UNIVERSITAT POLITÈCNICA DE VALÈNCIA

_ TELECOM ESCUELA TÉCNICA VLC SUPERIOR DE UPV INGENIEROS DE TELECOMUNICACIÓN

Máster Universitario en Ingeniería de Telecomunicación

Performing active noise control with wave field synthesis

Author:
Rubén Chuliá Mena

Tutor:
María de Diego Antón

Co-Tutor:
Miguel Ferrer Contreras

Valencia, September 10, 2018

Contents

Resum

Wave Field Synthesis (WFS) es un mètode proposat per primera vegada en la dècada dels 90 que, mitjançant una agrupació d'altaveus, permet replicar el camp acústic que una hipotètica font de soroll produiria. El principal avantatge enfront d'altres tècniques de reproducció multicanal de so es que WFS funciona en una àrea mes ampla, mentres que la resta estan limitades a un àrea més menuda, també denominada "sweetspot". Una de les aplicacions de WFS estudiades recentment es el Control Actiu de Soroll o Active Noise Control (ANC). El plantejament es identificar la posició y característiques d'una font de so y usar eixa informació per a sintetitzar un camp acústic que interfereixi destructivament amb el camp acústic original. Ens centrem en la viabilitat de realitzar ANC basat en WFS amb una agrupació de 96 altaveus distribuïts en forma d'octàgon en una sala reverberant. Estudiem diferents fonts d'error mitjançant simulació en Matlab. Partim de condicions ideals y després canviem alguns paràmetres amb l'objectiu d comprovar com influencien el funcionament del sistema. També realitzem mesuraments experimentals. Concloem que l'absència de condicions de propagació en espai lliure suposa la limitació principal per a aquest sistema.

Resumen

Wave Field Synthesis (WFS) es un método propuesto por primera vez en la década 90 que, mediante una agrupación de altavoces, permite replicar el campo acústico que una fuente de sonido hipotética produciría. La principal ventaja sobre otras técnicas de reproducción multicanal de sonido es que WFS funciona en un área más amplia, mientras que el resto están limitados a un área pequeña, también llamada "sweetspot". Una de las aplicaciones de WFS estudiadas recientemente es el Control Activo de Ruido o Active Noise Control (ANC). El planteamiento es identificar la posición y características de una fuente de ruido y usar esa información para sintetizar un campo acústico que interfiera destructivamente con el campo original. Nos centramos en la viabilidad de realizar ANC basado en WFS con una agrupación de 96 altavoces distribuidos en forma de octágono en una sala reverberante. Estudiamos diferentes fuentes de error mediante simulaciones en Matlab. Partimos de condiciones ideales, y luego algunos parámetros son cambiados con el objetivo de comprobar cómo influencian el funcionamiento del sistema. También realizamos mediciones experimentales. Concluimos que la ausencia de condiciones de propagación en espacio libre supone la limitación principal para este sistema.

Abstract

Wave Field Synthesis (WFS) was first proposed in the 1990s as a method that, by means of an array of loudspeakers, allows to generate the acoustic field that some hypothetical sound source would produce. The main advantage over other multichannel sound reproduction techniques is that WFS works over a wider area, whereas the rest are limited to a small area, often called sweetspot. One of the recently studied applications is Active Noise Control (ANC). The idea is to identify the position and characteristics of a noise source and use that information to synthesize an acoustic field that will interfere destructively with the original one. We focus on the viability of performing WFS based ANC with an octagon shaped 96-element loudspeaker array set up in a reverberant room. A series of Matlab simulations are done to study the possible sources of errors. Idealized conditions are first assumed, and then some parameters are changed to test how they influence performance. Then, some measures are carried out in a real environment. We conclude that non free-space conditions are the main limitation to the feasibility of this system.

List of Figures

Chapter 1

Introduction and State of the Art. What is wave field synthesis and active control noise?

1.1. Wave Field Synthesis

Wave Field Synthesis (WFS) is a method that, by means of an array of loudspeakers (large number of small and closely spaced loudspeakers) reproducing the proper audio signals, generates the acoustic wave field that a hypothetical source of sound would produce (virtual primary source). In other words, it is a way of accurately replicating temporal, spectral and spatial properties of a sound field. For example, in a room where one of this arrays is set up, a person situated in any point of the room could hear the voice of a person moving through the room, as if someone that is not there was actually talking and walking [1].

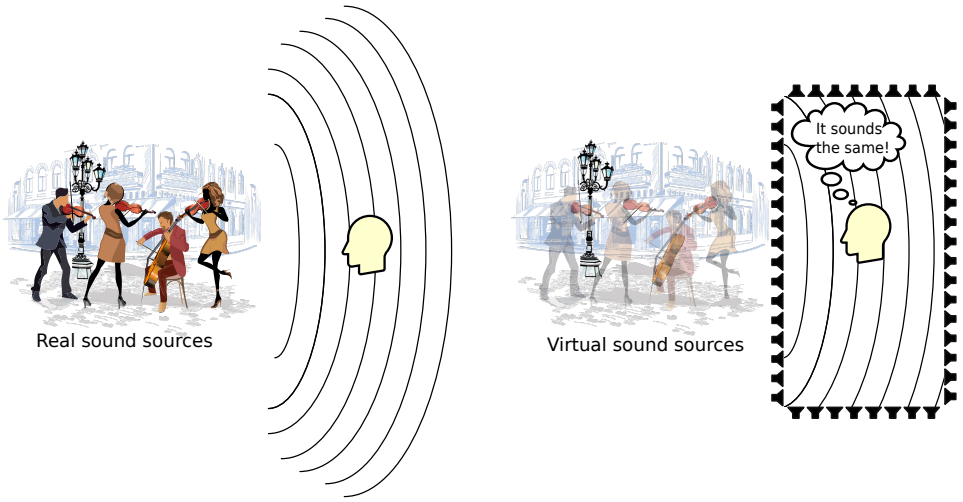


Figure 1.1: WFS explanation [2]

WFS takes advantage of a physical principle applied to wave fields (such as acoustic waves) expressed in Kirchhoff's integral. Before getting into the details of mathematical

expressions, let's just say that Kirchhoff's integral states that in a homogeneous wave propagation media, in any source-free volume V (fictive) delimited by a surface S, the wave field at any point in that space can be calculated if the wave field and its gradient on the surface are known. In other words, if we want to know the sound that one can hear at any point inside V, we just have to measure the acoustic pressure and its gradient on S.

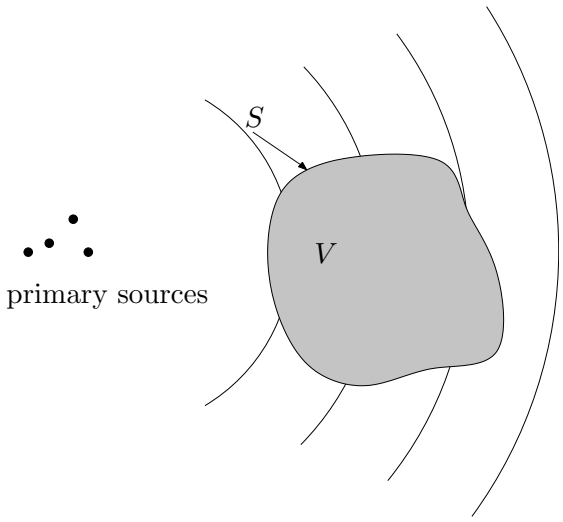


Figure 1.2: Kirchhoff integral [1]

This fact can be turned around so it is useful, not only for knowing the field, but to replicate it. If, in another time and place, we manage to generate a surface acoustic field identical to the measured one, the wave field inside the volume will be the same as previous one. For example, if we want that inside V one can hear the sound of a string quartet (located outside V) playing the Pachelbel's Canon, we have two options. On the one hand, we hire a string quartet, make them play and the problem is solved. On the other hand, we have a more interesting solution. We record the wave field on each point of S when the musicians play, then we build some audio reproducing system that can replicate it, and place a person inside.

How to build that system is the issue here. Kirchhoff's integral does actually provide some answers. It states that it could be done with a surface continuous distribution of an infinite number of monopole and dipole sources, called secondary sources. This means that if, at each point of S, there was one monopole and one dipole infinitesimal sources driven by the right signals, the replication of the virtual source field would be perfect inside V, and moreover, the field would be zero outside.

Of course, Kirchhoff-Helmholtz integral can not simply be put into practice due to obvious technical reasons. It is not practical (or even possible) to build a hollow volume with tons of tiny loudspeakers on the surface and place a listener inside. But thankfully, in a real scenario where a finite amount of real loudspeakers are used, in realizable and simple spatial distributions, with the presence of reflective objects, diffractions, where the air is not an ideal transport media for sound propagation (which is not, since it presents air damping effects), etc., we can still aim for some degree of accuracy. A common practical case is one where loudspeakers distributed as a straight line (not a closed surface) are used to synthesize a field only in the horizontal plane, and below a certain frequency that is inversely proportional by the separation between loudspeakers

(aliasing frequency). Simplifications such as that, are necessary to implement a feasible practical system. The price to pay is the limitation of the performance in terms of accuracy, bandwidth, spatial range where it works, etc. However, it still can provide good results, depending on the requirements of the system, which are usually defined by the human hearing capabilities.

Historically, WFS theory was developed in Delft University and first presented to the public in 1989 [3]. Since then, it has come a long way of development. During the 1990s, it was mainly a topic of research. Most of the research was focused in performing high fidelity sound reproduction to create a true immersive sound experience. High fidelity reproduction systems have been an important topic for many decades. Our auditory system plays a major role in how we experience our environment. It is continuously locating objects in distance and direction. Even in situations where visual cues are dominant, our ears help us analyse the environment and create the feeling of immersion. All stereo techniques (two-channel stereo, quadraphony, 5.1 and 7.1 surround sound) used in cinema or theatres share the shortcoming that only the listeners located in a very limited area, usually called sweet spot, experience good spatial immersion. In general, the more precise the spatial scene is, the smaller the sweet spot becomes. But WFS is able to synthesize a replica of the sound field over the whole listening area, and that is its biggest advantage, and the main motivation for the research [1]. Various cases of successful implementation proved that WFS could actually work to some degree at least on simple scenarios [4–6].

It was not until the 21st century when commercial applications were available [1]: in 2003, the first cinema based on WFS started daily operation in Ilmenau (Germany), the first WFS system in a sound stage was installed in 2004 in Studio City (California, US), and since 2008 a large WFS installation is at the Chinese Theatre in Los Angeles (US) (Figure 1.3). In living performance, it has been used to improve spatial coherence between audio and the visual part (Bregenz Festival, Austria), or to improve speech intelligibility (auditorium of the Technical University in Berlin [7]). Other application areas are theme parks, virtual reality, and even the music reproduction system of the car Audi Q7. Potential applications not fully implemented yet are adjustment of multi-purpose hall room acoustics and elimination of noise disturbance and unwanted echoes.

Figure 1.3: Chinese Theatre WFS installation by *IOSONO* (Los Angeles, US)

Some of the implementations are really complex and use hundreds of loudspeakers, especially in big installations, but usually they are simpler and very far from the ideal scenario that Kirchhoff's integral describes. This has the drawback that performance gets deteriorated. Despite that, since quality subjective experience is the goal in immersive sound reproduction, the criteria to evaluate whether the performance is actually good or bad are psychoacoustics. If the listener's experience is good, then it is considered

a good system and there's no point in aiming at a better performance that would not be appreciated by the listener. When psychoacoustic mechanisms for perceiving source position and width and spaciousness are considered, the necessary number of loudspeakers can be reduced drastically while maintaining good spatial listening experience and reducing computational demands. The precision of the physical replication and thus the amount of data to be processed can be reduced without audible effects [7]. For example, localization in the horizontal plane is much better than the perception of elevation, so using just a horizontal line array of loudspeakers does not influence that much the audio-visual experience [1]. Another approach is directional audio coding (DirAC): signals are separated into directional and diffuse components [7].

Despite all improvements over the years, there are remaining obstacles. WFS is still at a stage of research and development, and although several approaches already enable acoustic control to a certain degree, there is a long way to go. That's why nowadays WFS is usually employed as a complement to the main sound reproduction system, but not as a fully working stand-alone system.

Current research and development address mainly issues related to accessibility and improvement of performance. One important topic is the creation of easy accessible software and formats to simulate, create, control, store and play WFS content. Due to the lack of standardized loudspeaker configurations, object-based source material is necessary: combination of audio tracks with dynamic metadata that describes source position, trajectories, orientation and radiation characteristics and information on reflections and reverberation. Although several formats have been proposed, no standardized wave field synthesis format has established yet.

In order to derive the driving signals of secondary sources, the pressure field has to be known. Most approaches use mathematical models to estimate the field, because the problem of how to measure (record) a complete sound field still presents lots of limitation [8].

Another topic is the application of multiactuator panels (MAPs) so installations do not harm interior decoration due to their discreet appearance and, moreover, can be arranged continuously, preventing aliasing effects.

So far research has focused on the synthesis of virtual static monopole sources and plane waves. But more advanced features are beginning to be studied. For fast moving sources, the Doppler effect is important for an authentic sound experience. Radiation characteristics of musical instruments are complex, far from behaving as a monopole. There are also attempts to include room modes, early reflections and late reverberations [9]. Finally, the potential of psychoacoustics is regarded as very promising by many researchers because it can help to overcome current limited acoustic control and restrictions [7].

1.1.1. Active Noise Control

A problem related to sound, is the cancelation of noise. Active Noise Control (ANC) refers to a group of techniques that aim at reducing the effect of acoustic noise sources by means of an array of loudspeakers that generate a sound wave that interferes de-

structively with the noise wave field and, thus, cancels it. It past decades it has become a growing field of research, since passive methods are not as effective in cancelling low frequency noise [10].

ANC has had success in cases where the listening area is very small, e.g., inside a head phone, or around a listener with restricted head movement. However, for large spaces where listeners are allowed to move freely, the problem becomes much complicated. Traditional approaches require a huge number of sensors and sources distributed within the area of interest. Moreover, this high number of sensors would constitute a highly overdetermined multiple-input multiple-output system, which causes bad convergence of adaptive algorithms and very large loudspeaker driving signals [11]. Besides, the classical ANC adaptive filtering techniques (e.g., FxLMS and extensions) work well for minimizing the mean of some distortion measure of stationary Gaussian noise, but not for short duration noise because convergence is not achieved [12].

There have been proposals of the use of WFS to perform ANC as a solution to previous problems [11–14]. The use of WFS allows to control the sound field by using a distribution of sources and sensors only on the boundary of the listening space, and listeners are not restricted in their movement and no headphones or object tracking equipment is required [11].

1.2. Objectives

The objective of this thesis is the study, by means of simulations, of the possibilities of performing active noise control with WFS at the audio laboratory available in the Audio and Communications Signal Processing Group (GTAC) of the Institute of Telecommunications and Multimedia Applications (iTEAM) of the Universitat Politècnica de València (UPV).

The simulations are computed on the software programming platform *Matlab*. Simple scenarios are contemplated at first, and then complexity is increased in order to gain insight into the case at hand so we can provide a survey of the requirements, limitations and difficulties that may arise during the implementation of a true WFS based ANC system.

Chapter 2

Theoretical Basis. From Kirchhoff's Integral to Arrays of Loudspeakers.

In this chapter, the mathematical basis of Wave Field Synthesis (WFS) is explained. First, the monochromatic wave equation in three dimensions and Green's Theorem are combined to prove that the wave field in a source-free space volume is totally determined by the wave field on its surface. Specifically, Kirchhoff integral and Rayleigh integral are the expressions that allow us to calculate the wave field at any point inside the volume using just its surface information, but Rayleigh integral is simpler and, hence, more convenient.

The main idea of WFS is to use loudspeakers to generate a surface acoustic wave field identical to the one that would be created by a virtual sound source. Since the wave field inside the volume depends only on the surface one, it will be the same as the acoustic field that would be generated by the virtual source. The accuracy of WFS depends on how well we are able to replicate the surface wave field. We will see that we would need an infinite amount of monopole and dipole infinitesimal sources to generate an exact acoustic field, but it is obvious that in any real situation we can only use a finite number of loudspeakers that are not infinitesimal, nor they present ideal monopole or dipole radiation patterns. We will model a WFS system with discrete punctual sources and take a look at the problems and inaccuracies that derive from it.

