UNIVERSITY OF MIAMI

WAVE FIELD SYNTHESIS IN THREE DIMENSIONS BY MULTIPLE LINE ARRAYS

By

Matthew N. Montag

A THESIS PROJECT

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science in Music Engineering Technology

Coral Gables, Florida

May 2011

UNIVERSITY OF MIAMI

A Thesis Project submitted in partial fulfillment of
the requirements for the degree of
Master of Science in Music Engineering Technology

WAVE FIELD SYNTHESIS IN THREE DIMENSIONS
BY MULTIPLE LINE ARRAYS

Matthew N. Montag

Approved:

Colby N. Leider, Ph.D. Associate Professor of Music Engineering	Edward Asmus, Ph.D. Associate Dean of Graduate Studies
Will Pirkle Assistant Professor of Music Engineering	Christopher L. Bennett, Ph.D. Postdoctoral Associate of Anesthesiology

MONTAG, MATTHEW M.S., Music Engineering Technology
May 2011

Wave Field Synthesis In Three Dimensions By Multiple Line Arrays

Abstract of a Master's Research Project at the University of Miami

Research Project supervised by Colby N. Leider, Ph.D
No. of pages in text: 82

Wave field synthesis is a powerful method of spatial audio rendering that makes use of the Huygens principle to reproduce physically-accurate wavefronts for virtual sources. Large loudspeaker arrays can be used to synthesize the wavefront of a virtual source that exists outside of the listening room. The technique has traditionally been limited to the horizontal plane due to the prohibitive cost of planar loudspeaker arrays. Multiple line array wave field synthesis is proposed as an extension to linear WFS. This method extends the virtual source space into the vertical dimension using a fraction of the number of loudspeakers required for plane arrays. This thesis details the creation of a cross-platform software environment for wave field synthesis capable of driving a loudspeaker array according to the proposed extension, as well as the construction of a modular, low-cost loudspeaker array that can be adapted to linear, planar, or multiple line configurations.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

Since the invention of the phonograph, the goal of sound production technology has been to create the "perfect reproduction of the original sound." In 1935, the "acoustic curtain" was proposed (Snow 1935, 1955) to provide a window into an acoustic scene (Figure 1.1). It was imagined that a curtain of microphones at the venue could transmit the performance to a corresponding bank of loudspeakers in a remote listening room. At the time, it was not possible to implement the acoustic curtain, but the modern invention of wave field synthesis has brought this idea back to life.

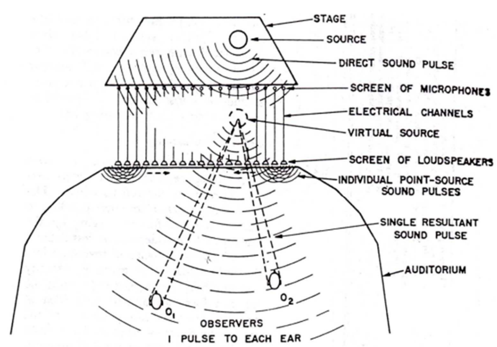


Fig. 2. Ideal stereophonic system. A very large number of very small microphones and loudspeakers would give a perfect reproduction of the original sound.

Figure 1.1: The original "acoustic curtain" concept (Steinberg and Snow 1934).

1.1 A Brief History of Spatial Audio Reproduction

Steinberg and Snow's original work inspired by the acoustic curtain took place at Bell Labs in the 1930s. They found good results by reducing the number of channels to three, a left, right, and center channel. Snow explained that the result of this three-channel configuration was fundamentally different because the three-source configuration does not create the original source wavefront, but rather produces its perceptual effect by the precedence effect (see Chapter 4). The distinction between a single source wave field and a stereo wave field was further explained by Alan Blumlein in 1931. Blumlein showed that a source could be positioned between two loudspeakers by amplitude difference alone. Surround-sound technology was initially commercially driven by the motion picture industry, just as it is today. *Fantasia* was the first motion picture to deliver a stereo soundtrack, in 1939. The rest of the industry did not catch up until ten years later, as widespread theater stereo was not in use until 1950. Stereo vinyl records as well as FM stereo radio became available to the public in the 1960s. Disparate multichannel theater formats evolved in the 1960s and 70s, finally culminating in Dolby Digital 5.1, inaugurated with the 1992 release of *Batman Returns*. Modern extensions to Dolby Digital and similar technologies like DTS and Sony's SDDS specify up to 12 channels.

Outside of the motion picture industry, Keibs and others developed ambiophonic techniques which focused on generating reverberant and ambient signals from separate surround speakers during the early 1960s. Modern ambiophonics is loosely defined as a method of widening the stereo image over two loudspeakers by crosstalk cancellation. In conventional stereo, sound from the left speaker reaches the right ear, and vice versa. Crosstalk cancellation is an attempt to remove this effect and widen the auditory image in a conventional 60 degree separation setup to 180 degrees. The technique may be expanded to multichannel setups by using recursive crosstalk cancellation. (Glasgal and Miller 2006)

Gerzon, Fellgett, Barton, and others pioneered the development of ambisonics during the 1970s. Ambisonics is intended as a general solution to directional sound capture and reproduction, based on extension of Blumlein's stereophonic principles to an arbitrary number of speakers. Ambisonics is capable of reproducing a full sphere spatial sound field, but the sound field is accurate only at the listener location. (Rumsey 2001)

Wave field synthesis made its first appearance in modern literature in the late 1980s with the work of Dr. A.J. Berkhout at TU Delft (Berkhout 1988).

1.2 Focus of the Thesis

I pursued three distinct objectives in my thesis research:

- Build a low-cost, modular, and rapidly configurable loudspeaker array
- Create an open-source, cross-platform wave field synthesis software environment
- Enhance wave field synthesis by expansion to the height dimension

I will leave behind a 48-channel modular loudspeaker array and a cross-platform software environment for wave field synthesis. Together they constitute a flexible experimental research instrument.

The first objective was motivated out of necessity since there was no existing wave field synthesis system available at the University of Miami. I determined to build a novel and versatile array of loudspeakers to remain with the music engineering technology program at the University of Miami, and which could be used for a number of different applications not necessarily related to wave field synthesis. Since this was a major component of the thesis, it warrants a brief introduction, and will be discussed in detail in Chapter 6.

The speaker array constructed for the thesis consists of 12 modules of 4 loudspeakers each. The loudspeaker modules are 20” by 5”, with equal loudspeaker spacing, such that each speaker occupies a 5” by 5” square. The modules can be stacked on the short edge or the long edge, and the individual loudspeakers will maintain positions on the 5” grid.

The modules were designed to be inexpensive, and the cost-per-channel, including amplification, totaled around $22.

Potential research applications of the loudspeaker array include acoustic beam steering, active listening room compensation, active noise control, room acoustic simulation, direct sound reinforcement, ambisonics, and other spatial audio reproduction techniques. The array is well-suited to investigation of listening room compensation and active noise control because of its ability to cover a large perimeter of the sound field. The array could be augmented with an array of microphones to enable real-time sound field control. Finally, the array is flexible enough to be used in creative performance applications.

Several software environments are currently available for wave field synthesis. These will be reviewed in Chapter 5. However, there is no freely available compiled solution for the Windows operating system. In addition, there is no available WFS software environment that performs 3D source positioning for use in plane arrays.

1.2.1 Emphasis on Virtual Reality

Spatial audio is frequently discussed in the context of sound recording and reproduction, and in particular, the reproduction of musical material. However, this field neglects a wide range of applications that motivate more powerful reproduction systems, and for which advanced spatial audio techniques are well-suited. (What constitutes a “powerful” system shall become clear momentarily.) These applications include virtual reality, video games, military simulation, and so on. A common factor among these applications is user interactivity, which stands in contrast to the concert reproduction scenario – a seated listener focusing on static sources in a static environment. The psychoacoustic experience of a concert emphasizes a different set of cues than a training simulation or video game wherein a participant must intently localize sound sources in order to succeed. Thus, the research is motivated to meet a more demanding and general perceptual criterion.

As an illustrative example, the “holy grail” of optical virtual reality is to build a system that can completely fool the eyes and the mind – to produce visual stimuli that are indistinguishable from reality. Whatever form the system took, it would be capable of an extreme dynamic range to match the range of light intensity found in nature, the purest color fidelity, a complete panoramic view, a fluid frame rate, and an image resolution to exceed the capability of the human eye. In addition, all the interactivity cues available In real life would be faithfully reproduced by such a system: binocular vision (a different image received at each eye), motion parallax (the apparent motion of foreground objects relative to background objects in response to a change in viewpoint) – even the focal depth cue available from accommodation (the action of muscles deforming the lens in the eye). Technology already exists to create a display that exceeds the resolving power of the eye. And while a display that exceeds the dynamic range of the eye seems unlikely, it is the replication of *interactive* cues that presents the most profound challenge.

In the same way, the ultimate acoustic virtual reality system will exceed the dynamic range and frequency response of the human ear, and additionally, provide the relevant perceptual cues to sustain the illusion of a physical acoustic scene. It is encouraging that the technological state of affairs appears much better for the auditory system than for the visual system: sound systems that exceed the fidelity, dynamic range, and frequency response of the human ear are widespread (Rumsey 1995). The remaining problems for acoustic virtual reality, then, lie mostly in the reproduction of spatial and interactive cues.

1.3 Introduction to Wave Field Synthesis

Wave field synthesis is based on the Huygens-Fresnel principle: a wavefront can be thought of as a superposition of numerous smaller wavefronts. Thus a wavefront for a virtual source can be approximated by overlapping wavefronts originating from actual sources at other positions. In practice, loudspeaker arrays are arranged in a line, plane, or a circle around the listener. Signals are emitted along the loudspeaker array at carefully measured delays to produce the desired composite wave front shape. In this fashion, wave field synthesis is a method of spatial audio reproduction that is capable of generating complete virtual acoustic environments, valid for an extended area, so that the listener is free to move around in the sound field and gain psychoacoustic depth cues from that motion. Sounds can be made to appear anywhere within or outside the listening area. The method does not rely on psychoacoustic exploits, because it synthesizes a stable physical wave field.

Wave field synthesis is currently in use at a number of venues around the world. In particular, WFS technology is practical in theaters. The Fraunhofer spinoff IOSONO has installed wave field synthesis systems at Disney World, Orlando, Florida; Bavaria Filmstadt, Munich; Odysseum Science Adventure Park, Cologne; at ToddAO mixing studios in Burbank, California; and at their offices in Los Angeles and Erfurt. (www.iosono.com)

The original suggestion of wave field synthesis came from A. J. Berkhout, who made the original leap from his research in geophysics and seismology to acoustics. In a 1988 paper titled “A Holographic Approach to Acoustic Control,” Berkhout argued that “acoustic holography, featuring the spatial reconstruction of direct and reflected wave fields with desired wavefront properties at each moment of time," are the ultimate in sound control, since "holographically reconstructed sound fields cannot be distinguished from true sound fields." This holographic approach became wave field synthesis and was elaborated in Berkhout's further work (Berkhout 1993).

Stereo and surround-sound audio systems can be seen as means to approximate the intended sound field. In the current convention of stereo and surround sound, importance is given to frequency response and total harmonic distortion of the system. These are the standard measures of the system that exclude the room response and spatial performance of the setup. When these perceptually relevant spatial attributes are included in the measure of a system's performance, the advantages of wave field synthesis over stereo and multichannel surround reproduction become clear.

Wave field synthesis is the only audio reproduction system that synthesizes a realistic sound field for an extended area. It produces superior source localization and a larger listener "sweet-spot" than other systems. (Lopez 2005)

The remainder of the thesis proceeds as follows: Chapter 2 of this thesis provides a detailed background of WFS theory. Chapter 3 discusses problems and limitations of WFS, and Chapter 4 discusses important perceptual properties relevant to WFS. Chapter 5 provides an overview of existing WFS implementations. Chapter 6 and Chapter 7 describe the novel implementation of a loudspeaker array and WFS software environment. Chapter 8 formally introduces wave field synthesis in the vertical direction using multiple line arrays. Finally, Chapters 9 and 10 discuss the results of the listening tests and future directions for research.

Chapter 2. Theoretical Foundation of Wave Field Synthesis

When deriving the mathematical basis for wave field synthesis, the objective is to start with fundamental equations and arrive at the driving functions for a loudspeaker array made up of discrete transducers. The result we arrive at will reveal the operation we must perform on a signal at a given virtual source position (the "primary source" position) to arrive at the signal that should be emitted from each loudspeaker (each "secondary source") in a given array configuration. The traditional derivation of wave field synthesis starts by generalizing the wave field synthesis scenario to continuous secondary sources on an arbitrary 3-dimensional surface, and ends with the driving function for discrete secondary sources arranged in a line. Figure 2.1 illustrates the derivation as a series of reductions and approximations to be performed on the bounding surface that separates the virtual source from the listener.

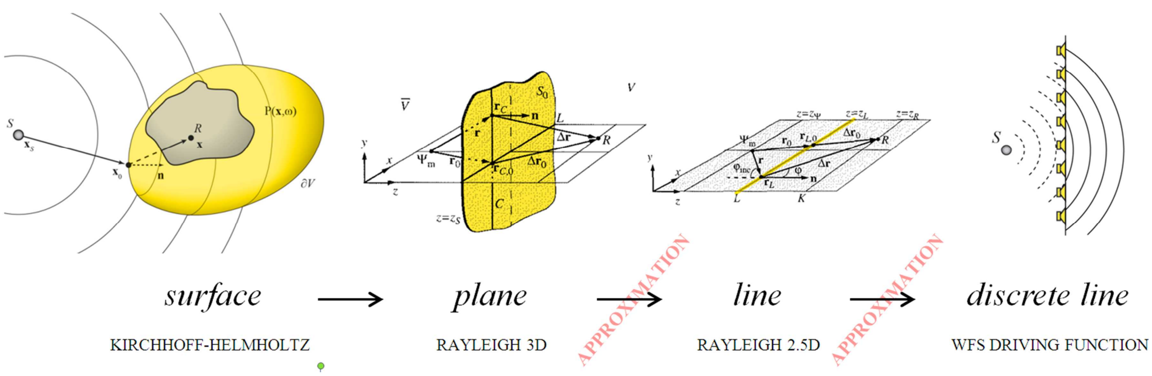


Figure 2.1: The derivation of WFS driving functions introduces approximations to sound field reproduction in the transformation from a plane to a line of secondary sources, and from continuous to discrete secondary sources.

The most general conception of wave field synthesis consists of a 3-dimensional wave medium containing a source, a receiver, and a hypothetical surface (such as a hollow sphere) that surrounds and encloses the receiver. This conception corresponds to the definition of the Kirchhoff-Helmholtz integral, explained as follows. The Kirchhoff-Helmholtz integral shows that we can calculate the sound pressure at any listening point within a source-free region enclosed by the surface ∂V if we know the pressure and particle velocity for every point on ∂V (a complete description of the state of the surface ∂V, for our intents). For clarity's sake, we mention that this takes place in a homogeneous wave medium, such as air. In applying the Kirchhoff-Helmholtz integral, our goal is to arrive at a general driving function for points along the surface ∂V so we can simulate sources outside the surface for listeners inside the surface. The Green's function G(·) is simply a place holder for an arbitrary field transfer function. (The field transfer function could vary depending on whether one wanted to simulate a dipole or monopole, or whether simulating in 2 or 3 dimensions, for example.) How should points on the surface ∂V behave, given that we are to produce a wave field in V matching the wave field produced by the source S?

The Kirchhoff-Helmholtz integral is defined by:

$$P(\mathbf{x},\omega)=\oint_{\partial V}\left(G(\mathbf{x}\,|\,\mathbf{x}_0,\omega)\frac{\partial}{\partial\mathbf{n}}P(\mathbf{x}_0,\omega)-P(\mathbf{x}_0,\omega)\frac{\partial}{\partial\mathbf{n}}G(\mathbf{x}\,|\,\mathbf{x}_0,\omega)\right)dS_0$$

Where P is the pressure at the receiver position $\mathbf{x}$, $\mathbf{x}_0$ is a position on the boundary surface, $\mathbf{n}$ is a normal vector for the surface at point $\mathbf{x}_0$ (pointing inwards), $\frac{\partial}{\partial\mathbf{n}}P(\mathbf{x}_0,\omega)$ is the directional gradient in the direction of $\mathbf{n}$. The surface integral adds the contribution from all points on the surface to arrive at a final sum pressure. This is illustrated in Figure 2.2.

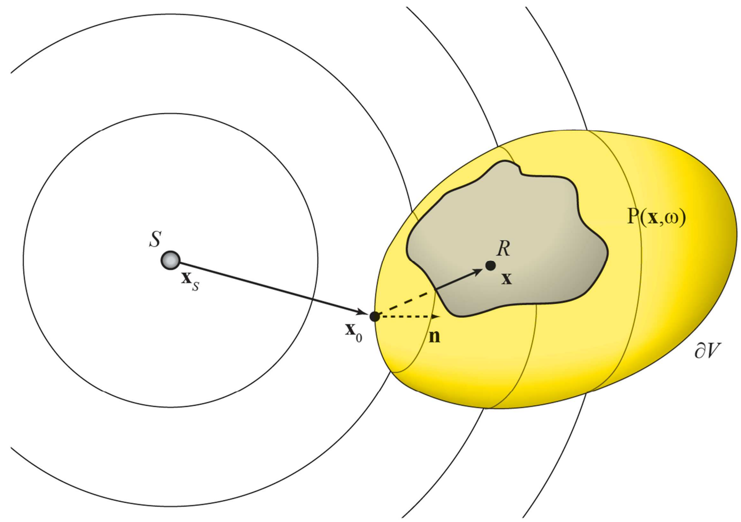


Figure 2.2: Conditions of the Kirchhoff-Helmholtz integral.

The Green's function $G(\mathbf{x}\,|\,\mathbf{x}_0,\omega)$ must be supplied. We will start with G_{3D}, which defines the spatial-temporal transfer function of a monopole source placed at $\mathbf{x}_0$. In other words, it tells us what a sensor placed at position $\mathbf{x}$ would receive from a source at $\mathbf{x}_0$. For a monopole source in a 3D medium,

$$G_{3D}(\mathbf{x}\,|\,\mathbf{x}_0,\omega)=\frac{1}{4\pi}\frac{e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|}$$

In simple terms, this describes how a monopole perturbs the medium that surrounds it. The observer at position $\mathbf{x}$ receives the source signal delayed by travel time $|\mathbf{x}-\mathbf{x}_0|/c$ and attenuated proportionally to its distance from the source, $|\mathbf{x}-\mathbf{x}_S|$. The directional gradient factor $\frac{\partial}{\partial \mathbf{n}}$ of G can be interpreted as the transfer function of a dipole placed at $\mathbf{x}_0$ with its axis aligned with normal vector $\mathbf{n}$. The solution given by the Kirchhoff-Helmholtz integral is null outside the volume V. The integral defines the behavior of both dipoles and monopoles on the boundary ∂V. The dipoles described by the directional gradient factor of G suppress the outward traveling wave from the surface ∂V . For the purposes of wave field synthesis inside the volume V, we do not need to suppress the outgoing wave and we can approximate the solution with a monopole surface. This results in the equation

$$P(\mathbf{x},\omega)=\oint_{\partial V}\mu(\mathbf{x}_0,\omega)G_0(\mathbf{x}\,|\,\mathbf{x}_0,\omega)dS_0\,.$$

The function $\mu(\mathbf{x}_0,\omega)$ is a source strength factor that must be determined. The appropriate source strength function can generally be described only by considering special geometries of the secondary source contour.

