

Enhanced Leak Localization in Headphones Through Beamforming and 3D Modelling

Master Thesis



Enhanced Leak Localization in Headphones
Through Beamforming and 3D Modelling

Master Thesis
March, 2024

By
Zhengmeng Li

Copyright: Reproduction of this publication in whole or in part must include the customary bibliographic citation, including author attribution, report title, etc.

Cover photo: Vibeke Hempler, 2012

Published by: DTU, Department of Electrical and Photonics Engineering, Ørsteds Plads, Building 352, 2800 Kgs. Lyngby Denmark
electro.dtu.dk

ISSN: [0000-0000] (electronic version)

ISBN: [000-00-0000-000-0] (electronic version)

ISSN: [0000-0000] (printed version)

ISBN: [000-00-0000-000-0] (printed version)

Approval

This thesis has been prepared over six months at the Section for Indoor Climate, Department of Civil Engineering, at the Technical University of Denmark, DTU, in partial fulfilment for the degree Master of Science in Engineering, MSc Eng.

It is assumed that the reader has a basic knowledge in the areas of statistics.

Zhengmeng Li - s212629

.....
Signature

.....
Date

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgements

Zhengmeng Li, MSc Acoustic Engineering, DTU
Student of Engineering acoustic, master thesis candidate

WooKeun Song, [Title], [affiliation]
[text]

Viecente, [Title], [affiliation]
[text]

Contents

Preface	ii
Abstract	iii
Acknowledgements	iv
1 Introduction	1
2 Theory	3
2.1 Intro	3
3 Mesh model	9
3.1 Manual registration	9
3.2 Global registration	10
3.3 From point cloud generate the mesh model	11
3.4 Normal vector estimation	11
4 Set of system	13
4.1 Device of measurement	13
4.2 Device under test	13
4.3 Lab environment	13
4.4 measurement setup	15
4.5 verification of 3d scan	15
4.6 verification of beamforming	15
4.7 Setup	15
4.8 Procedure	15
4.9 Stimulus	15
5 Measurement	19
5.1 Array alignment calibration	19
5.2 Array recording	19
5.3 Verification through Adam A5x	20
5.4 Measurement of Jabra evolve2 65 (HP2)	21
5.5 Measurement of Jabra Envolve2 75 (HP1)	30
5.6 Measurement of Vue smart glass	31
6 Reverse sound field simulation by BEM	33
6.1 Validation	33
6.2 Theory	34
6.3 Calculation Settings	37
6.4 Results	37
7 Conclusion	41
7.1 time consumption	41
7.2 accuracy	41
Bibliography	42
A Appendix	43
A.1 Additional Beamforming Plots	43

A.2	Adam A5x Beamforming test	44
A.3	Radiation Pattern simulation for HP2	44
A.4	Radiation pattern simulation and corresponding eigenvalues for HP1	44
A.5	Radiation pattern simulation and corresponding eigenvalues for Vue glass .	44

1 Introduction

The phenomenon of sound leakage in headphones, characterized by the escape of audio from the earcup area to the external environment, poses a multifaceted challenge in the realm of acoustic engineering. This issue not only breaches the listener's privacy by making personal audio content inadvertently accessible to others nearby but also detracts from the overall audio quality experienced by the user. Furthermore, sound leakage contributes to environmental noise pollution, complicating the quest for serene and noise-controlled spaces in urban settings. Addressing sound leakage is therefore paramount for headphone manufacturers focused on optimizing user experience and for consumers seeking superior audio performance across diverse environments.

The detection and analysis of sound leakage in headphones have evolved with advancements in acoustic technology. This study leverages a sophisticated approach that combines the use of a microphone array and 3D scanning capabilities provided by the Microsoft Azure Kinect. Through these technologies, we can accurately collect acoustical and geometrical information, enabling 3D beamforming analyses, and get the sound pressure reconstruction on the surface of scanned object. This method allows for precise identification of the location and strength of sound leakage, offering a intuitive and convenient insights into the acoustic performance of headphones.

Brief introduction to the Device Under Test System

Within this research framework, the Device Under Test (DUT) pertains to the headphones being evaluated for sound leakage. The study encompasses three distinct categories of DUTs: headphones, for their leakage detection capabilities; dual loudspeakers, to verify the system's performance across different frequency ranges; and smart glasses, to demonstrate the evaluation process for an "open radiation" system. The testing arrangement features a B&K 36 channel planar microphone array positioned to capture potential sound leakage information at relative far field (from 0.5m to 1m), with geometric data gathered via Azure Kinect. The mesh grid of scanned object will then be used as the beamforming grid, with different beamforming algrithrum reconstruct back to the individual grid point. The entire setup of the testing sytem as shown in (fig).

The precise virtual alignment with the microphone array. This setup guarantees a thorough and detailed acoustic field mapping around the headphones, ensuring accurate leakage detection.

Understanding General Leakage Mechanisms

Sound leakage in headphones can occur through various mechanisms, including diffraction, reflection, and transmission of sound waves through the headphone materials. Understanding these mechanisms is crucial for identifying potential leakage paths and developing strategies for their mitigation. This study explores the fundamental principles behind sound leakage, providing a theoretical framework that underpins the experimental analysis.

Overview of Measurement Systems

The measurement system is at the heart of this research, comprising hardware and software components tailored for acoustic analysis. The hardware includes the microphone

array and the Azure Kinect for 3D scanning, while the software component is responsible for data acquisition, processing, and analysis. This system is designed to be flexible and adaptable, capable of analyzing a wide range of headphone designs under various conditions.

Workflow and Measurement Environment

The workflow for measuring sound leakage involves several key steps, starting with the preparation of the DUT and the calibration of the measurement system. The process includes the placement of the DUT within the test area, 3D scanning to capture its geometrical attributes, and the recording of sound emissions using the microphone array. The measurement environment is carefully controlled to minimize external noise and reflections, ensuring the accuracy and reliability of the data collected. This controlled setting allows for the replication of real-world usage scenarios, providing valuable insights into the acoustic performance of headphones.

Conclusion

By employing a combination of cutting-edge technologies and rigorous methodologies, this study aims to shed light on the complex phenomenon of sound leakage in headphones. The detailed examination of leakage mechanisms, coupled with the innovative use of 3D beamforming, sets the stage for significant advancements in headphone design and acoustic analysis. Through this research, we aspire to not only understand the intricacies of sound leakage but also to propose effective solutions for its mitigation, ultimately enhancing the listening experience for users.

my point of introduction 1, what I want to include inside of the thesis

Why do this project, what will be the use of such system - intro to the sound leakage to the headphone, what is defined as a leakage, and what is the characteristic of a leakage Why and how people do to detect the noise leakage - from the simplest method to the modern way. - only about headphone leakage, but I have to admit that the beamforming method are inspired by car and sub-water acoustic system, and then some discussion about that Then I should not get too much into the detail, but I should give a introduction of beamforming - the 2d beamforming with 2d microphone array, and 3d beamforming - and then where to get the 3d model - scan by azure kinect. So in the end, this thesis is aim to introduce the setting of a measurement system, which can be used to obtain 3d beamforming result, including a detailed measurement workflow to give a brief idea to the users. In order to obtain a result with super resolution, the beamforming is not only limited with DAS, but MUSIC, deconvolution beamforming, Compressive sensing are included, this thesis will demonstrate and compare the result by different algorithm.

This template complies with the DTU Design Guide <https://www.designguide.dtu.dk/>. DTU holds all rights to the design programme including all copyrights. It is intended for two-sided printing. The \cleardoublepage command can be used to ensure that new sections and the table of contents begins on a right hand page. The back page always ends as an odd page.

All document settings have been gathered in Setup/Settings.tex. These are global settings meaning the settings will affect the whole document. Defining the title for example will change the title on the front page, the copyright page and the footer. A watermark can be enabled or disabled in Setup/Preamble.tex. You can edit the watermark to display draft, review, approved, confidential or anything else. By default the watermark is printed on top of the contents of the document and has a transparent grey colour.

2 Theory

There must be one section for the theory, note that the theory is not necessary to be full of equation, it also to be the concept, which will include the intro:

2.1 Intro

under the intro, I will have the introduction to the beamforming, and then the imagine function, as the conventional DAS, and DOA. then it comes to the Near field. which will be one of the final result. so I will have an additional section to have further explain.

Let's start with the basic wave equations.

The wave equation for sound propagation in a fluid medium can be derived from two fundamental principles:

1. Conservation of Mass, mass is Conservative, and we assume the fluid is incompressible, means the volume is a constant. so that,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \quad (2.1)$$

where ρ is the density of the medium and \vec{v} is the velocity vector of the fluid particles.

2. Conservation of Momentum (Euler's Equation):

$$\rho \left(\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \vec{v} \right) = -\nabla p \quad (2.2)$$

where p is the pressure in the fluid.

For sound waves in an ideal fluid (neglecting viscosity and heat conduction), we assume the fluid motion is irrotational, which allows us to introduce a velocity potential ϕ , such that $\vec{v} = \nabla \phi$. Combining the continuity equation and Euler's equation, and assuming small perturbations (linear acoustics), leads to the classical wave equation for the velocity potential:

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = 0 \quad (2.3)$$

where c is the speed of sound in the medium. The pressure perturbation p can be related to ϕ by $p = \rho \frac{\partial \phi}{\partial t}$.

Helmholtz Integral Equation (HIE) The Helmholtz equation arises when we consider time-harmonic oscillations of the wave equation, leading to a frequency-domain representation. For a time-harmonic pressure field $p(\vec{x}, t) = \Re \{P(\vec{x})e^{-i\omega t}\}$, where $P(\vec{x})$ is the spatial distribution of pressure and ω is the angular frequency, the wave equation transforms into the Helmholtz equation:

$$\nabla^2 P + k^2 P = 0 \quad (2.4)$$

where $k = \omega/c$ is the wave number. The Helmholtz Integral Equation (HIE) provides a solution to the Helmholtz equation in terms of boundary values, which is particularly useful for boundary element methods (BEM):

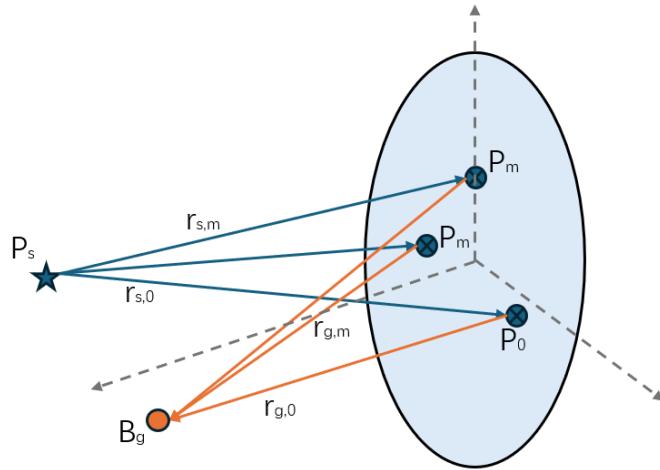


Figure 2.1: Demonstration of how microphone array receives the sound from source with sound strength P_s and B_g represent the autopower spectrum of the reconstructed output, which shall be used to represent the sound pressure level. P_0 is the reference point which is normally chosen at the center of the array.

2.1.1 Near field Delay and Sum Beamforming

so then I will intro the DAS in 3d object, this is for the explain of the dynamic resolution, how and why, and some demostration of the senario of dynamic resolution change under the distance.... to get the introduction of resolution, I have to give the resolution calculation.

For a Monopole in free-space injecting of a volume velocity at the source position. Additional volumes were added into the field at the location of the source r_0 , hence The equation of continuity is

$$\nabla \cdot \hat{\mathbf{u}}(\mathbf{r}) + \frac{j\omega \hat{p}(\mathbf{r})}{\rho c^2} = Q\delta(\mathbf{r} - \mathbf{r}_0) e^{j\omega t}.$$

- Euler's equation of motion is unchanged by the source, and these two equations result in the inhomogeneous Helmholtz equation:

$$\nabla^2 \hat{p}(\mathbf{r}) + k^2 \hat{p}(\mathbf{r}) = -j\omega \rho Q \delta(\mathbf{r} - \mathbf{r}_0) e^{j\omega t}.$$

The solution to this inhomogeneous Helmholtz equation under free-field conditions is

$$\hat{p}(\mathbf{r}) = \frac{j\omega \rho Q e^{j(\omega t - kR)}}{4\pi R}, \quad (2.5)$$

where $R = |\mathbf{r} - \mathbf{r}_0|$ is the distance to the source.

This equation describes how a pulsing source radiates sound into the field, and the sound pressure we can observe from this source is the pressure decaying with distance, and the phase homogeneous shifting with distance, depending on the wave number k .

In a practice scenario, we have the recordings from far field of unknown source location and unknown source strength. If we assume the source is a point source, Then we know the source's radiation and the sound pressure from this source that we can record in the field. Which then the eq. (2.5) can be reconstructed as,

$$P_m(r) \equiv P_0 v_m(r) = P_0 v (r_m - r) \quad (2.6)$$

where P_m is the recorded sound pressure at mth microphone at the microphone array, P_0 is the source strength, and $v(r)$ is the vector for the unknown source location, which is also called as "steering vector", given by the relation of the source to the :

$$v(r) = e^{-jk|r|/|r|}$$

Where k is the wave number, and r is the distance from the beamforming grid to the specific microphone.

Then, as in the fig. 2.1 shows, we can "delay and sum" the recorded sound pressure, scanning through the potential grid, which is applying the steering vector to all the microphone's recorded sound pressure in frequency, and sum the result and loop over all the points in the defined beamforming grid point to get the beamforming strength B_g , a higher B_g level represent a higher chance to have the actual source locate there.

From the recorded sound pressure at the microphone array point to the prediction of the specific grid point B_g , there are multiple approach to constructing such a steering vector. It is essentially projects the sound pressure from the microphone array to the grid point, as

$$p_F(x_t) = h(x_t)^H p \quad (2.7)$$

Where $h(x_t)$ is the steering vector, for turning the sound pressure matrix p from the array to the field point, H denotes for Hermitian transpose and $p_F(x_t)$ is the grid pressure

A general approach to derive B_g is to calculate the autospectrum of the projected field point, and we can find a relation between the steering vector

$$B(x_t) = E\{p_F(x_t)p_F^*(x_t)\} = h^H(x_t)E\{pp^H\}h(x_t) \quad (2.8)$$

$$= h^H(x_t)Gh(x_t) \quad (2.9)$$

Where G is the autospectrum of array pressure matrix. From this equation we can understand that now we have to find out the steering vector h for reconstruction.

A successful steering vector h should make beamforming output fulfill these two criterias [1]

Correct position: The predicted output should be maximum at the predicted source

$$B(x_t = x_s) > B(x_t \neq x_s).$$

Correct strength: The predicted source strength should be equivalent to the actual source strength.

$$B(x_t = x_s) = CE\{qq^*\}$$

where C is an arbitrary constant.

Where there is mutiple ways to reconstruct, but the most suitable approch for 3d beamforming scenario is by applying a normalization to the steering vector [1], Which will form an error function in between the actural crossspectrum of recording and the monopole projected crossspectum on the array. [2]

Here we should be able to construct an error function, that represents the error of which grid point gives the closest sound pressure to the actual CSM, which

$$E(a, r) = \sum_{m,n=1}^M |C_{nm} - C_{nm}^{mod}|^2 = \sum_{m,n=1}^M |C_{nm} - av_n^*(r)v_m(r)|^2 \quad (2.10)$$

Aiming of this equation is to find the source strength a to minimize the Error, one alternative way than directly solve this equation is to maximize the image function, which is in the form as below [2]

$$J^2(\omega, r) \equiv \frac{1}{\sqrt{M(M-1)}} \frac{\left| \sum_{m \neq n}^M C_{nm}(\omega) v_n(r) v_m^*(r) \right|}{\sqrt{\sum_{m \neq n}^M |v_n(r)|^2 |v_m(r)|^2}} \quad (2.11)$$

Where J is the image function with diagonal removal, and M represents the microphone, V_n and V_m is the steering vector. Since this equation represents the delay and sum of the array's microphone sound pressure, the J will then equally represent the source strength a .

so a is the source strength, from monopole in free space, it's related with the point source's pulsing velocity

$$P_0 = \frac{j\omega\rho Q}{4\pi} = a$$

then we can get Q

$$Q = \frac{4\pi a}{j\omega\rho}$$

Q is a idealized quantity, represent the monopole source pulsing strength at original point, which is not a realistic value. To make it more realistic, consider this Q as the vibration velocity of a piston on the baffle, then the volume velocity is the far-field approximation becomes

$$\hat{p}(r, \theta) = \frac{j\omega\rho Q e^{j(\omega t - kr)}}{2\pi r} \left[\frac{2 J_1(ka \sin \theta)}{ka \sin \theta} \right]$$

where the volume velocity of a piston has been introduced

$$(Q = \pi a^2 U)$$

The U here is the Q in monopole situation.

To calculate the velocity

$$\hat{\mathbf{u}} = \frac{-1}{j\omega\rho} \nabla \left(|\hat{p}| e^{j\varphi} \right)$$

$$\hat{\mathbf{u}} = \frac{-1}{j\omega\rho} (-e^{-j\varphi} \nabla |\hat{p}| + j e^{-j\varphi} |\hat{p}| \nabla \varphi)$$

So, the far-field pressure is

$$\hat{p}(r, \theta) = \frac{2I}{r} \cdot \pi a^2 \cdot e^{-jkr} \left[\frac{2 J_1(ka \sin \theta)}{ka \sin \theta} \right]$$

where a is the radius of piston, θ is the angle of observing point to the plane, k is the wave number, and I is the image function, as a result from beamforming reconstruction. r is the distance from the observing point to the grid point

2.1.2 Eigenvalue extraction of the Cross spectrum matrix

In the realm of acoustic signal processing, the Cross-Spectral Matrix (CSM) is pivotal for understanding the spatial distribution of signal energy and noise across a sensor array.

The CSM encapsulates the frequency-domain relationships between signals received at different sensor elements, thus serving as a foundational construct for eigenvalue-based signal enhancement techniques.

Eigenvalue Extraction and Signal Direction Amplification Eigenvalue extraction from the CSM reveals the principal directions of signal propagation and noise distribution. By decomposing the CSM through eigenvalue decomposition (EVD), we can separate the signal subspace from the noise subspace, which is crucial for isolating and enhancing signals of interest.

The EVD of the CSM is given by:

$$CSM = E\Lambda E^H$$

where E is a matrix whose columns are the eigenvectors of the CSM, and Λ is a diagonal matrix containing the corresponding eigenvalues. The eigenvectors represent the directions of signal and noise propagation, while the eigenvalues indicate the relative strength or energy along these directions.

Multiple Signal Classification (MUSIC) Beamforming

MUSIC Beamforming leverages the eigenvalue decomposition of the CSM to distinguish between the signal and noise subspace. It employs an inverse eigenvalue approach to amplify the directional of signal sources to their maximum extent by essentially applying a $1/\lambda$ weighting, where λ represents an eigenvalue associated with the signal subspace.