2.1. Kirchhoff-Helmholtz Integral

From three-dimensional wave equation and Green's Theorem, the Kirchhoff-Helmholtz Integral is deduced [15] [5]:

$$P(\mathbf{x}) = \frac{1}{4\pi} \int_S \left(P(\mathbf{x}_s) \frac{\partial G(\mathbf{x}_s|\mathbf{x})}{\partial \mathbf{n}} - G(\mathbf{x}_s|\mathbf{x}) \frac{\partial P(\mathbf{x}_s)}{\partial \mathbf{n}} \right) dS. \tag{2.1}$$

It expresses the pressure field $P(\mathbf{x})$ (particularized for a given frequency) inside

a free-source volume V bounded by the surface S, as a function of the pressure at S, and its directional derivative in the direction of $\mathbf{n}$, which is the inward pointing normal vector of S (Figure 2.1). $G(\mathbf{x}|\mathbf{x}_0)$ is called Green's function, and should obey the inhomogeneous wave equation for a source at position $\mathbf{x}_0$ ($\nabla^2 G - k^2 G = -4\pi\delta(\mathbf{x}-\mathbf{x}_0)$, where $\nabla^2 G = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} + \frac{\partial^2 G}{\partial z^2}$ is the laplacian of G).

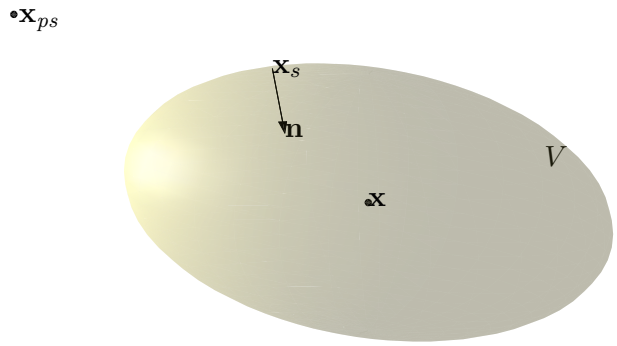


Figure 2.1: Scheme of Kirchhoff integral scenario

The general form of $G(\mathbf{x}|\mathbf{x}_0)$ is:

$$G(\mathbf{x}|\mathbf{x}_0) = \frac{e^{-jk\|\mathbf{x}-\mathbf{x}_0\|}}{\|\mathbf{x}-\mathbf{x}_0\|} + F(\mathbf{x}|\mathbf{x}_0), \tag{2.2}$$

where $F(\mathbf{x}|\mathbf{x}_0)$ is any function that satisfies the Helmholtz equation $\nabla^2 F - k^2 F = 0$.

If $F = 0$:

$$P(\mathbf{x}) = \frac{1}{4\pi}\int_S \left(P(\mathbf{x}_s)\frac{e^{-jk\|\mathbf{x}-\mathbf{x}_s\|}}{\|\mathbf{x}-\mathbf{x}_s\|}\left(jk + \frac{1}{\|\mathbf{x}-\mathbf{x}_s\|}\right)\cos\alpha - \frac{e^{-jk\|\mathbf{x}-\mathbf{x}_s\|}}{\|\mathbf{x}-\mathbf{x}_s\|}\frac{\partial P(\mathbf{x}_s)}{\partial \mathbf{n}}\right) dS, \tag{2.3}$$

where $\alpha = \left\langle \frac{\mathbf{x}-\mathbf{x}_s}{\|\mathbf{x}-\mathbf{x}_s\|}, \mathbf{n} \right\rangle$ is the angle between the inward normal vector and the vector that passes through the point at the surface $\mathbf{x}_s$ and $\mathbf{x}$.

It can be interpreted as if the field inside V was the result of the field generated by infinitesimal sources distributed over S. The first term of the integral represents a dipole source distribution driven by the pressure at the surface. The dipoles have the inward normal vector as broadside direction. The second term of the integral represents a monopole source distribution driven by the directional derivative of the pressure at the surface. The result of the integral outside V is 0. The original sources that generate the field are called primary sources. The surface monopole and dipole source distributions that can emulate the field inside the volume will be called secondary sources.

2.2. Rayleigh I and II Integrals

Kirchhoff-Helmholtz integral can be simplified at the cost of a fixed surface geometry and a non-zero field outside the volume, but those limitations are of little importance in practice for WFS. The simplified integrals, known as Rayleigh I and II integrals, are found by choosing a particular surface of integration and a suitable function F.

The new volume is a hemisphere. The surface is then constituted by a flat circle S_1 with radius R and the spherical surface S_2. All primary sources will be located behind the flat circle (Figure 2.2). When $R \to \infty$, the Sommerfeld condition is satisfied, which

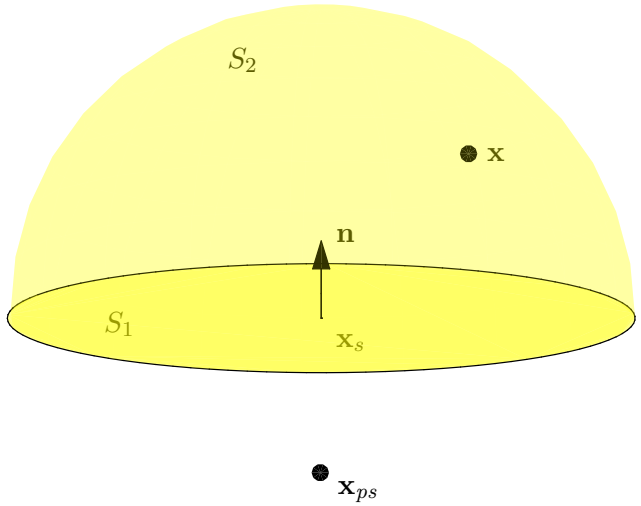


Figure 2.2: Scheme of Rayleigh integral scenario

means that the integral over S_2 becomes 0 [5].

If $F(\mathbf{x}_s|\mathbf{x}) = \frac{e^{-jk\|\mathbf{x}_s-\mathbf{x}_m\|}}{\|\mathbf{x}_s-\mathbf{x}_m\|}$, being $\mathbf{x}_m$ the mirrored image of $\mathbf{x}$ in the plane S_1, then the directional derivative of G becomes 0 at S_1, and Equation 2.1 transforms to:

$$P(\mathbf{x}) = \frac{-1}{2\pi} \int_{S_1} \frac{e^{-jk\|\mathbf{x}_s-\mathbf{x}\|}}{\|\mathbf{x}_s-\mathbf{x}\|} \frac{\partial P(\mathbf{x}_s)}{\partial \mathbf{n}} dS. \tag{2.4}$$

Previous equation is called Raileigh I integral and states that a secondary planar monopole source distribution can synthesize on one side of the plane S_1 the field of a primary source distribution located at the other side. The monopole sources are driven by two times the directional derivative of the pressure at the plane, in its perpendicular direction.

The function F can also be chosen to remove the monopole source distributions: $F = -\frac{e^{-jk\|\mathbf{x}_s-\mathbf{x}_m\|}}{\|\mathbf{x}_s-\mathbf{x}_m\|}$. In this case, Equation 2.1 transforms to the Raileigh II integral:

$$P(\mathbf{x}) = \frac{1}{2\pi} \int_{S_1} P(\mathbf{x}_s) \frac{e^{-jk\|\mathbf{x}_s-\mathbf{x}\|}}{\|\mathbf{x}_s-\mathbf{x}\|} \left(jk + \frac{1}{\|\mathbf{x}_s-\mathbf{x}\|} \right) \cos\alpha dS. \tag{2.5}$$

It presents a similar scenario as Raileigh I integral, but instead of monopole secondary sources, it uses dipole sources driven by two times the pressure at plane S_1.

2.3. Dimensionality reduction: Kirchhoff-Helmholtz 2.5D Integral

For practical reasons, it is convenient to reduce the surface S to a closed line L contained on a plane, that would correctly synthesize the field on that plane, as it is proposed in [6] and developed in [4] and, only for the Rayleigh integrals, in [5]. The way to do it is the next.

Let's assume that all primary sources and the positions where we want to synthesize their field are all located in the same plane P. We also assume that S is a surface formed by all the lines that are parallel to the normal vector of P, $\mathbf{n}_P$, and intersect with a closed 2D curved L contained on P. In other words, it is like a tube infinitely long that extends in the $\mathbf{n}_P$ direction and it's section (the intersection of S and the plane P) is L, as shows Figure 2.3.

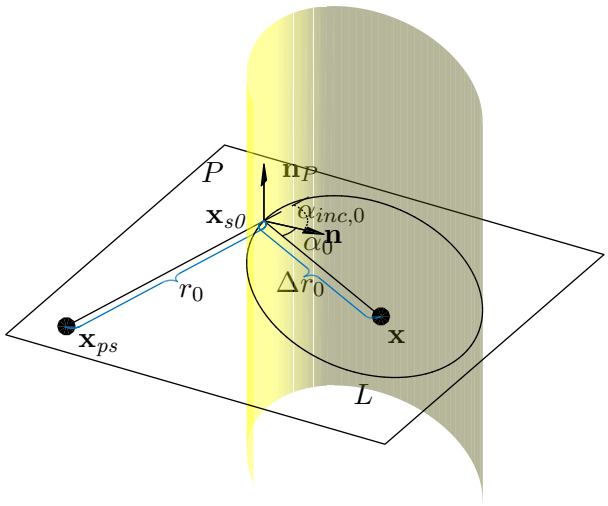


Figure 2.3: Scheme of Kirchhoff integral dimensionality reduction

In order to perform the dimensionality reduction, we calculate the contribution of each of lines that form S, one for each of the points $\mathbf{x}_{s0}$ that constitute the curve L.

$$P(\mathbf{x}) = \int_L \left(I_1(\mathbf{x}_{s0}, \mathbf{x}) + I_2(\mathbf{x}_{s0}, \mathbf{x})\right) d\mathbf{x}_{s0} \tag{2.6}$$

$$I_1(\mathbf{x}_{s0}, \mathbf{x}) = \frac{-1}{4\pi} \int_{-\infty}^{\infty} G\left(\mathbf{x}_{s0} + h\mathbf{n}_P \,\middle|\, \mathbf{x}\right) \frac{\partial P(\mathbf{x}_{s0} + h\mathbf{n}_P)}{\partial \mathbf{n}} dh \tag{2.7}$$

$$I_2(\mathbf{x}_{s0}, \mathbf{x}) = \frac{1}{4\pi} \int_{-\infty}^{\infty} P(\mathbf{x}_{s0} + h\mathbf{n}_P) \frac{\partial G\left(\mathbf{x}_{s0} + h\mathbf{n}_P \,\middle|\, \mathbf{x}\right)}{\partial \mathbf{n}} dh. \tag{2.8}$$

The field P is assumed to be produced by a punctual source:

$$P(\mathbf{x}) = P_{ps}(\mathbf{x}) = SD\left(\frac{\mathbf{x} - \mathbf{x}_{ps}}{\|\mathbf{x} - \mathbf{x}_{ps}\|}\right) \frac{e^{-jk\|\mathbf{x} - \mathbf{x}_{ps}\|}}{\|\mathbf{x} - \mathbf{x}_{ps}\|}, \tag{2.9}$$

where S is the signal strength of the source (complex scalar) and $D\left(\frac{\mathbf{x}-\mathbf{x}_{ps}}{\|\mathbf{x}-\mathbf{x}_{ps}\|}\right)$ is the directivity pattern of the source in the $\frac{\mathbf{x}-\mathbf{x}_{ps}}{\|\mathbf{x}-\mathbf{x}_{ps}\|}$ direction.

The approximated solution provided by the stationary phase method is ([4, Equation 3.17]):

$$I_1(\mathbf{x}_{s0}, \mathbf{x}) \approx \widetilde{I}_1(\mathbf{x}_{s0}, \mathbf{x}) = \frac{1}{2} SD\left(\frac{\mathbf{x}_{s0} - \mathbf{x}_{ps}}{\|\mathbf{x}_{s0} - \mathbf{x}_{ps}\|}\right) \cos\alpha_{inc,0} \frac{e^{-jkr_0}}{\sqrt{r_0}} \frac{e^{-jk\Delta r_0}}{\Delta r_0} \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}}, \tag{2.10}$$

where $r_0 = \|\mathbf{x}_{s0} - \mathbf{x}_{ps}\|$ is the distance between the primary source and the secondary source and $\Delta r_0 = \|\mathbf{x} - \mathbf{x}_{s0}\|$ is the distance between the secondary source and the receiving point, and $\alpha_{inc,0}$ is the angle between the inward normal vector $\mathbf{n}$ and the propagation direction from the primary to the source $\mathbf{x}_{s0} - \mathbf{x}_{ps}$. The result is valid for $kr_0 >> 1$.

Let's see that in previous equation there is the expression for the propagation of a monopole. So, we can express it as the strength of a secondary monopole source $Q_m(\mathbf{x}_{s0}, \mathbf{x})$ propagated to the receiving point $\mathbf{x}$:

$$\begin{aligned} \widetilde{I}_1(\mathbf{x}_{s0}, \mathbf{x}) &= Q_m(\mathbf{x}_{s0}, \mathbf{x}) \frac{e^{-jk\Delta r_0}}{\Delta r_0} \\ Q_m(\mathbf{x}_{s0}, \mathbf{x}) &= \frac{1}{2} SD\left(\frac{\mathbf{x}_{s0} - \mathbf{x}_{ps}}{\|\mathbf{x}_{s0} - \mathbf{x}_{ps}\|}\right) \cos\alpha_{inc,0} \frac{e^{-jkr_0}}{\sqrt{r_0}} \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}}. \end{aligned} \tag{2.11}$$

The same reasoning can be applied to Equation 2.8, and the result is [4, Equation 3.24]:

$$I_2(\mathbf{x}_{s0}) \approx \widetilde{I}_2(\mathbf{x}_{s0}) = \frac{1}{2} SD\left(\frac{\mathbf{x}_{s0} - \mathbf{x}_{ps}}{\|\mathbf{x}_{s0} - \mathbf{x}_{ps}\|}\right) \cos\alpha_0 \frac{e^{-jkr_0}}{\sqrt{r_0}} \frac{e^{-jk\Delta r_0}}{\Delta r_0} \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}}, \tag{2.12}$$

where α_0 is the angle between the inward normal vector $\mathbf{n}$ and the propagation direction from the secondary source to the receiving point $\mathbf{x} - \mathbf{x}_{s0}$. Hence, $\cos\alpha_0 = \left\langle \frac{\mathbf{x}-\mathbf{x}_{s0}}{\|\mathbf{x}-\mathbf{x}_{s0}\|}, \mathbf{n} \right\rangle$.

As with Equation 2.10, Equation 2.12 can also be expressed as the strength of a secondary source $Q_d(\mathbf{x}_{s0}, \mathbf{x})$ propagated to the receiving point, but this time the secondary

source is a dipole with the inward normal vector as broadside direction:

$$\begin{aligned} \widetilde{I}_2(\mathbf{x}_{s0}) &= Q_d(\mathbf{x}_{s0},\mathbf{x})\frac{e^{-jk\Delta r_0}}{\Delta r_0}\cos\alpha_0 \\ Q_d(\mathbf{x}_{s0},\mathbf{x}) &= \frac{1}{2}SD\left(\frac{\mathbf{x}_{s0}-\mathbf{x}_{ps}}{\|\mathbf{x}_{s0}-\mathbf{x}_{ps}\|}\right)\frac{e^{-jkr_0}}{\sqrt{r_0}}\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{\Delta r_0}{r_0+\Delta r_0}}. \end{aligned} \tag{2.13}$$

The synthesized field will be then:

$$\begin{aligned} P_{2.5D}(\mathbf{x}) = \int_{S_0}\left(\widetilde{I}_1(\mathbf{x}_{s0},\mathbf{x})+\widetilde{I}_2(\mathbf{x}_{s0},\mathbf{x})\right)\mathrm{d}\mathbf{x}_{s0} = \\ \int_{S_0}\left(Q_m(\mathbf{x}_{s0},\mathbf{x})\frac{e^{-jk\Delta r_0}}{\Delta r_0}+Q_d(\mathbf{x}_{s0},\mathbf{x})\frac{e^{-jk\Delta r_0}}{\Delta r_0}\cos\alpha_0\right)\mathrm{d}\mathbf{x}_{s0}. \end{aligned} \tag{2.14}$$

Equation 2.14 is known as Kirchhoff 2.5D integral, because it aims at synthesizing a wave field on a two dimensional plane, but using 3D wave propagation equations.

