

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

Acknowledgements

Zhengmeng Li, MSc Acoustic Engineering, DTU - Danmarks Tekniske Universitet
Student of Engineering acoustic, master thesis candidate

Vicente Cutanda Henriquez, Associate Professor, DTU - Danmarks Tekniske Universitet
Supervisor

WooKeun Song, Senior Research Engineer, HBK - Hottinger Brüel & Kjær
External Supervisor

Contents

Acknowledgements	ii
1 Introduction	1
2 Theory	5
2.1 Intro	5
3 Mesh model	10
3.1 Manual registration	10
3.2 Global registration	11
3.3 Mesh generation	12
4 Set of system	15
4.1 Device of measurement	15
4.2 Device under test	17
4.3 Lab environment	18
4.4 Stimulus	19
5 Measurement	23
5.1 Array alignment calibration	23
5.2 Array recording	23
5.3 Verification through Adam A5x	24
5.4 Measurement of Jabra evolve2 65 (HP2)	26
5.5 Measurement of Jabra Envolve2 75 (HP1)	34
5.6 Measurement of Vue smart glasses	35
6 Reverse sound field simulation by BEM	39
6.1 Theory	39
6.2 Validation	41
6.3 Calculation Settings	43
6.4 Results	43
7 Conclusion	47
Bibliography	49
A Appendix	53
A.1 Adam A5x additional Beamforming Plots	53
A.2 Comparison test for NSI and 3d beamforming on Adam A5x	54
A.3 Additional HP2 functional beamforming map	55
A.4 Radiation Pattern simulation for HP2	56
A.5 Radiation pattern simulation and corresponding eigenvalues for HP1	58
A.6 Radiation pattern simulation and corresponding eigenvalues for Vue glass .	60

1 Introduction

In the contemporary landscape of personal audio devices, the issue of sound leakage emerges as a significant challenge, detracting from user experience and privacy, which has already been addressed by some customer reviews [1] [2]. This leakage, traditionally associated with headphones, has broader implications as technology advances, encompassing a range of devices such as AR/VR glasses and bone conduction devices. These innovations, designed for intimate interaction with the user's auditory senses, inadvertently facilitate the escape of sound, potentially exposing sensitive information and disrupting the ambient auditory environment. This phenomenon underscores the need for a solution that addresses not only current but also future technologies, ensuring that personal audio devices can deliver immersive experiences without compromising privacy or comfort.

To further address the problem, Sound leakage, at its core, is a multifaceted issue that extends beyond simple mechanical gaps, such as those between headphone cushions and the user's ear. It encompasses structural vibrations within the device itself—vibrations that transform the entire device into an unintended radiating body of sound. Particularly in headphones, this can result from the resonance of components like the ear cup or the frame, which act in concert with the audio output to leak sound. Similarly, bone conduction devices, while bypassing the traditional ear path, do not entirely eliminate leakage, as vibrations can still be transmitted through the user's body or its own structure to the external environment. Moreover, certain designs incorporate intentional sound radiation to enhance the soundstage or user experience. Where as there are also some design choice, aimed at mitigating the occlusion effect [3] —a sensation of blockage in the ear canal leading to discomfort and altered perception of one's own voice — can inadvertently facilitate greater sound escape, thereby exacerbating the sound leakage issue. The evolution of personal audio devices necessitates precise leakage localization. For instance, identifying the optimal placement of microphones for Active Noise Cancellation (ANC) [4], or for feedback cancellation in voice microphones [5] [6], hinges on a thorough understanding of sound leakage localization and its intensity measurement. Such knowledge is crucial for the advancement and refinement of these technologies, ensuring they meet the dual objectives of delivering high-quality audio experiences while minimizing the challenge of design due to sound leakage.

Current solutions

The exploration of existing solutions to the sound leakage dilemma reveals a landscape marked by varied approaches, each with its inherent advantages and limitations. Among the most prevalent methods employed by headphone manufacturers is the utilization of a single microphone positioned near the headphone to gauge sound leakage, or utilizing the Ear Simulator to evaluate the overall performance and derive the amount of leakage [7]. This technique, while widely adopted, offers a rudimentary assessment, capturing only a snapshot of the sound pressure at specific points and directions, lacking the ability for comprehensive analysis and Leakage localization.

Further advancements have been made by companies like Klippel, which introduced a near-field scanner (NFS) [8] capable of meticulously scanning the sound field around a headphone. This device operates by rotating the subject wearing the headphones, with a fixed microphone that can scan the sound emitted from the subject, and thereby

mapping a directivity pattern of sound emission. While this approach is very accurate in detection capabilities, it is hampered by significant time consumption and flexibility.

Additionally, some entities have explored the use of 2D beamforming technology, such as the solutions provided by Brüel & Kjær, to visualize sound leakage. [9]. These systems employ acoustic cameras to depict the origin of sound in two dimensions. However, these solutions, too, grapple with issues of resolution and convenience, that are not as widely embraced by the headphone manufacturers due to their limitations.

The shortcomings of these existing methods—ranging from the lack of precision in single microphone techniques to the time-intensive nature of near-field scanning and the resolution constraints of 2D beamforming—underscore the necessity for a more advanced solution. It is here that the potential of 3D beamforming becomes evident. Despite the acknowledged promise of beamforming technologies, their application has been restricted by limitations in grid size, setup complexity, and resolution accuracy, especially in the measurement of headphone leakage.

This research proposes an innovative approach that integrates a 3D scanner to generate a precise mesh grid. This grid serves as the foundation for a sophisticated beamforming analysis, enabling the creation of a 3D sound pressure map that vividly illustrates the areas of highest sound leakage. Through this methodology, the study aims to surmount the prevailing challenges faced by current technologies, offering a novel perspective on sound leakage detection that combines precision, efficiency, and versatility.

The 3d beamforming

This thesis introduces a novel approach to detecting sound leakage from headphones by leveraging the capabilities of 3D beamforming, a technique previously applied in fields requiring precise noise detection, such as car internal design [10] [11] and aerodynamics [12]. While the foundational technology is not new, the application of headphone sound leakage detection presents unique challenges and opportunities for innovation. One such innovation is the utilization of the Microsoft Azure Kinect [13], an affordable and open-source 3D scanner, which enables rapid and precise surface scanning—a critical step in accurately identifying areas of sound leakage. Meanwhile, recognizing the intricacy and precision required for this task, we have conducted thorough verification to ensure measurement accuracy. At the heart of our methodology is a comparative analysis of various beamforming algorithms, including traditional Delay-and-Sum (DAS), Multi-source Classification (MUSIC) [14], CLEAN Based on Spatial Source Coherence (CleanSC) [15], and Functional Beamforming [16]. This evaluation aims to identify the most effective algorithm for this specific application, moving beyond conventional techniques to explore a wider range of advanced beamforming capabilities.

Beyond the identification of sound leakage points through beamforming maps, our research delves deeper into the mechanics of sound radiation. Acknowledging the industry's interest in understanding not just the location but also the directivity and the transmission of sound leakage, we employ Boundary Element Methods (BEM) [17] to simulate sound fields based on discrete mesh grids. This approach enables a nuanced analysis of sound radiation patterns, offering valuable insights into how sound leakage propagates from audio devices.

