HRTF-Based Sound Field Interpolation in the Presence of a Human Head

Anton Björkman

Master's thesis
2023:E10

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Contents

Acknowledgements

I would like to thank both of my supervisors, David and Andreas, for their support and guidance throughout this thesis. Not only has this been a great learning opportunity, it has also been the most enjoyable experience I have had throughout my studies thus far. Thank you both for all the hours we have spent discussing not only how to deal with all problems, but also further implementations that could be made. I would also like to express my gratitude to David for devoting a full day to setting up the measurements for the real data.

Abstract

There are different ways of evaluating the sound pressure in a sound field. Doing so is an essential part of several applications, such as active sound control and voice analysis. One way of evaluating the sound pressure in a sound field is by measuring a few discrete points in the field, combined with an interpolation method to find the sound pressure for the remaining field. Recently, it has been shown that kernel ridge regression (KRR) can accurately interpolate a sound field from discrete points in space. This report examines the effect of inaccurate measurements for the resulting estimated sound field by adding noise in all measurements. Further, to allow for accurate estimation of how a sound is perceived at a given location, one not only needs to form an estimate of the expected sound field at the location, but also to examine how the presence of the person's head will shape the resulting field. Thus, the use of a pre-calibrated head-related transfer function (HRTF) in combination with the interpolation technique to interpolate the in-ear sound field is also proposed. Using both simulated and anechoic audio data, the effect of different kernels as well as the added noise for the sound field estimate is studied, which is then illustrating how the combined sound field interpolation technique allows for an improved estimate of the sound field, for all frequencies. Finally, the combined sound field interpolation is shown to be more effective than the KRR method, for both the simulated and the anechoic audio data, with the presence of a person in the sound field.

Populärvetenskaplig sammanfattning på svenska

Uppskattning av ljudet mellan mikrofonpositioner, med fokus på känsligheten mot mätfel och brus

Arbetet presenterar en metod som uppskattar ljud så som en människa uppfattar det i ett rum. Metoden kan användas i exempelvis applikationer med augmented reality och brusreducering.

Att kunna uppskatta ljud i ett ljudfält är en nödvändighet för många applikationer. I exempelvis aktivt brusreducerade hörlurar får man in en signal som man vill tysta genom att skicka ut en motsatt signal, som i teorin tar ut den ursprungliga signalen och resulterar i att inget ljud når dina öron. Att så exakt och effektivt som möjligt kunna skatta hur ljudet beter sig i ljudfältet är därför nödvändigt för att få till effektiv brusreducering. I detta arbetet används en ny metod för att skatta ljud i ett ljudfält, där effektiviteten av metoden utvärderas med avseende på brus och mätfel. Utöver detta implementeras även en ny metod som skattar hur närvaron av en människa påverkar det resulterande ljudfältet.

Det finns olika sätt att skatta ljudtrycket i ett ljudfält. Ett sätt är att placera ut mikrofoner i fältet som sedan används för att skatta ljudet i övriga fältet, vilket är metoden i denna rapport. Både uppmätt data från ett ekofritt rum och simulerad data utvärderas, där vi i den simulerade datan lagt till både mätfel och brus i alla mätningar. Tanken med projektet är att kunna använda resultaten som en referenspunkt för framtida implementeringar av, för att använda det tidigare exemplet, aktiv brusreducering.

I en större skala, där man vill reducera ljudnivån i ett helt rum, blir problemet mer komplicerat. Mellan ett par hörlurar och dina öron är det inte mycket i vägen för ljudet att färdas fritt. I ett rum kan det å andra sidan finnas en mängd olika objekt utspridda; inte minst i form av en människa. En människas närvaro i rummet kommer påverka det resulterande ljudfältet, och av särskilt intresse är ljudfältets påverkan av människans närvaro när det gäller ljudet i människans eget öra. Om ljudet i hela rummet som helhet dämpas men ljudet i människans öra blir högre så blir ljudreduceringen i praktiken meningslös. Därför har vi implementerat en metod som tar hänsyn till en människas närvaro i rummet, och som då uppskattar ljudet i människans båda öron. Den implementerade metoden som presenteras i projektet är en blandning av den tidigare nämnda metoden och en funktion som beskriver hur ljudet förändras för en människa i rummet, head-related transfer function (HRTF).

Resultatet av experimenten med avseende på brus och mätfel presenteras först, där den implementerade metoden sedan utvärderas. Den implementerade metoden visar sig förbättra skattningen av ljudet i örat för alla använda frekvenser, jämfört med den

tidigare metoden.

HRTF-Based Sound Field Interpolation in the Presence of a Human Head

by Anton Björkman

LUND
UNIVERSITY

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Master's thesis, Mathematical Statistics for Engineers
Supervisors: Andreas Jakobsson, David Sundström
Examiner: Johan Lindström

Glossary

Term	Definition
KRR	Kernel ridge regression
u	Sound pressure in frequency domain
u^t	Sound pressure in time domain
λ	Regularization parameter for kernel ridge regression
β	Directional regularization parameter, directional Helmholtz kernel
Ω	The acoustic environment wherein the sound pressure is measured and interpolated
$\mathcal{H}$	Function space
$\|\cdot\|_{\mathcal{H}}$	Norm over the function space $\mathcal{H}$
$\|\cdot\|$	Euclidean norm
$(\cdot)^*$	Complex conjugate
r_m	Location of the mth microphone
s_m	Sound pressure recorded by the mth microphone
$\hat{u}$	Estimated sound pressure, given by the interpolation method
k	Wavenumber
c	Sound velocity
j_ν	Spherical Bessel function of order ν of the first kind
J_ν	Bessel function of order ν of the first kind
$Y_{\nu,\mu}$	Spherical harmonic function of order ν and degree μ
$P_{\nu,\mu}$	Legendre function of order ν and degree μ
κ	Kernel function
$\boldsymbol{\eta}$	Direction of which the sound is propagating
SNR	Signal-to-noise ratio, measures the impact of noise on the signal
SINAD	Signal-to-noise and distortion ratio, measures the impact of noise and distortion on the signal
$\bar{\boldsymbol{r}}_h$	Position of the sound source in relation to the head
HRTF	Head-related transfer function
HRIR	Head-related impulse response
$H_{\cdot,0}$	Head-related transfer function, the one used
$h_{\cdot,0}$	Head-related impulse response, the one used
$u_{\cdot}^t$	In-ear sound pressure, time domain
$\hat{u}_{\cdot}^t$	Estimated in-ear sound pressure, time domain

Chapter 1

Introduction

With recent development of spatial audio equipment, there has been a growing demand to accurately describe how sound propagates and is perceived. There are different ways to allow for an accurate representation of a sound field, one being an interpolation technique. Interpolating a sound field typically involves kernel ridge regression (KRR), with the Gaussian kernel being the most widely used today, which approximates the sound field roughly as a sum of Gaussian functions. Another approach to this is the use of a Bayesian framework [1]. In the first part of this report, KRR with a new type of kernel proposed by Ueno *et al.* [2] is investigated and compared to the Gaussian kernel [3], primarily to evaluate their robustness. A variation of the kernel proposed by Ueno *et al.*, which adds an assumed prior knowledge of the direction to the sound source, is also investigated [4, 5]. This is done with simulations as well as experiments consisting of real measurements. Interpolating a sound field has various applications, such as active sound control [6–9], acoustic environment analysis [10], voice analysis [11], and speech enhancement [12].

With increasing interest in three-dimensional applications such as virtual reality environments, augmented reality environments as well as video games, there has also been a growing demand for more natural and realistic experiences. In order to get a more immersive experience, being able to get an accurate representation of sound in a natural way is of great importance. Although the sound field interpolation methods presented above offer highly accurate estimates of the resulting sound field, the presence of the human head at the location of interest will shape the sound field, modifying the in-ear sound pressure from that of the interpolated sound field at the location. A common approach to account for this is to use a pre-calibrated head-related transfer function (HRTF), describing the spatial filtering effect resulting from the presence of the head and torso. As a consequence, several methods employing

the use of HRTFs have been introduced [13, 14], enhancing the measured HRTFs by interpolating between the discrete measurement positions [15–18] as well as accurately estimating the direction of sound [19].

The primary focus of this report is the sound field interpolation with KRR. Prior work with KRR and sound field interpolation has largely been focusing on active noise control (ANC) [4, 7, 20], where KRR allows the ANC to account for the interpolated field as opposed to the more commonly used approach, where only the sound at the recorded microphones is used. For smaller applications, such as active noise cancelling headphones, there is not a lot of space between the headphones and the ear canal; however, in a larger scale, if one for instance would want to perform ANC in an entire room, there could be plenty of objects altering the propagation of the sound. An object of particular interest is the human body. Lowering the noise level in the entire room would not necessarily be of interest if the purpose of the ANC is to lower the noise for the human in the sound field; if the overall noise level is lowered but at the same time increased for the person in the sound field, the noise reduction is essentially useless. Thus, a method that estimates the sound pressure as experienced by a human in the field is proposed, by combining KRR with pre-calibrated HRTF measurements, thereby enabling an improved estimate of the in-ear sound field, which will constitute the second part of this report. While prior work with HRTFs has mostly been working under the assumption that the in-ear sound pressure is available, this work is instead trying to estimate the in-ear sound pressure based on the free-field estimate of the sound field, specifically by interpolating the free-field estimate of the sound pressure to the center of the head, given measurements by microphones around the head, where the resulting interpolated sound field will be filtered through a pre-calibrated HRTF. In this formulation, the use of the HRTF implicitly assumes that the sound propagates from a single direction. To examine how this assumption affects the results, we here also examine how the used directionally weighted kernel affects the interpolation under this assumption.

1.1 Intuition behind the limitations of the Gaussian kernel

The Gaussian kernel that is widely used today has its limitations for sound field interpolation, as a sum of Gaussian functions will not necessarily capture the periodicity of a wave. An example of this can be seen in Figure 1.1, and shows how the Gaussian kernel does not capture the periodicity when the microphones are too far apart. This is an extreme example of microphones being evenly spread out to precisely miss the periodicity, but should give us some intuition behind the limitations of the Gaussian kernel. As can be seen, the Gaussian kernel gives us a fairly good estimate of the sound pressure very close to the microphones, but not necessarily a certain distance

away from the microphones. What can be seen is also that if only positive values of the sound pressure is recorded, the interpolation will only contain positive values.

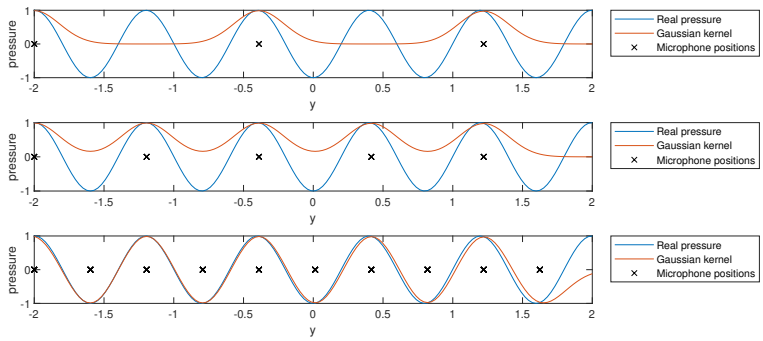

Figure 1.1: Simulation using regularization parameter $\lambda = 10^{-2}$, a frequency of 426, using 3, 5, and 10 microphones, respectively, and is done in a 2D space with the microphones being spread out evenly along the line $x = 0.0134$. The plot is examined along the line $x = 0.0134$ where the microphones are located, with $y \in [-2, 2]$. The wave used is a plane wave propagating in the y-direction. The grid is created with incremental changes in x and y of 0.05 m. The parameters will be explained later on in the report.

A plot for the interpolated sound field, as well as the real sound field, can be seen in Figure 1.2, using the same five microphone setup as for Figure 1.1.

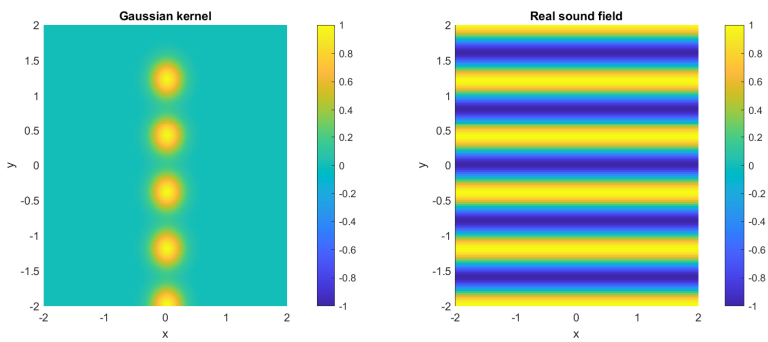

Figure 1.2: Simulation using regularization parameter $\lambda = 10^{-2}$, a frequency of 426, using 5 microphones, and is done in a 2D space with the microphones being spread out evenly along the line $x = 0.0134$. The wave used is a plane wave propagating in the y-direction. The grid is created with incremental changes in x and y of 0.05 m. In the left plot, the interpolated sound field using the Gaussian kernel is shown, and in the right plot the real sound field is shown. The parameters will be explained later on in the report.

The calculations necessary for this interpolation will be explained later on in the report (see the appendix, section 6.1, for the 2D case) - the plots are just shown to give some intuition for the idea behind the possible improvements that a new kernel can bring

to the sound field interpolation problem.