MUSIC Principle The essence of MUSIC lies in its ability to locate signal sources with high resolution by exploiting the orthogonality between the signal and noise subspace. The MUSIC spectrum is defined as:

$$P_{\text{Noise}} = a^H(\theta)E_nE_n^H a(\theta) \xrightarrow{x_t=x_s} 0 \quad (2.12)$$

$$P_{\text{MUSIC}}(\theta) = \frac{1}{a^H(\theta)E_nE_n^H a(\theta)} \xrightarrow{x_t=x_s} \infty$$

where $a(\theta)$ is the steering vector associated with a potential source direction θ , and E_n contains the eigenvectors corresponding to the noise subspace.

The peaks in the MUSIC spectrum indicate the estimated Directions of Arrival (DOAs) of the signal sources. This method excels in localization accuracy but does not directly provide estimates of source strength due to the inverse eigenvalue weighting.

Functional Beamforming

In contrast, Functional Beamforming aims to optimize localization while preserving estimates of source intensity. This approach modifies the traditional beamforming formula by incorporating eigenvalue-based weighting directly into the beamforming process to enhance signal detection and improve source strength estimation.

Functional Beamforming Approach Functional Beamforming is formulated as follows:

$$b_v(g) = \left[g' C_v^{\frac{1}{v}} g \right]^v \quad (2.13)$$

$$P_{\text{Functional}}(\theta) = \left| \sum_{i=1}^M \lambda_i \cdot a_i^H(\theta) S a_i(\theta) \right|^2$$

Here, λ_i represents the eigenvalues, serving as weights to adjust the contribution of each direction θ based on the signal's strength or energy. The matrix S is a modified version of the CSM that incorporates these eigenvalue-based weights, enhancing the ability to localize sources accurately while also estimating their intensity.

2.1.3 Clean based on Source Coherence

Besides of our previously introduced

3 Mesh model

Switching from 2D to 3D beamforming uses three dimensional grids instead of two dimensional ones to more accurately reconstruct sound fields and determine the source of sounds. This approach is based on the idea that 3D point cloud data contains more detailed information about objects, making it possible to achieve better sound field restoration and localization. The growing research in 3D point clouds, driven by both industry and academia, supports this work. Specifically, this paper utilizes the Microsoft Azure Kinect Developer Kit [3] and its related open source SDKs [4], alongside various projects focused on 3D point cloud processing. This includes using the OpenCV rgbd module [5], the Visual Tool Kit (VTK) [6] for image processing, OpenCL for accelerating the image processing [7], and Open3D [8] for tasks like point cloud processing, mesh generation, and manual point cloud registration. These technologies help measure distances between objects in a scene and accurately represent their shapes, thanks to ongoing improvements in software, hardware, and increased computational power, allowing for real time object scanning and straightforward post processing of scan results for beamforming.

The Microsoft Azure Kinect DK is essential to our approach, featuring a high-definition depth sensor and an Inertial Measurement Unit (IMU) for orientation and spatial tracking. These enable the capture of depth data at 15 frames per second, sufficient for our applications, and track the sensor's movement. This movement data provides a basic frame-by-frame registration matrix, which, through ICP registration, leads to an accurate 3D representation of scanned objects. This process relies on the Kinfu-example[4], incorporating OpenCV's rgbd module, VTK, and OpenCL technologies as we just mentioned.

3.1 Manual registration

The procedure described in section 3.1 starts with capturing the point cloud data for an object and a microphone array, with the next step being the identification of specific microphone points within the array's point cloud representation. Identifying these points is necessary for the calculation of the steering vector, which relies on the distances from grid points to those on the array. To facilitate this identification, manual registration is employed, allowing users to select corresponding points across two distinct point clouds. The accuracy of aligning the point clouds hinges on these selected points. In our approach, determining three pairs of corresponding points has been found to provide a transformation matrix with adequate accuracy for our needs. Nonetheless, accurately pinpointing the exact points on the scanned array can be challenging due to scanning limitations. However, any inaccuracies in initial registration can be adjusted in subsequent steps, as there will be a reference source for correction aids in ensuring the precise placement of virtual microphones at their intended positions, thus accommodating the initial selection's potential inaccuracies.

For the manual registration, it leverages Open3D's functionality for calculating the transformation matrix, a critical component in the manual registration process. The transformation matrix utilized is known as the homogeneous transformation matrix, characterized by a 4×4 configuration that encapsulates both rotation and translation information required for point cloud alignment. The matrix is structured as follows:

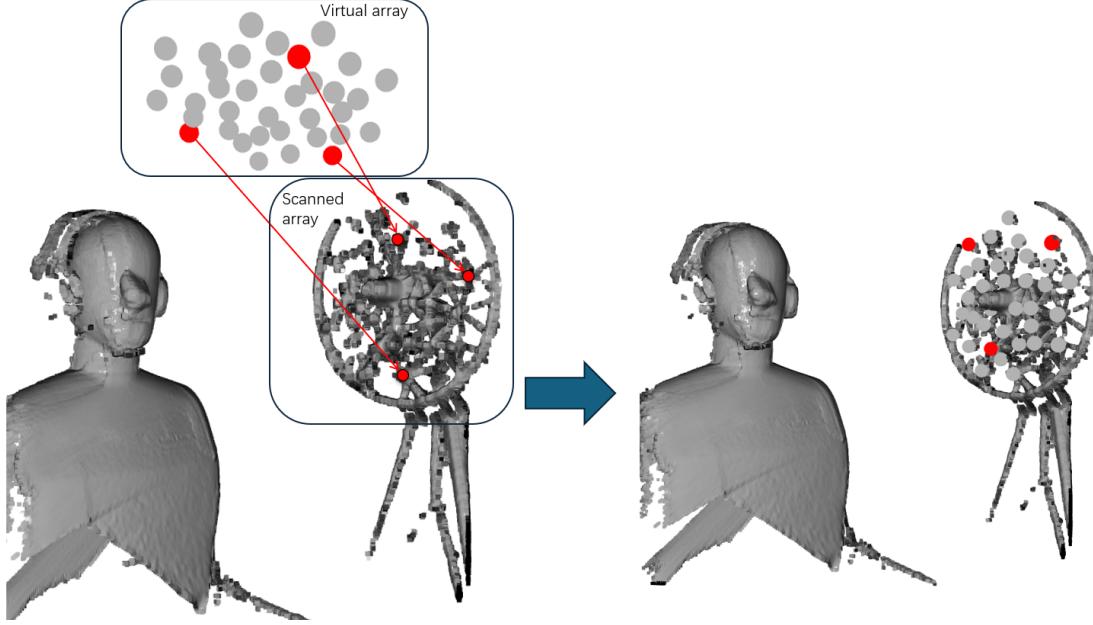


Figure 3.1: Process of manual registration. The red point in both the virtual array and the scanned array is the corresponding point that should be manually selected. After the translation, the virtual array will be panned 20mm towards the normal direction so that its center points fall exactly on the head of the microphone in the actual array

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Here, the 3×3 submatrix R represents rotation, and the vector $T = [T_x, T_y, T_z]$ represents translation. The last row is used to make the matrix compatible with homogeneous coordinates, allowing for both rotation and translation to be performed with a single matrix multiplication.

3.2 Global registration

The process of global registration in Kinect Fusion is a critical step towards synthesizing a coherent 3D model from sequential depth frames. This procedure is fundamentally composed of two stages: an initial coarse alignment using the 3D camera's trajectory, followed by a refinement phase employing Iterative Closest Point (ICP) registration. These stages are seamlessly integrated to facilitate accurate and efficient point cloud merging, accommodating the dynamic nature of the captured environments.

3.2.1 Primary registration based on 3d camera trajectory

At the outset, the primary registration capitalizes on the estimated trajectory of the Kinect camera to compute a transformation matrix that serves as a foundational alignment tool. This matrix is derived from the motion data of the camera, which encompasses its spatial movements and rotations over time. By applying this transformation, the current depth frame is preliminarily aligned with the previously accumulated global model. This coarse alignment is instrumental in establishing a unified coordinate framework for the sequential integration of depth frames, although it generally lacks the precision to account for fine geometric details.

3.2.2 ICP registration

To enhance the alignment precision, the process employs ICP registration as a subsequent refinement step. This algorithm iteratively adjusts the current frame's alignment by minimizing the geometric discrepancy between it and the global model. It does so through an optimization procedure that iteratively identifies corresponding points between the two datasets and minimizes the sum of squared distances between these pairs. This meticulous adjustment process results in the current frame being precisely overlaid onto the global model, with the transformation parameters being updated until the alignment error converges to a minimal value.

The Iterative Closest Point (ICP) algorithm is a fundamental method in the field of computer vision and 3D reconstruction, designed to align or register two clouds of points. The goal of ICP is to find the best possible alignment between a source point cloud and a target point cloud. This process involves minimizing the distance between the corresponding points in these two datasets.

In this project, We had used this method under the 3d scanning process, which is for merge mutiple 3d scanned point clouds image into one point cloud model. This approach is provided by open source resipo from Microsoft's example project '*Kinfu_example*', which is utilizing the OpenCV's model '*OpenCV_rgbd*', and Visual toolbox '*VTK*'. In this project, we do not involved in any development of the ICP algorithm or any opensource computer vision project, but only compile and use it. But to make this very essential approach clear to every reader, the following is a brief explanation of the theory of ICP registration used in this project,

The core of the ICP algorithm can be broken down into a few key steps, each associated with specific equations:

1. Point Matching: Initially, the algorithm identifies the closest point in the target point cloud for each point in the source point cloud. This step does not involve a specific equation but relies on geometric proximity.
2. Error Metric Minimization: The core objective of ICP is to minimize the total squared difference between the matched points. The error metric can be represented as follows:

$$E(R, t) = \sum_{i=1}^n \|(R \cdot p_i + t) - q_i\|^2$$

Here, $E(R, t)$ represents the error metric dependent on the rotation matrix R and the translation vector t . p_i are the points in the source point cloud, and q_i are the corresponding matched points in the target point cloud.

3. Optimization: The algorithm seeks the optimal rotation R and translation t that minimize the error metric $E(R, t)$. This optimization is typically performed using the Singular Value Decomposition (SVD) method or similar approaches to solve for R and t that minimize the distance between matched point pairs.
4. Iteration: The process repeats, with the source point cloud being transformed by the newly found R and t , and the steps of matching and optimization being iterated until convergence is achieved, or the changes fall below a predefined threshold.

3.3 From point cloud generate the mesh model

3.4 Normal vector estimation

4 Set of system

This chapter will explain the system setup, including the equipment used, and then it will list the objects that were to be measured, and the measurement parameters (in which meters are the most important), and also it will include the 3d scanning setup, because this is not a computer graphic thesis, so I will not provide further details of the scanning, but rather only give a simple list of parameter setting, and then the resolution/time consumption/accuracy verification - for verification, it will scan some simple object, like a loudspeaker, and distance to some object, and compare the result with a real ruler measurement, no further detailed shall be provided to avoid discussion which might get beyond of the scope of this thesis.

4.1 Device of measurement

The device of measurement is consist of two system, which is the Microsoft Azure Kinect as mentioned in chapter 3, and the B&K's Acoustic camera system. The system is the same

4.2 Device under test

4.3 Lab environment

The experiment was conducted in HBK's laboratory in Virum, which is a standard room rather than an anechoic chamber. Ideally, such experiments should be performed in an anechoic chamber to ensure the most reliable results; however, due to resource limitations, this was not feasible.

Within the lab, there are multiple windows and other experimental setups. The primary source of noise originates from the front end of the microphone array. By utilizing the microphone array, we can record and average the recordings from different channels to obtain the background noise frequency spectrum, as illustrated in fig. 4.2. The noise spectrum indicates that the majority of the noise is concentrated below 1000Hz. From 1500Hz to 10kHz, the noise floor remains roughly constant at 0dB.

For the room response, including floor and ceiling reflections, the impact is expected to be minimal in most scenarios due to the test object's proximity to the microphone array and its elevated position from the floor. Furthermore, the precision of 3D beamforming, which projects sound pressure onto a grid with accurately defined geometry and relative distances, inherently reduces the influence of room reflections. Additionally, the inherent noise floor limits the measurement's feasible frequency range to between 1kHz and 10kHz. Within this range, the wavelength—at a minimum of 0.343m at a 20°C room temperature—ensures that room reflections do not significantly affect the beamforming results. The subsequent results presented offer a comprehensive overview, enabling us to determine the appropriate calculation range accurately.

Moreover, the measurement environment has been controlled to ensure no significant unexpected noise sources could impact the array's measurement.

Jabra Evolve2 75/65 -Both are on-ear headphones, with the Evolve2 75 (right) being the larger one, and the Evolve2 65 (left) being smaller and more portable. -In some Figures, these two were referred to as HP1 and HP2, respectively.	
Adam A5x -A Loudspeaker set is used for verification of the results.	
Vue smart glass -Smart glass from a crowdfunding project on 'Kickstarter' in 2016. -This subject can be considered a mixture of bone conduction and "open radiation" device.	
B&K Head and Torso Simulator (HATS) Type 5128 -The simulator can reproduce the acoustic properties of an average adult human head and torso. -With the Type 4620 ear simulator, it has the acoustic properties close to the real human ear	

Figure 4.1: List of device under test

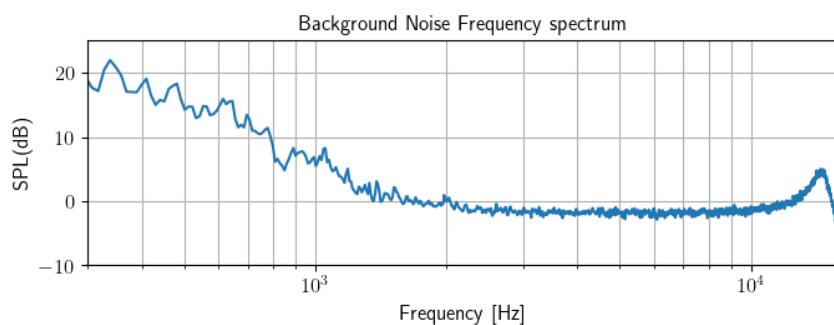


Figure 4.2: Background noise collected by the array, the frequency range is from 300Hz to 11kHz.

4.4 measurement setup

4.5 verification of 3d scan

4.6 verification of beamforming

in this section, it will include identification of two loudspeakers. here by it can be used to verify the accuracy of the measurement. No simulation will be included. Hope the measurement will be successful.

4.7 Setup

The setup is the setup of the lab, the data recurring system (bk connect, pulse labshop...) and the measurement signal the microphone array - tech detail the amplifier set of the signal - analog signal set of the mics which are used for reference signal measurement set of the signal - after correction, for the respiratory measurement.

4.8 Procedure

demo of the workflow of the measurement, include a workflow graph

4.9 Stimulus

To ensure accurate comparisons of sound leakage across different headphone systems, standardizing the output levels is essential. The variability in sound-emitting units among headphones affects sound leakage. Equalizing the output in the ear canal across various models normalizes the output levels [9], allowing for a direct comparison of leakage magnitudes.

The measurement of each headphone's output is critical, using white noise stimulus and assessing the frequency response through a microphone in a Brüel & Kjær Head and Torso Simulator (HATS). Consider the headphone's unit to the HATS microphone's microphone is an unknown system with transfer function $H(\omega)$, then when it is subjected to a white noise input signal $x(t)$, the output recorded at the HATS ear microphone $y(t)$ can be recorded. Transforming $x(t)$ and $y(t)$ to the frequency domain gives $X(\omega)$ and $Y(\omega)$, respectively. The transfer function $H(\omega)$ cannot be directly measured but can be estimated by analyzing the system's output in response to the known input. By Welch method, averaging the frequency response $\bar{Y}(\omega)$ obtained from multiple sequences of $x(t)$ to minimize noise and variations, resulting in an estimation of the system's behavior. The estimated transfer function $\hat{H}(\omega)$ can be approximated as:

$$\hat{H}(\omega) = \frac{\bar{Y}(\omega)}{X(\omega)}$$

where $\bar{Y}(\omega)$ represents the averaged output response to the white noise input.

With the estimated transfer function $\hat{H}(\omega)$, an FIR filter approximating $H^{-1}(\omega)$ is designed. This FIR filter acts as the inverse filter to equalize the system's response. The FIR filter is derived by seeking a filter that, when convolved with $Y(\omega)$, yields a response closely matching the original white noise input $X(\omega)$, effectively flattening the frequency response across the desired range. Applying the Inverse Filter for Equalization 1. Equalized Input Generation Applying the FIR filter approximating $H^{-1}(\omega)$ to $X(\omega)$ produces the equalized input $X^*(\omega)$:

$$X^*(\omega) = \text{FIR}_{H^{-1}(\omega)} * X(\omega)$$

where $*$ denotes convolution, and $\text{FIR}_{H^{-1}(\omega)}$ is the FIR filter designed to approximate $H^{-1}(\omega)$. 2. Achieving Flat Frequency Response Feeding $X^*(\omega)$ into the system results

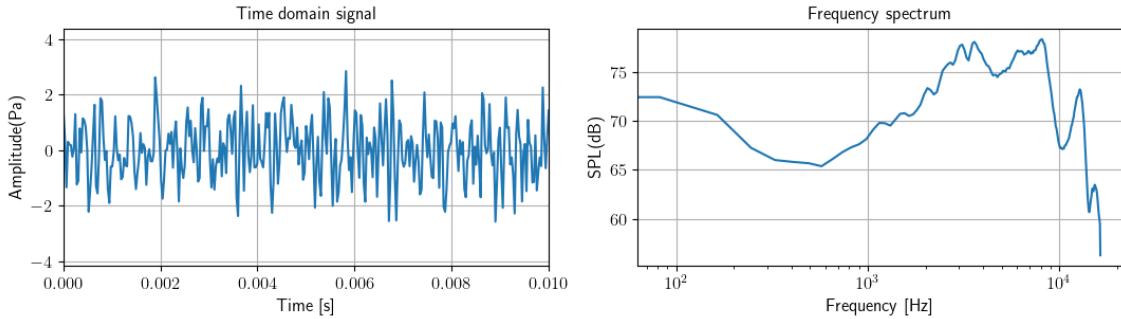


Figure 4.3: time and frequency response from the headphone to artificial ears microphone. The white band noise stimulus is band limited from 20Hz to 14kHz.

in an output $Y^*(\omega)$ that closely approximates the desired flat frequency response, ideally matching the input white noise in frequency content.

The reverse filter from the white noise frequency response as in the left plot of fig. 4.4. Where The inversion point in the filter frequency response is the 0dB point of the filter, which means that this point is the inversion reference point. Since it does not make sense to equalize parts that are too low or too high, for usability considerations, our strategy is to set all filter responses below the conversion point to 0dB, and for filter responses above 15dB are all set to 15dB so that we can build the filter within a reasonable range

As shown in the right panel of fig. 4.4, the frequency response of a finite impulse response (FIR) filter is derived from the square root of the frequency response of the filter shape. The derivation process is based on the minimum phase filtering principle, which aims to achieve the desired amplitude response while minimizing the phase response deviation. By taking the square root of the frequency response of the filter shape, we effectively prepare the FIR filter to be applied to a zero-phase filtering configuration. This method involves applying a filter forward and then in reverse, thus removing the phase distortion introduced during the first pass and ensuring that the phase of the signal is preserved. Since the filter is applied twice, the square root will give us the correct amplitude response in the final result.

The input stimulus to the Headphone are equalized, in order to obtain a flat response of all frequency range and in different devices. The procedure of equivalence are listed below All the recordings were collected by BK 4195 HATS, with the front-end and BK Connect Time data processing module. First, input a 20s band-limited white noise into the headphone, get the recorded time domain sound pressure in pascal. The result as below, can be seen as the response of the Headphone and HATS system's response.