Since we have removed the positive dipole contribution from the inside of volume V, the pressure inside V no longer matches the wave field produced by the virtual source. We compensate by multiplying the remaining monopole component by 2. If the volume V is so large that the curvature along the boundary between source and receiver approaches zero – like a plane – the Kirchhoff-Helmholtz integral degenerates to the Rayleigh I integral (that is, the Raleigh I integral is a specific case of the Kirchhoff-Helmholtz integral). It states that a wave field due to sources on one side of a plane can be reconstructed on the other side by a continuous distribution of monopoles on the plane. The result of this expression is denoted G_{2D}:

$$G_{2D}(\mathbf{x}\,|\,\mathbf{x}_0,\omega)=\sqrt{\frac{2\pi|\mathbf{x}-\mathbf{x}_0|}{j\frac{\omega}{c}}}\,\frac{1}{4\pi}\,\frac{e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|}$$

We have eliminated the dipole contribution of the Kirchhoff-Helmholtz integral, leaving only monopole secondary sources, and this has two consequences. First, the wave field outside the secondary source boundary will no longer be zero. The wave field will be emitted symmetrically from either side of the plane. This means the secondary source arrangement must be convex, so that the spurious outward traveling waves do not re-enter the listening area and corrupt the wave field. Second, the reproduced wave field will no longer exactly match the virtual source field, due to undesired “reflections” from the side of the secondary source boundary opposite of the virtual source. These reflections can be controlled with a loudspeaker selection function. The loudspeaker selection function $a_s(\mathbf{x}_0)$ for source s and a loudspeaker at position $\mathbf{x}_0$ can be defined as

$$a_s(\mathbf{x}_0) = \begin{cases} 1 & \text{, if } (\mathbf{x}_0 - \mathbf{x}_s)^T \mathbf{n}(\mathbf{x}_0) > 0 \\ 0 & \text{, otherwise.} \end{cases}$$

The final step of derivation is to transform the system into discrete driving functions for discrete line arrays. The specific driving function varies by virtual source type (spherical, plane wave, focused source). The general form of the driving function is given by

$$D(\mathbf{x}_0, \omega) = S(\omega) \cdot H(\omega) \cdot w(\mathbf{x}_0) \cdot e^{-j\omega\tau_0(\mathbf{x}_0)}$$

$$d(\mathbf{x}_0, t) = s(t) * h(t) * w(\mathbf{x}_0) * \delta(t - \tau_0(\mathbf{x}_0))$$

in the frequency and time domain, respectively. The driving function result $d(\mathbf{x}_0, t)$ gives the signal that should be emitted by the loudspeaker at position $\mathbf{x}_0$ at time *t*. The signal emitted by a virtual source is given by $s(t)$, and $h(t)$ is a static pre-equalization filter that compensates for the discrete line array approximation. The amplitude factor $w(\mathbf{x}_0)$

incorporates the speaker selection function $a_s(\mathbf{x}_0)$ (discussed above), the virtual source distance, and the reference listener distance. It also takes into account the angle of incidence of the virtual source wavefront at array position $\mathbf{x}_0$ – wavefronts approaching the secondary source contour at a shallow angle should be attenuated. $\tau_0(\mathbf{x}_0)$ is a time delay factor, and generally corresponds to time of sound travel from source position $\mathbf{x}_S$ to array position $\mathbf{x}_0$ $\left(\frac{|\mathbf{x}_0 - \mathbf{x}_S|}{c}\right)$. The driving functions are discussed in detail in the next section.

2.1 Virtual Source Classification

In WFS theory, virtual sources are classified as plane waves, spherical sources, or focused sources. Spherical and focused sources are assumed to be point sources with omnidirectional directivity characteristics. The class is determined by the location of the virtual source. They must be distinguished because the loudspeaker driving function is different for each class. These classes are illustrated in Figure 2.3.

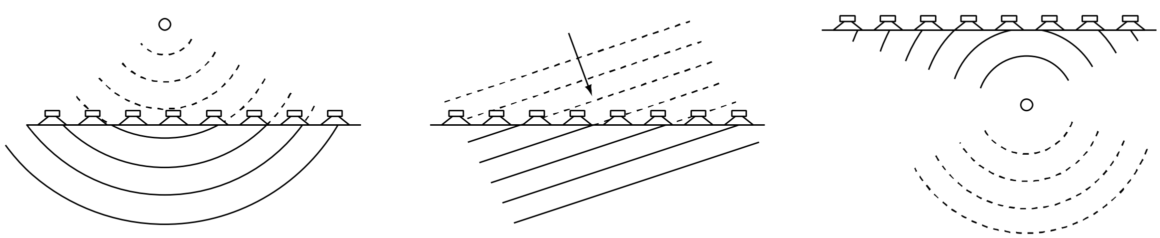


Figure 2.3: Spherical source, plane wave source, and focused source.

2.1.1 Plane Wave Sources

In the case of the plane wave source, the virtual source is positioned at an infinite distance beyond the array and the loudspeakers are driven at a linear delay based on the angle of incidence of the plane wave relative to the array. If the virtual source plane wave is propagating directly perpendicular to the array, there is no delay, and the loudspeakers are driven with the source signal uniformly. The plane wave driving function is given by

$$D_{\mathrm{pw,2.5D}}(\mathbf{x}_0,\omega) = -2a_{\mathrm{pw}}(\mathbf{x}_0)\sqrt{2\pi|\mathbf{x}_{ref}-\mathbf{x}_0|}\,\mathbf{n}_{\mathrm{pw}}^T\mathbf{n}(\mathbf{x}_0)\sqrt{j\tfrac{\omega}{c}}\,\hat{S}_{\mathrm{pw}}(\omega)e^{-j\frac{\omega}{c}\mathbf{n}_{\mathrm{pw}}^T\mathbf{x}_0}$$

$$d_{\mathrm{pw,2.5D}}(\mathbf{x}_0,t) = w_{\mathrm{pw}}\delta(t-\tfrac{\mathbf{n}_{\mathrm{pw}}^T\mathbf{x}_0}{c}) * (f_{\mathrm{pw}}(t) * \hat{s}_{\mathrm{pw}}(t))$$

in the frequency domain and time domain, respectively (Spors 2009). Note the weighting factor contains no reference to distance; the plane wave virtual source is identified only by direction. Weighting terms are combined; $w_{\mathrm{pw}} = -2a_{\mathrm{pw}}(\mathbf{x}_0)\sqrt{2\pi|\mathbf{x}_{ref}-\mathbf{x}_0|}\,\mathbf{n}_{\mathrm{pw}}^T\mathbf{n}(\mathbf{x}_0)$ and $f_{\mathrm{pw}}(t)$ is the inverse Fourier transform of $\sqrt{j\frac{\omega}{c}}$, which amounts to a 1/8 period delay, +3 dB per octave high-pass filter. This mysterious term is independent of source type, and must be understood as a consequence of using a line array to generate a wave field that can only really be generated by a plane array. Specifically, it is a result of the stationary phase approximation – approximating the wave field of an infinite line source, which has a frequency-varying far-field response, with a point source.

Simulating a plane wave with linear WFS geometries is equivalent to beam steering with a line array:

$$\vartheta_{NH} = \sin^{-1}\left(\frac{c}{c_s}\right) = \sin^{-1}\left(\frac{\lambda}{\lambda_s}\right)$$

where ϑ_{NH} is the beam direction (corresponding to the direction of plane wave propagation), λ indicates the wavelength of the signal in the direction of propagation, λ_s indicates the wavelength of the signal along the secondary source array, and c_s is the "sweeping speed" that describes how fast the signal is shifted across the line array.

2.1.2 Spherical Sources

Virtual sources positioned outside the boundary of the loudspeaker array are called spherical sources. This is the most general term for referring to the source types, as plane wave sources and focused sources can be thought of as species of spherical waves.

It can be shown that the driving function for spherical sources is given by

$$D_{\mathrm{sw,2.5D}}(\mathbf{x}_0,\omega) = -2a_{\mathrm{sw}}(\mathbf{x}_0)\frac{(\mathbf{x}_0-\mathbf{x}_S)^T\mathbf{n}(\mathbf{x}_0)}{|\mathbf{x}_0-\mathbf{x}_S|}\sqrt{2\pi|\mathbf{x}_{ref}-\mathbf{x}_0|}\left(\frac{1}{\sqrt{j\frac{\omega}{c}}|\mathbf{x}_0-\mathbf{x}_S|}+\sqrt{j\frac{\omega}{c}}\right)\hat{S}_{\mathrm{sw}}(\omega)\frac{e^{-j\frac{\omega}{c}|\mathbf{x}_0-\mathbf{x}_S|}}{|\mathbf{x}_0-\mathbf{x}_S|}$$

$$d_{\mathrm{sw,2.5D}}(\mathbf{x}_0,t) = w_{\mathrm{sw}}\delta(t-\tfrac{|\mathbf{x}_0-\mathbf{x}_S|}{c}) * (f_{\mathrm{sw}}(t) * \hat{s}_{\mathrm{sw}}(t))$$

in the frequency and time domain, respectively (Spors 2009).

2.1.3 Focused Sources

A focused source is a virtual source with a location between the secondary source array and the listener, or for circular arrays, a virtual source positioned inside the array. The synthesized wavefront of a focused source begins by converging on the virtual source location (the acoustic wave is "focused" on a point), and subsequently diverges from that point. The focused source is, therefore, only valid for listener positions enclosed by a region marked by lines extending from the edges of the array through the virtual source position. (Spors 2010) It has been proven that synthesis is not possible for virtual sources that cannot be seen through the “acoustic window” of the loudspeaker array. (Verheijen 1998). This includes focused sources.

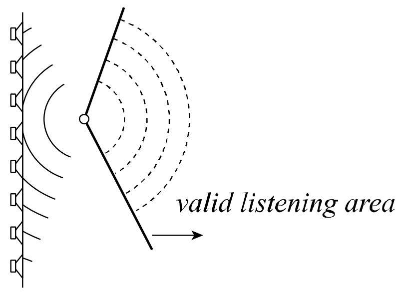


Figure 2.4: A focused source reproduces a sound field for a virtual source located in front of the line array.

The driving function for focused sources is given by

$$D_{2.5\mathrm{D}}(\mathbf{x}_0,\omega) = -g_0 \hat{S}_\mathrm{s}(\omega)\sqrt{\frac{\omega}{2\pi jc}}\,\frac{y_0 - y_s}{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|^{\frac{3}{2}}}\,e^{j\frac{\omega}{c}|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|}$$

$$d_{2.5\mathrm{D}}(\mathbf{x}_0,t) = s(t) * h(t) * \frac{y_0 - y_s}{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|^{\frac{3}{2}}}\,\delta\left(t + \frac{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|}{c}\right)$$

in the frequency and time domain, respectively (Spors 2009).

Chapter 3. Problems in Wave Field Synthesis

Fundamental WFS theory assumes a continuous distribution of secondary sources. In practice, this cannot be realized. The secondary sources are implemented by a limited number of discrete loudspeakers. This fact gives rise to these practical limitations of wave field synthesis.

3.1 Restriction to Horizontal Plane

Wave field synthesis demands many independently controlled loudspeakers. A line array may require hundreds of loudspeakers, and a plane array implementation might require thousands. For this reason, the majority of wave field synthesis research has concerned linear arrays. The restriction to linear arrays means that virtual sources can only move along the horizontal plane defined by the listener's ears and the loudspeaker array. Again, this is not a fundamental limitation of wave field synthesis, but a practical barrier in most implementations. However, there may be economic solutions to this problem that have not been explored.

In Chapter 8, I propose a solution to the horizontal plane problem. Multiple line arrays stacked at wide vertical intervals could be used to extend the virtual source space in the vertical direction without the drawbacks of implementing a complete plane array.

3.2 Amplitude Error

Most conventional wave field synthesis arrays are simple line arrays of equally spaced loudspeakers. In this common configuration, the amplitude falloff does not match the inverse law for natural sources. This is because the wavefront produced by the array is actually cylindrical instead of spherical, and so it follows the acoustic power function of a local line source instead of a planar (or far field plane wave) source. (Verheijen 1997, Spors 2009)

Figure 3.1 shows the desired amplitude level A_d and synthesized field amplitude level A_p plotted against the distance of the receiver from the array. The desired and synthesized amplitude levels match at the distance of the reference contour Δr_c (also labeled $\mathbf{x}_{\mathrm{ref}}$ by some authors, as in the discrete driving functions referenced in Chapter 3 of this thesis). This scalar value is used in the driving function to yield the correct amplitude level of the virtual source for a single listener distance.

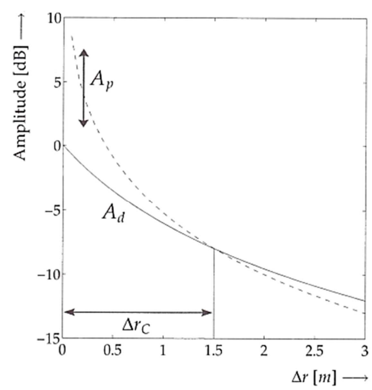


Figure 3.1: Amplitude error for loudspeaker line arrays. Desired amplitude A_d versus synthesized amplitude A_p. (Sonke, deVries 1998)

3.3 Truncation Effects/Diffraction

The WFS theory begins with the assumption that the secondary sources extend infinitely. In practice, the extent of the loudspeaker line array is limited. The truncation effect can be understood if the loudspeaker array is thought of as an aperture the virtual source wavefront must pass through on its way to the listener. As in conventional wave theory, the effect of this slit diffraction is akin to additional wave sources placed at both ends of the array. (Start 1997; DeVries, Start 1994)

3.3.1 Solution: Tapering

Tapering the power of the secondary sources to zero at the edges of the array solves the truncation problem. The effects of diffraction are reduced at the cost of a smaller effective array size. A one-sided squared cosine window is typically applied to achieve a smooth roll-off. (Spors 2009)

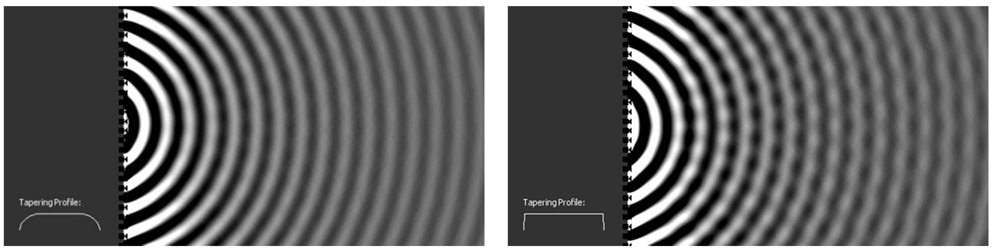


Figure 3.2: Diffraction due to loudspeaker array truncation. The line array depicted on the left uses a $\cos^2$ tapering profile to reduce diffraction. The array on the right uses no tapering. Diffraction effects are visible as faint superimposed waves that appear to originate at the off-screen edges of the array.

3.3.2 Solution: Surround Arrays

Loudspeaker array configurations that fully surround the listener (square or circular arrangements) also mitigate the boundary effects, since these array configurations do not have sharp discontinuities.

3.4 Spatial Aliasing

Spatial aliasing is best described as the fragmentation of the composite wavefront as a result of discretization of secondary sources. The summed wave is made up of discrete component waves and becomes incoherent for wavelengths shorter than the loudspeaker spacing. In other words, the individual component waves become audible. In the ideal condition of a continuous loudspeaker array, the composite wavefront would match the virtual source wavefront exactly. This is not realized in loudspeaker arrays. Perceptually, spatial aliasing results in a position-dependent comb filtering that causes a noticeable high frequency modulation or "flutter" as the listener moves around the listening area (Lopez 2005). Some authors report that the spatial aliasing artifacts are not very noticeable to human listeners (Spors, Teutsch, Rabenstein 2002; Oellers 2011). The extent of spatial aliasing is dependent on the position of the virtual source, the position of the listener, and the frequency of the source signal. Spors (2006) provides the cutoff frequency to avoid spatial aliasing for an array spacing Δx and a plane wave of angle α_{pw}:

$$f_{alias} = \frac{c}{\Delta x(1 + |\cos \alpha_{\mathrm{pw}}|)}$$

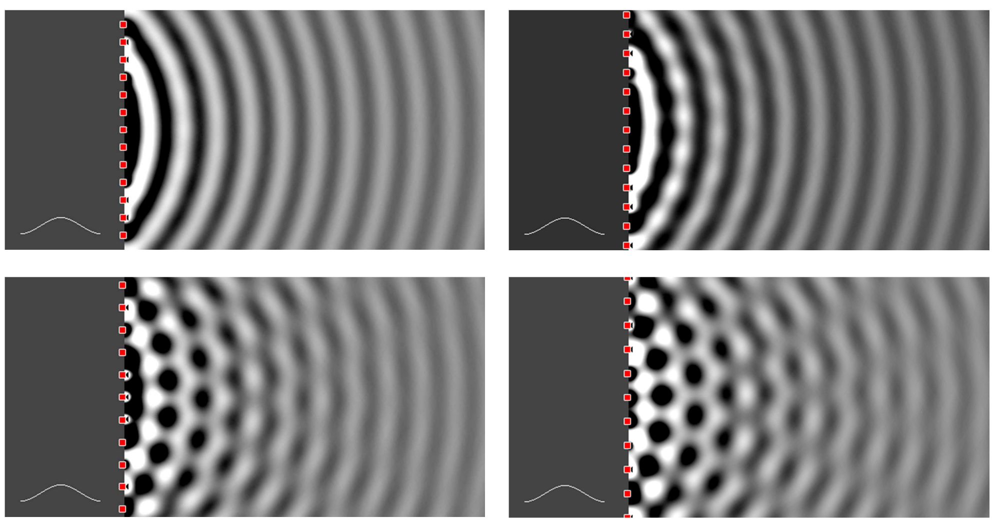


Figure 3.3: Spatial aliasing of a spherical source with progressively wider loudspeaker spacing.

Many authors suggest that spatial aliasing in content above 1.5 kHz does not significantly degrade direction cues (Start 1997; Boone, Verheijen 1995). Start compared the minimum audible angle (MAA) of real sources and virtual sources under various WFS array configurations.

The MAA for an array spacing of 11 cm was 0.8° for broadband and 1.1° for low-passed (<1.5 kHz) signals. The MAA increased to 1.6° for an array spacing of 22 cm, corresponding to a spatial aliasing frequency of 750 kHz. In Start's own words, "the localization accuracy of low-frequency noise stimuli is almost identical for synthesized and real sound fields. As expected, localization performance is seriously degraded for high-frequency noise stimuli."