Chapter 2

Methodology

Consider an acoustic environment, where the sound pressure is measured using M omnidirectional microphones and one seeks to determine the sound pressure. We denote the part of the acoustic environment where we want to find the resulting sound field as $\Omega \subseteq \mathbb{R}^3$, which is open and simply connected. Let $u(\boldsymbol{r}, \omega)$ denote the sound pressure for the temporal frequency ω at location $\boldsymbol{r} \in \Omega$ in the frequency domain. Let r_m and s_m denote the location of and the sound pressure at the mth microphone, respectively, for $m = 1, \ldots, M$. For the first part of this report, the frequency ω will be disregarded in the text as only one frequency will be investiagated at a time.

For the simulations, the microphones will be spread out evenly in Ω, unless otherwise specified. There will also be point sound sources placed outside of this region, from where a sound wave will be simulated, propagating spherically, which will then finally be interpolated, given the recorded signals of the microphones. The interpolation problem will be formulated as a regularized least square problem (see also [2]),

$$\hat{u}(\boldsymbol{r}) = \underset{u \in \mathcal{H}}{\operatorname{argmin}} \sum_{m=1}^{M} |u(\boldsymbol{r}_m) - s_m|^2 + \lambda \|u(\boldsymbol{r})\|_{\mathcal{H}}^2, \tag{2.1}$$

where $\hat{u}$ is the estimated sound pressure, $\mathcal{H}$ is some function space, $\|\cdot\|_{\mathcal{H}}$ the norm over $\mathcal{H}$, and λ is a user-specified regularization parameter. The goal is to find $\hat{u}(\boldsymbol{r})$ for all $\boldsymbol{r} \in \Omega$, given the measurements s_m.

2.1 The Helmholtz equation

This section is used as a short motivation for the choice of Hilbert space used for the kernels in the following section, as the Helmholtz equation will play a crucial role for the kernels used. The following derivation is mostly taken from [21]. Starting with the wave equation, we have

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) u(\boldsymbol{r}, t) = 0. \tag{2.2}$$

Assuming that the solution to this equation is space-time-separable, the solution to the wave equation can be re-written as

$$u(\boldsymbol{r}, t) = A(\boldsymbol{r})T(t). \tag{2.3}$$

Now substituting (2.3) into (2.2) yields an equation on the form of

$$\frac{\nabla^2 A(\boldsymbol{r})}{A(\boldsymbol{r})} = \frac{1}{c^2 T(t)}\frac{d^2 T(t)}{dt^2}. \tag{2.4}$$

The left hand side only depends on $\boldsymbol{r}$ and the right hand side only depends on t, which implies that both sides are equal to the same constant value. By letting both sides be equal to the constant value $-k^2$, two expressions can be found as

$$\frac{\nabla^2 A(\boldsymbol{r})}{A(\boldsymbol{r})} = -k^2, \tag{2.5}$$

$$\frac{1}{c^2 T(t)}\frac{d^2 T(t)}{dt^2} = -k^2, \tag{2.6}$$

and after some rearrangement, the expressions become

$$(\nabla^2 + k^2)A(\boldsymbol{r}) = 0, \tag{2.7}$$

$$(\frac{d^2}{dt^2} + \omega^2)T(t) = 0, \tag{2.8}$$

where $\omega = kc$, (2.7) is the homogeneous Helmholtz equation and (2.8) is a second-order ordinary differential equation. Consequently, under the assumption that the sound wave is space-time-separable, the spatial component will be a solution to the homogeneous Helmholtz equation.

2.2 Kernel ridge regression

For our interpolation problem, (2.1), the solution depends on the normed space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$. In our case, a reproducing kernel Hilbert space will be used, which turns the problem

into a kernel ridge regression problem as presented in [3]. The representer theorem states that the solution to (2.1), then, may be expressed as [22]

$$u(\boldsymbol{r}) = \sum_{m=1}^{M} \alpha_m \kappa(\boldsymbol{r}, \boldsymbol{r}_m), \tag{2.9}$$

where $\kappa : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ denotes a positive-definite kernel function on the normed space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ and $\alpha_m \in \mathbb{C}$ the corresponding magnitude at the location of the mth microphone. By introducing

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \dots & \alpha_M \end{bmatrix}^\mathsf{T}, \tag{2.10}$$

$$\boldsymbol{s} = \begin{bmatrix} s_1 & \dots & s_M \end{bmatrix}^\mathsf{T}, \tag{2.11}$$

$$\boldsymbol{K} = \begin{bmatrix} \kappa(\boldsymbol{r}_1, \boldsymbol{r}_1) & \dots & \kappa(\boldsymbol{r}_1, \boldsymbol{r}_M) \\ \vdots & \ddots & \vdots \\ \kappa(\boldsymbol{r}_M, \boldsymbol{r}_1) & \dots & \kappa(\boldsymbol{r}_M, \boldsymbol{r}_M) \end{bmatrix}, \tag{2.12}$$

and inserting (2.9) into (2.1), yields [2]

$$\boldsymbol{\alpha} = (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{s}, \tag{2.13}$$

such that the interpolated sound field finally is given by

$$\hat{u}(\boldsymbol{r}) = \left((\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{s}\right)^\mathsf{T} \boldsymbol{\kappa}(\boldsymbol{r}), \tag{2.14}$$

where $\boldsymbol{\kappa}(\boldsymbol{r}) = [\kappa(\boldsymbol{r}, \boldsymbol{r}_1), \dots, \kappa(\boldsymbol{r}, \boldsymbol{r}_M)]^\mathsf{T}$. The used kernel function κ may be selected in different ways; in this work, we will examine three alternative formulations, all depending on the chosen normed space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$. These kernels will be defined in the following sections.

2.2.1 The Helmholtz kernel

Under the assumption that there are no sound sources in Ω, the sound pressure u satisfies the homogeneous Helmholtz equation[1] [23]

$$(\Delta + k^2)\, u(\boldsymbol{r}, \omega) = 0, \tag{2.15}$$

where $\Delta = \nabla^2$ is the Laplace operator and $k = \frac{\omega}{c}$ is the wavenumber at sound velocity c. Shown in [24], any solution to this equation can be uniquely expanded

[1] There are certain assumptions made here, such as the medium being lossless and isotropic; essentially, the medium is modelled as vacuum.

around a point $\boldsymbol{r}_0 \in \Omega$ as

$$u(\boldsymbol{r}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{u}_{\nu,\mu}(\boldsymbol{r}_0)\varphi_{\nu,\mu}(\boldsymbol{r} - \boldsymbol{r}_0), \tag{2.16}$$

where $\tilde{(\cdot)}$ denotes the expansion coefficient at the specified location, in the expression above being r_0, and the basis function $\varphi_{\nu,\mu}(\cdot)$ is defined as

$$\varphi_{\nu,\mu}(\boldsymbol{r}) = \sqrt{4\pi} j_\nu(kr) Y_{\nu,\mu}(\hat{\boldsymbol{r}}). \tag{2.17}$$

Here, $\|\cdot\|$ is the Euclidean norm, $r = \|\boldsymbol{r}\|$ and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$. Furthermore, $j_\nu(\cdot)$ is the νth-order spherical Bessel function of the first kind [25], defined as

$$j_\nu(z) = \sqrt{\frac{1}{2}\frac{\pi}{z}} J_{\nu+\frac{1}{2}}(z). \tag{2.18}$$

Here, J_ν is the Bessel function of the first kind [25], defined as

$$J_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{(-\frac{1}{4}z^2)^k}{k!\Gamma(\nu + k + 1)}, \tag{2.19}$$

and $Y_{\nu,\mu}(\cdot)$ is the spherical harmonic function of order ν and degree μ [23], defined as

$$Y_{\nu,\mu}(\hat{\boldsymbol{r}}) = \sqrt{\frac{2\nu+1}{4\pi}\frac{(\nu-\mu)!}{(\nu+\mu)!}} P_{\nu,\mu}(\cos\theta_{\hat{\boldsymbol{r}}}) e^{i\mu\phi_{\hat{\boldsymbol{r}}}}, \tag{2.20}$$

where $\theta_{\hat{\boldsymbol{r}}}$ and $\phi_{\hat{\boldsymbol{r}}}$ are the polar and azimuth angles of $\hat{\boldsymbol{r}}$, respectively, and $P_{\nu,\mu}$ is the associated Legendre function, defined by an unpleasantly long expression in [23, eq. (6.30)], with the result of

$$P_{0,0}(z) = 1. \tag{2.21}$$

With all of the above, the Hilbert space is defined as [2]

$$\mathcal{H} = \left\{ u \in \mathcal{C}^2(\Omega; \mathbb{C}), \text{ s.t. } (\Delta + k^2)u = 0 \mid \|u\|_{\mathcal{H}} < \infty \right\},$$

where $\mathcal{C}^2(\Omega; \mathbb{C})$ denotes the set of twice continuously differentiable functions from Ω to $\mathbb{C}$, and the inner product and norm are defined as [2]

$$\langle u, v \rangle_{\mathcal{H}} = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{u}_{\nu,\mu}(\boldsymbol{r}_0)^* \tilde{v}_{\nu,\mu}(\boldsymbol{r}_0), \tag{2.22}$$

$$\|u\|_{\mathcal{H}} = \sqrt{\langle u, u \rangle_{\mathcal{H}}}, \tag{2.23}$$

where $(\cdot)^*$ denotes the complex conjugate. For this Hilbert space, the kernel is set as

$$\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2) = j_0(k\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|). \tag{2.24}$$

From here, we want to show that this kernel is a reproducing kernel, meaning that the inner product between the kernel and any function in the Hilbert space is equal to the function itself,

$$\langle \kappa(\boldsymbol{r}), u(\boldsymbol{r}) \rangle = u(\boldsymbol{r}), \tag{2.25}$$

where we let $\kappa(\boldsymbol{r}') = \kappa(\boldsymbol{r}, \boldsymbol{r}')$. We now use (2.20) and (2.21) to find that

$$Y_{0,0}(\hat{\boldsymbol{r}}) = \sqrt{\frac{1}{4\pi}} P_{0,0}(\cos\theta_{\hat{\boldsymbol{r}}}) e^0 = \sqrt{\frac{1}{4\pi}}. \tag{2.26}$$

By then combining (2.17) and (2.26), the basis function of order 0 and degree 0 can be expressed as

$$\begin{aligned} \varphi_{0,0}(\boldsymbol{r}_1 - \boldsymbol{r}_2) &= \sqrt{4\pi} j_0(k\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|) Y_{0,0}(\boldsymbol{r}) && (2.27)\\ &= j_0(k\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|) && (2.28)\\ &= \kappa(\boldsymbol{r}_1, \boldsymbol{r}_2), && (2.29) \end{aligned}$$

Here, we use that [26]

$$\varphi_{n,m}(\boldsymbol{r}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{\varphi}_{\nu,\mu}(\boldsymbol{r}') \varphi_{\nu,\mu}(\boldsymbol{r} - \boldsymbol{r}'), \tag{2.30}$$

for any arbitrary point $\boldsymbol{r}' \in \Omega$, and by letting $\boldsymbol{r}' = \boldsymbol{r}$ in combination with (2.29), yields

$$\begin{aligned} \kappa(\boldsymbol{r}) &= \varphi_{0,0}(\boldsymbol{0}) && (2.31)\\ &= \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{\varphi}_{\nu,\mu}(\boldsymbol{r}) \varphi_{\nu,\mu}(\boldsymbol{r} - \boldsymbol{r}), && (2.32) \end{aligned}$$

which leaves us with a unique solution when

$$\begin{aligned} \tilde{\kappa}_{\nu,\mu}(\boldsymbol{r}) &= \tilde{\varphi}_{\nu,\mu}(\boldsymbol{r}) && (2.33)\\ &= \delta_{\nu,0}, && (2.34) \end{aligned}$$

where $\delta_{\nu,\mu}$ denotes the Kronecker delta. From here, we use the inner product (2.22). It can be shown that this expression is invariant of coordinate rotation and translation

[26], leaving us with a new expression expanded around the arbitrary point $\boldsymbol{r}' \in \Omega$, as

$$\left\langle \kappa(\boldsymbol{r}'), u(\boldsymbol{r}') \right\rangle = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{\kappa}_{\nu,\mu}(\boldsymbol{r}')^* \tilde{u}_{\nu,\mu}(\boldsymbol{r}'). \tag{2.35}$$

By now substituting $\boldsymbol{r}' = \boldsymbol{r}$ into (2.35), the inner product becomes

$$\langle \kappa(\boldsymbol{r}), u(\boldsymbol{r}) \rangle = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \delta_{\nu,0}^* \tilde{u}_{\nu,\mu}(\boldsymbol{r}) = \tilde{u}_{0,0}(\boldsymbol{r}). \tag{2.36}$$

Using (2.18) and (2.19), it can be seen that spherical Bessel functions of all orders except for order 0 will have the function value 0 at $\mathbf{0}$, and function value 1 for order 0. Using this, and letting $\boldsymbol{r} = \mathbf{0}$ in (2.17), we find

$$\varphi_{\nu,\mu}(\mathbf{0}) = \sqrt{4\pi} j_\nu(0) Y_{\nu,\mu}(\hat{\mathbf{0}}), \tag{2.37}$$

which in combination with (2.26) yields $\varphi_{\nu,\mu}(\mathbf{0}) = \delta_{\nu,0}$. Combining this with (2.30), $u(\boldsymbol{r})$ is expanded around $\boldsymbol{r}$ instead of $\boldsymbol{r}_0$, which gives us

$$u(\boldsymbol{r}) = \tilde{u}_{0,0}(\boldsymbol{r}) \varphi_{0,0}(\mathbf{0}) = \tilde{u}_{0,0}(\boldsymbol{r}), \tag{2.38}$$

and combining (2.36) with (2.38), we finally get $\langle \kappa(\boldsymbol{r}), u(\boldsymbol{r}) \rangle = u(\boldsymbol{r})$, which means that κ is a reproducing kernel of $\mathcal{H}$. This kernel will be referred to as the Helmholtz kernel in the rest of the report. Note that this, in combination with (2.9), gives us that the solution will be a sum of waves.