The way to obtain the FIR filter response is the minimum phase, which means it's the square root of the FIR H, so when apply this filter we use the filtfilt method - the forward and reverse method to get the response without additional phase added onto the original signal.

So, the original white noise were filtered by this obtained FIR filter. To get the equalized stimuli. By this method, we can equalize the output from all measured headphones, so different headphone's sound leakage will become comparable.

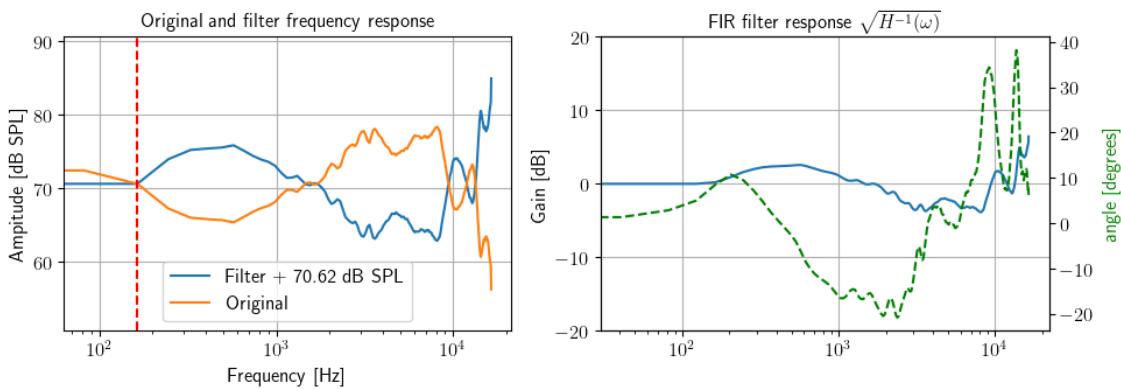


Figure 4.4: Magnitude and phase responses of a reverted filter and an FIR filter. The left plot shows the frequency response of both the filter and the original signal, with the filter's reversion point indicated by a red dashed line. The right plot displays the frequency response of the FIR filter, derived from the filter shape's square root value, alongside its phase response (green dashed line).

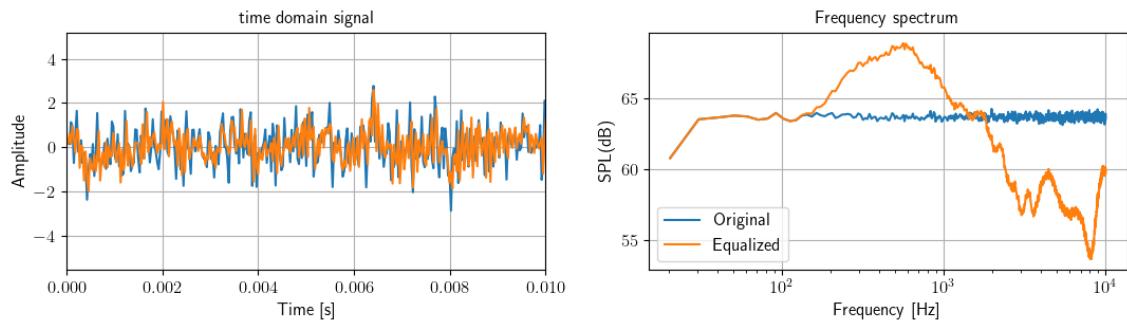


Figure 4.5: Original and Equalized white noise stimulus for HP2

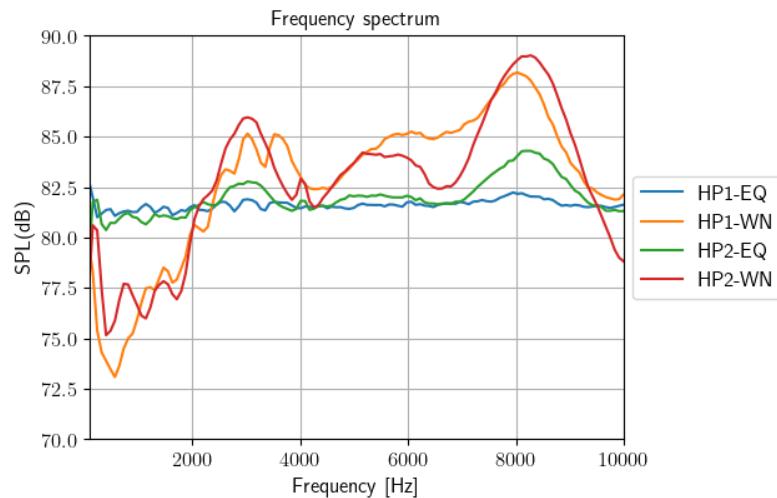


Figure 4.6: Comparison of the equalized response with the original frequency response under the white noise stimulus.

5 Measurement

In this chapter, I will repeat the measurement setup which are already listed in the previous chapter, but with more detail, in order to remind the reader and get a clear structure.

5.1 Array alignment calibration

Array alignment is a crucial step in correlating the recording position with the actual scanning location. A general manual alignment method is described in section 3.1. However, this method can suffer from inaccuracies due to the inherent error in manual selection and the subjective nature of choosing corresponding points. To mitigate errors associated with manual selection, we propose the introduction of a reference source with a known location, placed close to the object being measured. An additional step in the measurement process can then be implemented: by analyzing the beamforming image, we can identify the localization error of this known source. This error allows us to calibrate the microphone array's positioning to ensure it is correctly located.

For example, as shown in fig. 5.1, a small signal unit loudspeaker emitting an 8000Hz, 1/3 octave white noise stimulus was placed in front of the Head and Torso Simulator (HATS). At this frequency range, beamforming is known to be particularly robust. The measurement yielded a surface sound image map using the Beamforming Music method, with the dynamic range set to 1dB for precise localization. The source image was observed to be at some distance from the center of the loudspeaker membrane, which is the known actual source location. This discrepancy is attributed to the misalignment between the virtual array and the actual array. To correct this, we adjusted the microphone array's position to align with the correct location, as depicted on the right side of fig. 5.1. This adjustment indicates that we achieved correct array alignment.

This calibration procedure will be applied in all subsequent measurements, with the exception of the Adam A5x loudspeaker measurement. For the Adam A5x, its tweeter can also serve as a reference for array alignment calibration.

5.2 Array recording

In our experimental setup, the microphone array is positioned 0.75 meters away from the object under test. This distance, measured from the center of the array to the object's nearest point using a standard ruler, is rounded for convenience. Although this measurement method introduces a level of approximation, it is deemed sufficient for our purposes. The specified distance is a critical parameter in the 2D beamforming calculations performed by the NSI software. It is important to note that some degree of grid displacement inaccuracy is anticipated in 2D beamforming analysis, which is a common occurrence in such measurements.

The acoustic stimulus used varies depending on the device being tested. For headphones and smart glasses, the stimulus is equalized white noise, as detailed in section 4.9, at the HATS in-ear microphone it generates an 81dB sound pressure level. In contrast, the stimulus for loudspeaker measurements consists of two channels input which is uncorrelated white noise with a 0.1V voltage. This setup is expected to produce a sound pressure level of approximately 64dB at the location of the microphone array.

The recording duration for the microphone array is set to be 10 seconds, with a sampling

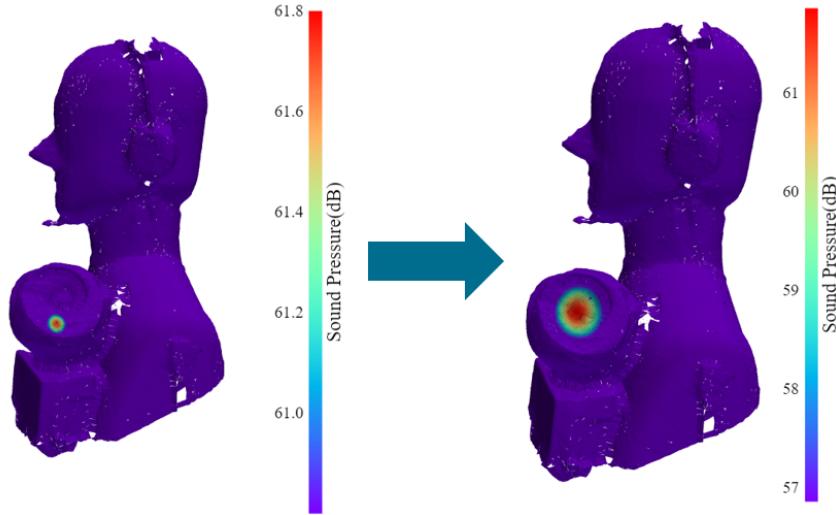


Figure 5.1: Array alignment calibration. The left plot is the beamforming image from uncalibrated array location, measured at 8000Hz, 1/3 octave bandwidth, and 1dB dynamic range; the right plot is the beamforming image after the array calibrated, with same frequency and bandwidth, but set to be 5dB dynamic range to visualize the source.

rate of 32,768 Hz. This configuration ensures that the data collected will be adequate for analyzing the measurement frequency range of interest.

The entire recording is processed by BK Connect, for the 2d beamforming calculation the recordings are transferred to NSI, and for the 3d beamforming calculation the recordings are exported in .mat format and processed in the python project dedicated created for this thesis in GitHub [10]

5.3 Verification through Adam A5x

As discussed in Section 2, the theory of beamforming involves two unknown variables: the location and the strength of the source. To validate our results, we conducted measurements using a known source, as the "ground truth", specifically the Adam A5x loudspeaker system. This system comprises two loudspeaker box, each box equipped with one woofer and one tweeter. According to a third-party measurements, which will be detailed further in Section 6, the woofer predominates at frequencies below 1000Hz, while the tweeter takes precedence above 2000Hz. In the intermediate frequency range of 1000Hz to 2000Hz, the vent is the dominant source, with significant contributions from both the woofer and the tweeter [11]

To further validate our measurement results, as well as form a comparison in regards of the performance of the accuracy and correction, we also do the measurement under the exact same configuration and calculate the 2d beamforming result by PULSETM NSI Array Acoustic post processing software [12]. This software provide a wide range of beamforming algorithm, those we will use is the DAS, Non-negative least square (NNLS), and CleanSC. We will have the 2d beamforming plot and then compare it with the 3d beamforming. In the NSI software, an initial photograph is taken at the start of the measurement using a camera mounted at the center of the array. This camera is calibrated to produce an image that can be overlapped with the beamforming image generated during the acoustic analysis. This overlapping of visual and acoustic data makes it directly correlates the identified acoustic source locations with the physical structure of the object

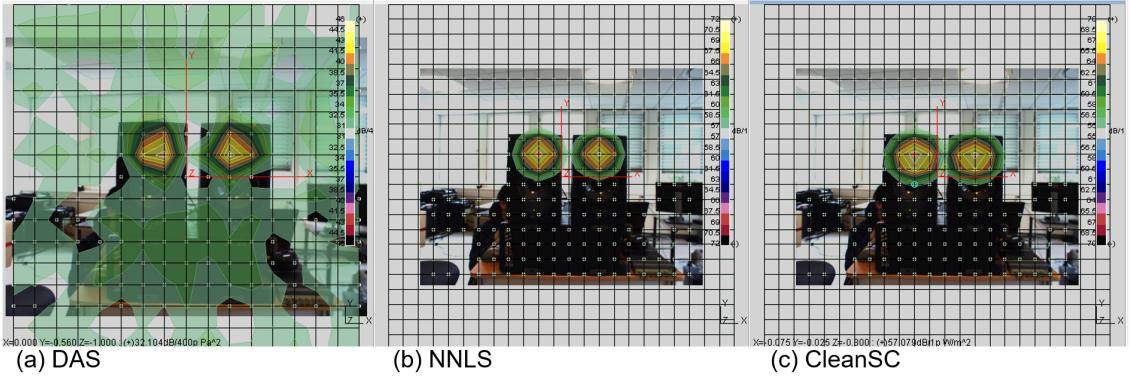


Figure 5.2: NSI-Array acoustic postprocessing result under 8000Hz, 15dB dynamic range

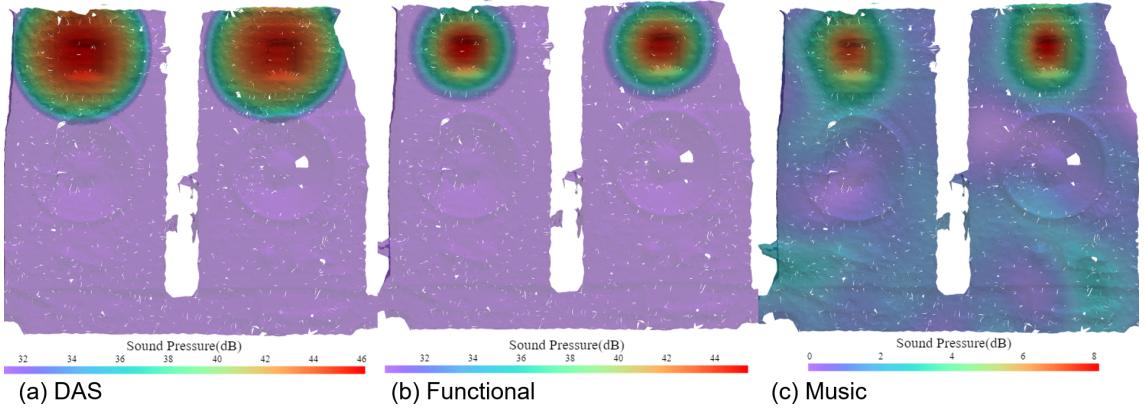


Figure 5.3: 3d beamforming result under 8000Hz, 15dB dynamic range.

under test.

The result from PULSE TM NSI Array Acoustic as shown in figs. 5.2, A.5 and A.6; The results by 3d beamforming as shown in figs. 5.3, A.7 and A.8. In the following text is only the results of 8000Hz, and the results of 2500Hz and 5000Hz is shown in appendix A.2

Beamforming CleanSC is filtering only the highest value out from the beamforming image, and suppress all the spacial correlated image to zero. So that the result from CleanSC if without the smoothing will only give a signal mesh point with cooresponding strength, which is very convenient to giving the location of the source, but then it will be very hard to visualize especially when doing the 3d beamforming because each point is very small and might be covered by others under specific viewing angle. But such algirtum is good for generating a general viewing of the entire frequency range, which means I can filter out the highest point location at every frequency point, and then plot them all togther within one beamforming plot. Beneficial from the naturally high resolution from the beamforming scanning, it can gives a clear trajectory of how the souce travel with frequency. Under that idea, the fig. 5.4 shows how the highest source travelling when measure only one box of the Adam A5x. Which is shows a very clear trend of movement, that under the low frequency the source (colored as blue)

5.4 Measurement of Jabra evolve2 65 (HP2)

One major distinction between headphone and loudspeaker measurements lies in their geometry. In loudspeaker measurements, we typically deal with two nearly flat surfaces

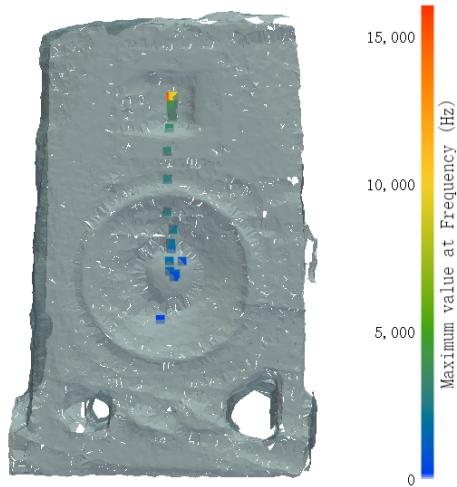


Figure 5.4: The highest beamforming value obtained by CleanSC as a location response to frequencies.

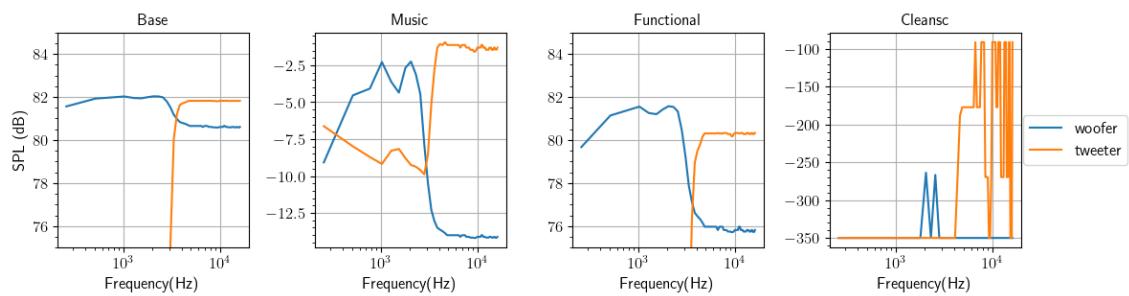


Figure 5.5: The frequency response of two point set, with one at the woofer, and another one at the tweeter location. Results from 4 different beamformers for comparison. The point set is picked based on the result shown in fig. 5.4, first 7 highest point index (as woofer) and last 7 highest point index (as tweeter) were picked and averaged by values.

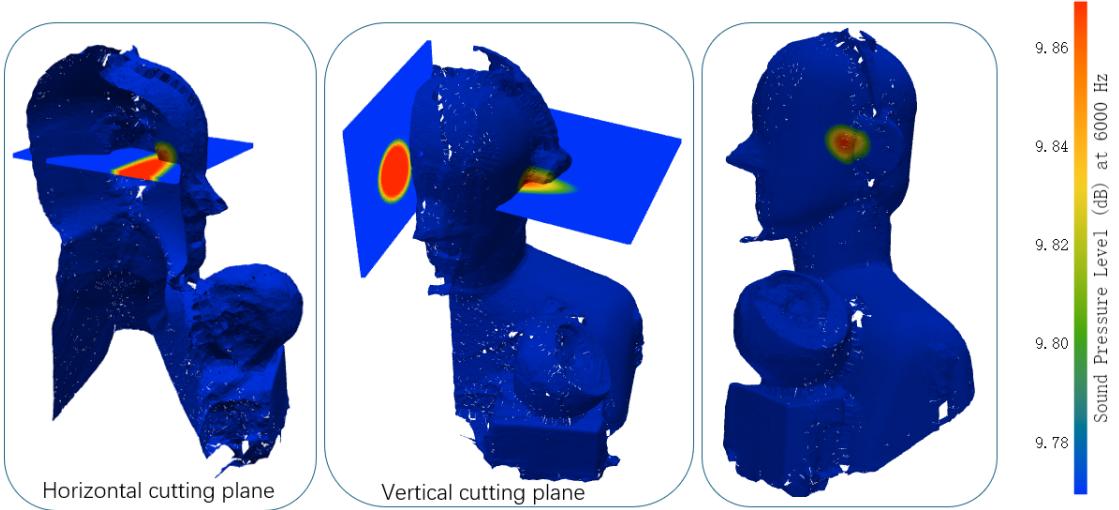


Figure 5.6: Beamforming reconstruction by functional Beamforming, to a scanned 3d grid, a Horizontal planar grid, and a vertical planar grid. The dynamic range is limited to 0.1dB to narrow the size of the main source.

aligned with each other, which for 2D beamforming calculations, a planar grid at a precise distance from the microphone array suffices to accurately cover the real source's location. However, headphone measurements, especially when performed on the HATS, the combined subject present a more complex 3D structure. A simple planar grid cannot encompass all potential source locations. As illustrated in fig. 5.6, sound leakage often occurs at the upper front of the headphone. By employing 3D beamforming, we can accurately calculate potential source locations within this geometry, ensuring the steering vector is precisely focused, resulting in an accurately located source image.