2.4. Rayleigh 2.5D I and II integrals

The application of previous dimensionality reduction to Rayleigh integrals is pretty straightforward. Some variables of the scenario get particularized. The integral over S gets substituted by an integral over a plane (S_1), and we assume that the plane of field synthesis P is orthogonal to S_1, so the curve L is actually an infinite straight line. Rayleigh I integral (Equation 2.4) uses a secondary monopole source distribution driven by the directional derivative of the pressure multiplied by 2, so the signal that feeds each monopole of L is the double of the one calculated for the 2.5D Kirchhoff integral. Rayleigh 2.5D I integral is:

$$P_I(\mathbf{x}) = \int_L Q_I(\mathbf{x}_{s0},\mathbf{x})\frac{e^{-jk\Delta r_0}}{\Delta r_0}\mathrm{d}\mathbf{x}_{s0}, \quad Q_I(\mathbf{x}_{s0},\mathbf{x}) = 2Q_m(\mathbf{x}_{s0},\mathbf{x}). \tag{2.15}$$

The same happens for Rayleigh II integral (Equation 2.5) with the dipole distribution. Rayleigh 2.5D II integral is:

$$P_{II}(\mathbf{x}) = \int_L Q_{II}(\mathbf{x}_{s0},\mathbf{x})\frac{e^{-jk\Delta r_0}}{\Delta r_0}\cos\alpha_0\mathrm{d}\mathbf{x}_{s0}, \quad Q_{II}(\mathbf{x}_{s0},\mathbf{x}) = 2Q_d(\mathbf{x}_{s0},\mathbf{x}). \tag{2.16}$$

Both integrals have the problem that the amplitude of the secondary source signals depends on the receiver position $\mathbf{x}$, which makes impossible to replicate the field of the primary source over the whole area simultaneously. However, one can replicate the field with high precision over a line parallel to the secondary source line and separated a distance d (the precision of the reconstruction will inevitably degrade as we move away from that line). It can be done by modifying the amplitude factor $g = \sqrt{\frac{\Delta r_0}{r_0+\Delta r_0}}$. Applying the stationary phase method, but now in the direction of the secondary source line L we would find that the main contribution to the pressure at a given point comes from the intersection of L and the line that goes from the primary source to the receiver

point . Taking advantage of this, we can change Δr_0 by d and r_0 by d_{ps} (distance between the primary source location $\mathbf{x}_{ps}$ and the closest point at L, Figure 2.4), so $g = \sqrt{\frac{d}{d+d_{ps}}}$, making it independent from the receiver point (implementable in real WFS systems) [5].

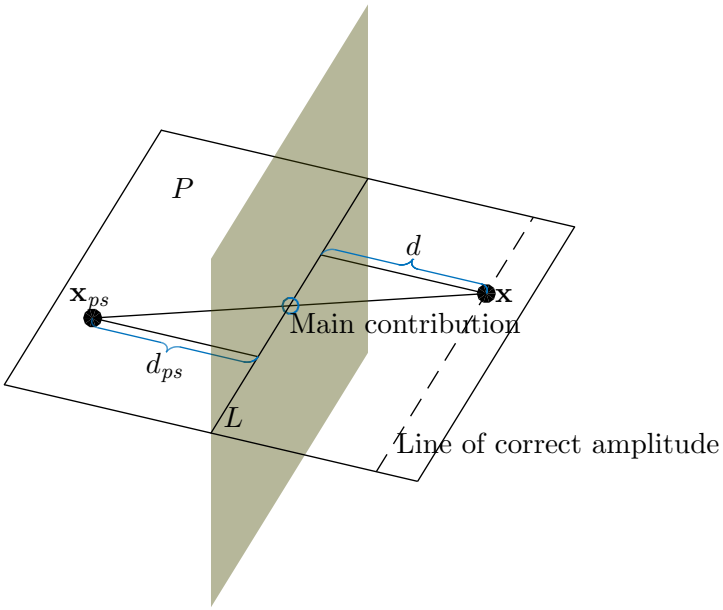


Figure 2.4: Scheme of Rayleigh integral dimensionality reduction

$$Q_I(\mathbf{x}_{s0}) = SD\left(\frac{\mathbf{x}_{s0}-\mathbf{x}_{ps}}{\|\mathbf{x}_{s0}-\mathbf{x}_{ps}\|}\right)\cos\alpha_{inc,0}\frac{e^{-jkr_0}}{\sqrt{r_0}}\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{d}{d+d_{ps}}}, \tag{2.17}$$

$$Q_{II}(\mathbf{x}_{s0}) = SD\left(\frac{\mathbf{x}_{s0}-\mathbf{x}_{ps}}{\|\mathbf{x}_{s0}-\mathbf{x}_{ps}\|}\right)\frac{e^{-jkr_0}}{\sqrt{r_0}}\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{d}{d+d_{ps}}}. \tag{2.18}$$

In conclussion, Rayleigh 2.5D I and II integrals (Equation 2.17 and Equation 2.18 respectively) allow to replicate the field by an infinite line distribution of monopole or dipole secondary sources, given the condition that all points where the field is replicated, as well as the primary source location and the secondary source line distribution (L), are all on the same plane, and that the secondary source line separates the reconstructed field region from the primary source.

2.5. Discretization and Spatial Aliasing

In practice, a continuous secondary source line is not realistic. An discrete linear secondary source array is a closer model to the practical cases where an array of loudspeakers is used. The issue with discretization is that aliasing effects appear. However, as it is exposed in [4], the synthesized wavefield will be exactly equal to the one with a continuous line source at those frequencies that respect next relation:

$$f < \frac{c}{2\Delta x}, \tag{2.19}$$

where Δx is the separation between contiguous loudspeakers.

This means that, for a sound signal with f_{max} as maximum frequency component, and with $\lambda_{min} = c/f_{max}$ as its corresponding wavelength, aliasing will be avoided as long as the separation between loudspeakers is smaller than half the wavelength: $\Delta x < \frac{\lambda_{min}}{2}$.

Discretization makes necessary to scale the feeding of sources by Δx:

$$P_I(x) = \Delta x \sum_n Q_I(\mathbf{x}_{s0n}) \frac{e^{-jk\Delta r_0}}{\Delta r_0} \tag{2.20}$$

$$P_{II}(x) = \Delta x \sum_n Q_{II}(\mathbf{x}_{s0n}) \frac{e^{-jk\Delta r_0}}{\Delta r_0} \cos\alpha_0. \tag{2.21}$$

Chapter 3

Analysis of potential sources of error

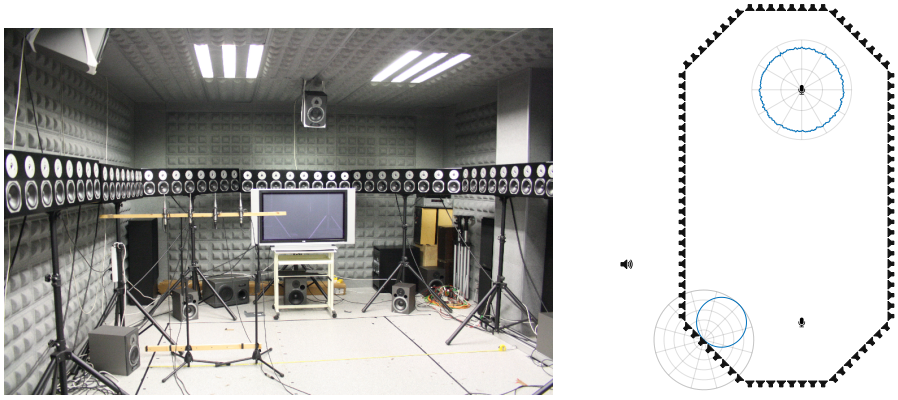
3.1. GTAC listening room modelling

In the GTAC listening room there is a 96-loudspeaker array distributed as an irregular octagon in the horizontal plane (Figure 3.1), 1.65m above the floor, with a separation between loudspeakers of $\Delta x = 0.18$m.

Many variables have an impact on the acoustic field generated by the loudspeakers: frequency dependent directivity of each individual loudspeaker, complex near-field radiation characteristics, non-linearities, reverberation of the chamber, diffraction, reflections on the floor (which is not recovered with absorbing material as the walls), etc. However, if we assume a simple model where the listening room is perfectly configured to emulate free-space conditions, and every loudspeaker is identical to the rest and behaves as an ideal monopole, then the similarities with the scenario presented by WFS theory become clear.

Under ideal conditions, the octagon can be interpreted as the closed curve of secondary sources L discretized with a step of $\Delta x = 0.18$m, so the aliasing frequency for a speed of sound $c = 340$m/s is $f_{alias} = 944.44$Hz. As we only count with monopole sources (loudspeakers with dipole characteristics are more difficult and expensive to manufacture), we should use the Rayleigh 2.5D I integral (Equation 2.15), but L should be an infinite line and not an octagon. Of course, an infinite array is not realizable in practice, so at some point we must truncate the array anyway. On the other hand, when dealing with a bent array, the amplitude factor $g = \sqrt{\frac{d}{d+|z_{ps}|}}$ that was calculated when L was a straight line might not be the best option any more.

The actual loudspeaker feeding signals that were used, were calculated applying formulas that were provided by the professors and are particularized for the specific geometry of the GTAC array. Specifically, they follow the Rayleigh 2.5D I Integral form (Equation 2.17), where the amplitude factor g is optimized for the middle line that



(a) Picture [16]

(b) Schematic of the loudspeaker array distribution

Figure 3.1: GTAC listening room

divides the octagon in two equal halves:

$$g = \begin{cases} \sqrt{\frac{d}{r_0+d}} & \alpha_{inc} \leqslant 90^\circ \\ 0 & \alpha_{inc} > 90^\circ \end{cases} \qquad (3.1)$$
$$d = \frac{1.44}{2} + 1.44\cos\left(\frac{\pi}{4}\right),$$

where α_{inc} and r_0 are represented in Figure 3.2. Other g could have been used, although different variations were tested and the general trend of results didn't change.

For simplification purposes, from now on it will be assumed that primary sources are monopoles.

3.2. Introduction to simulations

Traditionally, WFS has been used, not to cancel noise, but as a spatial audio reproduction system that competes with existing stereophonic systems as Dolby Surround. The main focus has been, then, not in replicating accurately a field, but in generating the subjective impression of natural hearing, this is, of sound heard from various directions. So, the evaluation of performance has been usually guided by the ability of subjects to localize virtual sound sources and other subjective measures.

Since human hearing has limitations, there are objective sound characteristics that it cannot perceive. We can take advantage of this, and use compression, downsampling and other techniques (common in mp3 and other compression formats) that lower the requirements of the system without worsening the subjective perception. One thing that humans cannot distinguish is constant phase shift with frequency. So, when implementing WFS, the term $\sqrt{j}$ can be omitted (as done in the real implementations in [5] and [6]) since it would require the implementation of a FIR filter that would not add anything to the experience of the audience. Depending on the requirements of the system, the

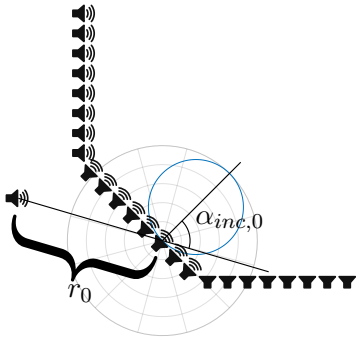


Figure 3.2: WFS calculation parameters

frequency dependent coefficient $\sqrt{k}$ can also be omitted, with the disadvantage that it would produce coloration of the sound ([6]).

Without both filters, the signal of a given loudspeaker can be calculated by just applying a delay to the virtual source signal and multiplying it by a scalar.

$$S^{(wfs)}(f) = S^{(ns,v)}(f) \cos \alpha_{inc,0} \frac{e^{-jkr_0}}{\sqrt{r_0}} g \tag{3.2}$$

$$s^{(wfs)}(t) = s^{(ns,v)}(t) \frac{g \cos \alpha_{inc,0}}{\sqrt{r_0}} * \delta(t - r_0/c), \tag{3.3}$$

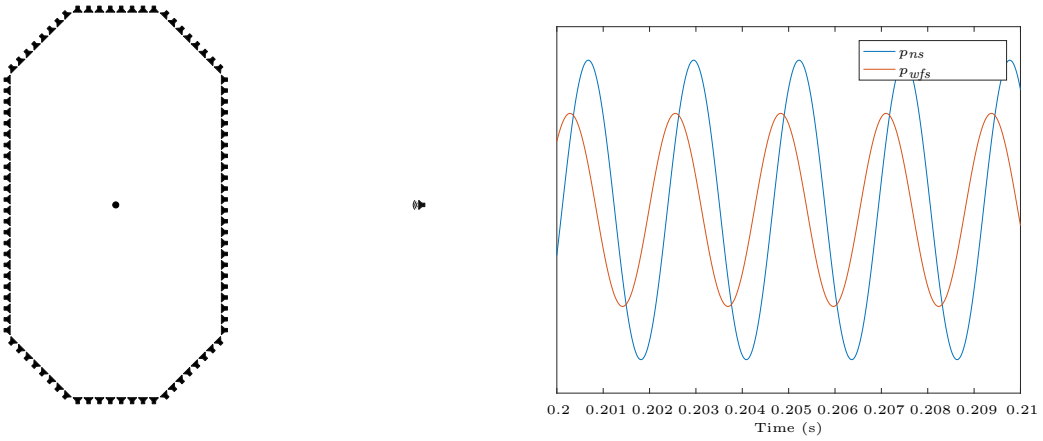
where $*$ is the convolution operator, and $s^{(ns,v)}(t) = -s^{(ns)}(t)$ is the signal transmitted by the primary source, now called virtual noise source because it is the same as the signal transmitted by the noise source multiplied by -1.

This is the first approach we used in simulations. As an example, let's consider the situation in Figure 3.3a. A sinusoidal signals of frequency 440Hz is transmitted by the noise source in free-space conditions. All loudspeakers transmit signals according to Equation 3.3. The signal received at the centre of the octagon is shown in Figure 3.3b. Ideally, the signal received from the noise source p_{ns} and the loudspeaker array p_{wfs} should be opposite, but it is clear they are not.

In order to quantify the performance, we will use the concept of gain G. It is a frequency dependent parameter, defined as the power of the received signal at a given point divided by the power of the signal received exclusively from the noise source.

$$G(f) = 10 \log \left| \frac{P_{wfs}(f) + P_{ns}(f)}{P_{ns}(f)} \right|^2 \quad \text{dB}, \tag{3.4}$$

where $P_{ns}(f)$ and $P_{wfs}(f)$ are the frequency components at fHz of the signals received from the noise source and loudspeaker array respectively. From now on, the frequency dependence will not be explicitly indicated: $G = G(f)$, $P_{ns} = P_{ns}(f)$, etc. Sometimes we will use the terms cancellation or attenuation, referring to the inverse of the gain. For example, saying "cancellation levels should be higher than 8dB" is equivalent to saying "gain levels should be lower than -8dB". Depending on the context, one or the other term will be used. The maximum cancellation (or minimum gain) will happen when $P_{ns} = -P_{wfs}$.



