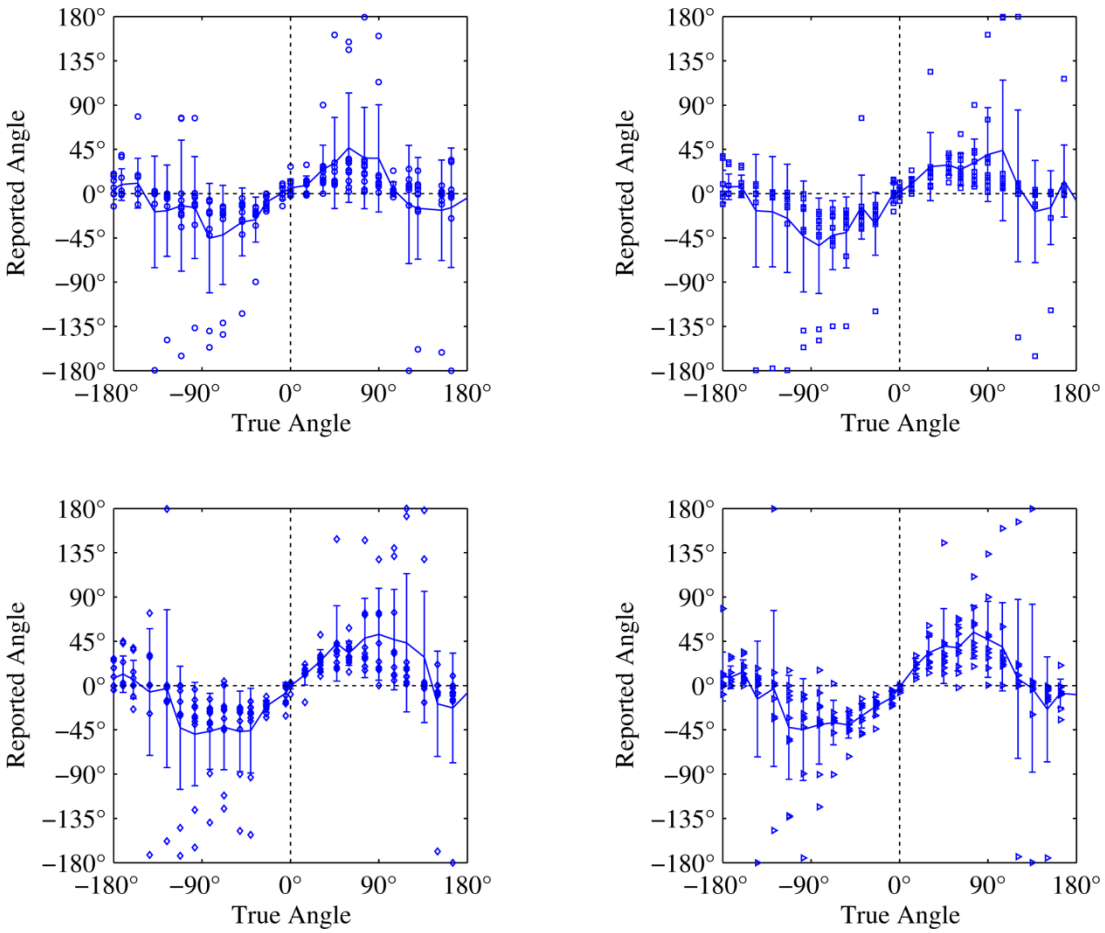


Fig. 5.10 Reported angle versus true angle of all 10 subjects for *speech* sources at a radius of *0.45m* (top left), *0.6m* (top right), *0.75m* (bottom left) and *0.9m* (bottom right). The error bars show the standard deviation of the reported angles.

The results for the speech source reveal a similar trend to that of the noise source, with the subjective view angle increasing as a function of source distance. It is apparent when comparing the results for the different source types that the speech source results in a greater standard deviation at larger angles (as shown in Fig. 5.11), suggesting that participants found it harder to localise the speech source than the noise source. This is in congruence with research in this area (Blauert, 1997; Wittek, 2009).

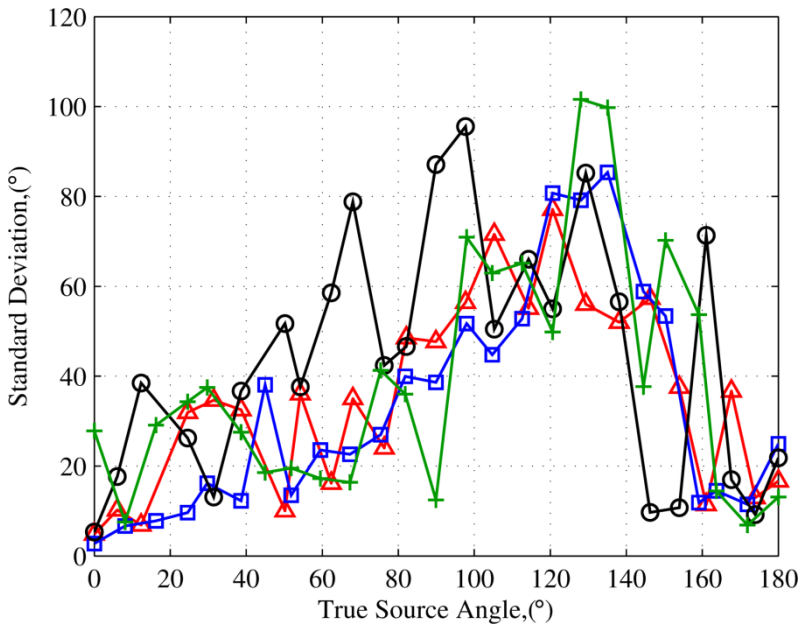


Fig. 5.11 Standard deviation of localisation results versus absolute virtual source angle for both test signal types. Blue: Noise at 0.6m, Red: Noise at 0.9m, Black: Speech at 0.6m, Green: Speech at 0.9m.

Fig. 5.11 not only demonstrates the increased uncertainty in the localisation of speech with respect to noise but also demonstrates how the standard deviation in the participants' responses increases as the absolute source angle increases. Interestingly, when the source is positioned at 180˚ (behind the listener), most of the participants localised the source as coming from in front of them and thus the standard deviation is decreased in this region. This front to back confusion is common in localisation tests as the source is placed in the so-called cone of confusion (see Fig. 2.3b). The reason for this is the very limited number of cues available to judge between sounds coming from in front or behind as described in section 2.2.1.

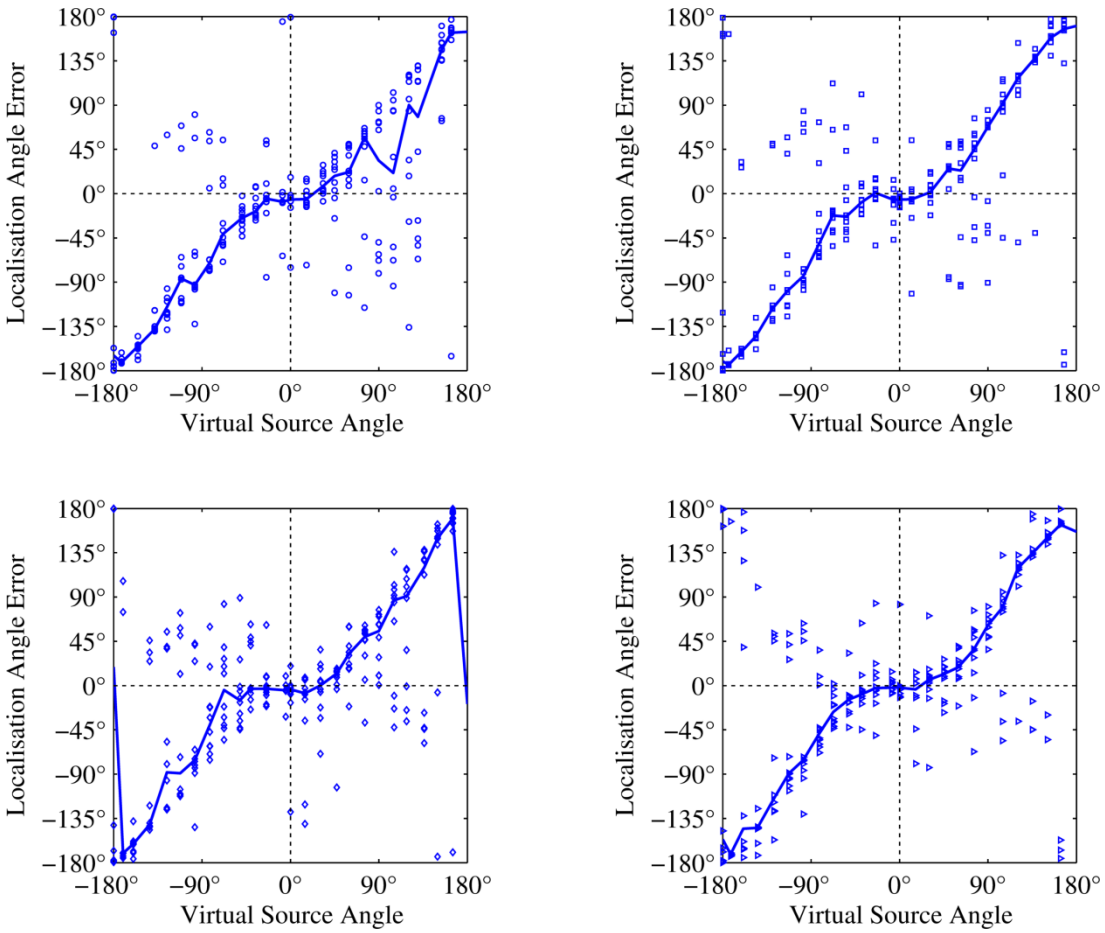


Fig. 5.12 Localisation angle error versus true angle of all 10 subjects for *speech* sources at a radius of *0.45m* (top left), *0.6m* (top right), *0.75m* (bottom left) and *0.9m* (bottom right)

Once again, the curve fitted median results are plotted on the same axes such that the relationship between the different test conditions can be compared more easily revealing the same trend as for the results with the noise source.

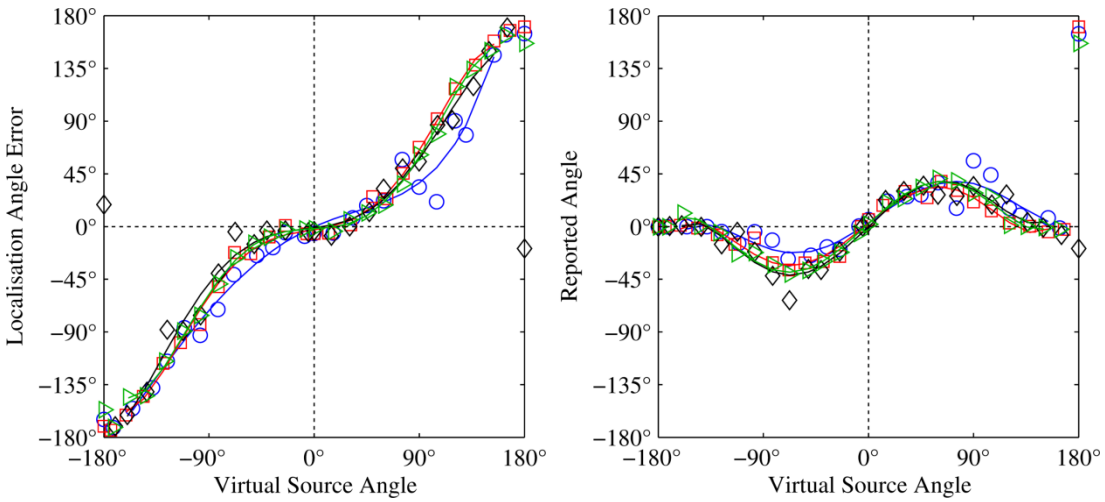


Fig. 5.13 Left: Median localisation error versus true angle, Right: Reported angle versus true angle. Both with 5th order polynomial curve fitting for the *speech* source at different distances. Blue 0.45m, Red: 0.6m, Black: 0.75m, Green: 0.9m. In each case, the line is the best fit line and the markers denote the median values

5.5.3.3 Distance perception

The results from the distance perception test reveal that there are large errors in the localisation for all virtual source distances. For each of the source distances evaluated, the minimum error occurs at approximately 180° where the listener is actually in the convergent region of the focused source, which is rendered behind them. This is most likely due to the increased amplitude of these sources with respect to the others as the listener is in the convergent rather than the divergent region.

The finding concurs with other research in this area (Wittek, 2009), that the wave front curvature alone does not provide enough auditory cues for an accurate localisation. For an accurate distance perception, listener movement is an important cue. For these experiments, the listeners were seated at a fixed position throughout the tests and are hence relying upon a limited set of distance cues (as described in section 2.2.3). Primarily the localisation cues here are wave front curvature and small spectral changes resulting from reflections from the pinna. Fig. 5.14 shows the results of the distance perception test using the noise source. Plotting the mean lines on these plots reveals that most subjects failed to accurately identify the distance although most subjects did in fact place the source in front of the loudspeaker array in each case.

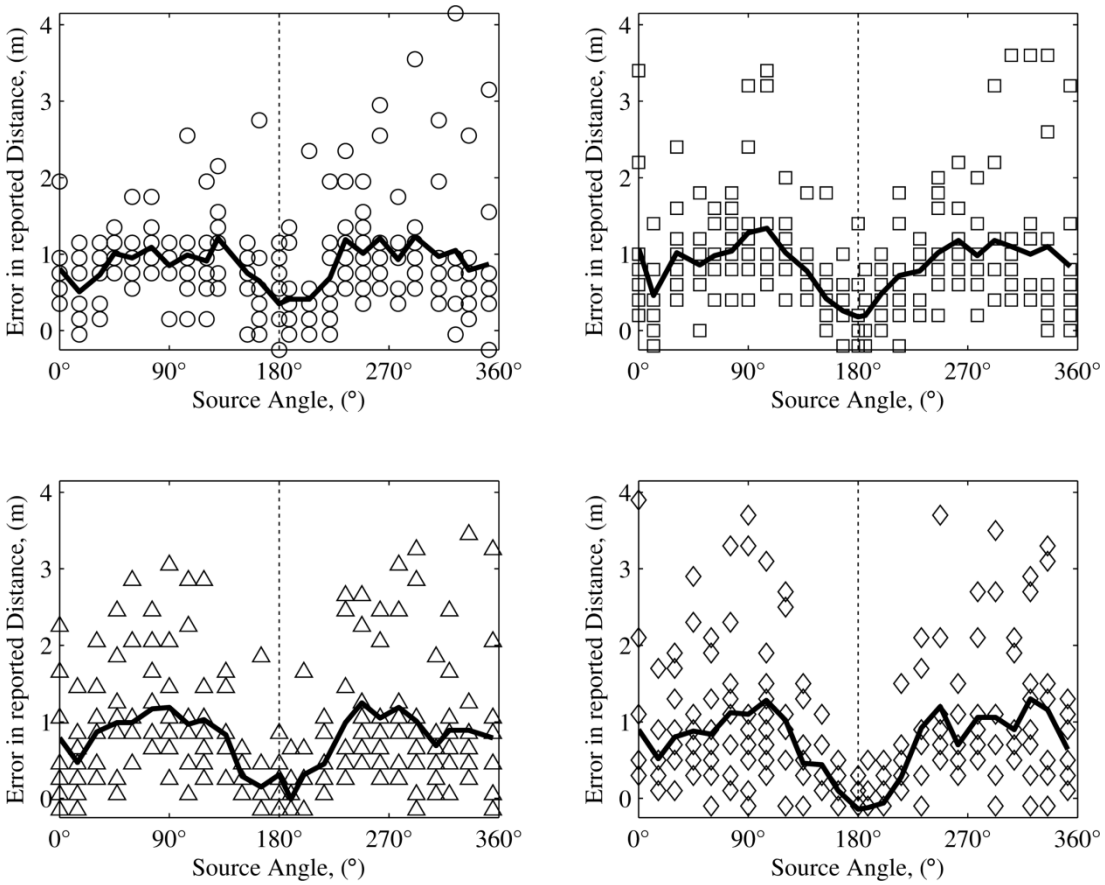


Fig. 5.14 Results from the distance perception tests, for source distances of 0.45m (top left), 0.6m (top right), 0.75m (bottom left), 0.9m (bottom right) with a *noise* source.

The same is true for the distance localisation results for the speech source as can be seen in Fig. 5.15. There is again no clear trend and no obvious difference between participants' ability to localise the source depth for different virtual source distances.

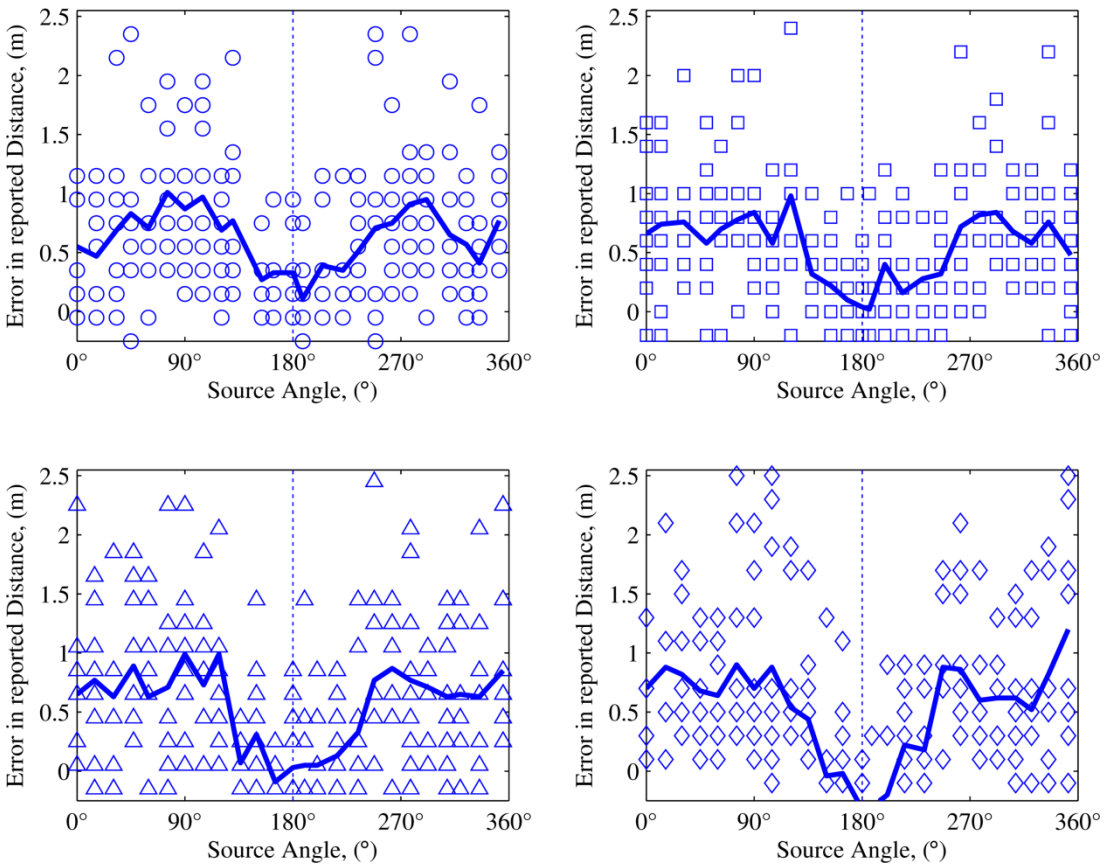


Fig. 5.15 Results from the distance perception tests, for source distances of *0.45m* (top left), *0.6m* (top right), *0.75m* (bottom left), *0.9m* (bottom right) with a *speech* source.

Despite providing some visual markers within the listening space, distance perception is often helped by the associated visual cues from the source. Thus localising a source within a room where there is no visual stimulus and no reflections constitutes an unusual scenario and may cause people to localise sources from behind the array as would be the case for traditional listening tests.

5.5.4 Discussion

These results of the horizontal localisation test show that the subjective view angle is not always the same as the objective view angle where the wave field can be considered to be reproduced correctly; this is unsurprising as humans don't necessarily need to have all of the correct cues in order to be able to localise correctly (hence even stereophonic reproduction is very effective). The subjective view angle for localisation however changes with distance from the source. The

localisation accuracy of a focused source decreases at smaller distances from source to receiver.

As the distance increases the rate at which the view angle increases with distance decreases, presumably asymptotically towards a value governed by the length of the array. Due to the size of the anechoic chamber it was not possible to render focused sources much further away from the listener than *0.9m* therefore this could not be tested. This finding is congruous with the results shown in the previous chapter, that the direct-to-pre-echo ratio (shown in Fig. 4.44) exhibits an increase in the angle over which this ratio is in excess of *15dB* as the distance from the source increases. This suggests that the localisation improves as the direct-to-pre-echo ratio increases.

The standard deviation of the reported results from all listeners is greater for the speech source than for the noise, this is most likely due to the transient nature of the speech signal emphasising the pre-echoes from the loudspeaker array and causing localisation confusion. The effect of pre-echoes on the noise stimulus doesn't seem to be as obvious, although the standard deviation in the noise case is also high suggesting confusion of the participants. This confusion is unsurprising as the cues received from a focused source at an oblique angle become less like those from a real source, and as the listener moves round to the side of the focused source the relative strength of the pre-echoes increases with respect to the direct energy from the virtual source, thus the localisation accuracy decreases; although some participants were still able to localise the source correctly even at large source-listener angles which is why the standard deviation of the results increases at larger angles.

A polar plot of the angular accuracy of the reported angles can be a useful representation of the results and demonstrates well the increase in view angle with an increase in source-listener distance. The radial axis in this case is the angular accuracy which is given by *(5.1)*.

$$\text{Angular Acuracy} = 1 - \left(\frac{median\left(\left|\theta_{error}\right|\right)}{180} \right) \qquad (5.1)$$

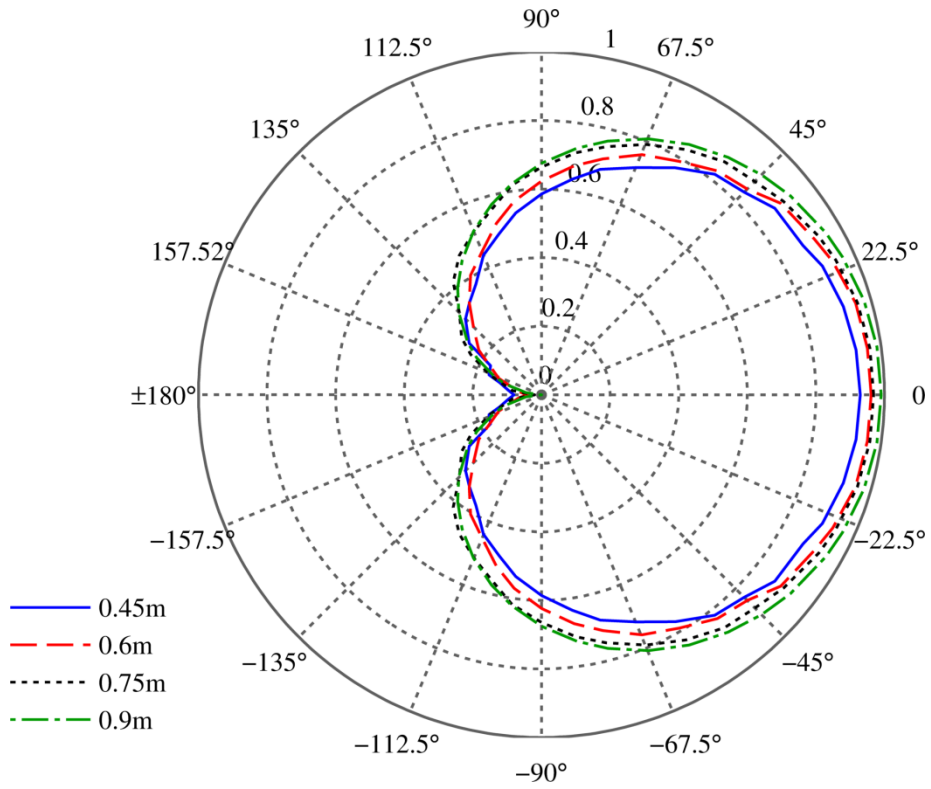


Fig. 5.16 Polar plot of localisation accuracy for a *noise* source

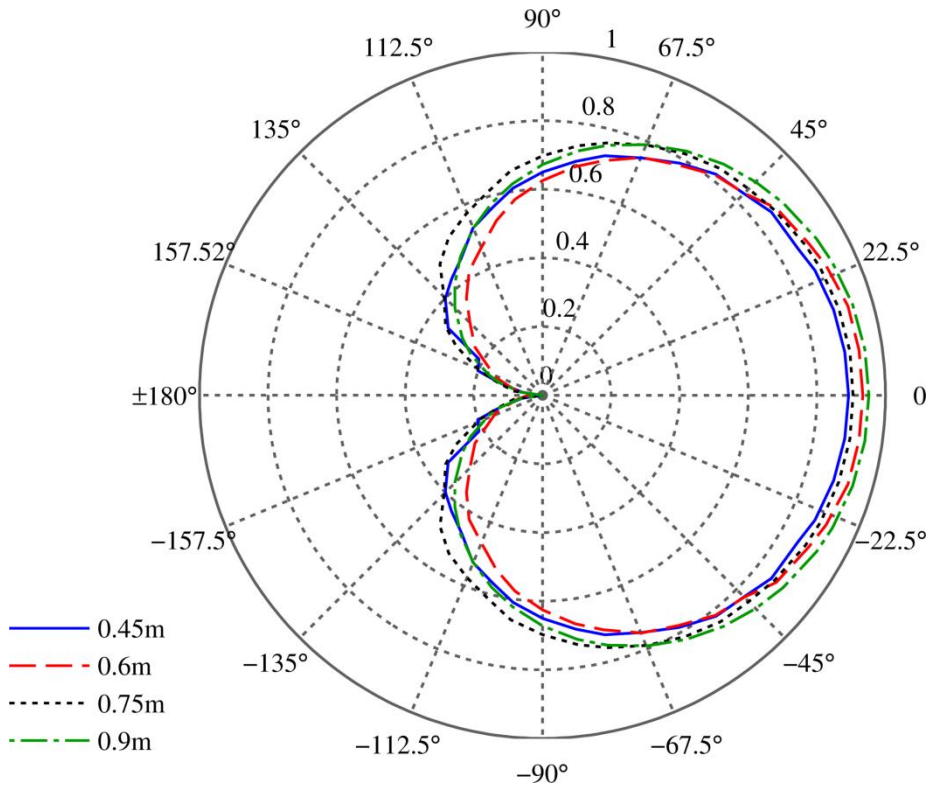


Fig. 5.17 Polar plot of localisation accuracy for a *speech* source

The polar plots show how the angular accuracy deteriorates as one moves to the rear of the focused source, it also demonstrates how there is a wider 'beam' for the larger source distances. This is seen both for the noise and for the speech sources in Fig. 5.16 and Fig. 5.17 respectively. The optimal polar response would be constant localisation accuracy of unity for all angles (i.e. an omnidirectional response), physical constraints of focused sources mean this is unlikely but subjective improvements can be made, as shown in section 5.6.