2.2.2 The directional Helmholtz kernel

Similarly to what was done for the Helmholtz kernel, we define our Hilbert space as

$$\mathcal{H} = \left\{ \frac{1}{2\pi} \int_{S_{(0,1)}} \tilde{u}(\boldsymbol{\eta}) \exp\left[i \boldsymbol{k}^\intercal \boldsymbol{r} \right] \mathrm{d}\boldsymbol{\eta} \;\middle|\; \tilde{u}(\boldsymbol{\eta}) \in L_2 \right\}, \tag{2.39}$$

where L_2 denotes the space of square-integrable functions and $S(0,1)$ denotes the unit sphere, again under the assumption that there are no sound sources in Ω. Although this Hilbert space does not include all solutions of the homogeneous Helmholtz equation, the solution of said equation can be approximated by functions in this Hilbert space in the sense of uniform convergence on compact sets [27]. As proposed in [27], a directional variation of the Helmholtz kernel is introduced, which incorporates prior knowledge of the direction from which the sound is propagating.

The solution of the homogeneous Helmholtz equation can be approximated by a sum of plane waves, as the Herglotz wave function [28]

$$u(\boldsymbol{r}) = \int_{S(0,1)} \tilde{u}(\bar{\boldsymbol{\eta}}) \exp\left[i\boldsymbol{k}^\intercal \boldsymbol{r}\right] \mathrm{d}\bar{\boldsymbol{\eta}}, \tag{2.40}$$

where $\tilde{u}$ is the expansion coefficient and $\boldsymbol{k} = -k\bar{\boldsymbol{\eta}}$ is the wave vector, propagating in the direction of $\bar{\boldsymbol{\eta}}$. The inner product and norm are set as [4, 5]

$$\langle u_1, u_2 \rangle_{\mathcal{H}} = \frac{1}{2\pi} \int_{S(0,1)} \frac{1}{w(\boldsymbol{\eta})} \tilde{u}_1(\boldsymbol{\eta})^* \tilde{u}_2(\boldsymbol{\eta}) d\boldsymbol{\eta}, \tag{2.41}$$

$$\|u\|_{\mathcal{H}} = \sqrt{\langle u, u \rangle_{\mathcal{H}}}, \tag{2.42}$$

where $(\cdot)^*$ denotes the conjugate, $\boldsymbol{\eta}$ is calculated as the sound wave propagation from chosen reference position to the sound source[2], and $w(\boldsymbol{\eta})$ is the directional weighting function, proposed in [5], based on the von Mises-Fisher distribution, defined as

$$w(\boldsymbol{x}) = \frac{1}{4\pi C(\beta)} \exp\left[\beta \boldsymbol{\eta}^\intercal \boldsymbol{x}\right], \tag{2.43}$$

$$C(\beta) = \begin{cases} 1, & \beta = 0 \\ \dfrac{e^{\beta} - e^{-\beta}}{2\beta}, & \beta \in (0, \infty) \end{cases}. \tag{2.44}$$

Here, $\beta \in [0, \infty)$ is a directional regularization parameter. The directional vector $w(\boldsymbol{x})$ uses prior knowledge of the direction of the sound sources, and makes the second term in equation (2.1) larger for waves propagating in another direction than the assumed direction of propagation for the sound wave, leaving us with a final interpolation more likely to propagate in the desired direction. Using this directional weighting, the kernel is defined as [5]

$$\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{1}{4\pi} \int_{S(0,1)} w(\boldsymbol{x}) \exp\left[-i\boldsymbol{k}^\intercal (\boldsymbol{r}_1 - \boldsymbol{r}_2)\right] d\boldsymbol{x}, \tag{2.45}$$

[2]This reference position is not necessarily easy to determine, and the choice of a reference position has certain consequences; these will be discussed later on in the report. In [4], where this kernel was first introduced, the propagating sound was assumed to be in the form of a plane wave, whereas our case will be a spherical wave.

and as shown in [5], by substituting $w(\boldsymbol{x})$ in (2.43) into (2.45), $\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2)$ can be expressed as

$$\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{1}{C(\beta)} j_0(\sqrt{\Psi}), \tag{2.46}$$

where

$$\Psi = \Big(i\beta \sin\theta_z \cos\phi - kr_x\Big)^2 + \Big(i\beta \sin\theta_z \sin\phi - kr_y\Big)^2 + \Big(i\beta \cos\theta_z - kr_z\Big)^2, \tag{2.47}$$

$j_0(\cdot)$ is the 0th-order *spherical* Bessel function of the first kind, ϕ and θ_z are respectively the azimuth and zenith angles of $\boldsymbol{\eta}$, and $[r_x, r_y, r_z]^\mathsf{T} := \boldsymbol{r}_1 - \boldsymbol{r}_2$. Now, by looking at

$$\langle \kappa(\boldsymbol{r}, \boldsymbol{r}_m), u(\boldsymbol{r}) \rangle_{\mathcal{H}} = \frac{1}{2\pi} \int_{S(0,1)} w(\boldsymbol{\eta}) \frac{1}{w(\boldsymbol{\eta})} \exp\left[i\boldsymbol{k}^\mathsf{T}\boldsymbol{r}_m\right] \hat{u}(\boldsymbol{\eta}) d\boldsymbol{\eta} \tag{2.48}$$

$$= \frac{1}{2\pi} \int_{S(0,1)} \hat{u}(\boldsymbol{\eta}) \exp\left[i\boldsymbol{k}^\mathsf{T}\boldsymbol{r}_m\right] d\boldsymbol{\eta} \tag{2.49}$$

$$= u(\boldsymbol{r}_m), \tag{2.50}$$

and letting $\boldsymbol{r}_m = \boldsymbol{r}$, we can confirm that this is a reproducing kernel of $\mathcal{H}$. As such, a larger value of β is expected to give a more accurate interpolation, granted that the assumed direction of the sound propagation used corresponds to the actual direction of the sound propagation. This kernel will be referred to as the directional Helmholtz kernel in the rest of the report. Note that by removing the directional weighting, i.e., letting $\beta = 0$ in (2.46), we get the previously obtained Helmholtz kernel.

2.2.3 The Gaussian kernel

The Gaussian kernel is defined as [3]

$$\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2) = e^{-\gamma^2 \|\boldsymbol{r}_1 - \boldsymbol{r}_2\|^2}. \tag{2.51}$$

With this, $\mathcal{H}_0$ denotes the function space

$$\mathcal{H}_0 = \left\{ f : \sum_{j=1}^{k} \alpha_j \kappa(\boldsymbol{r}_j, \boldsymbol{r}) \right\}. \tag{2.52}$$

Given this function space, an inner product and norm is defined as

$$\langle f_1, f_2 \rangle_{\mathcal{H}_0} = \sum_{i=1}^{k} \sum_{j=1}^{l} \alpha_i \alpha_j \kappa(\boldsymbol{r}_i, \boldsymbol{r}_j), \tag{2.53}$$

$$\|f\|_{\mathcal{H}_0} = \sqrt{\langle f, f \rangle_{\mathcal{H}_0}} = \sqrt{\sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_i \alpha_j \kappa(\boldsymbol{r}_i, \boldsymbol{r}_j)} = \sqrt{\boldsymbol{\alpha}^T K \boldsymbol{\alpha}}, \tag{2.54}$$

where we have $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_k]^\intercal$ and $\boldsymbol{K}_{ij} = \kappa(\boldsymbol{r}_i, \boldsymbol{r}_j)$. By now looking at the kernel expressed as a function in $\mathcal{H}_0$, as well as a general function $u(\boldsymbol{r})$, defined as

$$\kappa(\boldsymbol{r}) = \kappa(\boldsymbol{r}_m, \boldsymbol{r}), \tag{2.55}$$

$$u(\boldsymbol{r}) = \sum_{i=1}^{k} \alpha_i \kappa(\boldsymbol{r}_i, \boldsymbol{r}), \tag{2.56}$$

the inner product is found to be

$$\langle \kappa(\boldsymbol{r}), u(\boldsymbol{r}) \rangle_{\mathcal{H}_0} = \sum_{i=1}^{k} \alpha_i \kappa(\boldsymbol{r}_i, \boldsymbol{r}_m) \tag{2.57}$$

$$= u(\boldsymbol{r}_m). \tag{2.58}$$

By letting $\boldsymbol{r}_m = \boldsymbol{r}$, we can confirm that this is a reproducing kernel and ultimately leaves us with a reproducing kernel Hilbert space, $(\mathcal{H}_0, \|\cdot\|_{\mathcal{H}_0})$, with the Gaussian kernel. Note that this can be used for any dimension, unlike the two previous kernels that are only defined for the case of a three dimensional field [3].

2.3 Modelling the measurements

In order to evaluate the robustness of the interpolation with the different kernels to measurement errors, noise will be added to all of the measurements in our model. For all the sound pressures $u(\boldsymbol{r}_m)$ in the case of the simulated data, noise is added as

$$s_m = u(\boldsymbol{r}_m) + e_{p,m}, \tag{2.59}$$

where s_m denotes the recorded sound pressure, and $e_{p,m} \in \mathcal{N}(0, \sigma_p^2)$ denotes the noise in the recorded signal, both for microphone index m, and σ_p^2 is the specified

[3] The corresponding 2D kernels can be found in the appendix, section 6.1.

variance. For the microphone and sound source positions, we let

$$\hat{\boldsymbol{r}}_m = \boldsymbol{r}_m + \boldsymbol{e}_m, \tag{2.60}$$

$$\hat{\boldsymbol{r}}_s = (r_s, \theta_s + e_{s,\theta}, \phi_s + e_{s,\phi}) \tag{2.61}$$

where we let $\hat{\boldsymbol{r}}_m$ and $\boldsymbol{r}_m$ denote the estimated and real position of microphone m, respectively, $\hat{\boldsymbol{r}}_s$ and $\boldsymbol{r}_s$ denote the estimated and real position of the sound source, respectively, $\boldsymbol{e}_m \in \mathcal{N}^3(\mathbf{0}, \sigma_m^2 I)$ denote the error in microphone position estimate, $e_{s,\theta}, e_{s,\phi} \in \mathcal{N}(0, \sigma_s^2)$ denote the error in the measured θ and ϕ, respectively, where σ_s^2, σ_m^2 is the specified variances. All of these errors are assumed to be independent of the signal or measurements.

2.4 Performance measurements - SNR, SINAD

This section breaks down the different performance evaluation measurements, namely the signal-to-noise ratio (SNR) and the signal-to-noise and distortion ratio (SINAD). As both of these measures have several different definitions, we will define both of them in this section. As the similarity in their names imply, they're more or less the same thing. They both measure, in our case, one sound pressure measurement to another sound pressure measurement. When evaluating the recorded signal, we define the SNR as

$$\text{SNR} = \frac{|u(\boldsymbol{r}_m)|^2}{|s_m - u(\boldsymbol{r}_m)|^2} = \frac{|u(\boldsymbol{r}_m)|^2}{|u(\boldsymbol{r}_m) + e_{p,m} - u(\boldsymbol{r}_m)|^2} = \frac{|u(\boldsymbol{r}_m)|^2}{|e_{p,m}|^2}. \tag{2.62}$$

The SNR evaluates how much the noise alters the recorded sound pressure in relation to the real sound pressure. Further, the expected value of the SNR is then

$$\begin{aligned} \mathbb{E}(\text{SNR}) &= \mathbb{E}\left(\frac{|u(\boldsymbol{r}_m)|^2}{|e_{p,m}|^2}\right) \\ &= \frac{\mathbb{E}\left(|u(\boldsymbol{r}_m)|^2\right)}{\mathbb{E}\left(|e_{p,m}|^2\right)}. \end{aligned} \tag{2.63}$$

Since $e_{p,m} \in \mathbb{N}(0, \sigma^2)$, $|e_{p,m}|$ follows a half-normal distribution [29], which with

the definition of the variance yields

$$\begin{aligned}\mathbb{V}\left(|e_{p,m}|\right) &= \mathbb{E}\left(|e_{p,m}|^2\right) - \mathbb{E}\left(|e_{p,m}|\right)^2 \implies \\ \mathbb{E}\left(|e_{p,m}|^2\right) &= \mathbb{V}\left(|e_{p,m}|\right) + \mathbb{E}\left(|e_{p,m}|\right)^2 \\ &= \sigma^2(1 - \frac{2}{\pi}) + \left(\sigma\sqrt{\frac{2}{\pi}}\right)^2 \\ &= \sigma^2. \end{aligned} \tag{2.64}$$

Inserting (2.64) into (2.63), and then solving for σ^2 yields a final expression as

$$\sigma^2 = \frac{\mathbb{E}\left(|u(\boldsymbol{r}_m)|^2\right)}{\mathbb{E}\left(\text{SNR}\right)}. \tag{2.65}$$

This will later on be used to simulate an error that on average gives us a specified SNR[4]. To evaluate the interpolation, the SINAD is introduced, defined as

$$SINAD(\boldsymbol{r}) = \frac{|u(\boldsymbol{r})|^2}{|\hat{u}(\boldsymbol{r}) - u(\boldsymbol{r})|^2}. \tag{2.66}$$

The SINAD evaluates how much the noise in combination with the distortion created by the interpolation method affected the interpolated sound pressure in relation to the real sound pressure, hence the name. As we do not have access to the real signal $u(\boldsymbol{r})$ for the real data, neither the SNR nor the SINAD can be calculated. However, an approximation of the SINAD measurement is introduced by replacing the signal $u(\boldsymbol{r}_m)$ with s_m for the real data. For this measurement, as long as the distortion caused by the interpolation is small enough, the approximation is valid. Lastly, for the in-ear sound pressure measurements, the evaluation will be done in the time domain and is then defined as

$$SINAD(\boldsymbol{r}) = \frac{\sum_{t=0}^{N} |u^t(\boldsymbol{r}, t)|^2}{\sum_{t=0}^{N} |u^t(\boldsymbol{r}, t) - \hat{u}^t(\boldsymbol{r}, t)|^2}, \tag{2.67}$$

where $u^t(\boldsymbol{r}, t)$ denotes the sound pressure at time t at location $\boldsymbol{r}$, i.e., formulated in the time domain rather than the frequency domain. Again, the real signal u^t will be replaced with the in-ear sound pressure measurements for the real data.