(a) Scheme of basic scenario with one noise source and one point of measure

(b) Signal measured

Figure 3.3: Simulation in simple scenario with one noise source and one point of measure

Another useful parameter is what I call correction factor Ψ. It is conceived as the complex number by which loudspeakers signals $S^{(wfs)}$ should be multiplied in order to achieve maximum cancellation:

$$\Psi = -\frac{P_{ns}}{P_{wfs}}. \tag{3.5}$$

In the previous case, the gain and corrections factors are $G = 3.6\text{dB}$ and $\Psi = 1.1e^{j40.3/360}$.

Since the aim of WFS is achieving cancellation over an wide area, it is reasonable to measure the field not in just one location, but in a grid of points, and the cancellation will vary from one to another. In that case, it is convenient come up with some way of describing the overall performance with just one value that takes in account the field at every point of measure. The concept of average gain, G_a, can be used. Assuming there are M points of measure:

$$G_a = 10\log\left(\frac{1}{M}\sum_{m=1}^{M}\left|\frac{P_{wfs_{(m)}} + P_{ns(m)}}{P_{ns(m)}}\right|^2\right) \quad \text{dB}, \tag{3.6}$$

where $P_{wfs_{(m)}}$ and $P_{ns(m)}$ are the field produced at the m-th point of measure by the loudspeaker array and the noise source respectively.

In the same way, the correction factor has a global equivalent, the global correction factor $\Psi^{'}$, defined as the number that must be multiplied by the loudspeaker signals $S^{(wfs)}$ in order to minimize the average gain.

$$\Psi^{'}(f) = \arg\min_{\psi}\sum_{m=1}^{M}\left|\frac{\psi P_{wfs_{(m)}} + P_{ns(m)}}{P_{ns(m)}}\right|^2 = -\frac{\left(\sum_{m=1}^{M} P_{wfs(m)}/P_{ns(m)}\right)^*}{\sum_{m=1}^{M}\left|P_{wfs(m)}/P_{ns(m)}\right|^2}, \tag{3.7}$$

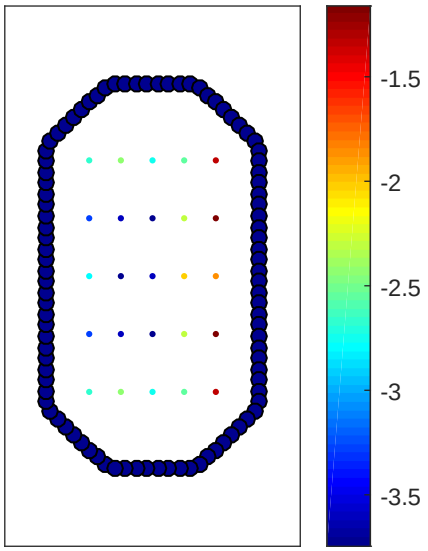


Figure 3.4: Gain 2D map in dB

where * is the complex conjugate symbol.

Justification:

$$\sum_{m=1}^{M}\left|\frac{\psi P_{wfs\,(m)}+P_{ns(m)}}{P_{ns(m)}}\right|^2=\sum_{m=1}^{M}\left|\psi\frac{P_{wfs\,(m)}}{P_{ns(m)}}+1\right|^2=\sum_{m=1}^{M}\psi\mathbf{P_{rel(m)}}+\mathbf{1}^2=\|\psi\mathbf{P_{rel}}+\mathbf{1}\|^2$$

$$\left\{\text{Least squares solution: } \arg\min_{\Psi}\|\mathbf{a}\psi+\mathbf{b}\|^2=(\mathbf{a}^H\mathbf{a})^{-1}\mathbf{a}^H\mathbf{b}=-\frac{\langle\mathbf{a}^*,\mathbf{b}\rangle}{\|\mathbf{a}\|^2}\right\}$$

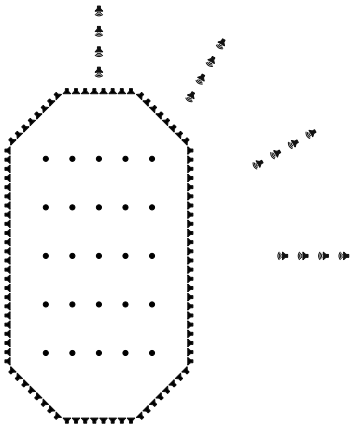
$$\text{so: } \arg\min_{\psi}\|\psi\mathbf{P_{rel}}+\mathbf{1}\|^2=-\frac{\langle\mathbf{P_{rel}}^*,\mathbf{1}\rangle}{\|\mathbf{P_{rel}}\|^2}=-\frac{\sum_{m=1}^{M}P^*_{rel(m)}}{\sum_{m=1}^{M}\left|P_{rel(m)}\right|^2}=-\frac{\left(\sum_{m=1}^{M}P_{wfs\,(m)}/P_{ns(m)}\right)^*}{\sum_{m=1}^{M}\left|P_{wfs\,(m)}/P_{ns(m)}\right|^2}. \tag{3.8}$$

So, instead of one point, now let's see what happens at a grid of points (Figure 3.4). It seems cancellation is bad at all of them. The average gain value is $G_a = 2.55$dB.

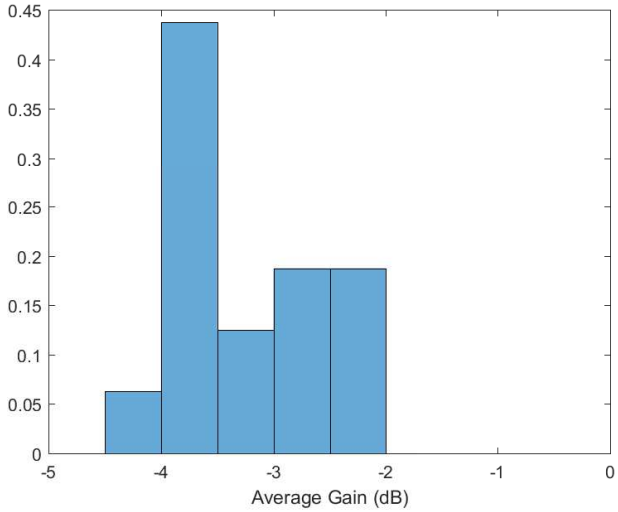
It might be that the bad results have to do with the location of the noise source. Let's then see what happens for different positions as shown in Figure 3.5a. Instead of just showing the 2D map of cancellations for each noise source positions, it will be more convenient for further analysis to visualize it as a histogram of average gain values (Figure 3.5b). The values express probability: number of occurrences in a given gain interval divided by the total number of occurrences.

However, all this has been tested just at one frequency. Maybe, the behaviour for the rest of frequencies is different. To check it out, we will make the noise source transmit a chirp signal from 20Hz to 940Hz (close to the aliasing spatial frequency) and then calculate the FFT. The average gain for one noise source position is shown in Figure 3.6. It is obvious that the values are bad over the whole bandwidth.

In order to understand what is happening for the whole bandwidth and multiple source position in just one snapshot, consider Figure 3.7a. It is a probability histogram



(a) Scheme of multiple points of measure and multiple noise source positions



(b) Histogram of average gains

Figure 3.5: Multiple positions of the noise source

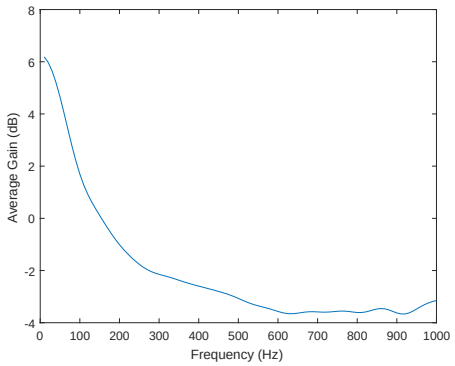
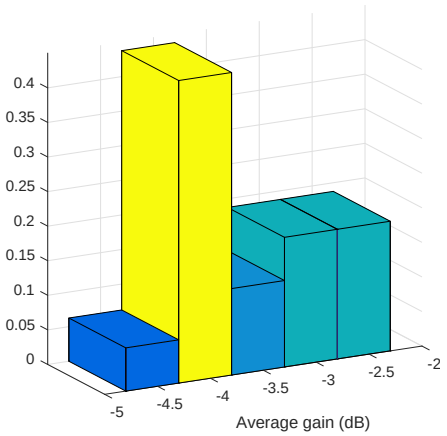


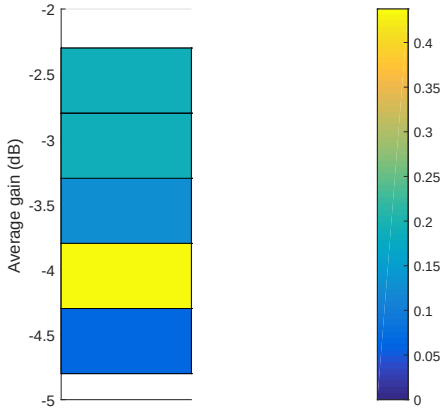
Figure 3.6: Average gain for the whole bandwidth and a given noise source position

of global gains, similar to Figure 3.5b, but seen in 3D, and each bar has been coloured according to the height. If we look at it from above (Figure 3.7b), we will not appreciate the heigh, but the colour will be enough. That histogram is particularized for just one frequency (440Hz). If, for each frequency, we generate a similar coloured histogram and stack them next to each other, we can form an image as Figure 3.7c. Each coloured rectangle is bounded horizontally (x axis) by two frequency values and vertically (y axis) by two gain values. The colour indicates the number of occurrences in that frequency and gain intervals divided by the total number of occurrences for that frequency interval; in other words, each column is an independent probability histogram. The histogram tells us that cancellation is bad at all frequencies, for every position of the noise source.

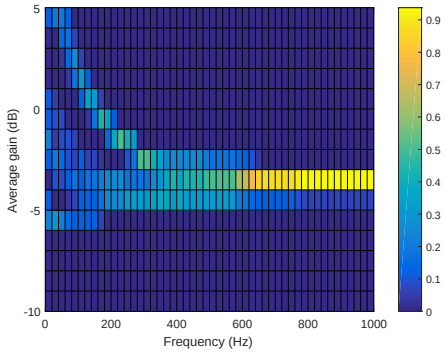
What is happening? The answer can be found if we represent the global correction factor $\Psi^{'}$ (Figure 3.8) . We can see that, even when there is some variance between the values for different noise source positions, the overall tendency follows the theoretical expression $\sqrt{jk/2\pi}$ that was omitted for being considered unnecessary. This result suggests that it is, indeed, necessary.



(a) Example of average gain histogram in 3D



(b) Example of average gain histogram in 3D seen from above

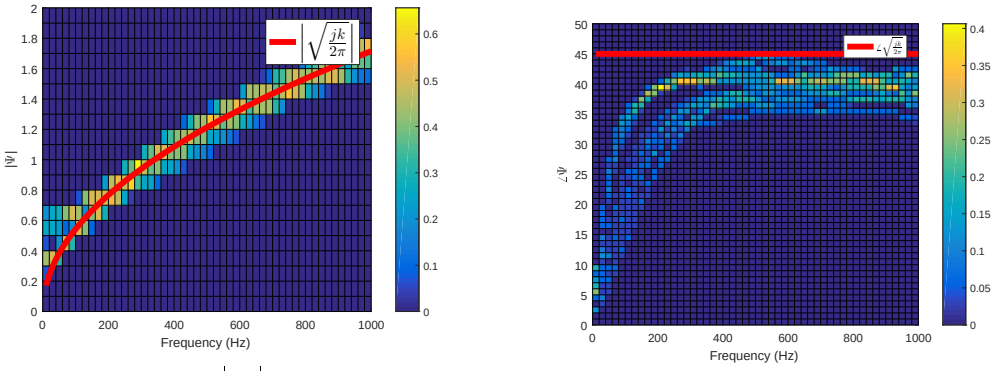


(c) Completed average gain histogram

Figure 3.7: Frequency dependent histogram of gain, step by step

The reason for this is that, even though that term is expendable when generating wave fields that are going to be heard by humans, it is required when the aim is to interference destructively with another existing wave field. For example, for a noise field $P_{ns} = 1$, the secondary source field should be $P_{wfs} = -1$ for perfect cancellation $A = -\infty$dB. But if the phase is shifted $-\pi/4$ ($1/\sqrt{j} = e^{-j\pi/4}$), the amplitude of the field is $|P_{ns} + P_{wfs}| = |1 - e^{-j\pi/4}| = 0.77$, which corresponds to a cancellation of $C = 2.32$dB. Amplitude variations of course also worsen the cancellation levels. So, the term $\sqrt{jk/2\pi}$ might irrelevant when for the human ear when listening to a signal, but are of vital importance if we are going to make a cancelling wave physically interfere with another. We can't get away with just ignoring it.

In conclusion, a $\sqrt{jk/2\pi}$ filter must be implemented in WFS when the intention is to perform active cancellation of noise. This is not a major problem in terms of computational efficiency because this filtering is common to all secondary loudspeakers, so it can be carried out just once on the primary source signal, and then apply the delay and amplitude modification corresponding to each secondary source independently. That is why we talk about "prefiltering". Nevertheless, it presents the problem that, theoretically, the prefilter has an anticausal response, but it must be implemented as a FIR digital filter, so inevitably it will introduce some amount of delay that must be compensated [17]. That sets a constraint on how close the noise source can be to the loudspeaker array, as we will see in next section.



(a) $\left|\Psi^{'}\right|$ (b) $\angle\Psi^{'}$

Figure 3.8: Global correction factor $\Psi^{'}$

3.3. Compromise between frequency filter length and accuracy

In previous section we've concluded that, in order to produce cancellation inside the area enclosed by the loudspeaker array, a filter with a frequency response $\sqrt{jk/2\pi} = \sqrt{jf/c}$ must be implemented, traditionally in the form of a FIR digital filter. This has a drawback, and it is that the ideal theoretical filter is anticausal. This means that, in order to calculate the output signal y at a time t, it needs to use input values that come after that time: $y(t)$ depends on $x(t + \Delta t)$, being $\Delta t > 0$. This is exactly like knowing the future. Mathematically speaking, that can make sense, and in simulations it is not a limitation either since we know exactly what the noise source signal is going to be. But in real-time systems it is, obviously, impossible to implement.

However, there is a nuance that can allow us to implement this filter in real-time. Let's complete Equation 3.3 with the new filter:

$$s^{(wfs)}(t) = s^{(ns,v)}(t)\frac{g\cos\alpha_{inc,0}}{\sqrt{r_0}} * \delta(t - r_0/c) * h(t) = s^{(ns,v)}(t)\frac{g\cos\alpha_{inc,0}}{\sqrt{r_0}} * h(t - r_0/c)$$
$$h(t) = \mathcal{F}^{-1}\left\{\sqrt{\frac{jk}{2\pi}}\right\}. \tag{3.9}$$

When we say that $h(t)$ is anticausal, it is the same as saying that there are negative values of t for which $h(t) \neq 0$. As long at this does not change, the practical implementation of this filter will not be possible. Nonetheless, let's notice that the loudspeaker signal $s^{(wfs)}(t)$ does not depend directly on $s^{(ns,v)}(t)$ filtered by $h(t)$, but on a delayed version of it: $h(t - r_0/c)$.

If the filter impulse response $h(t)$ was limited in time ($h(t) = 0, t < -\tau_1$), then, although $h(t)$ would be anticausal, $h(t - r_0/c)$ will be causal as long as $r_0/c > \tau_1$. However, since the ideal $h(t)$ has an infinite response, the only option left is to use an approximation ($\widetilde{h}(t)$) that satisfies that $\widetilde{h}(t) = 0$ for $t < -\tau_1$, where $\tau_1 < r_0/c$. The delayed filter $\widetilde{h}(t - r_0/c)$ is causal, and hence could work in a real time system.

If, in addition, it has a finite duration, it would be implementable as a FIR digital

filter. This means it must satisfy that $h(t) = 0, t \notin [-\tau_1, \tau_2]$. However, as it works as an approximation, the smaller r_0/c gets, the shorter the duration of the impulse response needs to be, and hence the worse the approximation and the accuracy of WFS will be. So, there is actually a trade-off between performance and the distance between the source and the closest loudspeaker r_0. The smaller the distance, the worse the performance. At a given sampling frequency, which is 44100Hz in our simulations, the duration of the impulse response translates directly to number of coefficients of the digital filter.