Assuming correct localisation to be where the angular error is ±10º it is possible to define a subjective view angle for focused sources from the results at each distance as can be seen in Fig. 5.18 where the *±10º* error limits are plotted as black dotted lines.

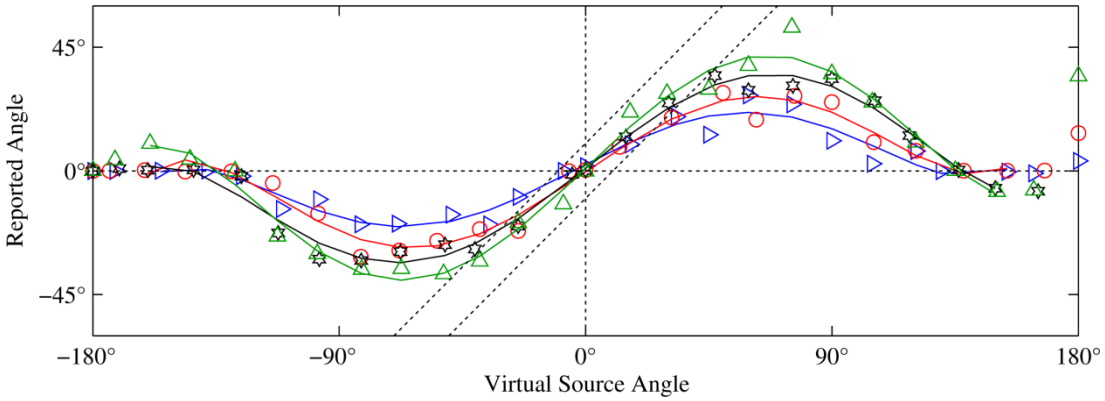


Fig. 5.18 Best fit of median results from all participants at all source angles for noise. Blue 0.45m, Red: 0.6m, Black: 0.75m, Green: 0.9m.

From Fig. 5.18 the perceptual view angle for focused sources is given in Table 5.1.

Distance	***Subjective View angle***
0.45m	-16˚ → 23˚ = 39˚
0.6m	-28˚ → 25˚ = 53˚
0.75m	-31˚ → 36˚ = 67˚
0.9m	-43˚ → 47˚ = 90˚

Table 5.1 Changes in subjective view angle with listener distance noise

These results demonstrate that the geometric view angle (70˚ in this case – see Fig. 5.6) is indeed a good approximation of the subjective view angle for moderate source distance. When the source is very close to the listener localisation seems to be more difficult and this is reflected in the particularly small subjective view angle.

The same procedure can be applied to the results from the test using the speech source as shown in Fig. 5.19 and Table 5.2.

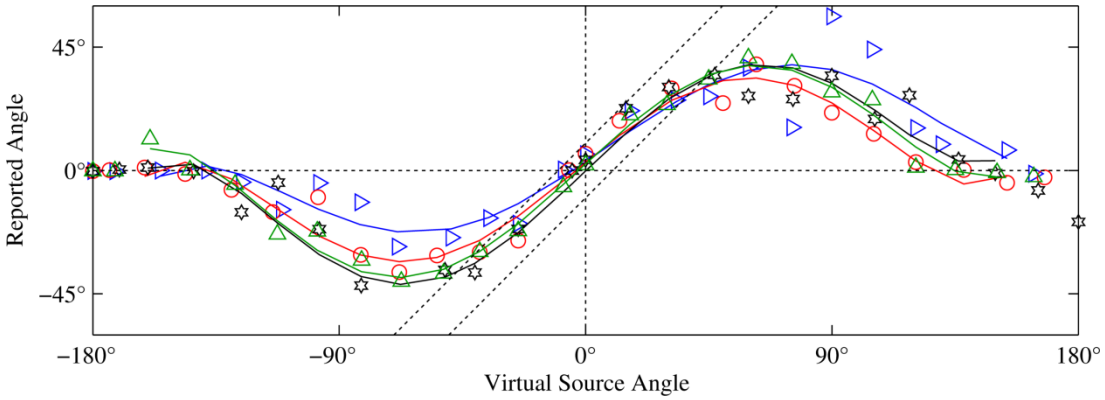


Fig. 5.19 Best fit of median results from all participants at all source angles for speech. Blue 0.45m, Red: 0.6m, Black: 0.75m, Green: 0.9m.

Distance	***Subjective View angle***
0.45m	-19˚ → 37˚ = 56˚
0.6m	-30˚ → 38˚ = 68˚
0.75m	-47˚ → 43˚ = 90˚
0.9m	-40˚ → 45˚ = 85˚

Table 5.2 Changes in subjective view angle with listener distance speech

The results from the speech signal are similar to those of the noise experiment, exhibiting a similar trend. The one anomaly in this case is that view angle at *0.75m* is greater than at *0.9m,* this is a result of the abrupt nature of the imposed error region which was assigned being unable to allow for fluctuations in the median of subjects' responses which are particularly prevalent in the speech results due to the higher standard deviation. The method of deriving the subjective view angle is fairly arbitrary in that the range of allowed angular errors (±10º) could be defined as any

angle depending upon the required accuracy, however the results do show how the view angle changes with distance and provide a definition of the subjective view angle that can be used as a means of measuring the success of any attempt to increase the view angle as described in the next section.

5.6 Improving the subjective view angle

It is apparent from these results that there is a need to broaden the subjective view angle of focused sources rendered using WFS. Although for these tests only one array was active, thus it could be considered a worst case scenario, there is still need for improvement as there is a large region where localisation is incorrect. Various methods have been suggested for improving the limitations of focused sources as described by Melchior et al. (2008) who suggested using a listener tracking system to update the WFS driving function such that the listener was always in the zone of correct localisation. Additionally Caulkins et al. (2003) suggested the inclusion of image sources to improve the localisation of such sources and to make the sensation more akin to natural listening conditions although they did not present any subjective tests to substantiate this idea, the purpose of the following is to substantiate this claim.

5.6.1 Adding first order image sources

This section presents an initial test performed in order to test the image source hypothesis, 10 subjects participated using the same setup as described above. All the sources were placed at $0.75m$ from the listener but with two different rendering scenarios which were compared with the normal WFS focused source. For the two rendering scenarios the same focused source was produced but with the four principal first order image sources included as the room model (Allen and Berkley, 1979). These sources were added in the rendering as normal WFS point sources, corresponding to the position and weighting of the reflections from the four walls as shown in Fig. 5.20.

The image sources were added with the correct delay applied to the source to account for the extra distance that would be travelled by the image source. The two

scenarios that were tested differed only with respect to the absorption coefficient of the ‘walls’. For the first scenario this was set to 0.5 and for the second it was set to 0.8. The location of the image sources was governed by the distance of the focused source from the acoustically transparent curtain such that the curtain acted as a wall in terms of visual and auditory perception.

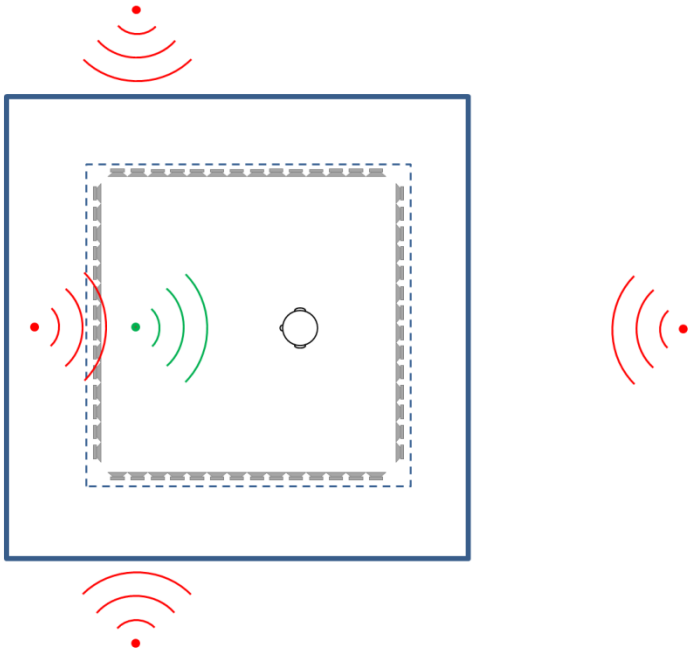


Fig. 5.20 Focused source (green) with the four first order image sources (red). The dashed line shows the virtual room from which the position of the images are calculated

As the test took place in an anechoic chamber there was no conflict between synthesised and real reflections.

5.6.1.1 Results

Again, the results from the ten participants are shown as scatter plots with the median line and error bars. As for the initial test, both noise and speech signals were used.

Noise

Once again the noise stimulus was two bursts of pink noise with the amplitude envelope shown in Fig. 5.5.

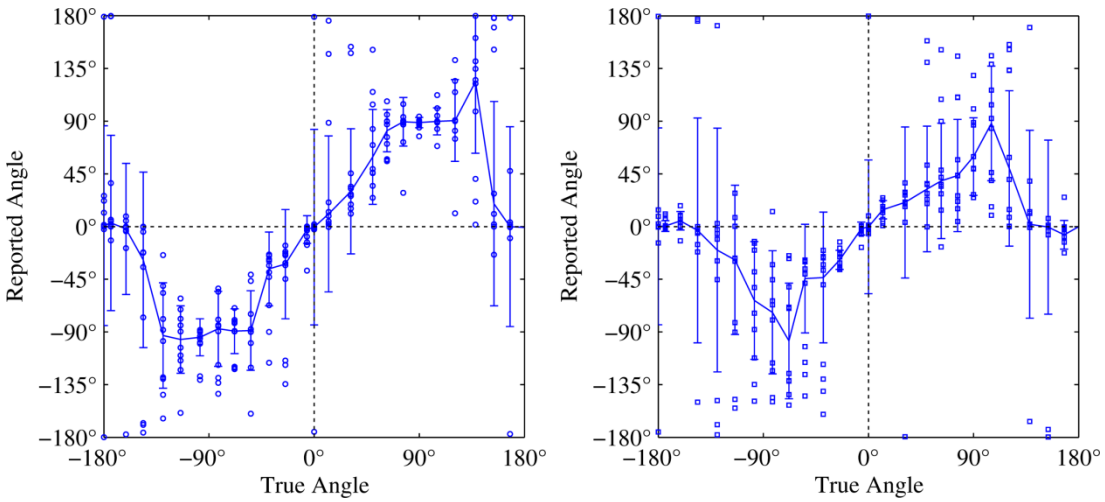


Fig. 5.21 Reported angle versus true angle for a focused *noise* source at 0.75m from the listener with images sources (Left: $\alpha = 0.5$ Right: $\alpha = 0.8$)

Plotting the curve fitted median results for the two image source scenarios and the normal WFS focused source show a large increase in the area of correct localisation of the focused source with added image sources as shown in Fig. 5.22 with a 5^{th} order polynomial fitted to the data once again.

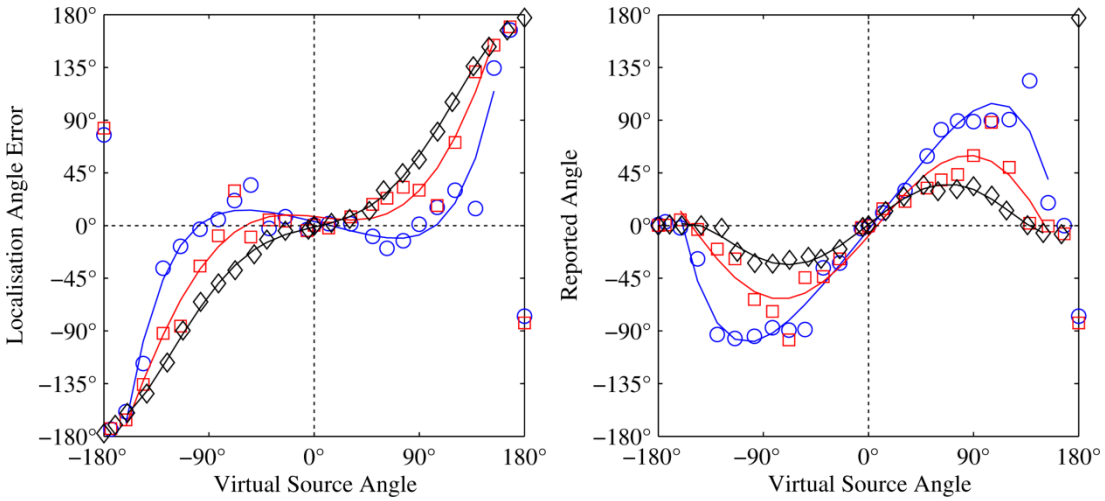


Fig. 5.22 localisation error versus virtual source angle (left) and reported angle versus virtual source position (right) for *noise source* at 0.75m from the listener (Blue: $\alpha = 0.5$ Red: $\alpha = 0.8$, Black: no image sources)

Distance	*Subjective View angle*
Normal	-16˚ → 23˚ = 39˚
Images08	-72˚ → 64˚ = 136˚
Images05	-108˚ → 113 = 221˚

Table 5.3 Changes in subjective view angle with listener distance noise

Speech

Plotting the same set of curves for the speech source reveals a similar increase in the area of correct localisation as shown in Fig. 5.23 and Fig. 5.24.

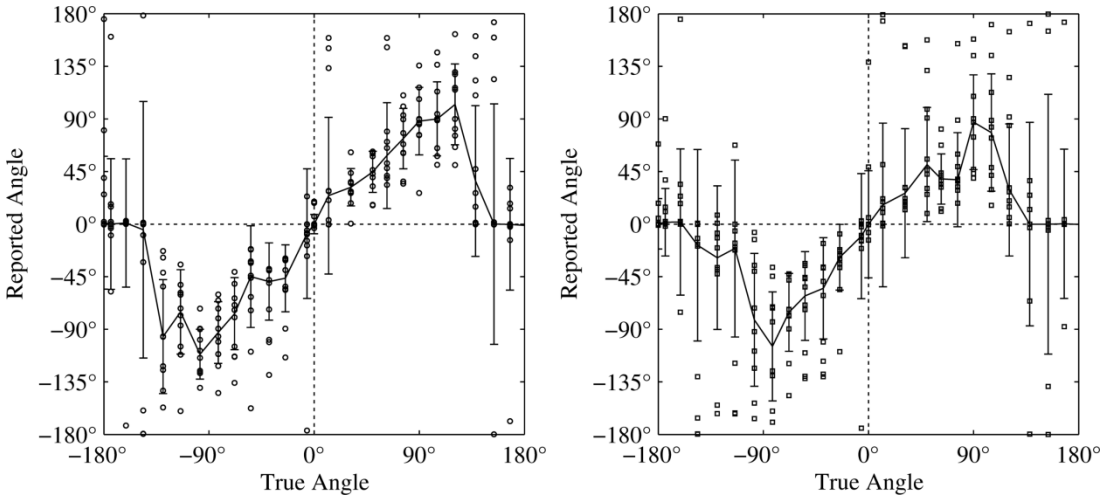


Fig. 5.23 Reported angle versus true angle for a focused *speech* source at 0.7m from the listener with images sources (Left: $\alpha = 0.5$ Right: $\alpha = 0.8$)

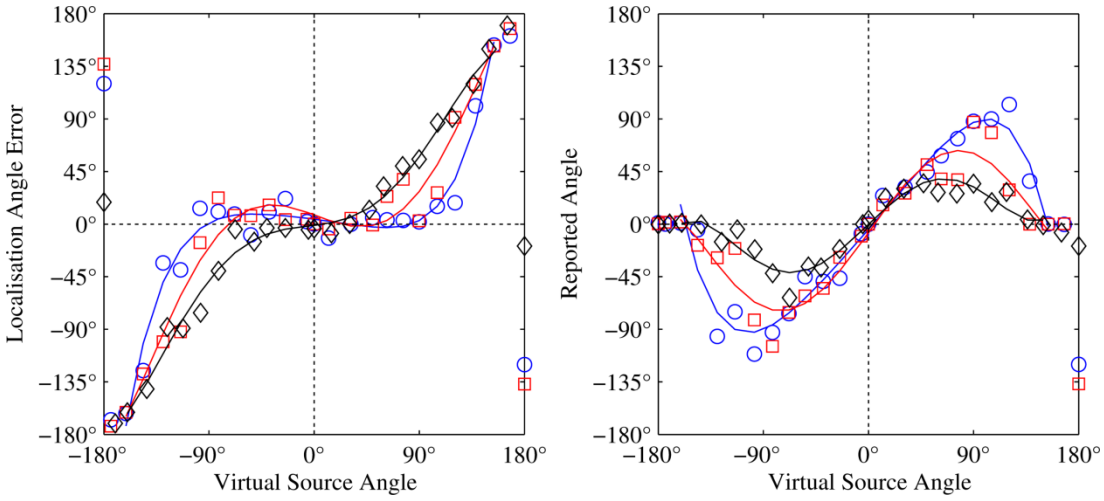


Fig. 5.24 localisation error versus virtual source angle (left) and reported angle versus virtual source position (right) for *speech source* at 0.7m from the listener (Blue: $\alpha = 0.5$ Red: $\alpha = 0.8$, Black: no image sources)

If the same error limits are applied, the subjective view angle with the inclusion of first order images sources can be seen in Table 5.4. It should be noticed that there is a greater variation in the results when adding the first order image sources for both the speech and noise stimuli, thus the strict criterion of ±10° as the allowed error limits for the definition of the subjective view angle has to be relaxed slightly as there are points where the median line has an error greater than ±10° but then moves back within these error margins at a larger source angle. Thus the definition of the subjective view angle here is defined by the greatest angle at which the line crosses the ±10° boundary without coming back again within these limits.

Distance	***Subjective View angle***
Normal	-47˚ → 43˚ = 90˚
Images08	-83˚ → 72˚ = 155˚
Images05	-102˚ → 99˚ = 201˚

Table 5.4 Changes in subjective view angle with listener distance speech

5.6.1.2 Discussion

It can be seen from these results that the addition of the principal first order image sources to the rendering of a focused source greatly widens the view angle. With the image sources with an absorption coefficient of 0.5, the subjective view angle increases to about 221° for noise and 201° for speech which corresponds to a *82%* increase for noise and *55%* for speech. It can be seen from Fig. 5.23 however that with the increased amplitude of the images sources, people began to localise the image sources rather than the focused source itself. The reason for this is due to the very low level of the direct energy from the focused source at angles of 60 - 90°. This means that with the image sources set at the higher level the lateral image source begins to dominate the localisation. This is less of a problem with the image sources set with an absorption coefficient of 0.8, whilst localising the image sources leads to an error in localisation, the error is less than at the same angles with no image sources. It could be argued that as the positions of the image sources are related to the actual source location, the errors will less disturbing and when combined with listener movements will give greater localisation accuracy as the

listener puts the whole 'picture' together, so the method can be considered advantageous. Further investigations will assess the validity of this method with listeners permitted to move around the space.

Again the standard deviation of the reported results is high. This is caused by the conflicting cues and people making localisation errors due to the pre-echoes and in this case the image sources. Front to back confusion also causes a larger than expected standard deviation in the results especially at 0°, where the ILD and ITD cues are the same for a source at 180° except for some spectral differences from shadowing of the pinna. These spectral differences can't be used for localisation in this case as the direct energy comes only from the front loudspeaker array, thus the front to back confusion can be an even greater problem. For the image source test, energy from the rear loudspeaker array reproducing one of the image sources can also cause people to localise sources mistakenly from the rear.

Differences between the noise and speech sources are surprisingly small. It was predicted that rendering the image sources for a speech signal would be more effective as the transients in the source would provide more localisation cues and draw attention to the image sources. Also humans are very accustomed to listening to people talking in a room, where there will be real reflections, thus it was predicted that there would be a cognitive expectancy of how the source should sound based on previous experience of listening to people speaking in small rooms, i.e. subconsciously expecting a certain reflected energy pattern. However it seems that the effect of the image sources on the noise is equally as effective as with the speech source, suggesting that the addition of the image sources works well for both transient and more steady state signals.

5.6.2 Multi-band WFS approach

A multi-band approach was also postulated and piloted during the subjective test. The principle was to split up the WFS source file into its high and low frequency components and render the low frequency with the rear loudspeakers and the high frequency content from the front loudspeaker array. This concept is illustrated in Fig. 5.27. Both the high and low frequency sources are rendered to the same location with

the idea that for any source angle the listener will be in the divergent region of one of the sources. The cut-off frequency was set at *500Hz*, the principle was to separate out the different localisation methods to see if the brain could use different cues to localise the source at the different angles around the head. It was hoped that by doing this there would be enough localisation cues at all angles to localise the source. The localisation accuracy in this case does come at the cost of colouration however, thus this method would be more useful for an application where the listeners were stationary such that the detection of these colouration artefacts would not be as apparent.

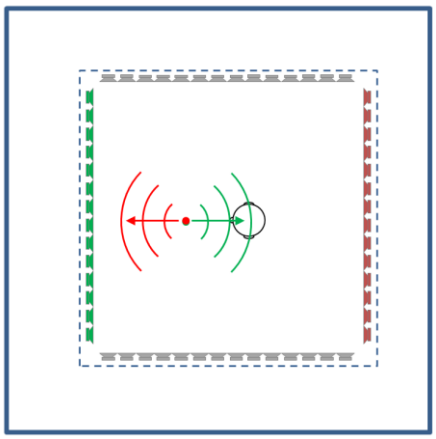


Fig. 5.25 Diagram of the split frequency source concept, the red array renders the red (low frequency source) and the green array renders the red (high frequency source) at the same position

5.6.2.1 Results

The spit source experiment again shows an increase in the subjective view angle (as shown in Fig. 5.26 and Fig. 5.27) although there is an increase in angular error which is roughly proportional to the true angle of the source suggesting that this approach is not valid or at least needs some improvement before it can become a viable solution.

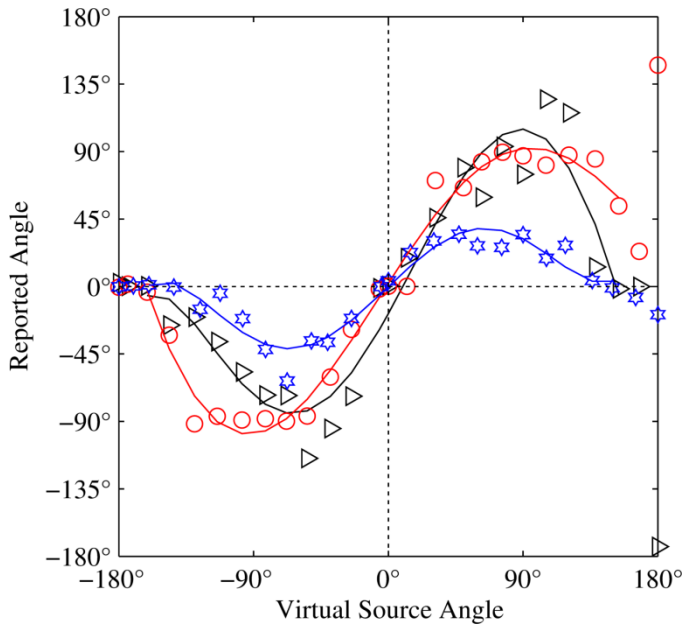


Fig. 5.26 Reported angle versus true angle for a split frequency *speech source* with. Blue: Standard rendering, Red: Split source rendering, Back: Split source rendering with 1st order images ($\alpha = 0.8$)

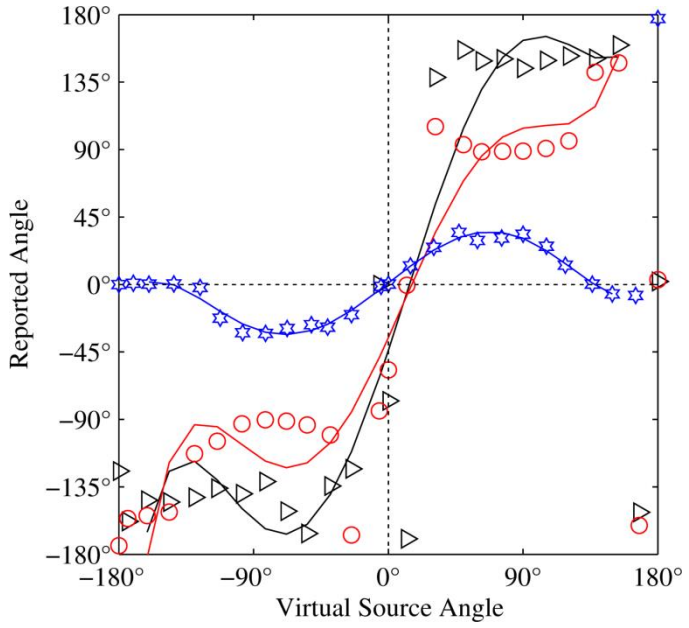


Fig. 5.27 Reported angle versus true angle for a split frequency *noise source* with. Blue: Standard rendering, Red: Split source rendering, Black: Split source rendering with 1st order images ($\alpha = 0.8$)

A polar response showing the four new rendering methods is shown in Fig. 5.28 for noise and Fig. 5.29 for speech and reveals the relative merits of the different techniques.