Furthermore, the integration of 3D scanning and beamforming challenges traditional perceptions of beamformer performance, necessitating a reevaluation in light of altered scenarios and applications. This shift underscores the computational intensity of processing potentially tens of thousands of grid points simultaneously, a factor that significantly impacts the feasibility of employing certain advanced beamformers. Notably, some de-

convolution Beamforming methods which rely on the iteratively calculation will become very hard to apply due to the increasing computational consumption under this application.

Through this multifaceted approach, the thesis not only advances our understanding of sound leakage and its detection but also illuminates the comparative efficacy of various beamforming algorithms in a 3D context. Further more, by marrying detailed beamforming maps with sound field simulations, we provide a holistic view of sound leakage, encompassing both its localization and its radiative behavior.

2 Theory

This chapter delves into the theoretical underpinnings essential for understanding the phenomena and methodologies central to this thesis.

2.1 Intro

The introduction to beamforming, imaging functions, conventional Delay-And-Sum (DAS), and Direction of Arrival (DOA) techniques lay the groundwork for our exploration. These concepts are pivotal in navigating the complexities of sound wave manipulation and spatial audio capture, serving as the foundation upon which more advanced topics, such as near-field Beamforming, are built. This section aims to scaffold our discussion, leading to a nuanced understanding of how these principles are applied in practice and their relevance to the overarching goals of this research.

2.1.1 Near field Delay and Sum Beamforming

Consider a monopole source in free space, characterized by a volume velocity injection at position r_0 . This scenario introduces additional volume into the field at the source location, governed by the equation of continuity:

$$\nabla \cdot \hat{\mathbf{u}}(\mathbf{r}) + \frac{j\omega \hat{p}(\mathbf{r})}{\rho c^2} = Q\delta(\mathbf{r} - \mathbf{r}_0) e^{j\omega t}. \quad (2.1)$$

where $\nabla \cdot \hat{\mathbf{u}}(\mathbf{r})$ represents the divergence of the particle velocity, $\hat{p}(\mathbf{r})$ is the sound pressure, ρ is the air density, c is the speed of sound, and $Q\delta(\mathbf{r} - \mathbf{r}_0)$ denotes the source strength as a function of position.

Euler's equation of motion, which remains unaffected by the source, combined with the continuity equation, yields the inhomogeneous Helmholtz equation:

$$\nabla^2 \hat{p}(\mathbf{r}) + k^2 \hat{p}(\mathbf{r}) = -j\omega \rho Q \delta(\mathbf{r} - \mathbf{r}_0) e^{j\omega t}. \quad (2.2)$$

where k is the wave number, representing the spatial frequency of the sound wave.

The solution to this equation under free-field conditions is elegantly captured as:

$$\hat{p}(\mathbf{r}) = \frac{j\omega \rho Q e^{j(\omega t - kR)}}{4\pi R}, \quad (2.3)$$

where $R = |\mathbf{r} - \mathbf{r}_0|$ denotes the distance from the observation point to the source. This solution describes the sound pressure radiated by a pulsating source, highlighting the inverse relationship between pressure and distance, alongside a phase shift that is contingent upon the wave number k and the distance R .

In practical scenarios, when dealing with recordings from the far field of an unknown source location and strength, assuming the source as a point source simplifies the process of understanding its radiation and the sound pressure that can be recorded in the field. This assumption allows for the reconstruction of the sound pressure equation as follows:

$$P_m(r) \equiv P_0 v_m(r) = P_0 v(r_m - r), \quad (2.4)$$

where P_m represents the sound pressure recorded at the m^{th} microphone within the array, P_0 denotes the source strength, and $v(r)$ is the steering vector, which encapsulates

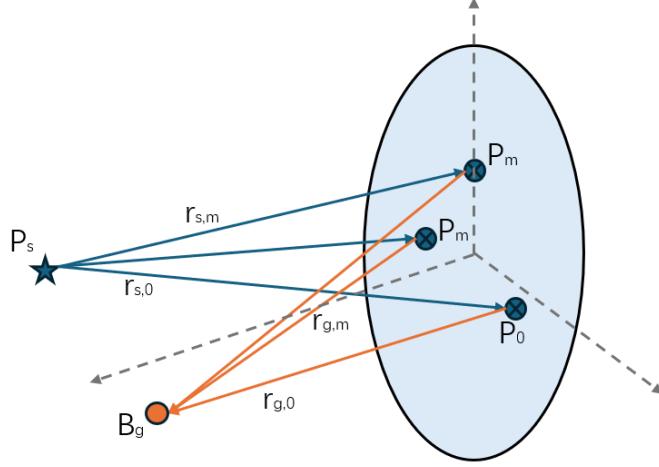


Figure 2.1: Demonstration of how microphone array receives the sound from source with sound strength P_s and B_g represent the autopower spectrum of the reconstructed output, which shall be used to represent the sound pressure level. P_0 is the reference point which is normally chosen at the center of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

the relative position of the unknown source. The steering vector here is defined by the equation:

$$v(r) = e^{-jk|r|/|r|} \quad (2.5)$$

with k being the wave number, and r the distance from the beamforming grid to a specific microphone.

As illustrated in fig. 2.1, the "delay and sum" technique involves scanning through a potential grid by applying the steering vector to the recorded sound pressure at each microphone across frequencies. By summing these results and iterating over all points in the predefined beamforming grid, we obtain the beamforming strength B_g . A higher value of B_g indicates a greater likelihood of the actual source being located at that grid point.

The culmination of this process is a spatially resolved map, highlighting potential source locations through peaks in B_g .

Despite we already have a steering vector shown in eq. (2.5), From the recorded sound pressure at the microphone array point to the prediction of the specific grid point B_g , there are multiple approaches to constructing such a steering vector. It essentially projects the sound pressure from the microphone array to the grid point, as

$$p_F(x_t) = h(x_t)^H p \quad (2.6)$$

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

A general approach does not directly calculate the value of the projected field point $p_F(x_t)$, but it is more conventional to calculate the B_g is to take the autospectrum of $p_F(x_t)$, as

$$B(x_t) = E \{p_F(x_t) p_F^*(x_t)\} = h^H(x_t) E \{pp^H\} h(x_t) \quad (2.7)$$

$$= h^H(x_t) G h(x_t) \quad (2.8)$$

Where G is the cross-spectrum matrix of the recorded microphone array signal. From this equation, we can understand that now we have to find out the steering vector h for reconstruction.

There are two criteria were introduced as a guide for making an accurate steering vector h .[18]

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

$$B(x_t = x_s) > B(x_t \neq x_s). \quad (2.9)$$

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

$$B(x_t = x_s) = CE\{qq^*\} \quad (2.10)$$

where C is an arbitrary constant.