We should differentiate between the filter that implements the magnitude of the frequency response $h_1 = \mathcal{F}^{-1}\left\{\sqrt{f/c}\right\}$ and the one that implements the frequency independent phase shift $h_2 = \mathcal{F}^{-1}\{\sqrt{j}\}$, because they present different requirements. The real implementations will be called $\widetilde{h_1}$ and $\widetilde{h_2}$ respectively.

A way of generating this filters is by using the analytical discrete impulse responses that were provided in [18][Equation 2.192 and 2.194] for the phase and magnitude filters respectively:

$$h_{mag}[n] = \begin{cases} -\frac{S(\sqrt{2n})}{\sqrt{2\pi}n^{3/2}}, & n \neq 0 \\ \frac{2}{3}\sqrt{\pi}, & n = 0 \end{cases} \tag{3.10}$$

$$h_{pha}[n] = \begin{cases} 0, & \text{if } n \text{ is even and } n \neq 0 \\ \frac{-\sqrt{2}}{\pi n}, & \text{if } n \text{ is odd} \\ \frac{1}{\sqrt{2}}, & n = 0 \end{cases} \tag{3.11}$$

where S is the fresnel sine integral function (fresnels in Matlab).

The frequency response of the magnitude filter h_{mag} when the number of coefficients is infinite is:

$$H_{mag}(f_{norm}) = \frac{2}{5}\sqrt{f_{norm}}, \quad f_{norm} \in [-0.5, 0.5], \tag{3.12}$$

where f_{norm} is the normalized frequency in periods per sample.

As the ideal filter we want to get has a magnitude response $\sqrt{f/c}$ at a sampling frequency f_s, the magnitude filter must be scaled in order to fit our purposes:

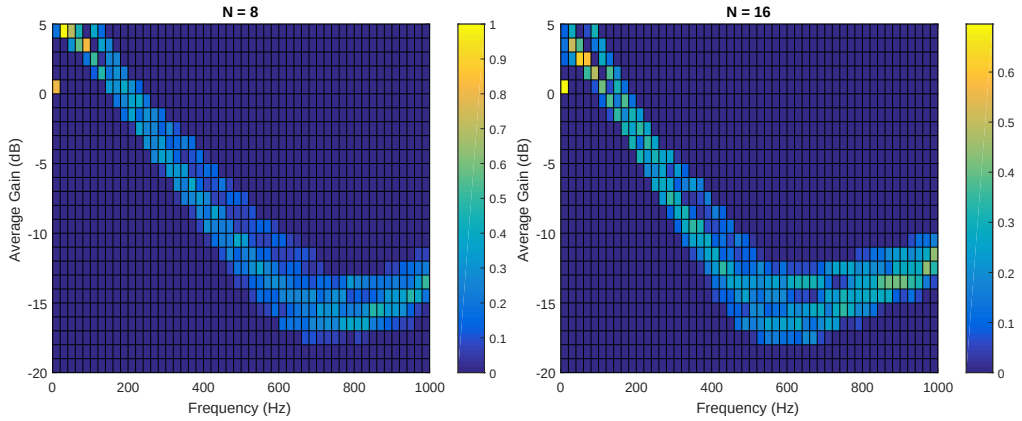
$$h_1 = h_{mag}\frac{5}{2}\sqrt{\frac{f_s}{c}}. \tag{3.13}$$

Finally, we must choose a window function in order to make it finite.

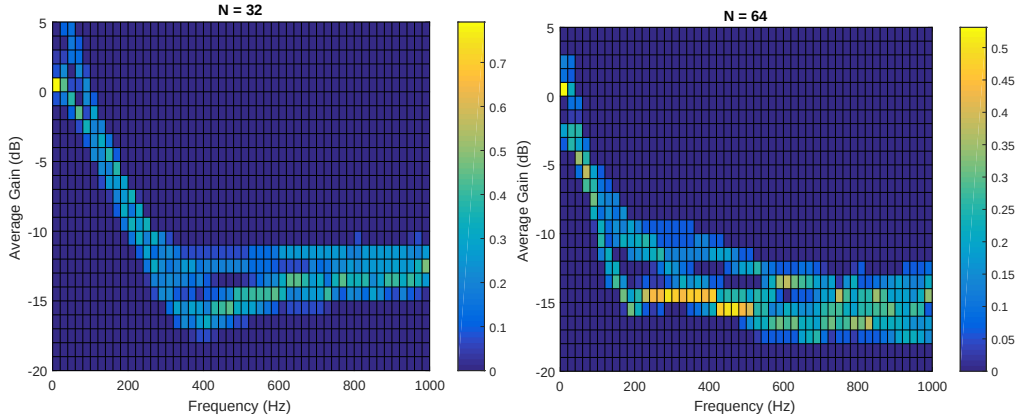
Matlab also has its own internal functions to design FIR filters based on their frequency response and the number of coefficients, and in the case of these filters, it also returns symmetrical impulse responses ($\tau_1 = \tau_2 = \tau$). Using anyone of both ways (Matlab built-in function or analytical equations with rectangular window or Kaiser-Bessel window as proposed in [18]), results are pretty similar. Using other window functions may improve filter response. Next simulations use the Matlab functions because, as they use optimization algorithms to find the most accurate response possible for a given number of coefficients, it is very likely that the generated filter it's close to the best possible configuration. But, as said before, results are actually pretty similar anyway.

We have plotted the histograms of gain for different lengths of the implementation of $\widetilde{h_1}$ (Figure 3.9), keeping $\widetilde{h_2}$ with a high enough length so it resembles the ideal case

and does not produce any noticeable distortion. We can see that, below an order of 64 (the order of a FIR filter is its length minus 1), the performance starts to get worse and worse. Above it, the improvement is not too noticeable.



(a) Order: 8. $\tau = 0.091$ms. $r_{0min} = 3$cm (b) Order: 16. $\tau = 0.181$ms. $r_{0min} = 6$cm



(c) Order: 32. $\tau = 0.363$ms. $r_{0min} = 12$cm (d) Order: 64. $\tau = 0.726$ms. $r_{0min} = 25$cm

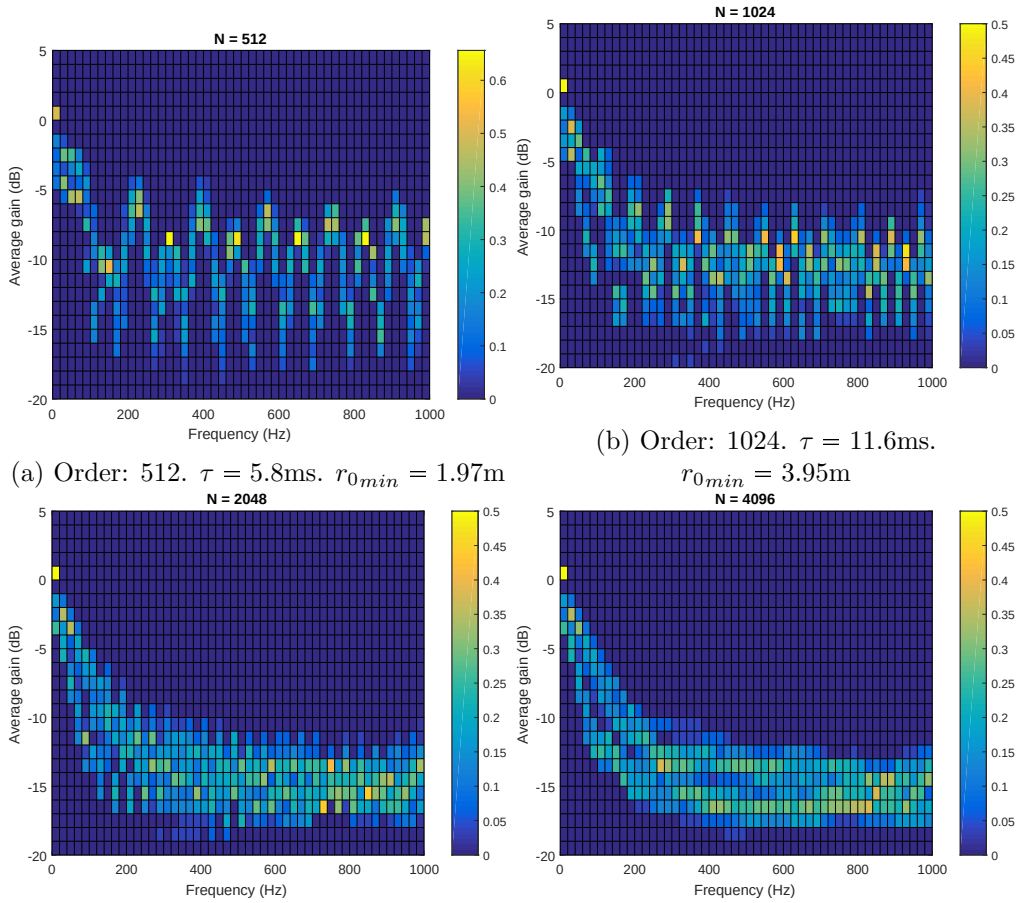
Figure 3.9: Average gain for different lengths of h_1

We've done the same with $h_2 = \mathcal{F}^{-1}\{\sqrt{j}\}$. In this case, a Hilbert filter must be implemented and, as we see in (Figure 3.10), its length has to be much bigger than with the magnitude filter length. It entails a more severe constraint. $\widetilde{h_1}$ order is high enough so produced distortions are negligible. Orders of 4096 and above get practically the same levels of attenuation, so it makes no sense to go beyond 4096. However, such a long filter implies that the distance from the noise source to the closest loudspeaker should be at least 15.79m, a restriction that can not always be met.

3.3.1. Alternatives

In case the noise source is located close to the array, an alternative solution must be seek. On the one hand, one could just reduce the length of the filter as long as the system requirements allow such a reduction in performance.

Another option is available if we know what the sound is, and it's limited in time (sound pulse), but we don't know when it happens. For example, in [12] the cancellation



(a) Order: 512. $\tau = 5.8$ms. $r_{0min} = 1.97$m

(b) Order: 1024. $\tau = 11.6$ms. $r_{0min} = 3.95$m

(c) Order: 2048. $\tau = 23.2$ms. $r_{0min} = 7.89$m

(d) Order: 4096. $\tau = 46.4$ms. $r_{0min} = 15.79$m

Figure 3.10: Average gain for different lengths of h_2

of gun shots was simulated. In a shooting range, it's an unknown when a gun will be fired, but the sound of every gun shot shares a lot of similarities. It was found that, when using one gun shot standard signal to cancel the sound of a different gun shot (similar but not equal), even though the sound cancellation performance is significantly lower than if the same sound was used, it's still consistently better than if no WFS was used. The fact that the signal used for WFS cancellation is not the actual noise signal but a known model that resembles the actual signal, allows us to compute the filtered signal in advance. The performance then depends on the similarity between pulse sounds (the model and the real), and the precision at calculating the time the signal starts, which is the actual unknown. This can be used for sounds produced by gun shots, mechanical machines that produce repetitive pulse sounds that are similar to one another, etc.

If this condition is not met, there is still a last option: not using this filter at all (h_2). By means of a delay, we can apply the right phase shift for a certain frequency f_0, ideally located in the centre of the bandwidth of interest. The idea is to substitute $h_2 = \mathcal{F}^{-1}\{\sqrt{j}\}$ by a simple delay $h_2' = \delta(t-\tau)$, and so $H_2 = \sqrt{j} = e^{j\pi/4}$ by a linear phase shift $H_2' = e^{-j2\pi f\tau}$. The value of τ must be the smallest positive delay that causes a phase shift of $\pi/4$ or equivalent ($\pi/4 + n2\pi$, $n \in \mathbb{Z}$) at f_0:

$$-2\pi f_0\tau = \pi/4 + n2\pi \rightarrow \tau = \frac{1}{f_0}\frac{8n-1}{8} = \{n=1\} = \frac{7}{8f_0}. \tag{3.14}$$

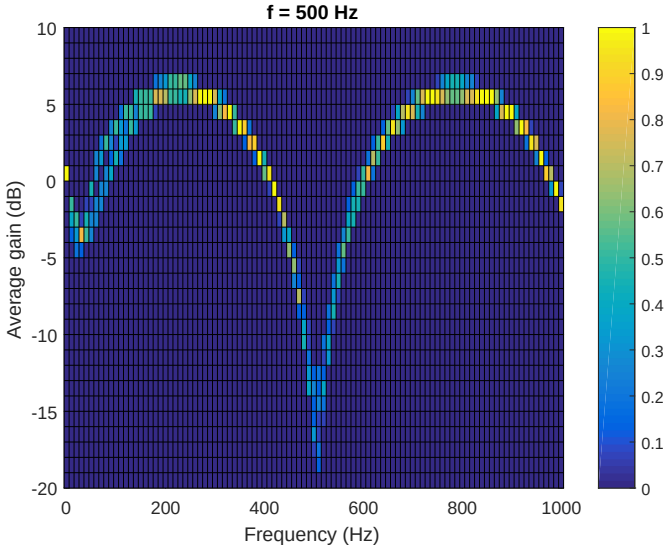


Figure 3.11: Average gain with $h_2 = \delta(t - \frac{7}{8 \cdot 500})$

In Figure 3.11, a frequency of $f_0 = 500$Hz has been chosen. Cancellation is high at $f_0 \pm 25$Hz approximately, but it decays quickly outside that band. This can be a good solution for narrowband noise signals.

As a side-note, a IIR filter design was proposed [18], but the response was shown to be poor at low frequencies (the few more examples of IIR implementations found in literature are already taken in account in the mentioned thesis). However, we did not tested the actual performance in our system and and it can be an interesting line of research.

3.4. Performance in non-free-space conditions

Every simulation that has been done so far has assumed free-space conditions. However, in a real set-up there are all kinds of reflections, diffraction effects, etc. Simulations of more real conditions are interesting. In order to do so, a Room Impulse Response generator was used. It has been developed by International Audio Laboratories Erlangen. It allows to define the dimensions of a room in the shape of a rectangular box ($[x, y, z]$), the reflection coefficients of each one of the six walls, and calculate, using the image method, the impulse responses between points specified by the user [19]. Those impulse responses have been used instead of the free space ones.

The order of the filters $\widetilde{h}_1$ and $\widetilde{h}_2$ are high enough so they don't cause any relevant inaccuracies. The same reflection coefficient β has been used for every wall. Room dimensions have been chosen to resemble the actual listening room dimensions: $[4.48, 9.13, 2.64]$m, and all sources, loudspeakers and receiving points are located at a height of 1.65m (Figure 3.12).

Figure 3.13 shows the average gain histogram for different values of β. When it goes above 0.2, the average gain reaches levels above -10dB.

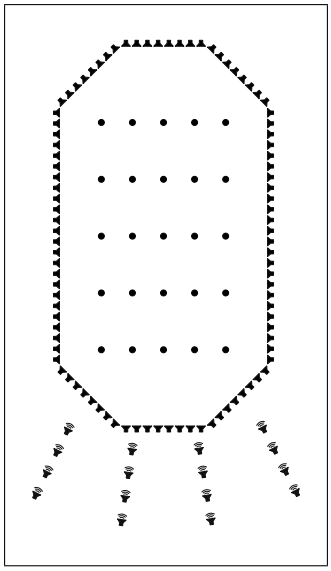


Figure 3.12: Scheme of the simulated room

3.4.1. GTAC listening room

The real impulse responses in the listening room are available in the research group website [20]. They were measured with high precision for a rectangular grid of points of size 24x15 separated a distance of 20cm between adjacent nodes [16]. We could use those responses in calculations instead of the simulated ones. However, we found the limitation that this measures were done only for the loudspeaker array, but not for loudspeakers outside the array in arbitrary positions. So, responses for potential noise sources are not available.