3.4.1 Solution: Optimized Phantom Source Imaging

One approach to solving the spatial aliasing problem is to separate the signal into high- and low-frequency components. The low frequency component is sent to the loudspeaker array and processed using wave field synthesis as usual, while the high-frequency component is sent to a separate set of tweeters and processed with conventional amplitude panning. This is the Optimized Phantom Source Imaging (OPSI) method, proposed by Wittek in 2002.

The high frequency speaker array can be spaced more widely than the WFS array, because it is not used to generate a coherent wavefront in the higher frequencies, and therefore uses fewer source elements. OPSI generates two different auditory cues for each virtual source: one consisting of the physical reproduction of the wave field in the low frequencies, and the other being the perceptual phenomenon of phantom source position appearing between the driving elements, as in conventional stereo imaging. The general problem in implementing OPSI then is to reduce the difference between these two cues; to make the source direction presented by the high frequency array match the WFS array as closely as possible.

The perceived phantom source location depends on the listener's location relative to the driving loudspeakers, a familiar problem in stereo audio reproduction. This introduces a localization error, defined by Wittek as the directional difference between the low-frequency virtual source and the high-frequency phantom source. Because of the dependence of phantom source localization on the listener's location relative to the loudspeakers, the localization error is dependent on listener location. Wittek found that a phantom source direction that deviates less than 7.5° from the virtual source direction does not lead to an audible localization shift of the combined source image in his original study, but this tolerance is dependent on signal content.

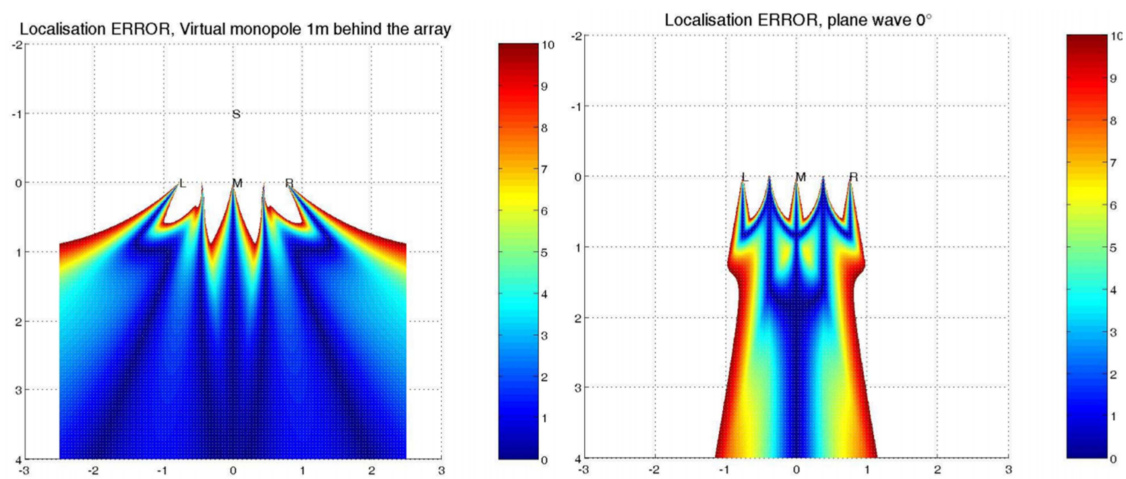


Figure 3.4: Illustration of localization error using the OPSI method for a linear array at y=0 and listening area where y>0. X and Y axis show coordinates in meters. Spherical source indicated by S on left; plane wave shown on right. High frequency content is emitted from loudspeakers at locations L, M, and R. (Wittek 2002)

3.4.2 Solution: Sub-Band Approach

The sub-band approach proposed by Lopez et al. (2005) is similar to the OPSI method. The high frequency phantom source component of the signal is not sent to a separate loudspeaker array but is instead recombined with the low frequency WFS result and sent to the main loudspeaker array. This has the advantage of eliminating the interdependence of crossover frequency and spatial aliasing frequency. Most of the discussion about OPSI also applies to the sub-band approach, since the method differs by output implementation only. WFS Designer implements the sub-band approach, elaborated in section 7.5.2.

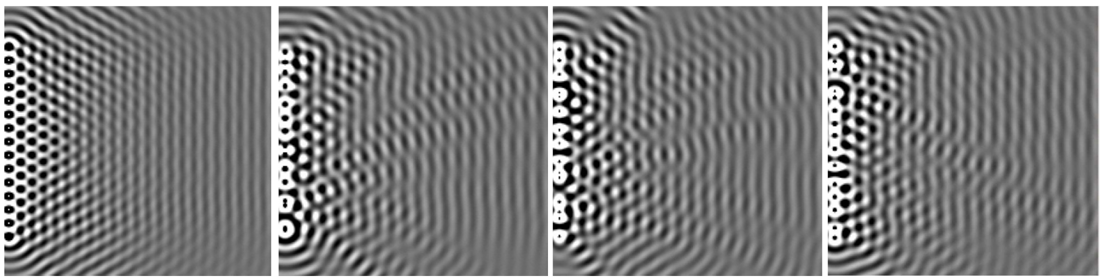
3.4.3 Solution: Distributed Mode Loudspeaker/Multi-Actuator Panels

The distributed mode loudspeaker (DML) is a flat-panel loudspeaker technology that produces sound by inducing uniformly distributed vibration modes through a prescribed panel shape. (NXT http://www.nxtsound.com/420.html) Boone and others have found the DML to be a good candidate for WFS reproduction because it does not have any unexpected distortion and exhibits a wide directivity (Boone 2004). A multi-actuator panel (MAP) consists of a flat acoustic radiation panel attached to a number of exciters. It is like the distributed mode loudspeaker, except that a DML uses a single voice coil and is intended for a single channel of reproduction. The multi-actuator panel has been explored for application to wave field synthesis (Boone 2004, Corteel 2006). These authors have found MAP suitable for application to WFS because their diffuse spatial characteristic helps to smear the effect of spatial aliasing.

3.4.4 Solution: High-Frequency Randomization

Start attempted to break up the irregularities produced by spatial aliasing by inserting random time offsets (less than a period of the frequencies in question) in high-frequency content delivered to each loudspeaker (Start 2007) . The process alters the pattern of spatial aliasing but does not reduce its overall strength. Corteel's experiments indicate that randomization of secondary source positions, equivalent to randomized time delays, does not improve accurate source localization (Corteel 2006). Figure 3.5 illustrates the effect of randomization. The first image shows a plane wave under aliasing conditions.

The most notable effect is the undesired plane waves that appear to propagate at +/-60 degrees. The second, third, and fourth images show the same plane wave with randomized source positions. The randomization seems to "break up" the spatial distribution of aliasing, but the perceptual effect is unclear.



(a) *(b)* *(c)* *(d)*

Figure 3.5: Aliasing of a plane wave with uniform (a) and randomized (b, c, d) loudspeaker positions.

3.4.5 Solution: Spatial Bandwidth Reduction

Recall from section 3.4 that f_{alias} is dependent on the the virtual source wave front's angle of incidence upon the secondary source array; f_{alias} decreases with a more oblique wave front. Spatial bandwidth reduction is an attempt to exploit this angular dependence by decreasing the angle of incidence of high-frequency virtual source content (DeVries 1994, Start 1997). In the case of a line array configuration, this smears the virtual source image in a frequency-dependent manner so that the high-frequency content is steered toward the center of the array.

3.4.6 Solution: The Spatial Antialiasing Loudspeaker

Authors Spors (2010) and Start (1997) have written on the possibility of an "antialiasing" loudspeaker. The antialiasing loudspeaker would function by limiting the bandwidth of the speaker in a spatially dependent way, focusing high-frequency content in a forward direction.

The antialiasing loudspeaker remains a theoretical proposition at this point in time because no transducer has been developed with the required properties.

Spatial bandwidth reduction and the theoretical spatial antialiasing loudspeaker are both ways of selectively omitting source signal content, which inevitably leads to an incomplete wave field and errors in perception.

3.5 Room Acoustics

Recall that wave field synthesis assumes an anechoic listening environment. In the context of problems for wave field synthesis, room acoustics refers to the unintended and unwanted reflection and reverberation of the listening environment that corrupts the intended sound field. Many authors consider this the most perceptually detrimental artifact (Spors 2005; Wittek 2007; Oellers 2010; Corteel, Nicol 2003). The interference comes not from the room reflection of the virtual source, but the room reflections of the secondary sources.

3.5.1 Solution: Room Effect Compensation

Several authors have investigated techniques for mitigating this effect. Fortunately, the loudspeaker array itself is a useful tool for active sound field compensation, including global noise control applications (Kuntz 1999). Two basic approaches toward room compensation in WFS are active monitoring compensation, requiring numerous microphones for multi-point sensing, and model-based compensation based on an accurate mathematical representation of room acoustics. An analytical approach seems to be thus far unexplored.

Active compensation experiments by Spors, et al. (2003) make use of circular loudspeaker arrays paired with circular microphone arrays. They have shown effective results, but the accurate compensation by destructive interference is limited by the spatial aliasing frequency of the array. Spors, Renk, and Rabenstein elaborated on their efforts in 2005 and clarified the limiting effects of active room compensation were the spatial aliasing frequency, the inability to correct for vertical reflections with a linear array configuration, and the variance in listeners’ vertical position.

Chapter 4. Perceptual Properties of Wave Field Synthesis

Since WFS generates a physical sound field, the perception of the virtual auditory scene presented should occur just as natural hearing of natural sources as long as the wave field is sufficiently accurate. The artifacts of WFS interfere with the perceptual cues, but the cues left intact will still be perceived through known conventional mechanisms for natural wave fields.

4.1 Challenges for Assessment

In evaluating the performance of a system designed to reproduce a physical sound field – not just a source in a particular direction – there are many perceptual attributes to define and organize. Some of these attributes are not consistently defined by the literature. This author follows the lead of Wittek (2003) in proposing the following system of attributes, which may be investigated only after a thorough description.

First, performance attributes can be classified according to three top level categories:

- Attributes of location and dimension of a virtual source
- Attributes of the virtual source content, the signal itself
- Attributes of the virtual environment

The reasoning behind this scheme is that these three groups reflect the perceptual stratification of information derived from the sound field, and they represent perceptually independent dimensions. The perceptual separation of source information from environmental information (i.e., "Where am I?") is investigated by Rumsey (2002) and the perceptual separation of source object information (i.e., “Where is the speaker located?”) and signal content (i.e., “What is the speaker saying?”) is preceded by Theile’s association model (Thiele 1980).

Localization (Source Attributes)	Content (Signal Attributes)	Environment (Global Attributes)
Direction Distance Width Focus Locatedness Stability Robustness Externalization	Loudness Sound color (timbre) Familiarity Plausibility	Depth Room dimension Envelopment Presence Naturalness Room timbre Reverberance

Table 4.1: Categorization of Perceptual Sound Source Attributes (Wittek 2003)

A survey of the definitions of these terms in the literature, as summarized by Wittek, follows:

Property	Definition
Localisation	General mapping law between the location of an auditory event and a certain attribute of the sound source (definition according to Blauert, 1997). Mechanism/Process that maps the location of an externalised auditory event to certain characteristics of one or more sound events (definition according to Theile, 1980).
Direction	The direction in which the source is perceived.
Distance	Perceived range between listener and reproduced source (definition according to Rumsey’s (2002) ‘individual source distance’).
Depth	Sense of perspective in the reproduced scene as a whole (definition according to Rumsey’s (2002) ‘environment depth’).
Stability	The degree to which the perceived location of a source changes with time.
Robustness	The degree to which the perceived location of a source changes with movement of the listener.
Accuracy	The degree to which the intended and the actually perceived source agree with each other. This ‘agreement’, unless defined differently, involves all attributes of the source. Often, the term accuracy is used only for the ‘directional

	accuracy', which means the agreement concerning the source direction. The relevant measure for this attribute is the 'directional error' of a source/system.
Resolution	The achievable precision of the synthesized sound field in terms of direction and/or distance.
Individual source width ISW, Apparent source width ASW	Perceived width of the source (definition according to Rumsey 2002).
(Image) focus	The degree to which the energy of the perceived source is focused in one point.
Definition of the image	Similar to image focus
Diffuseness	Inverse of image focus
Blur	Inverse of image focus
Locatedness	Spatial distinction of a source (definition according to Blauert 1997). The degree to which an auditory event can be said to be clearly perceived in a particular location.
Certainty of source localisation	Similar to 'locatedness', used by Lund (2000) .
Localisation quality, Localisation performance	These terms describe a mix of attributes. They describe the overall performance of localisation. They should be defined individually, because they can have ambiguous meanings ('quality' of the directional accuracy, sound color, focus, locatedness or an 'average' quality?).
Externalisation	The degree to which the auditory event is outside the head.
Spaciousness	Often used in the same meaning as 'apparent source width' ASW, but also used to describe the perceived size of the environment.
Presence	Sense of being inside an (enclosed) space or scene (definition according to Rumsey 2002). Often also used as an attribute of sound color.

Table 4.2: Definitions of Perceptual Sound Source Attributes (Wittek 2003)

4.2 Wave Field Synthesis Compared to Other Spatialization Methods

	Stereo/Multichannel	**HRTF Binaural**	**WFS**	**Ambisonics**
Transducers	Few loudspeakers	Headphones	Loudspeaker array	Arbitrary
Type of Reproduced Source	Phantom source	Virtual source	Virtual source	Virtual source
Perceptual reproduction principle	Psychoacoustic integration of correlated sources	Reproduction of the signal at the ear drum	Physical synthesis of wave field produced by discrete virtual	Reproduce the directional sound pressure field captured at one location

			sources	
Freedom of listener movement within virtual sound field	Small sweet spot; phantom source shifts toward nearest loudspeaker	Zero	Large degree of freedom. Virtual source is stable as listener moves. Virtual source level changes closer to real source under listener motion	Limited sweet spot corresponds to the microphone recording position
Stable sources	Loudspeakers themselves	Zero	Unlimited	Loudspeakers themselves
Source direction	Phantom source can be presented at any location between speakers	Perfect spatial reproduction with constraints	Virtual source can be presented at any direction within the extent of the array	Perfect spatial reproduction with constraints
Source distance	Phantom source is positioned along the line between loudspeakers	Perfect spatial reproduction with constraints	Distance is reproduced through wavefront curvature	Perfect spatial reproduction with constraints

Table 4.3: Comparison of Spatialization Methods (Wittek 2007)

Stereo and multichannel reproduction relies on the psychoacoustic merging of strongly correlated source signals coming from two loudspeakers into a phantom source located between the two loudspeakers.

HRTF binauralization and ambisonics rely on capturing the final psychoacoustic cues at a single listening location, and reproducing those cues to the listener. As such, the cues provided in the sound output are embedded with auditory cues for a single listener location. Synthesizing these cues for arbitrary audio scenes is another matter.

4.3 Phantom Source vs. Virtual Source

References to phantom sources and virtual sources are made throughout this thesis. The difference between the two should be clarified. The phantom source and virtual source differ in their psychoacoustic operation. Wittek describes a phantom source as an auditory event that is not based on natural binaural cues, but instead is assumed to result from “coactions of specific mechanisms of the auditory system which have developed during natural listening” (Wittek 2008). For the purposes of this thesis, a virtual source is defined as a source produced by physical reconstruction of the wave field (wave field synthesis, in this case) and a phantom source is a source produced by amplitude panning.

4.4 The Perception of Distance

Auditory distance is one of the most challenging perceptual phenomena to reproduce, which makes it a good measure of the performance of virtual audio displays (Wittek 2008). Perception of distance is especially important to making advances in wave field synthesis. It is important to understand what cues lead to the perception of distance in order to successfully present the distance of a virtual source.

4.4.1 Loudness

Loudness is widely accepted as the most important distance cue. In the anechoic chamber, it is often the only distance cue. For sources more than 1 meter away from the listener, the interaural level and time differences may be sufficient for conveying source direction, but are not useful for conveying source distance (Zahorik 2002). This is a side effect of simple trigonometric properties. It should be noted that this cue depends on prior knowledge of the natural intensity of source. If the source is unfamiliar, or if volume levels are manipulated, the listener can easily be misled.

4.4.2 Interaural Difference

For sources less than 1 meter away, interaural differences provide meaningful information about source distance. That is, a source at a given direction 0.5 meters away will produce different interaural cues than a source at the same direction at 1 meter away, because the distance between the ears becomes significant. This cue breaks down on the listener's median plane as the interaural differences diminish.

4.4.3 Direct-to-Reverberant Energy Ratio

Once we get out of the anechoic chamber and into a natural reflective acoustic environment, we gain more information about source distance from the reverberant environment. This cue invokes an important dependence on the knowledge of the room (Mershon 1979); this is an aspect that warrants more discussion. For now, we suffice to say that in general, the ratio of direct to reverberant energy decreases as the source moves away from the listener (Nielsen 1993).

4.4.4 Initial Time Delay Gap

Initial time delay gap (ITDG) is defined as the delay in arrival time between the direct source sound wave and the first reflection wave at the listener's position. The behavior of ITDG as a function of source distance is shown in Figure 4.1.

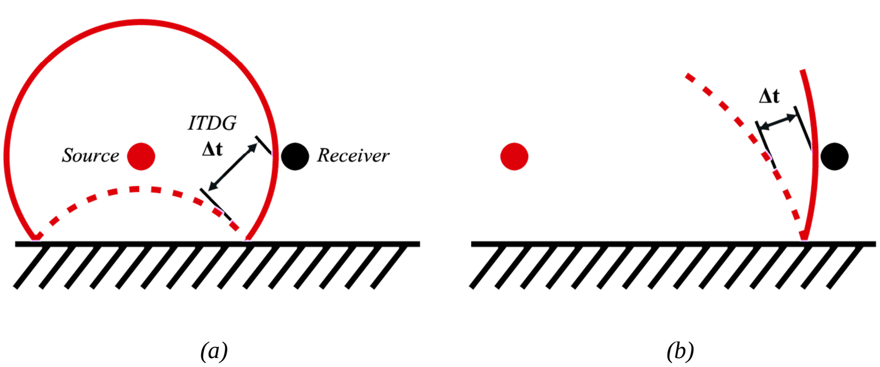


*Figure 4.1: Initial time delay gap (**Δt**) is large for a nearby source (a) and small for a far-away source (b).*

The delay between direct and reflected sound is greater when the source is nearer to the observer. Again, this cue means that perception of distance depends on the room/listening environment.