[4]Note that this can be done both for real and complex-valued numbers; in this project, it was done only with real numbers.

2.5 Using real data

In order to use real data, microphones will of course record the sound pressure in the time domain rather than the frequency domain, giving us $u^t(\boldsymbol{r}, t)$. In order to perform the proposed interpolation methods, one needs to account for this in some way. In this work, we will only work with single frequency sinusoids, as those were the only signals used for the real measurements. Thus, the recordings will be transformed to the frequency domain through a Fourier transformation, and only the largest found frequency will be considered for the calculations. This could easily be expanded to work for all frequencies in the signal, but as we know the signal used only contains a single frequency, the rest will be considered noise.

2.6 In-ear sound pressure interpolation

While the previous parts of the report consisted of free-field sound interpolation, i.e., an acoustic environment where the sound waves propagate freely, we here apply this method to find an estimate of the in-ear sound pressure with the presence of a human body in the field. Notable efforts have gone into modelling how the presence of the listener affects the sound field. As the physical properties of the listeners head significantly impact the sound pressure inside of the ear canal, a generalized HRTF cannot be used if one wants an accurate representation of the sound pressure. While there are several datasets of HRTFs available [30–32], these are often cumbersome to use, and requires notable efforts to measure. As a result, most datasets are limited to HRTFs for a sound source at a given distance, measured at incremental steps in both azimuth and elevation. In order to take the presence of the head and torso into account for the sound field interpolation, a pre-calibrated HRTF is often used. Considering a sound source at $\bar{\boldsymbol{r}}_h = (r_h, \theta_h, \phi_h)$, where r_h denotes the radial distance, θ_h the elevation, and ϕ_h the azimuth angle to the center of the listener's head, the HRTFs from the position of the center of the listener's head in a free field, to the listener's left and right ears, $H_{l,0}$ and $H_{r,0}$, as defined in [31], may be expressed as

$$H_{i,0}(r_h, \theta_h, \phi_h, \omega) = \frac{H_i(r_h, \theta_h, \phi_h, \omega)}{H_0(r_h, \theta_h, \phi_h, \omega)}, \tag{2.68}$$

with $i \in \{l, r\}$, where u_l and u_r denote the sound pressure in the left and right ear, H_l and H_r the transfer function from the sound source to u_l and u_r, u_0 the sound pressure in the position of the center of the head if the head was absent, H_0 the transfer function from the sound source to u_0, and $H_{l,0}$ and $H_{r,0}$ the HRTF from u_0 to u_l and u_r. It should be stressed that $\bar{\boldsymbol{r}}_h$ depends on the orientation of the head; in our case, the left and right ears are placed on the positive and negative $y-$axis,

respectively, with the head center being placed at the origin. The HRTF can then be expressed in the time domain, as

$$h_{i,0}(r_h, \theta_h, \phi_h, t) = \mathcal{F}^{-1} H_{i,0}, \tag{2.69}$$

with $i \in \{l, r\}$, where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform, and $h_{l,0}$ and $h_{r,0}$ the head related impulse response (HRIR) from the center of the listener's head to the left and right ear, as was also done in [31]. This allows the sound pressure in each ear to be expressed as

$$u_i^t(r_h, \theta_h, \phi_h, t) = h_{i,0}(r_h, \theta_h, \phi_h, t) * u_0^t(r_h, \theta_h, \phi_h, t), \tag{2.70}$$

with $i \in \{l, r\}$, where u_l^t and u_r^t corresponds to the sound pressure in the left and right ear in the time domain, respectively, u_0^t corresponds to the sound pressure in the position of the center of the head in the time domain with the head absent, and $*$ denotes the convolution operator.

In order to find an estimate for the in-ear sound pressure, one needs to find an estimate of u_0^t. By setting up microphones in the acoustic environment, we here interpolate the sound pressure to the center of the head using the recorded signals of the microphones. This is done by first taking the Fourier transform of the signals in every microphone, which we then use to find an estimate of the sound pressure, in the frequency domain, at the position of the center of the head with the KRR method from the previous parts of the report, yielding $\hat{u}_0(r_h, \theta_h, \phi_h, \omega)$.

The proposed method for the in-ear sound pressure interpolation is then to use the estimated sound pressure using the KRR interpolation method (see also [3]), expressed in the time domain, $\hat{u}_0^t(r_h, \theta_h, \phi_h, t)$, to find the sound pressure estimate inside of the ears, as (for $i \in \{l, r\}$)

$$\hat{u}_i^t(r_h, \theta_h, \phi_h, t) = h_i(r_h, \theta_h, \phi_h, t) * \hat{u}_0^t(r_h, \theta_h, \phi_h, t). \tag{2.71}$$

2.6.1 Different HRTF datasets

The HRTF dataset used in this project is from [31]. This dataset consists of HRTF measurements defined as the transfer function for the sound pressure in the center of the head, with the person absent, in a free field, to the sound pressure inside of both of the ear canals in the same field but with the person present. As the dataset from [31] also introduced a group delay of 100 samples, further changed by windowing and filtering, which has to be accounted for, the estimated delay for the HRIRs was approximately 108 samples.

An example plot of the HRIR from this dataset can be seen in Figure 2.1. The shape of the different HRIR measurements differ a lot depending on the angles to the sound

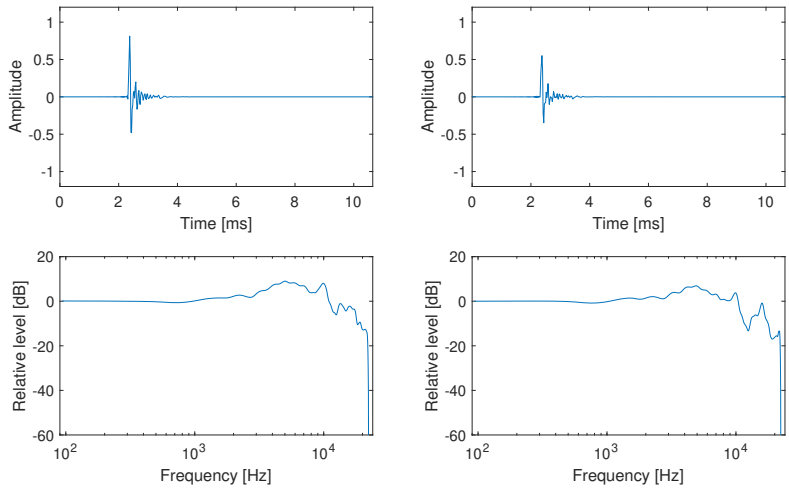

Figure 2.1: Example plot of an HRIR from [31], using the KEMAR manikin for the measurements, for the azimuth of $180°$and the elevation of $60°$. In the upper plots are the HRIRs for the left and right ear, in the left and right plot, respectively. The relative level of the respective sound pressure is shown below the HRIRs. The relative level of the respective sound pressure is shown below the HRIRs, comparing the in-ear sound pressure to the reference sound pressure, in this case the sound pressure in the center of the head, with the absence of the KEMAR manikin.

source; however, worth noting is that the relative level of the sound pressure does not significantly change for lower frequencies, and this behaviour is similar for all of the HRIRs.

However, HRTF datasets are defined in different ways, and another example of this is the HRTF dataset from [30]. This is instead defined as the transfer function from the speaker to the sound pressure in both ear canals, and consequently, the relative sound pressure levels are inversely proportional to the distance to the sound source, given the spherical form of the wave. An example plot of the HRIR from this dataset can be seen in Figure 2.2. Note that the HRIRs are very different from the ones in Figure 2.1, as is expected since the HRIRs are defined through completely different transfer functions.

For the former figure, the HRIR needs to be compensated by the approximate 108 samples of introduced extra delay, while the latter figure does not need this. However, the relative change in sound pressure, by both amplitude and phase shift, is still intact for both datasets. A common approximation for HRIRs defined for distances larger than 1 m is as a far-field approximation, where the distance no longer is relevant for the relative changes in phase shift and amplitude for the HRIRs [33]. In the case with the dataset used in this project, as long as the mentioned delay is adjusted for, the HRIRs can be used as they are defined for any distance greater than 1 m. For the

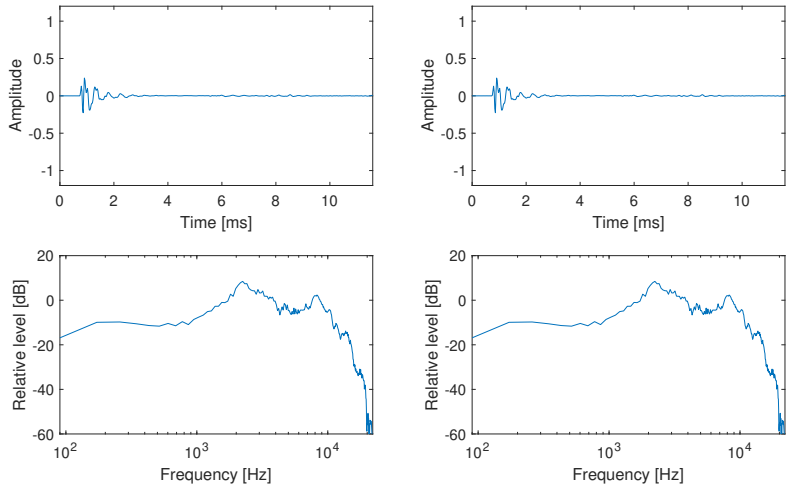

Figure 2.2: Example plot of an HRIR from [30], using the KEMAR manikin for the measurements, for the azimuth of $180°$and the elevation of $60°$. In the upper plots are the HRIRs for the left and right ear, in the left and right plot, respectively. The relative level of the respective sound pressure is shown below the HRIRs, comparing the in-ear sound pressure to the reference sound pressure, in this case the signal omitted from the speaker.

HRIRs from [30], the distance would still be needed for the amplitude scaling as well as the delay. In conclusion, HRTFs defined as in [31] will be easier to use for this project, which motivates our choice of the HRTF dataset actually used.

Chapter 3

Experimental evaluation

Before we dive into the real experiments, we look back at the 2D example from the introduction. What was used in the example was five microphones placed along a line, evenly spread out. What could be seen was that the Gaussian kernel did not at all capture the periodicity of the sound wave. Now, we've introduced the new kernels[1], and the same experiment but with the directional Helmholtz kernel as well can be seen in Figure 3.1, where the directional parameter β was set to a value of 10. It is evident from the plot that the directional Helmholtz kernel introduces a significant improvement to the interpolation problem, at least in the two-dimensional case for this setup.

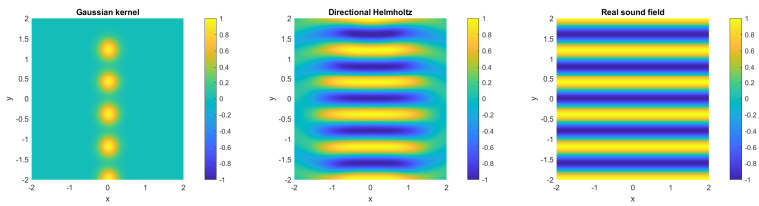

Figure 3.1: Simulation using regularization parameter $\lambda = 10^{-2}$, a frequency of 426, using 5 microphones, and is done in a 2D space with the microphones being spread out evenly along the line $x = 0.0134$. The wave used is a plane wave propagating in the y-direction. The grid is created with incremental changes in x and y of 0.05 m. In the leftmost and middle plot, the interpolated sound field using the Gaussian kernel and directional Helmholtz kernel ($\beta = 10$), respectively, is shown, and in the rightmost plot the real sound field is shown.