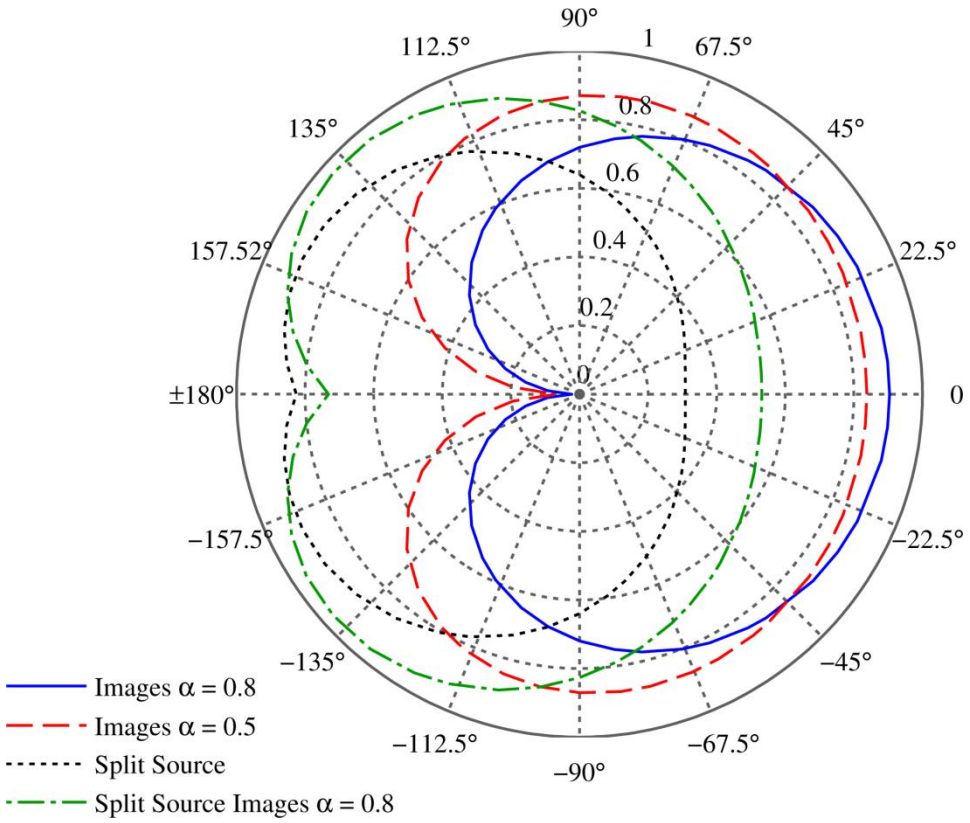


Fig. 5.28 Polar plot of angular accuracy for the 4 new rendering methods for *noise*

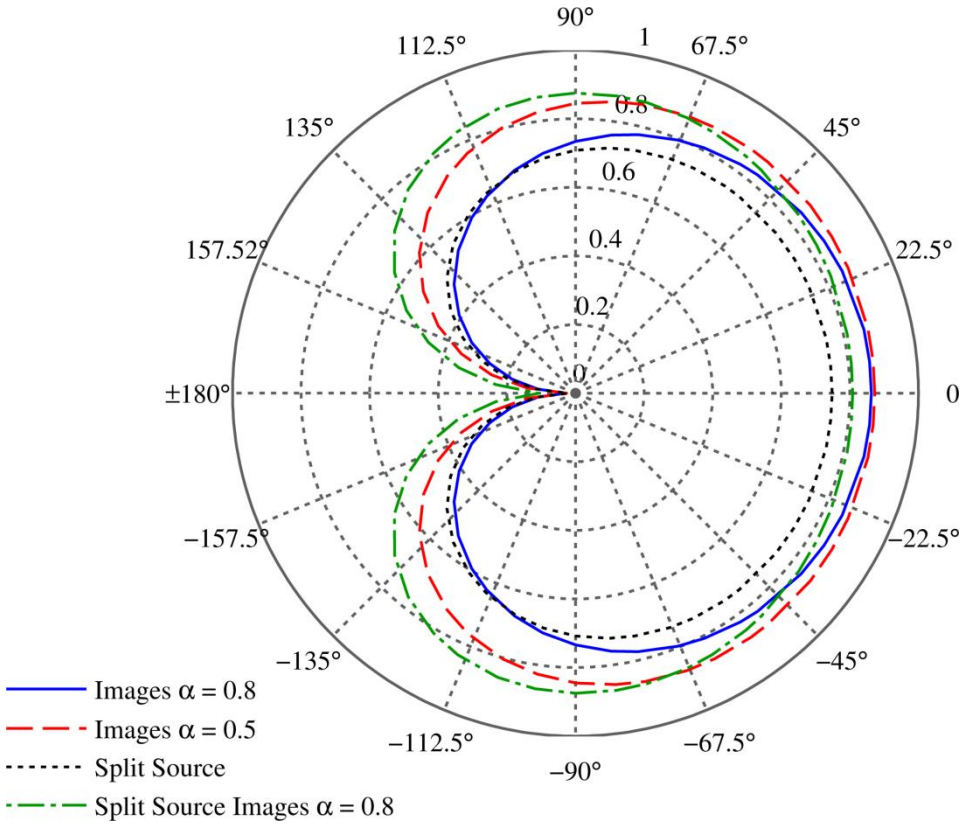


Fig. 5.29 Polar plot of angular accuracy for the 4 new rendering methods for *speech*

5.6.2.2 Discussion

These polar plots show the data with a 5th order polynomial curve fitting applied to the data. The results demonstrate that the scenario with the images sources with an absorption coefficient of 0.5 has the greatest view angle but the split source with images also performs well, especially at orthogonal angles. It could be postulated that having image sources with an even lower absorption coefficient might be beneficial, however experience of listening to the system during the pilot test phase revealed that as the level of the image sources increased, the locatedness (Rumsey, 2001) of the source decreased such that it is difficult to position the source at a specific location.

5.7 Conclusions

This chapter has described some of the interesting perceptual properties of focused sources in WFS. It has highlighted some of the key experiments and results found in the literature. The chapter also describes a localisation test that was performed in order to determine the subjective 'view angle' of focused sources. Results from this test demonstrate that the subjective view angle is a function of virtual source distance; decreasing for sources very close to the listener. This chapter has also shown that the inclusion of the first order image source in the rendering of a focused source can dramatically increase the subjective view angle, although the level of the reflections needs to be sufficiently low so as to insure that the listeners don't localise the images over the direct sound. A multi-band approach was also proposed, rendering two sources for each virtual source location with the high frequencies from one direction and the low frequencies from the other, such that the listener would be in the divergent region of one of the source at any given location, it was hoped this would provide enough cues to improve the localisation over a wide range of angles. However, results suggest that this technique introduces too many colouration artefacts and the standard deviation of the results is too great to consider it a viable improvement to focused source rendering. The work described here can also be found in (Oldfield et al., 2010).

6 Focused Sources in Compromised Reproduction Scenarios

6.1 Introduction

As has been seen from the preceding chapters, wave field synthesis is a spatial sound reproduction which offers a very accurate representation of sound scenes. However there are strict requirements upon the system, namely that the loudspeakers should be setup at ear level as a continuous array surrounding the listener. Secondly the reproduction room should ideally be source-free as dictated by the Kirchhoff-Helmholtz equation; this includes any scattering objects in the space as the resulting reflections approximate an acoustic source. These conditions are rarely met in real WFS systems and the respective reproduction environments and as such it is important to analyse the effect of not adhering to the requirements on the reproduced sound field. This chapter will analyse some of the errors of WFS sound fields in non-ideal reproduction situations. The errors that will be covered are those associated with the reproduction room, errors as a result of elevated loudspeakers and the effect

of scattering objects in the reproduction space. All of these errors are more critical for focused sources as they are generally rendered close to the listeners who have a clear frame of reference for the source as it has a tangible location within the room. The close proximity of the source to the listeners means that any errors are more keenly felt and as such should be characterised and if possible, corrected for.

Room Effects

The fundamental principles of wave field synthesis demand that the reproduction volume be source free and thus theoretically WFS systems should be installed in anechoic environments where no reflections/sources are present. For several practical reasons this is rarely the case apart from test/research systems such as the one described in section 3.6.2. Consequently there are several errors associated with the reproduction of sound fields in real rooms resulting from the reflections from the room's boundaries and objects therein. Investigations by Gauthier and Berry (2007) show experimental data of sound field distortion as a result of room effects on their WFS system and Wittek (2009) in his thesis made some brief observations on the real reflections from focused sources and the corresponding effect on the reproduced wave field but other than these publications there has been relatively little attention to objective and subjective effects of room reflections on the practical use of wave field synthesis systems. Work has however been done to try and alleviate these effects by using active control techniques (Corteel and Nicol, 2003; Gauthier and Berry, 2006; Spors et al., 2004a). WFS is well suited to these techniques as there is such a large number of sources available for controlling the sound field in the room. In fact some of the very earliest papers on WFS were specifically focused on the topic of controlling the sound field and acoustics of the reproduction room (Berkhout et al., 1993; Berkhout, 1988), many other active control techniques have also been presented, an overview of some has already been described in section 3.7.1. Most of the techniques presented offer rudimentary control in the 2D plane that the WFS system operates in. This can be problematic as room reflections and scattering off objects happen in three dimensions requiring that an infinite number of loudspeakers be placed on a continuous surface surrounding the reproduction space in order to accurately control the entire sound field and eliminate the temporal and spectral errors resulting from these reflections. There are also problems associated

with the number of transducers required to sample the sound field such that the necessary corrections can be determined and applied.

This chapter demonstrates the usefulness of the finite difference time domain (FDTD) method for predicting the sound field produced by a WFS system as part of a given room. Running the model in both free-field conditions and as part of the room enables the acoustic effects of the room to be quantified; this presents a significant advantage when designing a system as it provides an estimation of the amount of absorption needed in the reproduction room for sound field reproduction of sufficient accuracy. The model not only allows arbitrary room geometries but also enables the absorption profiles of the room's surfaces to be changed allowing optimisation for a given application. Results show that there is significant degradation in the synthesised sound field spatially, temporally and with respect to frequency when the reproduction takes place in a room as opposed to free field conditions (as the WFS definition dictates). The positioning and amount of absorption in the room is consequently crucial to a successful WFS system unless active room compensation is employed but this presents practical difficulties and significant additional computation complications.

Elevated Loudspeakers

In many WFS systems (as with the system described in section 3.6.5) it is not possible to position the loudspeakers at ear height surrounding the listener due to visual requirements such as the inclusion of projection screens in the reproduction environment. The consequence of this is that the rendered source position will be different to the intended source position. This is particularly problematic for focused sources where the source is rendered close to the listener and small changes/errors in position are more noticeable.

Scattering Objects

Also covered in this chapter is the effect of scattering objects on the reproduction of focused sources. Scattering objects within a room are common in real listening environments and for real source radiation and consequently the human auditory system is good at interpreting the auditory scene and localising sources even in these

scenarios. It could thus be argued that it is not necessary to analyse the effects of scattering on the sound field (assuming the sound field is accurately rendered) as the situation is approximating real listening conditions anyway. For focused source reproduction however there are additional considerations for the scattering of the sound field as the scattering object may be in the convergent region of the focused source and may therefore have a tangible effect on the rendering of the said source. This chapter concludes with an analysis of the scattering effects of objects both in the convergent and divergent region of a focused source. The scattering from a rigid sphere is initially considered followed by the scattering introduced by more complex objects using an FDTD approach. A method for correcting the effects of this scattering using time reversal techniques is also presented.

6.2 Finite difference time domain (FDTD) method

It is difficult to create an analytical model that can predict the performance of a system including the reflections from boundaries in the room, which means that real reproduction scenarios are hard to predict. This section presents a finite difference time domain (FDTD) model that can be used to predict the sound field produced by a given system either in an anechoic environment or when used in a real room (Oldfield et al., 2009; Drumm and Oldfield, 2010). The model described here allows the arbitrary assignment of room geometry (Drumm and Lam, 2007) as well as the application of frequency dependent absorption characteristics to the surfaces and objects within the room. As the modelling is done in the time domain, it is easy to produce a visualisation of the waves propagating through the room enabling the determination of the directions of particularly problematic reflections. Consequently it can be used to determine the effects that the room has on the reproduced wave field in the temporal, frequency and spatial frequency domains, thus making it a very powerful tool for the analysis of WFS systems.

6.2.1 FDTD Theory

The FDTD method has attracted substantial interest in the last decade for modelling complex acoustic wave propagation situations. The technique was originally

developed by Yee (1966) to solve Maxwell's equations for electromagnetic wave propagation through isotropic media by approximating derivatives by finite differences and has since been adapted and applied to acoustic propagation scenarios (Botteldooren, 1995; Maloney and Cummings, 1995).

6.2.1.1 Staggered grid

Yee's classic method for simulating electromagnetic wave propagation uses spatially and temporally staggered grids where electric and magnetic fields are stored at interleaved locations. The acoustic FDTD method on the other hand uses spatially staggered grids of pressure and particle velocities, as shown in Fig. 6.1. Multiple pressure generators/sources (e.g. loudspeakers) and surfaces can be placed within this grid and hence the interaction of waves with scatterers and each other can be predicted in the time domain.

The acoustic implementation of FDTD considers sound propagation as two coupled equations *(6.1)* and *(6.2)*

$$\frac{\partial v}{\partial t} = -\frac{1}{\rho}\nabla p \qquad (6.1)$$

$$\frac{\partial p}{\partial t} = -K\nabla \cdot v \qquad (6.2)$$

where p is the acoustic pressure, v is the particle velocity and ρ and K are the density and bulk modulus of the medium respectively. Equations *(6.1)* and *(6.2)* can be expressed as centre difference equations *(6.3), (6.4), (6.5)* and *(6.6)* for the velocities in each orthogonal direction and pressure respectively where n is the time step and i, j and k represent the spatial grid location index in each Cartesian direction (x, y, and z).

$$v_x^{n+1}\left(i+1/2,\, j, k\right) = v_x^{n}\left(i+1/2,\, j, k\right) - \frac{\delta t}{\rho \delta x}\left\{p^{n+1/2}\left(i+1,\, j, k\right) - p^{n+1/2}\left(i,\, j, k\right)\right\} \qquad (6.3)$$

$$v_y^{n+1}\left(i,\, j+1/2,k\right)=v_y^{n}\left(i,\, j+1/2,k\right)-\frac{\delta t}{\rho\delta y}\left\{p^{n+1/2}\left(i,\, j+1,k\right)-p^{n+1/2}\left(i,\, j,k\right)\right\} \tag{6.4}$$

$$v_z^{n+1}\left(i,\, j,k+1/2\right)=v_z^{n}\left(i,\, j,k+1/2\right)-\frac{\delta t}{\rho\delta z}\left\{p^{n+1/2}\left(i,\, j,k+1\right)-p^{n+1/2}\left(i,\, j,k\right)\right\} \tag{6.5}$$

$$\begin{aligned} p^{n+1/2}\left(i,\, j,k\right)=p^{n-1/2}\left(i,\, j,k\right)-K\delta t\Bigg\{&\frac{v_x^{n}\left(i+1/2,\, j,k\right)-\left(i-1/2,\, j,k\right)}{\delta x}+\\ &\frac{v_y^{n}\left(i,\, j+1/2,k\right)-\left(i,\, j-1/2,k\right)+}{\delta y}\\ &\frac{v_z^{n}\left(i,\, j,k+1/2\right)-\left(i,\, j,k-1/2\right)\Bigg\}}{\delta z} \end{aligned} \tag{6.6}$$

The half steps in the equation indicate that the pressure and velocity grids are interleaved in a ‘staggered grid’ arrangement. Hence via a leap-frog algorithm the pressure at a given time step is calculated from the previous pressure and adjacent velocities and conversely the velocity calculated from previous velocity and adjacent pressures. This principle can be represented for a simple 1 dimensional case diagrammatically as Fig. 6.1.

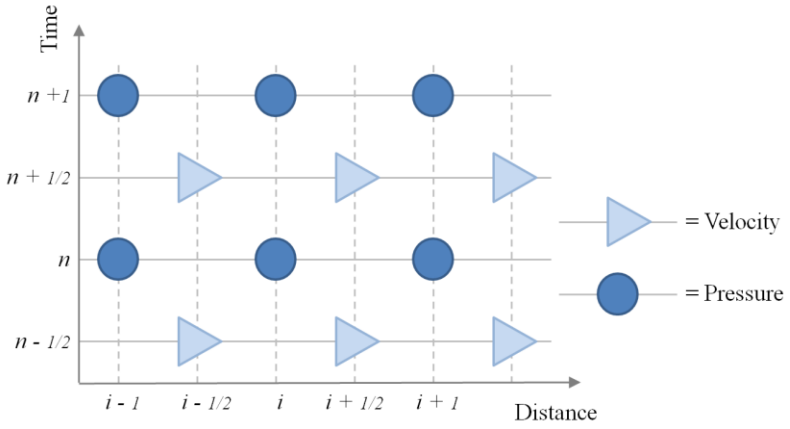


Fig. 6.1. The ‘staggered grid’ representation of the FDTD algorithm

6.2.1.2 Stability

The temporal and spatial grid size as well as the wave speed imposes limits on the stability of the FDTD algorithm which can be related to the Courant number (Maloney and Cummings, 1995; Taflove and Hagness, 2000) as follows. In order for the algorithm to be stable the condition in *(6.7)* must be upheld, where *s* is the Courant number, *c* is the wave propagation speed Δt is the temporal spacing and Δx is the spatial separation (assuming a uniform grid in all Cartesian directions), *N* is the number of dimensions in the model.

$$\left(s = \frac{c\Delta t}{\Delta x} \right) < \frac{1}{\sqrt{N}} \qquad (6.7)$$

6.2.1.3 Perfectly matched layer

In order to implement FDTD in practical software application there has to be limits to the FDTD medium to insure that there is no scattering or reflections back from the terminations of the medium causing errors in the prediction. This problem can be greatly reduced by implementing a theoretically reflection-free boundary around the medium based on modified FDTD equations which include anisotropic absorption and pressures. A practical implementation of this principle can be realised by progressively increasing absorption at the boundaries using a technique detailed in the literature as perfectly match layers (PML) (Berenger, 1994; Yuan et al., 1999). A PML is necessary especially in models such as the one described here where the boundaries such as walls etc. have absorbing characteristics as leakage through the walls would cause unwanted reflections that would propagate back into the room being modelled.

6.2.1.4 Source Type

Another potential source of undesirable scattering in FDTD affecting the accuracy of the prediction is caused by the source itself. There are principally, three possible source types that can be used in FDTD modelling, *hard sources*, *soft sources* and *transparent sources*. Hard sources are the most basic sources; the principle here is simply to set the pressure at a given grid location using some kind of time dependant function such as a sinusoid or a Gaussian pulse etc. This means that the pressure or

velocity at the node is not controlled by equations *(6.3)* - *(6.6)* but are rather controlled by the source function. The problem with this simplistic source is that the source itself will scatter incident waves in the grid. Consequently so called *soft sources* are sometimes used which add pressure to the grid rather than actually setting its value. These sources allow sound to propagate freely through the source location thus don't exhibit the problems due to scattering that hard sources do. The problem associated with this source type is that the net effect of the source is modified by the medium's response, effectively introducing a filtering into the source function. A transparent source (Schneider et al., 1998) has thus since been suggested that attempts to compensate for this filtering by convolving the source function with the inverse response of the medium. Transparent sources incur extra computational cost as the simulation effectively has to be carried out twice and also exhibit a lack of stability as there is no guarantee that the medium's impulse response will be minimum phase therefore be invertible. The FDTD model used here does allow the capability for transparent and soft sources, but the negative effects of hard sources were found to have negligible effect of the predictions of the reproduced sound field when computed in 3D, so have been used for all of the FDTD predictions in this thesis.

6.2.2 Implementation

The simulations presented here have been made using the WaveTank software developed at the University of Salford (Drumm and Lam, 2007). The software allows not only to position surfaces in the FDTD grid but also allows the assigning of frequency dependent absorption coefficients to those surfaces.

6.2.2.1 Incorporating scene geometry

Arbitrary room geometries can be imported into the WaveTank software in a variety of common formats including CATT acoustic, ODEON or X3D files (Drumm and Lam, 2007). These are then voxelised using a ray based approach (Drumm, 2005) to form surfaces in the 3 dimensional FDTD medium. Voxelising the planes in this manner requires a process of firing rays at the plane and testing whether it falls within the FDTD grid. Depending on how coarse the grid is, the object's smooth

edges will become more like a step in a similar manner to pixilation in a 2D image. Using this approach also allows arbitrary geometric objects to be placed within the FDTD grid as scattering/absorbing bodies (as shown in section 6.6.4.1).

6.2.2.2 Incorporating frequency dependant boundaries

The frequency dependent absorption characteristics of any surfaces within the room are modelled using an approach analogous to IIR filtering. It is assumed that a velocity component at the surface is given as the function of the history of the normal velocity components incident and the history of the outputs from this function such that:

$$v = f_n\left(n_x \cdot v_x^n, n_x \cdot v_x^{n-1}, n_x \cdot v_x^{n-2}, n_x \cdot v_x^{n-2}..., f_{n-1}(...), f_{n-2}(...),...\right) \quad (6.8)$$

Where v_x^n is given by *(6.5)* and n_x is the absolute value of the surface's unit normal. The output from this equation becomes an input of *(6.6)* hence the surface effect can easily be included in the FDTD iterative cycle. Equation *(6.8)* can be considered as the difference equation of an IIR filter which can be written in the form:

$$v = \left[h_0 + \sum_{i=1}^{N} b_i w_{n-1} - \sum_{j=1}^{N} a_j f_{n-1} \right] \quad (6.9)$$

The coefficients h_0, $a_1, a_2 \ldots a_N$ and $b_1, b_2 \ldots b_N$ can determined from the plane's octave banded absorption data using a quasi-Newton optimisation algorithm to minimise the Lp-norm objective function.

$$E(x) = \int \left| H(\omega) - H_d(\omega) \right|^2 d\omega \quad (6.10)$$

Where the desired magnitude response of the filter as a function of octave banded absorption coefficients , $a(\omega)$ is given as:

$$H_d(\omega) = 1 - \sqrt{1 - \alpha(\omega)} \tag{6.11}$$

And

$$H(\omega) = h_0 \frac{B}{A} \tag{6.12}$$

Where

$$\begin{aligned} B &= 1 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_N z^{-n} \\ A &= 1 + a_1 z^{-1} + b_2 z^{-2} + \ldots + a_N z^{-n} \end{aligned} \tag{6.13}$$

Hence in conjunction with the gradient

$$\nabla E(x) = 2 \int \mathrm{Re}\left[\bar{e}(\omega) \nabla H(\omega) \right] d\omega, \tag{6.14}$$

an iterative scheme based on the Davidon-Fletcher-Powell update formula (Davidon, 1991) is employed.

6.2.3 Simulations

Single audio sources or arrays thereof can be positioned virtually into a room with an arbitrary geometry such as a concert hall, providing a very useful audio design tool. The entire FDTD grid can be exported allowing auralisation at any point within the room before the installation of a system or to demonstrate other acoustic effects. Even relatively complex geometries can be imported as shown in Fig. 6.2, demonstrating the usefulness and versatility of the FDTD method in modelling WFS reproduction.

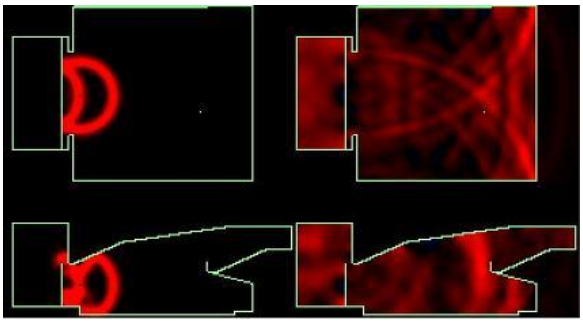


Fig. 6.2. WFS source reproduction in a room with complex geometry

It is also possible to include Direct3D models into the simulation as will be seen in section 6.6.4.1 where the effects of scattering objects on the reproduced wave field are analysed. Direct3D is the graphics API within DirectX and models 3D objects as a series of interconnected triangles to form a more complex geometries that can be navigated around using DirectX applications.

Showing the propagation through the room in the time domain using FDTD is very important as it provides a useful diagnostic tool to determine which reflections are causing the majority of problems. The problems could be seen as colouration in the frequency domain, temporal smearing in the time domain or distortion of the spatial wave front leading to a reduced localisation ability.

6.3 FDTD for determining Room effects on Focused Sources

As has been shown, FDTD is a very powerful acoustical modelling tool allowing the prediction of the performance of WFS systems that traditional methods would not permit. This is particularly important for the analysis of the effect of real room reflections for focused sources.

6.3.1 Room acoustic principles

A typical impulse response of a source in a room is generally split into three sections corresponding to the direct energy, the early reflections and the late reflections or reverberation, as shown diagrammatically in Fig. 6.3.

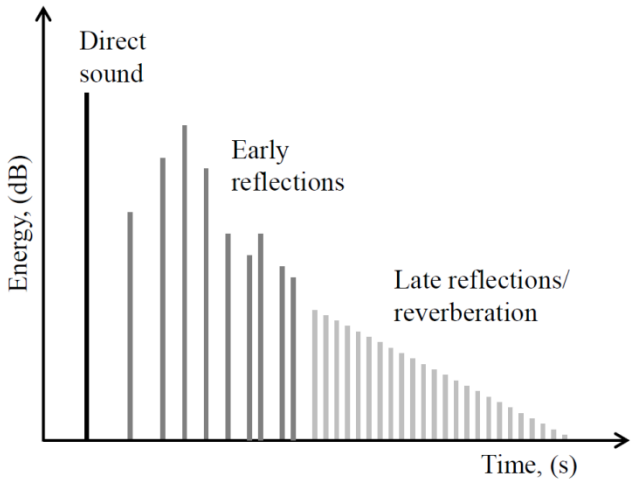


Fig. 6.3 Diagram of a typical impulse response of a source in a room

The first impulse represents the direct sound emitted by the source. This part defines the timbre and spatial properties of the source and the binaural cues generated therein inform the auditory system of the location of the source. The early reflections are normally distributed in all directions depending on the surrounding surfaces of the room and arrive within a time range of approximately *20ms* to *200ms*. These early reflections are important for the characterisation of the size and type of environment. It is well known that the early reflections provide the auditory system with information as to the room dimensions, help invoke a sense of spaciousness (Griesinger, 1997) and also help the perception of source directivity (Toole, 2008). The late part of the impulse responses is normally characterised by a diffuse field (Kuttruff, 2000) and can be largely described by its statistical properties (Jot et al., 1997). The later reverberant part of the impulse response is the most perceptually obvious cue for room size and is often characterised by the RT60 which is the time in which it takes for the acoustic energy to decay by 60dB (as discussed in section 6.3.1.1).