Nonetheless, there is an approach that can help us estimate what the performance in the real set up would be. We have seen that the existence of reflective surfaces worsens performance. The impulse response $h(t)$ from a source to a measuring point does not follow any more the ideal monopole free-space propagation equation in which the whole WFS theory is based.

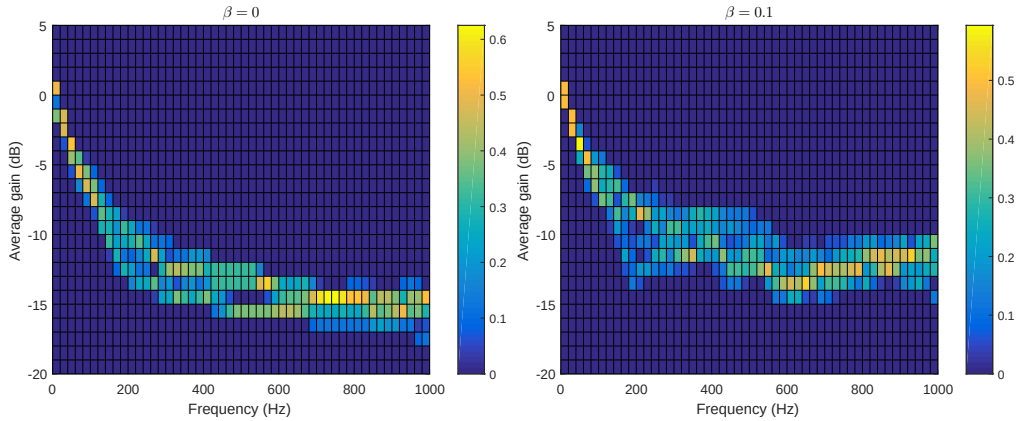
In the time domain, the propagation in free-space conditions of a signal $s^{(ns)}(t)$ emitted by a monopole source to a given measure point separated a distance d, consists just of a delay equal to the time it takes to the signal to arrive at the measure point ($\tau = d/c$, where c is the speed of sound) and an amplitude reduction proportional to the distance between the source and the measure point ($1/d$):

$$p(t) = s^{(ns)}(t) * h(t) = s^{(ns)}(t) * \frac{1}{d}\delta(t-\tau). \tag{3.15}$$

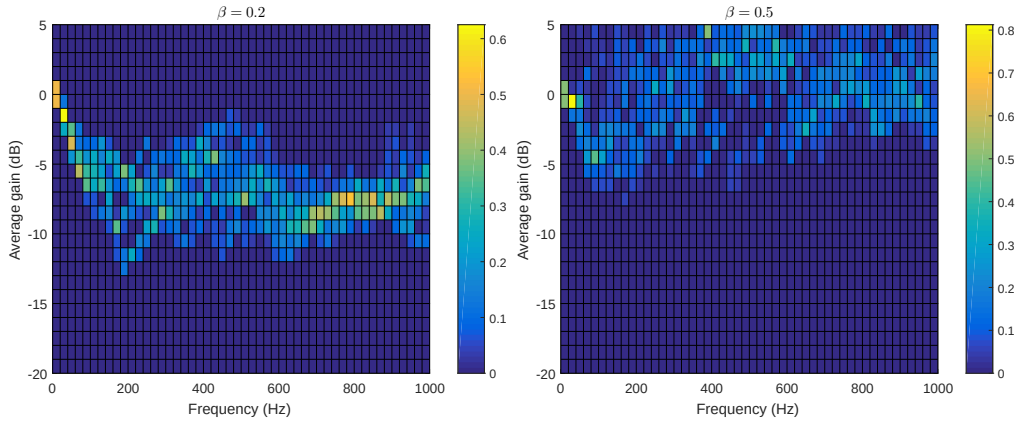
In the frequency domain, this translates to

$$P(f) = S^{(ns)}(f)H(f,d) = S^{(ns)}(f)\frac{e^{-j\frac{2\pi f}{c}d}}{d}. \tag{3.16}$$

A way of visualizing the deterioration is comparing the magnitude of the ideal response for a given distance, with the magnitude of the actual one. In Figure 3.14a the



(a) $\beta = 0$ (b) $\beta = 0.1$

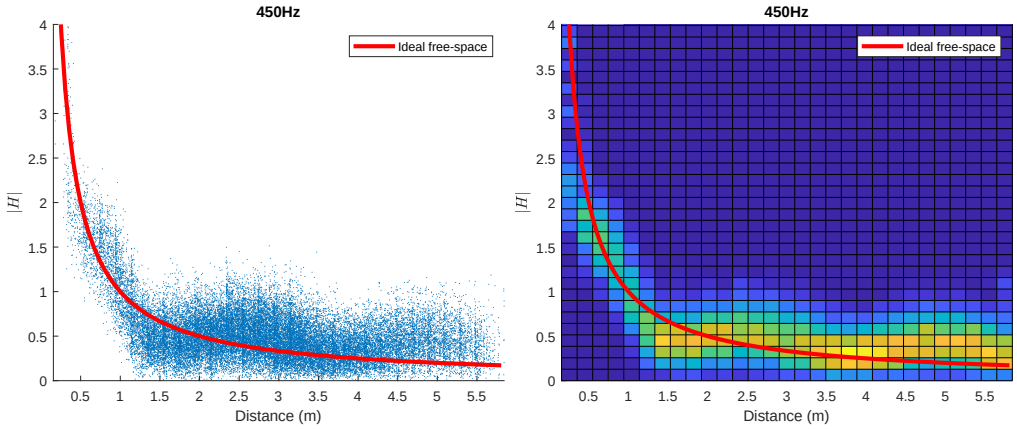


(c) $\beta = 0.2$ (d) $\beta = 0.5$

Figure 3.13: Average gain for different reflection coefficients β

ideal response is represented by a line and the real one, particularized for 450Hz, is represented with dots, where each dot represents a pair loudspeaker-measure point (96 loudspeakers, 360 measure points, $96 \cdot 360 = 34560$ dots). The x value is the distance between source and measure point, and the y value is the magnitude of the frequency response $|H|$. A clearer way of visualizing it is a composed histogram as the ones used for the gain (Figure 3.14b). Each coloured rectangle is delimited horizontally by two distance values (x axis), and vertically by two magnitude values (y axis). The colour represents the proportion of points that lie in that distance and magnitude interval with respect to the total amount of points for that given distance interval. In other words, each column is an independent probability histogram.

The difference is evident. Ideally, all dots (or all coloured rectangles) would lie on the line, but they don't. And the further away they are from the source, the more scattered they appear. But, how would this affect performance? We can make a rough estimate by comparing it with the deterioration produced in simulated scenarios. Figure 3.15 uses the same type of representation, but in this case, the impulse responses have not been measured in the real set up, but calculated with the Room Impulse Response generator using the listening room dimensions. A visual comparison over different values of the reflection coefficient β, suggests that the one that generates a most similar scenario to Figure 3.14b is around $\beta = 0.8$ (magnitude values have been normaized). The average gain histogram simulated with that value, for the noise source configuration already



(a) (b)

Figure 3.14: Relation of frequency response and distance from source and receiver. Measured responses in the listening room.

used in previous subsection (Figure 3.12) is shown in Figure 3.16. As we can see, in most of the cases the average gain is even positive. This suggests that the possibility of performing noise cancellation with WFS techniques is very unlikely.

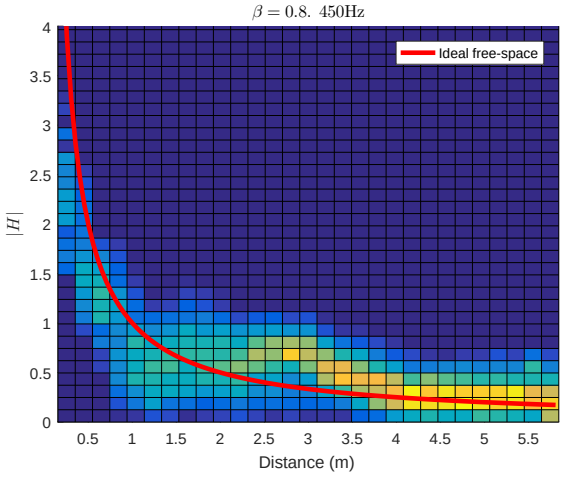


Figure 3.15: Relation of frequency response and distance from source and receiver. Simulated responses for $\beta = 0.8$.

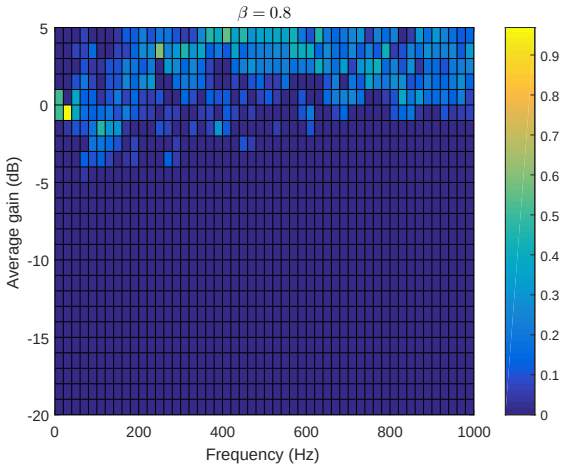


Figure 3.16: Average gain for $\beta = 0.8$.

3.5. Truncation and lower cut-off frequency

There is a remarkable phenomenon that can be observed through all simulations. If we take a look at any of the previously shown attenuation histograms, we will see that from approximately 200Hz down, the performance gets poorer as the frequency decreases. It is as if the system had a lower cut-off frequency of 200Hz below which it doesn't work properly (Figure 3.17). That transition region is there due to the fact that in practice, the line of secondary sources (the loudspeaker array) has a finite length, in other words, it is truncated. In order to explain this, it is convenient to recall some of the theoretical expressions of WFS.

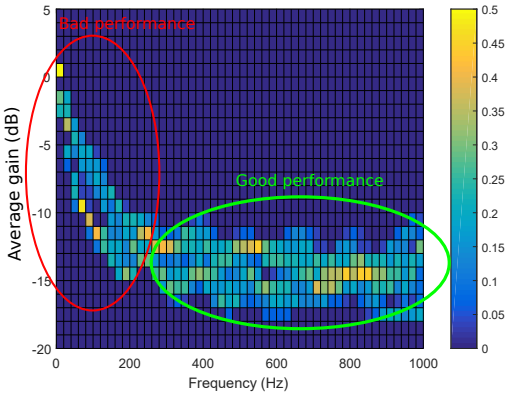


Figure 3.17: Example of performance degradation for low frequencies

As we've already seen, the field created by a monopole source (primary source) located at $\mathbf{x}_{ps}$ at a receiving point $\mathbf{x}$ is:

$$P_{ps}(\mathbf{x}) = S(f)\frac{e^{-jk\|\mathbf{x}-\mathbf{x}_{ps}\|}}{\|\mathbf{x}-\mathbf{x}_{ps}\|}, \tag{3.17}$$

where f is the frequency, $S(f)$ is the feeding coefficient to the source, and $k = \frac{2\pi}{\lambda}$ is the propagation constant.

Rayleigh I 2.5D integral allows to replicate the field by an infinite line distribution of monopole sources (secondary sources), given the condition that all points where the field is replicated, as well as the primary source location and the secondary source line distribution (L), are all on the same plane, and that the secondary source line separates the reconstructed field region from the primary source. The expression is:

$$\begin{aligned} P(\mathbf{x}) &= \int_L Q_I(\mathbf{x}_{s0}, \mathbf{x}) \frac{e^{-jk\Delta r_0}}{\Delta r_0} \mathrm{d}\mathbf{x}_{s0}, \\ Q_I(\mathbf{x}_{s0}) &= S \cos \alpha_{inc,0} \frac{e^{-jkr_0}}{\sqrt{r_0}} \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{d}{d + d_{ps}}}, \end{aligned} \tag{3.18}$$

where $Q(\mathbf{x}_{s0}, \mathbf{x})$ is the feeding of the differential secondary monopole source located at $\mathbf{x}_{s0}$, $r_0 = \|\mathbf{x}_{s0} - \mathbf{x}_{ps}\|$ is the distance from the primary source to the secondary source and $\Delta r_0 = \|\mathbf{x}_{s0} - \mathbf{x}\|$ is the distance from the secondary source to the reconstruction point and d_{ps} is the distance between the primary source and the secondary source line. It is valid for values of $kr_0 >> 1$. The field is replicated with correct amplitude over a line parallel to the secondary source line and separated a distance d.

One of the reasons why Rayleigh I 2.5D integral is still far from practical applications is that the line distribution of monopoles L has an infinite longitude, which is of course unrealistic. At some point the line has to be truncated, and hence, the accuracy of the reconstructed field will be affected.

In order to study this limitation, we can look at a simple scenario where a line of secondary sources D meters long is located at the x axis from from $-D/2$ to $D/2$, a primary source is located on the negative y axis at $\mathbf{x}_{ps} = [0, -d_{ps}, 0]$ and the receiving point on the positive y axis at $\mathbf{x} = [0, d, 0]$ (Figure 3.18). Depending on those four parameters, three distances (D, d_{ps} and d) plus the frequency f, the accuracy of the synthesized field will vary.

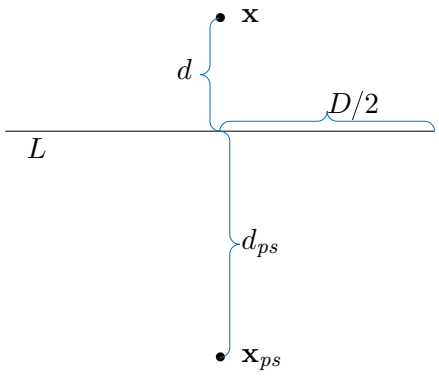


Figure 3.18: Scheme of truncation scenario

In this scenario, the reconstructed field at point $\mathbf{x}$ is:

$$P(\mathbf{x}) = \int_{-D/2}^{D/2} Q_I(x_{s0}, \mathbf{x}) \frac{e^{-jk\Delta r_0}}{\Delta r_0} \mathrm{d}x_{s0}, \tag{3.19}$$

where $r_0 = \|\mathbf{x}_{s0} - \mathbf{x}_{ps}\| = (d_{ps}^2 + x_{s0}^2)^{1/2}$ is the distance from the primary source to the secondary source, $\Delta r_0 = \|\mathbf{x}_{s0} - \mathbf{x}\| = (d^2 + x_{s0}^2)^{1/2}$ is the distance from the secondary

source to the reconstruction point and $\cos\alpha_{inc,0} = d_{ps}/r_0$ is the cosine of the angle of incidence.

A useful measure that facilitate the evaluation of the accuracy is the relative field $P_{rel}(\mathbf{x})$, this is, the resulting field divided by the ideal one produced by the primary source.

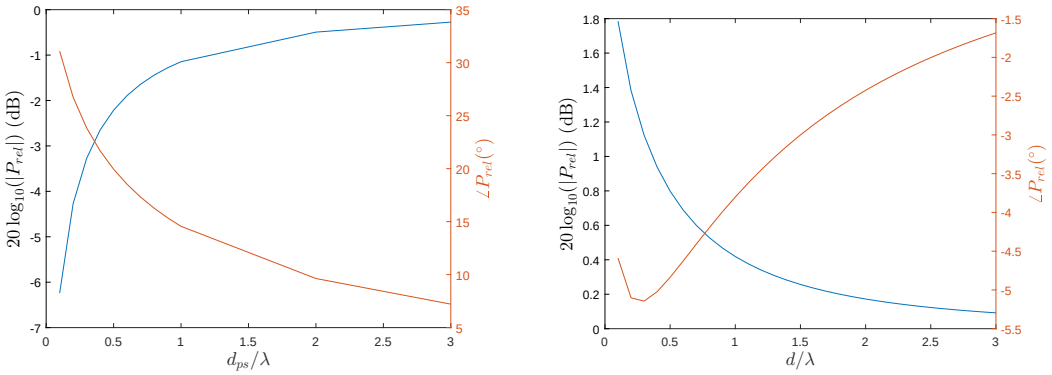
$$P_{rel}(\mathbf{x}) = \frac{P(\mathbf{x})}{P_{ps}(\mathbf{x})}. \tag{3.20}$$

The closer P_{rel} is to 0dB the better the accuracy. The starting point will be a very simple case, and then we will add some complexity.