Where there is mutiple ways to foam a steering vector for the reconstruction, but the most suitable approach for a 3d beamforming scenario is by applying a normalization to the steering vector [18], which is derived by form an error function in between the actual crossss-pectrum of recording and the monopole projected crossss-pecturm on the array [19], as

$$E(a, r) = \sum_{m,n=1}^M |C_{nm} - C_{nm}^{mod}|^2 = \sum_{m,n=1}^M |C_{nm} - av_n^*(r)v_m(r)|^2 \quad (2.11)$$

Within this equation, the $E(a, r)$ is the error function that we aim to minimize. It represents the total squared error between the actual cross-spectral matrix elements C_{nm} and the modeled cross-spectral matrix elements C_{nm}^{mod} . The goal is to find values of a (source strength) and r (source location) that minimize this error. The $\sum_{m,n=1}^M$ is summation runs over all microphones in the array, where M is the total number of microphones.

C_{nm}^{mod} is the modeled cross-spectral matrix elements, expressed here as

$$C_{nm}^{mod} = av_n^*(r)v_m(r) \quad (2.12)$$

This model assumes a point source with strength a located at position r , with $v_n^*(r)$ and $v_m(r)$ being the steering vectors for microphones n and m , respectively.

By minimizing $E(a, r)$, the beamforming algorithm seeks to align the modeled sound field with the actual sound field as closely as possible, thereby identifying the most probable source location and its strength.

One way is to find the source strength a to minimize the Error $E(a, r)$, one alternative way is to maximize the image function, which is derived from the eq. (2.12). Maximize the image function is equivalent to minimizing the error function [19]. The image function with diagonal elements removal (autospectra exclusion) were constructed as,

$$J^2(\omega, r) \equiv \frac{1}{\sqrt{M(M-1)}} \frac{\left| \sum_{m \neq n}^M C_{nm}(\omega)v_n(r)v_m^*(r) \right|}{\sqrt{\sum_{m \neq n}^M |v_n(r)|^2 |v_m(r)|^2}} \quad (2.13)$$

Where J is the image function with diagonal removal, and M represents the microphone, Vn and Vm 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 .

As a revision of the eq. (2.7), it is clear that this image function actually forms a normalization to the steering vector, with the diagonal removal of the cross-spectrum matrix, it has been proven to have a good ability to reflect the correct location of the source, yet, can only result in an estimated source strength map. [18]

2.1.2 Eigenvalue extraction of the Cross spectrum matrix

Another term besides the Steering vector in the Beamforming map is the Cross-Spectral Matrix (CSM). The CSM encapsulates the frequency-domain relationships between signals received at different sensor elements, which as a matrix, abstractly should be considered as a representation of the sound field, in the frequency domain. Since then, because it is a maximization of the sound field recordings, it is very natural that we want to extract the main component from it if we understand the major component of CSM came from the contribution of the source.

So, the 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.

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. The MUSIC is based on one idea, that is taking the reciprocal of the noise map, which the noise map is derived by the eigenvector matrix of CSM without the predefined numbers of eigenvectors, which means we manually removed the eigenvectors contributed by the source outside of the CSM, which then it will form a noise map, as

$$B(x_t)_{\text{Noise}} = h^H(x_t) E_n E_n^H h(x_t) \xrightarrow{x_t=x_s} 0 \quad (2.14)$$

where $E_n E_n^H$ form the Cross-spectrum without the signal contributed eigenvector, the x_t is the grid point, x_s is the source location. When the grid point is equal to the source point, the noise map $B(x_t)_{\text{Noise}}$ will then get very close to zero. Then we can take the reciprocal of the noise map, and then,

$$B(x_t)_{\text{MUSIC}}(\theta) = \frac{1}{h^H(x_t) E_n E_n^H h(x_t)} \xrightarrow{x_t=x_s} \infty \quad (2.15)$$

where $h(x_t)$ is the steering vector associated with a potential source location x_t , and E_n contains the eigenvectors corresponding to the noise subspace.

The peaks in the MUSIC spectrum indicate the estimated source location, in the idealized situation it will reach infinity. This method excels in localization accuracy and low frequency super-resolution but does not directly provide estimates of source strength.

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:

$$B(x_t)_{\text{Functional}}(x_t) = |h^H(x_t) \cdot S \cdot h(x_t)|^v \quad (2.16)$$

The matrix S is a modified version of the CSM that incorporates these eigenvalue-based weights

$$S = CSM^{\frac{1}{v}} \quad (2.17)$$

Where v is a constant defined by the user, If we understand the CSM as a transfer matrix that represents the sound field, then take the order of $\frac{1}{v}$ is essentially suppressing the vectors that contributed by noise. Unlike the Music, the estimated source strength has been recovered by applying the v th power after steering through the grid.

2.1.3 Clean based on Source Coherence

Besides our previously introduced traditional DAS or beamforming through the modification to the CSM, there is another method that we will use in the following measurement, which is the Clean based on Source Coherence, CleanSC Beamforming. This is an iterative signal source isolation method, in which the source identification is based on their spatial coherence. The algorithm proceeds by mathematically modeling the identified coherent signals. These models are then subtracted from the original composite signal. This step is repeated, each time focusing on the next set of dominant signals, gradually filtering out noise and enhancing the clarity of the signal of interest.

As a result, the CleanSC in our application will result in only a few very small points, that represent the highest value point within each coherent area. One of the key advantages of CleanSC lies in its targeted approach to signal processing. Unlike NNLS and similar iterative methods that scan across the entire grid iteratively, CleanSC only focuses on areas exhibiting the most coherent signals, and will not have to iterate many times through the entire grid. Since in our application, the 3d scanned mesh grid in most the time is in a very large size, the NNLS or many other iterative methods will become un-applicable.

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 [13] and its related open source SDKs [20], alongside various projects focused on 3D point cloud processing. This includes using the OpenCV rgbd module [21], the Visual Tool Kit (VTK) [22] for image processing, OpenCL for accelerating the image processing [23], and Open3D [24] 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[20], 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:

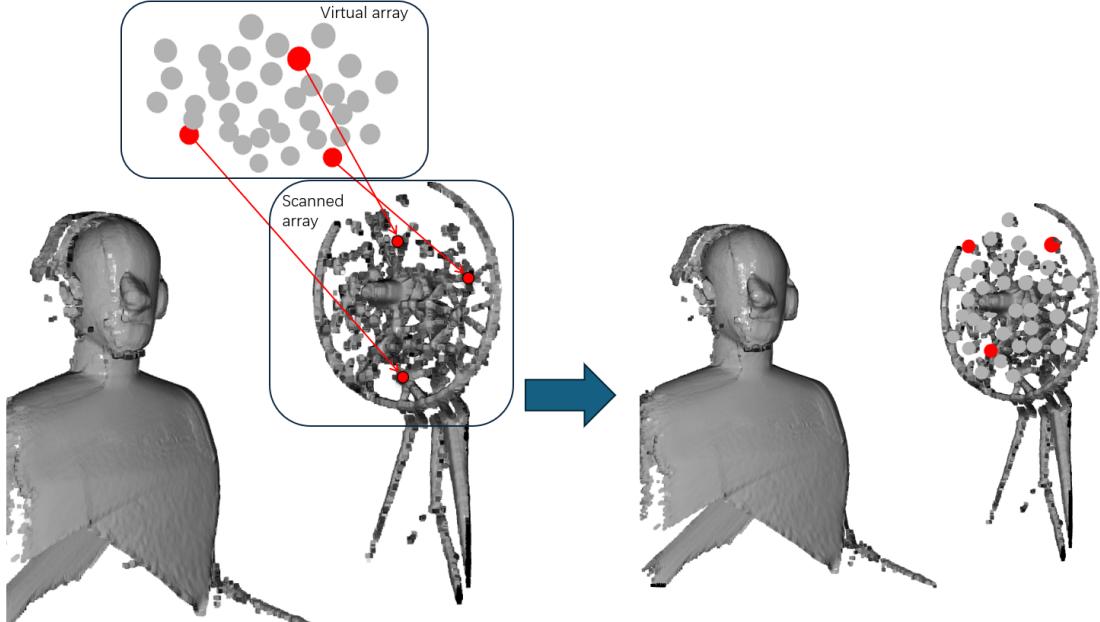


Figure 3.1: Process of manual registration. The red point in both the virtual array and the scanned array is the corresponding point that should be manually selected. After the translation, the virtual array will be panned 20mm towards the normal direction so that its center points fall exactly on the 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 employed the Iterative Closest Point (ICP) method during the 3D scanning process to amalgamate multiple 3D scanned point clouds into a singular point cloud model. This approach leverages resources from the open-source repository provided by Microsoft's *Kinfu_example* project, which utilizes the OpenCV module '*OpenCV_rgbd*' and the Visualization Toolkit (VTK). It's important to clarify that our involvement with the ICP algorithm and any open-source computer vision project was limited to compilation and application, without contributing to the development of these technologies. To ensure comprehensive understanding, a succinct overview of the ICP registration theory, as applied in our project, is presented below.

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:

$$E(R, t) = \sum_{i=1}^n \|(R \cdot p_i + t) - q_i\|^2 \quad (3.1)$$

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 Mesh generation

In this project, the Open3D library's capabilities for surface reconstruction and mesh generation play a pivotal role [25]. Among the available algorithms, the Ball Pivoting Algorithm (BPA) [26] was selected for its efficiency in constructing meshes from point clouds. The BPA simulates the traversal of a virtual ball across the point cloud, forming a triangle

whenever it contacts three points. This iterative process continues until all potential points within the point cloud are examined. The algorithm's effectiveness is particularly notable when the ball size is optimally set, which largely depends on the point-to-point distance within the cloud.

Given that the Azure Kinect scanner produces point clouds of fixed resolution, the distance between points remains relatively consistent. With appropriately configured parameters, the BPA can rapidly construct meshes, typically completing the process within 10 seconds for our scans.

The ball size, a crucial parameter in BPA, directly impacts the mesh's fidelity and completeness. Within our application, the parameters governing the ball size are specified as follows:

$$[\alpha_{min} = 0.0005, \alpha_{max} = 0.006, \hat{\alpha} = 0.001, \alpha' = 0.0005] \quad (3.2)$$

Here, α_{min} and α_{max} represent the minimum and maximum ball diameters, respectively. $\hat{\alpha}$ is the initial average diameter, serving as the starting point for the ball size, and α' denotes the incremental step, the unit is in meters. This parameter setting ensures that the BPA can adapt to the specific characteristics of the point cloud generated by the Azure Kinect scanner, facilitating efficient and accurate mesh construction.

4 Set of system

This chapter outlines the comprehensive system setup utilized for the acoustic measurements, detailing the equipment employed, the objects selected for measurement, and the specific measurement parameters. Also the generation of the stimulus to the system, and the lab environment. Additionally, it provides an overview of the microphone array performance.

4.1 Device of measurement

The measurement apparatus comprises two main systems. The first is the Microsoft Azure Kinect, as discussed in Chapter [reference], which utilizes optical and depth sensors to capture and amalgamate geometric information, resulting in a point cloud. The second system is a Brüel Kjær Acoustic Camera System, consisting of a tripod for support, a LAN-XI System with three 3660-C frames of the 3053-B 12-channel as for the front-end in order to operate the microphone and store the recordings. And a Sliced Wheel Array equipped with $36 \times$ B&K 4959 quarter inch Microphones.

A photograph depicting the laboratory setup with all mentioned devices is presented in fig. 4.1. The actual experimental setup is somehow more complex, as it necessitates the connection of the laptop recording and operating of the Azure Kinect scanner. Additionally, an extra Lan-XI front-end serves as an amplifier for the Adam A5x loudspeaker when it needs to be used, and another Lan-XI front-end is used with an in-ear microphone in HATS to record the in-ear sound pressure levels in the case for headphone equalization. For brevity, the description focuses on the primary measurement devices.

During the measurement, the system recording the About the B&K acoustic camera, or mentioned as the microphone array, we do not involve into any development of this system, but on

Microphone Array performance

The microphone array, configured as a 36-channel sliced wheel, is subject to various performance evaluation methods. Disregarding potential variations among individual microphones, we proceed under the assumption that all microphones in the array record without bias. Consequently, the most definitive method to verify the array's performance is through direct experimental testing.

To gauge the array's effectiveness, we executed a test aimed at differentiating between two distinct, incoherent virtual sources positioned 0.7m away from the array, located at $\pm 0.3\text{m}$ along the x-axis, respectively. The source strength is 1pa rms, which corresponds to 94dB SPL. As a result, The beamformer's map at 8000Hz, illustrated in fig. 4.2, showcases the capability of all beamformers to clearly separate the sources within an idealized, noise-free scenario. However, the peak values represented in the map slightly deviate from the expected, which is anticipated due to the limitations inherent in the steering vector.

Another metric, the Maximum Side Lobe Level (MSL), was calculated to identify the highest lobe excluding the main lobes, is for the demonstration of the ability of the array to separate the main lobe and the side lobe. The results shown in fig. 4.3, indicate a general decrease in MSL with frequency, except for the Music beamformer. Intriguingly, Music's MSL remains relatively constant at lower frequencies and increases at higher frequencies.