4.4.5 Frequency Spectrum

Air damping is stronger at high frequencies. The spectrum of a sound may provide information about its distance; but, as with loudness, this feature requires prior knowledge of what the source nominally sounds like. (Nielsen 1993)

4.4.6 Reflection Pattern

Information is conveyed in the timing, level, and direction of each early reflection. It could be said that every cue mentioned so far can be mimicked by a stereo system. This is not so for the reflection pattern; the reflection pattern contains directional qualities and it is a considerably powerful cue that cannot easily be duplicated without reproduction of an entire sound field. (Pellegrini 2001)

4.4.7 Motion Parallax

Motion parallax is the corresponding change of perspective with movements. In auditory perception, many of the aforementioned cues change with listener motion. Therefore it can be considered a compound cue, incorporating the way other cues change as a function of listener position. For example, if one moves 2 meters toward a source, how much louder does it get? This provides a distance cue. (Wittek 2008)

4.5 Gestalt/Associative Model

We can see that the perception of source distance is hardly an intrinsic feature of the auditory system: it relies on a spatial framework, and distance emerges from the scene.

The brain is constantly integrating sensory information and rejecting conflicting information in order to build a sensible scene model. It may be that the real perception of distance arises only from the complete integration of overall scene analysis, listening space, relationship to prior auditory percepts, perception of distinct reflections, and relationship to visual percepts (Thiele 1980). This is illustrated in Figure 4.2.

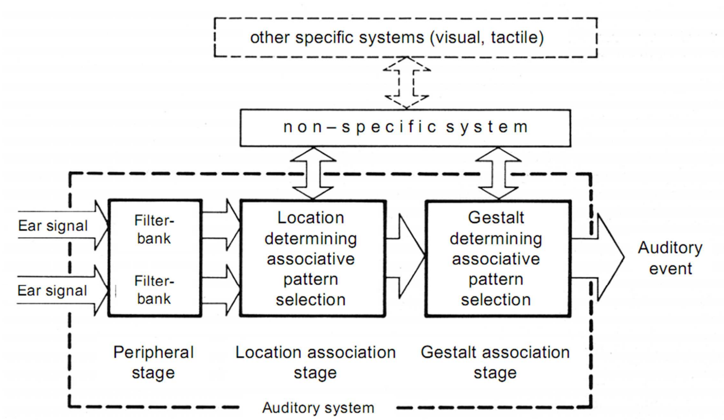


Figure 4.2: Associative Model (Theile 1980)

A different kind of auto-association makes experimental testing more difficult: the “ventriloquism effect” is the tendency for listeners to associate sound with visible anchors (Blauert 1997). This causes responses to be biased toward visible markers or loudspeakers in the test area. Careful design is required to remove the influence of visual perception on auditory perception. Several authors use an acoustically transparent canvas or curtain to obscure the test area from view. (Wittek 2007, Usher 2004)

4.6 WFS Distance and Depth

Wave field synthesis is particularly well-suited to generating motion parallax and spatially accurate reverberant field cues. How does perception of distance and depth improve with the addition of these cues? Wittek anticipated this result in his 2007 thesis:

“There is another possible property of WFS that might give rise to an enhanced distance perception: WFS can very accurately simulate position and level of the early reflections…The enhanced possibility for the auditory system to distinguish between distinct reflections [in WFS as opposed to stereo] may, however, give rise to a better spatial perception.” According to de Bruijn (2001), “Although it is already well known that with WFS the source location is extremely stable when observed from different listener locations, depth or distance perception can only be obtained in combination with reflections and reverberation.”

Wittek posed another question regarding involuntary parallax motion: “It is of vital importance whether spontaneous, unconscious head movements would then be sufficient for depth perception or if conscious movements are required which give rise to a significant change of the perceived scene perspective in the sound scene.” However, his experiments focused on isolating the effects of wavefront curvature on perception of distance and locatedness, and excluded parallax motion cues. This leaves an interesting opportunity for investigation.

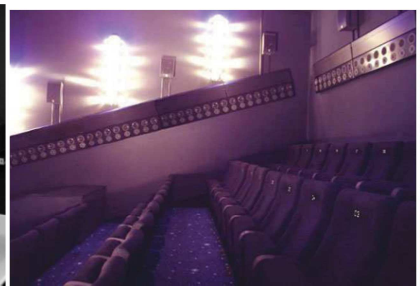
Chapter 5. Existing Implementations

There are several reference designs available in the literature demonstrating how to implement the hardware and software WFS production system. This chapter provides a brief overview of some notable systems.

5.1 Hardware

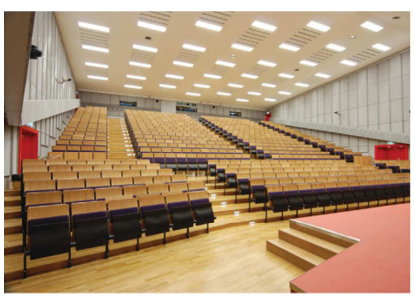
Almost all experimental setups for WFS employ line arrays with loudspeaker spacing between 10 and 25 cm, with the notable exception of "Loudspeaker Walls" by Ono (1997). The largest WFS installations exist in Germany, where most WFS research has been carried out. Somewhat more modest arrays have been constructed for experimental research purposes at other universities around the world.

ISONAR / 24 channel (Salvador 2010)

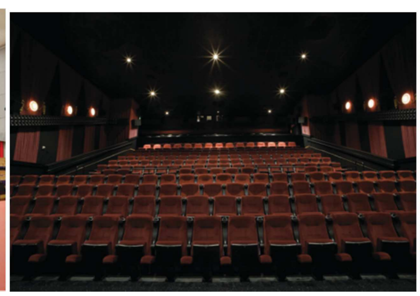


Fraunhofer IDMT / 192 channel

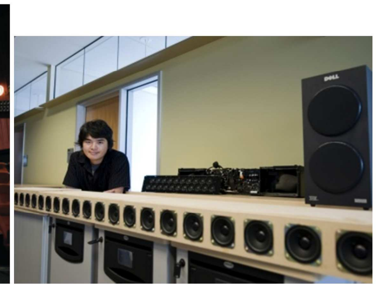
T-Labs / 56 channel



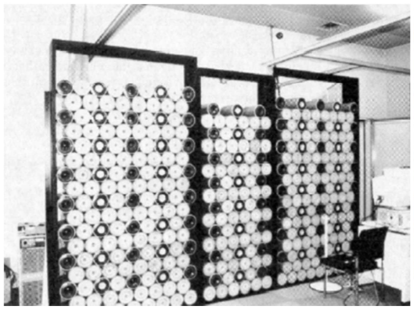
TU Berlin / 832 channel



IOSONO / 378 channel



UCSD / 24 channel (Yamada 2008)



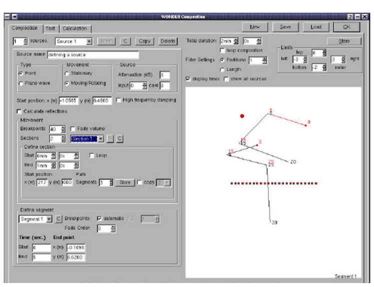
Ono 1997 (not strictly WFS)

University of Surrey / 18 (56) channel (Wittek 2007)

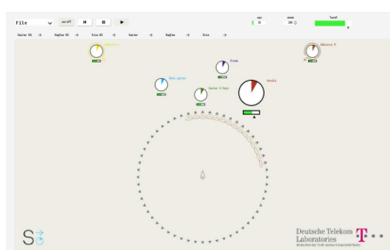
Figure 5.1: Example WFS installations and experimental configurations.

5.2 Software

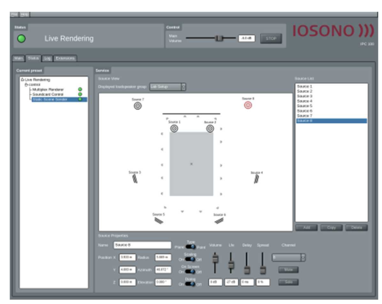
Several groups have developed model-based rendering applications for WFS, pictured in Figure 5.2. Some authors have developed WFS engines for audio processing environments such as Supercollider, Max/MSP, and Pure Data (Salvador, 2010). Three well-known WFS applications are WONDER, developed at TU Berlin (Baalman, Plewe 2004); The SoundScape Renderer, developed by Deutsche Telekom and TU Berlin (Geier et al., 2007); and IOSONO's commercial software Spatial Audio Workstation (IOSONO 2011). These systems are described below.



WONDER (Balmaan 2004)



SoundScape Renderer (Geier, et al. 2007)



Spatial Audio Workstation (IOSONO)

Figure 5.2: Example WFS software applications.

5.2.1 *WONDER*

WONDER consists of a 2-dimensional composition canvas, a grid specification tool and a play function. The program runs under Linux and is controllable via OpenSoundControl (OSC), making it possible to control the system from most commonly used programs for composition or live performance. The program allows users to define virtual source motion paths and simulate room reflections. WONDER calculates virtual source filters on a user-defined grid before playback is begun. Room reflections are incorporated in the resulting filters. The user is then free to move virtual sources within the calculated grid region during performance. The discretization of virtual source space for pre-calculated filters is further described in section 5.2.4. (Baalman, Plewe 2004)

5.2.2 *SoundScape Renderer*

The SoundScape Renderer (SSR) is capable of rendering 2-dimensional virtual acoustic scenes using wave field synthesis, binaural rendering, ambisonics, and vector-based amplitude panning. SSR provides a 2-dimensional sound composition area. Virtual sources are defined as plane wave or point source types. The loudspeaker array and virtual source scene are specified by an Audio Scene Description Format (ASDF) XML file, conceived by the authors. SSR uses a modular network-based architecture so that the audio engine and user interface can cooperate in real-time on separate machines. SSR can be compiled for Linux or Apple OS X systems. (Geier et al., 2007)

5.2.3 *Spatial Audio Workstation*

IOSONO's Spatial Audio Workstation is a commercial software package for WFS that integrates with Steinberg's Nuendo digital audio workstation. All audio tracks and events appear on the 2-dimensional stage view of the Spatial Audio Workstation as selectable sound objects providing a clear overview of the entire project and allowing quick access to any track or event. The mixer can create spatial sound scenes by using a variety of automation features. Moving, rotating, scaling, and grouping functions are provided to manipulate virtual sources. The IOSONO company and Spatial Audio Workstation developed out of research at Fraunhofer IDMT. (IOSONO 2011)

5.2.4 Sampling the Virtual Source Space

The just noticeable difference for source location azimuth has been found to be 3.6° for sources in front of the listener and 10° for sources to the side. Since the WFS listening environment allows an arbitrary head rotation, we must exceed the threshold for frontal source location of 3.6° (Blauert 1997).

A naïve implementation of a WFS filtering program will calculate the transfer function for any virtual source location on demand, in a continuous space. A more efficient implementation will calculate a bank of filters ahead of time for virtual sources at a fixed grid of locations. For purposes of generating this bank of filters it is possible to sample the virtual source space without a perceptible loss in location resolution. The precise sampling period is dependent on the array size and the listening area, as the goal is to reduce the maximum direction angle between any two virtual source points to below 3.6° from the perspective of any valid listener position. Figure 5.3 shows the source space sampling strategy employed by Corteel (2006).

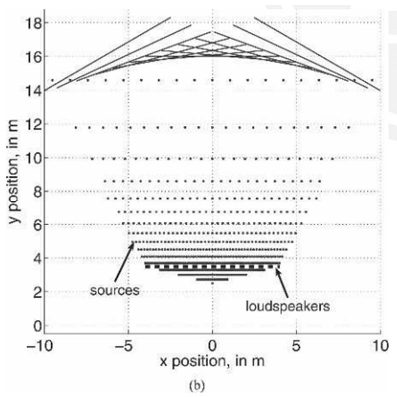


Figure 5.3: Virtual source space sampling (Corteel 2006). The spherical and focused source grid is shown by points, and plane wave discretization is shown with lines in the upper portion of the figure.

Chapter 6. Implementation: Loudspeaker Arrays

6.1 48-Channel Modular Array

6.1.1 Design Rationale

The modular array, shown in Figure 6.1, is easily reconfigured for different wave field synthesis array geometries. The enclosure is 5” by 20” so that it is stackable to form a 2-dimensional array with even spacing. I decided to build modules of four speakers each because this offers a good compromise between the two competing design goals of 1) array flexibility and 2) convenience/build simplicity. The modules were designed with low-cost in mind. The cost of all construction materials, including amplifier boards but excluding the audio interfaces, amounted to roughly $22 per channel.

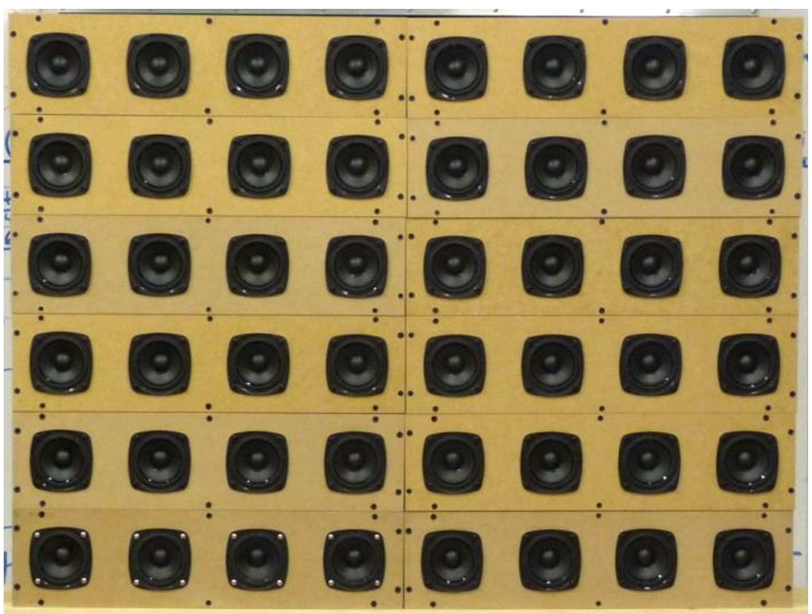


Figure 6.1: 48 channel loudspeaker array in planar configuration.

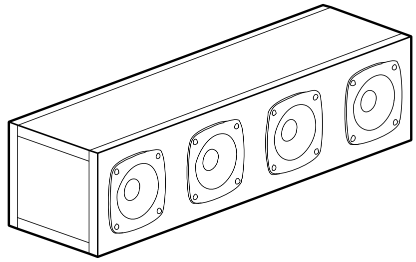


Figure 6.2: 3D CAD drawing of loudspeaker module.

6.1.2 Loudspeaker Characteristics

A 4 in. speaker, spaced 5 in. on center, was decided upon as the best compromise between the design goals of 1) small speaker spacing in order to reduce spatial aliasing and 2) full-range frequency response. A 5 in. spacing produces an f_{alias} of 2.7 kHz. The Tang Band W3-1053SC was selected for its flat overall frequency response (100–20000 Hz ±3dB) and high efficiency (86 dB). The enclosures were constructed from ½ in. MDF, with partitions between each speaker to create 4 sealed chambers. It was important to use a sealed chamber as opposed to a ported design to ensure the speaker would behave as much like an omnidirectional point source as possible.

The enclosure was designed in Bass Box Pro 6 and optimized for the flattest frequency response. The solution derived from Bass Box was a closed box design with an internal volume of 1.27 liters with heavy fill. This box coupled with the TB-1053SC speaker has an F3 of 138.5 hz.

The internal box dimension for each speaker in my module is approximately 4.0 x 4.25 x 4.5 in. (1.25 L) with heavy internal damping that boosts the volume seen by the speaker somewhere around 10 to 15% (this is accounted for in Bass Box). The damping also reduces standing waves inside the box, smoothing out bumps in midrange frequency response, and helps it to act as a simple air spring.

6.1.3 *Audio Interface and Amplification*

The MOTU 24I/O audio interface was chosen to drive the loudspeaker array. Two MOTU 24I/O interfaces are connected through a PCI424 PCI board. These rack-mounted interfaces were chosen because they have 24 onboard analog outputs, simplifying the hardware configuration and eliminating the need for ADAT converters.

The loudspeakers are amplified on 48 discrete channels with the Sure Electronics TPA3123 2 x 8 watt Class-D audio amplifier board.

6.2 16-Channel Desktop Array

An array was designed with small speaker distance to reduce spatial aliasing to a minimum. This array uses 2 in., 4-ohm drivers spaced at 2 in. This results in an f_{alias} of 6.8 kHz. The purpose of this array is to test the feasibility of compact WFS arrays for personal use. The array is designed to lay flat on a desktop or bookshelf and occupies about the same total volume of a pair of large bookshelf speakers.

Figure 6.3: 16-channel compact loudspeaker array.

Chapter 7. Implementation: WFS Designer

7.1 Introduction

An important component of my research was to build a flexible and user-friendly software platform for wave field synthesis. The objectives in this software were 1) use open source libraries in its construction, 2) maintain cross-platform compatibility, and 3) release the software into the community under an open-source license. An implicit goal is to make Windows binaries available for download. Finally, building the software from scratch offers the flexibility to extend the WFS engine for multiple line array synthesis, described in detail in Chapter 8.

Other solutions exist for performing wave field synthesis within other environments, such as Supercollider, Max/MSP, Pure Data, and MATLAB. In contrast to these, one of my goals was to develop a stand-alone software program to make wave field synthesis accessible outside of the research and experimentation domain. There are no turn-key wave field synthesis applications for the Windows operating system; this hinders the exploration and adoption of wave field synthesis outside the academic world. The Sound Scape Renderer from TU Berlin (Geier 2007) is a flexible program, but it must be compiled on either Mac OS X or Linux. It is sufficient then to say there is an opportunity to provide a stand-alone software wave field synthesis environment.

The wave field synthesis makes high demands of system resources and requires low-level access to sound hardware. A simulation might demand five virtual sources routed to 48 channels of output. The engine must be efficient in order to support as many WFS channels as possible, so C++ was selected as the development language.

7.2 WFS Visualizer

In preparation for development of the WFS Designer application, I developed an animated graphical application for simulating 2D wave field synthesis. The application helps to visualize the impact of various changes to the synthesis parameters in real-time. The WFS Visualizer is a Processing sketch/Java applet available online at http://www.mattmontag.com/wfs-visualizer. It allows the user to interact with the wave field synthesis simulation of a virtual source that follows the position of the mouse cursor.