6.3.1.1 Reverberation time (RT60)

The reverberation time is perhaps the most perceptually significant attribute of a room impulse response and was first formally introduced by Wallace Sabine (1922) as a method of characterising a room's response to sound based upon its volume and absorption coefficient of the room's surfaces. His formula was derived empirically and can be expressed as:

$$RT60 = \frac{55.3V}{cA} \tag{6.15}$$

Where V is the volume of the room, c is the speed of sound and A is the total absorption of all surfaces given as:

$$A = \sum_{i=1}^{N} S_i \alpha_i \tag{6.16}$$

with S_i and α_i as the surface area and absorption coefficient of the i^{th} surface of the room respectively. There have been several extensions to this formula (see Kuttruff, 2000) to include for example the different absorbing characteristics of the surfaces of the room (e.g. Millington, 1932) but it is still widely adopted as a simple means of predicting the reverberation time in diffuse conditions of a room with a given volume.

The reverberation time is defined as the time taken for the acoustic energy in the room to decay by *60dB*. In many cases the dynamic range of the impulse is not sufficiently high to achieve an accurate measure of a *60dB* decay, thus in many practical situations, the decay is measured between *-5dB* and *-35dB* (this is often termed T30) and the time multiplied by 2 to give the equivalent RT60, it is also not uncommon to see reverberation time measurements based on a *20dB* decay (T20), but in general lower decay values introduce further inaccuracies in the measurement.

Estimating the reverberation time from a measured impulse response is done by computing the Schroeder curve (Schroeder, 1965) which is computed by taking a backwards integral of the impulse energy response given in *dB* by *(6.17)*.

$$S(t) = 10\log_{10}\left[\int_{t}^{\infty} h^2(t)\,dt\right] \tag{6.17}$$

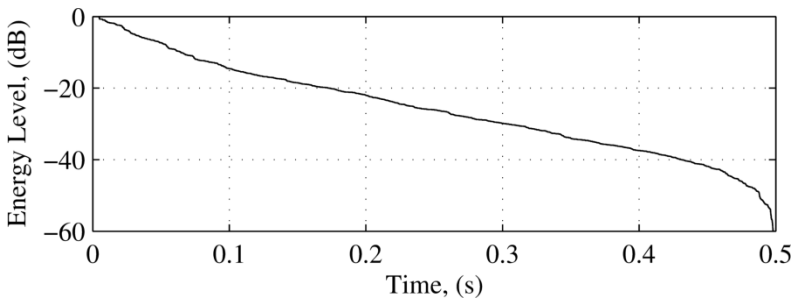


Fig. 6.4 Example Schroeder decay curve

It can be seen from Fig. 6.4 that due to the logarithmic decay of the energy of a typical impulse response of a point source in a room, the Schroeder decay curve exhibits a clear linear region, which can be used to estimate the RT60. By taking a line of best fit of this curve between *-5dB* and *-35dB* from the maximum value the T30 can be calculated and hence the RT60 can be determined (as shown in Fig. 6.5).

The decay curve in Fig. 6.4 demonstrates also that for a typical room impulse response the early part of the curve exhibits a steeper gradient which is a perceptually significant feature of sound decay in rooms and has hence lead to the development of the *early decay time* as described below.

6.3.1.2 Early decay time (EDT)

The early decay time (Jordon, 1970) is generally considered to be a perceptually better measure of reverberation time when compared with the perception of sound attenuation in rooms (Gade, 1994). Like the RT60, EDT is a measure of the time it takes for the sound energy in the room to decay by *60dB* but rather than performing an extrapolation based on the decay rate between *-5dB* and *-35dB*, the extrapolation is done over the first *10dB* decay of the Schroeder curve as shown in Fig. 6.5.

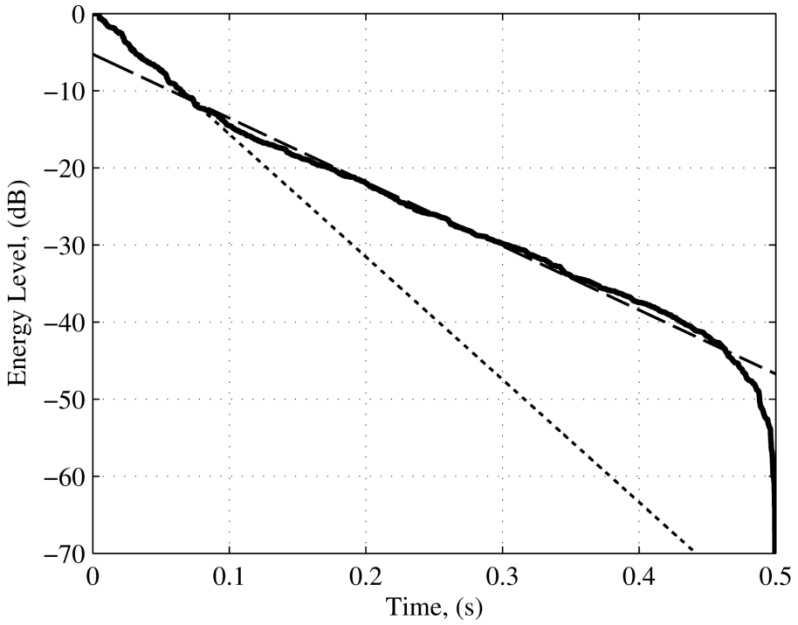


Fig. 6.5 Calculating the RT60 and EDT from the Schroeder decay curve. Solid line is the Schroeder decay curve, the dashed line is the best fit line between -5dB and -35dB used for calculating the RT60 and the dotted line is the best-fit line between 0dB and -10dB used for calculating the EDT

As EDT is measured initial region of the impulse response is predominantly affected by early reflections. As such it is exhibits more variation with listener/measurement position and is also affected more by the geometry of the room in question. EDT can be both longer or shorter than the RT60 depending on the room.

6.3.1.3 Deutlichkeit (D_{50})

Deutlichkeit or 'definition' is a measure of clarity and can be calculated directly from the room impulse response, $h(t)$. It is the ratio between the acoustic energy in the first *50ms* and the entire acoustic energy of the impulse response at the receiver location as a percentage. D_{50} was initially proposed by (Thiele, 1953) as a measure of clarity in concert halls and can be calculated by:

$$D_{50} = 100\left[\frac{\int_0^{50ms} h(t)^2\,dt}{\int_0^{\infty} h(t)^2\,dt}\right]\% \qquad (6.18)$$

Both the integrals in *(6.18)* must contain the direct sound of the impulse. A high value of D_{50} corresponds to good clarity of the received signal. If there are no reflected components arriving after *50ms*, the D_{50} would be *100%* and the clarity good. The relationship between D_{50} and syllable intelligibility was first established by Boré (1956) and shows that the respective measures have a high correlation.

6.3.1.4 Clarity Index (C_{80})

The clarity Index (C_{80}) was first presented by Reichardt et al. (1974) and is an alternative measure of the clarity/distinctness. Whereas D_{50} is particularly aimed at calculating the definition of speech, C_{80} is used to calculate the clarity or transparency of music. Whilst this parameter was originally designed for characterising the acoustics of concert halls, it is also widely used for smaller room scenarios (Zahorik, 2009). The C_{80} is the ratio in dB of the acoustic energy arriving in the first *80ms* with the energy arriving from *80ms* to the end of the impulse response and can be calculated using *(6.19)*.

$$C_{80} = 10\log_{10}\left[\frac{\int_{0}^{80ms} h(t)^2\,dt}{\int_{80ms}^{\infty} h(t)^2\,dt}\right] dB \qquad (6.19)$$

As C_{80} is mainly used for the clarity of music a higher integration limit is employed, allowing for the fact that reflections are less detectable for music than for speech. As with D_{50}, a high value of C_{80} corresponds to good clarity. Typical values of C_{80} are considered to be between *-5dB* and *+3dB* (Gade, 1994).

6.3.1.5 Centre time

The integration times defined in both the D_{50} and C_{80} are intended as a crude approximation of the how human hearing and are thus designed to approximate perceptual experience. The *centre time* (Kürer, 1969) is similar to the aforementioned parameters although in this case the integration time limit is infinite.

The centre time is defined as the first moment of the squared impulse response and can be calculated by:

$$Tc = \left[\frac{\int_0^\infty h(t)^2 \, t dt}{\int_0^\infty h(t)^2 \, dt} \right] ms \qquad (6.20)$$

Essentially the centre time measures the centre of gravity of the impulse response. The centre time is a useful room acoustic parameter as it has a high correlation with the speech intelligibility word score (Kuttruff, 2000). The centre time accounts for the fact that a later arriving reflection is potential more problematic in terms of intelligibility or transparency. The longer a reflection is delayed, the greater effect it will have on the centre time and the higher the calculated value will be, consequently it is apparent that a low value corresponds to high transparency or good speech intelligibility.

Many of the important room acoustic parameters and the procedures by which they are measured and calculated are described in ISO-3382 (1997).

6.3.1.6 Listener envelopment

Listener envelopment (LEV) (Bradley and Soulodre, 1995a, 1995b) is the sense of being in a large space. It is not so much related to localisation but more to being surrounded by many diffuse sound sources. It is generally considered to be a positive attribute of concert halls and can often be improved with the addition of diffusers in the listening space.

6.3.1.7 Apparent source width

The definition of apparent source width (ASW) is generally ascribed to (Barron, 1971) although in his paper he referred to it as *spatial impression*. It is a similar measure to LEV although it is characterised predominantly by an apparent broadening of a sound image which has a specific location. Whereas LEV is affected

predominantly by the later reverberation, ASW is affected more by the early reflections.

6.3.1.8 Precedence effect

Fundamental to the perception of sound in reflective environments is the 'law of the first wavefront' (Cremer, 1948), 'precedence effect' (Wallach et al., 1949) or 'Haas effect' (Haas, 1951). All three of these theories, although founded by different authors essentially describe the same phenomena and aid the understanding of human auditory perception in reflective environments.

Many reflections reach the listener after the arrival of the initial signal, however, surprisingly the presence of these reflections seem not to affect the human localisation of sound. This is a puzzling phenomenon as one would expect there to be many source locations discerned by the brain corresponding to the cues from each of the reflections. The truth of the matter is that even with a sound in a room where there are many reflections, still one auditory event is localised depending on the relative delay of the reflection with respect to the direct sound.

If two loudspeakers are placed in front of a listener (as per a stereo setup), then the position of the phantom source they create can be shifted on a continuum between the loudspeakers by altering the relative delays between the reproduced signals. In order for it to be clear that one source is a delayed form of the other, the signals should be non-periodic, coherent sounds. This is the case for small delays ($td<0.63ms$), when the delay is increased in the region of ($0.63ms<td<1ms$), the phantom source can be localised to the position of the loudspeaker that radiates the signal first. Above 1ms, the phantom source is localised using only the cues from the loudspeaker that radiates the signal first, this effect was first highlighted by Cremer in 1948 (Cremer, 1948) and is commonly referred to as the *law of the first wavefront*.

The smallest time delay at which the law of the first wavefront becomes the primary localisation method is defined by the minimum time delay between the two sounds that the perceived phantom source location begins to move away from the location of the first arriving sound. This value is between *0.63ms* and *1ms*.

6.3.2 Room acoustics of focused sources

The reflections from a focused source are different from the reflections that would come from a real source at the same position due to their causal nature, being generated by a remote array of secondary sources. This difference in reflection pattern will provide confusing localisation information to the listener and may affect the way in which focused sources should be rendered for optimal localisation accuracy. Of particular interest is the early reflection pattern produced by a focused source as this is the region in which the impulse responses of real and WFS focused sources differ the most (as will be discussed later and can be seen from Fig. 6.7).

A problem with WFS reproduction in real spaces is that the loudspeaker arrays are typically placed near the boundaries of the room which equates to the position of the pressure maxima of the room modes. Thus the arrays easily excite these room modes particularly as line arrays are used hence the area over which the mode is excited is greater. Consequently for WFS reproduction there is a need to think carefully about absorption, or active control techniques (e.g. Corteel and Nicol, 2003; Spors et al., 2004) specifically in these lower frequency modal regions. Modal problems can cause colouration in the frequency domain and temporal ringing in the time domain as well as an uneven spatial response (Kuttruff, 2000); consequently it is imperative that they are damped sufficiently for applications of WFS. The FDTD model described here allows for the inclusion of a band-pass filter absorbing boundary to be placed at the corner, simulating a membrane absorber such that the absorbers used within the room can be tuned and optimised (Oldfield, 2006) to suit the reproduction environment.

6.3.3 FDTD model implementation

A simulation was made of a WFS 2.5D focused source reproduced in a room approximating a standard listening room, with frequency dependent absorbing boundaries matching the characteristics of plaster (for the walls and ceiling) and a thick carpet (for the floor) (see Table 6.1). Setting the frequency dependant absorbing boundaries of the room enables more accurate predictions to be made rather than simply applying a fixed absorption coefficient to each surface. The

surfaces in the model represent real materials that are used in construction but exhibit low absorption coefficients. In a real room environment the acoustic situation is more complex with various soft furnishings which will provide significant absorption. However, modelling in this way, highlights the effects of the room acoustics on the wave field as many of the rooms in which WFS systems are installed do not have significant soft furnishings etc. The materials used in these simulations approximate those of the ITU listening room described in section 3.6.4.

		Octave Band Absorption Coefficient					
Surface	*Material*	*63 Hz*	*125 Hz*	*250 Hz*	*500 Hz*	*1000 Hz*	*2000 Hz*
Floor	Carpet	0.02	0.04	0.08	0.2	0.35	0.4
Ceiling and side walls	Plaster	0.2	0.15	0.1	0.08	0.04	0.02

Table 6.1 Octave band absorption coefficients of surfaces in the room

The WFS system was modelled to reproduce a focused source at different positions across the room as shown in Fig. 6.6.

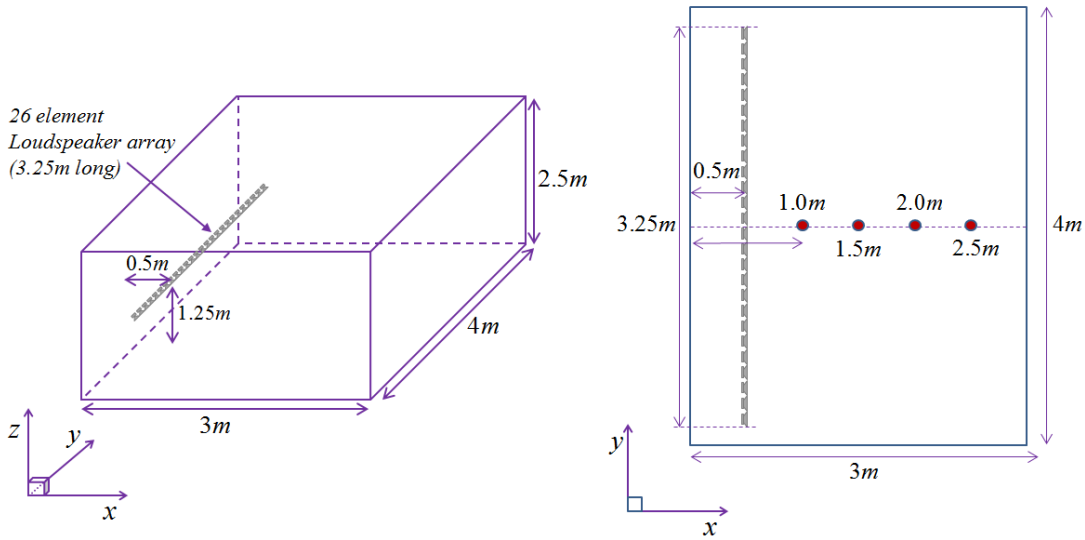


Fig. 6.6 Geometrical layout of room modelled with FDTD, the red circles show the virtual source positions modelled

The FDTD model described in section 6.2.2 was used to analyse the effect of boundary reflections on WFS focused source reproduction in the temporal, spectral and spatial domains. Simulations were performed with 4 scenarios for each source position:

1) Anechoic non-WFS source
2) Real room non-WFS source
3) Anechoic WFS focused source
4) Real room WFS focused source

By modelling source propagation with and without boundaries and for WFS and normal point sources enables firm comparisons to be made and an analysis both of the effect the reproduction room has on the wave field and also the effect of reproducing focused sources as virtual point sources.

6.3.4 Temporal effects introduced by the reproduction room

It is well known in room acoustics that the reflections from the room's boundaries create a feeling of spaciousness and that many reflections cause reverberation, providing the sense of a natural listening environment – in fact many people when they first venture into an anechoic chamber find the experience disorientating as the lack or reflections is so alien. Reverberation in the time domain is a smearing of the impulse response at a given location; this has an interesting effect on the temporal attributes of focused sources, most notably pre-echoes. In an anechoic chamber these pre-echoes are detected by the auditory system as discrete reflections, and their close proximity causes colouration (Wittek, 2009) rather than noticeable echoes. It is in fact the change in colouration as the listener moves around the listening space that has proved to be the most disturbing effect of pre-echoes, thus making moving focused sources particularly problematic in WFS.

Informal listening tests have shown that these artefacts of focused sources are more intrusive for an anechoic rendering than in a real room environment, this can be seen in Fig. 6.10 and Fig. 6.11 where the spatial fluctuations in the frequency domain are

far more pronounced for the anechoic case when compared with the system modelled in a room. This is consistent with the preliminary work done by Start (1997) who also pointed out that an anechoic environment was not the optimal reproduction environment for WFS as it highlighted the errors inherent in the system. The room therefore acts as an averaging filter in the time, frequency and spatial domain which reduce the perception of the WFS artefacts.

This section looks at the characteristics of focused source radiation in reflective environments in the time domain. Fig. 6.7 shows a comparison between the impulse response from a real source in a room and a WFS focused source at the same location.

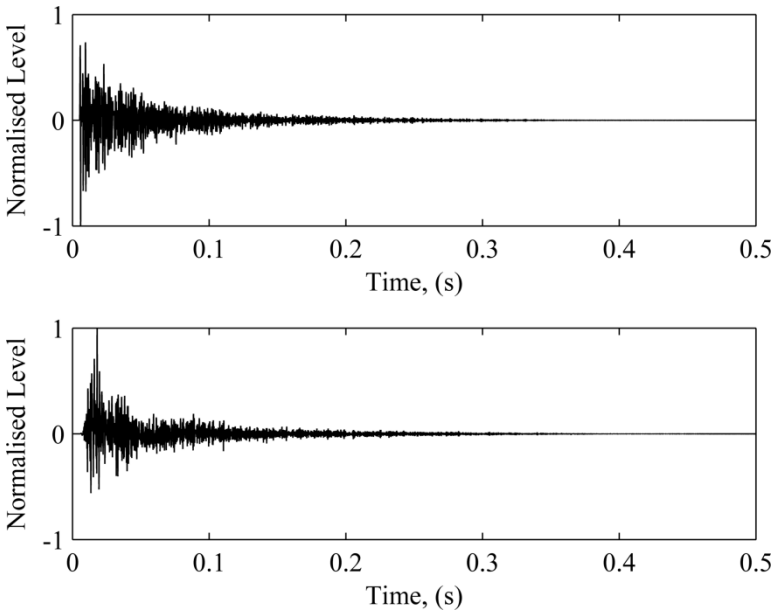


Fig. 6.7 Example impulse response predictions made using FDTD model for a real point source (top) and a WFS focused source (bottom). Note the pre-echo energy arriving before the direct sound for the WFS case.

Looking at the impulse responses simulated for both the real source and the WFS focused source it is apparent that the temporal decay in both cases is very similar but more processing is needed in order to determine the differences in the other room acoustic parameters mentioned in section 6.3.1. From the derived energy decay curve and Schroeder curve it can be seen that there is a significant noise floor in the FDTD simulation as shown in Fig. 6.8. This noise is a consequence of the FDTD method and is compounded in the WFS models as there are many sources in this case each

one adding noise to the simulation, resulting in an apparent non-linear energy decay (as seen in the Schroeder curve in Fig. 6.8) which would not be expected for a real room. To reduce the effect of this modelling noise a half-Hanning window is applied to the first *0.5s* of each impulse such that the Schroeder curve has a linear decay, resulting in more accurate assessment of room acoustic parameters. If the impulse response is windowed like this to remove the noise floor, the RT60 and EDT times are within the difference limen shown in Table 6.3. Thus the conclusion is that the energy decay of a real source and a WFS focused source are equal.

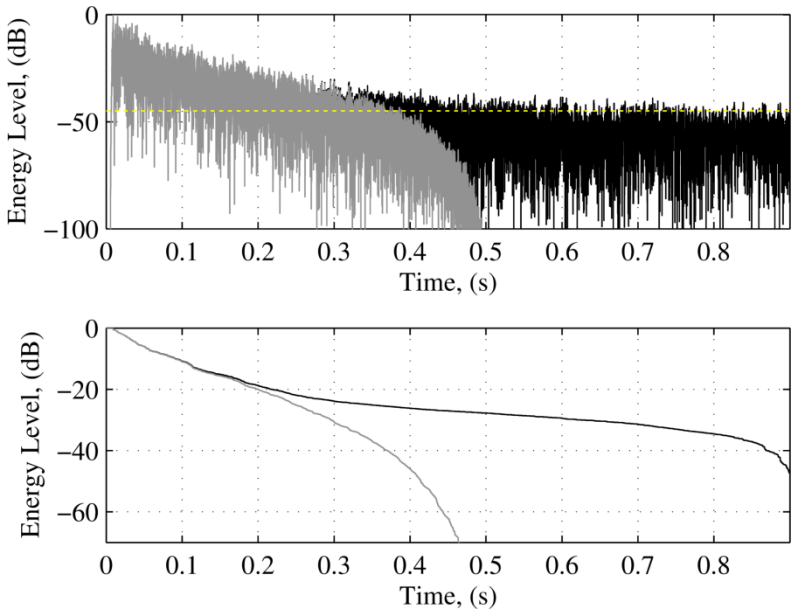


Fig. 6.8 Top: Energy decay curve, Bottom: Schroeder curve, for an exemplar FDTD modelled focused source for the un-windowed case (black traces) and the windowed case (grey traces). The approximate noise floor is shown by the dashed yellow line in the top plot.

A line of 155 receiver positions across the width of the room and *1m* in front of the source position were chosen from the FDTD simulations for both the real source case and the WFS focused source, with the geometry of the model defined in Fig. 6.6. In both cases the RT60, EDT, D_{50}, C_{80} and *Tc* were averaged across all 155 receiver locations to give the mean results shown in Table 6.2. The results from the source at a distance of *1.5m* from the array for each receiver position are shown in Fig. 6.9. This figure shows that there is no overall trend in how the parameters change throughout the room but does demonstrate that the overall trend of the parameters for the real and focused source relative to each other remain consistent across the room.

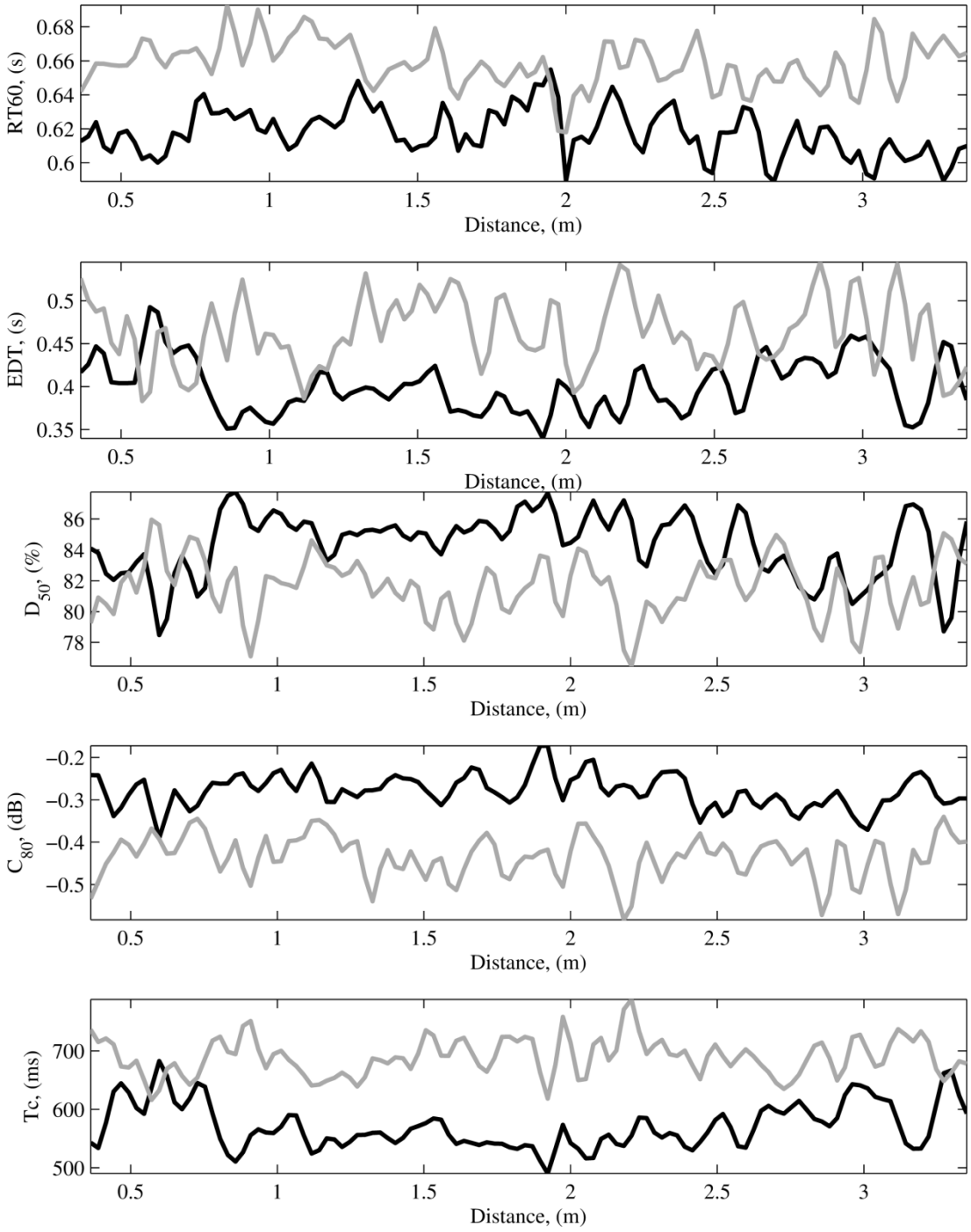


Fig. 6.9 Acoustic parameters calculated across a line of 155 receivers *1m* in front of the source. Black: Real source, Grey: WFS focused source

Distance (m)	***RT60 (s)***		***EDT (s)***		***D_{50} (%)***		***C_{80} (dB)***		***Tc (ms)***	
	Real	*WFS*	*Real*	*WFS*	*Real*	*WFS*	*Real*	*WFS*	*Real*	*WFS*
0.5	0.6	0.61	0.35	0.51	87.4	77.77	-0.24	-0.56	24.3	43.7
1.0	0.62	0.66	0.4	0.46	84.35	81.61	-0.28	-0.44	29.7	37.3
1.5	0.61	0.64	0.42	0.42	83.21	80.23	-0.32	-0.41	30.1	35.2
2.0	0.61	0.65	0.45	0.44	82.5	81.68	-0.39	-0.40	35.4	34.2

Table 6.2 Predictions of RT60 and EDT for both the real source and WFS focused source modelled scenarios in FDTD simulations

It can be seen from Table 6.2 that both the RT60 and EDT predictions are very similar for both real source and WFS focused source for each distance. For both of these parameters the values from the simulation fall within the perceptual difference limen (Bork, 2000) as given in Table 6.3.