Now we move on to the 3D case, which will be the case for the rest of the report. To assess the effectiveness of the interpolation method, we first examine the simulated

[1]The kernels introduced so far have been the 3D kernels, but the corresponding 2D kernels can be found in the appendix, section 6.1.

audio environment. A point sound source was placed uniformly at random on a distance of a specified interval away from the center of Ω, outside of the region, where the audio signal was formed as a single tone with a random phase shift, propagating as a monochromatic spherical wave in a free field to all microphones. The signal omitted is given by

$$v_{wave}(\boldsymbol{r}, t) = \frac{A}{|\boldsymbol{r}|} \exp(i(\omega t - k|\boldsymbol{r}|)), \tag{3.1}$$

where A is a complex-valued constant set to $A = 1$ for all simulations. Then, 12 microphones were placed in Ω. The regularization parameter λ was set to 10^{-2} for all of the simulations [2]. The interpolation was done with all three, previously defined, kernels, with $\gamma = 0.5k$ for the Gaussian kernel, as it was shown to perform the best for this value in [2]. For all the plots, we let "no interpolation" denote an interpolation where all interpolated values are set to 0.

3.1 Grid evaluation

For this section, all the measurements were evaluated in all of Ω, by setting up a grid with incremental changes in the x, y, and z direction of 0.05 m. In this section, we defined Ω to be a sphere of radius 1 m. There were 64 simulations for each frequency, all of which had different positions of the microphones as well as the point sources. The microphone positions were based on the grid coordinates, but as noise was added to the measurements, their real positions was off the grid; this does not change anything for the computations, only the recorded sound pressure values. For these grid measurements, the reference position chosen for the directional Helmholtz kernel was the center of Ω, meaning that the directional weighting was based on the direction from the center of the sphere to the point sound source. Unless otherwise specified, the point sound source was placed uniformly at random on a distance of 10 to 12 meters from the center of Ω. The directional weighting parameter β is specified in the plots.

[2]This could be examined further; however, this parameter is added to model the noise, and as such the optimal parameter value would depend on the added noise. Thus, in the case of no noise in any measurement, the optimal value would be 0, but this is not realistic for real measurements. The value 10^{-2} was used in [2], which is why it was used here as well.

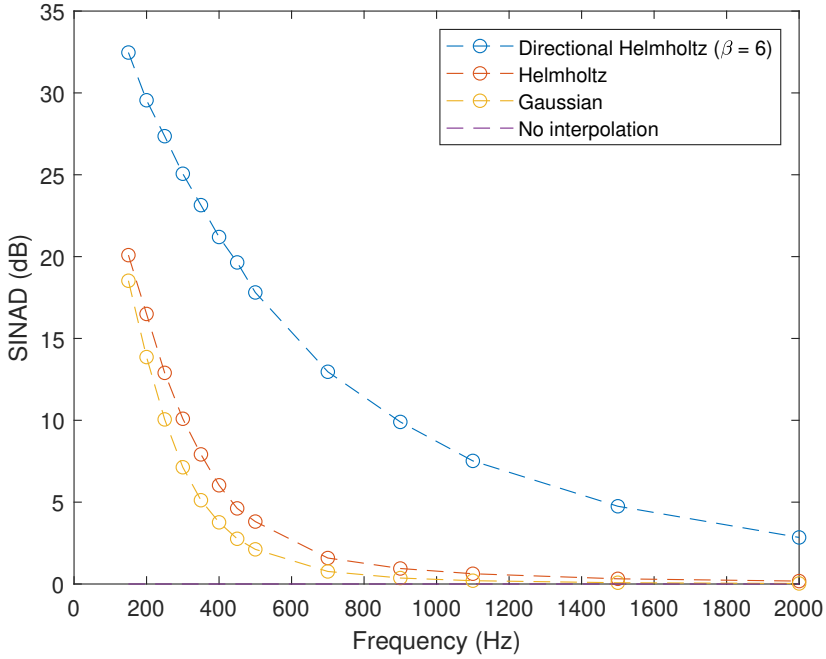

Figure 3.2: No noise in any measurements. Frequencies used are frequencies between 150 and 500 with incremental steps of 50, as well as 700, 900, 1100, 1500, and 2000.

In Figure 3.2, we see the resulting average SINAD for all points in Ω, given the interpolation by the specified kernel, for all measured frequencies[3]. As expected, the directional Helmholtz kernel outperforms both other kernels, while the Helmholtz kernel slightly outperforms the Gaussian kernel.

In order to evaluate the robustness to noise in the sound pressure measurements, the average SINAD in dB scale for the frequencies between 150 and 500 with incremental steps of 50, when noise is added by the specified SNR, was calculated. These were chosen as the interpolation was fairly good for all kernels for these frequencies.

[3]Note that these are the frequencies measured in the real experiment. This is why we use a combination of these frequencies for the plots in this section.

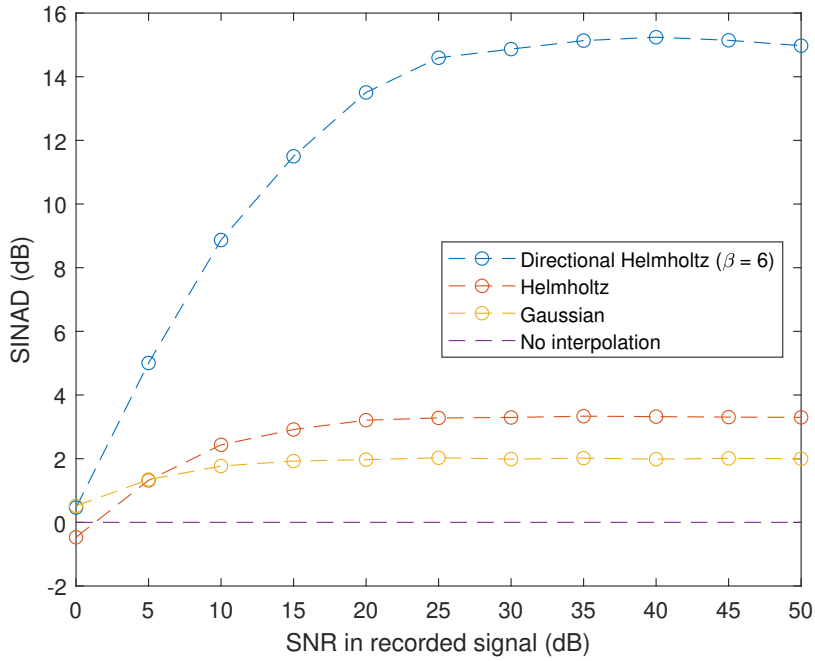

Figure 3.3: Noise is added to the sound pressure measurements by the microphones, by the level of the specified SNR. Frequencies used are frequencies between 150 and 500 with incremental steps of 50, where the SINAD is calculated as the average SINAD for all frequencies.

As can be seen in Figure 3.3, once again, the directional Helmholtz kernel outperforms both other kernels for all SNR values above 0, while the Helmholtz kernel slightly outperforms the Gaussian kernel for SNR values above 5. A note on this is that an SNR of 0, in dB scale, corresponds to having the same amount of power in the recorded sound pressure coming from the noise as is coming from the real signal, or with other words - the recording itself is extremely noisy.

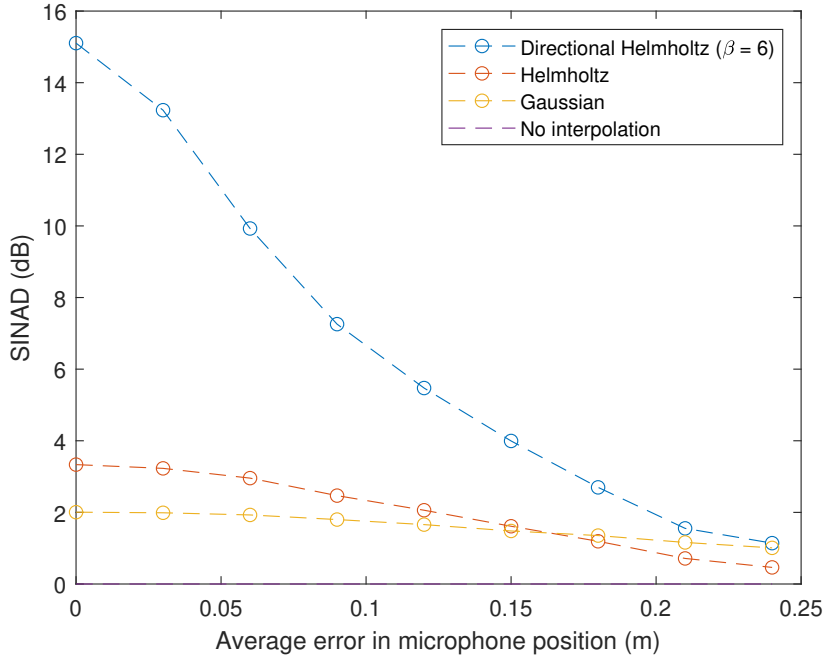

Figure 3.4: Noise is added to the measured position of the microphones. Frequencies used are frequencies between 150 and 500 with incremental steps of 50, where the SINAD is calculated as the average SINAD for the frequencies from 150 to 500, with incremental steps of 50.

In Figure 3.4, the average SINAD for the measured frequencies can be seen as a function of the added expected error in the microphone position measurements. As shown, the SINAD decreases most significantly for the directional Helmholtz kernel. Since the directional Helmholtz kernel not only gets an error in the measured sound pressure due to the positional change, it also tries to approximate the measured sound pressure, with the error, as a wave propagating in a specific direction, the relative change in the SINAD is expected to be larger for this kernel as the error grows larger. Conversely, the non-directional kernels put no emphasis on the direction of which the sound wave was propagating, and thus should suffer smaller relative changes to the SINAD as the measured positions have an increasing error.

The relative change for the SINAD is larger for the directional Helmholtz kernel, compared to the two other kernels. The small relative change for the SINAD with the Gaussian kernel could be explained by the fact that the kernel puts no emphasis on the physical properties of a wave, while the others do, which in turn could lead to the addition of errors in the measurements not having a too big of an impact on the interpolation, as the interpolation is already relatively bad. Conversely, the impact of the other two kernels is expected to be larger, as they rely more heavily on the

physical properties of the actual sound wave; the directional Helmholtz kernel for both the wave itself and its direction of propagation.

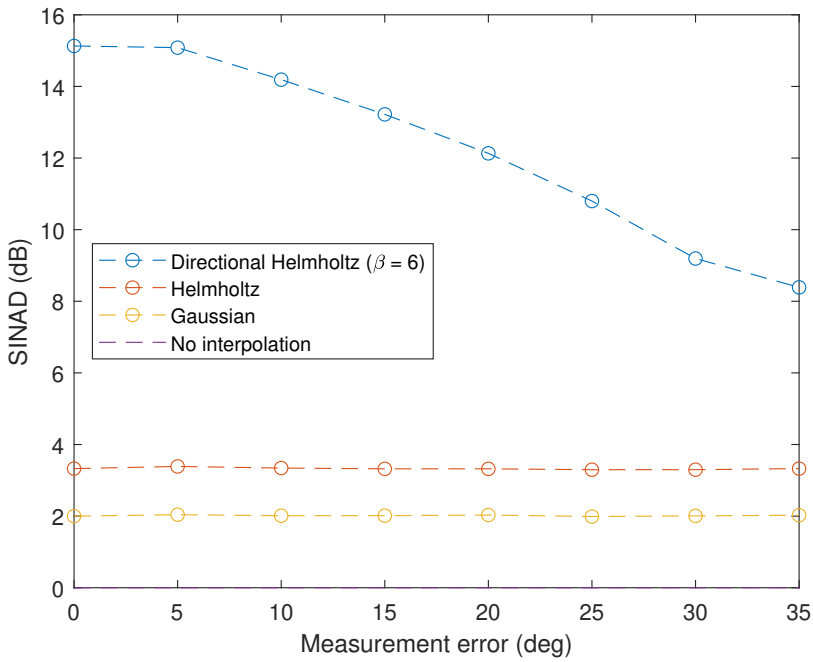

Figure 3.5: Noise is added to the measured position of the point sound source position. Frequencies used are frequencies between 150 and 500 with incremental steps of 50. The measurement error is added such that the elevation and azimuth angles for the sound source position are both off by the specified measurement error, in degrees, on average.

Figure 3.5 shows the SINAD plotted against the added average measurement error for the sound source position. Here, the measurement error was added such that the elevation and azimuth angles were both off by the specified measurement error, in degrees, on average. For the Helmholtz and Gaussian kernel, the point sound source position is irrelevant for the interpolation algorithm, while the directional weighting for the directional Helmholtz kernel depends on the direction of the propagating sound for its weighting function. As shown in the plot, while the performance does decrease for the directional Helmholtz kernel, it is still outperforming the other two kernels for very large changes in the measured position of the point source in relation to the real position.