3.5.1. Ideal case: $D = \infty$

First, we consider an array of infinite length $D = \infty$. Theoretically, any inexactitude must necessarily be produced by the application of the stationary point method in the dimensionality reduction from a plane to a line, and to the assumption that the primary source is in the far field $kr_0 >> 1$ (section 2.3).

In numerical calculation of the integral in Matlab (Figure 3.19a) shows that when d_{ps} gets smaller, the reconstructed field deteriorates, and so WFS is actually not useful under such conditions. A similar behaviour but much less severe is found when the primary source is in the far field but the receiving point is very near the secondary source line (Figure 3.19b). In conclusion, as long as both the primary source and the



(a) $d/\lambda \gg 1$ (b) $d_{ps}/\lambda \gg 1$

Figure 3.19: Magnitude and phase of the relative field for an infinite line array $D = \infty$

receiving point are not too close to the secondary source line, the precision will not be too distorted.

3.5.2. Primary source in the infinite

Let's take the case where the primary source is far: $d_{ps} \gg 1$. Then,

$$\cos\alpha_{inc,0} \approx 1$$
$$r_0 \approx d$$
$$\sqrt{\frac{d}{d_{ps}+d}} \approx \sqrt{\frac{d}{d_{ps}}},$$

for all the segment $x_{s0} \in [-D/2, D/2]$, so Equation 3.19 simplifies to:

$$P(\mathbf{x}) = \int_{-D/2}^{D/2} Q(x_{s0},\mathbf{x})\frac{e^{-jk\Delta r_0}}{\Delta r_0}\mathrm{d}x_{s0} =$$
$$= \left\{ Q(x_{s0},\mathbf{x}) = \frac{e^{-jkd_{ps}}}{\sqrt{d_{ps}}}\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{d}{d_{ps}}} \right\} =$$
$$= \frac{e^{-jkd_{ps}}}{\sqrt{d_{ps}}}\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{d}{d_{ps}}}\int_{-D/2}^{D/2}\frac{e^{-jk\Delta r_0}}{\Delta r_0}\mathrm{d}x_{s0}. \quad (3.21)$$

The integral is actually the field generated by a linear source. When we are dealing with an infinite line source ($D \to \infty$), the exact solution is a scaled version of the zero-th order of the Hankel function of second type. Nonetheless it approximates very well to another much more useful expression:

$$\int_{-\infty}^{\infty}\frac{e^{-jk\Delta r_0}}{\Delta r_0}\mathrm{d}x_{s0} = -\pi j\mathrm{H}_0^{(2)}(\mathrm{kd}) \approx \frac{e^{-jkd}}{\sqrt{kd}}\sqrt{\frac{2\pi}{j}}. \quad (3.22)$$

Substituting Equation 3.22 in Equation 3.21:

$$P(\mathbf{x}) \approx \frac{e^{-jkd_{ps}}}{\sqrt{d_{ps}}}\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{d}{d_{ps}}}\frac{e^{-jkd}}{\sqrt{kd}}\sqrt{\frac{2\pi}{j}} = \frac{e^{-jk(d_{ps}+d)}}{d_{ps}} \approx \frac{e^{-jk(d_{ps}+d)}}{d_{ps}+d} = P_{ps}(\mathbf{x}). \quad (3.23)$$

So, when $D \to \infty$, the synthesized field is the same as the field from the primary source, as stated by WFS theory. What happens when the length of the secondary source line gets shorter? It all comes dowm to the integral:

$$I(\lambda/D, d/D) = \int_{-D/2}^{D/2}\frac{e^{-jk\Delta r_0}}{\Delta r_0}\mathrm{d}x_{s0} = \int_{-1/2}^{1/2}\frac{e^{-j\frac{2\pi}{\lambda/D}\sqrt{(d/D)^2+x_{s0}^2}}}{\sqrt{(d/D)^2+x_{s0}^2}}\mathrm{d}x_{s0}. \quad (3.24)$$

Figure 3.20 shows an example of how the integral evolves when increasing D. As we see, the magnitude and the phase oscillate and converges towards the value in Equation 3.22 when D increases. It is pretty obvious that there is a transition period where D is too small to produce accurate enough results.

As we've seen in Equation 3.24, I depends actually just on d/D and D/λ. In Figure 3.21 we've represented the value of D/λ where the phase (and also the magnitude) starts to converge.

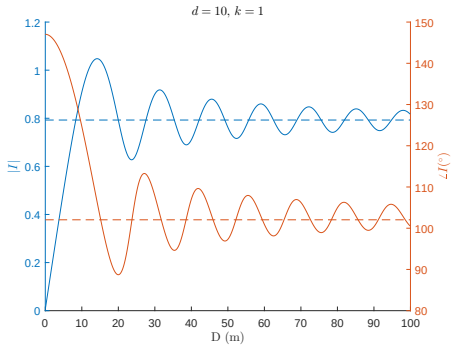


Figure 3.20: Field generated by a linear source of length L ($d = 10$, $k = 1$)

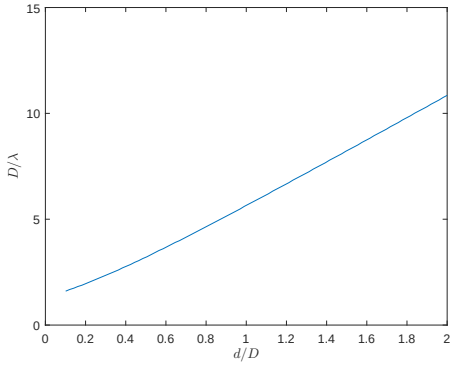


Figure 3.21: Minimum value of D/λ where the value of I (Equation 3.24) starts to converge

All this means that there is a lower cutoff frequency below which WFS is not useful. For example, Figure 3.22 shows the cutoff frequency for a linear array of length $D = 0.18 \cdot 23 = 4.14$m, as one of the sides of the loudspeaker array in the GTAC listening room.

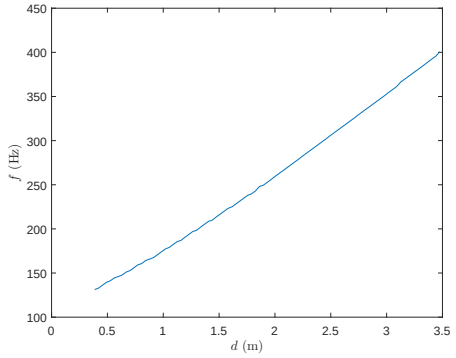


Figure 3.22: Cutoff frequency for a linear array whose length is $D = 0.18 \cdot 23$, as two of the sides of the loudspeaker array in the GTAC listening room

It's worth noticing that this analysis is done with a geometrically very simple scenario where the primary source is in the infinite, the receiving point is centred with respect to the secondary source line, we don't use complicated geometries as octagons, etc. The more variations we add, the more complex the results are. However, given what we observed in the simulations of the GTAC scenario, the existence of a lower cut-off

frequency seems to be maintained even in those more complicated cases.

A deeper theoretical analysis of the truncation artefacts can be found in [4, Section 4.3], where various analytical approximations are proposed. Some techniques, as tapering, are proposed to reduce the effects of truncation, although they have only shown to work in mid and high frequencies.

Chapter 4

Experimental measures

WFS theory is based on a propagation model where the acoustic field is generated by punctual primary sources and the cancellation field is generated by punctual monopole secondary sources, everything in an homogeneous media and free-space condition, and the location of every source is perfectly known.

A real situation, as the one we find in a listening room with real loudspeakers, is very different. A great variety of phenomena not contemplated by the WFS simple model occur: reflections, diffractions produced by obstacles, loudspeakers are not punctual sources so the near-field does not follow the far-field approximation, the directivity is not the one of an ideal monopole and depends on the frequency, the frequency response of loudspeakers is not flat, which would not be that big of a problem if it was the same for every loudspeaker, but it might actually be very different from one to the other, their exact locations are not accurately known, non-linearities, etc.

All these differences influence the way the acoustic waves propagate and, in general, they worsen the performance in a real situation. Experimental measures help us understand how this not contemplated differences limit the possibilities of using WFS in the real world. In a real situation, all we can certainly know is that, if we transmit a signal through a loudspeaker and we measure at some point with a microphone, we receive a modified version of the signal. That modification depends on all the conditions previously mentioned (multiple reflections, etc.), and together they form what is often called acoustic path. An acoustic path between a loudspeaker and a point of measure acts as a filter characterized by an impulse response (time domain) or frequency response (frequency domain).

When we use N_{WFS} secondary loudspeakers, N_{NS} noise sources and M points of measure, the relation between transmitted signals and received ones in the frequency domain is:

$$\mathbf{P}(f) = \boldsymbol{A}(f)\mathbf{S}(f) = \left\{ \begin{array}{l} \mathbf{S} = \begin{bmatrix} \mathbf{S}^{(wfs)} \\ \mathbf{S}^{(ns)} \end{bmatrix} \\ \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{WFS}, & \boldsymbol{A}_{NS} \end{bmatrix} \end{array} \right\}$$
$$= \boldsymbol{A}_{NS}(f)\mathbf{S}^{(ns)}(f) + \boldsymbol{A}_{WFS}(f)\mathbf{S}^{(wfs)}(f) = \mathbf{P}_{ns}(f) + \mathbf{P}_{wfs}(f), \quad (4.1)$$

where $\mathbf{P}_{ns} = [P_{ns(1)}, P_{ns(2)}, ..., P_{ns(M)}]^T$ and $\mathbf{P}_{wfs} = [P_{wfs_{(1)}}, P_{wfs_{(2)}}, ..., P_{wfs_{(M)}}]^T$ are the acoustic pressure vectors (M is the number of points of measure), $\mathbf{S}^{(ns)} = [S_1^{(ns)}, S_2^{(ns)}, ..., S_{N_{NS}}^{(ns)}]^T$ (N_{NS} is the number of noise sources) and $\mathbf{S}^{(wfs)} = [S_1^{(wfs)}, S_2^{(wfs)}, ..., S_{N_{WFS}}^{(wfs)}]^T$ (N_{WFS} is the number of secondary sources) are the transmitted signal vectors, and $\boldsymbol{A}_{NS(M\times N_{NS})}$ and $\boldsymbol{A}_{WFS(M\times N_{WFS})}$ are matrices whose (m,n)-th element is the frequency response of the acoustic path between the n-th loudspeaker and the m-th point of measure. If the cancellation is successful, $\mathbf{P}_{ns}(f) \approx -\mathbf{P}_{wfs}(f)$ and so, the elements of $\mathbf{P}$ become really small. Cancellation will be successful as long as the acoustic paths are the same or very similar to the ideal ones. In the ideal case, the (m,n)-th element of $\mathbf{a}$ would apply a simple delay and an amplitude attenuation:

$$a_{m,n}(f) = \frac{e^{-jkd_{m,n}}}{d_{m,n}}, \quad (4.2)$$

where $d_{m,n}$ is the distance between the n-th loudspeaker and the m-th point of measure.

In order to perform a simple experiment, a loudspeaker will act as a noise source that transmits a known signal. Specifically, it is a chirp signal of duration 40 seconds, preceded and followed by two seconds of silence. The frequency increases linearly from 20Hz to 1250Hz (Figure 4.1). We have chosen to work at a sample rate of 44100.

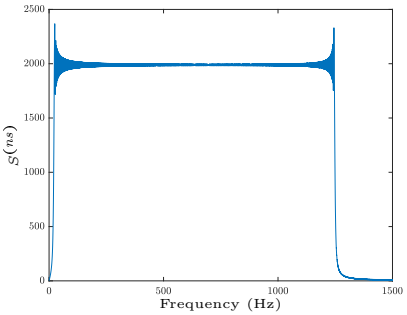


Figure 4.1: Noise source signal spectrum

The signal transmitted by the WFS array loudspeakers $\mathbf{s}^{(wfs)} = [s_1^{(wfs)}, s_2^{(wfs)}, ..., s_{N_{WFS}}^{(wfs)}]^T$ are calculated using Equation 3.9, where the prefilter $h(t) = \mathcal{F}^{-1}\left\{\sqrt{\frac{jk}{2\pi}}\right\}$ has been implemented using a magnitude filter $h_1 = \mathcal{F}^{-1}\left\{\sqrt{f/c}\right\}$ of order 1024, and a phase filter $h_2 = \mathcal{F}^{-1}\left\{\sqrt{j}\right\}$ of order 4096. Since the noise signal is known beforehand, we have been able to calculate $\mathbf{s}^{(wfs)}$ before reproduction, and the delay introduced by the prefilter has been artificially compensated by shifting signals by a right amount of samples. This obviously would not be possible in a real-time application. The estimated noise source

position (the positions that is used to calculate WFS signals) is $\mathbf{x}_{ns} = [x_{ns}, y_{ns}, z_{ns}] = [3.83, 1.40, 1.65]$m (remember all loudspeakers are situated 1.65m above the floor). Two microphones are located at $[1.97, 3.69, 1.65]$m and $[1.44, 5.42, 1.65]$m. The room has dimensions $(4.48 \times 9.13 \times 2.64)$m (Figure 4.2).

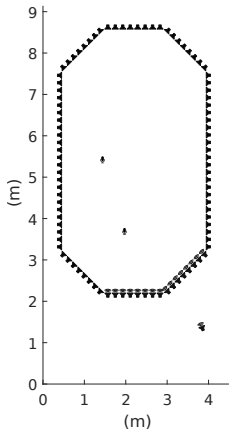
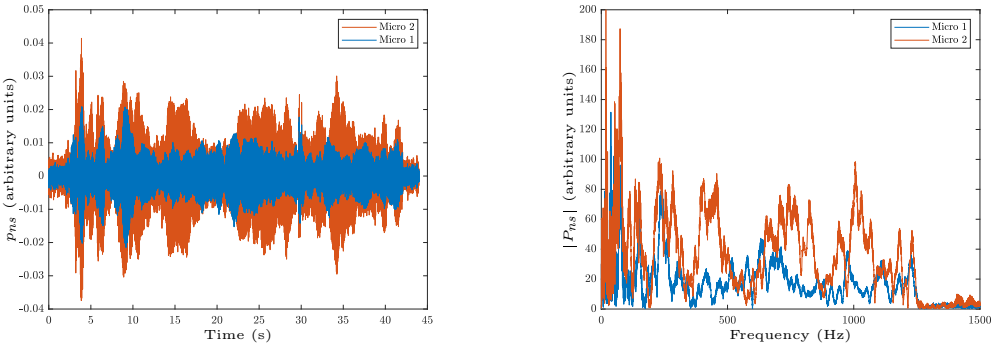


Figure 4.2: Scheme of the measure scenario

When playing the noise source signal only, the received signals by the microphones are shown in Figure 4.3. The spectrum has been calculated by applying the Fast Fourier Transform (FFT) to the received signal. The number of points used is the same as the number of samples of the signal, 1940400. The rest of spectra are calculated also in this way. An ideal response would present the same amplitude during the whole pulse. The real one presents significant variations due to the fact that the acoustic path response between loudspeaker and microphones is frequency selective. This is an indicator of the present multiple path phenomenon, diffractions, etc. The same type of variations are found in the received signal from the other loudspeakers.