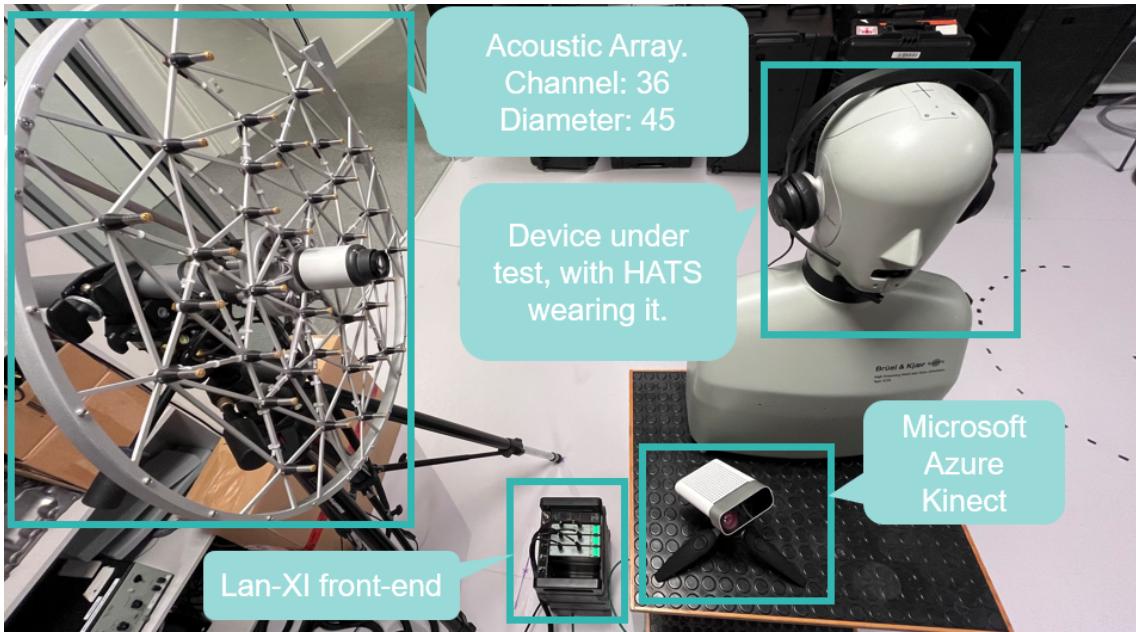


Figure 4.1: Main devices for the measurement

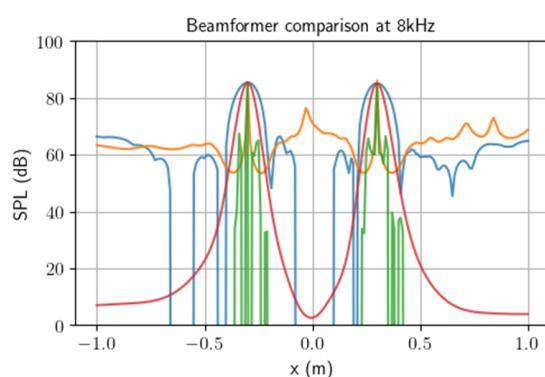


Figure 4.2: Beamformer grid and source at $z = 0.7\text{m}$

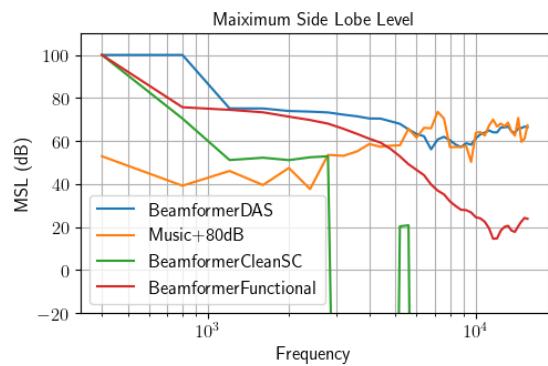


Figure 4.3: MSL of beamformers. If less than two peaks are found, then the MSL will be set to be 100dB.

Further investigation at 200Hz, as shown in fig. 4.5, reveals that while DAS and Functional Beamformer display a flat response and CleanSC fails to accurately locate the sources, Music accurately identifies the source locations. However, as expected, Music does not accurately represent the source strength due to its intrinsic limitations.

This outcome can be attributed to the principle of the Music algorithm, which, unlike other methods, its projects back to the noise rather than the source, and then it take the reciprocal of the result map to extract the source location. By this way of operation, it is almost unlimited by the beamwidth, or the resolution of the mainlobe, because its mainlobe is pointing towards to the noise, instead of the source. In our noise-free simulation, Music theoretically performs optimally. Alternatively, this can be understood as Music sacrificing source strength identification for superior source separation capability.

It's important to note, however, that MUSIC's effectiveness is contingent upon accurate grid placement. An illustrative example of this limitation is observed when positioning two sources at 1m on the z-axis but directing the beamformers towards a grid at 0.7m on the

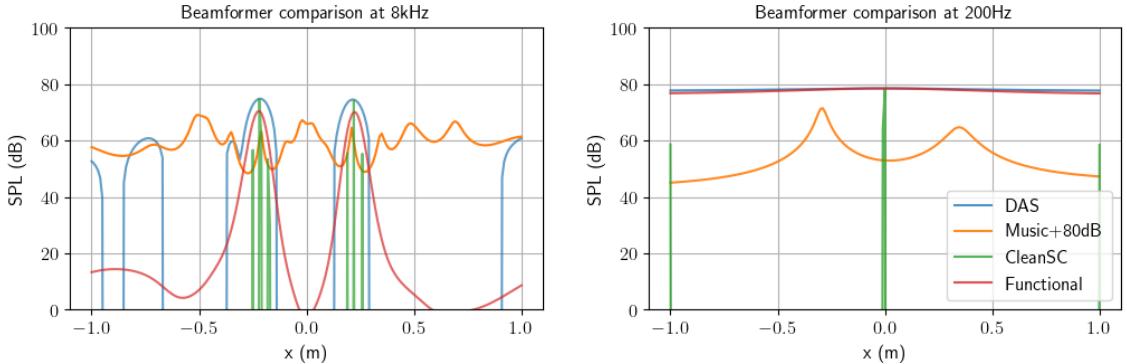


Figure 4.4: Beamformer grid at $z = 0.7\text{m}$, Figure 4.5: Beamformer grid and source at source located at $z = 1\text{m}$

z -axis. At 8000Hz, as depicted in fig. 4.4, while other beamformers accurately localize the sources, MUSIC erroneously identifies the highest peaks at $\pm 0.5\text{m}$ instead of the actual $\pm 0.3\text{m}$, leading to incorrect localization.

This observation underscores that all beamformers in this array effectively localize sources situated 0.7m away across frequencies beyond 1000Hz. Specifically, MUSIC demonstrates exceptional resolution at lower frequencies, provided the array's positioning relative to the grid is precise. The algorithm's super-resolution capabilities are significantly dependent on the accurate placement of the grid.

4.2 Device under test

The devices under examination are depicted in fig. 4.6. Among them are two headphone models by Jabra: the Jabra Evolve2 75 [27] and the Evolve2 65 [28]. The Jabra Evolve2 75, referred to as HP1 in subsequent discussions, is the larger model equipped with a more substantial cushion and an Active Noise Cancellation (ANC) feature. Its ANC functionality is dependent on external microphones, which may be susceptible to sound leakage from the device itself. The Jabra Evolve2 65, designated as HP2, offers greater portability with a collapsible hinge for compact storage and lacks the ANC feature, potentially leading to more significant sound leakage due to its lighter construction.

Additionally, the Adam A5x [29], a pair of loudspeaker boxes, is included as a verification device due to its known tweeter and woofer locations.