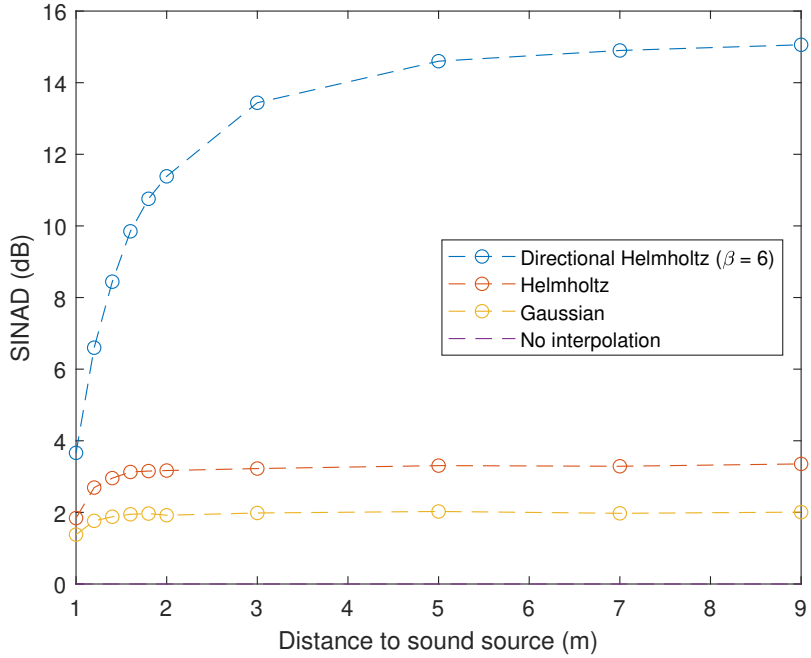

Figure 3.6: The plot shows the average SINAD as a function of the distance to the sound source from the center of Ω. Frequencies used are frequencies between 150 and 500 with incremental steps of 50.

As stated previously in the report, for both variations of the Helmholtz kernel, the region Ω must be source free in order for the solution to be valid. Further, while the sound source is located close to Ω, the amplitude of the wave changes rapidly as it is inversely proportional to the distance to the sound source. Far away, the wave can be approximated as a plane wave; this is not the case when the sound source is close. Thus, it seems likely that we would need more measurement points in Ω to more accurately model the behaviour of the wave when the sound source is nearby. With this in mind, we want to investigate the dependence of the distance to the sound source. Further, the directional Helmholtz kernel has a reference position from where it calculates the direction of which the sound wave propagates; this angle is used for all positions in the sound field, which naturally means that a point sound source very far away from Ω will result in fairly similar direction of propagation for all points in Ω, while a point sound source being close will not.

Note that Ω is a sphere of radius 1 m, which is why we place the sound source no closer than 1 m to the center of the sphere. In Figure 3.6, we plot the average SINAD for the specified frequencies against the distance to the point sound source. While the resulting SINAD for all kernels improves by increasing the distance, the most signif-

icant improvement can be seen for the directional Helmholtz kernel, as is expected given the reasoning above.

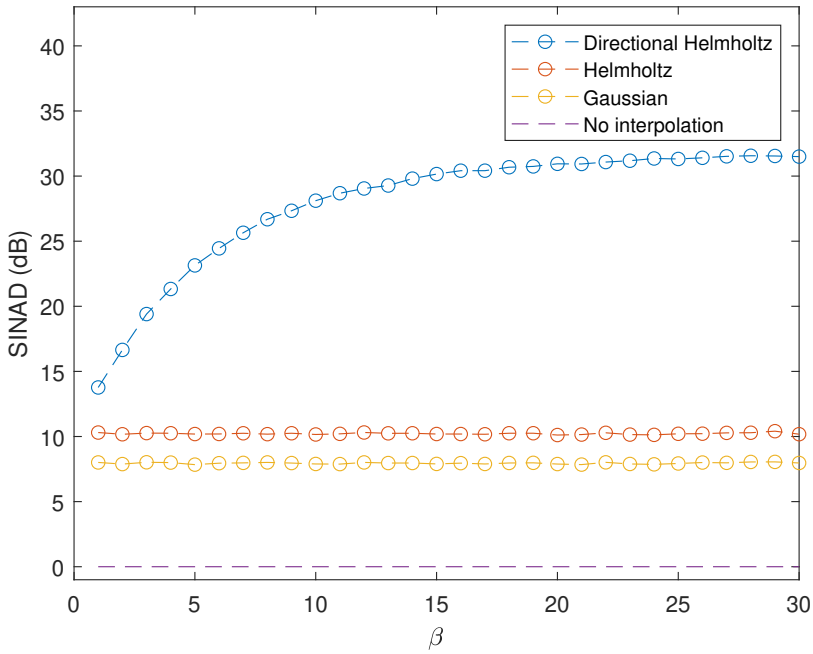

Figure 3.7: The plot shows the average SINAD as a function of the directional regularization parameter β. Frequencies used are frequencies between 150 and 500 with incremental steps of 50. The SINAD of the Helmholtz kernel and the Gaussian kernel will be constant on average, as they do not depend on β.

Lastly, an evaluation of the performance as a function of the directional regularization parameter β can be seen in Figure 3.7. As the sound source was located rather far away from Ω, the direction of sound wave propagation in relation to each of the microphones should be fairly similar, and as there was no added noise or measurement error, an increasing β value should result in a better performance, as was also the case for the simulation.

3.2 Spatial point evaluation

As opposed to the previous section, this section will not be using a grid to measure the interpolation performance. Instead, the microphones will be placed in Ω based on real-life measurements, taken from the anechoic chamber at the Dept. of Construc-

tion Sciences at Lund University, and the evaluation will be done by iterating through every microphone as the validation microphone, with the remaining 11 microphones used to form the interpolation to the validation microphone. The setup can be seen in Figure 3.8, including the in-ear measurement positions, which were excluded for this part of the project. As the purpose of this section is to evaluate the real-life measured data, the simulated data is used to mimic the real-life experiment as a comparison; as such, the simulated data will be evaluated in the same way, despite being able to evaluate the interpolated data for the entire field, which we would be able to do since we have access to the simulated sound pressure in the entire field.

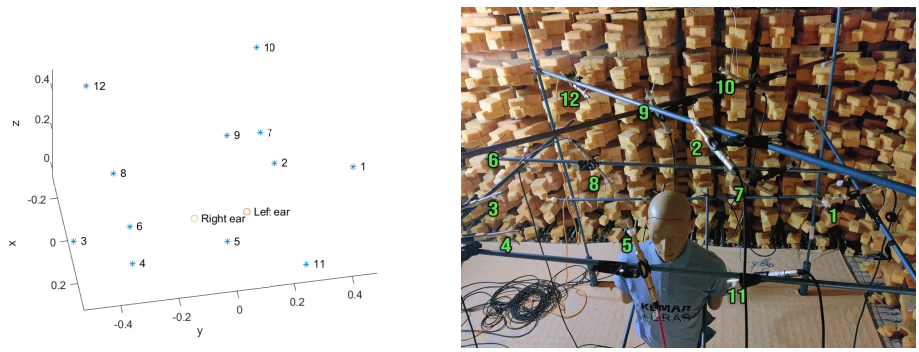

Figure 3.8: On the left is a plot of the resulting microphone positions, including the in-ear microphones, and on the right is a picture of the real setup of the microphones, taken from the anechoic chamber at the Dept. of Construction Sciences at Lund University. Exact positions can be found in the appendix.

The distances between the microphones were measured and used with classical multidimensional scaling [34] in order to find the relative positions. The speaker position was found by trilateration, using an eigenvalue solver as described in [35]. The root-mean-square error (RMSE) of the microphone positions was measured to be ± 0.02 m, by comparing the estimates given by the multidimensional scaling with that of the distance measurements. As certain distances were hard to measure due to the presence of the KEMAR manikin, two microphone positions were also found using the trilateration method from [35].

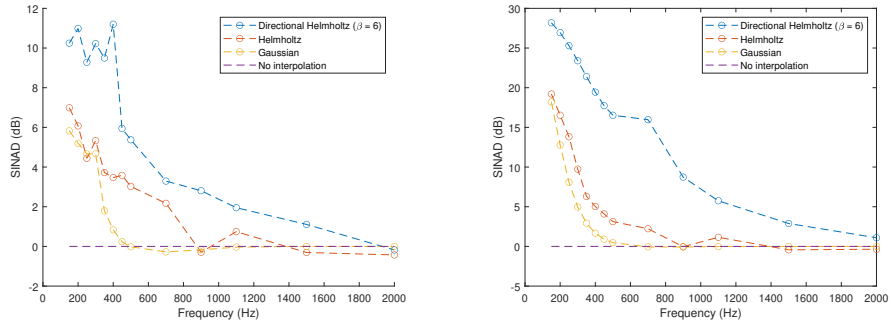

Figure 3.9: On the left is the resulting SINAD for the real data. On the right is the plot for the simulated data, with no measurement error or noise.

The resulting average SINAD for the real data as well as the simulated data using the same microphone and sound source position can be seen in Figure 3.9. While there was no noise in the simulated data, there is still a fairly large resemblance between the two plots, most notably for the Helmholtz kernel for the frequency of 900. For this frequency, the SINAD drops below 0 for both the real data and the simulated data, which is likely due to the placement of the microphones, which also implies that the model is likely fairly accurate.

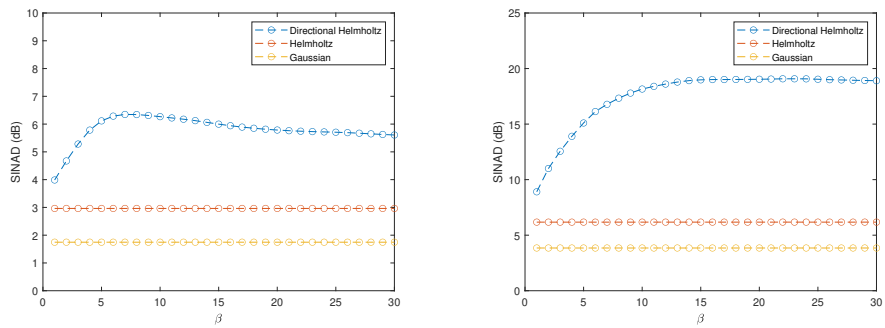

Figure 3.10: On the left is the resulting average SINAD as a function of β for the real data. On the right is the corresponding plot for the simulated data, with no measurement error or noise. Frequencies used are frequencies between 150 and 500 with incremental steps of 50, as well as 700, 900, 1100, 1500 and 2000.

Finding an optimal β value depends on both microphone positions as well as the measurement errors and noise. In Figure 3.10, the average SINAD for all measured frequencies is shown as a function of β. For the real data, the best SINAD for the specified frequencies was found for $\beta \approx 7$, while for the simulated data it was found for $\beta \approx 17$. A similar experiment was done for the frequencies between 150 and 500 with incremental steps of 50, where the best SINAD for the specified frequencies was

found for $\beta \approx 6$ for the real data, which motivates the use of $\beta = 6$ for the previous section.

3.3 Spatial point evaluation, in-ear sound pressure

In order to get a more accurate interpolation of the in-ear sound pressure, an HRTF was used, which essentially describes the spatial filtering effect resulting from the presence of the head and torso. The way the HRTF used in this report [31] is defined is as a function of $\bar{\boldsymbol{r}}_h = (r_h, \theta_h, \phi_h)$, where r_h denotes the radial distance, θ_h the elevation, and ϕ_h the azimuth angle from the center of the listener's head to the sound source. As such, the orientation of the listeners head in the space is essential to determine the angles. Thus, the position of the ears were used to rotate all measured positions such that the position of the left and right ears were on the positive and negative y-axis, respectively, as was done for the HRTF measurements in [31]. Once the rotation was complete, the left and right ears were placed at $[0 \ 0.09 \ 0]^\intercal$ and $[0 \ -0.09 \ 0]^\intercal$, respectively, being the ear locations of the KEMAR manikin. Worth noting here is that the RMSE of the microphone positions was roughly 0.02 m, which for the initial interpolation is not necessarily a lot - however, for the purpose of rotating all the measured positions based on the position of the left and right ear, granted that the distance between the ears is only 18 cm, this error can lead to a fairly large error in the rotation. This will only affect the HRTF used, but as HRTFs are known to be very sensitive to the $\bar{\boldsymbol{r}}_h$, the impact could be quite large [4]. The HRTF dataset used had measurements for 835 different positions for the sound source, all of which had the sound source being located 1.5 m away from the position of the center of the head.

The setup for the microphones was still as shown in Figure 3.8. The only addition from before was that we here evaluated the performance of the final interpolation based on the average SINAD using the recorded sound pressure in the left and right ear, forming the interpolation based on all recordings from the 12 regular microphones as test data. In addition to the propagating sound wave being modelled as a spherical wave in a free field, to mimic the presence of the head, the signal was further filtered using the HRTFs for the corresponding direction in order to form the sound pressure inside each of the ears, thus using the exact same HRTF for the simulated signal as for the interpolated signal for the simulated data. The simulations were done for each of the 835 measured point source locations for the HRTF data, while the real data only consisted of data from one sound source position. The HRTF used for both the real and simulated data was measured with a KEMAR manikin, which is the same type of manikin used in our experiment.

[4] Due to this, the sound source position was manually slightly adjusted, read more about this in the appendix.

For the measured data, only one sound source location was used, located approximately 2.8 m away from the center position of the head. As the elevation and azimuth angles from the sound source to the center of the head in our measurement were not part of the HRTF dataset detailed in [31], the used HRTF was interpolated according to the method described in [16]. As the distances to the sound source both in the HRTF measurements and in our experiment were larger than 1.0 m, the measured HRTF was used as a far field approximation of the HRTF [33].