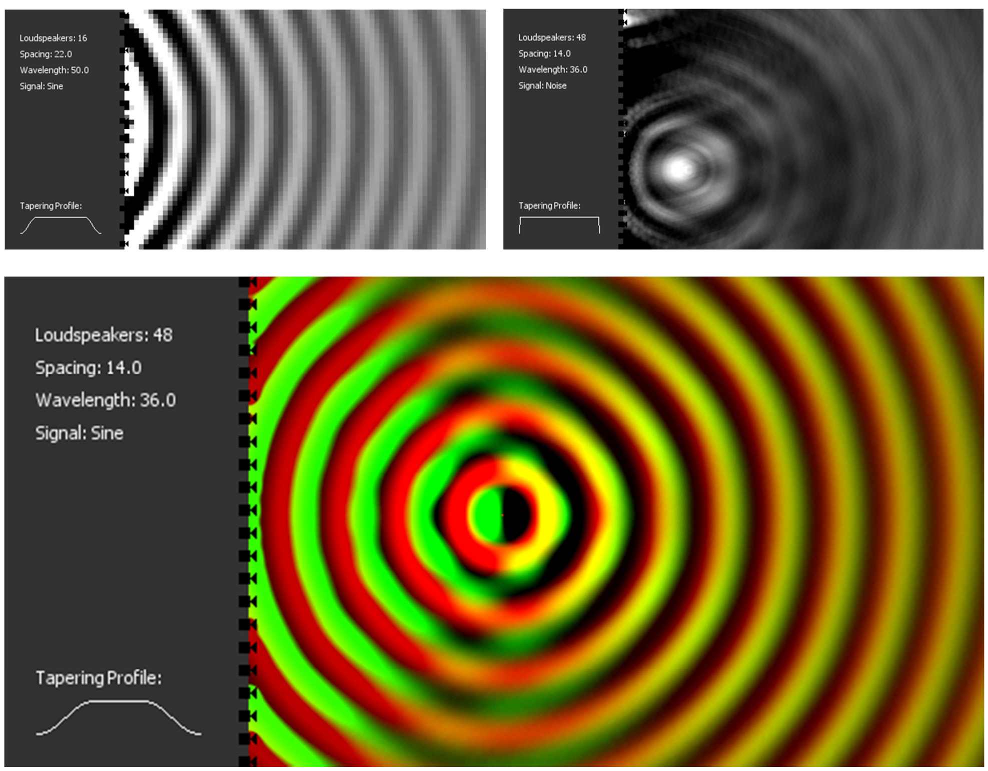


Figure 7.1: Various simulation modes of WFS Visualizer.

Figure 7.1 shows the WFS Visualizer in various display modes. The top left image shows synthesis of a pure tone spherical source. The top right image shows a broadband focused source. The bottom image shows the primary wave of the virtual source in red and the synthesized wave field in green, for comparison. Under ideal conditions, the red and green wave fields would perfectly coincide and result in a yellow image color. Thus, a persistent yellow area indicates a region of accurate synthesis. The user has control over the tapering profile, number of loudspeakers, array spacing, wavelength, signal waveform, and simulation resolution using the following keyboard shortcuts:

p	Toggle primary wave
1/2	Increase/decrease resolution
q/w	Adjust tapering profile
Left arrow/Right arrow	Decrease/increase number of loudspeakers
Up arrow/Down arrow	Increase/decrease array spacing
[/]	Decrease/increase signal wavelength
s	Change signal waveform (sine, noise, and saw)

Table 7.1: WFS Visualizer hotkeys

This application has proved to be a useful sandbox for experimenting with WFS parameters. WFS Visualizer is useful as an instructional tool because it allows the user to gain an intuitive understanding of what matters in reproducing an accurate sound field (although the apparent accuracy of the sound field does not always coincide with its perceptual validity). The limitations of the WFS, the extent of the valid listening area, and the effects of spatial aliasing immediately become clear when the array parameters are manipulated.

7.3 Software Libraries

WFS Designer makes use of several open-source, cross-platform libraries. The libraries and their role in WFS Designer are briefly described below.

7.3.1 Qt

Qt is a cross-platform GUI development framework. Several cross platform GUI libraries were evaluated, including MFC, Juce, WxWidgets, and Qt. Qt 4.7.1 was selected because of its maturity, robustness, and very good documentation. Qt is maintained by Nokia and is available under GPL and LGPL software licenses, which makes it suitable for academic purposes. Qt is available online at http://qt.nokia.com/.

7.3.2 FFTW

FFTW is a C library for computing the discrete Fourier transform of multidimensional, real or complex valued signals. The library provides optimized, cross-platform FFT and IFFT functions. WFS Designer uses fast convolution to apply filters, so it is heavily reliant on the Fourier transform. For M sources and N loudspeakers, WFS designer performs 2*M*N FFT operations per buffer. (Frigo 2005)

7.3.3 Libsndfile

Libsndfile is a C library for reading and writing audio file formats by Erik de Castro Lopo. It is released under the LGPL version 2.1 and version 3, depending on the needs of the developer. The role of Libsndfile in WFS Designer is to supply signal from local audio files to virtual sources in the WFS environment. (de Castro Lopo 2005)

7.3.4 PortAudio

PortAudio is a C library for audio I/O originally proposed by Ross Bencina and Phil Burk. It also is free and cross-platform, licensed under a GNU GPL-compatible MIT license. PortAudio was selected because it is a lightweight interface designed to permit real-time audio applications that run on more than one platform. PortAudio serves as a unified proxy for Windows MME, DirectX, ALSA, ASIO, and other audio host APIs. The role of PortAudio in WFS Designer is to provide a low-level, low-latency interface to audio devices across all platforms, regardless of the host API. (Bencina 2001)

7.4 Architecture

Software for real-time wave field synthesis control should allow the user to specify a physical loudspeaker arrangement, and configure the position of one or more virtual sources in the context of the loudspeaker arrangement. This section describes how these features are accomplished in WFS Designer.

WFS Designer is comprised of 6 primary C++ classes: WFSDesigner, WFSPortAudio, WFSFilter, PhysicalModel, VirtualSourceModel, and BiQuad. The WFSDesigner class is the application controller, and its methods contain logic necessary for the core wave field synthesis implementation. It owns instances of WFSPortAudio, VirtualSourceModel, PhysicalModel, and a 2-dimensional array of WFSFilter instances.

The loudspeaker array is represented by the PhysicalModel. The PhysicalModel holds a list of Loudspeaker instances, corresponding to the number of active output channels (Note: Loudspeakers are synonymous with output channels in this context, and the terms are used interchangeably). Changes to audio interface configuration are handled in WFSPortAudio, and the virtual sources are managed in VirtualSourceModel.

The VirtualSourceModel contains a list of VirtualSources. Each VirtualSource has properties for defining its position, its volume, its signal (a local audio file from disk), and its type (spherical source or plane source). If a source is a plane source, the direction of plane wave propagation is determined by the vector pointing from the source to the center of the listening scene. In this case, the distance of the source from the center has no effect. VirtualSource subclasses the QGraphicsItem class, so that it can be attached to the WFSCanvas, which is derived from the QGraphicsView widget. This means the majority of the user interaction behavior is taken care of by Qt. For example, a QGraphicsItem can simply set a flag that says “isMoveable” to acquire mouse draggable behavior. The “mouseMoved” method is overridden to provide a notification of new position to the WFSDesigner, which then propagates the event by updating the WFSFilters. Additionally, instances of the VirtualSource class are entirely responsible for maintaining their own signal buffer. This is implemented as a ring buffer.

Qt provides a built-in Observer pattern with its “Signals and Slots” mechanism. Briefly, this allows object instances to communicate while remaining loosely coupled. The observer pattern is better known in other languages as event dispatch or notification. The key feature is that multiple objects can attach themselves as listeners for a given event.

Although the operating system may furnish interaction behind the scenes in multiple threads, WFS Designer primarily operates in three logical threads: the event-driven UI thread, the PortAudio callback thread, and the VirtualSource ring buffer I/O thread. The ring buffer I/O thread is created at application startup and checks on a timer interval to see if any VirtualSource ring buffer needs to be filled with new audio data.

At application startup, PortAudio is initialized with default output device options, a callback function is specified, and then the stream is started. This creates the PortAudio thread, which calls the callback function at the interval set by the output buffer length.

The callback function calls a method in WFSDesigner, processAllChannels. The PortAudio thread is now executing in the scope of the controller.

ProcessAllChannels is a method responsible for filling the entire PortAudio output buffer with signal data for every loudspeaker in the array. The method contains a nested for-loop. The outer loop iterates over every active VirtualSource in the scene, calling each source's copyFromRingBuffer to obtain a chunk of audio data equal to the PortAudio buffer length. For each source M, the inner loop filters the source signal with the WFSFilter[M][N] (the filter that maps source M to channel N) and adds the result to the output buffer for channel N. The filtering operation will be discussed in detail momentarily.

WFS Designer also implements a fast delay-line-only wave field synthesis engine. It is appropriate for large numbers of loudspeakers and slower computers. In this implementation, WFS Designer maintains a simple table of delays, calculating a delay time for each source-loudspeaker pair.

When the user drags a VirtualSource within the WFS Designer user interface, a few things happen. First, the underlying position data of the VirtualSource instance is updated through the QGraphicsScene framework. Then, a sourceMoved signal is emitted by the VirtualSource instance. The controller is instructed to listen for this signal when any VirtualSource is created. The controller handles the event by running updateFilterBank on this VirtualSource. The updateFilterBank method is where the principles of wave field synthesis are carried out, in order to create the appropriate filters representing the transfer function of each VirtualSource M to each Loudspeaker N. Alternatively, if delay-line implementation is enabled, the controller calls updateDelays and updateGains

Processing power limits the length of the FIR. For example, a 48-channel array simulating four virtual sources with an FIR length of 2048 samples will demand, by multiplication, the equivalent of a 393216-point convolution in real time. If this operation were performed with direct convolution, it would require 17 billion multiply-accumulate operations per second for a typical sample rate of 44100 Hz. Fast convolution with FFTW makes this a feasible operation.

7.4.1 Creating Filters for WFS

The filter update procedure can be discussed in straightforward language. The signal must be delayed, amplitude scaled, low-pass filtered, and processed through a virtual room acoustics model. We can apply these operations to the signal itself, or apply them to an impulse and convolve the signal with the result. The operations can be applied in any order since it is a linear time-invariant system. The intent is to create a finite impulse response representing the transfer function from a virtual source M to the loudspeaker N. Representing the transfer function with an impulse response gives us flexibility to apply long reflection and reverberation impulse responses generated by room acoustic models.

7.4.1.1 Constructing the FIR Filter

An FIR filter is instantiated for each source-loudspeaker pair. This will be convolved with the audio signal for virtual source M to arrive at a solution for the output of speaker N. First we check that the inward normal vector of the array at the position of speaker N shares a positive component with the direction of the source's wave propagation (their dot product is greater than zero). If this is not the case, then the source has no wavefront contribution from the direction of loudspeaker N. We begin with a unit impulse at time 0; the first sample in the FIR. The impulse is delayed by a number of samples corresponding to the distance of the virtual source M from the loudspeaker N. For a source at distance of 4 meters, this corresponds to about 500 samples. Fractional delay is not currently implemented in WFS Designer because the phase distortion contributed by this temporal quantization is small. When rounding to the nearest sample, the signal delay will be off a maximum of half a sample period – typically $\frac{1}{88,200}$ of a second; the wavefronts from the loudspeakers may deviate from the ideal by 4 millimeters. For a loudspeaker spacing of 12 cm, the spatial aliasing frequency is 2875 Hz. A $\frac{1}{88,200}$-second delay at this high-frequency limit is a phase difference of 3.26%. This intuition is supported by the thorough simulation results of Daniel Salvador (2010). He states "the fractional delays [do] not affect [the] discretization, where integer delays or 1st order FIR or IIR filters are enough." (Salvador 2010) Finally, the smallest delay in each source group is subtracted from the entire group; the loudspeaker closest to a source will have a delay of zero for that source.

7.4.1.2 Amplitude Scaling

The amplitude scaling is performed by defining a reference listener distance, as discussed in section 3.2. In circular array configurations, the listener position is the center of the array. In linear array configurations, this reference distance defaults to 2 meters. The distance gain is calculated for each source-loudspeaker pair as $\sqrt{\frac{B}{A+B}}$, where A is the source to loudspeaker distance, and B is the nominal loudspeaker to listener distance. The wavefront obliqueness gain is calculated as the dot product of the inward loudspeaker array normal and the vector pointing in the direction of wave propagation. These two factors are multiplied to give the final amplitude scale factor.

7.4.1.3 Room Acoustics Modeling and Room Compensation

At this point in the process, the impulse response still contains one impulse, delayed and attenuated. A room acoustics model can be excited with the impulse. The FIR length is adjusted to fit the result of the virtual room impulse response. Additionally, a room compensation filter can be applied. This step is mentioned for completeness. At the time of writing, WFS Designer does not implement a room acoustics model or room compensation tool.

7.4.1.4 Low-Pass Filter

As previously discussed, the output signal must be low-pass filtered to avoid artifacts caused by spatial aliasing. This is performed with a 4^{th} order Linkwitz-Riley crossover. The crossover is performed once per source. The high-frequency content is directed to the active VBAP loudspeakers, and the low-frequency content is passed to the WFS filter bank.

7.4.1.5 Array Windowing

To reduce the diffraction artifacts mentioned in Chapter 3, the loudspeakers toward the edge of the array are attenuated in amplitude. If the array is circular, no attenuation is applied because the array is continuous. The user can apply custom attenuation to each loudspeaker in irregular arrays.

7.5 WFS Designer Features

7.5.1 Vector-Base Amplitude Panning

Vector-base amplitude panning (VBAP) is a method for positioning virtual sources to arbitrary directions using multiple loudspeakers (Pulkki 1997). The concept is the same as that used in conventional stereo panning, expanded to a multichannel speaker configuration.

Vector-base amplitude panning is a prerequisite for sub-band mixing, so it is incorporated in the application by necessity. For purposes of testing and comparison, the wave field synthesis engine can be bypassed and the original signal can be processed using VBAP.

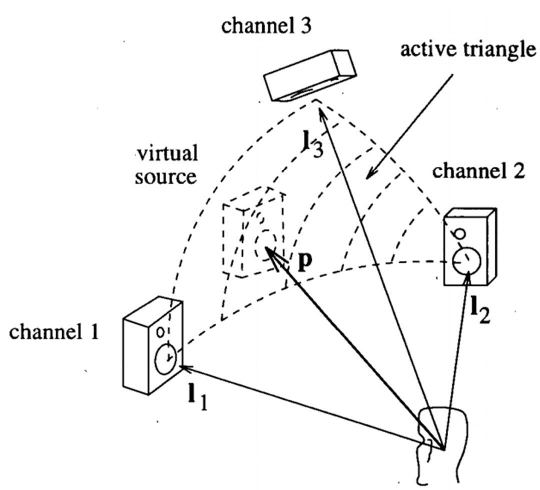


Figure 7.2: Vector base amplitude panning. (Pulkki 1997)

7.5.2 Sub-Band Mixing/High-Frequency Amplitude Panning

The signal content is split into high- and low-frequency paths at the spatial aliasing frequency. A matrix of coefficients is maintained, mapping each source to each loudspeaker. The coefficients store the solution of the panning algorithm. The VBAP solution results in equal-power panning. If a source is positioned between neighboring speakers A and B, the gain factor will be $\frac{2}{\sqrt{2}}$ for both loudspeakers A and B, and 0 for the rest of the speakers. The high-frequency content is multiplied by these mixing coefficients and summed with the WFS result at each output channel. The WFS delay normalization makes the delay time zero in the direction of the source, which will always match the loudspeaker placement of the high frequency content. This means it is not necessary to delay the VBAP solution. The low- and high-frequency content will be coincident.

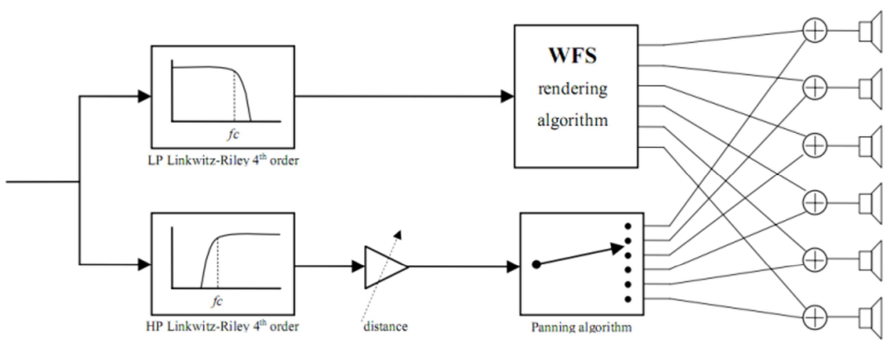


Figure 7.3: Sub-band mixing model. (Lopez 2005)

7.5.3 Virtual Room Acoustics/Image Source Model

WFS Designer supports room acoustics simulation derived from the image source model. The room acoustics model should be sophisticated enough to take source and listener position into account as well as the virtual room parameters. For the purpose of creating the room responses, each loudspeaker is considered a listener. Note that for accurate wave field synthesis application of room modeling, the system must be aware of the actual listener location. It is correct to consider each WFS loudspeaker position as the listener position in computing the room response, but the important difference is that the WFS loudspeaker is a “directional” listener – and should only hear waves traveling toward the actual listener in the WFS listening room. Waves traveling in a direction from the listener to the array should not be emitted. Virtual room acoustics for WFS has successfully been implemented by Brix et al. (2010).

Loudspeakers must be selected for image source synthesis in the same way as they are selected for primary sources – based on the source's direction of propagation relative to the normal of the array. In Figure 7.4, the situation is simplified to one reflection for illustration. the primary source wave (red) is emitted from the loudspeakers shown in red. Likewise, the source's reflected wave from the top wall (green) is emitted only from the loudspeakers shown in green. The loudspeakers indicated in orange emit both the red and green wave. No sound is emitted from the loudspeakers shown in black. Reflections from other walls are omitted for clarity.

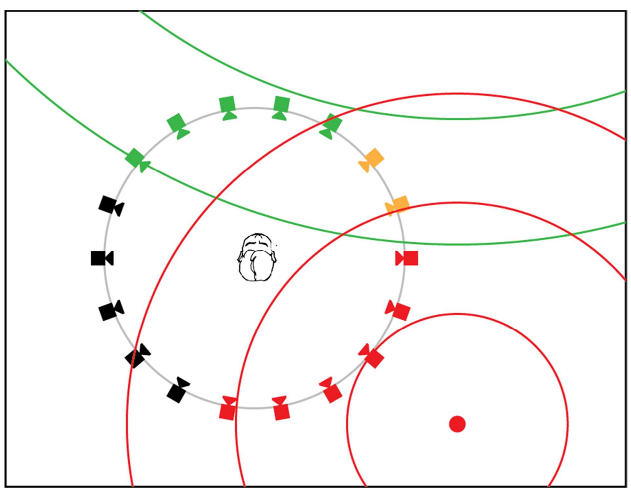


Figure 7.4: Synthesizing room acoustics with wave field synthesis.

7.5.4 3-Dimensional Virtual Environment

WFS Designer allows the user to place spherical sources in a three-dimensional space. This requires a loudspeaker array with a 2-dimensional component. The built-in Qt classes QGraphicsScene, QGraphicsView, and QGraphicsItem are designed to handle 2-

dimensional position, transforms, and display, and must be extended to take on a z-position. This is handled in an ad-hoc manner so as to least disturb the functioning of base class methods like pos() for which Qt expects a return value of a 2D QPoint. The Scene Layout view can be toggled between a 2D view and a 3D view. The 3D view is rendered using the OpenGL graphics library.

7.5.5 Loudspeaker Positioning

WFS Designer features an array configuration control panel that allows the software representation of the array to easily be manipulated to match the physical array setup by specifying listener distance, loudspeaker spacing, and array height. The control panel, shown in Figure 7.5, allows configuration of six different arrangements: line, circle, double line, double circle, box, and U-shape. The double line and double circle options correspond to multiple line array configurations.