The Vue smart glasses, which emerged from a crowdfunding campaign in 2016 [30], introduce a unique method for personal audio delivery. Utilizing bone conduction technology, these glasses transmit sound through a vibrator positioned on the temple, just beneath the helix and in direct contact with the wearer's head. Despite the innovative design, the bone conduction vibrator's placement allows sound to radiate into the surrounding environment during use. This characteristic suggests potential design challenges in effectively minimizing sound leakage, suggesting an area for further refinement in bone conduction audio devices.

For simulating the acoustic properties of a human head and torso, the Brüel & Kjær Head and Torso Simulator (HATS) [31] is utilized. This simulator is equipped with a Type 4620 ear simulator, capable of replicating the real pinna and ear canal's shape and acoustic performance. It also features a built-in microphone to measure the sound pressure level within the ear canal. This capability is crucial for later stages of the study, where it will be

used to equalize the output of the two headphone models for a reciprocal comparison.

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 4.6: List of 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. 4.7. The noise

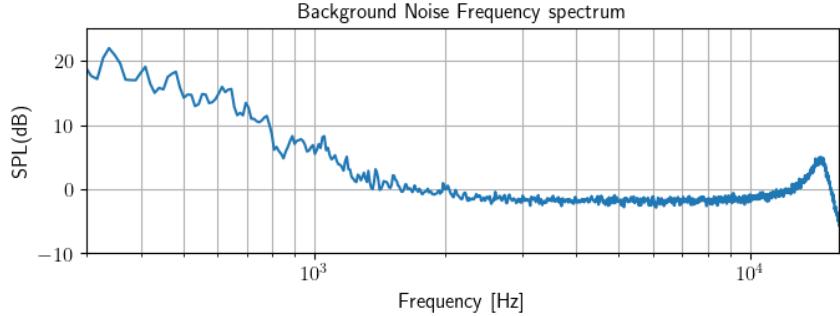


Figure 4.7: Background noise collected by the array, the frequency range is from 300Hz to 11kHz.

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.

4.4 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 [32], 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 $Y(\omega)$ 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 in an output $Y^*(\omega)$ that closely approximates the desired flat frequency response, ideally matching the input white noise in frequency content.

The process of equalizing the input stimulus for headphones involves creating a reverse filter based on the frequency response of the system, including both the headphones and the Head and Torso Simulator (HATS). This approach ensures a flat frequency response across all ranges and consistency across different devices. The methodology for constructing this reverse filter and applying it for equalization is detailed below, illustrated with reference to fig. 4.8. The right panel of fig. 4.8 demonstrates how the frequency response of a Finite Impulse Response (FIR) filter is derived. This process involves taking the square root of the system's frequency response, adhering to the principles of minimum phase filtering. The goal here is to adjust the amplitude response to the desired level while minimizing any deviations in the phase response. By applying the square root operation, the FIR filter is tailored for a zero-phase filtering configuration, which involves filtering the signal forwards and then in reverse. This dual application ensures the elimination of phase distortions encountered in the initial pass and preserves the signal's phase integrity. The square root operation is critical because the filter is applied twice, ensuring the amplitude response is accurately reflected in the final output.

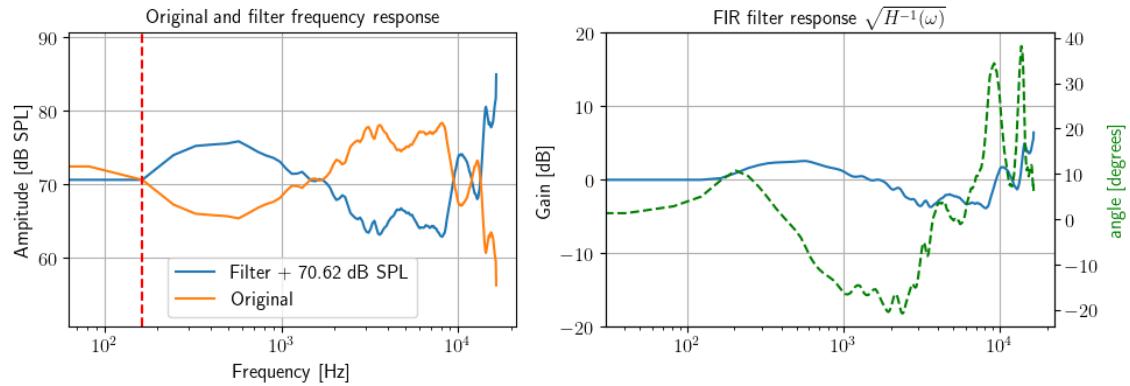


Figure 4.8: Magnitude and phase responses of a reverted filter and an FIR filter. The left plot shows the frequency response of both the filter and the original signal, with the filter's reversion point indicated by a red dashed line. The right plot displays the frequency response of the FIR filter, derived from the filter shape's square root value, alongside its phase response (green dashed line).

The original band-limited white noise signal is processed through the obtained FIR filter. This step modifies the white noise, adjusting its frequency content to counterbalance the

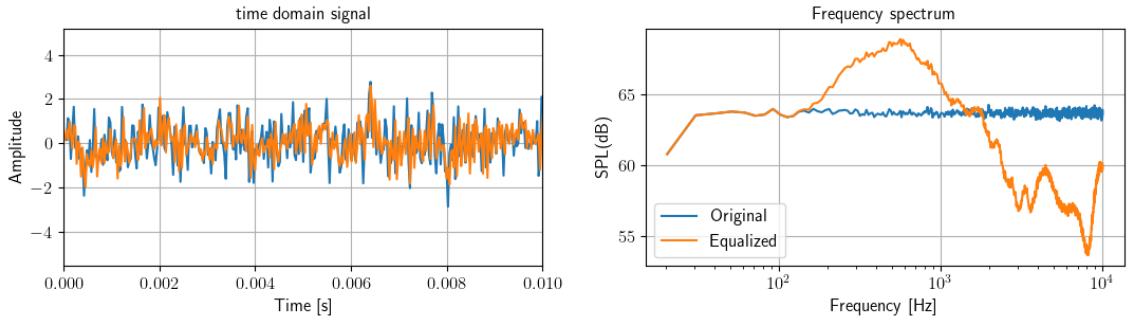


Figure 4.9: Original and Equalized white noise stimulus for HP2

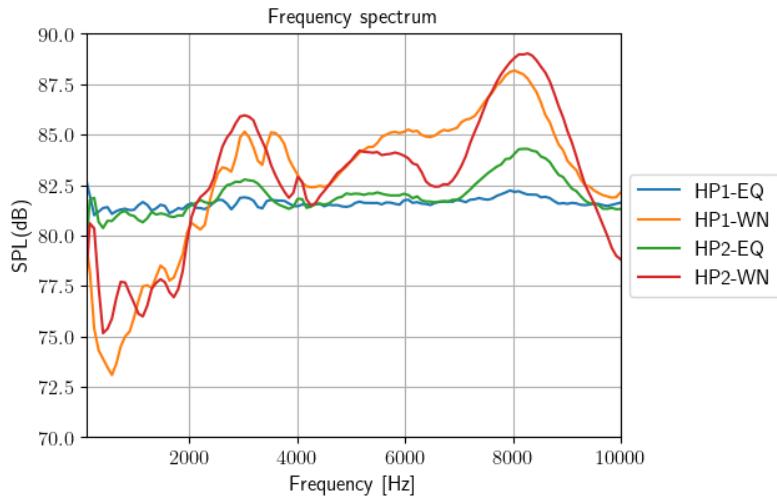


Figure 4.10: Comparison of the equalized response with the original frequency response under the white noise stimulus.

identified discrepancies in the headphone's response. The result white noise wave shape and the frequency response as shown in fig. 4.9.