Parameter	**Difference Limen**
RT60(t30)	5% (>2s), 0.1s (<2s)
EDT	5% (>2s), 0.1s (<2s)
C_{80}	1dB
D_{50}	5%
Tc	10ms

Table 6.3 Difference limen of some common room acoustic parameters, from (Bork, 2000)

Confirming these difference limen, Kuttruff (2000) also states that there is little point in quoting RT60 measurements to an accuracy greater than *0.1s* due to inaccuracies in measurement errors therefore perceptually real and focused sources can be considered equivalent in terms of reverberation time. The same can be considered the case for EDT. Thus it can be concluded that for the modelled room there will not be a perceptible difference between the temporal decay of the energy from a real source or a WFS focused source. The other parameters in Table 6.2 show that there is a small difference in the clarity of the sound in the room produced by a real source and a WFS focused source. The focused source is subject to pre-echoes as described in the previous chapter and as such there is a larger difference in the measured centre

time. The pre-echoes are also responsible for the decrease in clarity as determined from the lower values of D_{50} and C_{80}. These differences are small however and it should be remembered that these results are only based on one particular room geometry. There is however evidence to suggest (when combined with perceptual tests mentioned in Chapter 5) that the clarity of the source is affected by the pre-echo artefacts anyway and therefore the room may not have such an effect on the clarity.

6.3.5 Room Effects in the frequency domain

WFS is formulated such that the reproduction space is assumed to be source free and as such it is commonly thought to only be an appropriate to reproduction technique in anechoic environments. However, as will be demonstrated in this section, room effects can improve the perception and colouration as a result of the aliasing.

The WFS system was modelled in FDTD both in anechoic conditions (Fig. 6.10) and also in the room (Fig. 6.11) and the results compared.

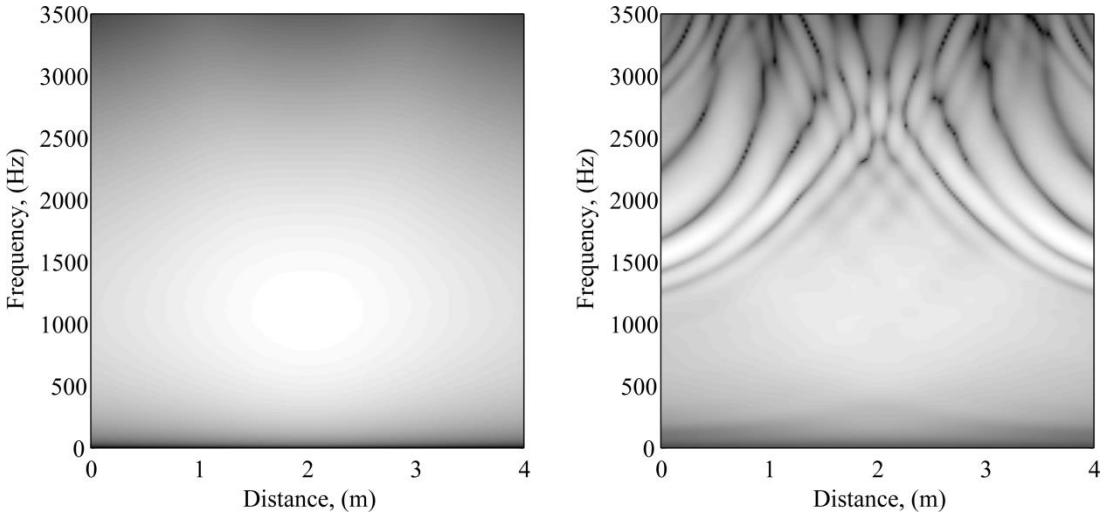


Fig. 6.10. FDTD prediction of a real source (left) and a WFS focused source (right) in anechoic conditions with source geometry given by Fig. 6.6. The colour bar is clipped between *-50dB* (black) and *0dB* (white)

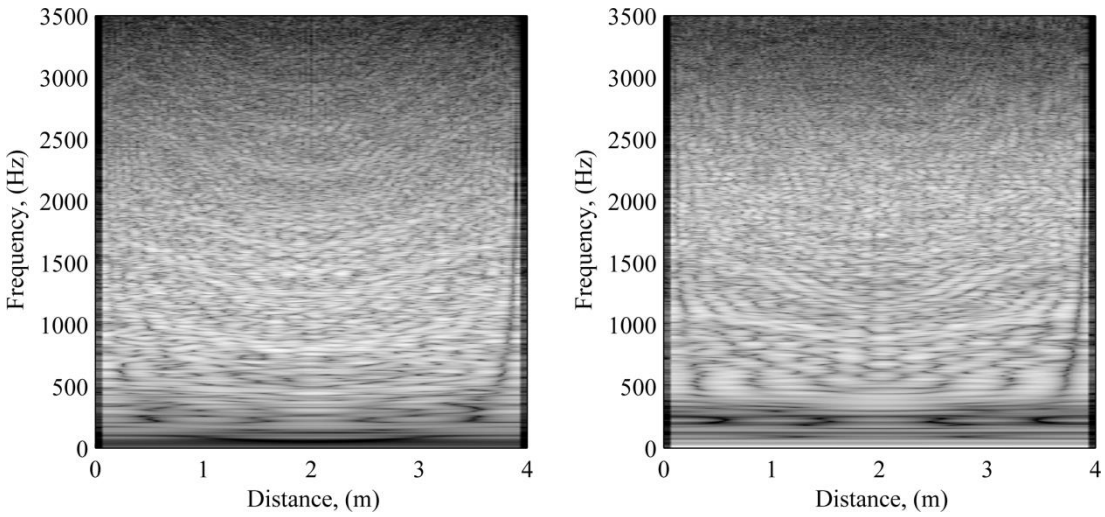


Fig. 6.11. Frequency response across a line or 155 receivers, 1m in front of a real source (left) and a WFS focused source (right) modelled using FDTD with a room geometry as shown in Fig. 6.6 with the source at 1.5m. The colour bar is clipped between -50 (black) and 0dB (white)

Comparing the results from Fig. 6.10 and Fig. 6.11 it can be seen that the effect of the WFS reproduction in the room is to even out the effects of colouration resulting from spatial aliasing effects. In the anechoic case the level of the reproduced wave field with respect to frequency varies greatly across the listening space, whereas for the real room case the multiple reflections from the surfaces of the room mean that the global effect of the colouration of the WFS system is evened out. This is emphasised still further when the difference in between the two cases is plotted as shown in Fig. 6.12 where once again there is no obvious pattern in the energy distribution across the room and the error is largely diffuse in nature.

This finding is in agreement with informal listening tests performed both in an anechoic environment and also in a listening room in which participants complained about the colouration and ‘phasing’ effect as they moved in the anechoic chamber but not when presented with the same audio scene in the listening room environment.

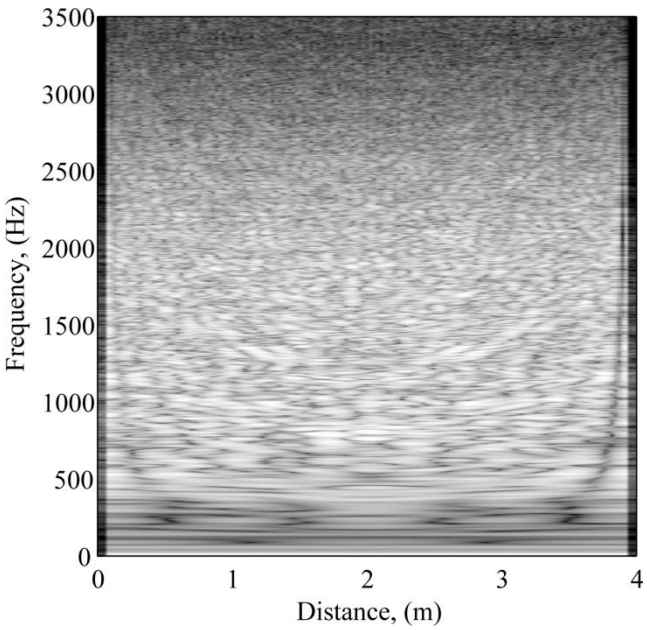


Fig. 6.12 Difference in frequency response between a real and focused source radiating in a room registered across a line or 155 receivers, 1m in front of the source modelled using FDTD with a room geometry as shown in Fig. 6.6 with the source at 1.5m. The colour bar is clipped between -50 (black) and 0dB (white)

6.3.6 Room Effects in the spatial domain

The effects of room reflections can be shown clearly in the spatio-temporal domain by employing a multi-trace impulse response technique such that all of the impulse responses measured by a line of receivers are aligned side by side to demonstrate the temporal evolution of the spatial response along the line of receivers. This can be seen in Fig. 6.13 for both a real source and WFS focused source modelled in an anechoic environment. From this figure it is clear that the aliased focused source produces pre-echoes seen as additional wave fronts arriving at a given listener location before the direct wave front. These additional wave fronts produce multiple reflections when radiating in a room as seen in Fig. 6.15.

Fig. 6.13 exhibits some low level reflections in the reproduction space (seen as faint lines arriving after the direct pulse); these are not artefacts of the focused source rendering but rather are caused by the imperfection of the FDTD PML (see section 6.2.1.3) resulting in reflections at the medium boundary.

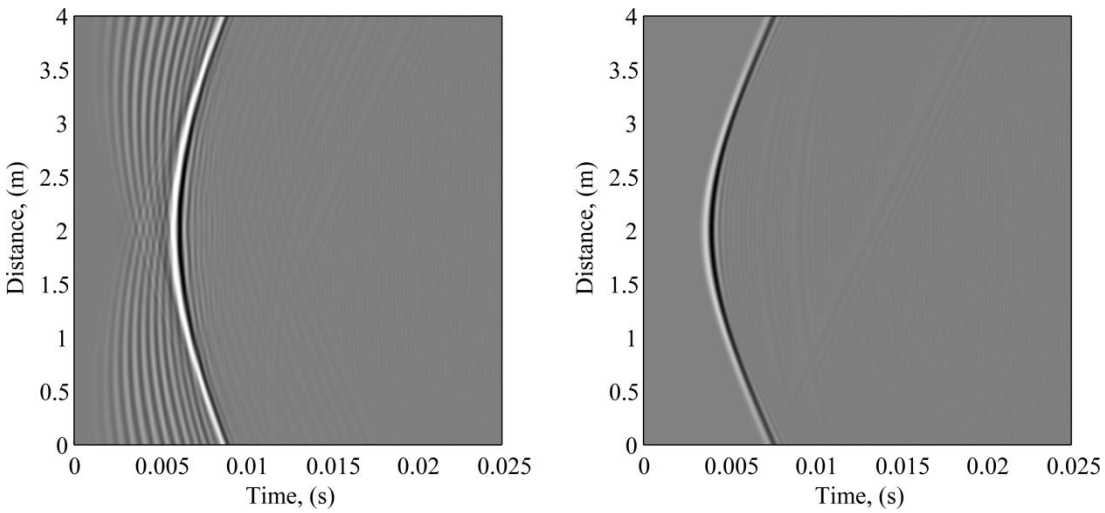


Fig. 6.13 Space time plot of a real source (right) and a WFS focused source (left) at *[0m, 0.5m]* according Fig. 6.6 with a *4m* long array of 155 receivers, *1m* in front of the source position in anechoic conditions

Modelling both an WFS focused source and a real source in a real room as shown in Fig. 6.15 and Fig. 6.14 respectively, demonstrates the effects of the room on the spatio-temporal response of the system. In this case the source was placed 1m in front of the loudspeaker array.

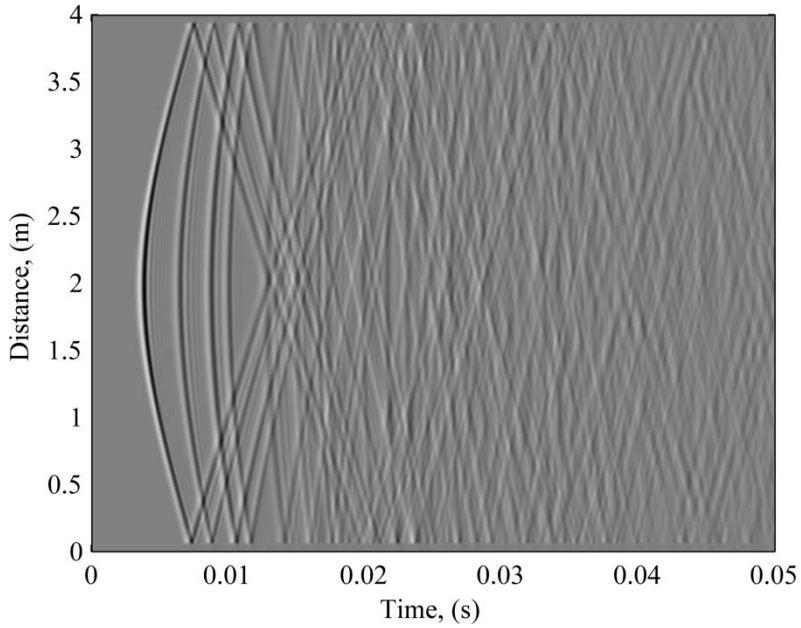


Fig. 6.14 Space time plot of a real source at *[0, 1m]* with a *4m* long array of 155 receivers, *1m* in front of the source position

It can be seen from Fig. 6.14 how the pattern of reflections changes with time. It clearly shows the different areas of the room impulse response. The direct sound wave front arrives first at all positions along the line of receivers and is followed by several early reflections from the boundaries of the room which can be seen to come from specific directions. These reflections get close together as time goes on until such a point that they are almost indistinguishable from each other as the sound field becomes progressively diffuse in nature.

When the same scenario is plotted for a focused source as shown in Fig. 6.15, the effects of spatial aliasing are clearly visible as additional unwanted wave fronts radiating in all directions. In this case, the direct sound no longer arrives first at all locations, on the contrary, it is the pre-echo artefacts that arrive first from the edges of the loudspeaker array, these can be seen particularly clearly when they are spatially aliased. The consequence of these aliased contributions is that there are many more wave fronts propagating in all directions and as such the sound field becomes diffuse more quickly than for the real source case. Reproducing the point source with secondary sources as a WFS focused source means that additional reflections are seen in the room as the radiation from the loudspeakers is reflected form the wall behind the array. Consequently there reflection density again increases faster than the real source case as there are more acoustic sources in the room. Interestingly however all of the same reflections as for the real source are present in the reflection pattern from the focused source, there are just extra reflected components resulting from the array and the spatially aliased contributions.

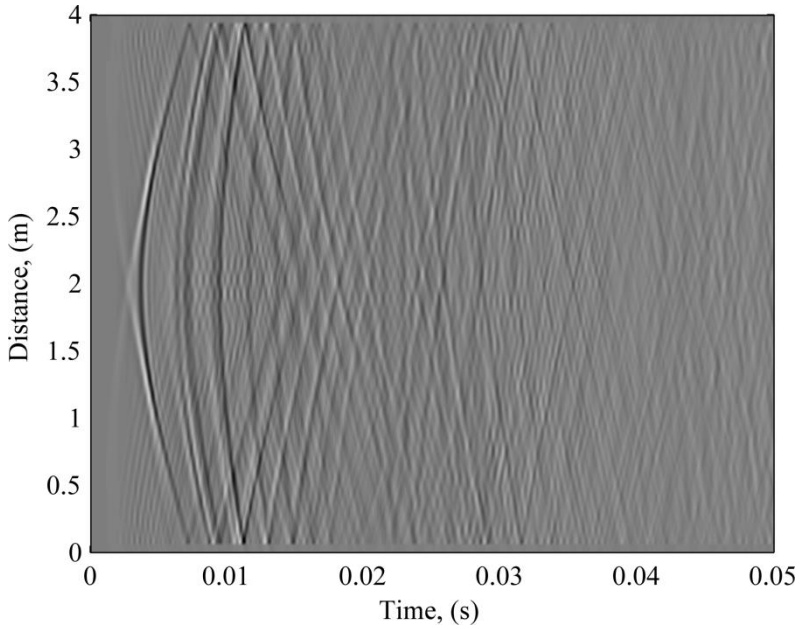


Fig. 6.15 Space time plot of a focused source at *[0, 0.5m]* with a *4m* long array of 155 receivers *1m* in front of the source position

The wave front curvature at the edges is also slightly less than for a true spherical source as a result of the truncated length of the secondary source array. Fig. 6.15 also shows that the lateral reflections are not as strong for the focused source case as for the real source. This is a consequence of focused sources not being omnidirectional and thus not radiating as much lateral energy as a true point source.

A further interesting feature of these plots is the effect of pre-echoes (as can be seen when comparing Fig. 6.15 and Fig. 6.13a). The prominence of the pre-echoes when the source is reproduced in a room is less that when reproduced anechoically. The reason for this is the many lateral reflections from the side walls reflecting the individual pre-echo contributions from the loudspeakers. The superposition of these many reflections (with the direct loudspeaker contributions) cause a more diffuse sound field and thus the pre-echoes are less defined. This provides some explanation as to why colouration as a result of the pre-echoes is less pronounced with a focused source in a room than in the anechoic case (see Fig. 6.10 and Fig. 6.11).

6.4 A qualitative view of real reflections from focused sources

As previously mentioned, Wittek (2009) pointed out that the real reflections from a focused source are not necessarily the same as those from a real source and are actually dependant on the listening position, this means the real reflections can provide incorrect localisation cues and in particular can affect the accuracy of the distance perception of focused sources. The reflections from WFS focused sources will also depend on the source geometry and the shape of the loudspeaker array. Fig. 6.16 shows diagrammatically the first order reflections from both a real source and a focused source. It can be seen that some of the reflections are in the correct positions but there will be additional reflections from the edges of the array and also from the pre-echo artefacts (as shown in the previous section). The position of the reflection behind the loudspeaker array is the most problematic as it will not necessarily line up with the respective reflection from a real source (except for very specific geometries).

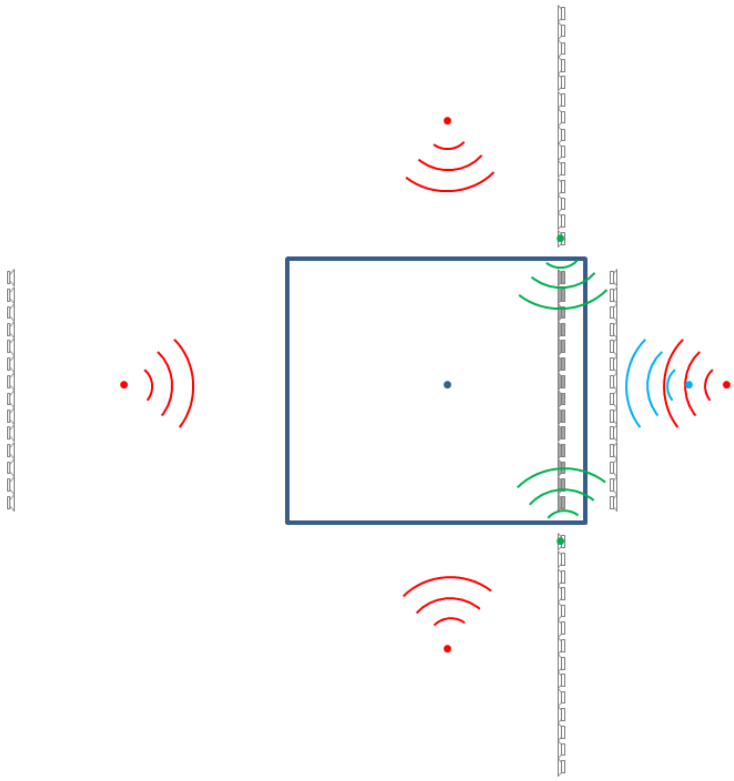


Fig. 6.16 Diagram of the reflections of a rectangular room in the horizontal (reproduction plane). The red image sources show reflections, common to both real and synthesised sources and green image sources will only be present for the WFS focused sources, blue sources are only present for real sources

When curved arrays are used for the reproduction, the reflections produced by the focused source will differ from the linear case, as the radiation from the rear of the loudspeakers will not be a mirror image of the virtual source but will rather have a characteristic more similar to a real source (depending on the position of the virtual source with respect to the array). For the special case of a circular array reproducing a focused source in the centre of the array the reflected sources will be similar to those of a real source Fig. 6.16.

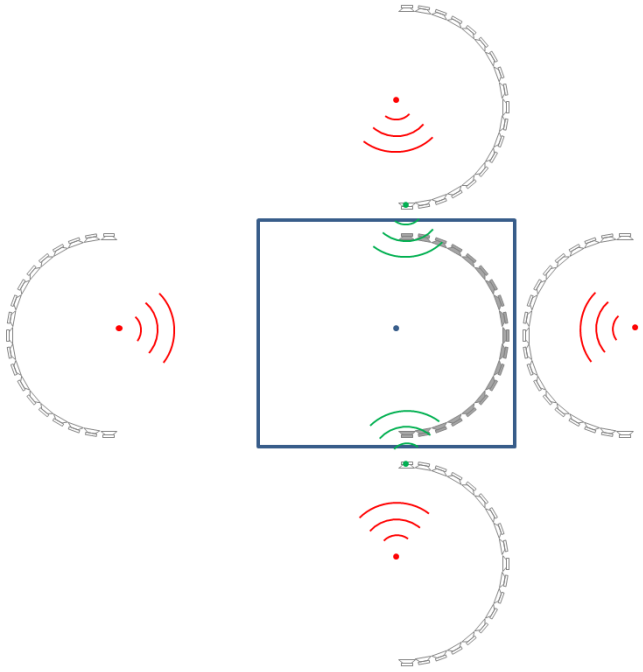


Fig. 6.17 Diagram of the reflections of a rectangular room in the horizontal (reproduction plane) with a focused source rendered by a circular array. The red image sources show reflections, common to both real and synthesised sources and green image sources will only be present for the WFS focused sources, the blue image source are present only for real sources

This special case is interesting as it very closely approximates the case of true source reproduction and subsequent reflections. For the example case that a point source is reproduced at a location in the centre of a loudspeaker array, the source reflections from the back wall will exactly match those from a real source in the same location although there will be additional reflections from the edges of the arrays and from the floor and ceiling which will not match those of a real source which could provide

miss-leading cues especially in terms of source depth information particularly above the spatial aliasing frequency.

6.5 Analysing the effect of elevated loudspeakers for focused source reproduction

For many practical applications of WFS systems, the loudspeakers cannot be placed at ear height as would be desired, due to the location of visual equipment such as screens etc. The consequence of this is that the system generates additional elevation cues such that all sources are localised above the intended reproduction plane. For scenarios such as the Octave system described in section 3.6.5, where the sound field is accompanied by a fully immersive visual display this can be less of a problem as the localisation cues are complimented by the visual cues. The additional elevation cues may not be problematic for sources modelled far away from the listeners but for focused source reproduction however there will be an apparent source position miss-match as a result of using the elevated loudspeakers which will be noticeable as the source is close to the listener. The consequence of rendering focused sources with elevated arrays therefore is that the location of the pressure maximum will not be at the intended position, as shown in the diagram in Fig. 6.18.

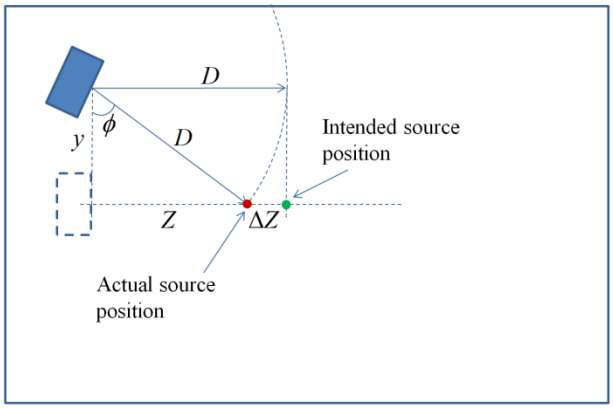


Fig. 6.18 Diagrammatic representations of source position miss-match by using elevated loudspeakers

Fig. 6.18 demonstrates the geometry of the problem of using elevated loudspeakers. If the intended source position is a given distance, D from the loudspeaker array, then depending on the height of the loudspeakers, y the position of the source will actually be, Z where

$$Z = \sqrt{D^2 - y^2} \tag{6.21}$$

So, if elevated loudspeakers are used for the intended reproduction in the horizontal plane a correction, ΔZ in the focused source position of the source must be applied in the driving function where:

$$\Delta Z = D - \sqrt{D^2 - y^2} \tag{6.22}$$

The errors in the position rendering of focused sources using elevated loudspeakers can be plotted as shown in Fig. 6.19 as a function of loudspeaker height and distance of virtual source from the loudspeaker array. As the intended distance of the focused source from the loudspeaker array increases, the error in source position increases.