In order to make an easier distinction between the methods, we will hereafter denote the interpolation method from the previous sections as kernel ridge regression (KRR), while the same interpolation method but through the filtering of the HRTF will be denoted as the proposed method.

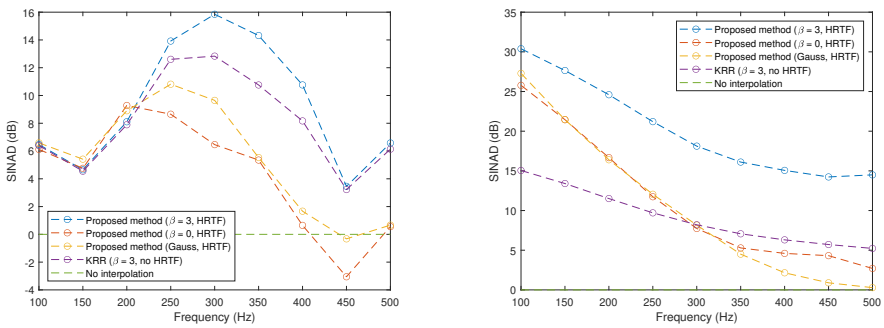

Figure 3.11: The figure show the estimated average SINAD for the sound pressure in both ears for different frequencies. The plot on the left is obtained using the real data, while the plot on the right is obtained using simulated data. The purple line corresponds to the interpolation method without the use of an HRTF, while the proposed method is using the HRTF filtering of the interpolated signal from the center position of the head.

In Figure 3.11, the resulting average SINAD for the real data on the left, as well as the simulated data on the right, is shown. The purple line is using the interpolation method of the previous sections, i.e., KRR without the use of an HRTF. From this, there are two primary comparisons to be made. First, the resulting average SINAD for the three plots labeled as the proposed method. It is clear that the directional Helmholtz kernel outperforms the other two kernels. The other comparison is between the proposed method with the directional Helmholtz kernel and the use of the same kernel but without the HRTF filtering; i.e., the blue and purple lines. Again, the proposed method with the directional Helmholtz kernel improves the average SINAD for both the real and simulated data. While the behaviour of the measured data was quite different from the simulated data, the same conclusion still holds; the proposed method, with the directional Helmholtz kernel, performed better than the previously used interpolation method. The difference between the behaviour of the two plots for the lower frequencies is likely due to measurement errors.

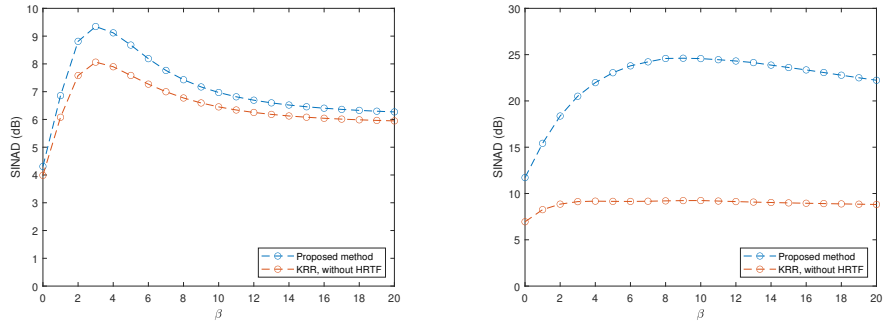

Figure 3.12: The figures show the estimated average SINAD for the sound pressure in both ears for different values of the regularization parameter β. The plot on the left is obtained using the real data, while the plot on the right is obtained using simulated data.

Lastly, the effect of the directional regularization parameter β on the resulting average SINAD for the measured frequencies was investigated, and is shown in Figure 3.12. For the simulated data, the resulting SINAD was improved by increasing the β value until $\beta \approx 9$ for the proposed method, while with the KRR method, an increased β value essentially made no difference for β values larger than 3. Since the KRR method does not take the presence of the head and torso into account, this is to be expected, as a more accurate free field estimate does not imply a more accurate estimate while the head and torso is present. What is evident from this figure is also that the proposed method, for the measured frequencies, on average, outperforms the KRR method for all values of β at least up to 20, for both simulated and real data.

Chapter 4

Discussion

4.1 Robustness of the HRTF implementation

What would seem like a natural addition to the in-ear sound pressure simulations is the addition of either noise in the HRTF measurements, or measurement error in the measurement of the sound source position, as was done for the KRR method. Ideally, this should also be done, but in order to get a meaningful result out of this, there are some issues. First, in the case of noise in the HRTF measurements, we would either need to set up a new experiment with a very precise change in position of the sound source, which is far beyond the scope of this project, or get access to such data from somewhere else. As the HRTF datasets are usually presented in terms of the finished HRTF, rather than the measured signals, this type of data is not easily accessible.

In the case of the robustness in relation to the sound source position estimate, this could be done for the simulated data fairly easily. However, as mentioned previosuly, the HRTF data available is measured with incremental steps in elevation and azimuth angles, and to get the HRTF for the intermediate values, an interpolation of the HRTF needs to be done, as was the case in this project, using the HRTF interpolation method described in [16].

For the simulations for the in-ear sound pressure in this report, the real signal was simulated as a spherical wave. In order to get the real in-ear sound pressure, the sound pressure in the position of the center of the head was filtered through

$$H_{i,0}(r_h, \theta_h, \phi_h, \omega).$$

Then, the recorded sound pressure from the microphones in the simulation was used

to form the interpolated value in the center of the head, once again filtered through

$$H_{i,0}(r_h, \theta_h, \phi_h, \omega)$$

to get the estimated in-ear sound pressure. By adding a noise to the position of the sound source, the latter HRTF used would have to be changed into

$$H_{i,0}(\hat{r}_h, \hat{\theta}_h, \hat{\phi}_h),$$

where $\hat{\hat{\boldsymbol{r}}}_h = (\hat{r}_h, \hat{\theta}_h, \hat{\phi}_h)$ denotes the position of the sound source with the added measurement error. As the HRTF was only measured for incremental steps in the elevation and azimuth angles, the added error either has to be exactly the same as the incremental steps for the measured data, or the HRTF has to be interpolated. However, by interpolating the HRTF, the resulting robustness to evaluate would be a mixture of several things at once, including the HRTF interpolation method, the KRR interpolation as well as the real HRTFs robustness to measurement error. Therefore, neither the robustness to noise in the HRTF measurement nor the addition of measurement error in the sound source position was evaluated for in-ear sound pressure estimates.

For the real data, we already use an interpolation method for the HRTF itself; this could very well be shifted by some degrees to evaluate the robustness, but once again, the same problem occurs.

A final note on this topic is that the effectiveness as well as the robustness of the proposed method, i.e., the in-ear sound pressure interpolation, depends on several things, including the interpolation method used, the measurement errors, the robustness of the real HRTF, as well as the robustness of the HRTF interpolation method. As all of these things are interdependent, they would have to be evaluated all together, meaning that if for instance an optimal β value for the directional Helmholtz kernel was found for a specific HRTF interpolation method, this might not be the ideal β value for another HRTF interpolation method.

4.2 Measurement errors for the real data

The measured data from the anechoic chamber at the Dept. of Construction Sciences at Lund University had several issues. While the gain for the regular microphones were set to be the same for all microphones, this gain differed from the KEMAR manikin measurements. On top of this, the KEMAR manikin recordings were not properly synced with the microphone recordings, and had to be fixed manually. This was done by the use of a clapperboard and estimated positions of the microphones. Thus, the

recorded sound pressure had to be re-scaled and synced manually. This could explain the behaviour for the lower frequencies in Figure 3.11 for the real data. Another point is the rotation of all coordinates; the HRTF used could be off by quite a large margin due to the measurement error for the ears, which were used for the rotation. This could be fixed by measuring the positions of something along the same axis as the ears but with a larger distance between the points, as the relative changes in the angles between the points would be significantly lower the farther away the measured points were, assuming that the measurement error remains the same.

Further, the assumption used for the interpolation is that the microphones are omnidirectional, which was not the case for the microphones used. This could impact the sound pressure recorded by each microphone, essentially adding distortion to the recorded sound pressure. There are likely ways of dealing with this, but would require measurements of the direction of the microphones, which we never measured.

Lastly, the presence of the KEMAR manikin as well as everything else used to set up the microphones in the sound field will likely affect the sound field, as it introduces absorption, diffraction and reflections. In the simulations, this was completely ignored, while this introduces distortion to the recorded, real, data. There are several ways of dealing with this, including the finite element method and the boundary element method [36], however, this was deemed to be beyond the scope of this project. Another point on the same topic is that the anechoic chamber at Lund University is not perfectly anechoic, resulting in minor reflections of the sound waves from the walls, in which case a more accurate representation of the sound wave used in the measured data could be simulated using the mirror image method, as originally described in [37]. This would however require additional measurements, as the reflection coefficients of the walls would be needed. This was also deemed to be beyond the scope of this project.

4.3 Angles used for the directional Helmholtz kernel

The angles used for the directional Helmholtz kernel are calculated from the same reference position to the sound source location for all positions. While the initial idea behind the implementation of this kernel was to use it for plane waves, in this project we used it for spherical waves, which is why the reference position was introduced. In the case of a plane wave, the direction of the sound wave propagation can be estimated through various methods and will be the same for all microphones, but for the spherical wave this will not be the case. There are some possibilities to improve the directional Helmholtz kernel by introducing a new weighting function, which would then use the direction and possibly even the distance to the sound source into account.

However, for real applications, it is commonly the case that the distance to the sound source is unknown and hard to estimate, while the general direction to the sound source, from all microphones, is much easier to estimate. Thus, this implementation was never fully explored.

4.4 Improvements to the Helmholtz kernel

The kernel used in this work uses the Bessel functions, spherical and non-spherical for the 3D and 2D case, respectively. The reasoning behind the kernel comes from the expansion of a wave around a point r_0, as defined earlier in the report, as

$$u(\boldsymbol{r}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{u}_{\nu,\mu}(\boldsymbol{r}_0)\varphi_{\nu,\mu}(\boldsymbol{r} - \boldsymbol{r}_0).$$

What was later shown was that this could be expanded around any point in Ω, and by expanding around an arbitrary point $\boldsymbol{r}' \in \Omega$, the resulting expression becomes

$$u(\boldsymbol{r}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{u}_{\nu,\mu}(\boldsymbol{r}')\varphi_{\nu,\mu}(\boldsymbol{r} - \boldsymbol{r}') \tag{4.1}$$

$$= \tilde{u}_{0,0}(\boldsymbol{r}')\varphi_{0,0}(\boldsymbol{r} - \boldsymbol{r}') + \sum_{\nu=1}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{u}_{\nu,\mu}(\boldsymbol{r}')\varphi_{\nu,\mu}(\boldsymbol{r} - \boldsymbol{r}'). \tag{4.2}$$

Now, by looking at the expansion around the point of evaluation, i.e., letting $\boldsymbol{r}' = \boldsymbol{r}$, the second term becomes 0, and we get

$$u(\boldsymbol{r}) = \tilde{u}_{0,0}(\boldsymbol{r})\varphi_{0,0}(\boldsymbol{r} - \boldsymbol{r}) = \tilde{u}_{0,0}(\boldsymbol{r})j_0(k\|\boldsymbol{r} - \boldsymbol{r}\|). \tag{4.3}$$

Essentially, by defining the kernel as $\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2) = j_0(k\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|)$, as was done in this report, we model the behaviour in the measured points, i.e., where we have microphones, by letting the expansion coefficient be the measured sound pressure. For these discrete points, the estimate gives us the actual expansion of the wave, with the expansion coefficient $\tilde{u}_{0,0}$ being the sound pressure in the measured positions. However, the resulting sound field is approximated by a sum of waves expanded around the points of the microphones; when evaluating the points that are not the microphone positions in the field, the second term in (4.2) does not disappear. Thus, the resulting interpolation method used in this report only accounts for the zeroth-order expansion of the real wave, expressed in terms of the basis function $\varphi_{\nu,\mu}$. In order to get a more accurate representation of the real sound field, higher order estimates of the basis functions would be needed. Something similar, but by using a Bayesian framework with infinite-order analysis, was done in [38].

Worth pointing out is that while the resulting sound field is not accurate in the entire field, the zeroth-order component used corresponds to the only component in the expression that is independent of the direction of propagation of the wave, which likely is why it was proposed this way to begin with; using higher orders makes the problem significantly much more complex to deal with.

Chapter 5

Conclusion

A simulated acoustic environment was set up, and an interpolation technique through kernel ridge regression was implemented. A similar experiment was also set up in a real acoustic environment. Three different kernels were introduced, and the directionally weighted kernel performed the best in essentially all aspects for all of the experiments. However, this kernel is dependent on prior knowledge of the direction of the propagating sound wave. The robustness of the KRR technique was evaluated for simulated data with noise in the measured signals, measurement errors for the microphones, measurement error for the the sound source position, distance to the sound source as well as different values of the directional regularization parameter β for the directional Helmholtz kernel.

The measured data was also evaluated as a function of the directional regularization parameter β, and compared to a simulated case using the same setup as the measured data. The optimal β value was found to be around 7 and 17 for the measured and simulated data, respectively.