For comparison, introducing a horizontal cutting plane and conducting beamforming analysis on this plane reveals a tapered curve. This curve widens as it moves away from the source location, reflecting strength adjustments as described in eq. (2.11). This phenomenon is akin to the principle of perspective: a small object appears larger as it approaches the lens, compared to a larger object that is farther away. The vertical cutting plane yields similar errors to the horizontal plane. These inaccuracies, partly due to limiting the dynamic range to match that of our 3D beamforming estimates, highlight the challenges of accurately estimating source size and location in complex geometries.

5.4.1 Leakage localization under different frequency

Adopting a methodology akin to that used in Figure fig. 5.4, we conducted a "travel" plot analysis to delineate the highest sound source locations on the HP2 headphones' surface under three distinct configurations. To elucidate sound leakage with greater clarity, this analysis tracked the locations of the three highest value points, as illustrated in Figure fig. 5.7. The configurations examined were: the headphones worn in their standard manner, with a small porous foam inserted at the front between the ear and the ear cup to serve as a sound absorber, and with the foam similarly positioned at the back of the headphone. This approach enabled us to chart how the highest sound source locations on the surface varied in response to frequency changes.

The analysis employed a color-coded scheme to represent frequency ranges: frequencies below 2kHz were denoted in blue, from 2kHz to 6kHz in green, from 6kHz to 10kHz in yellow, and frequencies above 10kHz in red. This color-coding facilitated an observation

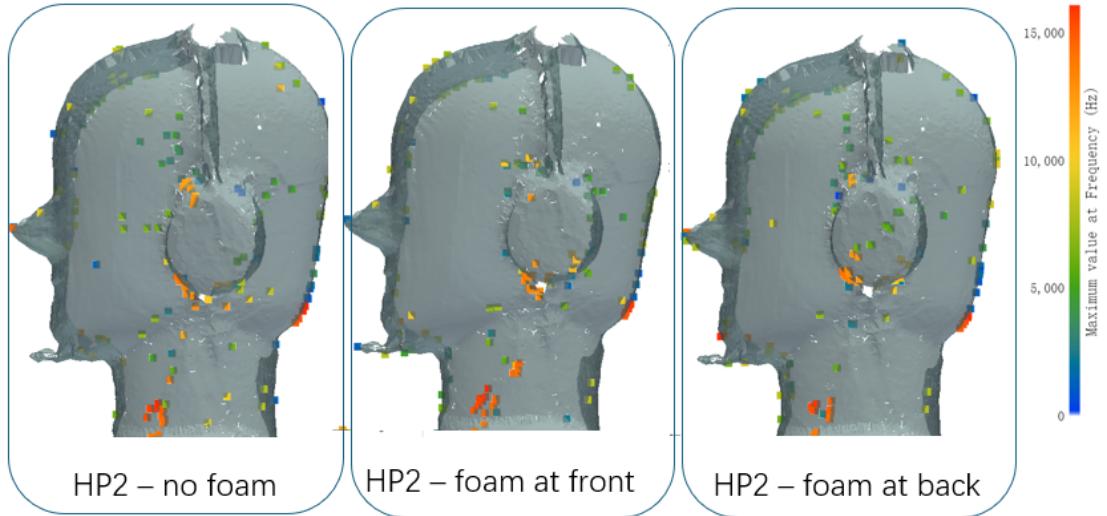


Figure 5.7: The three highest beamforming values obtained by CleanSC as a location response to frequencies. Due to the complexity of plotting, the order of point levels is not demonstrated in this plot, but we list them equally as a square. From this plot, we can understand how the sound leakage “travels” under different frequencies.

of the sound leakage dynamics: without any foam, sound leakage at frequencies under 2kHz predominantly occurred at the upper back of the ear cup. At around 3kHz, leakage was primarily observed at the front of the headphone. Notably, at higher frequencies, leakage was detected at both the upper front and upper back of the headphone, with the most significant leakage at the highest frequencies occurring at the upper front and bottom of the headphone. It is crucial to acknowledge that measurements at lower frequencies were more susceptible to environmental noise interference, which compromised localization accuracy. Moreover, certain points were identified on the HATS body, suggesting that at specific frequencies, environmental noise either overshadowed the sound leakage or contributed as a secondary or tertiary source, rendering the sound leakage at these frequencies indistinct.

The SPL level response is demonstrated in fig. 5.8, from this plot it revealed that two frequency ranges—2500Hz to 6000Hz (green in ‘travel’ plot) and 12500Hz to 13000Hz (orange in ‘travel’ plot)—where the primary sources of leakage were most distinguishable. Only within these two frequency ranges do the values exceed 0 dB, which we’ve previously established as the noise floor in our measurements. In other frequency ranges, the distinction between the primary source and the secondary and tertiary sources is less pronounced, suggesting that actual sources of sound may be obscured by noise and, therefore, not significantly detectable. This observation implies that outside these key frequency ranges, the acoustic environment may be predominantly influenced by noise rather than by direct sound leakage from the headphones. Notably, the addition of foam appears to have a minimal impact on mitigating leakage within these critical ranges.

When foam is placed at the front, the impact on high-frequency sound leakage appears straightforward: high-frequency leakage is mitigated, likely absorbed by the porous foam. This absorption effect is consistent across the bottom of the ear cup, but improvements are also noted at the upper front. For low-frequency leakage, the impact of foam placement seems negligible, as evidenced by the continued observation of blue points at both the front bottom and the upper back of the ear cup. This analysis suggests that the addi-

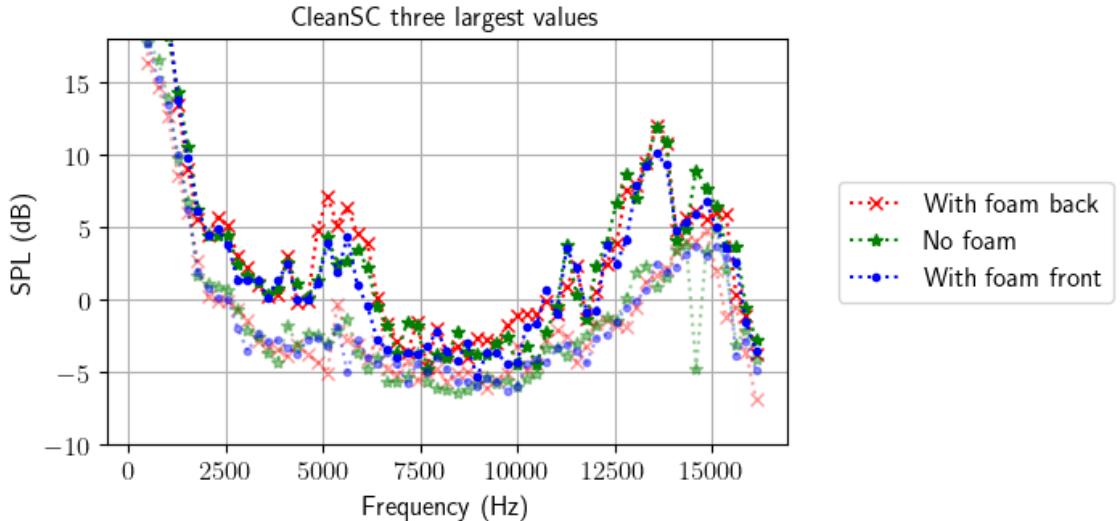


Figure 5.8: As illustrated in fig. 5.7, we present the SPL of the top three highest value points, as determined by the CleanSC beamforming, across various foam configurations and their frequency response. To facilitate easier identification within the figure, the highest value point is depicted with a solid line, while the second and third highest value points are represented with transparency in same color.

tion of foam at front can effectively reduce high-frequency sound leakage, especially for frequencies above 6000Hz.

However, around 6000Hz, notable differences in sound leakage location were observed depending on whether foam was attached. Without foam, primary leakage was detected at the upper back of the ear cup. However, with foam placement, leakage shifted to occur directly on the ear cup itself, showing an approximate 1dB increase in SPL compared to configurations without foam.

To delve deeper into the impact of foam placement on sound leakage, we generated detailed SPL maps using functional beamforming at frequencies of 5376Hz, 5632Hz, and 5888Hz. These analyses, depicted in figs. A.9 and A.10, were performed under two conditions: the standard wearing scenario and with additional foam placed at the back of the headphones. Normally, sound leakage primarily exits through the back of the headphone, likely due to a gap between the ear helix and the cushion. Without foam, this leakage might disperse upon reflection off the Head and Torso Simulator (HATS) surface before reaching the microphone array, leading to a diminished sound pressure and making the leakage seem less significant.

Securing foam with tape to seal this gap leads to an unexpected outcome: higher sound pressure levels. This counterintuitive increase in SPL might be attributed to structural vibrations induced by the tape's application. While the tape effectively seals the gap, it may also constrain the headphone structure, leading to enhanced structural vibrations. Consequently, these vibrations amplify the SPL, highlighting the complexity of mitigating sound leakage through the addition of materials alone.

This observation leads to an intriguing conclusion: the dynamics of sound leakage from headphones are intricate and cannot be straightforwardly mitigated by merely adding more sound-absorbing materials. Furthermore, if the leaked sound is directional, allowing it to escape through a gap and be absorbed by the wearer's hair and skin might actually

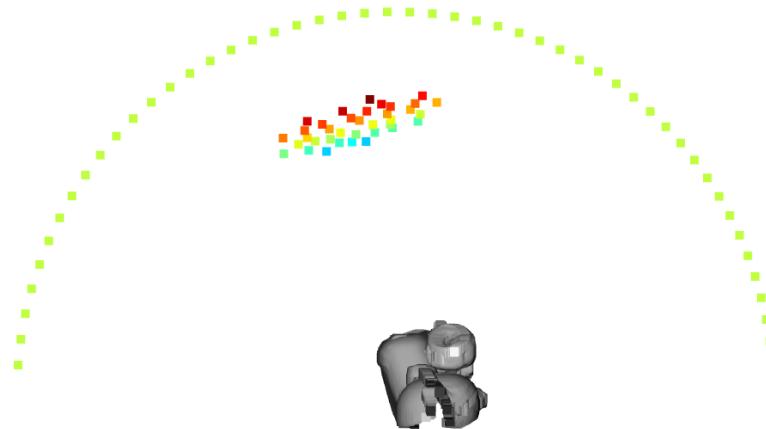


Figure 5.9: A total of 50 points are evenly spaced in a horizon plane perpendicular to the earcups of the headset under test, with the angle to the normal vector of the earcups arranged clockwise from -90 degrees to 90 degrees. The point set in between the circular gird and mesh model is the virtual microphone array.

reduce leakage more effectively than sealing the gap. In contrast, structural vibrations induced by sealing modifications could pose a greater risk for unintentional sound leakage, as they directly transmit sound into the environment.

5.4.2 Simulation of the sound radiation pattern

With the sound pressure level (SPL) data obtained at the surface of the headphones, and the detailed scanned mesh available, it becomes feasible to employ the Boundary Element Method (BEM) for further analysis. BEM allows for the calculation of how sound radiates from the headphone surface into the surrounding field through the discrete element method. For an in-depth explanation of this approach, please see Section chapter 6. In our leakage sound test, the BEM simulation utilizes the SPL data at the headphone surface to generate a radiation pattern.

As illustrated in Figure fig. 5.9, the experimental setup includes a semi-circular grid that serves as the observation point field. Additionally, to validate the simulation, it's crucial to simulate the sound pressure at individual points of the microphone array and compare these simulated results with actual measurements. The objective is to verify and align the simulation outcomes with the array recordings, subsequently simulating the SPL across the semi-circular grid to deduce the angular radiation pattern emitted by the headphones.

Through simulation, we have derived significant eigenvalues from the simulated cross-spectrum matrix of the array, which we then compared with the actual cross-spectrum matrix obtained from the initial array recordings, as depicted in Figure fig. 5.10. Given that various beamforming techniques yield different sound pressure patterns, our approach involved simulating each technique to identify which most accurately reflects the actual results.

The analysis reveals that the Beamformer Functional most closely approximates the actual eigenvalues' frequency response in terms of level. However, when considering the trend of the eigenvalues, the Beamformer CleanSC demonstrates a closer alignment. This trend closely mirrors the frequency response of the recordings, primarily because CleanSC typically reconstructs discrete points on the surface of the mesh. Consequently, simulations based on CleanSC are significantly influenced by the strength of individual points, leading to two main outcomes: 1)The reconstructed sound pressure is notably

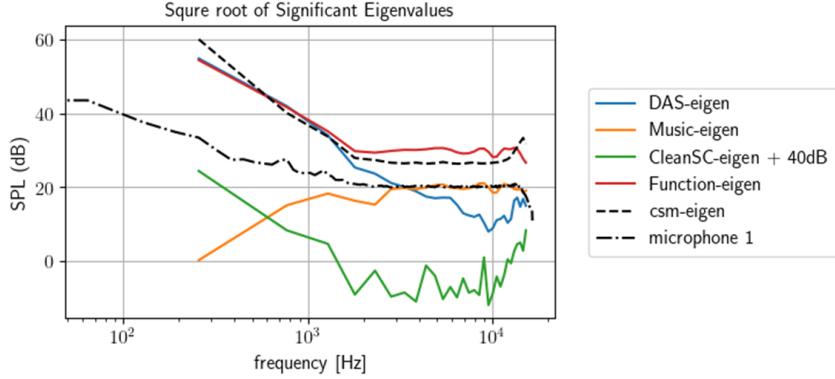


Figure 5.10: Significant eigenvalues from the square root of cross-spectrum matrix of the array by actual recording and simulation of HP2 measurement. And the actual recorded frequency response at microphone 1 position.

smaller than the actual results, as the actual area of noise generation is much larger in reality. 2)The reconstruction tends to follow the general frequency response of the recordings, given that individual points interact minimally with the mesh, thus closely tracking the recording trends.

Regarding Functional Beamforming and Delay-and-Sum (DAS) beamforming, both techniques exhibit similar characteristics at low frequencies. However, the DAS beamformer's response decays more rapidly at higher frequencies and lacks precision in this range. In contrast, the Functional Beamformer aligns more accurately with the actual eigenvalues' frequency response, showcasing its effectiveness in capturing the nuanced acoustic behavior at higher frequencies.

Through simulation, we have analyzed the eigenvalues derived from the simulated cross-spectrum matrix of the microphone array and compared these with the actual cross-spectrum matrix obtained from initial array recordings, as depicted in fig. 5.10. This comparison is pivotal as different beamforming techniques yield varied sound pressure patterns, necessitating a comprehensive evaluation to identify which method aligns most closely with actual results.

To further dissect these findings, we plotted the simulation and actual pressure frequency response for microphone 1 in the array, as shown in fig. 5.10. Here, the Music Beamformer's level matches the actual recordings remarkably well, particularly at higher frequencies, with negligible discrepancy. The analysis reveals that the Music Beamformer exhibits a frequency response at high frequencies almost identical to that of the Functional Beamformer, suggesting a high level of accuracy in these regions. However, at lower frequencies, the Music Beamformer does not adhere as closely to the observed trends.

Interestingly, the trend observed in the individual recording at microphone 1's position indicates a decay at very high frequencies, aligning with the Functional Beamforming result. Given that the stimulus signal is equalized white noise with a bandwidth that decays at high frequencies, as referenced in fig. 4.2, an accurate simulation should reflect this decay in the eigenvalues if they genuinely represent the signal's direction rather than noise. The observed increase in eigenvalue frequency response at high frequencies, contrary to the stimulus signal's decay, suggests an influence from noise not associated with headphone sound leakage. The CleanSC and DAS Beamformer's increasing trend

at high frequencies indicates it included those sources outside the simulation area. In contrast, both the Functional and Music Beamformers exhibit a decaying trend, implying they do not incorporate external sound sources into the simulation, which aligns with the expected behavior of an accurate simulation.

Based on these observations, we conclude that both the Music and Functional Beamformers provide the most accurate fit for the simulation task at hand. Their ability to closely match the actual eigenvalues' frequency response, especially at high frequencies, and to correctly reflect the decaying trend of the stimulus signal, underscores their effectiveness in simulating the sound radiation pattern due to leakage from headphones.

5.4.3 Radiation directivity pattern

Additionally, simulating the original array's location aids in calibrating the simulation results, allowing for the application of a compensatory filter across different frequencies to align the beamforming results with actual recordings. This approach, akin to the equalization method described in section 4.9, is not detailed here. The resulting sound radiation patterns, with foam placed at the front and without foam, are illustrated in fig. 5.11 and fig. A.11, respectively. It's important to note that the equalization process utilizes the eigenvalue frequency response as a reference point, applying a filter to all simulation results to ensure their trend aligns with that of the actual recordings. However, it is critical to understand that while equalization adjusts the strength of the simulation output, it does not alter the angular response. This means that the simulation receives a specific level adjustment at each frequency, based on the difference between its original response and the recording's eigenvalues on the cross-spectrum matrix of the array. This adjustment ensures consistency in the trend of the simulation output across all directions, albeit without affecting the directional characteristics of the sound radiation pattern.

In our investigation of sound radiation patterns, we simulated four distinct beamforming algorithms for comparative analysis. All results were equalized with reference to the eigenvalue frequency response, ensuring a uniform strength across the simulations. For the final analysis, we selected six frequencies for detailed examination: 1kHz, 2kHz, 4kHz, 6kHz, 8kHz, and 10kHz. These simulations revealed varying radiation predictions among the different beamformers. Notably, after correction, the results from the Functional and Music Beamformers nearly overlapped, suggesting these methods provide the most reliable outcomes. The DAS Beamformer displayed a close resemblance to the Functional Beamformer at 1kHz, but discrepancies emerged starting from 2000Hz. The CleanSC results, however, appeared to be inconsistent with expected patterns.

Focusing on the Music and Functional Beamformers, we observed an increasing order of spherical harmonics with rising frequency. At lower frequencies, radiation was predominantly directed towards the front, exhibiting a broad main radiation pattern to the left side. By 2000Hz, the radiation pattern became flatter, with the highest sound pressure of 20dB observed at the front. At 4000Hz, directivity began to manifest in various directions, showing a highest 17dB SPL radiation to the left front. Higher frequencies demonstrated more distinct directivity patterns in different directions. Additionally, irregular radiation gaps were noted, particularly at 6000Hz and 8000Hz, where the lowest radiation dropped to -5dB. This could be attributed to the simulations not accounting for noise or structural contributions, making such patterns challenging to detect in real measurements. Comparatively, conventional radiation pattern measurements conducted using the Klippel system revealed similar trends overall [13]. However, the accuracy of these similarities remains to be verified.