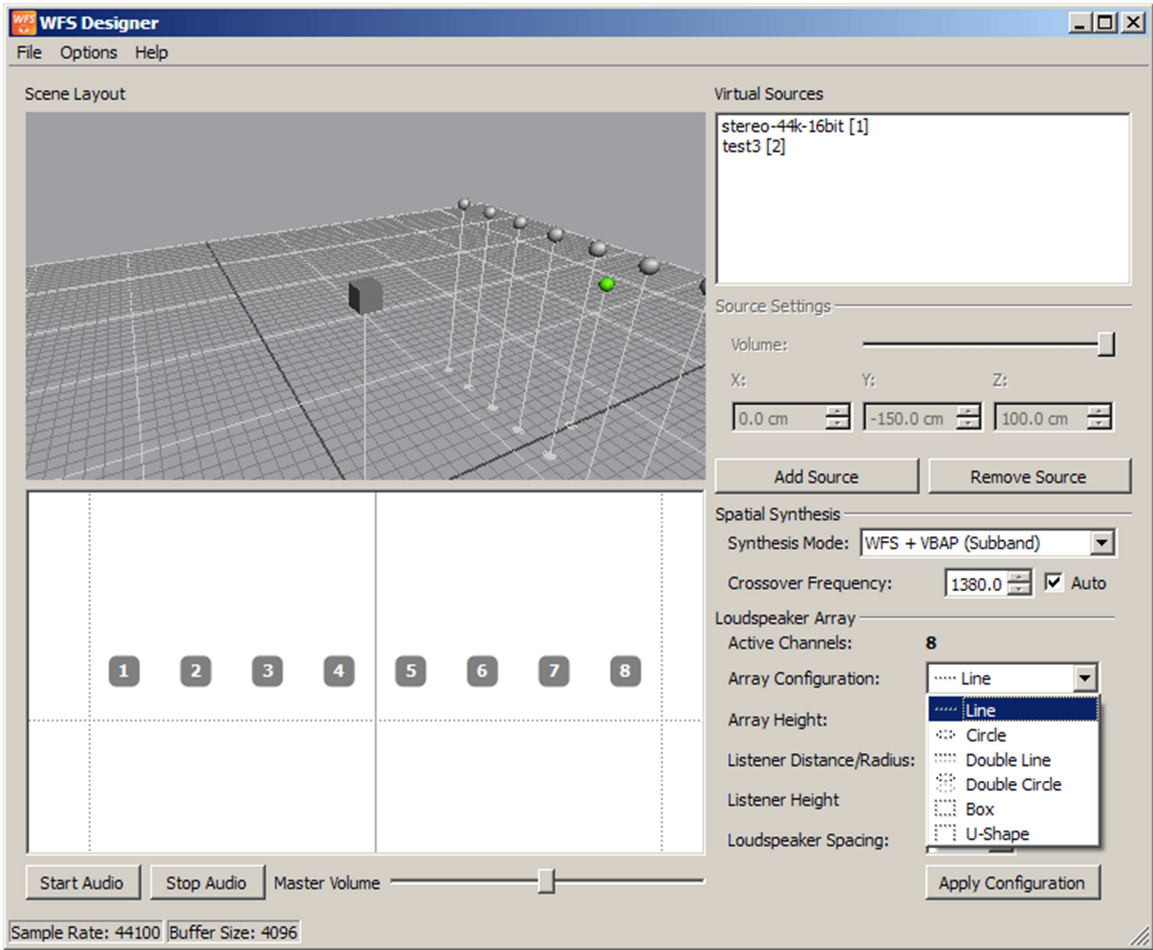


Figure 7.5: WFS Designer's 3D scene layout and array configuration tool.

WFS Designer allows some flexibility in the array configuration, but the synthesis parameters (spatial aliasing frequency, amplitude factor, array tapering) are determined when the array is initialized (at application startup and when the user clicks the Apply Configuration button). This allows the user to tweak loudspeaker positions in the GUI to match bent array configurations and other unique layouts. WFS Designer will calculate the appropriate delay times and source relative gains for each loudspeaker. However, the global array parameters are not updated and are not guaranteed to maintain a valid result. For example, array tapering gains are statically assigned to the first and last range of loudspeakers in a linear array. If the user moves the first loudspeaker from its original position to a position in the center of the array, it will still have a tapering attenuation applied as if it were on the end of the array. This may be addressed in future versions of the program.

Flexibility in loudspeaker placement removes the constraint of line or circle arrays. Circle arrays are optimal for a listener at the center of the array, but line arrays are not optimal for listeners centered in front of the array. With a line array placed in front of the listener, the apparent azimuth angle between adjacent speakers is wider at the center of the array than at the edges. To avoid artifacts from spatial aliasing, the wave field synthesis signal must be low passed at a frequency inversely proportional to the largest distance between any two adjacent speakers in the array. However, the crossover frequency can be overridden for arbitrary array configurations.

One situation where this is advantageous is in the line array configuration with a center-biased listener position. Line arrays designed for theater usage must present a valid spatial field for all occupants, from one side of the listening room to the other. In this situation an equally spaced line array is appropriate. However, for a listening situation that tends to situate the listener toward the center of the room, a line array of speakers on the front wall might be spaced such that the apparent angle between adjacent speakers remains constant. This means the speakers near the center should be spaced closer together, and the advantage comes from the fact that the spatial aliasing cutoff can be raised.

It must be reiterated that this freedom allows the user to very easily arrive at problematic and invalid array configurations, as described in Chapter 2.

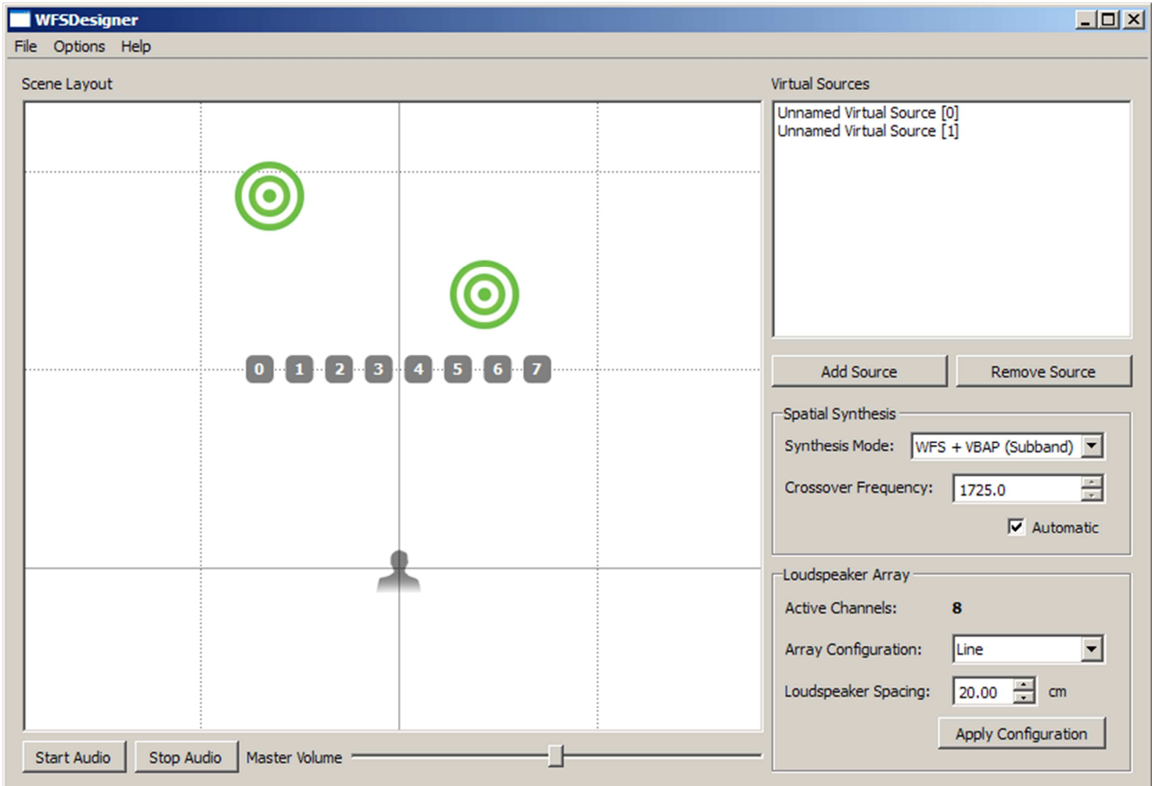


Figure 7.6: WFS Designer manipulating two virtual sources.

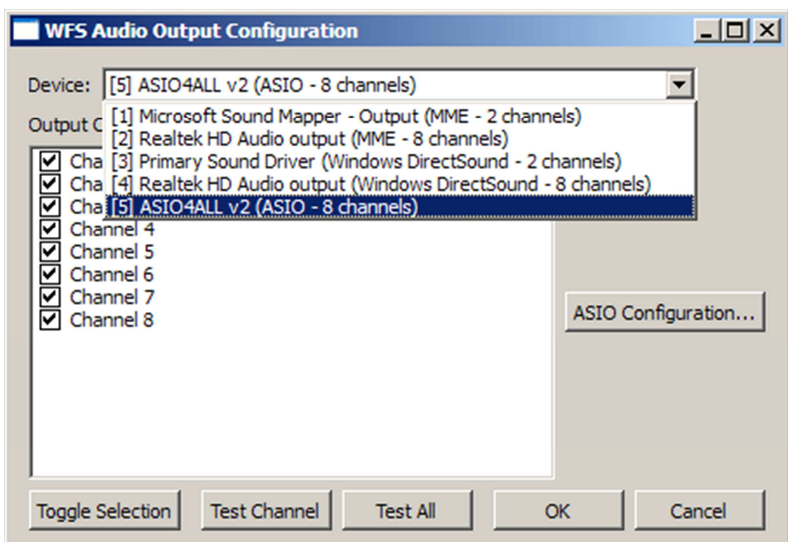


Figure 7.7: WFS Designer audio output configuration screen displaying available host APIs.

Chapter 8. Proposed Enhancement to WFS: Multiple Linear Arrays

As highlighted in section 3.1, one of the principal limitations of wave field synthesis is the restriction of virtual sources to the horizontal plane. It remains impractical to perform wave field synthesis using planar loudspeaker arrays not only due to the sheer number of loudspeakers involved, but also due to the discrete output channel and amplification requirements. To escape limitation to the horizontal plane without a geometric increase in the number of required loudspeakers, I propose a configuration of vertically stacked linear arrays to expand spatialization to the height dimension. It is proposed that this can be achieved without distortion or compromise to the listening area. It extends existing approximations without introducing new errors into the synthesized wave field, and without introducing unacceptable limitations.

When wave field synthesis is performed with a horizontal line array, the listener receives a physically valid wavefront in the horizontal direction. But the listener is not presented with a physically accurate wavefront in the vertical direction. This is a well-known and accepted limitation of wave field synthesis. The wave field is only valid on the horizontal plane that contains the entire loudspeaker array This is an acceptable approximation because the listener, while free to move about the listening area, can be expected to remain in a fairly constant vertical position. By the same token, I propose that spatialization for vertical sources can be performed without setting off any "perceptual alarms."

In the proposed model, a loudspeaker line array positioned at the ear level is duplicated at the floor, at the ceiling, or at both the floor and ceiling. The virtual source space can be imagined in spherical polar coordinates, with the listener at the origin. Virtual sources are "steered" with wave field synthesis in the horizontal (azimuth) axis and in depth, and with vector base amplitude panning in the vertical (elevation) axis. The proposal can best be understood by example: if a virtual source is positioned between a line array and another identical line array duplicated 6 feet above it, a single horizontal WFS solution is calculated and emitted at equal gain from both top and bottom line arrays. If the virtual source moves closer to the top array, the WFS solution is attenuated in the bottom array and intensified in the top array, just as the phantom source in conventional stereo amplitude panning. Therefore, the virtual source, based on the description of its synthesis method, is now both a phantom source and a virtual source.

Start (1997) investigated the effects of vertical deviations in listener position in the context of direct sound enhancement with wave field synthesis. The consequences of vertical displacement are relevant to direct sound enhancement in concert applications because the audience cannot be expected to occupy a single vertical position.

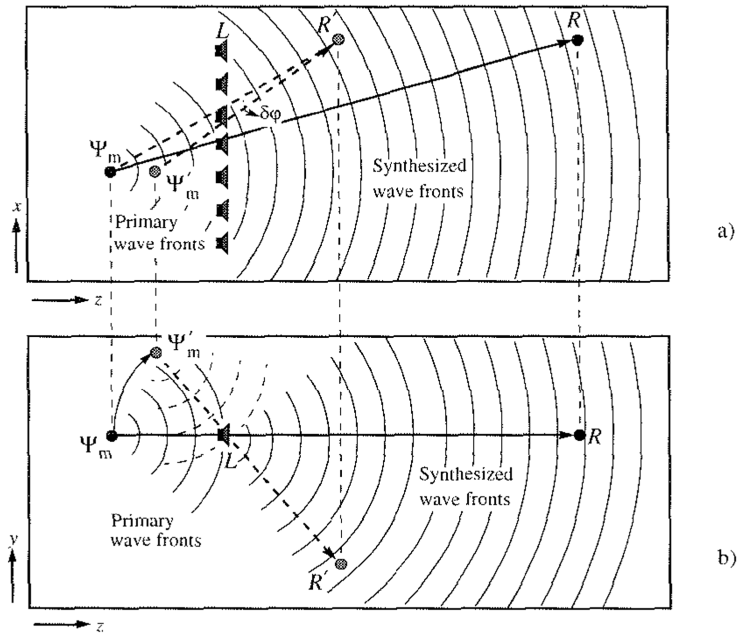


Figure 8.1: Apparent source position for listeners at different heights. (Start 1997)

As shown in Figure 8.1, the apparent source position is different for two listeners R and R' at unequal heights. Diagram (a) shows a top-down view, and diagram (b) shows a side-view. For receiver R, who is positioned at the same height as the array L, the apparent source is positioned correctly with respect to azimuth and elevation. For receiver R' at a lower position, the apparent source is rotated around the array L, yielding an elevation of the source and a small azimuthal deviation δφ. (Start 1997) The important insight is that the virtual source is always positioned in the direction of the array from the listener's perspective.

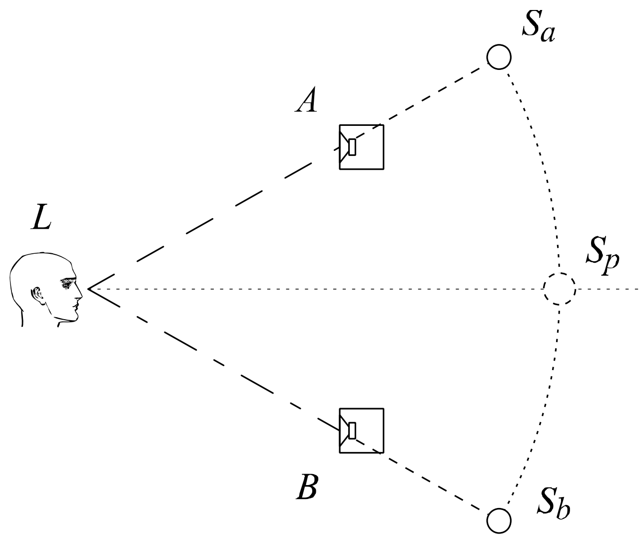


Figure 8.2: The "phantom virtual source."

In multiple linear WFS, this property is employed as shown in Figure 8.2. Line array A creates a virtual source at position S_a, and line array B produces a virtual source at S_b. S_a and S_b perceptually merge at phantom source position S_p. In this scenario, there are two simultaneous WFS planes, one defined by the plane containing line array A and listener L, the other defined by the plane containing line array B and the listener. By virtue of the WFS performed in the loudspeaker array planes, S_p is endowed with a stable distance characteristic not observed in traditional phantom sources. I have characterized *Sp* as a "phantom virtual source" because it inherits properties from both the phantom source and virtual source. The effective virtual source region is now the entire pyramidal volume extending from the listener through the corners of arrays A and B.

The listener will receive the sum of the WFS result in plane AL and plane BL. If array A and B are driven with the same WFS solution, then in order to receive a coherent synthesized wave field, the listener must be an equal distance from array A and B.

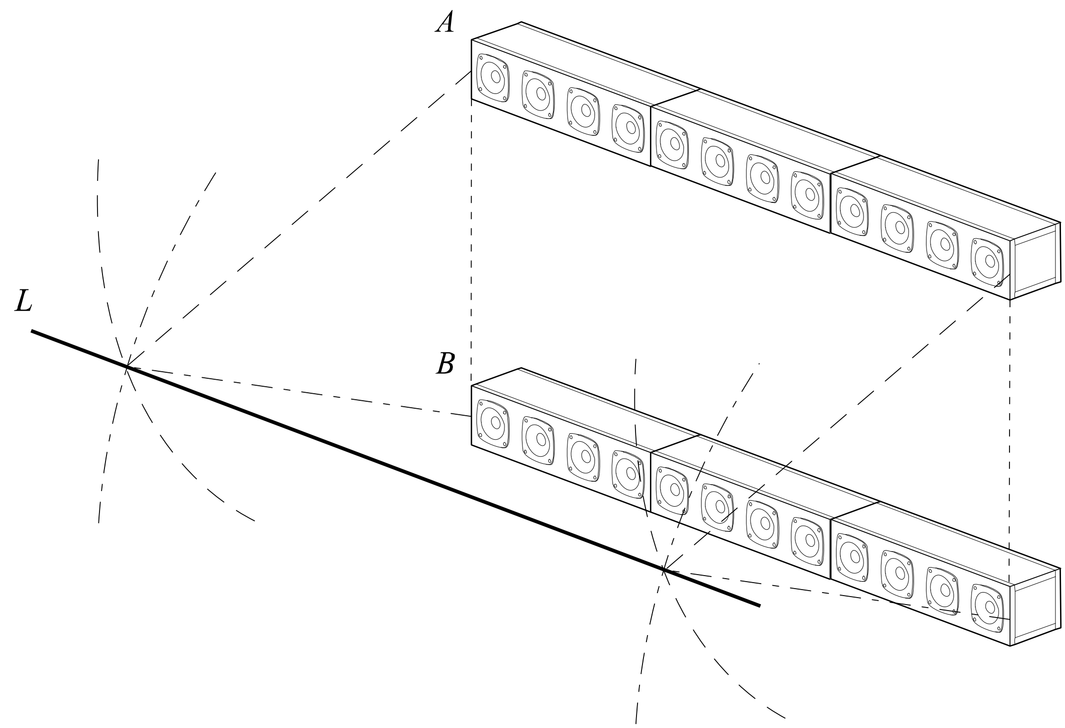


Figure 8.3: Example of a multiple line array WFS loudspeaker configuration.