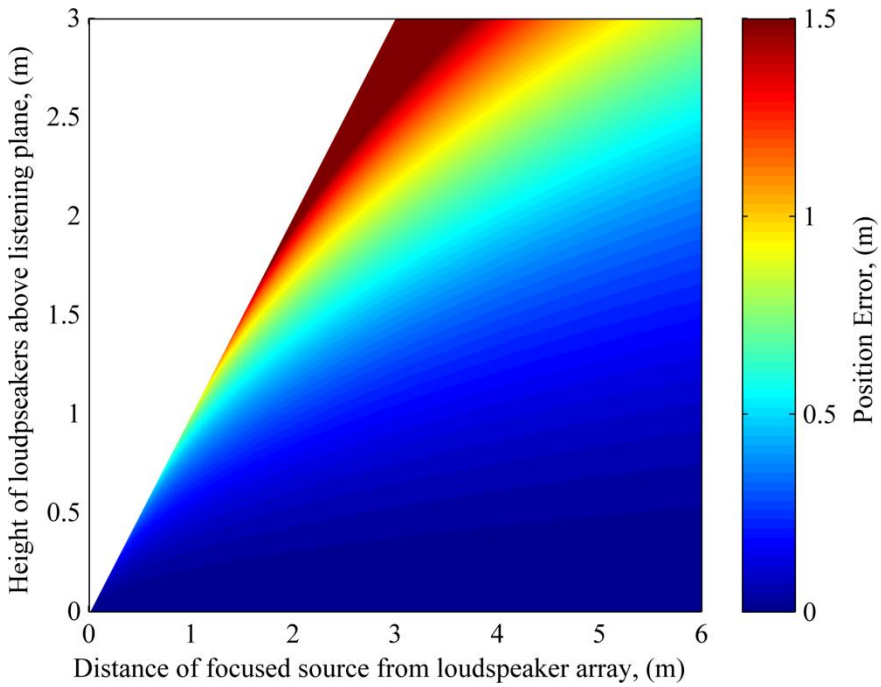


Fig. 6.19 Errors in rendered focused source position as a function of loudspeaker height above listening plane and distance of source in front of loudspeakers. Large error values are clipped as shown by the colour bar

Fig. 6.19 demonstrates that as a result of the elevated loudspeakers there is a region or set of distances at which it is not possible to render a focused source, this is

because the focal length in this region is less than the distance, *D*, meaning that the pressure maxima categorising the position of the focused source is not able to be in the horizontal/listener plane. This is likely to cause listeners to localise the source based on the elevation cues rather than in the listening plane.

The problems associated with elevated loudspeakers are more significant for focused source reproduction than for normal WFS sources because the angular error will be greater as is shown in Fig. 6.20. The angular error for a non-focused source, ϕ_{nf} is always the angle between the listener and the loudspeakers regardless of source position, whereas the angular error for a focused source, ϕ_f is dependent on the virtual source position and is greater than ϕ_{nf} .

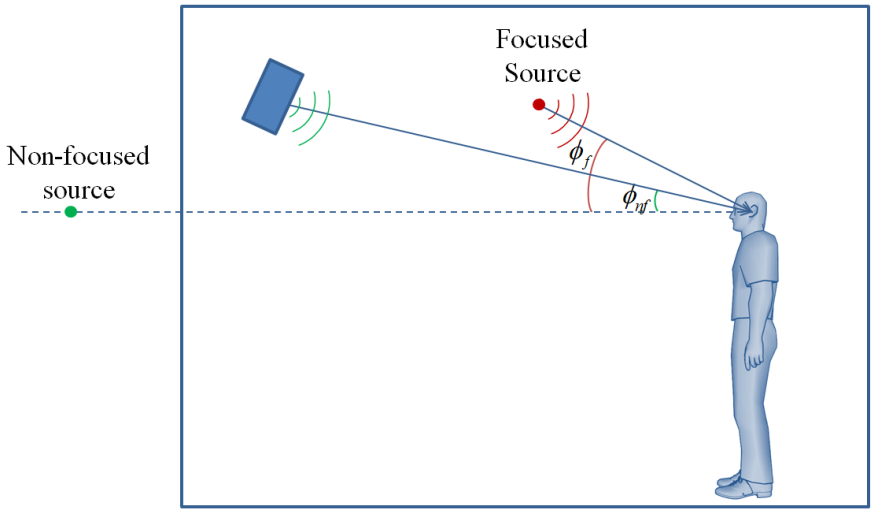


Fig. 6.20 Illustration of increased problems of elevated loudspeakers when rendering focused sources as opposed to non-focused sources

Elevated loudspeakers will therefore introduce an error in vertical localisation, the effect of which may be decreased to some extent by using additional loudspeakers at a lower elevation to bring the localised sound image down a bit. This would not exhibit correct wave front curvature at the lower level but may provide enough additional acoustic energy to prevent all sound source being localised too high. This is a topic of further research.

6.6 Scattering of the focused wave field

The scattering of synthetic sound fields has been investigated by Ahrens and Spors (2011) who used a spherical harmonic description of the intended sound field and the scattered field to produce plots of the total summed field. Their results demonstrated that the scattering off a sphere placed in a sound field of a plane wave produced by a spherical array of loudspeakers within the reproduction space was equivalent to the scattering produced by the same sphere by an analytical plane wave. They did however demonstrate that for the case of 2.5D sound field reconstruction that the shadow region produced by the presence of the sphere was more pronounced than in the natural case.

Here the approach to analysing the scattered field is different and is more versatile for analysing the scattered pressure for different source types and positions of the sphere.

6.6.1 Theory

The scattering from a rigid sphere is a topic that has been well documented when the incident field is that of a plane wave, the theory for which was first formulated by Rayleigh (1896). In this thesis the theory for the scattering by a sphere is given when the incident field is that of a *spherical source*, following (Bowman et al., 1987; Kirkeby et al., 1998).

The boundary condition for a rigid sphere states that there will be zero velocity on the surface of the sphere, such that.

$$\left.\frac{\partial S_{Tot}}{\partial r}\right|_{r=a} = 0 \qquad (6.23)$$

Where S_{Tot} is the total pressure given as

$$S_{Tot}(\omega, r) = S_{FF}(\omega, r) + S_{Scat}(\omega, r) \qquad (6.24)$$

With S_{FF} as the free field pressure of a monopole and S_{Scat} is the scattered pressure and the source geometry given in Fig. 6.21.

The sound radiation from a monopole source is given as (Morse and Ingard, 1987):

$$S_{FF}(\omega, R) = j\omega\rho_0 q \frac{e^{-jkR}}{4\pi R} \tag{6.25}$$

With ρ_0 being the density of air and q being the source strength. As before, k is the acoustic wavenumber given as $k = \omega/c$ where ω is the angular frequency of the source and R is the distance of the source to the receiver point. The spatially invariant terms can be grouped together as an amplitude constant, V such that

$$S_{FF}(\omega, R) = V \frac{e^{-jkR}}{R}, \tag{6.26}$$

where

$$V = \frac{j\omega\rho_0 q}{4\pi}. \tag{6.27}$$

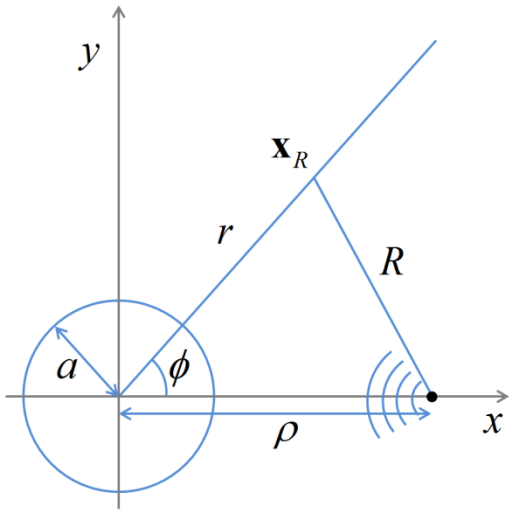


Fig. 6.21 Source geometry for the derivation of the field of a point source scattered by a rigid sphere

Equation *(6.26)* can be expanded using the series expansions found in Abramowitz and Stegun (1965, eq.10.1.45 and 10.1.46, p440),

$$\frac{\sin(kR)}{KR} = \sum_{m=0}^{+\infty}(2m+1)\, j_m(kr) j_m(k\rho) P_m(\cos\phi) \tag{6.28}$$

and

$$-\frac{\cos(kR)}{KR} = \sum_{m=0}^{+\infty}(2m+1)\, j_m(kr) y_m(k\rho) P_m(\cos\phi) \tag{6.29}$$

where

$$R = \sqrt{r^2 + \rho^2 - 2r\rho\cos\phi} \tag{6.30}$$

and the source geometry is given in Fig. 6.21. From *(6.29)* and *(6.30)*, the free field pressure in *(6.26)* can be expanded such that the sound field at the receiver position, *r* given as

$$S_{FF}(\omega, r) = -jkV\sum_{m=0}^{+\infty}(2m+1)\, j_m(kr)\left[\, j_m(k\rho) - j y_m(k\rho)\right] P_m(\cos\phi) \tag{6.31}$$

Where $j_m(z)$ and $y_m(z)$ are the m^{th} order spherical Bessel function of the first and second kind respectively and are related to the m^{th} order Bessel functions of the first and second kind ($J_m(z)$ and $Y_m(z)$) by *(6.32)*.

$$j_m(z) \equiv \sqrt{\frac{\pi}{2z}} J_{m+1/2}(z), \quad y_m(z) \equiv \sqrt{\frac{\pi}{2z}} Y_{m+1/2}(z) \tag{6.32}$$

(6.31) is not used directly for the calculation of the total field of a point source scattered by a rigid sphere as it converges very slowly and thus requires a high value, *m* so is very computationally expensive. Thus *(6.26)* is used for the calculated free field pressure produced by the point source. However *(6.31)* is useful for deriving the scattered pressure if one imposes the boundary condition in *(6.23)* to *(6.31)*.

The scattered pressure from the sphere can be considered as an infinite sum of waves propagating outwards from the centre of sphere and is given by (Crighton et al., 1992, p.533-535) as.

$$S_{Scat}(\omega,r) = kV\sum_{m=0}^{+\infty} b_m\left[j_m(kr) - jy_m(kr)\right]P_m(\cos\phi) \qquad (6.33)$$

The constants b_m can be determined by applying the boundary condition in *(6.23)* to *(6.31)* which is given by (Kirkeby et al., 1998) as:

$$b_m = j(2m+1)\frac{j_m(k\rho) - jy_m(k\rho)}{1 - j\left(y'_m(ka)/j_m(ka)\right)} \qquad (6.34)$$

Where, j'_m and y'_m are the first order differentials of the first and second order spherical Bessel functions expressed as:

$$\begin{aligned} j_m{}'(z) &= \frac{1}{2}\left(j_{m-1}(z) - \frac{j_m(z) + z j_{m+1}(z)}{z}\right) \\ y_m{}'(z) &= \frac{1}{2}\left(y_{m-1}(z) - \frac{y_m(z) + z y_{m+1}(z)}{z}\right) \end{aligned} \qquad (6.35)$$

6.6.2 Implementation

The scattered pressure can thus be computed easily in MATLAB®. It should be noted that the theory presented here is a simplification based on the specific geometry given in Fig. 6.21 but as a sphere is symmetrical and the result will be the same irrespective of the source angle it is a simple matter to perform a coordinate transformation such that the source and receiver can be placed anywhere in the modelled reproduction space. For the non-simplified case of the scattered pressure from a rigid sphere the reader is referred to (Bowman et al., 1987).

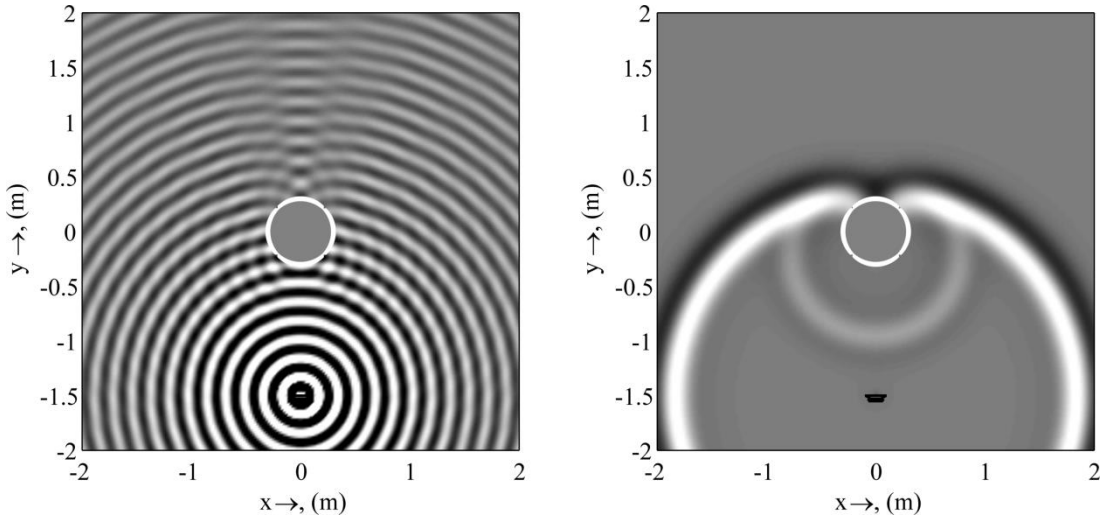


Fig. 6.22 A snapshot in time of a monopole point source scattered by a rigid sphere for both the monochromatic case $f = 2000Hz$ (Left) and the narrowband case (Right). In both cases the source position is $[0m, -1.5m]$ with the sphere of radius, *0.3m* positioned at $[0m, 0m]$.

The simulation in Fig. 6.22 shows that some of the energy is diffracted around the sphere with a smaller portion being reflected. As expected, for the monochromatic case a clear shadow region can be identified behind the sphere with the propagation in this area being delayed due to the extra distance travelled by the diffracted waves. The narrow band case demonstrates that the resultant wave field exhibits direct sound, diffracted components and some specula reflections off the surface of the sphere as expected.

6.6.3 Modelling the scattering of WFS sources

This model can easily be extended to model the sound field of multiple point source scattered by a rigid sphere such that the total sound field is a summation of the free-field and scattered pressures of all of the point sources:

$$S_{Tot}(\omega, r) = \sum_{n_{PS}=0}^{N_{PS}} \left[S_{FF,n_{PS}}(\omega, r) + S_{Scat,n_{PS}}(\omega, r) \right] \qquad (6.36)$$

6.6.3.1 Spheres in the divergent zone

Utilising *(6.36)* it is possible to place a sphere at an arbitrary location in the wave filed created by the superposition of many secondary sources. Placing a sphere in the divergent region of a focused source reveals very similar results to those obtained using an analytical point source.

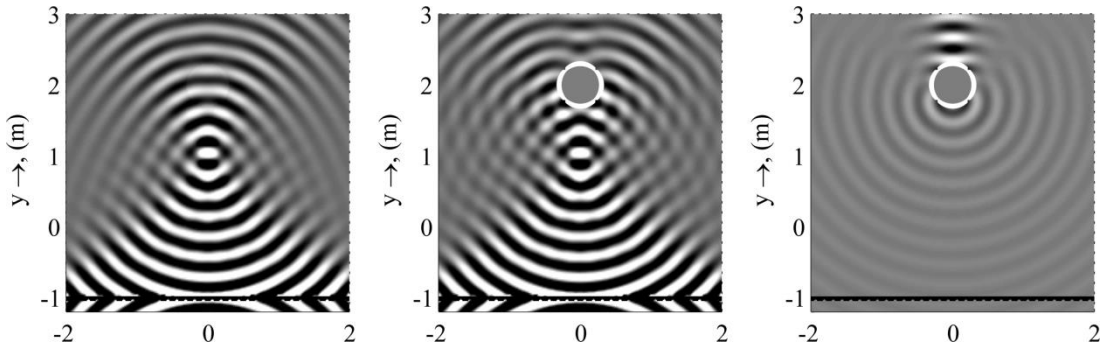


Fig. 6.23 Monochromatic WFS focused source at $\left[0m,1m\right]$ and $f=1200Hz$. Free-field case (left), total pressure (middle) and scattered pressure (Right), with a sphere of radius *0.3m* in the divergent region of the focused source at $\left[0m,2m\right]$.

As with Fig. 6.22a Fig. 6.23b exhibits a clear shadow region behind the sphere where the amplitude is reduced. Additionally the wave front is delayed with respect to the rest of the wave front in this region as a consequence of the diffraction around the sphere. This result is unsurprising as it has already been shown that WFS synthesised wave fields below the aliasing frequency are physically accurate.

6.6.3.2 Spheres in the convergent zone

Whilst the scattering effects of sphere in the divergent region of a focused source field has been shown to be equivalent to that of a real source it useful to see what the effects of a scattering object are in the convergent region of a focused source as this could affect how well the acoustic energy focuses at the intended location. This analysis is important as it reveals whether objects and in particular, people in the room affect the rendering of the focused sources. The results from the simulations (shown in Fig. 6.24) demonstrate the case of a sphere with a radius of *0.0875m* which is often used as the radius of a sphere to approximate a human head.

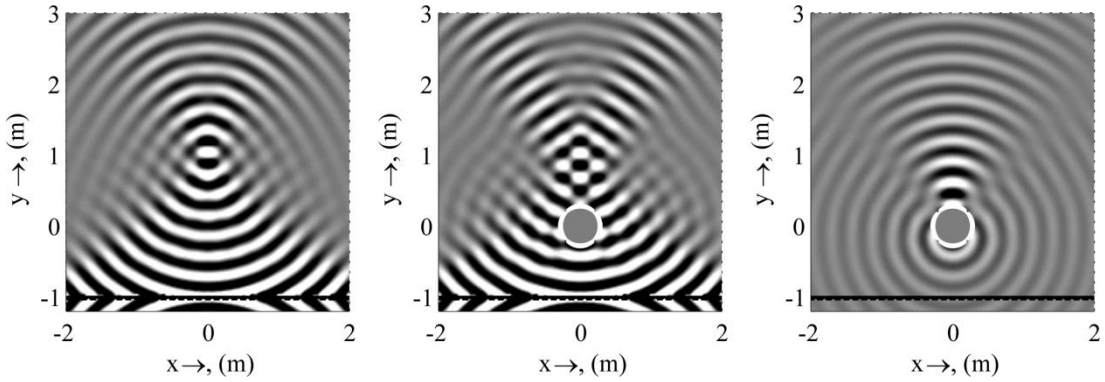


Fig. 6.24 Monochromatic WFS focused source at $[0m, 1m]$ and $f = 1200Hz$. Free-field case (left), total pressure (middle) and scattered pressure (Right), with a sphere of radius *0.3m* in the convergent region of the focused source at $[0m, 0m]$.

It interesting to note from Fig. 6.24 that when a scattering object is placed in the convergent region of a focused source, the shadow region leads rather than lags the free-field energy. This is due to the time reversal involved with the rendering of point sources such that the scattered energy arrives before the free-field energy. As expected once again the shadow region exhibits a reduction in sound pressure level as the rigid sphere scatters some of the energy specularly causing it to act as a barrier.

6.6.4 Correcting for the scattering

To some degree it is possible to correct for the scattering of an object that is in the convergent region of a focused source, by combining standard time-reversal/phase conjugation techniques with the WFS driving function. The correction demands that the transmission path between the focused source position and each of the loudspeakers is known with the scattering object present. For this, a source is placed at the intended focused source position and the pressures at each of the loudspeaker locations are recorded with the scattering object in place and time-reversed (or phase-conjugated in the frequency domain). These time reversed signals are then radiated from the loudspeakers with the scattering object at the same location as before. Due to acoustic reciprocity (Kinsler et al., 2000b) the signals from the loudspeakers will retrace their paths back to focus at the virtual source position. The

result is that the radiated pressure at the focused source position more closely approximates that of the intended source.

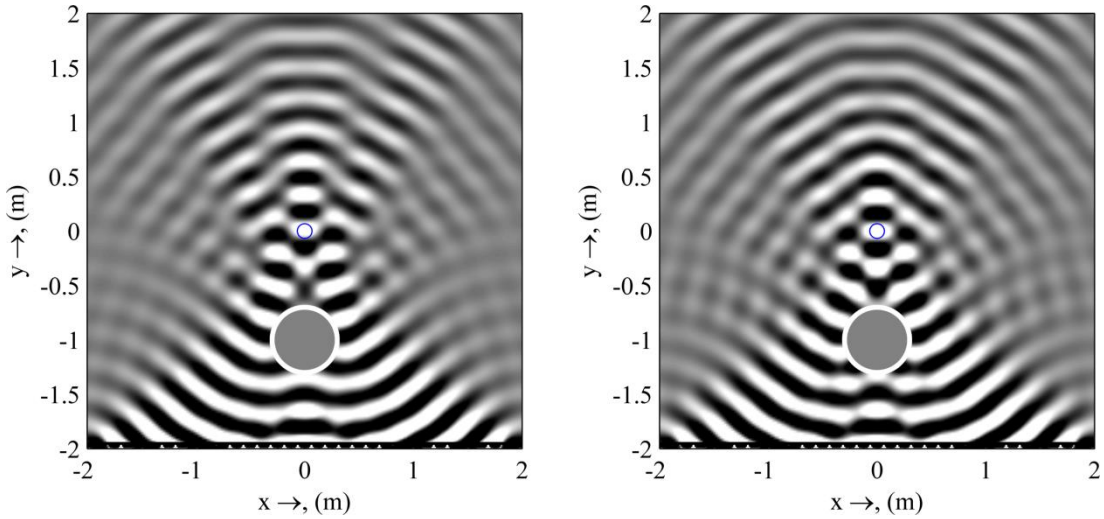


Fig. 6.25 Uncorrected (left) and corrected field (right) of a monochromatic WFS focused source at $[0m, 0m]$ and $f = 1200Hz$ with a sphere of radius *0.3m* in the convergent region of the source at $[0m, -1m]$.

Fig. 6.25 shows the corrected field from a focused source with a sphere present in the convergent region of the source. The simulation shows that there is now no longer a phase lead in the shadow region of the sphere as this has been corrected for through the use of the phase conjugation method. There is however still a decrease in the amplitude of the reproduced wave field in the shadow region of the sphere. Thus the effects of the scattering of the sphere cannot be completely removed. The reason for this is because some of the energy incident to the sphere is scattered specularly and thus the transfer function between the source position and the loudspeaker positions does not completely describe the propagation of the scattered focused source. An analysis of the scattered pressure component in specula directions would allow for this energy to be added into the system to improve the shadowing effect of the sphere but this is difficult to do for objects that do not have an analytical solution like a sphere. If the same procedure is performed on a focused source radiating in a room, the results can be more favourable as the reflections from the surfaces of the room mean that the specular reflections are included in the transfer function between in the virtual source and the loudspeakers so that shadow region is less well

pronounced and the wave front curvature more closely approximates that of the ideal focused source where no scattering object is present.

A further limitation of this technique is that it only allows for the correction of acoustically rigid objects. When the scattering object is to some degree absorbing, the situation becomes more complicated as some energy is lost and therefore the complete transfer function at the loudspeaker positions cannot be determined.

Combining these principles within a WFS rendering requires some consideration, for example it should be noted that the frequency correction as described by equation *(3.28)* needs to still be applied to correct for the rendering of a 3D field using a 2D array of secondary sources. Additionally it should be noted that correcting for scattering using the above mentioned time reversal technique is only valid below f_{alias} as the time reversal process measured the transfer function between just one source and one receiver and as such is independent of spatial aliasing. This may not be too problematic because, as shown in section 4.4.1, a characteristic of focused sources is an area near the source position which is aliasing free thus the time reversal approach can still be applied if the sphere is in the un-aliased region of the source.

6.6.4.1 Correcting for the scattering of objects with more complex geometry

More complicated scattering objects can also be taken into account using the FDTD methods presented earlier in this chapter. As described in section 6.2.2.1 the FDTD model enables DirectX models to be voxelised into the FDTD grid allowing for the prediction of scattering and diffraction of objects with complex geometry. Thus the FDTD model can be used to test the method of correcting for scattering. The principles remain the same of calculating the pressures from a point source at the intended source location at each loudspeaker with the scattering object in place, time reversing the signals and the re-radiating them from the loudspeaker positions with the scattering object still in place.

Initially a sphere is modelled in the FDTD simulation in the same manner as described in section 6.2.2.1. The sphere is situated in the centre of a room *1.5m* in front of a 25 element loudspeaker array reproducing a focused source to a point *0.5m* behind the sphere (the sphere is in the convergent region of the focused source). A slice is taken through the centre of the room in the horizontal plane as shown in Fig. 6.26.

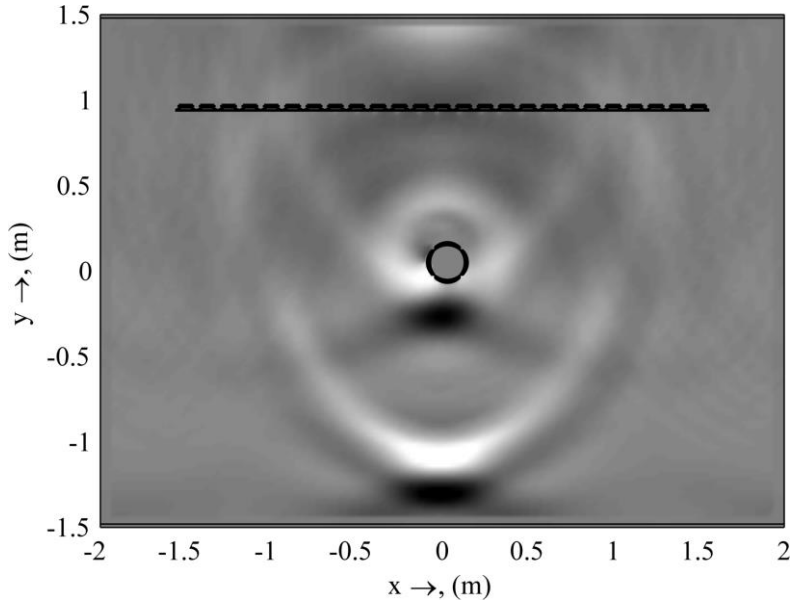


Fig. 6.26 FDTD simulation of a sphere in the convergent region of a focused source being reproduced in a 3D room environment with realistic reflecting boundaries.

The simulation demonstrates that for a sphere of this radius in the convergent region of a focused source being reproduced in a real room environment, the scattering from the sphere has little effect upon the focused source wave front. The specular reflection is seen as a small wave front travelling back towards the loudspeaker array, but there is not much deformation of the direct/intended wave front.

Fig. 6.27 3D model of a sofa imported into the FDTD model

For larger, more complicated geometries such as the sofa shown in Fig. 6.27 the effects on the rendering of the focused source can be much more pronounced and can prevent an accurate focusing. When the effects of scattering are not addressed and corrected for as shown in Fig. 6.28, the wave front curvature of the intended focused source is incorrect and in fact more closely approximates a source at the position of the scattering object (in this case, a sofa). If correction is applied using the time reversal approach described above, it is possible to once again render a focused source at the correct location as can be seen from Fig. 6.29 although there is a compromise in the rendering with the addition of contributions arriving before the direct energy.