The comparison between the sound radiation patterns in the normal wearing scenario and

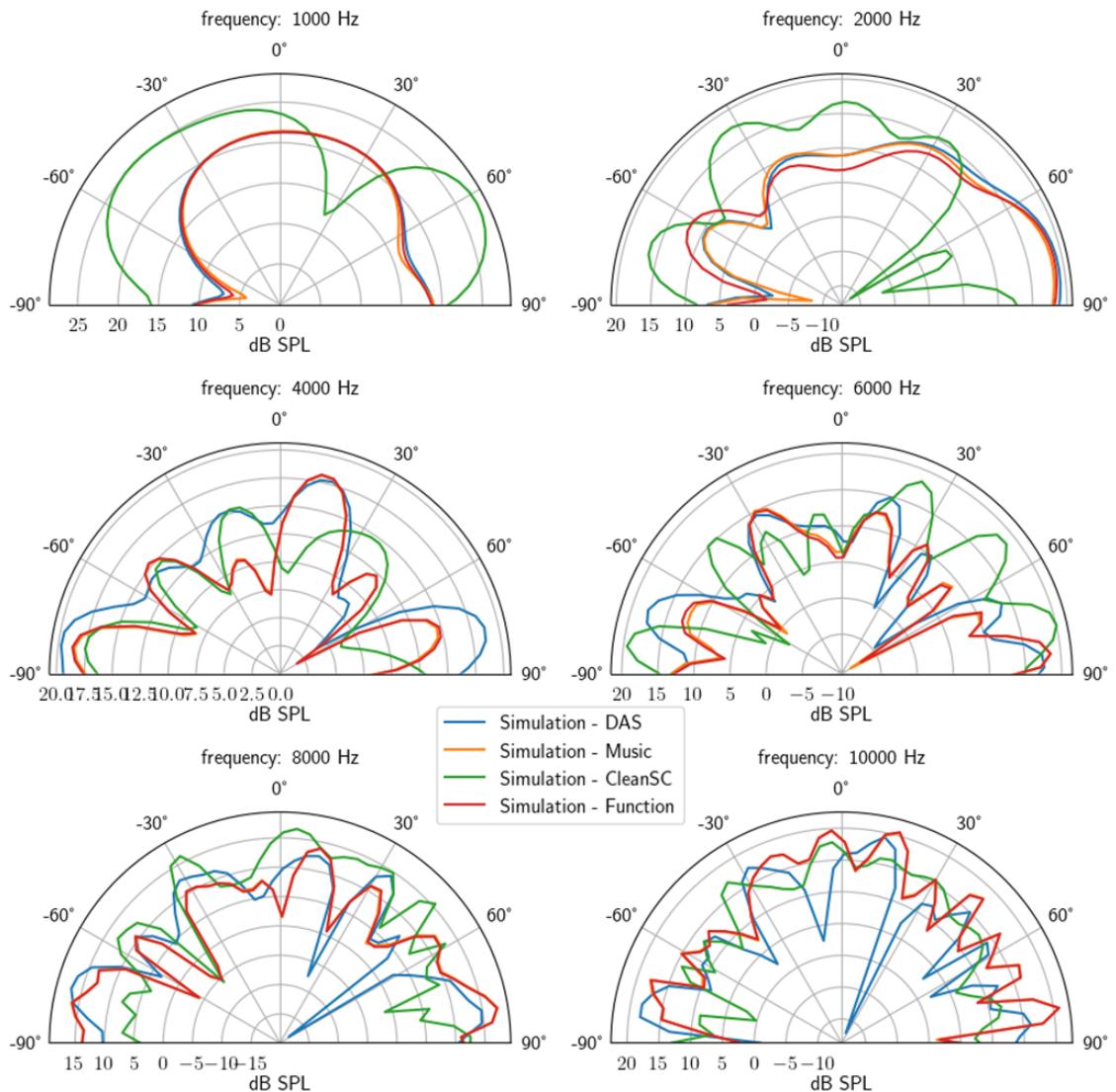


Figure 5.11: Simulated sound radiation pattern when the foam is attached at the front gap of HP2. All the directive response curves are equalized based on the response shown in fig. 5.10. Note that the yellow curve which represents the simulation result from Music almost overlaps with the simulation result from Function, especially under high frequency.

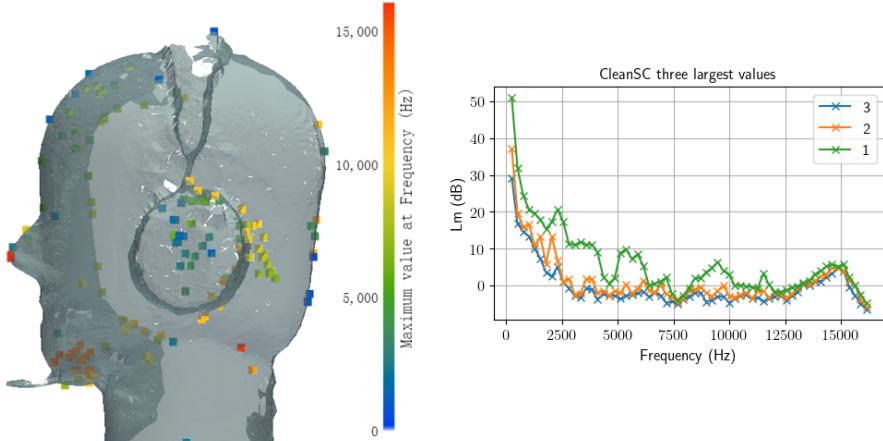


Figure 5.12: The 3 highest source strength SPL and it's location with response to the frequency.

the configuration with added foam reveals a crucial insight: both setups exhibit essentially identical radiation patterns. This observation indicates that the sound leakage, and consequently the radiation pattern it causes, is either unaffected or only marginally influenced by the introduction of additional foam. Such a finding suggests that while foam may alter certain acoustic properties within the headphones, its impact on the external radiation of sound, particularly that stemming from leakage, is minimal.

5.5 Measurement of Jabra Envolve2 75 (HP1)

5.5.1 Leakage localization under different frequency

Continuing with the analysis, we focused on the HP1. The process of scanning the geometry shape and array calibration is the same as HP2. For HP1, we only examine its performance in a standard wearing configuration without the addition of foam. The investigation, as illustrated in fig. 5.12, reveals how sound leakage varies with frequency alongside the corresponding pressure response. This analysis identifies three primary frequency ranges where sound leakage is most pronounced: from 2500Hz to 4500Hz, from 5000Hz to 6000Hz, and from 8000Hz to 10000Hz. In the travel plot, these frequency ranges are represented by the colors Blue, Green, and Yellow, respectively.

The findings indicate that low-frequency leakage predominantly results from structural transmission through the earcup. In contrast, mid-frequency leakage appears to be a combination of structural transmission and leakage at the back of the ear cup, potentially through the gap between the ear helix and the cushion. As for high-frequency leakage, it is primarily observed at the back of the ear, with indications of sound reflection on the HATS, evident at both the upper and lower back regions.

To provide a more detailed view of the sound pressure level on the surface, Figure fig. 5.13 showcases a sound pressure level map. Building on our initial analysis, we selected three frequencies to represent the low, middle, and high frequency ranges: 2560Hz, 5376Hz, and 9472Hz, respectively. Employing the Functional beamformer as usual since the performance is proven to be reliable and with good resolution, allowed us to observe the different modes of sound leakage.

5.5.2 Simulated radiation directivity pattern

The sound radiation pattern simulation for HP1 adheres to the methodology previously established for HP2. For brevity, detailed figures on eigenvalues and radiation patterns

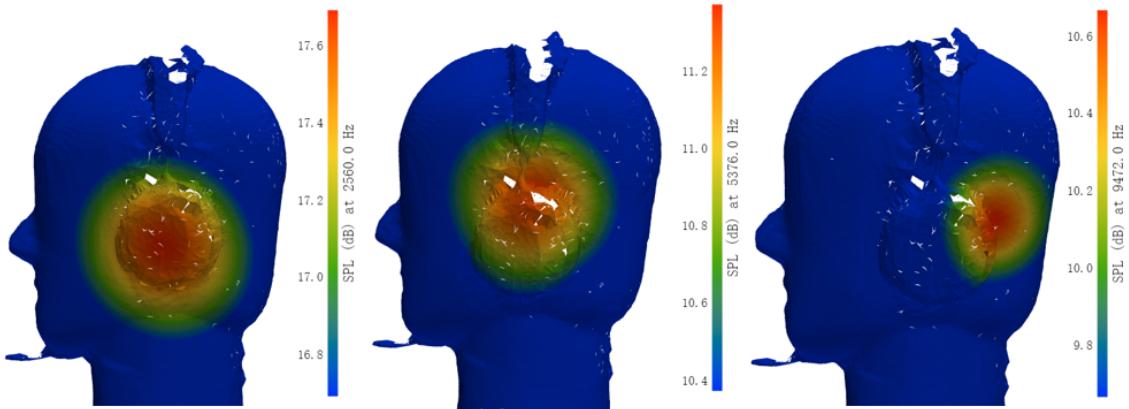


Figure 5.13: HP1 SPL map reconstructed by the Beamformer Functional. The dynamic range is limited to 1dB.

across frequencies have been relocated to appendix A.4. The simulated and measured eigenvalues, alongside the recorded array frequency response, are presented in fig. A.12, while the radiation pattern is detailed in fig. A.13. The simulation setup mirrors that of HP2, utilizing an identical circular grid and orientation relative to the Head and Torso Simulator (HATS), where -90 degrees represents the back of the headphone, +90 degrees the front, and 0 degrees corresponds to the left surface of HP1.

Analysis of the radiation pattern reveals an increase in radiation strength at lower frequencies, with the most significant radiation observed in the front and back directions. As frequencies rise, radiation strength begins to decay. Compared to HP2, HP1 exhibits a similar trend but with less pronounced side radiation, indicating that the main radiation predominantly occurs at the front and back. Specifically, dominant radiation at the back is noted around -60 degrees for frequencies above 1000Hz. At the front, significant radiation is observed at either 30 degrees or both 30 and 60 degrees. Side radiation from the left is generally less contributory, except at 1000Hz where it becomes notably significant.

Overall, the radiation pattern displays a degree of randomness, yet it is evident that -60 degrees and +30 degrees are major contributors to radiation. Radiation from the left side, potentially due to structural vibration, is primarily significant at lower frequencies.

5.6 Measurement of Vue smart glass

The Vue glasses, introduced to the public in 2016 through a crowdfunding campaign [14], represent an innovative approach to personal audio. Constructed from hard plastic, these glasses feature vibrators mounted on each glass temple, designed to transmit sound through bone conduction. Notably, the glasses are quite audible during operation, indicating significant sound radiation into the surrounding environment. For our analysis, we utilized white noise spanning from 20Hz to 14000Hz as the stimulus without equalization since this glass is a different type of sound device compared with the headphone. The volume was set to maximum to fully assess the device's radiation capabilities, as illustrated in fig. 5.14. Compared to traditional headphone leakage, the Vue glasses exhibit more pronounced radiation across the frequency range of 2500Hz to 12500Hz.

Analysis of the sound pressure levels (SPL) at various points allowed us to identify three key frequency ranges of significant radiation: 8000Hz to 12500Hz (yellow and orange), 5000Hz to 8000Hz (green), and around 2500Hz (blue). The distribution of radiation sources is also noteworthy; high-frequency radiation predominantly occurs at the glass

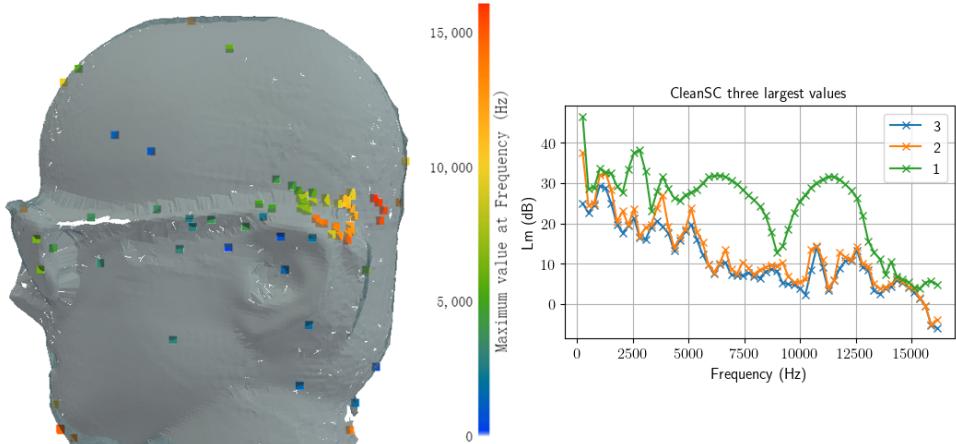


Figure 5.14: Enter Caption

temple tips, while low-frequency radiation is more central along the glass temple. Given the placement of the bone conduction vibrator beneath the helix, it's evident that many identified sound sources are located near this area. The middle frequency leakage, represented by green, appears to not only at the vibrator's location but also span through the entire body of the glass temple. The distinct primary source strength at middle frequencies suggests minimal influence from noise, confirming the observed distribution. From this source distribution pattern we can tell that the glass has both structural vibration and direct radiation from the vibrator. Also, from the point strength we observed a significant dip around frequency 8500Hz, which might be due to the anti-resonance of the structure.

5.6.1 Simulation of Radiation directivity pattern

The radiation directivity of Vue glass as shown in

6 Reverse sound field simulation by BEM

The beamforming reconstruction technique facilitates the elucidation of the acoustic field's characteristics by providing detailed insights into the surface sound pressure levels. Through the application of this methodology, it becomes feasible to derive the volume velocity across individual mesh surfaces, predicated upon the differential pressure observed at these surfaces. This derivation is fundamentally anchored in the evaluation of the pressure gradient at the mesh surface.

Further, by leveraging the pressure gradient alongside the gradient of Green's function, one can employ the Boundary Element Method (BEM) to perform forward calculations of the sound pressure level at any designated point within the sound field. This computational approach enables the synthesis of simulated data, emanating from the beamforming reconstruction, for subsequent analytical comparison with empirical data sets.

The simulation is based on one assumption: can we capture enough information by a planar microphone located in front of the sound source, and can we reconstruct the sound field also with the help of geometry information? The answer should be mixed, first of all, the planar array is located at a relatively far field, hence the near field sound pressure shall be unable to capture, also the geometry information is not exactly the truth, so our reconstruction method will be inaccurate, but such method is so convenient that may make it attractive in many scenarios, e.g., real-time scanning.

6.1 Validation

To validate the efficacy of the proposed beamforming reconstruction method in acoustic analysis, we have conducted a comparative study utilizing an ADAM A5X powered speaker as the test subject. This investigation incorporates three distinct datasets for comprehensive evaluation:

Sound Pressure Levels Recorded by Microphone Arrays: This dataset serves as the foundational input for beamforming calculations, facilitating the subsequent simulation processes. By applying beamforming techniques, we generate simulated data that allows for reverse calculation back to the original microphone positions. This reverse engineering approach enables us to establish a cost function by comparing the original microphone data with the reverse-calculated data. Ideally, the original data should equate to the reverse-calculated data plus noise, allowing us to deduce the noise or loss function's magnitude across different frequencies.

Third-party Assessment using Klippel Near-field Scanner (NFS): The NFS, a robotic measurement system capable of rotating the object and measuring from various angles (as referenced in fig. 6.1a), provides a benchmark for directivity and sound field measurements. Data obtained from the NFS is considered the standard against which our simulated outcomes are gauged, offering a precise assessment of the ADAM A5X's acoustic performance. We don't have a Klippel NFS system in our own lab, so the data is from a review to the Adam A5x in the "audiosciencereview.com" [11], The NFS system can subtract room reflections, so the measurement shall be independent to their measurement environment and able to reflect the actual sound radiation from the Adam A5x.

Simulated Data: By computing the sound field pressure at measurement positions aligned with the aforementioned datasets, we generate a third set of data. This simulation allows

for a direct comparison between our model's predictions and both the empirical microphone array recordings and the standard NFS measurements.

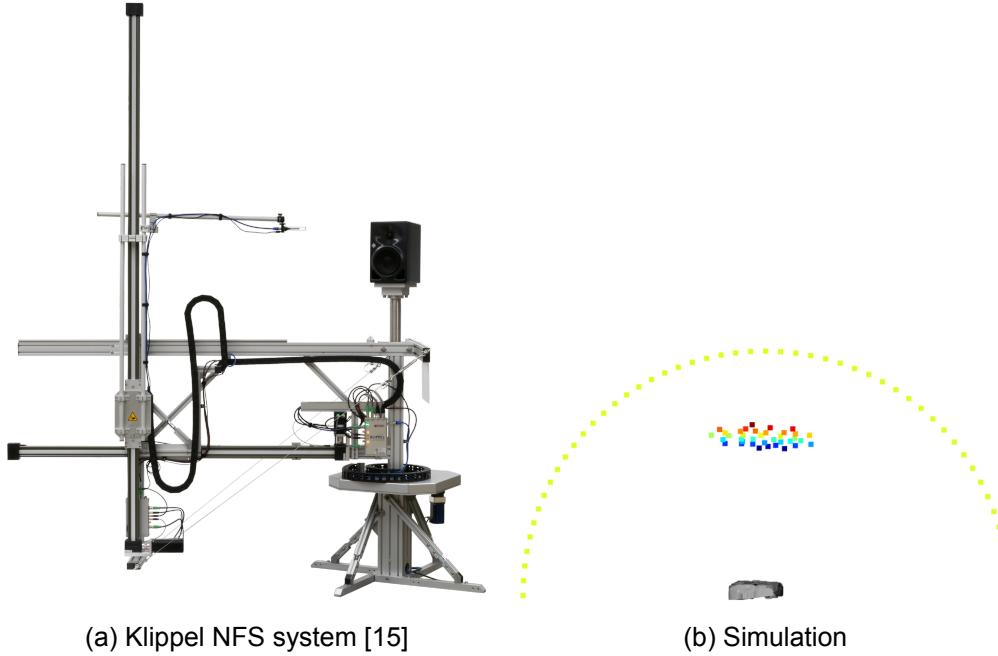


Figure 6.1: (a) The Klippel NFS measurement system. (b) The mesh model of Loudspeaker and circular grid for simulation in order to replicate the same grid set in the Klippel measurement.

The forthcoming discourse will elucidate the theoretical underpinnings of forward beam-forming, emphasizing its application in reconstructing surface sound pressure and facilitating simulation. Subsequently, a comparative analysis of the simulation results against real measurement data will be presented, underlining the efficacy and applicability of this method in acoustic research and practical implementations. And then simulate and analyse the sound field from the leakage of the DUT, to have a more comprehensive understanding of how the leakage affect the far field perception.

6.2 Theory

$$P(\vec{x}) = \int_S \left[G(\vec{x}, \vec{x}') \frac{\partial P}{\partial n}(\vec{x}') - P(\vec{x}') \frac{\partial G}{\partial n}(\vec{x}, \vec{x}') \right] dS' \quad (6.1)$$

where S is the surface of the object, \vec{x} and \vec{x}' are points in space, $\frac{\partial}{\partial n}$ denotes the derivative normal to the surface S , and $G(\vec{x}, \vec{x}')$ is the Green's function for the Helmholtz equation, representing the response at point \vec{x} due to a point source located at \vec{x}' .