Recall that 2.5D wave field synthesis suffers from an amplitude decay error arising from the reduction of the 2D Rayleigh integral plane to a line. This must be corrected by choosing a reference distance in front of the array at which synthesized sources will match natural output levels, as explained in section 3.2. In a multiple-line-array WFS configuration as shown in Figure 8.3, the reference distance for array A coincides with the reference distance for array B at line L. The net result is that there is no additional penalty restricting listener movement. If the listener strays from the reference line L, he will perceive the same amplitude error as he would in a single-array configuration.

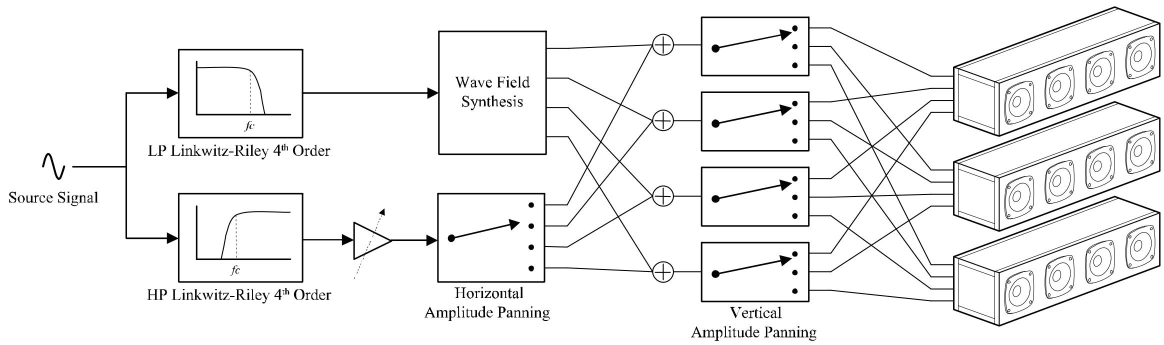


Figure 8.4: Signal flow in multiple line array WFS.

The perceptual validity of performing WFS in the horizontal axis and amplitude panning in the vertical axis is explored in the experimental chapter (Chapter 9).

Chapter 9. Experiments

9.1 Listening Tests

A listening test was devised to evaluate the perceptual validity of the phantom virtual source. The listening test was performed in the semi-anechoic recording loft in the Maurice Gusman Concert Hall on the University of Miami campus.

A loudspeaker array was configured in two 20-loudspeaker rows at a height of 82 and 216 cm. The bottom row was positioned 290 cm away from the listener location. The top row was offset a further 33 cm away due to constraints of the support apparatus. A curtain was placed between the listener and the loudspeaker array to reduce the "ventriloquist effect," the tendency for visible objects to influence sound source localization. The curtain also provided a surface for listeners to shine a laser pointer at, indicating the perceived direction of a sound source.

Figure 9.1: Setup of the listening test environment.

Thirteen subjects participated in the test. Subjects were given a brief introduction to the experiment and asked to indicate the perceived direction of test tones by shining a laser on the curtain, and rate the locatedness and distance of test tones on a scale of 1 to 7. The position of the laser on the curtain was photographed for each test tone.

Some listeners were asked to take the test from multiple listening positions. There were 7 tests taken from the left listening position, 5 from the center, and 5 from the right. In total, 170 test tone evaluations took place in 17 test runs among 13 listeners.

Listeners were presented with 10 test tones rendered as spherical sources in the WFS Designer software environment. The test tone consisted of six 0.5 second pulses of white noise separated by 0.5 seconds of silence. The noise was band-limited to the spatial aliasing frequency of 2.7 kHz. Listeners were able to repeat each test tone upon request. Each of the 10 test tones varied in its virtual source position. The virtual sources were positioned as indicated in Figure 9.5. These 10 positions were chosen so that 1) the stable source positioning characteristic of traditional wave field synthesis could be validated, and 2) phantom virtual sources could be evaluated independently of virtual sources for analysis and comparison. Test tones 3, 4, 5, and 9 are placed on the extreme upper or lower edge of the valid virtual source space so that they activate only the top or bottom row of loudspeakers. This makes these sources *virtual sources* as produced by traditional WFS. Test tones 1, 2, 6, 7, 8, and 10 are positioned somewhere between top and bottom so that they are reproduced by both rows, and qualify as *phantom virtual sources*.

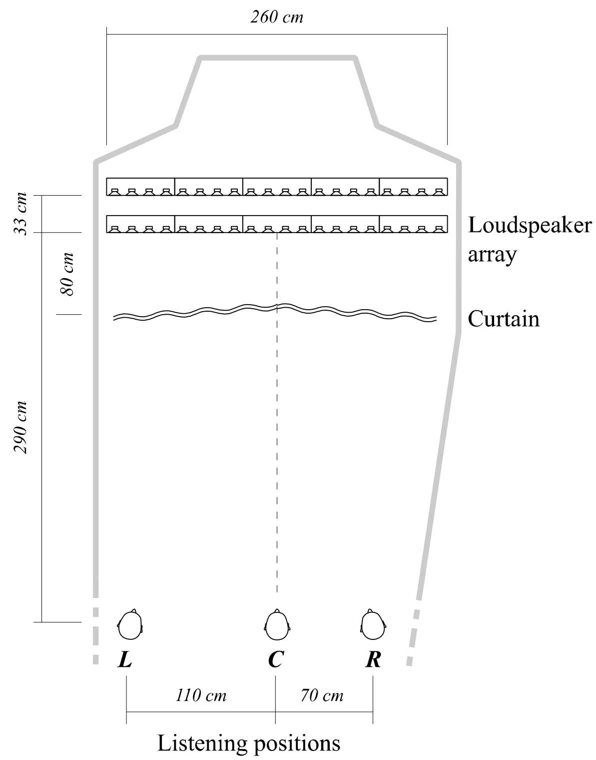


Figure 9.2: Plan view of listening test environment.

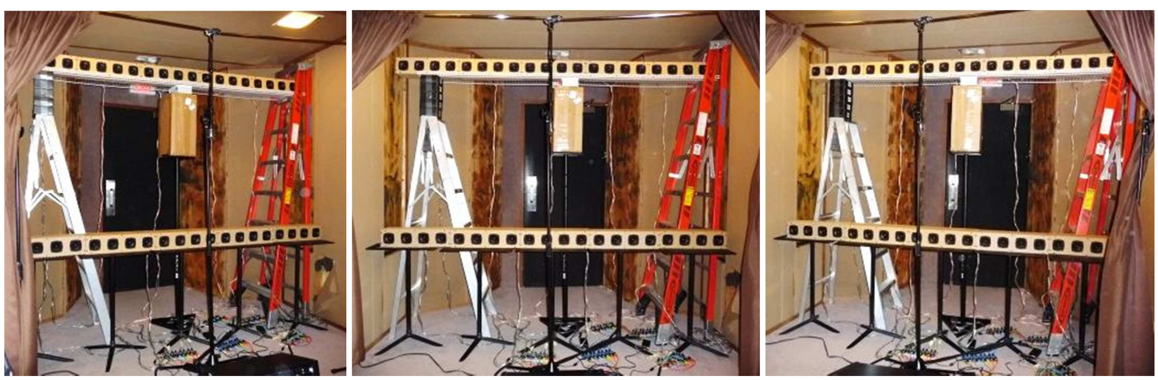


Figure 9.3: Unconcealed loudspeaker array seen from left, center, and right listening positions.

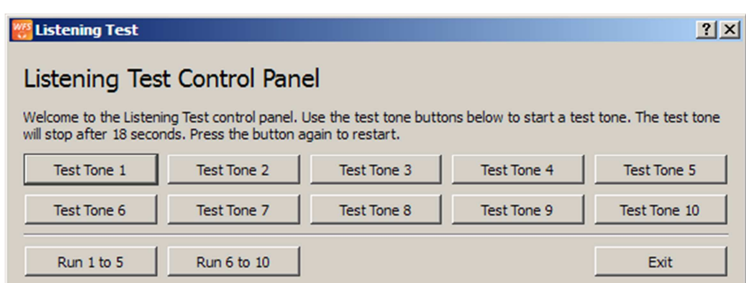


Figure 9.4: Listening test control panel in WFS Designer.

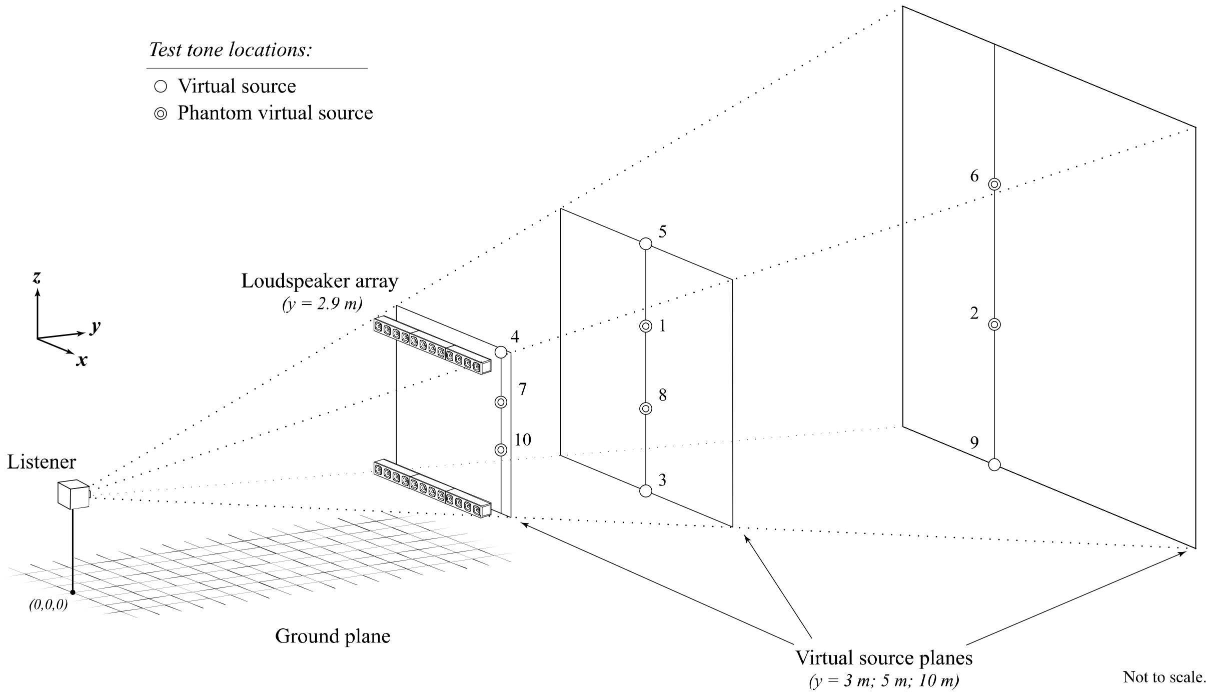


Figure 9.5: Test tone virtual source locations.

9.2 Confirmation of Stable Distant Sources

The validity of stable distant virtual sources is validated by comparing the localization pattern for distant virtual sources at different distances. Correct representation of source cues by wave field synthesis will result in localization vectors that converge at the appropriate distance; closer for nearby virtual sources, further for distant sources.

This is informally confirmed in the test results. Taking two example test tones, 5 and 10, we expect to see the localization pattern indicated in Figure 9.6. While there appears to be significant localization error, notice that the average horizontal position of listener responses closely matches the reference source direction, validating the stable source cue.

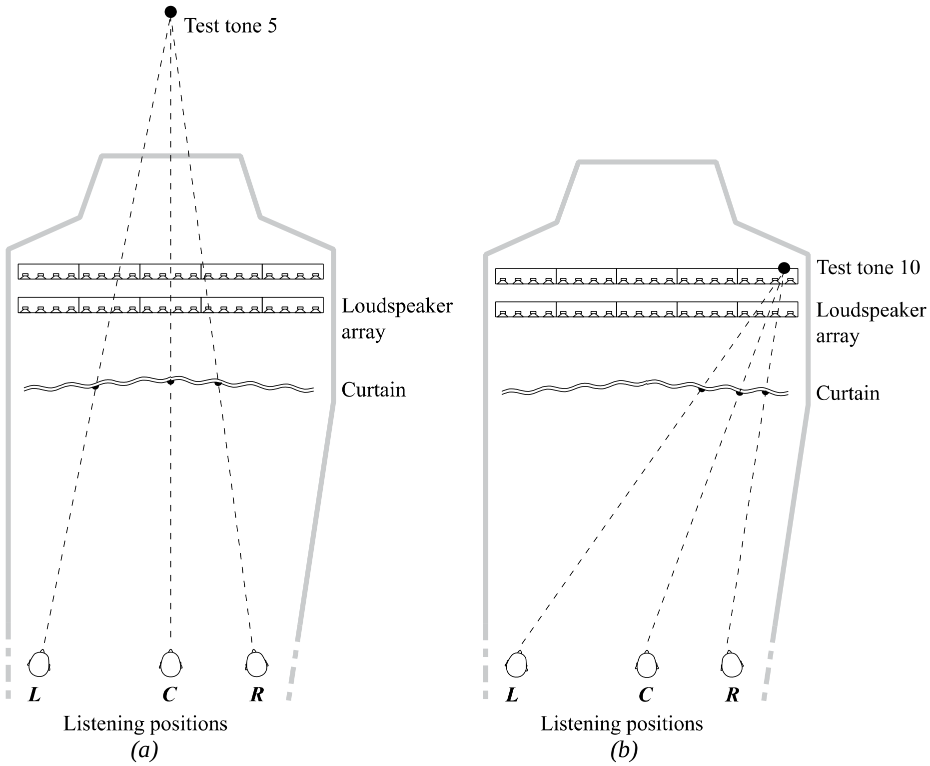


Figure 9.6: Example virtual source positions for test tone 5 (a) and test tone 10 (b). Test tone 5 has widely spaced apparent locations (indicated on the curtain), while test tone 10 has tightly grouped apparent locations.

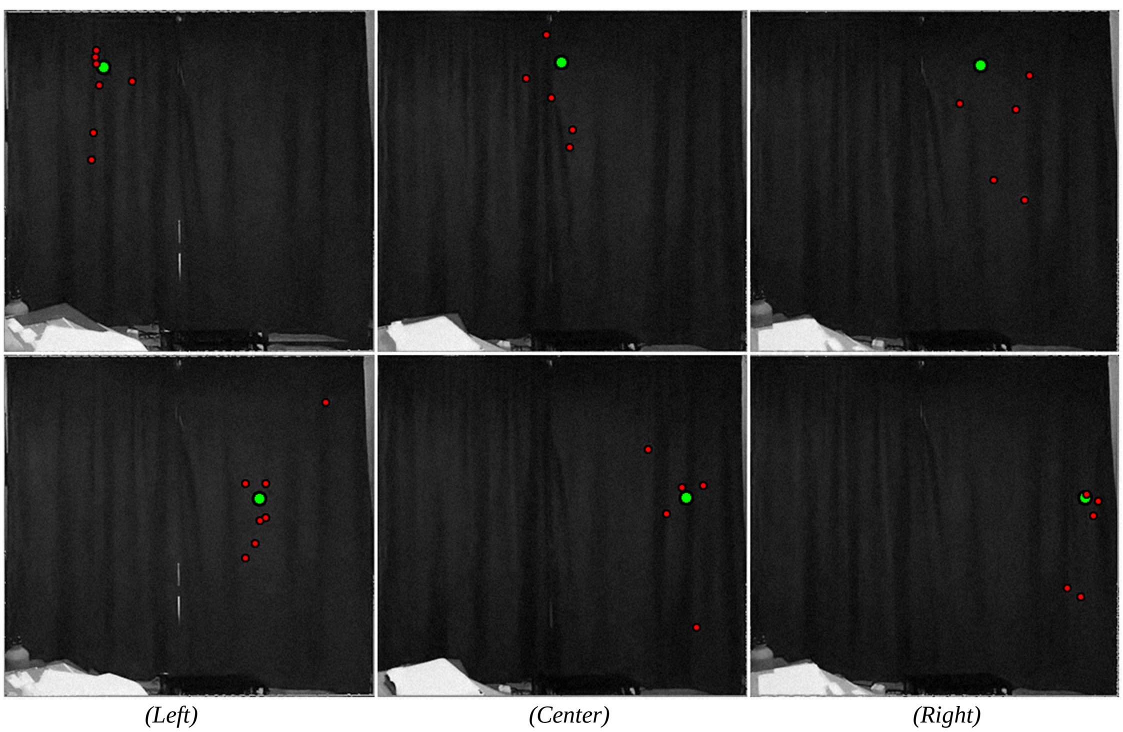


Figure 9.7: Source direction test results for test tones 5 (top row) and 10 (bottom row) from all listening positions. The reference position is shown as a green marker, and the clustered red markers show listener responses.

9.3 Localization Error

The reported source positions of all 170 listening trials are superimposed in Figure 9.8. The reference position is subtracted so that only the localization error remains. The data points are separated into two groups; blue dashes represent trials for virtual sources, while red crosses represent trials for phantom virtual sources.

The overall vertical localization error was significantly higher than the horizontal localization error, with a sample standard deviation of 6.5 degrees. The horizontal error exhibited a standard deviation of 3.3 degrees. This is consistent with the established tenet that humans are better at horizontal localization.

It is apparent from Figure 9.8 that the results of virtual sources are not significantly different from results for phantom virtual sources. This discussion is followed by analysis of variance tests to confirm the equivalence of error between the two groups.

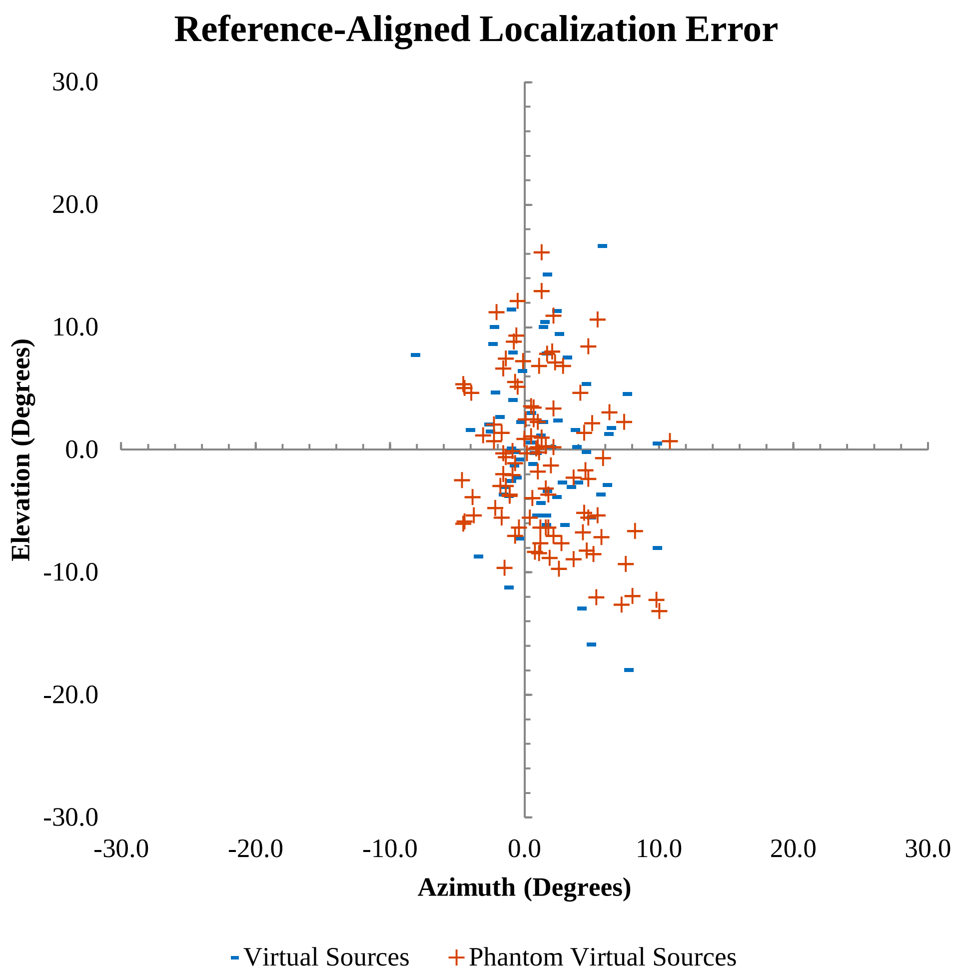


Figure 9.8: Localization error for all trials grouped by source type.