By applying this procedure onto the two headphone HP1 and HP2, we can obtain a equalization FIR filter and the corresponding equalized white noise. Use these signal as the input stimulus for the microphone, we can obtain the sound pressure recorded by the HATS in-ear microphone, the frequency response of the recorded signal as in fig. 4.10.

After obtaining the flat response, by manually adjusting the signal output strength, which is the audio volume level, we can ensure that the input stimulus to each headphone is consistent, allowing for an objective comparison of their sound leakage characteristics. By equalizing the signal to a flat frequency response, we mitigate the influence of inherent discrepancies in headphone responses on the measurement outcomes. This method provides a reliable basis for assessing and comparing the acoustic performance of different headphone models under uniform testing conditions.

5 Measurement

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 glasseses, the stimulus is equalized white noise, as detailed in section 4.4, 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 rate of 32,768 Hz. This configuration ensures that the data collected will be adequate for analyzing the measurement frequency range of interest.

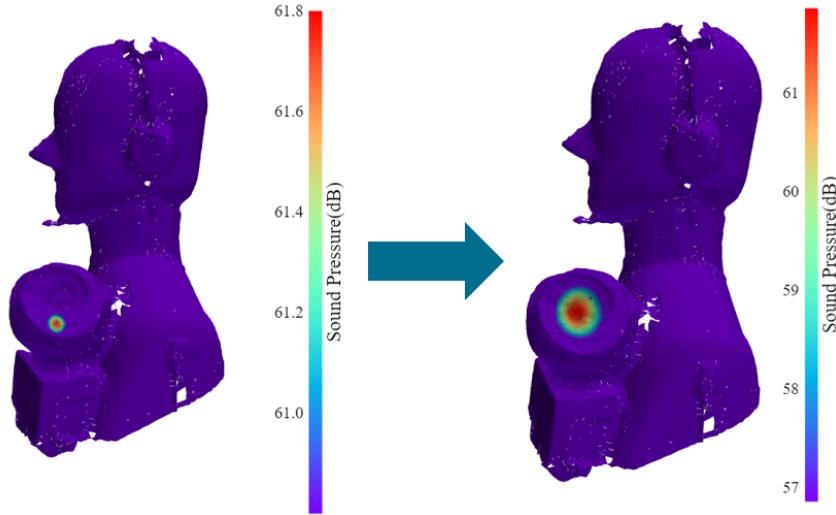


Figure 5.1: Array alignment calibration. The left plot is the beamforming image from uncalibrated array location, measured at 8000Hz, 1/3 octave bandwidth, and 1dB dynamic range; the right plot is the beamforming image after the array calibrated, with same frequency and bandwidth, but set to be 5dB dynamic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

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 [33]

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 [29]

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 [9]. 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 under test.

The result from PULSETM NSI Array Acoustic as shown in figs. 5.2, A.7 and A.8; The

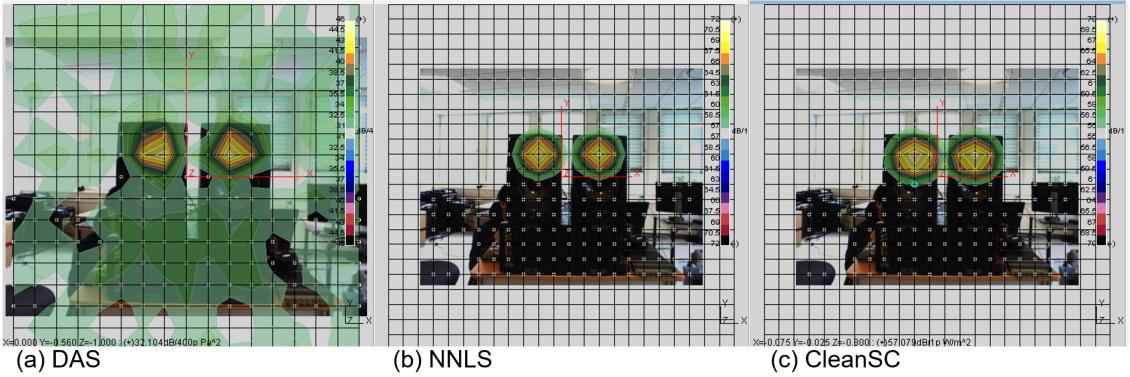


Figure 5.2: NSI-Array acoustic postprocessing result under 8000Hz, 15dB dynamic range

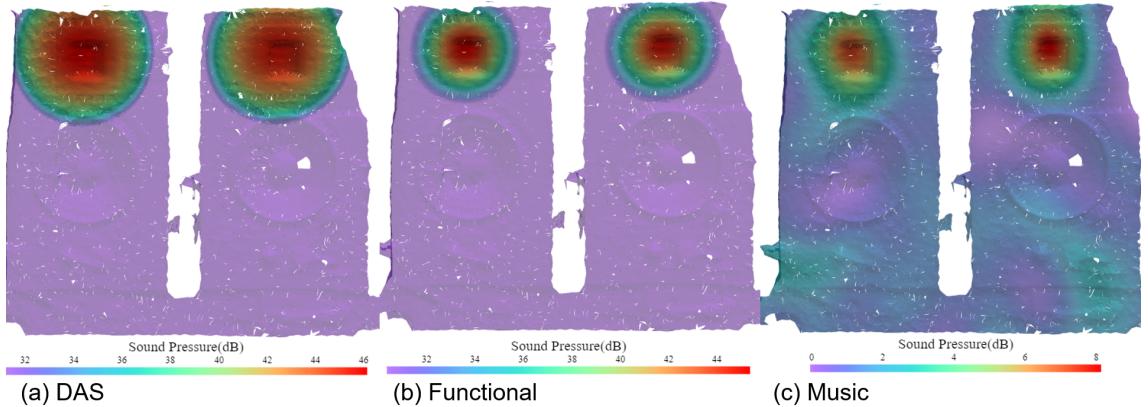


Figure 5.3: 3d beamforming result under 8000Hz, 15dB dynamic range.

results by 3d beamforming as shown in figs. 5.3, A.9 and A.10. For brevity, in the following text is only the results of 8000Hz, and the results of 2500Hz and 5000Hz is shown in appendix A.2

Additionally, as we addressed in section 4.1, The Music under the idealized situation will show a very good super-resolution ability under low frequency. To verify that, a comparison of the Music with two do-convolution Beamformer, CleanSC and NNLS under 1500Hz frequency has the results as shown in fig. 5.4. From the result, it is clear that neither the NNLS nor CleanSC can not have the correct detection of the source location, yet the Music still can separate two woofers, though the left woofer has around 3dB less than the right one. The Music still shows an extinct ability under the low frequency source localization.