6.2.1 Estimation of gradient to pressure and green's function and their normal derivative

The Green's function in 3d is

$$\nabla^2 G + k^2 G = -4\pi\delta(Q - P) \quad (6.2)$$

$\delta(Q - P)$ represents a point source, Q, P are points in space

$$G = G(P, Q) = G(R), \quad R = |P - Q| \quad (6.3)$$

Expressions of the Green's function:

$$\text{In 3-D: } G(P, Q) = \frac{e^{-jkR}}{R} \quad (e^{iwt} \text{ convention is used}) \quad (6.4)$$

To calculate the parcial differential of G to the normal vector n:

$$\frac{\partial G}{\partial n} = \left(-\frac{e^{-jkR}}{R^2} - jk \frac{e^{-jkR}}{R} \right) \left(\frac{x}{R} n_x + \frac{y}{R} n_y + \frac{z}{R} n_z \right) \quad (6.5)$$

Pressure Gradient Estimation Using Beamforming Results Beamforming results provide a spatial map of acoustic pressure P at discrete points (or meshes) in the field. To estimate the gradient of pressure ∇P at a specific mesh point with a normal vector n_0 , one can utilize the acoustic pressures and positions of neighboring mesh points. This method involves the following steps: 1. Identification of Nearest Neighbors: Utilize a KD-tree algorithm to identify the four nearest neighboring mesh points to the point of interest. The KD-tree approach efficiently searches for the nearest neighbors in a multidimensional space, making it particularly suitable for spatial data encountered in beamforming analysis. 2. Calculation of Relative Positions and Pressures: For each of these neighboring points, calculate the relative position vectors D_1, D_2, D_3 and the directional vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$, which are derived from the coordinates of the mesh points' centers. Obtain the pressure values P_1, P_2, P_3 from the beamforming results for these neighbors, along with P_0 for the point of interest. 3. Gradient Estimation: The pressure gradient at the point of interest ∇P_0 can be approximated as an average rate of change of pressure with respect to the distances to the neighboring points. Mathematically, this can be expressed as:

$$\nabla P_0 = \left(\frac{\partial P_0}{\partial x_0}, \frac{\partial P_0}{\partial y_0}, \frac{\partial P_0}{\partial z_0} \right) \approx \frac{\sum_{i=1}^3 (P_0 - P_i)}{D_i} \quad (6.6)$$

Calculation of Pressure Derivative with Respect to the Normal Vector Once the pressure gradient ∇P_0 at the mesh point is estimated, the next step is to calculate the derivative of pressure with respect to the normal vector at that point, $\frac{\partial P}{\partial n}(\vec{x})$. This is achieved by taking the dot product of ∇P_0 and the normal vector n_0 at the point of interest:

$$\frac{\partial P}{\partial n}(\vec{x}) = \nabla P_0 \cdot n_0 \quad (6.7)$$

This calculation provides the rate of change of pressure in the direction perpendicular to the surface at the point. After get the differential of Green's function \dot{G} and differential of Pressure \dot{P} at individual node point, we can then calculate the discrete form of the HIE over the scanned surface.

$$P(x) = \sum_{i=0}^{i=N} \left[G_i \cdot \dot{P}_i - \dot{G}_i \cdot P_i \right] S_i \quad (6.8)$$

Where $P(x)$ is the specific point in the sound field, G_i , P_i , and S_i is the Green's function, Pressure and area at individual mesh point.

6.2.2 Significant eigenvalues and eigenvector

Building on the discretized BEM detailed in equation eq. (6.8), our focus shifts towards validating simulation accuracy through direct comparison with actual measurements. To achieve this, we initially back-project the sound pressure to the locations of the individual

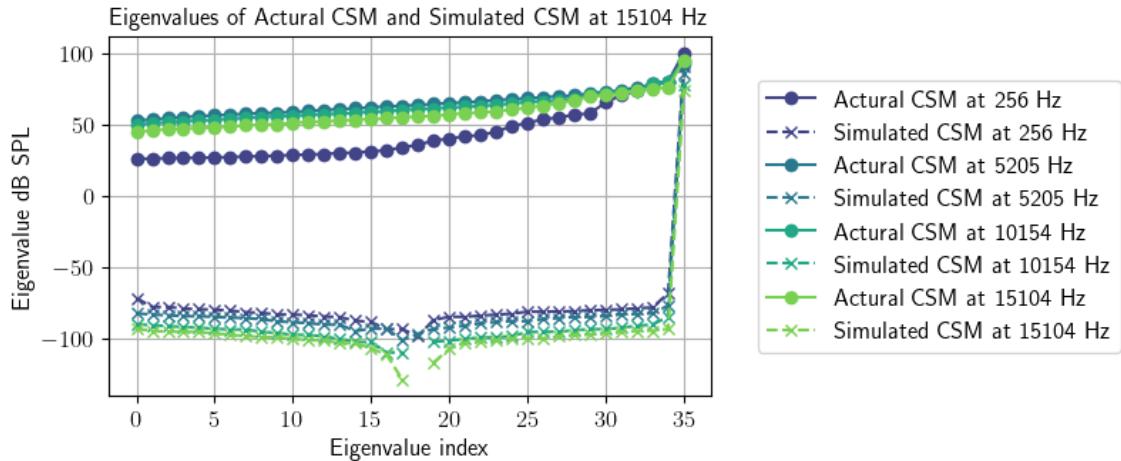


Figure 6.2: Eigenvalue comparison of the actual recording and simulate sound pressure at microphone array's position, under different frequency. This simulation is conducted from the sound pressure at the surface which is reconstructed by Beamforming Functional

microphones within our array. This step is crucial for juxtaposing simulated sound pressures with those recorded before. Ideally, these two sets of data should align closely, the recorded sound pressure should be the composition of room reflection, noise, and the sound pressure from the source which we simulated, if we assume a high fidelity in our simulations. However, given our setup encompasses an array of 36 microphones, a straightforward comparison becomes overwhelmingly complex.

To streamline this process and ensure a meaningful analysis, dimensionality reduction becomes essential. One effective method to achieve this is through eigenvalue decomposition of the cross-spectrum matrix (CSM) derived from both simulated and recorded sound pressures. The decomposition reveals eigenvalues and eigenvectors that encapsulate the primary modes of sound interaction within the environment, by compare the significant eigenvalues in two configurations, we can simplifying the comparison through a focus on these dominant features. Mathematically, this can be expressed as follows:

$$Cv_i = \lambda_i v_i$$

where C denotes the cross-spectrum matrix, v_i the eigenvectors, and λ_i the corresponding eigenvalues.

As a example, the eigenvalue CSM of the ADAM A5x with the simulation and actual measurement under different frequency as shown in fig. 6.2. It shows that from the actural measurement the CSM is contain both the sound from the actural source and the noise, but for the simulation since we do not have any noise contribution, so the eigenvalue has only one very significant component.

So, by extracting and only compare the significant eigenvalues can be helpful for focusing on the principal eigenvalues and their corresponding eigenvectors, which enables a more noise-resistant comparison, enhancing the reliability of our findings.

In practice, this methodology not only simplifies the comparative analysis by condensing it to the examination of key sound field components but also improves the robustness of our validation process against noise interference. By employing this eigenvalue-based framework, we are better equipped to differentiate between the essential characteristics

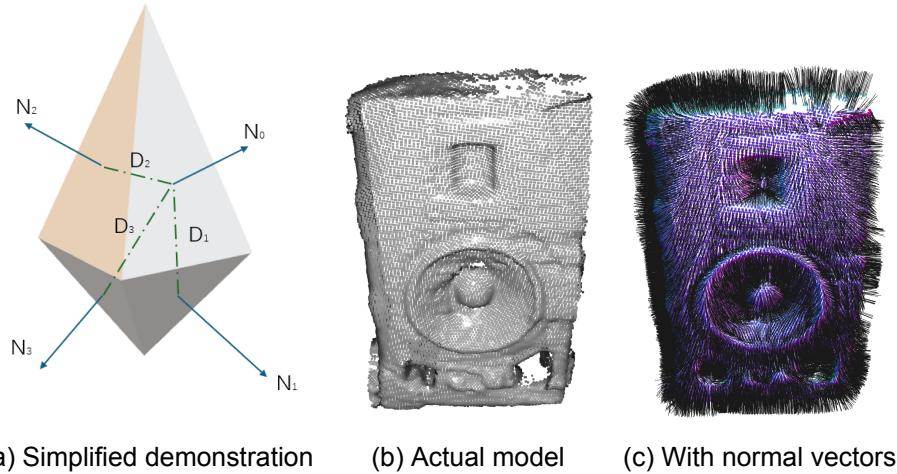


Figure 6.3: Demonstration of calculating the gradient differential to the normal vector. (a) demonstrate the simplified scenario, N_i represent the normal vector, and D_i is the distance from Mesh 0 to the i th mesh; (b) is the actual scanned loudspeaker model that we shall use for calculation; (c) is the model with normal vector. The calculation will iterate through all the mesh points on the scanned model to calculate individual's green's function and the partial derivative of the pressure with respect to the normal vector

of the simulated and actual sound fields, ensuring our simulations accurately reflect real-world conditions.

6.3 Calculation Settings

In this study, we focus on utilizing the results obtained from beamforming to estimate the sound pressure at the surface of the object in question. This initial step is critical as it directly influences the accuracy of subsequent simulations. It's important to note that different beamforming algorithms yield varying source distributions and effective areas of influence. Although Near-Field Acoustic Holography (NAH) is identified as potentially more suitable for precise volume velocity estimation, our study does not employ NAH. Instead, we aim to assess the capability of existing beamforming techniques in reconstructing surface pressure accurately.

We analyze the performance of four distinct beamforming methods as same as the chapter 5: DAS, MUSIC, Functional, and CleanSC. As previously detailed, these algorithms demonstrate varied proficiencies in source localization. This examination extends to include the effectiveness in reconstructing the sound field under the previous sound pressure results, and assert their performance when it comes to the sound field reconstruction task. The computational analysis is carried out using a corresponding GitHub project [3d_beamforming].

6.3.1 level correction for the simulation

The beamformer is for the source localization, but not for the actual sound field reconstruction, so the reconstructed surface pressure might be unable to reflect the actual source strength. so it is necessary to have the correction to the final result, that is achieved by compare the simulated eigenvalues with the recorded eigenvalues.

6.4 Results

6.4.1 Simulation results

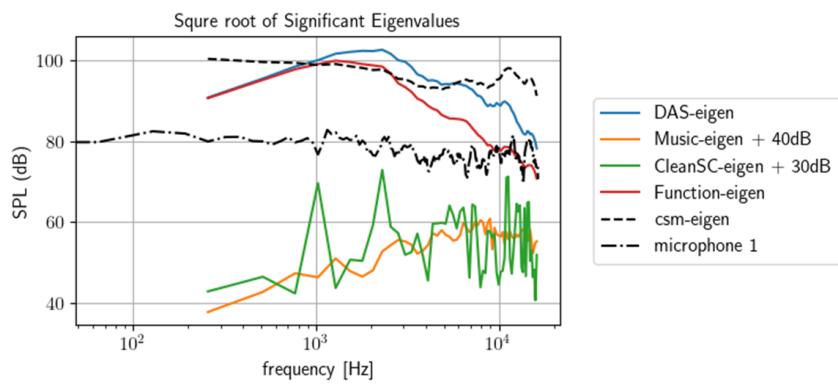


Figure 6.4: Significant eigenvalues comparison between different algorithms and the microphone array recordings. Eigenvalues are extracted from cross spectrum matrix so the eigenvalue magnitude in the unit of sound power in dB scale

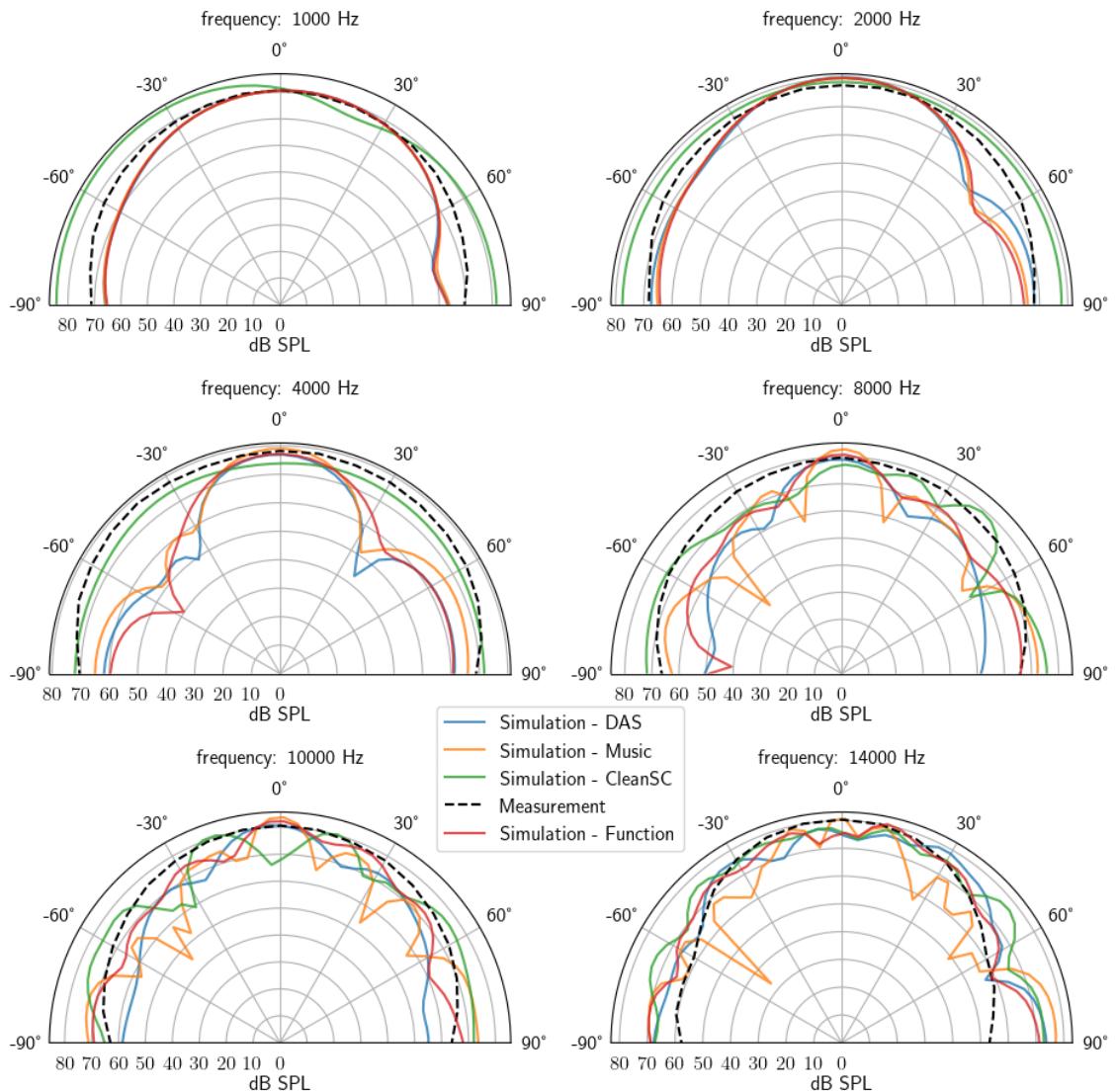


Figure 6.5: Directivity pattern from simulation from Clean-SC beamforming after correction on the Sound Pressure Level by vs Klippel measurement in different frequency range.

7 Conclusion

This chapter will include the conclusion of the measurement result, and the conclusion of the entire system

7.1 time consumption

compare the time consumption of each method. include a chart of the computing time

7.2 accuracy

get the accuracy of each method. and how frequency affect the result

Bibliography

- [1] Jafar Saniie, Mario Kupnik, and Erdal Oruklu. "Advances in Acoustic Sensing, Imaging, and Signal Processing". In: *Advances in Acoustics and Vibration* 2012 (Sept. 2012), pp. 1–2. ISSN: 1687-627X. DOI: 10.1155/2012/901547. URL: <http://dx.doi.org/10.1155/2012/901547>.
- [2] *Technical Review: No. 1 2004 Beamforming (BV0056)*. <https://www.bksv.com/media/doc/bv0056.pdf>. Accessed: 2024-03-19. 2004.
- [3] *Azure Kinect DK – Develop AI Models | Microsoft Azure*. <https://azure.microsoft.com/en-us/products/kinect-dk/#layout-container-uid3944>. Accessed: 2024-03-20. 2024.
- [4] *Azure-Kinect-Samples/opencv-kinfu-samples at master · microsoft/Azure-Kinect-Samples*. <https://github.com/microsoft/Azure-Kinect-Samples/tree/master/opencv-kinfu-samples>. Accessed: 2024-03-20. 2024.
- [5] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).
- [6] Will Schroeder, Ken Martin, and Bill Lorensen. *The Visualization Toolkit (4th ed.)* Kitware, 2006. ISBN: 978-1-930934-19-1.
- [7] John E. Stone, David Gohara, and Guochun Shi. "OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems". In: *Computing in Science Engineering* 12.3 (2010), pp. 66–73. DOI: 10.1109/MCSE.2010.69.
- [8] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. "Open3D: A Modern Library for 3D Data Processing". In: *arXiv:1801.09847* (2018).
- [9] Jussi Rämö. "Equalization Techniques for Headphone Listening". In: (). URL: <https://aaltodoc.aalto.fi/bitstreams/a5681c9a-ca63-49e5-84cd-648ecc87352a/download>.
- [10] *1ieng2en/3d_beamforming*. https://github.com/1ieng2en/3d_beamforming. 2024.
- [11] *Adam A5X Review (Powered Studio Monitor) | Audio Science Review (ASR) Forum*. <https://www.audiosciencereview.com/forum/index.php?threads/adam-a5x-review-powered-studio-monitor.22860/>. Accessed: 2024-03-13. 2024.
- [12] Brüel Kjær Sound Vibration Measurement A/S. "PULSE Array-based Noise Source Identification Solutions: Beamforming Type 8608, Acoustic Holography Type 8607 and Spherical Beamforming Type 8606". In: (). URL: <https://www.bksv.com/media/doc/bp2144.pdf>.
- [13] Robert Werner. *Measuring 3D Directivity of Microphones and Radiated Sound in Headsets and Wearables*. Online. Dresden, Germany, 2019. URL: [https://klippe.de/fileadmin/klippe/Files/Know_How/Literature/Transparancies/Workshop%203d%20Directivity%20Data%20in%20Headsets%20and%20Hearables%20-%20R.Werner%20\(KLIPPEL%20GmbH\)%20-%20online.pdf](https://klippe.de/fileadmin/klippe/Files/Know_How/Literature/Transparancies/Workshop%203d%20Directivity%20Data%20in%20Headsets%20and%20Hearables%20-%20R.Werner%20(KLIPPEL%20GmbH)%20-%20online.pdf).
- [14] *Vue: Your Everyday Smart Glasses by Vue — Kickstarter*. <https://www.kickstarter.com/projects/vue/vue-your-everyday-smart-glasses>. Accessed: 2024-03-29. 2024.
- [15] KLIPPEL GmbH. *Near Field Scanner System (NFS)*. <https://www.klippe.de/products/rd-system/modules/nfs-near-field-scanner.html>. Accessed: 2024-03-13. 2024.

A Appendix

A.1 Additional Beamforming Plots

The following plots shows the results under different beamforming algrithum with freuecy from 1000Hz to 8000Hz. The Dynamic range is limited to be 5dB for comparison.