9.3.1 ANOVA of Localization Error in Virtual vs. Phantom Virtual Sources

Single factor analysis of variance can be used to test whether two groups of data represent truly different probability distributions. The test data were grouped into two batches: localization error of all trials with a virtual source, and localization error of all trials with a phantom virtual source. ANOVA was performed on these two groups, shown in Table 9.1 for the horizontal error and Table 9.2 for the vertical error.

SUMMARY

Groups	*Count*	*Sum*	*Average Error*	*Variance*
Virtual Sources	68	78.70	1.16	10.98
Phantom Virtual Sources	102	144.00	1.41	11.55

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.64	1	2.64	0.23	0.63	3.90
Within Groups	1902.07	168	11.32			
Total	1904.71	169				

Table 9.1: ANOVA of horizontal localization error in virtual vs. phantom virtual sources

The F value for horizontal localization error is well below F_{crit}, so we accept the null hypothesis; namely, the horizontal localization error is not significantly different for phantom virtual sources.

SUMMARY

Groups	*Count*	*Sum*	*Average Error*	*Variance*
Virtual Sources	68	26.64	0.39	44.64
Phantom Virtual Sources	102	-88.18	-0.86	41.16

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	64.39	1	64.39	1.51	0.22	3.90
Within Groups	7148.20	168	42.55			
Total	7212.59	169				

Table 9.2: ANOVA of vertical localization error in virtual vs. phantom virtual sources

Again, the F value is below F_{crit} for vertical localization error. We conclude that there is no difference in vertical localization error between virtual sources and phantom virtual sources based on the data collected.

9.4 Locatedness and Distance

The locatedness and distance survey did not yield valid results. It is likely that the dominant loudness cue confounded distance perception. Some test tones were noticeably louder than others depending on the listening position, and this was not measured or corrected. Test tones 4, 7, and 10 were rated significantly more distant than other sources, when in fact they were the nearest virtual sources (as indicated in Figure 9.5), most likely because of attenuation due to being positioned along the far right edge of the array. This invalid result corroborates Wittek's report that a source distance cue is not perceived through the curvature of the wavefront alone (Wittek 2007).

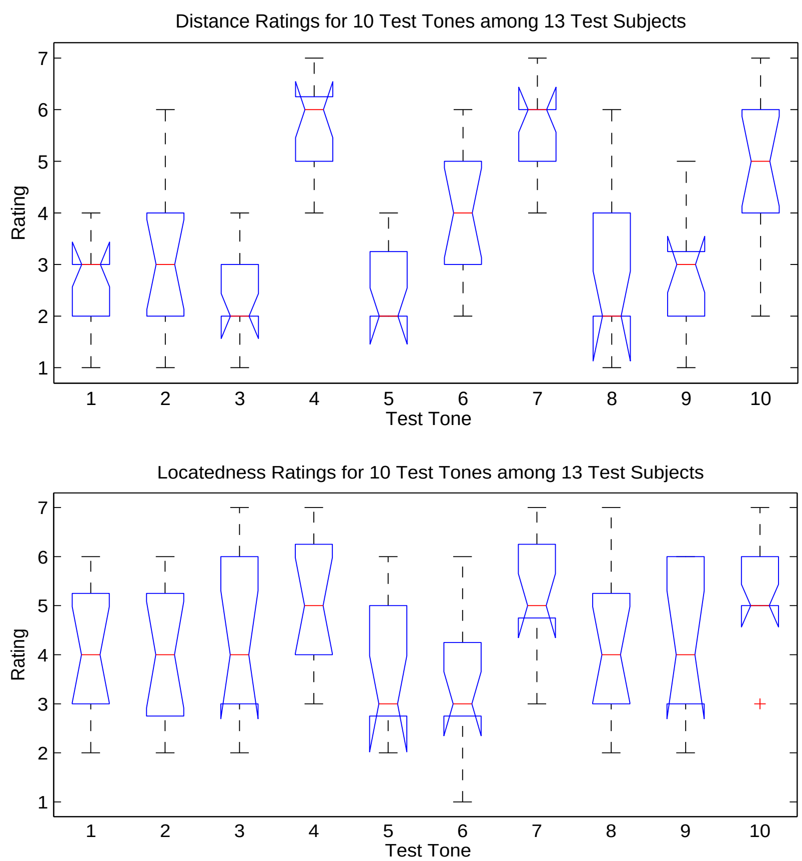


Figure 9.9: Distance and locatedness survey boxplots. The median rating for each test tone is indicated by the red line.

Chapter 10. Conclusion and Future Work

The objectives of this thesis were to build a low-cost, modular, and rapidly configurable loudspeaker array, create an open-source, cross-platform wave field synthesis software environment, and enhance wave field synthesis by expansion to the height dimension. In all three respects, the project was successful. The modular loudspeaker array will remain at the University of Miami to support future research. Work will continue on WFS Designer and it will be made available for public download. Finally, the proposed enhancement to WFS was experimentally validated.

10.1 Validity of Multiple Line Array Wave Field Synthesis

The validity of the proposed method, wave field synthesis in three dimensions by multiple line arrays, was confirmed by the listening test. Listeners did not exhibit greater localization error for virtual sources positioned between the upper and lower loudspeaker arrays. Further experiments are necessary to determine the largest acceptable vertical spacing between line arrays.

10.2 Future Listening Test Improvements

A crucial objective in further listening tests is to reduce overall localization error in order to produce clearer results. It is apparent that the lack of visual association target disturbs the localization process. This could be improved if listeners were given a short training session to get past the initial confusion involved with blind localization. The error may also have been due to the unfamiliar signal content; Start's original WFS experiments showed that subjects localized speech content more accurately than noise and musical content (Start 1997). The room acoustics of the test environment may have contributed adversely; although the "semi-anechoic" testing room is acoustically treated, it is unknown how effective the treatment is or how it compares to a true anechoic chamber. The room's small size could have contributed to error; the loudspeaker array had close walls at the top and sides. The loudspeaker support apparatus was also less than ideal. The music stands used to suspend the bottom row of loudspeakers could have introduced amplitude or diffraction errors. Future listening tests should also be done with the curtain positioned close in front of the array to boost the effective curtain area.

10.3 Future Research

Work on WFS Designer is ongoing, and outside contributions to the project are welcomed and encouraged. WFS Designer could be extended to support a performance playback system with dynamic virtual source motion paths.

There are many opportunities for future research applications. These include acoustic beam steering, active listening room compensation (Corteel, Nicol 2003; Spors et al. 2003; Fuster et al. 2005; Corteel 2006), active noise control (Kuntz, Rabenstein 1999), room acoustic simulation, direct sound reinforcement (Start 1997), ambisonics, and other spatial audio reproduction techniques. In recent years, listening room compensation has been one of the most active areas of wave field synthesis. Our array is well-suited to investigation of listening room compensation and active noise control because of its modularity and ability to cover a large perimeter of the sound field. The array could be augmented with an array of microphones to enable real-time sound field control. Finally, the array supports creative electronic music performance applications and interactive sound installations, as described by Baalman (2004).

LIST OF REFERENCES

Ahrens, J., and S. Spors. 2008. Reproduction of a plane-wave using planar and linear arrays of loudspeakers. *2008 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP).*

Ahrens, J., and S. Spors. 2009. Sound Field Reproduction Employing Non-Omnidirectional Loudspeakers. *126th AES Convention*.

Ahrens, J. 2010. Wave Field Synthesis Employing Linear and Circular Distributions of Secondary Sources. *129th AES Convention*.

Allen, J., D. Berkley, and M. Hill. 1979. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*: 943-950.

Baalman, M., and D. Plewe. 2004. WONDER–a software interface for the application of wave field synthesis in electronic music and interactive sound installations. *Proc. ICMC, Miami, 2004*.

Bacivarov, I., P. Bazzana, M. Beckinger, J. Ceng, A. Franck, W. Haid, K. Huang, et al. 2007. SHAPES-A Scalable Parallel HW/SW Architecture Applied to Wave Field Synthesis. *Journal of the Audio Engineering Society* 4, no. viii: 1-13.

Bencina, R., and P. Burk. 2004. PortAudio – an Open Source Cross Platform Audio API. *Proc. 2001 Intl. Computer Music Conf.(ICMC-01)*.

Berkhout, A.J. 1988. A Holographic Approach to Acoustic Control. *Journal of the Audio Engineering Society* 36, December 1988, pp. 977–995.

Berkhout, A. J., D. De Vries, P. Vogel. 1993. Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America* 93, no. 5: 2764–2778.

Boone, MM. 2004. Multi-actuator panels (maps) as loudspeaker arrays for wave field synthesis. *Journal of the Audio Engineering Society* 52, no. 7-8: 712–723.

Boone, MM, and E. Verheijen. 1995. Spatial sound-field reproduction by wave-field synthesis. *Journal of the Audio Engineering*.

Brix, S., F. Melchior, A. Partzsch, and C. Sladeczek. 2010. Design and Implementation of an Interactive Room Simulation for Wave Field Synthesis. *Journal of the Audio Engineering Society*.

Caulkins, T., and O. Warusfel. 2006. Characterization of the reverberant sound field emitted by a wave field synthesis driven loudspeaker array. In *120th AES Convention*, 1-11.

Corteel, E. 2006. Equalization in an extended area using multichannel inversion and wave field synthesis. *Journal of the audio Engineering Society* 54, no. 12: 1140–1161.

Corteel, E. 2006. On the use of irregularly spaced loudspeaker arrays for wave field synthesis, potential impact on spatial aliasing frequency. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx'06)*, 209–214.

Corteel, E., and T. Caulkins. 2004. Sound Scene Creation and Manipulation using Wave Field Synthesis. *Development*.

Corteel, E., and R. Nicol. 2003. Listening room compensation for wave field synthesis. What can be done? *23rd AES International Conference*.

De Castro Lopo, E. 2005. Libsndfile. Accessed February 20, 2011. http://www.mega-nerd.com/libsndfile.

Densil Cabrera, A., B. Daeup Jeong, C. Hyun Jeong Kwak, and D. Ji-young Kim. 2005. Auditory Room Size Perception for Modeled and Measured Rooms. In *The 2005 Congress and Exposition on Noise Control Engineering*.

Escolano, J., S. Bleda, B. Pueo, and J.J. López. 2004. Wave field synthesis simulation by means of finite-difference time-domain technique. *Proceedings of 12th European Signal Processing Conference*: 1777-1780.

Franck, A., K. Brandenburg, and U. Richter. 2008. Efficient Delay Interpolation for Wave Field Synthesis. *Journal of the Audio Engineering Society* i, no. 1.

Frigo, M. and S. G. Johnson, 2005. "The Design and Implementation of FFTW3," Proceedings of the IEEE 93 (2), 216–231 (2005). Invited paper, Special Issue on Program Generation, Optimization, and Platform Adaptation. http://www.fftw.org

Fuster, L., J. Lopez, A. Gonzalez, and P. Zuccarello. 2005. Room compensation using multichannel inverse filters for wave field synthesis systems. *118th AES Convention*.

Gauthier, P.A., and A Berry. 2007. Adaptive Wave Field Synthesis for Sound Field Reproduction: Theory, Experiments, and Future Perspectives. *Journal of the Audio Engineering Society*: 1-22.

Geier, M., J. Ahrens, A. Mohl, S. Spors, J. Loh, and K. Bredies. 2007. The Soundscape Renderer: A versatile software framework for spatial audio reproduction. *WFS Workshop*.

Geier, M., H. Wierstorf, J. Ahrens, I. Wechsung, A. Raake, and S. Spors. 2010. Perceptual evaluation of focused sources in wave field synthesis. In *128th Convention of the Audio Engineering Society*.

Glasgal, R., and R. Miller. 2006. True-to-Life Sound Reproduction using Recursive Ambiophonic Crosstalk Elimination (RACE).

IOSONO. “Spatial Audio Workstation 2.0 Upmix.” Accessed May 9, 2011. http://www.iosono-sound.com/products-and-services/spatial-audio-workstation-20-upmix/.

Kapralos, B., M. R. Jenkin, and E. Milios. 2008. Virtual Audio Systems. *Presence: Teleoperators and Virtual Environments* 17, no. 6 (December): 527-549.

Kerber, S., H. Wittek, H. Fastl, and G. Theile. Experimental investigations into the distance perception of nearby sound sources: Real vs. WFS virtual nearby sources. *Companion paper, CFA/DAGA*, no. Table 1: 2-3.

Kim, Y.H., and S.T. Ahn. 2001. A reverberation model based on objective parameters of subjective perception. *Journal of the Audio Engineering Society* 49, no. 9: 786–794.

Kolund, M., C. Faller, and M. Vetterli. 2009. Sound Field Reconstruction: An Improved Approach for Wave Field Synthesis. *126th AES Convention*.

Kuntz, A., and R. Rabenstein. 1999. An approach to global noise control by wave field synthesis. In *European Signal Processing Conference (EUSIPCO)*, 1999-2002.

Lopez, J., S. Bleda, and J. Escolano. 2005. A sub-band approach to wave field synthesis rendering. *Journal of the Audio Engineering Society*.

Melchior, F., and S. Spors. 2010. Spatial Audio Reproduction: From Theory to Production. In *128th AES Convention*.

Molino, J. 1973. Perceiving the range of a sound source when the direction is known. *The Journal of the Acoustical Society of America* 53, no. 5 (May): 1301-4.

Nielsen, S.H. 1993. Auditory distance perception in different rooms. *Journal of the Audio Engineering Society* 41: 755–755.

Oellers, Helmut. 2011. “Wave Field Synthesis and Holophony.” Accessed February 20, 2011. http://www.syntheticwave.de.

Ono, K., S. Komiyama, and K. Nakabayashi. 1997. A Method of Reproducing Concert Hall Sounds by Loudspeaker Walls. *Journal of the Audio Engineering Society*, no. 11.

Park, J. 2003. The effects of early decay time on auditory depth in the virtual audio environment. *Journal of the Audio Engineering Society*.

Petrausch, S., S. Spors, and R. Rabenstein. 2005. Simulation and visualization of room compensation for wave field synthesis with the functional transformation method. *119th AES Convention*.

Pulkki, V. 1997. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society* 45, no. 6.

Rabenstein, R., and S. Spors. 2006. Evaluation of the Circular Harmonics Decomposition for WDAF-based Active Listening Room Compensation. *Journal of the Audio Engineering Society*, no. 2: 1-16.

Rumsey, F. 2006. Spatial audio and sensory evaluation techniques – context , history and aims. *Spatial Audio & Sensory Evaluation Techniques*.

Salvador, C. 2010. Discrete Driving Functions for Horizontal and Higher Order Ambisonics. *Journal of the Audio Engineering Society*: 1-14.

Santarelli, S., V. Best, and B. Shinn-Cunningham. 2007. Simulating distance cues in virtual reverberant environments. *Hearing Research*, no. September.

Shinn-Cunningham, B. 2000. Distance cues for virtual auditory space. In *Proceedings of the IEEE-PCM*, 227–230.

Sonke, J.J., D. de Vries, and J. Labeeuw. 1998. Variable acoustics by wave field synthesis: A closer look at amplitude effects. *Journal of the Audio Engineering Society*.

Spors, S., and J. Ahrens. 2010. Reproduction of focused sources by the spectral division method. *2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*.

Spors, S., A. Kuntz, and R. Rabenstein. 2003. An approach to listening room compensation with wave field synthesis. *AES 24th International Conference on*: 49-52.

Spors, S., and R. Rabenstein. 2006. Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis. In *120th AES Convention*, 1-14.

Spors, S., R. Rabenstein, and J. Ahrens. 2008. The Theory of Wave Field Synthesis Revisited. *124th AES Convention*.

Spors, S., M. Renk, and R. Rabenstein. 2005. Limiting effects of active room compensation using wave field synthesis. *118th AES Convention*.

Spors, S., H. Teutsch, and R. Rabenstein. 2002. High-quality acoustic rendering with wave field synthesis. In *Vision, Modeling, and Visualization*, 101–108.

Spors, S., H. Wierstorf, M. Geier, and J. Ahrens. 2009. Physical and Perceptual Properties of Focused Sources in Wave Field Synthesis. *127th AES Convention*.

Start, E. 1997. Direct Sound Enhancement By Wave Field Synthesis. PhD dissertation, Delft University of Technology.

Steinberg, J. C., W. B. Snow. 1934. Auditory Perspective – Physical Factors. In *Transactions of the American Institute of Electrical Engineers* 53, no. 1 (Jan): 12-17.

Theile, G., and T. Neher. 1980. On the Localisation in the Superimposed Soundfield. PhD dissertation, Berlin Institute of Technology.

Theile, G., and H. Wittek. 2007. Wave field synthesis - a promising spatial audio rendering concept. *Journal of the Institute of Image Information and Television Engineers*.

Vries, D. de. 1994. The Wave Field Synthesis Concept Applied to Generation of Reflections and Reverberation. *96th AES Convention*.

Wittek, H. 2003. Perception of spatially synthesized sound fields–Literature review about WFS. Technical report, http://www. hauptmikrofon.de/wittek.htm.

Wittek, H. 2007. Perceptual differences between wavefield synthesis and stereophony. PhD thesis, Department of Music and Sound Recording School of Arts, Communication and Humanities, University of Surrey, 2007.

Wittek, H., S. Kerber, F. Rumsey, and G. Theile. 2004. Spatial perception in Wave Field Synthesis rendered sound fields: Distance of real and virtual nearby sources. In *116th AES Convention*, 1-20.

Yamada, T. 2008. Wave Field Synthesis Research [Video] Retrieved May 9, 2011, from http://www.youtube.com/watch?v=_je7qaJsttU.

Zahorik, P. 2002. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America* 111, no. 4: 1832.

Zahorik, P. 2002. Auditory display of sound source distance. In *Proceedings of the International Conference on Auditory Displays*, 326–332.