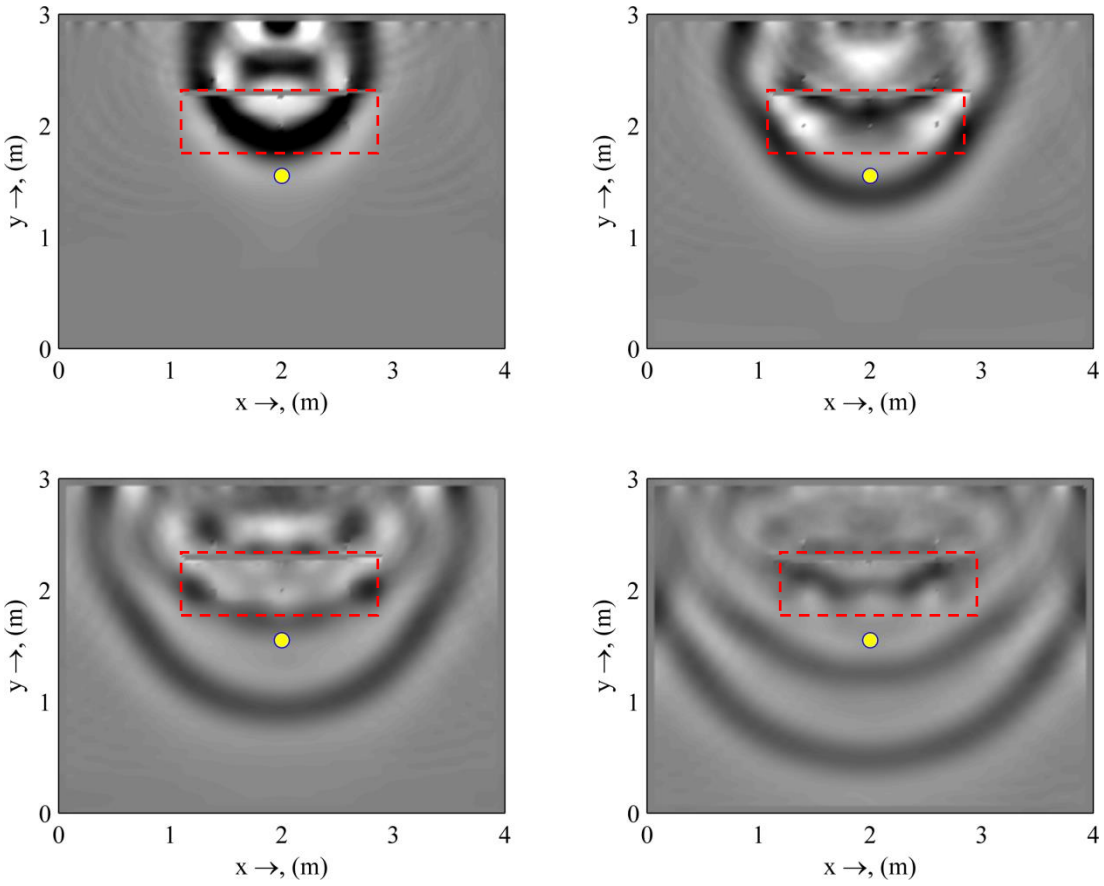


Fig. 6.28 Time evolution of the sound field reproduced from the scattering of a sofa in the convergent region of a focused source. The dashed red box shows the position of the sofa and the yellow circle shows the intended virtual source position.

Fig. 6.29 demonstrates a focused source rendered at the position of the yellow circle with a sofa in between the loudspeakers and the source position. The time evolution plot shows that using the time reversal approach for reproduction in a room, the scattering can, to some extent, be corrected for. Performing this correction does produce a focused source at the correct location (with the correct wave front curvature) but there are additional spatial artefacts caused by the reflections from the early part of the time reversed impulse which cause a spatial distortion of the wave field. This an unavoidable consequence of using time reversal techniques in rooms where there is a significant portion of energy radiated into the room before the intended impulse (Yon et al., 2003a). This also has consequences in terms of the perception of the sound especially for transient signals and for highly reverberant environments where the contributions arriving before the main impulse will be particularly pronounced.

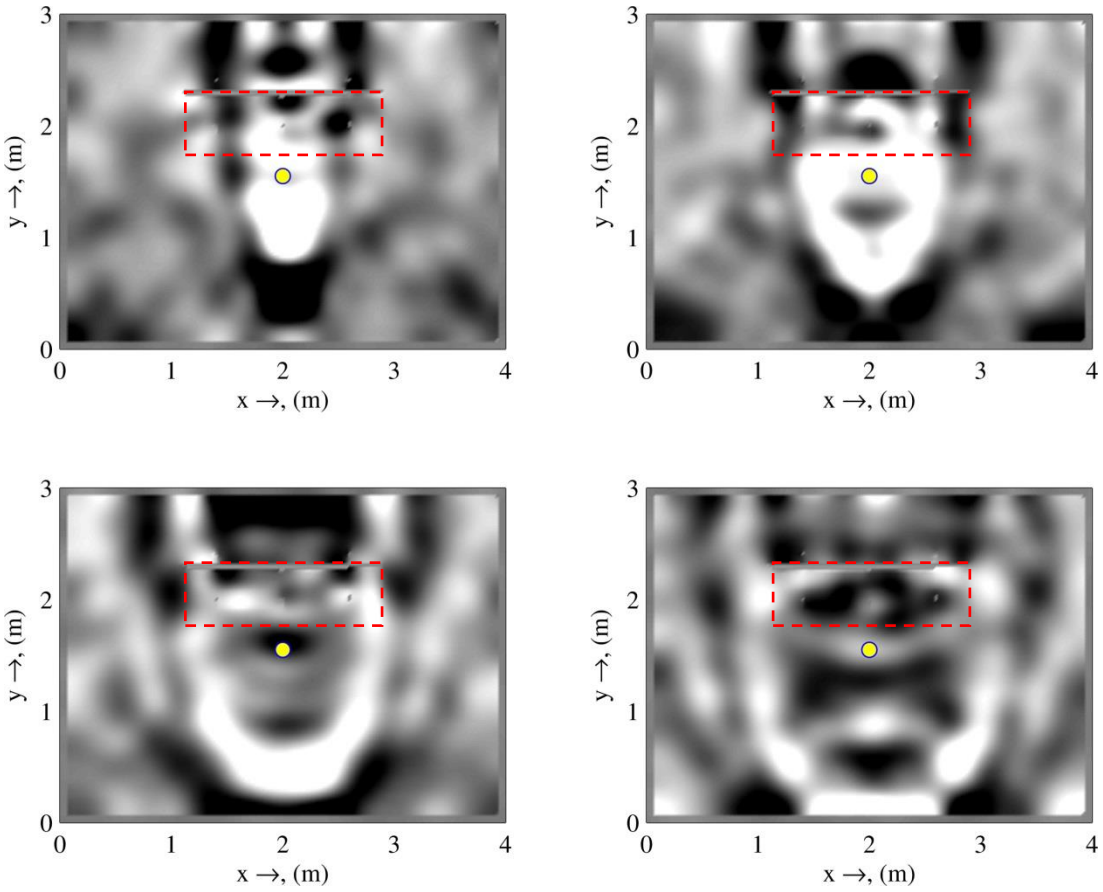


Fig. 6.29 Time evolution of the sound field reproduced from the correction for the scattering of a sofa in the convergent region of a focused source. The dashed red box shows the position of the sofa and the yellow circle shows the virtual source position.

In order to apply this technique in the strictest sense, one would need an infinite amount of time to account for all of the possible reflected paths, this is clearly not possible or desirable and thus a compromise has to be reached for the number of reflections taken into account and the time taken for the rendering of the source. As an initial simulation, the first order reflection only can be taken into account as these are the most problematic and can in particular lead to comb filtering and other adverse perceptual attributes in the reproduced sound field. Thus the amount of time between when the loudspeakers become active and when the focused source radiates, depends on the longest path of the first order reflections and hence upon the dimensions of the reproduction room. The merits of increasing the order of reflections will reap diminishing returns as the temporal structure of the signal becomes less important, the amplitude of the reflections becomes less and small changes within the room (such as people) will have a larger temporal affect and hence the effects of the reflections may not be accurately corrected for.

Clearly there is more work to be done in order to fully correct for scattering objects in the convergent region of an intended focused source, but the time reversal technique presented here demonstrates a proof of concept which could be used to improve the source localisation in compromised situations. This approach may even be used to remove the effect of the room's reflection on the rendering of a focused source. The goal in this case would not be to have a rendered focused source with no room effects at all but rather to have a focused source where only the divergent part was affected by the room response, thus better approximating a real source.

Application

An application of such a correction would be the use of focused sources as virtual loudspeakers as so-called *virtual panning spots* (Theile et al., 2003). Fig. 6.30 shows an example of using virtual panning spots for a variable loudspeaker configuration in a recording studio.

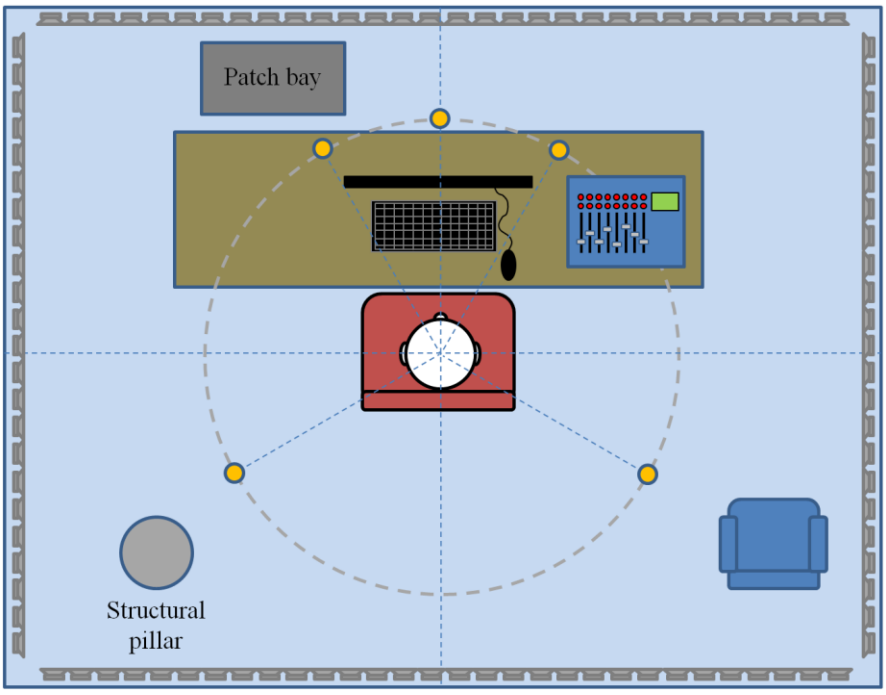


Fig. 6.30 WFS focused source used as virtual loudspeakers (represented as orange circles), with scattering objects in the convergent region of the virtual sources

A setup like this would allow for content to be mixed in a variety of different spatial audio formats such as 2-channel stereo or 5.1 as shown here, by switching between virtual source positions (shown by the orange circles in Fig. 6.30). It is conceivable that in such an example there may be pieces of furniture or other objects in the room that would be in the convergent region of the focused sources and would hence have to be corrected for. In this case the position of the focused sources would not change as they could be placed in the optimum position with respect to the listener which would mean that the transfer function between virtual source and secondary source would only have to be determined once. In this case it may be worthwhile to perform a laser scan of the object and to input the geometry into an FDTD model to work out the required transfer function between each loudspeaker and the intended focused source so that the scattering objects could be corrected for.

This technique would be unsuitable for situations where moving focused sources were part of the intended soundscape as one would first have to calculate the transfer function between the primary and secondary source positions before the source could be rendered. However, if the source positions are stationary, the transfer function calculations (or measurements) can be made in advance and can then be applied to

the signal before rendering. Practically, this would most likely involve convolving the source signal with a measured or calculated impulse response between the source positions. It should be noted that the source function must first be reversed before the simulation take place so that when it is time reversed, the signal plays as intended.

Another application of correcting for the scattering of WFS focused sources is for spatial audio in combination with a video screen as shown in Fig. 6.31. A problem with standard techniques is that it is difficult to make audio sources match the location of the visual sources because it would require audio sources to be in front of the screen or for the use of an acoustically transparent screen, both of which are difficult to achieve in practice. However using WFS focused sources it would be possible to render sources just in front of the screen such that there was no audio and visual position miss-match. In order to achieve this, the scattering effects of television screen would have to be corrected for as described above. By rendering an array of these scattering corrected sources, a virtual WFS array could be realised (as shown in Fig. 6.31) which could then be used to create virtual sources at any location that are not affected by the scattering from the TV screen. This would require techniques similar to the local sound field synthesis methods proposed by Spors and Ahrens (2010) which use focused sources as virtual secondary sources to control the sound field within a smaller area.

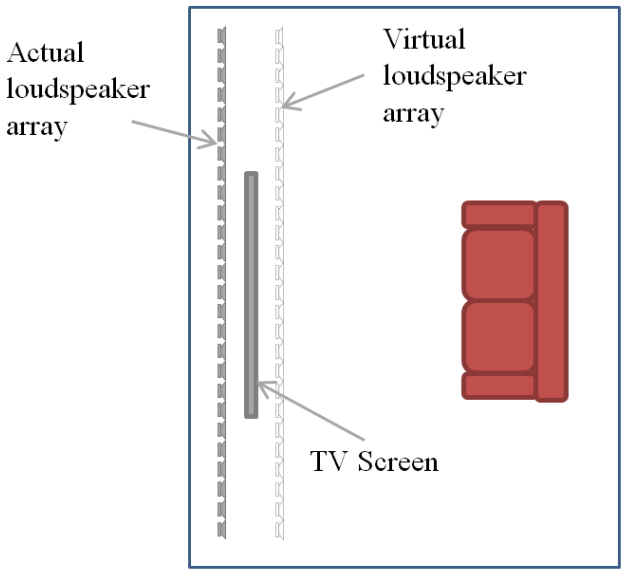


Fig. 6.31 Conceptual diagram, using an array of focused sources to create a virtual WFS array

If the target audio format was a technique other than WFS, the focused sources could act as virtual loudspeakers (Theile, 2004). This would only require the measurement of one set of transfer functions as the focused sources would always be rendered to the same positions.

6.7 Conclusions

This chapter has presented the theoretical and practical aspects of rendering WFS focused sources in non-ideal reproduction situations. It has been shown that the finite difference time domain (FDTD) is a useful modelling technique that can be used for the analysis of the effects of the real room acoustics on the WFS rendering. This is particularly important for focused sources where listeners have a tangible source within the room aiding their comparison between the virtual source and an equivalent real source which draws attention to any incongruity between the virtual and real sources. Several FDTD models have been run that allow different room and source geometries to be modelled. The models demonstrate that despite being generated with remote arrays of secondary source, the room acoustic parameters in the time domain are the same for both real and focused sources. Analysing the effects in the frequency domain demonstrate how the room serves to even out the spatial aliasing artefacts produced by the WFS array and consequently make them less perceptible than for the same sources in an anechoic environment. Showing the results of the FDTD predictions in the spatial domain using the multi-trace impulse response technique demonstrates how the reflections from a focused source become more diffuse quicker than for a real source, helping a sense of spaciousness. It has also been shown that the reflection pattern from a focused source is different from a real source at the same position although with some strong similarities.

The effects of elevated loudspeakers used for the WFS rendering of focused source has also been analysed with a correction factor derived to account for the position error that will be in inherent in the reproduction of focused sources. It has been further illustrated the inability of the system to render focused sources close to the

array if elevated sources are used. The position error is a function of the height of the secondary sources and the intended position of the focused source in the space.

The chapter concluded with an analysis of the scattering of focused sound fields with spherical objects using an analytical approach and more complex geometries using further FDTD models. It has further been demonstrated that to some extent the scattering in the convergent region of a focused source can be corrected for by using time reversal principles which has applications of using WFS for rooms with obstacles in the way of the intended listening area.

7 Conclusions and Further Work

7.1 Conclusions

This thesis has presented a detailed analysis of focused source reproduction using wave field synthesis (WFS) and demonstrated several improvements to the rendering thereof. The second chapter of this thesis put the described work into context by highlighting some of the key spatial audio principles and current techniques relevant to the work undertaken here. The third chapter presented a detailed theory of wave field synthesis demonstrating the derivation of the Kirchhoff-Helmholtz equation, Rayleigh I integral and driving functions of the major source types used in WFS. This chapter also introduced some of the errors inherent in WFS systems including spatial aliasing and truncation artefacts and described some analysis techniques used throughout the rest of the thesis. Following this a detailed analysis of the physical attributes of focused sources has been presented. Particular attention has been given

to the derivation of the important driving functions, including point sources, dipole and piston sources. A new secondary source selection criterion has been demonstrated which allows focused sources to be angled in a chosen direction, additionally the criterion also prevents asymmetric propagation of virtual sources. The concept of focal point shift has also been presented and the derivation of a geometrical correction that can be applied in the driving function to better position focused source particularly in the low frequency region or when the rendering is done with short arrays. This chapter also analysed and quantified some of the errors and characteristics specific to focused sources including spatial aliasing, truncation, colouration and amplitude errors. The spatially aliased field is analysed and a formula for predicting the un-aliased region of focused source reproduction is given. An analysis of pre-echoes is given with the derivation of the direct-to-pre-echo ratio which can be used as a rough metric of the localisation performance of a given source, this is used for an analysis of the use of large arrays for the virtual source reproduction. Other issues such as the phase response, focal spot size, the rendering of multiple focused sources and moving sources are also covered.

Chapter 5 reviewed the perceptual characteristics of focused sources and described a series of perceptual tests that were undertaken to determine the localisation accuracy of focused source with respect to listener angle. It is further shown that improvements to the rendering can be made by including first order image sources into the WFS rendering as virtual point sources. This technique has been shown to greatly increase the range of listener angles at which the virtual source can be accurately localised.

Chapter 6 covered the rendering of focused sources in compromised (non-ideal) scenarios. Much of the analysis was done using the finite difference time domain method of which the theory is presented. The effects on focused source reproduction in real rooms are presented. It is shown that pre-echo artefacts cause more reflections and increase the reflection density of the reproduced sound field. It is however shown that room acoustic parameters are similar for both the real and virtual source cases when modelled in a real room. Results do however suggest that there is a decrease in clarity resulting from the rendering of focused sources with respect to

real sources. Colouration artefacts are shown to be averaged out by the room's reflections, making the perception of focused sources in real rooms better than in an anechoic environment. This chapter also covers the rendering of focused sources using elevated loudspeakers and demonstrates how this limits the possible positions of the source. Also covered is the scattering of the reproduced field with a scattering object in both the convergent and divergent field of the source. It is shown that to some extent the scattering in the convergent region can be corrected for using time reversal methods. The theory is extended to scattering objects with more complex geometry with the finite difference time domain model being used to generate the time reversed signals and subsequent re-radiation to correct for the scattering.

7.2 Further work

- More work should be carried into the optimal method of rendering moving focused sources and the resulting subjective preferences.

- Further perceptual tests are needed, especially in real room environments and to quantify the perceptual effects of colouration in the reproduced wave field.

- New methods of pre-echo reduction should also be investigated

- An analysis of the perceptual effects of elevated loudspeakers and possible methods of reducing the localisation errors inherent in the rendering.

- More work is also required on the combining time reversal methods to help alleviate reproduction errors resulting from the scattering of objects with complex geometry.

References

Abramowitz, M., and Stegun, I. A. (**1965**). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Dover publications.

Ahrens, J. (**2012**). *Analytic Methods of Sound Field Synthesis*, Springer, Heidelberg, Germany.

Ahrens, J., and Spors, S. (**2007**). "Implementation of directional sources in wave field synthesis," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 66 – 69.

Ahrens, J., and Spors, S. (**2008**). "Reproduction of Moving Virtual Sound Sources with Special Attention to the Doppler Effect," 124th Conv. Audio Eng. Soc., Amsterdam, The Netherlands.

Ahrens, J., and Spors, S. (**2008**). "Notes on Rendering Focused Directional Virtual Sound Sources in Wave Field Synthesis," 34th Annual German Conference on Acoustics (DAGA), Dresden.

Ahrens, J., and Spors, S. (**2008**). "Focusing of virtual sound sources in higher order Ambisonics," 124th Conv. Audio Engi. Soc., Amsterdam.

Ahrens, J., and Spors, S. (**2010**). "Sound Field Reproduction Using Planar and Linear Arrays of Loudspeakers," IEEE Transactions On Audio Speech And Language Processing, **18**, 2038–2050. doi:10.1109/TASL.2010.2041106

Ahrens, J., and Spors, S. (**2011**). "On the Scattering of Synthetic Sound Fields," 130th Conv. Audio Eng. Soc., London, UK.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am., **65**, 943 – 950.

Baalman, M. A. J. (**2008**). *On Wave Field Synthesis and electro-acoustic music, with a particular focus on the reproduction of arbitrarily Shaped Sound Sources* PhD Thesis, TU Berlin,TU Berlin, p. PhD Thesis, TU Berlin.

Barron, M. (**1971**). "The subjective effects of first reflections in concert halls—The need for lateral reflections," J. Sound and Vibration, **15**, 475 – 494.

Batke, J.-M., Spille, J., Kropp, H., Abeling, S., Shirley, B., and Oldfield, R. G. (**2011**). "Spatial Audio Processing for Interactive TV Services," 130th Conv. Audio Eng. Soc., London, UK.

Bech, S. (**1998**). "Spatial aspects of reproduced sound in small rooms," J. Acoust. Soc. Am., **103**, 434 – 445.

Begault, D. (**1994**). *3-D sound for virtual reality and multimedia*, Academic Press.

Begault, D. R., Wenzel, E. M., and Anderson, M. R. (**2001**). "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," J. Audio Eng. Soc, **49**, 904 – 916.

Bell, A. G. (**1876**). *Improvement in telegraphy*, US Patent,174465.

Berenger, J. P. (**1994**). "A perfectly matched layer for the absorption of electromagnetic waves," J. comp. phys., **14**, 185 – 200.

Berkhout, A. ., De Vries, D., and Vogel, P. (**1993**). "Acoustic control by wave field synthesis," J. Acoust. Soc. Am., **93**, 2764 – 2778.

Berkhout, A. J. (**1987**). *Applied Seismic Wave Theory*, Elsevier Science Pub. Co. Inc.,, New York, NY.

Berkhout, A. J. (**1988**). "A Holographic Approach to Acoustic Control," J. Audio Eng. Soc, **36**, 977–995.

Berkhout, A., Vogel, P., and Vries, D. de (**1992**). "Use of wave field synthesis for natural reinforced sound," 92nd Conv. Audio Eng. Soc, Vienna.

Bianchi, L., Antonacci, F., Canclini, A., Sarti, A., and Tubaro, S. (**2012**). "A Psychoacoustic-Based Analysis of the Impact of Pre-Echoes and Post-Echoes in Soundfield Rendering Applications," International Workshop on Acoustic Signal Enhancement,.

Blauert, J. (**1997**). *Spatial Hearing (rev. ed.)*, MIT Press, Cambridge, MA.

Blumlein, A. D. (**1931**). "Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems," British Patent,.

Boone, M., and Horbach, U. (**1999**). "Virtual surround speakers with wave field synthesis," 106th Conv. Audio Eng. Soc., Munich.

Boone, M. M., and De Bruijn, W. P. J. (**2000**). "On the Applicability of Distributed Mode Loudspeaker Panels for Wave Field Synthesis Based Sound Reproduction," 108th Conv. Audio Eng. Soc., Paris.

Boone, M., Verheijen, E. N. G., and Jansen, G. (**1996**). "Virtual Reality by Sound Reproduction Based on Wave Field Synthesis," 100th Conv. Audio Eng. Soc., Copenhagen, Denmark.

Boone, M., Verheijen, E., and Tol, P. Van (**1995**). "Spatial sound-field reproduction by wave-field synthesis," J. Audio Eng. Soc, **43**, 1003–1012.

Bork, I. (**2000**). "A comparison of room simulation software-The 2nd round robin on room acoustical computer simulation," Acta Acustica United With Acustica, **86**, 943 – 956.

Born, M., and Wolf, E. (**1999**). *Principles of Optics*, Cambridge University Press, Cambridge, UK, 7th ed.

Boré, G. (**1956**). "Kurzton-Meßverfahren zur punktweisen Ermittlung der Sprach verständlichkeit in lautsprecherbeschallten Räumen / ['Short tone measurement techniques for point-wise determination of speech intelligibility in speakers echoed heritage areas'],," Dissertation, Technische Hochshule, Aachen,.

Botteldooren, D. (**1995**). "Finite-difference time-domain simulation of low frequency room acoustic problems," J. Acoust. Soc. Am., **98**, 3302–3308.

Bowman, J. J., Senior, T. B. A., and Uslenghi, P. L. E. (**1987**). *Electromagnetic and acoustic scattering by simple shapes (Revised edition)*, Hemisphere Publishing Corp., New York, NY, USA.

Bradley, J. S., and Soulodre, G. A. (**1995**). "The influence of late arriving energy on spatial impression," J. Acoust. Soc. Am., **97**, 2263 – 2271.

Bradley, J. S., and Soulodre, G. A. (**1995**). "Objective measures of listener envelopment," J. Acoust. Soc. Am., **98**, 2590 – 257.

De Bruijn, W. (**2004**). *Application of Wave Field Synthesis in Videoconferencing*, PhD Thesis, TU Delft, The Netherlands,.

De Bruijn, W. P. J., and Boone, M. M. (**2003**). "Application of Wave Field Synthesis in Life-size Videoconferencingv," 114th Conv. Audio Eng. Soc., Amsterdam.

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources II Localization of a broadband source," J. Acoust. Soc. Am., **106**, 1956 – 1968.

Caulkins, T., Corteel, E., and Warusfel, O. (**2003**). "Wave field synthesis interaction with the listening environment, improvements in the reproduction of virtual sources situated inside the listening room," 6th Int. Conf. on Digital Audio Effects (DAFx-03), London.

Chen, X., Schwarz, K. Q., and Parker, K. J. (**1993**). "Radiation pattern of a focused transducer: A numerically convergent solution," J. Acoust. Soc. Am., **94**, 2979 – 2991.