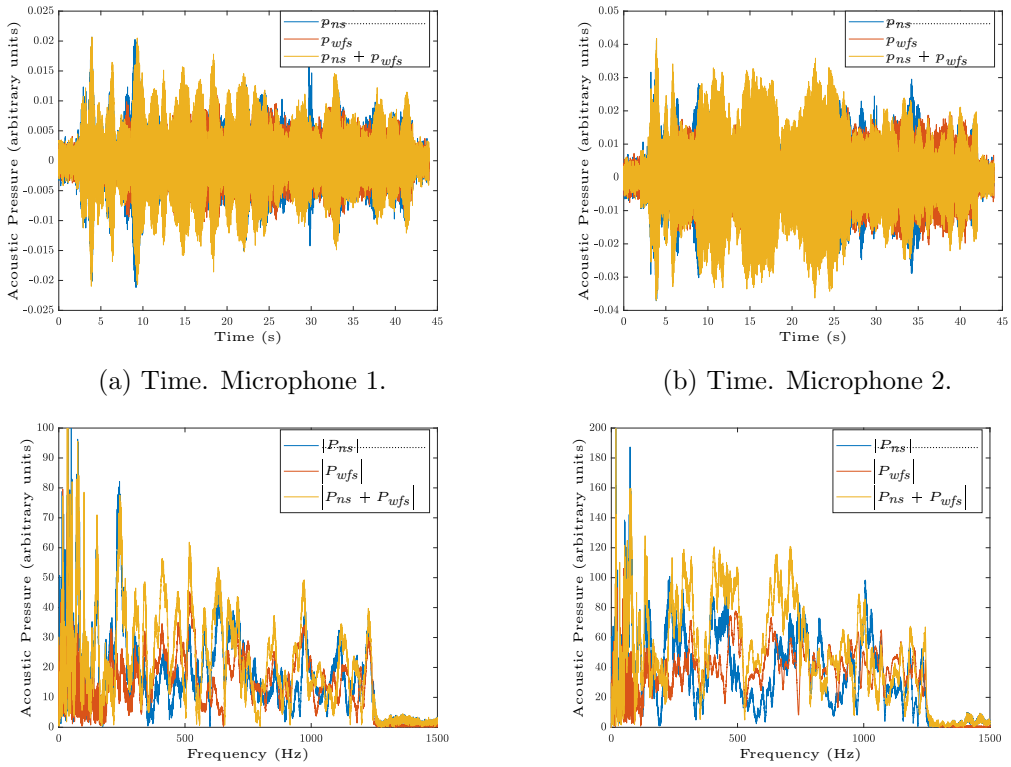
When playing the noise source and WFS signals simultaneously, the received signals (in comparison with the ones received only from the noise source) are shown in Figure 4.4. As expected, no cancellation is achieved, since the real acoustic path responses $\boldsymbol{A}(f)$ are too different from the ideal ones.

It could be possible that noise cancellation was spoiled because there is a sound vol-



(a) Time (b) Frequency

Figure 4.3: Received signal from noise source



(a) Time. Microphone 1.

(b) Time. Microphone 2.

(c) Frequency. Microphone 1.

(d) Frequency. Microphones 2.

Figure 4.4: Received signal

ume mismatch between the noise loudspeaker and the rest of loudspeakers. Unlike the secondary array loudspeakers, which were all adjusted to have similar behaviour (same volume and response, though in reality there are of course inevitable variations), the noise loudspeaker volume can be independently adjusted manually by means of a potentiometer. This introduces an unknown variable that definitely affects noise cancellation, even in the ideal model. Fortunately, this is relatively easy to model and compensate.

Basically, regarding the volume of the source as a separate variable when dealing with real loudspeakers, we transform Equation 4.1 in:

$$\mathbf{P} = \beta_{ns} \boldsymbol{A}_{NS} \mathbf{S}^{(ns)}(f) + \beta_{wfs} \boldsymbol{A}_{WFS} \mathbf{S}^{(wfs)}(f), \tag{4.3}$$

where β_{ns} and β_{wfs} are scalar real numbers that represent the volume of the noise and the secondary loudspeakers respectively.

In order to perform cancellation, we must compensate for this volume difference by multiplying the amplitude of secondary signals by $\Psi_g = \beta_{ns}/\beta_{wfs}$. It could be estimated in different ways. A simple one is measuring the field generated by the noise loudspeaker and the secondary array separately. The result are two acoustic pressure signals for each microphone: $p_{ns}(t)$ and $p_{wfs}(t)$. Then, we must minimize the energy of the sum of both signals:

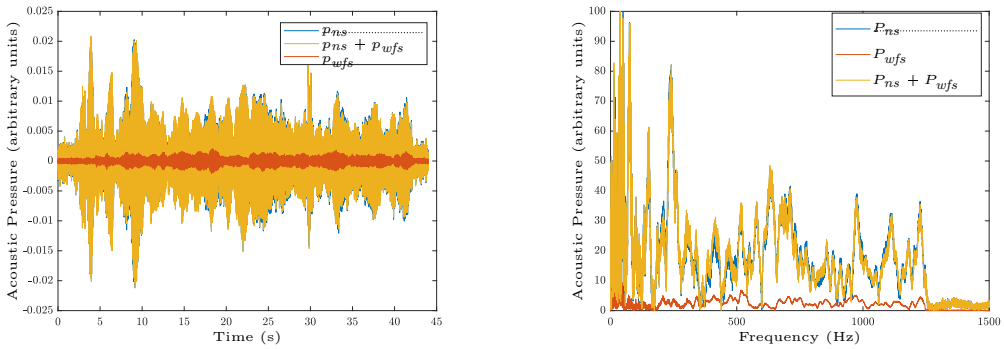
$$\Psi_g = \min_{\Psi} \int (p_{wfs}(t)\Psi + p_{ns}(t))^2 \mathrm{d}t. \tag{4.4}$$

In reality, this signals are actually discrete vectors with as many elements as recorded

samples ($\mathbf{p}_{wfs}[n]$ and $\mathbf{p}_{ns}[n]$, where n is the index of the sample). Hence, previous optimization becomes a simple vector operation.

$$\Psi_g = \min_{\Psi} \|\mathbf{p}_{wfs}\Psi + \mathbf{p}_{ns}\|^2 = -\frac{\langle \mathbf{p}_{wfs}, \mathbf{p}_{ns} \rangle}{\|\mathbf{p}_{wfs}\|^2}. \tag{4.5}$$

Let's notice that this estimation get's a value for each microphone. As we already know that for low frequencies cancellation is not good even in simulated scenarios, and that above the spatial aliasing frequency the synthesis is not correct, we have optimized considering just the interval that transmits frequencies between 500Hz and 850Hz. In Figure 4.5 there is the resulting signal after volume correction (the result is similar for the other microphone). The total signal is actually very similar to the contribution from just the noise source. This is not strange if we take in account that the volume correction factor we have applied is very small ($\Psi_g = 0.1518$), so the contribution from the secondary array is minimal. What seems to have happened is that the correlation between $\mathbf{p}_{wfs}$ and $\mathbf{p}_{ns}$ is too small. In the ideal scenario, the correlation coefficient would be very close to one, so the estimated value would actually very close to β_{ns}/β_{wfs}. However, both variables are so uncorrelated that the best way of minimizing the total power of the sum of both is making the secondary signals very small. This means that the cause of the low cancellation levels is not the volume mismatch, but probably a combination of previously mentioned phenomena (reverberation, diffractions, irregular directivity patterns, etc.).



(a) Time (b) Frequency

Figure 4.5: Received signal after volume correction. Microphone 1. $\Psi_g = 0.1518$.

Chapter 5

Summary and future research

Wave Field Synthesis (WFS) theory was developed in the 1990s as a new sound reproduction paradigm. Unlike stereophonic techniques, that can produce a sound image similar to that of the original sources on a small area or sweet-spot, WFS aimed at the synthesis of sound wave fronts over a volume or area. WFS theory is derived from Kirchhoff-Helmholtz equation. It states that the wave field produced by a sound source (primary source) inside a free source volume can be perfectly replicated by a surface distribution of monopole and dipole sources (secondary sources) that enclose the volume. In order to do so, secondary sources must reproduce signals that are directly proportional to the surface acoustic pressure and its directional gradient. This information can be derived from the surface geometry, and position of the primary source, it's directivity, and the signal it transmits.

A simplification can be made if the mentioned surface is a plane, and the primary sources are on one side of that plane. In that case, either a distribution of monopoles or dipoles is sufficient to replicate a field on the other side of the plane (Rayleigh I and II integrals). If, in addition, the sound sources as well as the listeners are on the same plane, just a line distribution of secondary sources, either monopoles or dipoles, is required (Rayleigh 2.5D I and II integrals). A linear array of loudspeakers can actually be approximately modelled by this last theoretical scenario and, indeed, it is the most typical type of WFS implementation in commercial applications and research so far.

Of all applications, this study has been focused on Active Noise Control (ANC). It refers to the idea of using loudspeakers to create sound fields that interfere destructively with the field generated by noise sound sources. If the sound signal and location of a noise source are known, WFS would allow us to synthesize a replica of the noise field, but with opposite sign, so it will produce noise cancellation over the area that the WFS system covers.

In order to study the possibility of using WFS based ANC in a real listening room as the one in the GTAC facilities, a series of simulations was carried out. The starting point was a simplified model were loudspeakers are substituted by ideal monopole sources and free-space conditions are assumed. At first, the main limitation we have found is the necessity of implementing a prefilter for the virtual noise source signal with frequency response $H = \sqrt{jk/2\pi}$. The WFS available literature that does not refer specifically to

ANC, does not mention that filter or understate the importance of it. Under subjective perception criteria, this filter does not have a big effect on source location, coloration or spaciousness when synthesizing virtual sources because the human auditory system is tolerant to some types of distortions. However, if the purpose is to interfere destructively with another field, accuracy is critical. Slight phase inaccuracies can completely undermine system performance.

This prefilter is usually implemented as a FIR digital filter, but the ideal response is anticausal, so it makes necessary to delay the generation of secondary source signals by a given amount of time. The higher the FIR order (and therefore precission), the longer the delay time needed. In practice, this sets a trade-off between the system performance and the distance between the loudspeaker array and the noise source.

Since the real system is located inside GTAC's listening room, non-free space conditions were tested in simulations. A box shaped room was assumed as an idealized model of the actual listening room. A Matlab tool was used to generate acoustic path responses for different wall reflection coefficients. Of course, the higher the coefficient, the poorer the result was.

The effects of truncation were also studied. It was proven that the bad performance that we systematically got at low frequencies was caused by the fact that the length of the loudspeaker array is finite. A simple scenario was used. It was formed by a finite line of secondary sources, a primary source located at an infinite distance and a centred point of measure. It was shown that there is an almost linear relation between the distance from the measure point to the secondary line, and the minimum frequency at which the performance starts to converge.

In measures, we were able to prove that, as simulations of highly reverberant rooms suggested, it was not possible to achieve high noise cancellation levels. So, due to this technical limitation, a practical demonstration of active noise cancellation based on WFS indoors was not possible to be carried out.

5.1. Future research

During this work, multiple questions remained unanswered. The design of the prefilter is an aspect that should be studied more thoroughly because it establishes a critical constraint. Some possible alternatives were mentioned. It is especially interesting the use of a IIR filter design as proposed in [18] since it would drastically reduce the required filter order, and hence, the delay of the system.

Due to the reverberant nature of the listening room, we could not perform a good demonstration of noise cancellation with WFS. Some examples of experimental measures outdoors are available in the literature, but most of them use linear arrays, none with an octagon shaped array and the dimensions we use. Experiments in an environment that resembles more to free-space (outdoors, anechoic chamber...) are interesting. They would provide new insights that can't be drawn from computer simulations. Only after understanding the difficulties that arise in such experimental conditions, would be profitable to try the system in more practical ones. Moving directly from idealized sim-

ulations to measures in a real environment, make appear too many unknowns that are complicated to analyse and understand. A step by step process would be more reliable. For example, the issue of ground reflections has been addressed as a separate problem that can be compensated with an additional filter [17].

Regarding truncation issues, it is convenient to explore techniques other than tapering in order to overcome bad performance at low frequencies. Other geometries for the secondary source distribution may produce different artefacts, like circular arrays or arc arrays, which have received some attention in literature. Another possibility can be to use different secondary signal processing strategies for low and mid-high frequencies. For example, apart from WFS, the other most known sound field synthesis method nowadays is Near-field Compensated Higher Order Ambisonics (NFC-HOA). Altough it is an approach theoretically restricted to spherical and circular secondary source distribution geometry and narrow-band synthesis [21], it has been analytically proved that WFS is a generalized, high-frequency/far-field approximation of NFC-HOA [18]. Hence, it is not strange that at low frequencies/near-field conditions, NFC-HOA can show better accuracy than WFS.

Bibliography

[1] Karlheinz Brandenburg, Sandra Brix, and Thomas Sporer. WAVE FIELD SYNTHESIS. (1):1–4, 2009. (Cited in pages 1, 2, 3 and 4.)

[2] Icons made by "pixel perfect" from www.flaticon.com, and www.hooley.ie. (Cited in page 1.)

[3] A. J. Berkhout and Diemer de Vries. Acoustic holography for sound control. In *Audio Engineering Society Convention 86*, Mar 1989. (Cited in page 3.)

[4] Evert Walter. Start and Beeld en Grafisch Centrum Technische Universiteit Delft). *Direct sound enhancement by wave field synthesis*. s.n.], 1997. (Cited in pages 3, 9, 10, 12 and 34.)

[5] Edwin Verheijen. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1997. (Cited in pages 3, 6, 8, 9, 12 and 15.)

[6] P Vogel. *Application of Wave Field Synthesis in Room Acoustics*. PhD thesis, Delft University of Technology, 1993. (Cited in pages 3, 9, 15 and 16.)

[7] Systematic Musicology. *Springer Handbook o Systematic Musicology*. (Cited in pages 3 and 4.)

[8] Amir Avni, Jens Ahrens, Matthias Geier, Sascha Spors, Hagen Wierstorf, and Boaz Rafaely. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. *The Journal of the Acoustical Society of America*, 133(5):2711–2721, 2013. (Cited in page 4.)

[9] Jens Ahrens. Challenges in the Creation of Artificial Reverberation for Sound Field Synthesis: Early Reflections and Room Modes. *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, (April):3–5, 2014. (Cited in page 4.)

[10] S. M. Kuo and D. R. Morgan. Active noise control: a tutorial review. *Proceedings of the IEEE*, 87(6):943–973, June 1999. (Cited in page 5.)

[11] A Kuntz and R Rabenstein. 12th European Signal Processing Conference (EUSIPCO), Vienna , Austria , September 2004 AN APPROACH TO GLOBAL NOISE CONTROL BY WAVE FIELD SYNTHESIS. *Most*, (September):2–5, 2004. (Cited in page 5.)

[12] Alessandro Lapini, Massimiliano Biagini, Francesco Borchi, Monica Carfagni, and Fabrizio Argenti. Active Noise Control for Pulse Signals by Wave Field Synthesis. pages 883–887, 2016. (Cited in pages 5 and 23.)

[13] Michele Zanolin, Paolo Podini, Angelo Farina, Stefano De Stabile, and Paolo Vezzoni. Active Control of Noise by Wave Field Synthesis. (Cited in page 5.)

[14] Antonio Cano Morcillo, Miguel Ferrer Contreras, María De Diego Antón, and González Salvador. PERFORMING ACTIVE NOISE CONTROL WITH WAVE FIELD SYNTHESIS. (July):12–16, 2015. (Cited in page 5.)

[15] A.J. Berkhout. Applied seismic wave theory. (Cited in page 6.)

[16] Krzysztof Czy, Communications Signal, Processing Group, and Multimedia Applications. Modeling of 2-D sound pressure distribution in laboratory enclosure. 2011. (Cited in pages 15 and 26.)

[17] Alessandro Lapini, Francesco Borchi, Monica Carfagni, and Fabrizio Argenti. Application of wave field synthesis to active control of highly non-stationary noise. *Applied Acoustics*, 131(July 2017):220–229, 2018. (Cited in pages 20 and 42.)

[18] Frank Schutz. Sound Field Synthesis for Line Source Array Applications in Large-Scale Sound Reinforcement. *Dissertation*, 2015. (Cited in pages 22, 25, 41 and 42.)

[19] Emanuel Habets. Room impulse response generator. https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator, 2010. (Cited in page 25.)

[20] Loudspeaker and microphone setup at gtac listening room. http://www.gtac.upv.es/room.php. (Cited in page 26.)

[21] Jens Ahrens. *Applications of Sound Field Synthesis*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. (Cited in page 42.)