Finally, the proposed method, combining the KRR method with a pre-calibrated HRTF, was evaluated and compared to the KRR method without the use of the HRTF, as well as a function of β. For the real experiments, the optimal β value was found to be around 3, while for the simulated case with the same setup, the parameter increased the performance for very high values of β. The proposed method was shown to improve the performance for both simulated and real data, for all frequencies, and for all values of the directional parameter β.

While it is hard to generalize the result to any microphone and sound source setup, this project could be used as a foundation to similar experiments.

Chapter 6

Appendix

6.1 The Helmholtz kernels in two dimensions

As the kernels primarily used in the project were defined for the three-dimensional (3D) case, this section will contain a brief comment on the two-dimensional (2D) case. The deduction of the kernel is close to identical to the previously presented kernels, where the only difference is in the initial differential equation that is used to define the basis function $\varphi_{\nu,\mu}(\cdot)$. For the 3D case, spherical coordinates are used, while in the 2D case polar coordinates are used, which slightly alters the solution. For more details on the derivation of the solution, the reader is referred to [4, 7], as this kernel will only be used for illustrative purposes in this report. However, this kernel can be used to model for instance water waves and seismic waves. The 2D Helmholtz kernel is defined as

$$\kappa(\boldsymbol{r}) = J_0(k\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|), \tag{6.1}$$

where J_0 denotes the zeroth-order Bessel function of the first kind. The only difference between the 2D case and the 3D case is that the 3D case uses the zeroth-order *spherical* Bessel function of the first kind. Similarly, the directional Helmholtz kernel is defined as

$$\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2) = J_0\Big(\sqrt{\big(i\beta\cos\phi - kr_x\big)^2 + \big(i\beta\sin\phi - kr_y\big)^2}\Big), \tag{6.2}$$

as shown in [4]. Note that the directional weighting function in this case does not contain $C(\beta)$, as this is a normalizing constant for three dimensions. Here, ϕ is the polar angle from a chosen reference position to the sound source.

6.2 Measurements from the anechoic chamber

Below are the coordinates of the microphones, the KEMAR microphones as well as the speaker. These are the final positions after the rotation is complete, as well as the change in the ear positions to fit the real measurements of the KEMAR manekin. Note that after the rotation, the speaker position was changed by -0.2 in the x-direction, -0.2 in the y-direction and -0.1 in the z-direction to more accurately represent where it actually was located in the room. This had a very minor impact on the final result, and the same conclusion would still hold without this change.

Table 6.1: Position of all the calculated points from the measurements in the anechoic chamber.

	x-coordinate	y-coordinate	z-coordinate
Microphone 1	-0.1437	0.4849	-0.0015
Microphone 2	0.1834	0.1575	0.4284
Microphone 3	0.1163	-0.5258	0.0914
Microphone 4	0.3162	-0.3577	0.1624
Microphone 5	0.2920	-0.0273	0.1867
Microphone 6	0.3256	-0.3689	0.3600
Microphone 7	-0.2272	0.1765	0.1367
Microphone 8	-0.1844	-0.3365	0.0739
Microphone 9	0.0975	0.0052	0.5025
Microphone 10	-0.3049	0.1771	0.4842
Microphone 11	0.2443	0.2557	-0.0309
Microphone 12	-0.2794	-0.4140	0.4280
Left ear	0	0.09	0
Right ear	0	-0.09	0
Speaker	0.8396	-0.8508	2.4578

In Figure 6.1, the speaker and the microphones can be seen.

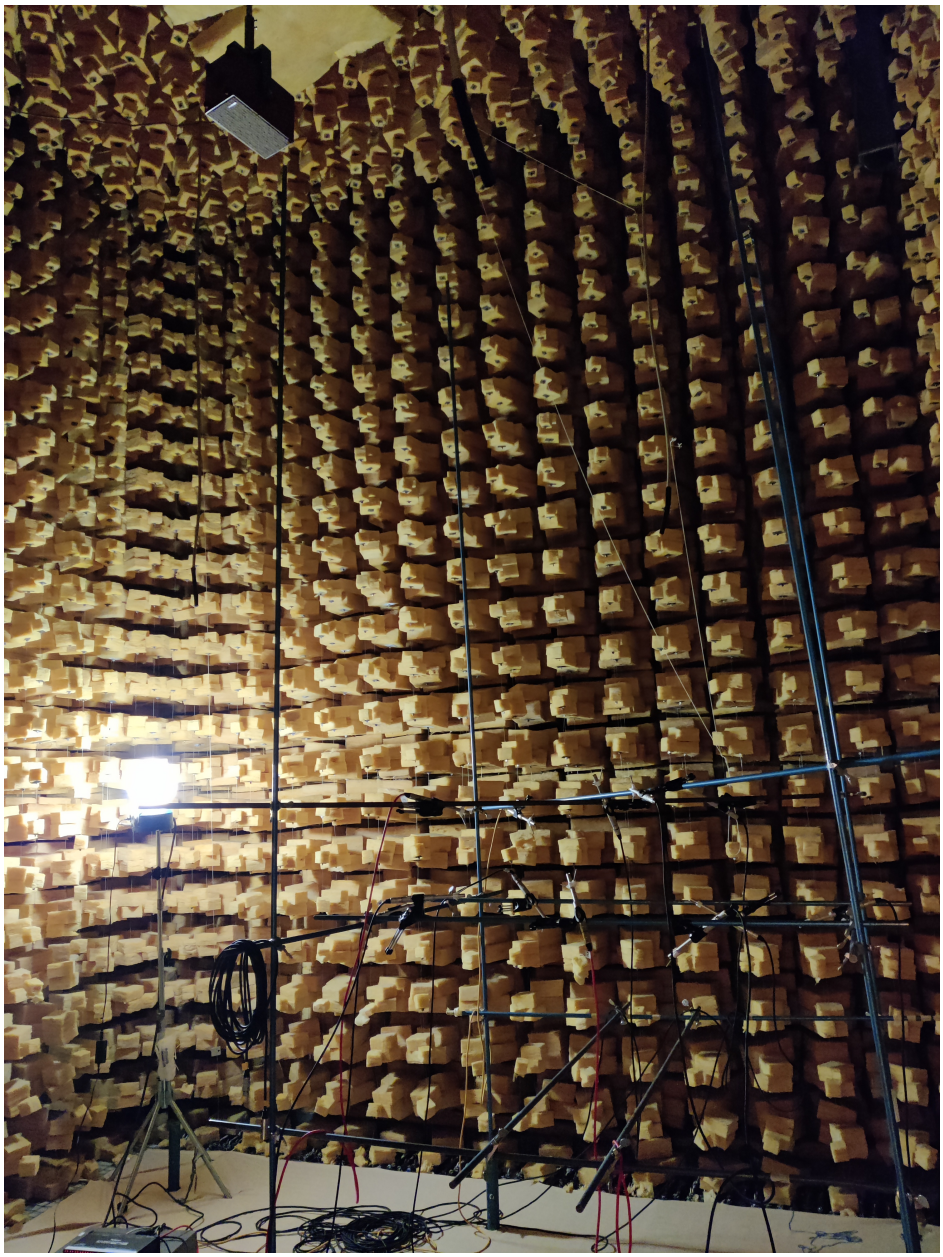

Figure 6.1: The speaker, as well as the microphones, without the KEMAR manikin present.

References

[1] J. Antoni, "A bayesian approach to sound source reconstruction: optimal basis, regularization, and focusing," *The Journal of the Acoustical Society of America*, vol. 131, no. 4, pp. 2873–2890, 2012.

[2] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation," in *16th International Workshop on Acoustic Signal Enhancement*, Tokyo, September 2018, pp. 436–440.

[3] K. P. Murphy, *Machine learning: A probabilistic perspective*, MIT Press, Cambridge, USA, 2012.

[4] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Spatial active noise control based on kernel interpolation with directional weighting," in *IEEE Int. Conf. on Acoustic, Speech and Signal Proc.*, 2020, pp. 8404–8408.

[5] R. Horiuchi, S. Koyama, J. G. C. Ribeiro, N. Ueno, and H. Saruwatari, "Kernel learning for sound field estimation with L1 and L2 regularizations," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, October 2021, pp. 261–265.

[6] S. J. Elliott, W. Jung, and J. Cheer, "Causality and robustness in the remote sensing of acoustic pressure, with application to local active sound control," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, May 2019, pp. 8484–8488.

[7] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Feedforward Spatial Active Noise Control Based on Kernel Interpolation of Sound Field," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Brighton, May 2019, pp. 511–515.

[8] S. Koyama, J. Brunnström, H. Ito, N. Ueno, and H. Saruwatari, "Spatial Active Noise Control Based on Kernel Interpolation of Sound Field," *IEEE/ACM Trans. on Audio Speech and Language Processing*, vol. 29, pp. 3052–3063, 2021.

[9] F. M. Heuchel, D. Caviedes-Nozal, J. Brunskog, F. T. Agerkvist, and E. Fernandez-Grande, "Large-scale outdoor sound field control," *J. of the Acoustical Society of America*, vol. 148, no. 4, pp. 2392–2402, 2020.

[10] F. M. Heuchel, *Sound Field Control for Outdoor Concerts*, Ph.D. thesis, Technical

University of Denmark, 2020.

[11] J. Donley and C. Ritz, "An efficient approach to dynamically weighted multizone wideband reproduction of speech soundfields," in *Summit and International Conference on Signal and Information Processing*, Chengdu, China, 2015, pp. 60–64.

[12] H. Gustafsson, *Speech enhancement for mobile communications*, Ph.D. thesis, Blekinge Institute of Technology, 2000.

[13] G. Ramos, M. Cobos, B. Bank, and J. A. Belloch, "A parallel approach to HRTF approximation and interpolation based on a parametric filter model," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1507–1511, 2017.

[14] R. Wu and G. Yu, "Improvements in HRTF dataset of 3D game audio application," in *Int. Conf. on Audio, Language and Image Proc.*, Shanghai, 2016, pp. 185–190.

[15] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs [head related transfer functions]," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Montreal, 2004, vol. 4, pp. iv–iv.

[16] F. P. Freeland, L. W. P. Biscainho, and P. S. R. Diniz, "Interpolation of head-related transfer functions (HRTFS): A multi-source approach," in *European Signal Processing Conf.*, Vienna, 2004, pp. 1761–1764.

[17] Y. Ito, T. Nakamura, S. Koyama, and H. Saruwatari, "Head-related transfer function interpolation from spatially sparse measurements using autoencoder with source position conditioning," in *Int. Workshop on Acoustic Signal Enhanc.*, Bamberg, 2022, pp. 1–5.

[18] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.

[19] F. Keyrouz, "Advanced binaural sound localization in 3-D for humanoid robots," *IEEE Trans. on Instr. and Measurement*, vol. 63, no. 9, pp. 2098–2107, 2014.

[20] J. Brunnström and S. Koyama, "Kernel-interpolation-based filtered-X least mean square for spatial active noise control in time domain," in *International Conference on Acoustics, Speech and Signal Processing*, Toronto, 2021, pp. 161–165.

[21] M. C. Saleh, B. E. A.; Teich, *Fundamentals of Photonics*, John Wiley Sons, New York, 2019.

[22] F. Dinuzzo and B. Schölkopf, "The representer theorem for Hilbert spaces: a necessary and sufficient condition," in *Advances in Neural Information Processing Systems*. 2012, vol. 25, Curran Associates, Inc.

[23] E. G. Williams, "Chapter 2 - plane waves," in *Fourier Acoustics*, pp. 15–87. Academic Press, London, 1999.

[24] A. Kirsch and F. Hettlich, *The Mathematical Theory of Time-Harmonic Maxwell's Equations: Expansion-, Integral-, and Variational Methods*, Applied mathematical sciences. Springer, New York, 2014.

[25] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, 1964.

[26] P. A. Martin, *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles*, Cambridge University Press, New York, 2006.

[27] N. Ueno, S. Koyama, and H. Saruwatari, "Directionally weighted wave field estimation exploiting prior information on source direction," *IEEE Trans. on Signal Processing*, vol. 69, pp. 2383–2395, 2021.

[28] F. G. Vasquez and C. Mauck, "Approximation by Herglotz wave functions," 2017.

[29] J. Bland, "The half-normal distribution method for measurement error: two case studies," January 2005, University of York.

[30] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," vol. 97, no. 6, pp. 3907–3908, 1995.

[31] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, and S. Sato, "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoustical Science and Technology*, vol. 35, no. 3, pp. 159–165, 2014, available at `https://www.riec.tohoku.ac.jp/pub/hrtf/`.

[32] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," pp. 99–102, 2001.

[33] J. Arend and C. Pörschmann, "Synthesis of near-field HRTFs by directional equalization of far-field datasets," March 2019.

[34] S.T. Birchfield and A. Subramanya, "Microphone array position calibration by basis-point classical multidimensional scaling," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 5, pp. 1025–1034, 2005.

[35] M. Larsson, V. Larsson, K. Åström, and M. Oskarsson, "Optimal trilateration is an eigenvalue problem," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Brighton, 2019, pp. 5586–5590.

[36] S. Kopuz and N. Lalor, "Analysis of interior acoustic fields using the finite element method and the boundary element method," *Applied Acoustics*, vol. 45, no. 3, pp. 193–210, 1995.

[37] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[38] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 135–139, 2018.

Master's Theses in Mathematical Sciences 2023:E10
ISSN 1404-6342

LUTFMS-3468-2023

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/