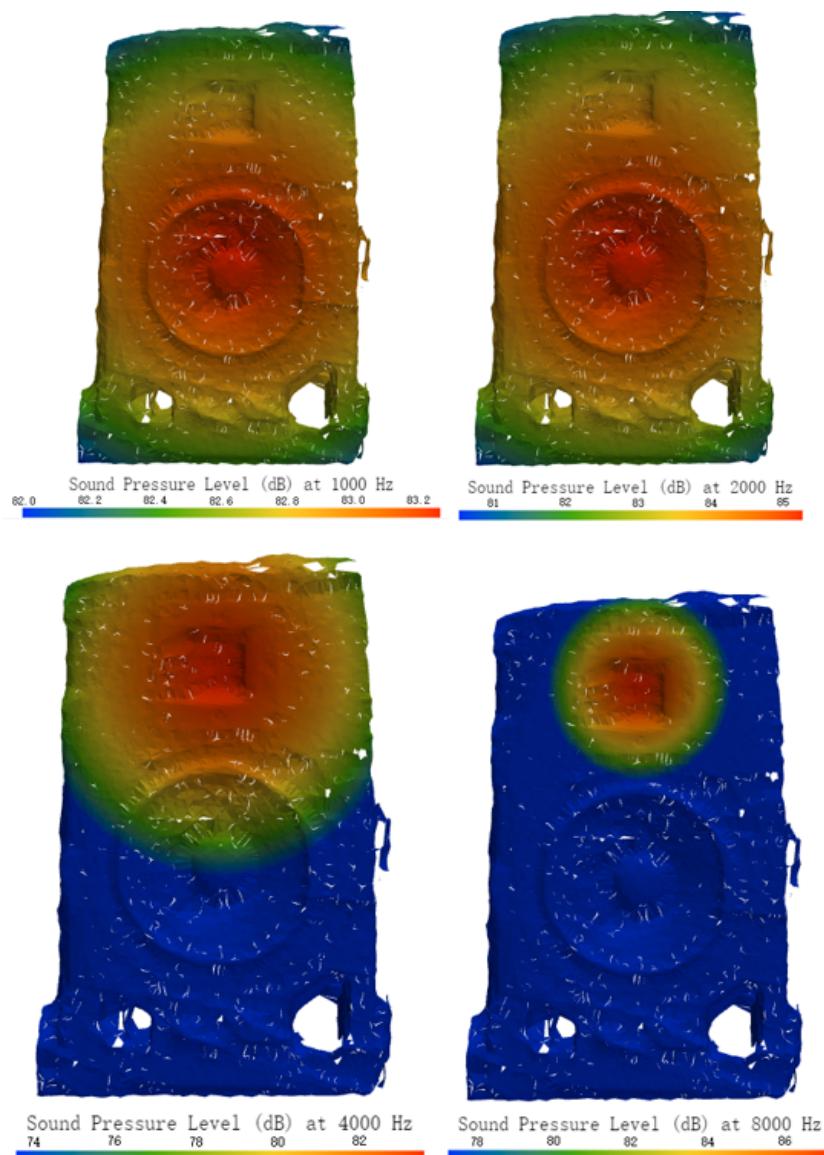


Figure A.1: Delay and Sum Beamforming reconstructed sound pressure level to the surface of the loudspeaker

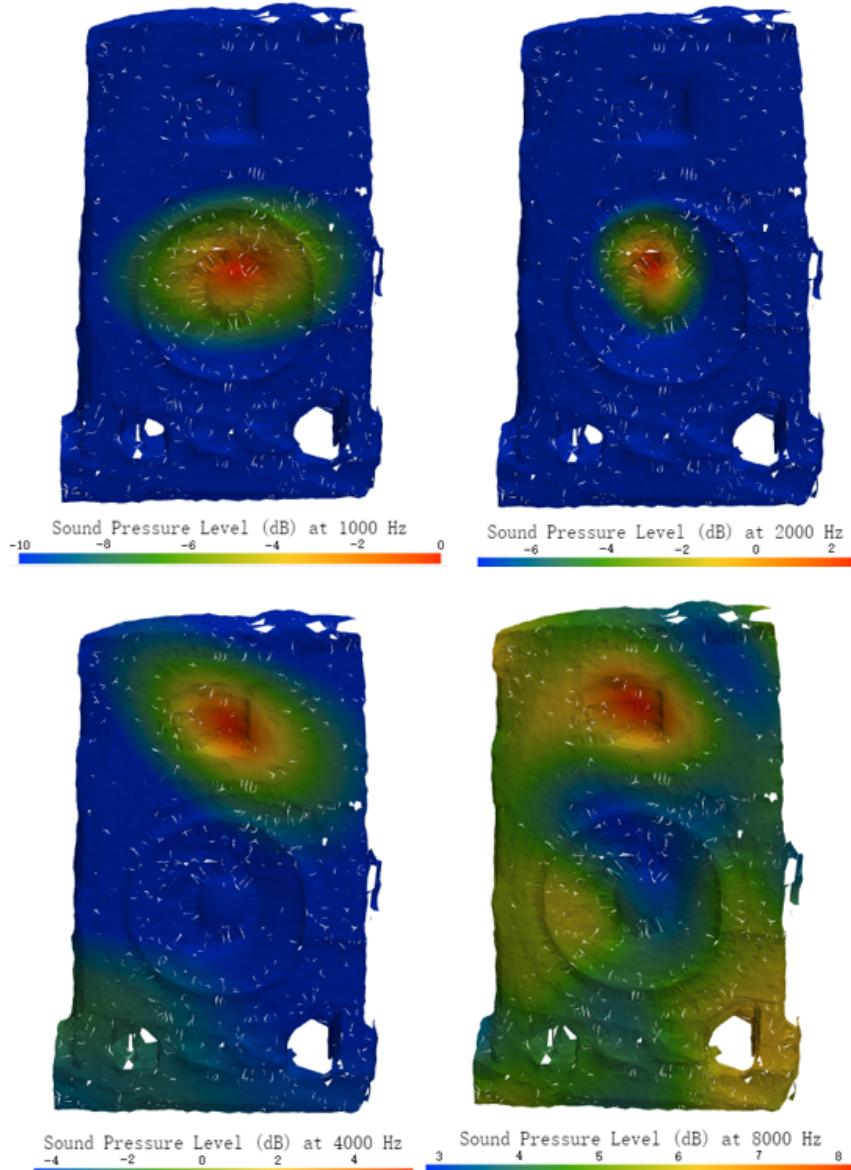


Figure A.2: Music Beamforming reconstructed sound pressure level to the surface of the loudspeaker. It's worth to note that the sound pressure level in this plot is not representing the actual level, but only the reciprocal level of noise index.

A.2 Adam A5x Beamforming test

From NSI Acoustic array postprocessing software

A.2.1 HP2 functional beamforming plots

A.3 Radiation Pattern simulation for HP2

A.4 Radiation pattern simulation and corresponding eigenvalues for HP1

A.4.1 Simulation of the sound radiation pattern

A.4.2 Radiation directivity pattern

A.5 Radiation pattern simulation and corresponding eigenvalues for Vue glass

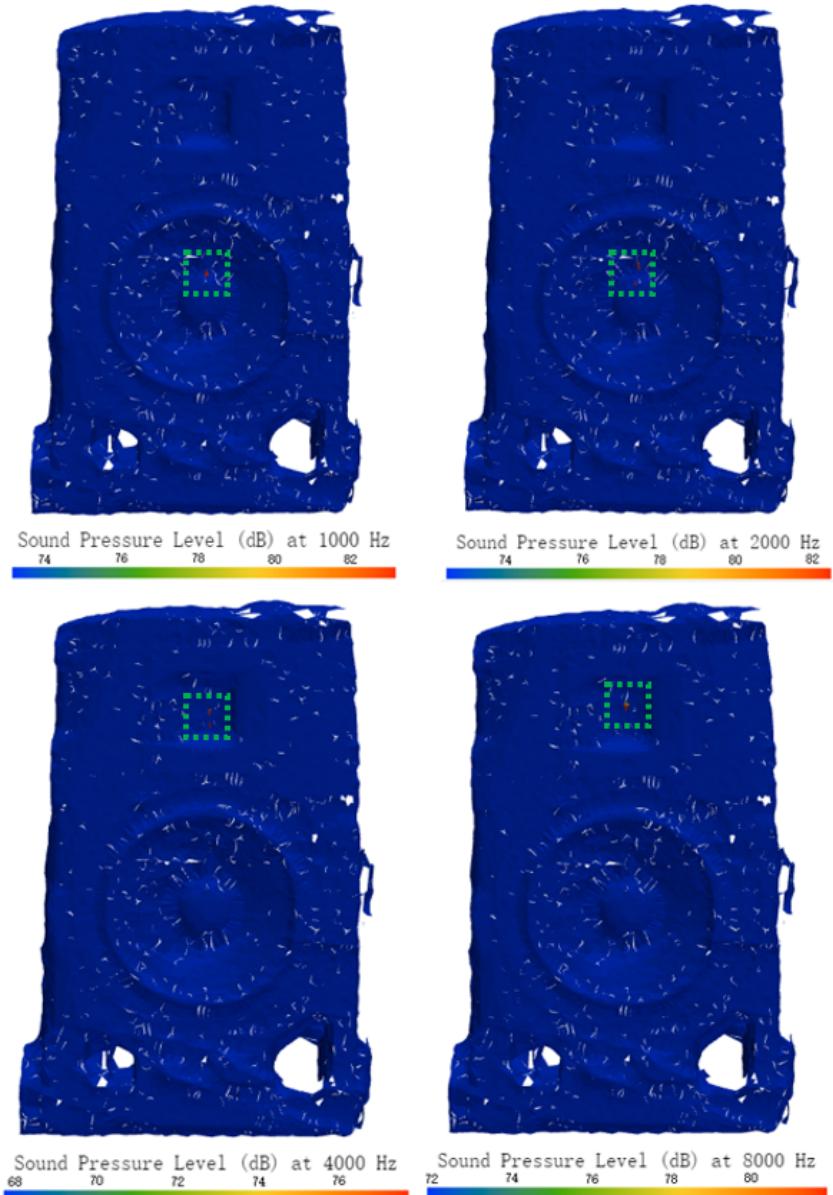


Figure A.3: CleanSC Beamforming reconstructed sound pressure level to the surface of the loudspeaker. The green dash box were added in order to help to identify the source location.

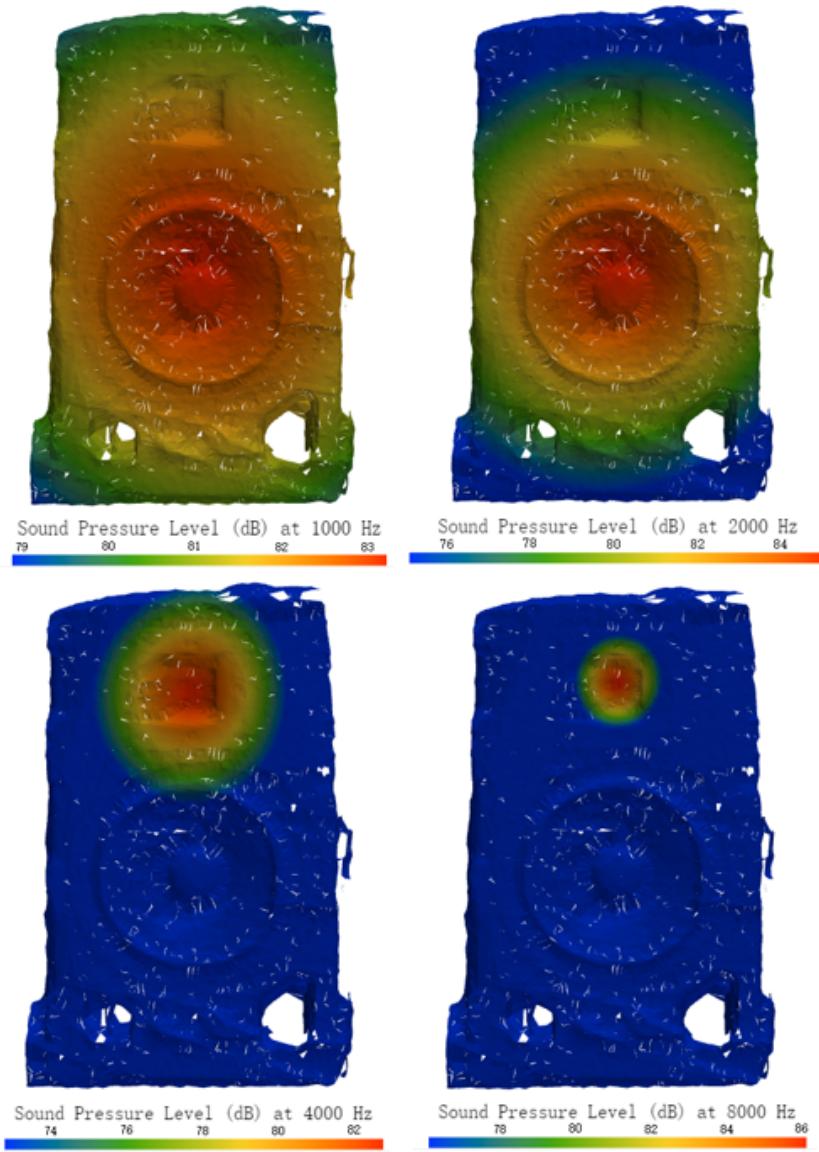


Figure A.4: Functional Beamforming reconstructed sound pressure level to the surface of the loudspeaker

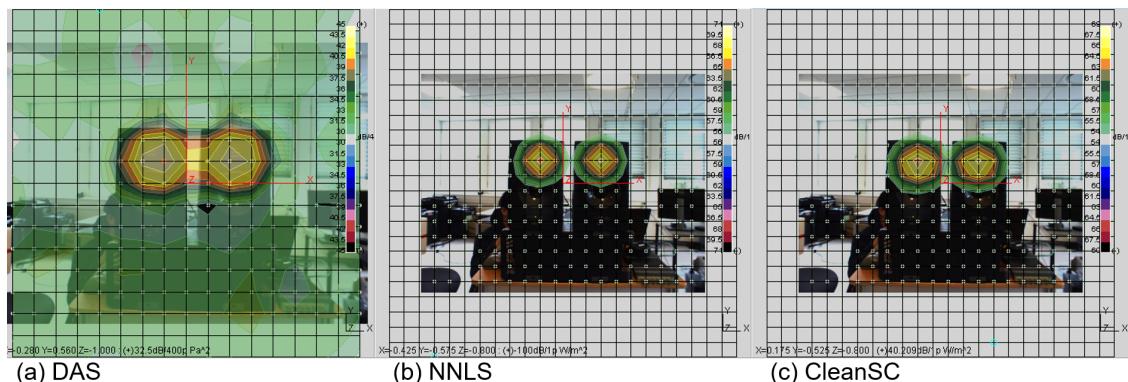


Figure A.5: NSI-Array acoustic postprocessing result under 5000Hz, 15dB dynamic range

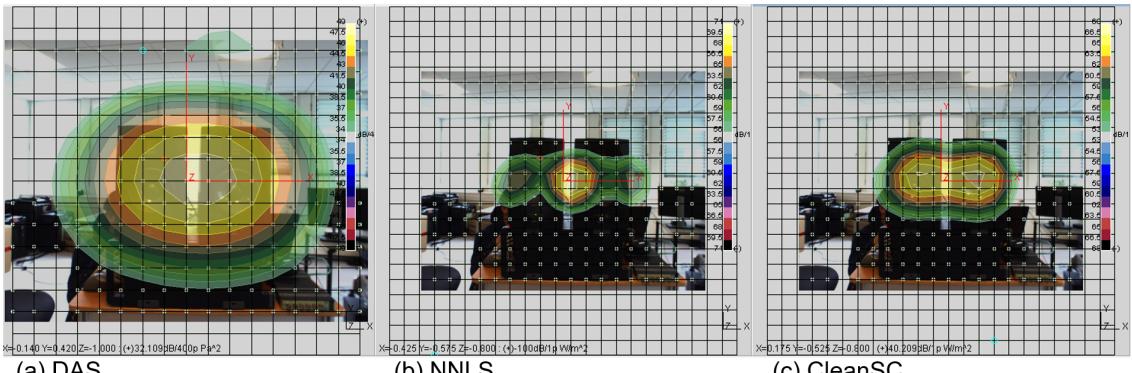


Figure A.6: NSI-Array acoustic postprocessing result under 2500Hz, 15dB dynamic range

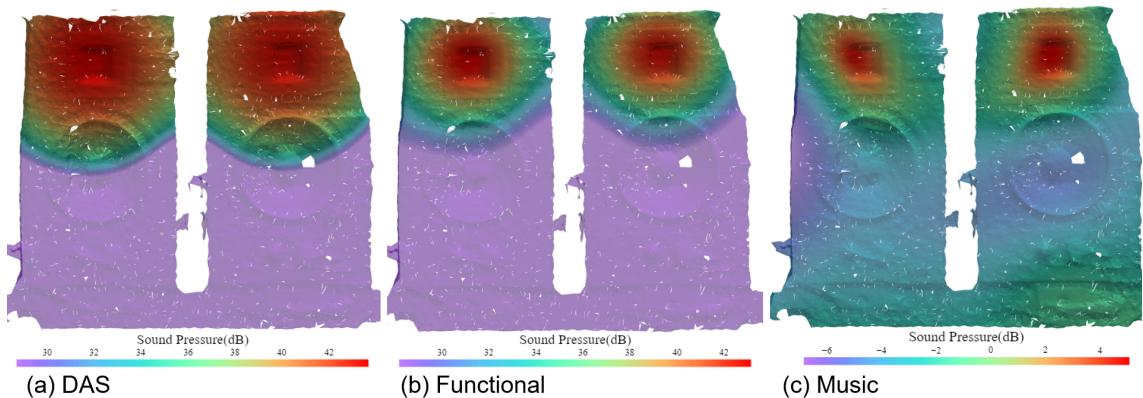


Figure A.7: 3d beamforming result under 5000Hz, 15dB dynamic range.

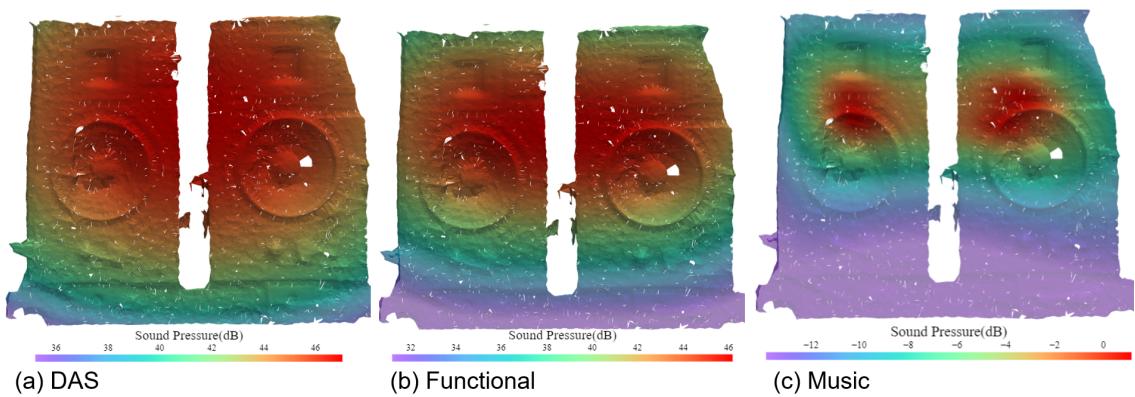


Figure A.8: 3d beamforming result under 2500Hz, 15dB dynamic range.