Clement, G., White, J., and Hynynen, K. (**2000**). "Investigation of a large-area phased array for focused ultrasound surgery through the skull," Physics in Medicine and Biology, **45**, 1071–1083.

Cooper, D. H., and Bauck, J. L. (**1989**). "Prospects for Transaural Recording," J. Audio Eng. Soc, **37**, 3 – 19.

Corteel, E. (**2006**). "On the Use of Irregularly Spaced Loudspeaker Arrays for Wave Field Synthesis, Potential Impact on Spatial Aliasing Frequency," 9th

International Conference on Digital Audio Effects (DAFx-06), Montréal, PQ, Canada.

Corteel, E. (**2006**). "Equalization in an Extended Area Using Multichannel Inversion and Wave Field Synthesis," J. Audio Eng. Soc, **54**, 1140 – 1161.

Corteel, E. (**2007**). "Synthesis of directional sources using wave field synthesis, possibilities, and limitations," EURASIP Journal on Advances in Signal Processing,.

Corteel, E., NGuyen, K.-V., Warusfel, O., Caulkins, T., and Pellegrini, R. (**2007**). "Objective and subjective comparison of electrodynamic and map loudspeakers for wave field synthesis," 30th Int. Conf. Audio Eng. Soc., Saariselkä.

Corteel, E., and Nicol, R. (**2003**). "Listening room compensation for wave field synthesis What can be done?," 23rd Int. Conf. Audio Eng. Soc., Copenhagen, Denmark.

Corteel, E., Warusfel, O., Van Zon, R., and De Vries, D. (**2004**). "Multi-Actuator Panel (MAP) Loudspeakers: How to Compensate for Their Mutual Reflections?," 116th Conv. Audio Eng. Soc., Berlin.

Cremer, L. (**1948**). *Die wissenschaftlichen Grundlagen der Raumakustik / ['Lectures on technical acoustics']*, S. Hirzel Verlag, Stuttgart, Germany.

Crighton, D. G., Dowling, A. P., Ffowcs-Williams, J. E., Heckl, M., Leppington, F. G., and Bartram, J. F. (**1992**). *Modern methods in analytical acoustics lecture notes*, Springer-Verlag, London, UK, pp. 533 – 535.

Daniel, J. (**2001**). *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia,* PhD thesis Universitè Paris,PhD thesis Universitè Paris.

Daniel, J. (**2003**). "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format," 23rd Int. Conf. Audio Eng. Soc., Helsingør, Denmark.

Daniel, J., Nicol, R., and Moreau, S. (**2003**). "Further Investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," 114th Conv. Audio Eng. Soc., Amsterdam.

Davidon, W. C. (**1991**). "Variable metric method for minimization," SIAM Journal on Optimization, **1**, 1 – 17.

Dickins, G., Williams, M., and Hanlen, L. W. (**2005**). "On the dimensionality of spatial fields with restricted angle of arrival," IEEE Internation Symposium on Infomation Theory (ISIT), 1033 – 1037.

Drumm, I. A. (**2005**). "The application of adaptive beam tracing and managed DirectX for the visualisation and auralisation of virtual environments," 9th Int. Conf. on Information Visualisation, 961 – 965.

Drumm, I. A., and Lam, Y. W. (**2007**). "Development and assessment of a finite difference time domain room acoustic prediction model that uses hall data in popular formats," Proc. Internoise 07, Istambul, Turkey.

Drumm, I. A., and Oldfield, R. G. (**2010**). "The prediction of synthesised wavefields within realistic room acoustics scenarios," 20th Int. Congress on Acoustics (ICA), Syndey, Australia.

Drumm, I. A., and Oldfield, R. G. (**2013**). "The optimisation of Wave Field Synthesis with adaptive recursive filtering for real deployment scenarios," J. Audio Eng. Soc (Accepted for publication February 2013),.

Edison, T. A. (**1878**). "The phonograph and its future," The North American Review, **126**, 527 – 536.

Elrod, S., Khuri-Yakub, B., and Quate, C. (**1988**). "Acoustic lens arrays for ink printing," US Patent 4,751,530,.

Fazi, F. M., Nelson, P. A., Christensen, J. E. N., and Seo, J. (**2008**). "Surround system based on three dimensional sound field reconstruction," 125th Conv. Audio Eng. Soc.,.

Feddersen, W. E. (**1957**). "Localization of High-Frequency Tones," J. Acoust. Soc. Am., **29**, 988 – 1991.

Fink, M. (**1992**). "Time reversal of ultrasonic fields I Basic principles," IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, **39**, 555 – 566.

Fink, M. (**1999**). "Time-reversal mirrors," J. Phys. D: Appl. Phys., **26**, 1333–1350.

Fink, M. (**2006**). "Time-reversal acoustics," Contempory Mathmatics, **408**, 151 – 175.

Fink, M., Montaldo, G., and Tanter, M. (**2004**). "Time reversal acoustics," IEEE Ultrasonics Symposium, **2**, 850 – 859.

Franck, A., Graefe, A., Korn, T., and Strauss, M. (**2007**). "Reproduction of Moving Sound Sources by Wave Field Synthesis: An Analysis of Artifacts," 32nd Int. Conf. Audio Eng. Soc., Hillerød, Denmark.

Gade, A. C. (**1994**). *Proceedings of the Sabine Centennial Symposium, Cambridge, Mass.,*, Acoustical Society of America, Woodbury, New York.

Gardner, M. (**1969**). "Distance Estimation of 0° or Apparent 0° Oriented Speech Signals in Anechoic Space," J. Acoust. Soc. Am., **45**, 47 –53.

Gauthier, P. A., and Berry, A. (**2006**). "Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory," J. Acoust. Soc. Am., **119**, 2721 – 2737.

Gauthier, P. A., and Berry, A. (**2007**). "Objective Evaluation of Room Effects on Wave Field Synthesis," Acta Acustica united with Acustica, **93**, 824 – 836.

Geier, M., Ahrens, J., and Spors, S. (**2008**). "The Soundscape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," 124th Conv. Audio Eng. Soc.th Conv. Audio Eng. Soc., Amsterdam, The Netherlands.

Geier, M., Wierstorf, H., Ahrens, J., and Wechsung, I. (**2010**). "Perceptual evaluation of focused sources in wave field synthesis," 128th Conv. Audio Eng. Soc, London, UK.

Gerzon, M. A. (**1973**). "Periphony: With-height sound reproduction," J. Audio Eng. Soc, **21**, 2 – 10.

Griesinger, D. (**1997**). "The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces," Acta Acustica united with Acustica, **83**, 721 – 731.

Haas, H. (**1951**). "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache / ['On the influence of a single echo on the audibility of speech']," Acustica, **1**, 49 – 58.

Hamasaki, K., Hiyama, K., and Okumura, R. (**2005**). "The 222 multichannel sound system and its application," 118th Conv. Audio Eng. Soc., Barcelona, Spain.

Hannemann, J., and Donohue, K. D. (**2008**). "Virtual sound source rendering using a multipoleexpansion and method-of-moments approach," J. Audio Eng. Soc, **56**, 473 – 481.

Hartman, W. M. (**1983**). "Localization of sound in rooms," J. Acoust. Soc. Am., **74**, 1380 – 1391.

Hofman, P. (**1999**). "Relearning sound localization with new ears," J. Acoust. Soc. Am., **105**, 1035–1035.

Von Hornbostel, E. M., and Wertheimer, M. (**1920**). "Uber die Wahrnehmung der Schallrichtung [On the perception of the direction of sound]," Sitzungsberichte der preussischen Akademie der Wissenschaften,.

Hulsebos, E. (**2004**). *Auralization using Wave Field Synthesis,* PhD Thesis, TU Delft, The Netherlands,.

Hulsebos, E., De Vries, D., and Bourdillat, E. (**2002**). "Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis," J. Audio Eng. Soc, **55**, 779 – 790.

Hurwitz, M. (**1977**). *Acoustic lens system*, US Patent 4,001,766,4001766.

Ise, S. (**1991**). "A Principle of Sound Field Control Based on the Kirchhoff–Helmholtz Integral Equation and the Theory on Inverse Systems," Acta Acustica united with Acustica, **85**, 78 – 87.

ISO-3382 (**1997**). *Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters*, International Organization for Standardization, Geneva.

Jensen, G. (**1997**). *Focused Wavefields and Moving Virtual Sources by Wavefield Synthesis* TU Delft.

Jordon, V. L. (**1970**). "Acoustical criteria for auditoriums and their relation to model techniques," J. Acoust. Soc. Am., **47**, 408 – 412.

Jot, J.-M., Cerveau, L., and Warusfel, O. (**1997**). "Analysis and synthesis of room reverberation based on a statistical time-frequency model," 103rd Conv. Audio Eng. Soc., New York, Ny, USA.

Katz, B. F. G. (**2001**). "Boundary element method calculation of individual head-related transfer function II Impedance effects and comparisons to real measurements," J. Acoust. Soc. Am., **110**, 2449 – 2455.

Katz, B. F. G. (**2001**). "Boundary element method calculation of individual head-related transfer function I Rigid model calculation," J. Acoust. Soc. Am., **110**, 2440 – 2448.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. (**2000**). *Fundamentals of Acoustics*, John Wiley & Sons, New York, NY, USA, 4th ed.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, John Wiley & Sons, New York, NY, USA, 4th ed.

Kirkeby, O., and Nelson, P. A. (**1993**). "Reproduction of plane wave sound fields," J. Acoust. Soc. Am., **94**, 2992 – 3000.

Kirkeby, O., Nelson, P. A., and Hamada, H. (**1998**). "Local sound field reproduction using two closely spaced loudspeakers," J. Acoust. Soc. Am., **104**, 1973 – 1981.

Komiyama, S., Morita, A., Kurozumi, K., and Nakabayashi, K. (**1991**). "Distance Control System for a Sound Image," 9th Int. Conf. Audio Eng. Soc., Detroit, MI, USA.

Kuntz, A., and Rabenstein, R. (**2004**). "An approach to global noise control by wave field synthesis," 12th European Signal Processing Conference (EUSIPCO), Vienna.

Kuperman, W., Hodgkiss, W., and Song, H. (**1998**). "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am., **103**, 25 – 40.

Kuttruff, H. (**2000**). *Room Acoustics*, (4th Edition, Ed.) Spon Press, Oxford, UK.

Kürer, R. (**1969**). "Zur gewinnung von eizahlkriterien bei impulsmessungeg in der raumakustik," Acustica,.

Lee, S.-R., and Sung, K.-M. (**2003**). "Generalized encoding and decoding functions for a cylindrical ambisonic sound system," IEEE Signal Processing Letters, **10**, 21 – 23.

Lemons, R., and Quate, C. (**1974**). "Acoustic microscope—scanning version," Applied Physics Letters, **24**, 163 – 165.

Lindau, A., Hohn, T., and S.Weinzierl (**2007**). "Binaural resynthesis for comparative studies of acoustical environments," 122nd Conv. Audio Eng. Soc., Vienna, Austria.

Lopez, J. J., Cobos, M., and Pueo, B. (**2010**). "Elevation in wave-field synthesis using HRTF Cues," Acta Acustica United With Acustica, **96**, 340 – 350.

Lucas, B. G., and Muir, T. G. (**1982**). "The field of a focusing source," J. Acoust. Soc. Am., **72**, 1289 – 1296.

López, J. J., González, A., and Fuster, L. (**2005**). "Room compensation in wave field synthesis by means of multichannel inversion," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 146 – 149.

Makov, Y. N., Espinosa, V. J., Sánchez-Morcillo, V. J., Ramis, J., Cruañes, J., and Camarena, F. (**1982**). "Strong on-axis focal shoft and its nonlinear variation in low-Fresnel-Number ultrasound beams," J. Acoust. Soc. Am., **116**, 3618 – 3624.

Makov, Y. N., Sánchez-Morcillo, V. J., Camarena, F., and Espinosa, V. (**2008**). "Nonlinear change of on-axis pressure and intensity maxima positions and its relation with the linear focal shift effect," Ultrasonics, **48**, 678 – 686.

Malham, D. (**2003**). *Space in Music – Music in Space*, PhD thesis University of York, UK,.

Maloney, J., and Cummings, K. (**1995**). "Adaptation of FDTD techniques to acoustic modeling," 11th Annual Review of Progress in Applied Computational Electromagnetics, 724 – 731.

Melchior, F., Brix, S., Sporer, T., Roder, T., and Klehs, B. (**2003**). "Wave Field Syntheses in Combination with 2D Video Projection," 24th Int. Conf. Audio Eng. Soc. on Multichannel Audio, Banff, Canada.

Melchior, F., Sladeczek, C., De Vries, D., and Frohlich, B. (**2008**). "User-dependent optimization of wave Field synthesis reproduction for directive sound Fields," 124th Conv. Audio Eng. Soc., Amsterdam, The Netherlands.

Menzel, D., Wittek, H., Theile, G., and Fastl, H. (**2005**). "The binaural sky: A virtual headphone for binaural room synthesis," 1st VDT-Symposium,.

Meyer, J., and Elko, G. (**2002**). "highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),.

Miller, D. (**1983**). *Transducer acoustic lens*, US Patent 4,387,720,4387720.

Millington, G. (**1932**). "A modified formula for reverberation," J. Acoust. Soc. Am., **4**, 69 – 82.

Moncel, T. du (**1881**). "The telephone at the Paris opera," Scientific American,.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, Academic Press, 5th ed.

Morimoto, M., and Maekawa, Z. (**1989**). "Spaciousness and envelopment in musical acoustics," 13th International Congress on Acoustics (ICA),.

Morse, P. M., and Ingard, K. U. (**1987**). *Theoretical Acoustics*, Princeton University Press, Princeton, NJ, USA.

Nicol, R., and Emerit, M. (**1999**). "3D-sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonics," 32nd Int. Conf. Audio Eng. Soc., Hillerød.

Nielsen, S. (**1991**). *Distance perception in hearing*, Aalborg University Press, Aalborg, Denmark.

Nilsson, M., Soil, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am., **95**, 1085 – 1099.

Oldfield, R., Drumm, I., and Hirst, J. (**2010**). "The Perception of Focused Sources in Wave Field Synthesis as a Function of Listener Angle," 128th Conv. Audio Eng. Soc., London, UK.

Oldfield, R. G. (**2006**). "Improved Membrane Absorbers," MSc (By Research) Thesis, University of Salford, UK,.

Oldfield, R. G., Drumm, I. A., and Hirst, J. (**2009**). "A Finite Difference Time Domain Approach to Analysing Room Effects on Wave Field Synthesis Reproduction," 124th Conv. Audio Eng. Soc., Amsterdam, The Netherlands.

Oldfield, R. G., Hargreaves, J. A., Drumm, I. A., and Moorhouse, A. T. (**2011**). "Implementing Wave Field Synthesis in an ITU Spec Listening Room Part 1: Keeping it at Ear Level," IoA Reproduced Sound, Brighton.

O'Neil, H. T. (**1949**). "The theory of focusing radiators," J. Acoust. Soc. Am., **21**, 516 – 524.

Peretti, P., Cecchi, S., Palestini, L., and Piazza, F. (**2007**). "A novel approach to active noise control based on wave domain adaptive filtering," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Platz, 307 – 310.

Petrausch, S., Rabenstein, R., and Spors, S. (**2005**). "Simulation and visualization of room compensation for wave field synthesis with the functional transformation method," 119th Conv. Audio Eng. Soc., New York, NY.

Poletti, M. A. (**1996**). "The design of encoding functions for stereophonic and polyphonic sound systems," J. Audio Eng. Soc, **44**, 948 – 963.

Pulkki, V. (**1997**). "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," J. Audio Eng. Soc, **45**, 456 – 466.

Rayleigh, J. W. S. (**1896**). *The Theory of Sound, Vol. 2*, Reprinted, Dover, New York, 1945.

Reichardt, W., Abdel, O. A., and Schmidt, W. (**1974**). "Defintion und Meßgrundlage eines objektiven Maßes zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit bei Musikdarbietung," Acustica, **32**, 126 – 137.

Rumsey, F. (**2001**). *Spatial Audio*, Focal Press, Oxford, UK.

Sabine, W. C. (**1922**). *Collected papers on acoustics*, Harvard university press.

Salvador, C. D. (**2010**). "Discrete Wave Field Synthesis Using Fractional Order Filters and Fractional Delays," 128th Conv. Audio Eng. Soc., London, UK.

Schneider, J. B., Wagner, C. L., and Broschat, S. L. (**1998**). "Implementation of transparent sources embedded in acoustic finite-difference time-domain grids," J. Acoust. Soc. Am., **103**, 136 – 142.

Schroeder, M. R. (**1965**). "New Method of Measuring Reverberation Time," J. Acoust. Soc. Am., **37**, 409 – 412.

Snow, W. B. (**1953**). "Basic principles of stereophonic sound," J. SMPTE, **61**, 567 – 589.

Song, M.-H., Choi, J.-W., and Kim, Y.-H. (**2012**). "A selective array activation method for the generation of a focused source considering listening position," J. Acoust. Soc. Am., **131**, EL156 – EL162.

Song, M.-H., and Kim, Y.-H. (**2011**). "Acoustic Measure of Causality Artifacts for Generating Focused Source," 43rd Conf. Audio Eng. Soc., Pohang, Korea.

Sonke, J. J., De Vries, D., and Labeeuw, J. (**1998**). "Variable acoustics by wave field synthesis: A closer look at amplitude effects," 104th Conv. Audio Eng. Soc., Amsterdam.

Sporer, T. (**2004**). "Wave field synthesis-generation and reproduction of natural sound environments," 7th Int. Conf. on Digital Audio Effects (DAFx-04),, Naples.

Spors, S. (**2003**). "Extension of an analytic secondary source selection criterion for wave field synthesis," 123rd Conv. Audio Eng. Soc., New York, NY.

Spors, S., and Ahrens, J. (**2007**). "Analysis of near-field effects of wave field synthesis using linear loudspeaker arrays," 30th Int. Conf. Audio Eng. Soc., Saariselkä, Finland.

Spors, S., and Ahrens, J. (**2008**). "A Comparison of Wave Field Synthesis and Higher-Order Ambisonics with Respect to Physical Properties and Spatial Sampling," 125th Conv. Audio Eng. Soc., San Francisco, CA, USA.

Spors, S., and Ahrens, J. (**2010**). "Local Sound Field Synthesis by Virtual Secondary Sources," 40th Conf. Audio Eng. Soc., Tokyo.

Spors, S., and Ahrens, J. (**2010**). "Analysis and Improvement of Pre-equalization in 25-Dimensional Wave Field Synthesis," 128th Conv. Audio Eng. Soc., London, UK.

Spors, S., and Ahrens, J. (**2010**). "Reproduction of focused sources by the spectral division method," 4th International Symposium on Communications Control and Signal Processing ISCCSP, , doi: 10.1109/ISCCSP.2010.5463335. doi:10.1109/ISCCSP.2010.5463335

Spors, S., and Ahrens, J. (**2011**). "Efficient range extrapolation of head-related impulse responses by wave field synthesis techniques," International Conference on Speech and Signal Processing (ICASP), 49 – 52.

Spors, S., and Buchner, H. (**2004**). "An approach to massive multichannel broadband feedforward active noise control using wave-domain adaptive filtering," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Platz.

Spors, S., Buchner, H., and Rabenstein, R. (**2004**). "Efficient active listening room compensation for Wave Field Synthesis," 116th Conv. Audio Eng. Soc.,.

Spors, S., Buchner, H., Rabenstein, R., and Herbordt, W. (**2007**). "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," J. Acoust. Soc. Am., **122**, 354 – 369.

Spors, S., Buchner, H., and Rabenstien, R. (**2004**). "A novel approach to active listening room compensation for wave field synthesis using wave-domain adaptive filtering," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, (ICASSP'04)., 29 – 32.

Spors, S., and Rabenstein, R. (**2006**). "Spatial Aliasing Artifacts Produced by Linear and Circular Loudspeaker Arrays used for Wave Field Synthesis," 120th Conv. Audio Eng. Soc., Paris, France.

Spors, S., Rabenstein, R., and Ahrens, J. (**2008**). "The theory of wave field synthesis revisited," 124th Conv. Audio Eng. Soc., Amsterdam, The Netherlands.

Spors, S., and Rabenstien, R. (**2006**). "Spatial Aliasing Artifacts Produced by Linear and Circular Loudspeaker Arrays used for Wave Field Synthesis," 120th Conv. Audio Eng. Soc., Paris.

Spors, S., Seuberth, D., and Rabenstein, R. (**2005**). "Multi-Exciter Panel Compensation for Wave Field Synthesis," Annual German Conference on Acoustics (DAGA), Munich.

Spors, S., Wierstorf, H., Geier, M., and Ahrens, J. (**2009**). "Physical and Perceptual Properties of Focused Sources in Wave Field Synthesis," 127th Conv. Audio Eng. Soc., New York, NY, USA.

Springer, J. P., Sladeczek, C., Scheffler, M., Hochstrate, J., Melchior, F., and Frohlich, B. (**2006**). "Combining wave field synthesis and multi-viewer stereo displays," IEEE Virtual Reality Conf., 237 – 240.

Start, E. W. (**1997**). "Direct sound enhancement by wave field synthesis," PhD Thesis, TU Delft, The Netherlands,.

Steinberg, J. C., and Snow, W. B. (**1934**). "Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: Physical Factors," Bell System Technical Journal, **13,** 245 – 258.

Strauss, M., Wagner, A., Walther, A., and Melchior, F. (**2004**). "Generation of highly imersive atmospheres for wave field synthesis reproduction," 116th Conv. Audio Eng. Soc., Berlin.

Strutt, J. (**1907**). "On our perception of sound direction," Philosophical Magazine, **13**, 214 – 232.

Sutin, A., and Johnson, P. (**2005**). "Time reversal acousto-seismic method for land mine detection," Proc. SPIE: Defense and Security, 706–716.

Taflove, A., and Hagness, S. C. (**2000**). *Computational electrodynamics: The Finite Difference Time Domain*, Artech House, Boston, USA, 3rd ed.

Teutsch, H. (**2007**). *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, Springer-Verlag.

Theile, G. (**2004**). "Wave Field Synthesis - A Promising Spatial Audio Rendering Concept," 7th Conf. on Digital Audio Effects (DAFx'04), Naples.

Theile, G., Wittek, H., and Reisinger, M. (**2003**). "Potential Wavefield Synthesis Applications in the Multichannel Stereophonic World," 24th Int. Conf. Audio Eng. Soc. on Multichannel Audio, Banff, Canada.

Thiele, R. (**1953**). "Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen," Acustica, **3**, 291 – 302.

Thomas, J., Wu, F., and Fink, M. (**1996**). "Time reversal focusing applied to lithotripsy," Ultrasonic imaging, **18**, 106 – 121.

Toole, F. (**2008**). *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*, Focal Press, Oxford, UK.

Verheijen, E. N. G. (**1998**). *Sound Reproduction by Wave Field Synthesis*, PhD Thesis, TU Delft, The Netherlands,TU Delft, The Netherlands.

Vogel, P. (**1993**). *Application of Wave Field Synthesis in Room Acoustics*, PhD Thesis, TU Delft, The Netherlands,.

De Vries, D. (**1996**). "Sound reinforcement by wavefield synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics," J. Audio Eng. Soc, **44**, 1120 – 1131.

De Vries, D. (**2009**). *Wave field synthesis*, AES Monograph, New York.

De Vries, D., Start, E. W., and Valstar, V. G. (**1994**). "The Wave Field Synthesis concept applied to sound reinforcement: Restrictions and solutions," 96th Conv. Audio Eng. Soc., Amsterdam.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," The American Journal of Psychology, **62**, 315 – 336.

Wanrooij, M. Van, and Opstal, A. Van (**2005**). "Relearning sound localization with a new ear," The Journal of neuroscience, **25**, 5413 – 5424.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am., **94**, 111 – 123.

Wierstorf, H., Geier, M., and Spors, S. (**2010**). "Reducing artifacts of focused sources in Wave Field Synthesis," 129th Conv. Audio Eng. Soc., San Francisco, CA, USA.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am., **91**, 1648 – 1661.

Williams, E. G. (**1999**). *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic Press, London, UK.

Wittek, H. (**2009**). *Perceptual differences between wavefield synthesis and stereophony*, PhD Thesis, University of Surrey, UK,PhD Thesis, University of Surrey, p. PhD Thesis, University of Surrey.

Wittek, H., Kerber, S., Rumsey, F., and Theile, G. (**2004**). "Spatial perception in Wave Field Synthesis rendered sound fields: Distance of real and virtual nearby sources," 116th Conv. Audio Eng. Soc., Berlin.

Wittek, H., Rumsey, F., and Theile, G. (**2007**). "Perceptual enhancement of wavefield synthesis by stereophonic means," J. Audio Eng. Soc, **55**, 723 – 751.

Wittek, H., Rumsey, F., and Theile, G. (**2007**). "Perceptual Enhancement of Wavefield Synthesis by Stereophonic Means," J. Audio Eng. SocJ, **55**, 723 – 751.

Yee, K. S. (**1966**). "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," IEEE Transactions on Antennas and Propagation, **14**, 302 – 307.

Yon, S., Tanter, M., and Fink, M. (**2003**). "Sound focusing in rooms: The time-reversal approach," J. Acoust. Soc. Am., **113**, 1533 – 1543.

Yon, S., Tanter, M., and Fink, M. (**2003**). "Sound focusing in rooms II The spatio-temporal inverse filter," J. Acoust. Soc. Am., **114**, 3044–3052.

Yuan, X., Borrup, D., Berggeren, M., Wiskin, J., and Johnson, S. (**1999**). "Simulation of acoustics wave propagation in dispersive media with relaxation losses by using FDTD method with PML absorbing boundary condition," IEEE Transactions on Ultrsonics, Ferroelectrics and Frequency Control,.

Zahorik, P. (**2009**). "Perceptually relevant parameters for virtual listening simulation of small room acoustics," J. Acoust. Soc. Am., **126**, 776 – 791.

Zotkin, D. N., and Gumerov, N. A. (**2004**). "Interpolation and range extrapolation of HRTFs [head related transfer functions]," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, (ICASSP'04).,.

Zumpano, G., and Meo, M. (**2007**). "A new nonlinear elastic time reversal acoustic method for the identification and localisation of stress corrosion cracking in welded plate-like structures–A simulation," International journal of solids and structures, **44**, 3666 – 3684.