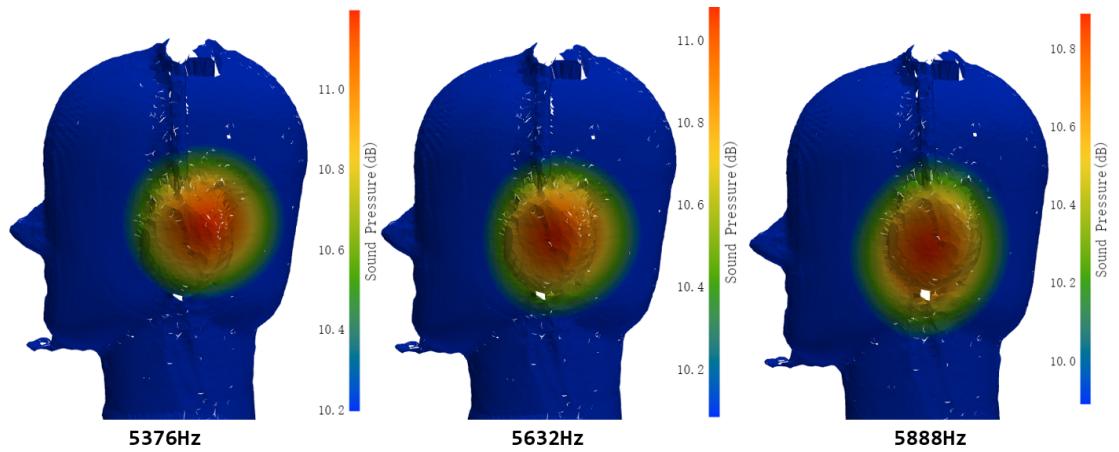


Figure A.9: Functional Beamforming result for the HP2 when foam placed at the back, The dynamic range is limited to be 1dB.

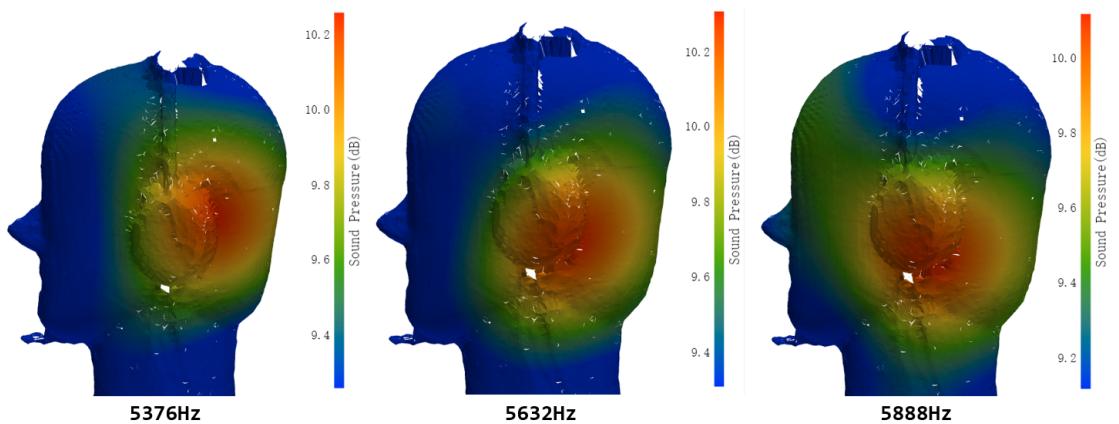


Figure A.10: Functional Beamforming results for the HP2 when the HP2 normally wore without foam, The dynamic range is limited to be 1dB.

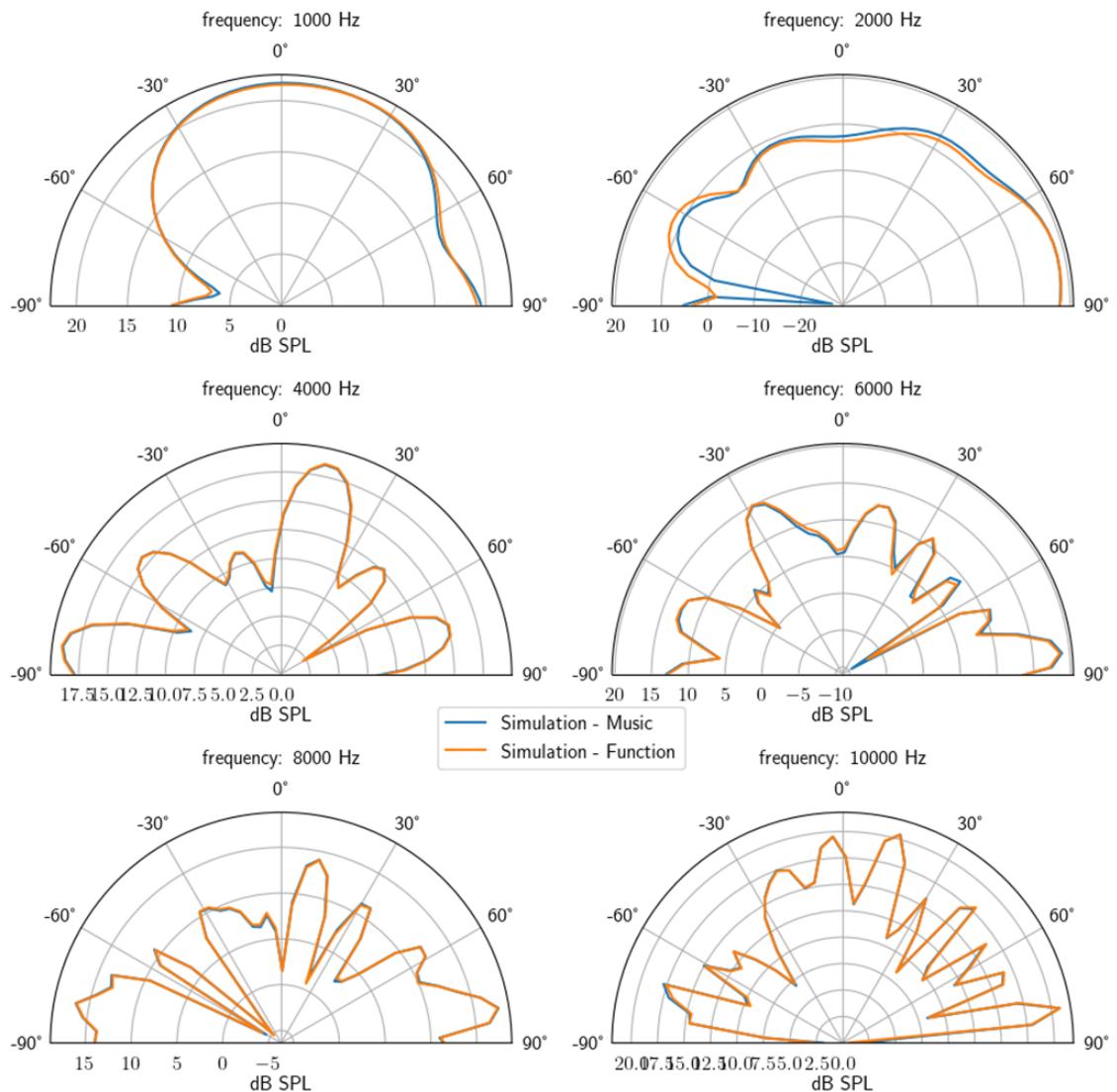


Figure A.11: Simulated sound radiation pattern when no foam is attached on HP2.

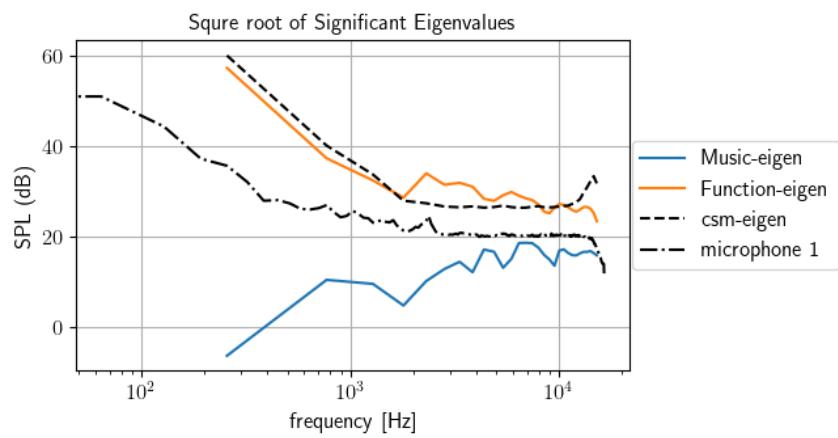


Figure A.12: Significant eigenvalues from the cross-spectrum matrix of the array by actual recording or simulation of HP1 measurement.

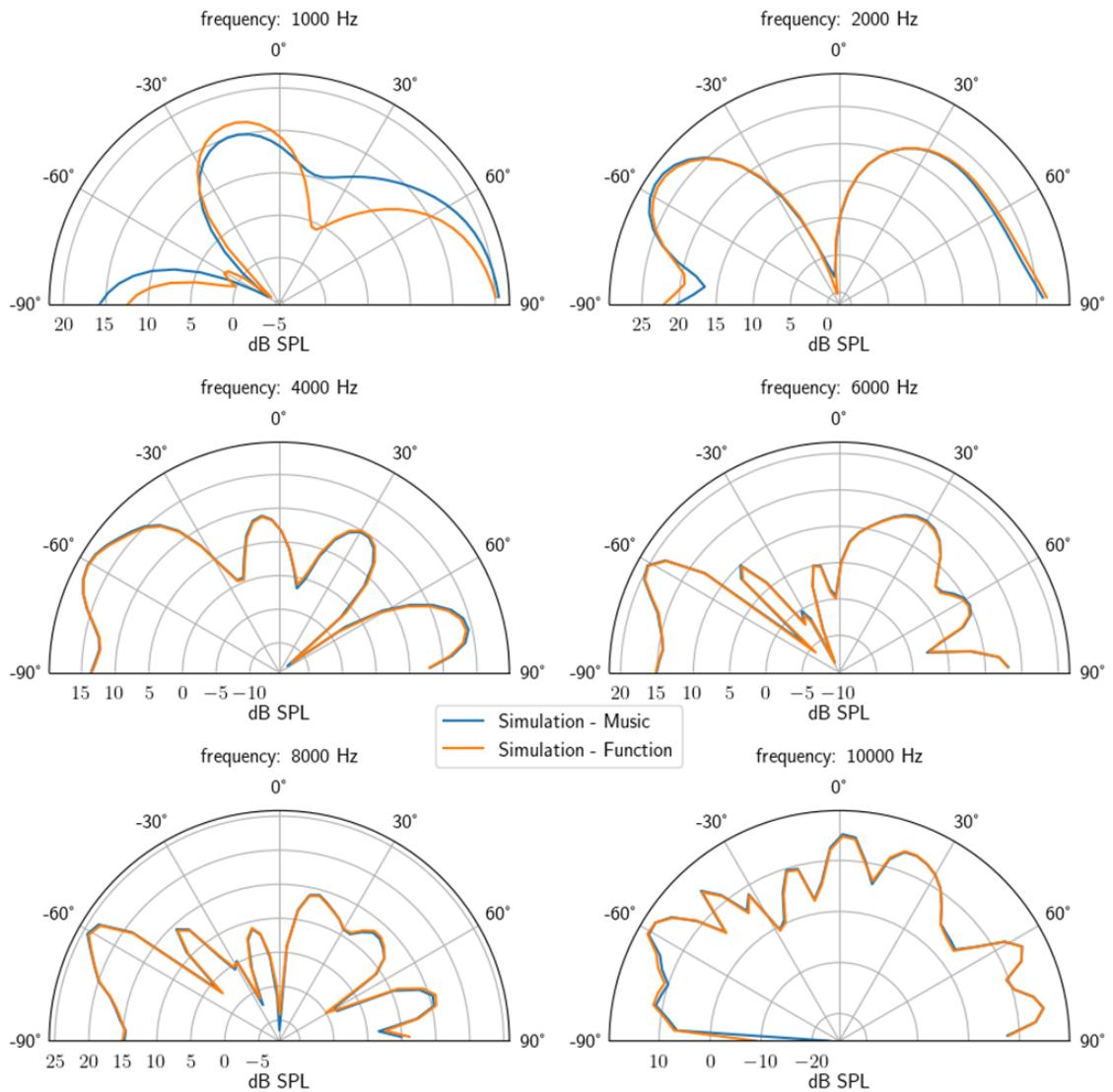


Figure A.13: Radiation directivity pattern of HP1. Both Music and Functional simulation result is equalized based on the eigenvalue at the origional array position.

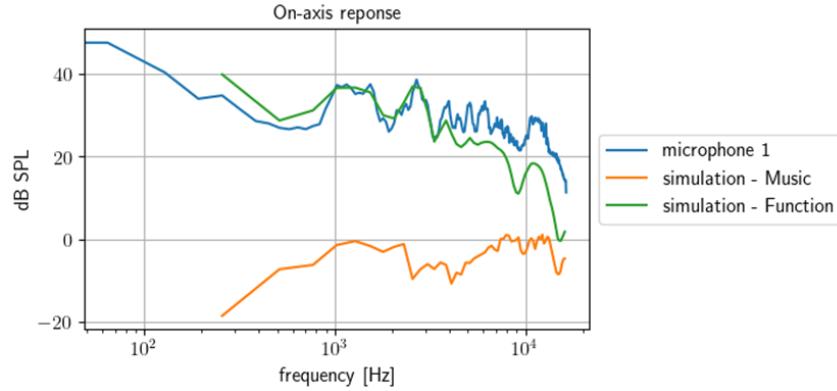


Figure A.14: Simulated and measured frequency response of Vue smart glass.

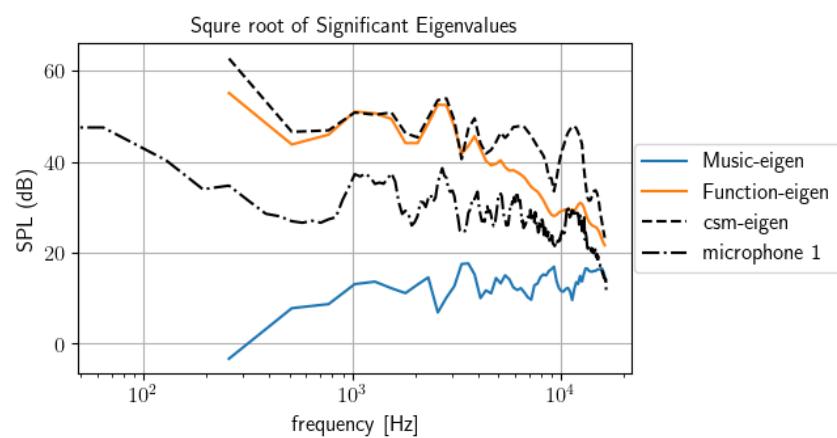


Figure A.15: Simulated and measured eigenvalue with frequency response of Vue smart glass.

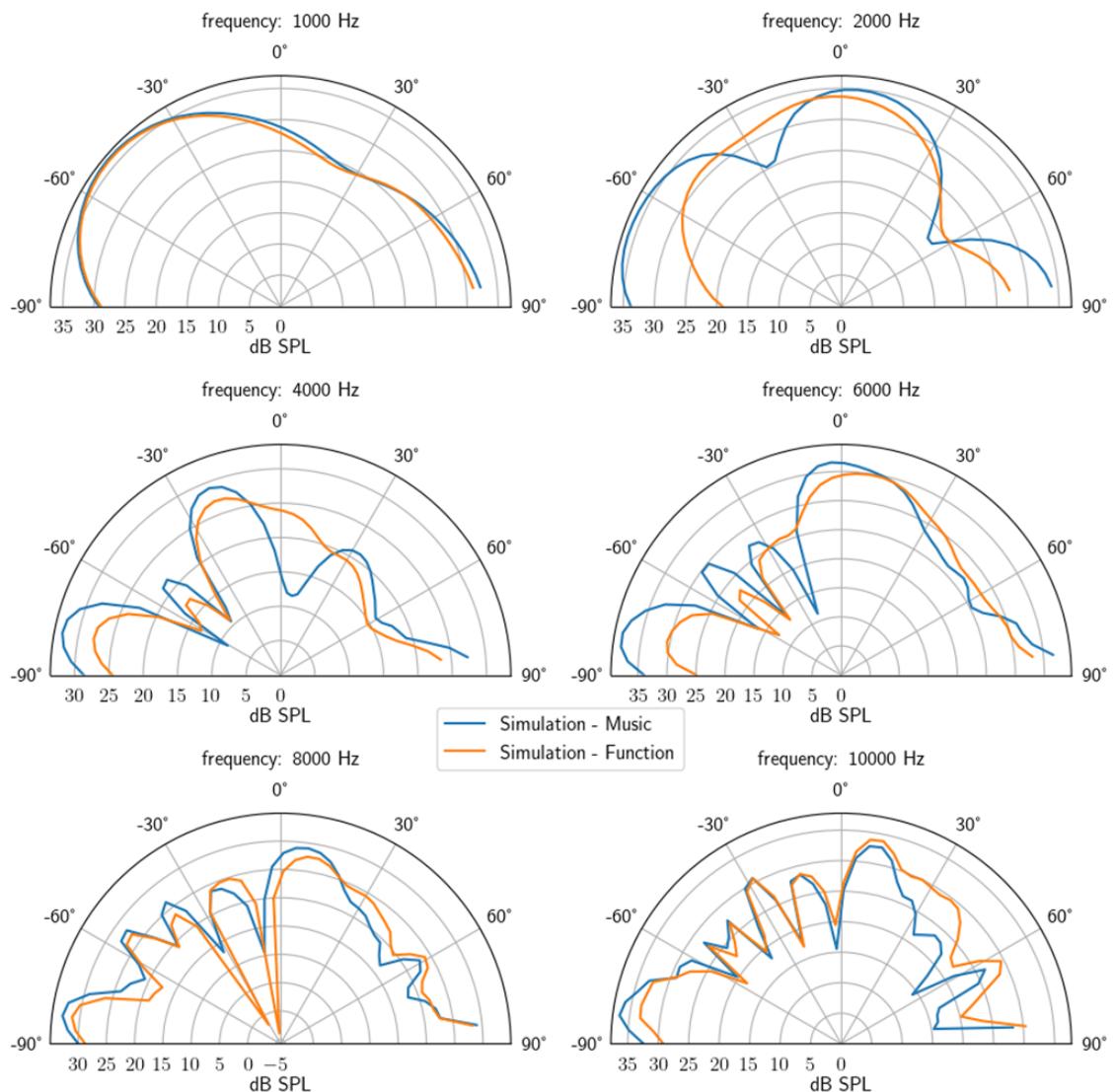


Figure A.16: Radiation directivity pattern of the Vue glass.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Technical
University of
Denmark

Ørsteds Plads, Building 352
2800 Kgs. Lyngby
Tlf. 4525 1700

electro.dtu.dk