Beamforming CleanSC filters only the highest value out from the beamforming image, and suppresses all the spacial correlated images to zero. So that the result from CleanSC without the smoothing will only give a signal mesh point with corresponding strength, which is very convenient for 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 an algorithm 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 together within one beamforming plot. Beneficial from the naturally high resolution of the beamforming scanning, it can give a clear trajectory of how the source travels with frequency. Under that idea, the fig. 5.5 shows how the highest source traveling when

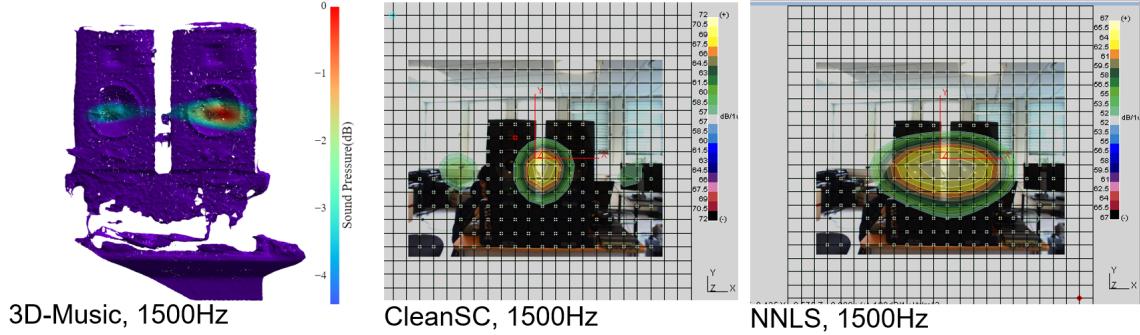


Figure 5.4: Comparison of Beamformers under a very low frequency

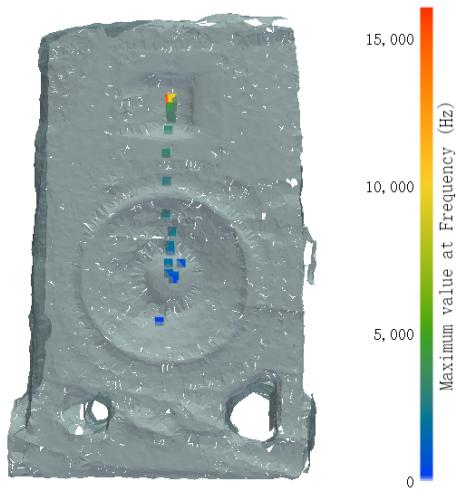


Figure 5.5: The highest beamforming value obtained by CleanSC as a location response to frequencies.

measuring only one box of the Adam A5x. It shows a very clear trend of movement, that under the low frequency, the source (colored as blue) mainly stayed within the woofer's range, towards to higher frequency the source gradually moved to the tweeter's location.

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 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. 5.7, sound leakage occurs at the upper front of the HP2 at 8000Hz. 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 functional 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. (2.13). 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

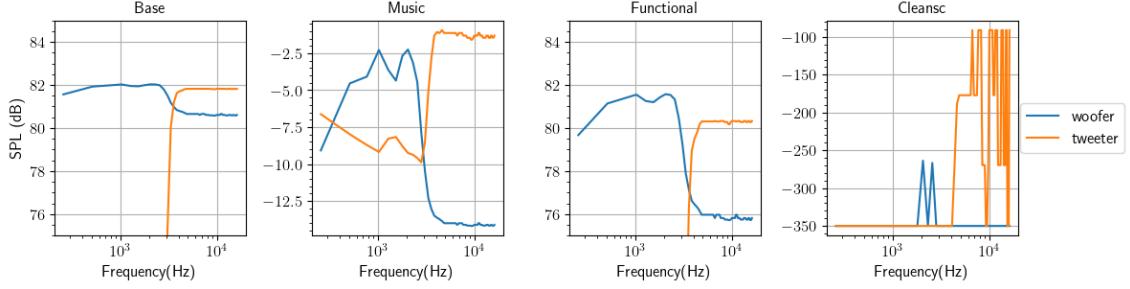


Figure 5.6: The frequency response of two point set, with one at the woofer, and another one at the tweeter location. Results from 4 different beamformers for comparison. The point set is picked based on the result shown in fig. 5.5, first 7 highest point index (as woofer) and last 7 highest point index (as tweeter) were picked and averaged by values.

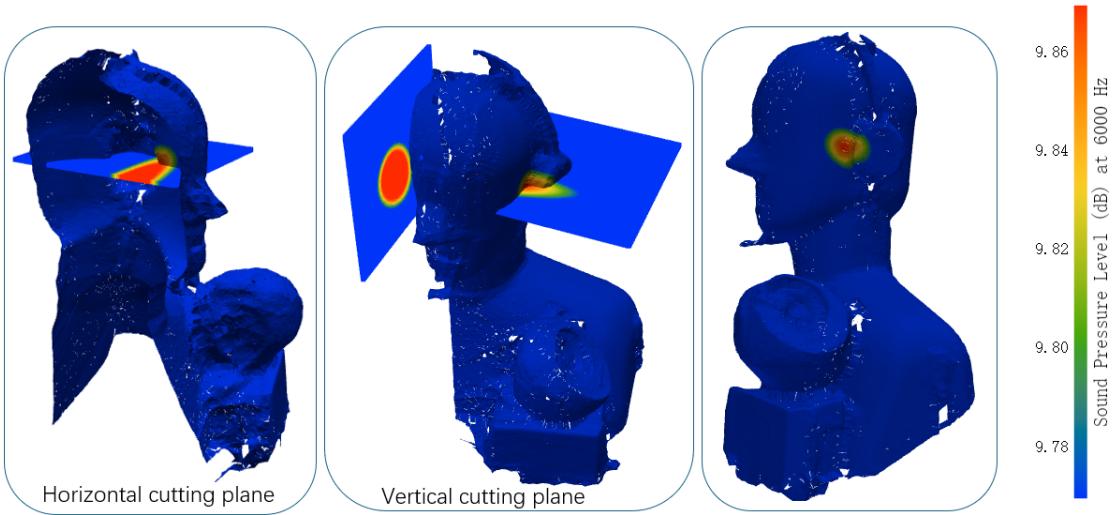


Figure 5.7: Beamforming reconstruction by functional Beamforming, to a scanned 3d grid, a Horizontal planar grid, and a vertical planar grid. The dynamic range is limited to 0.1dB to narrow the size of the main source.

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 fig. 5.5, 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 fig. 5.8. The configurations examined were: the headphones worn in their standard manner, with a small porous foam attached by tape at the front between the ear and the ear cup to serve as a sound absorber, and with the foam similarly taped 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

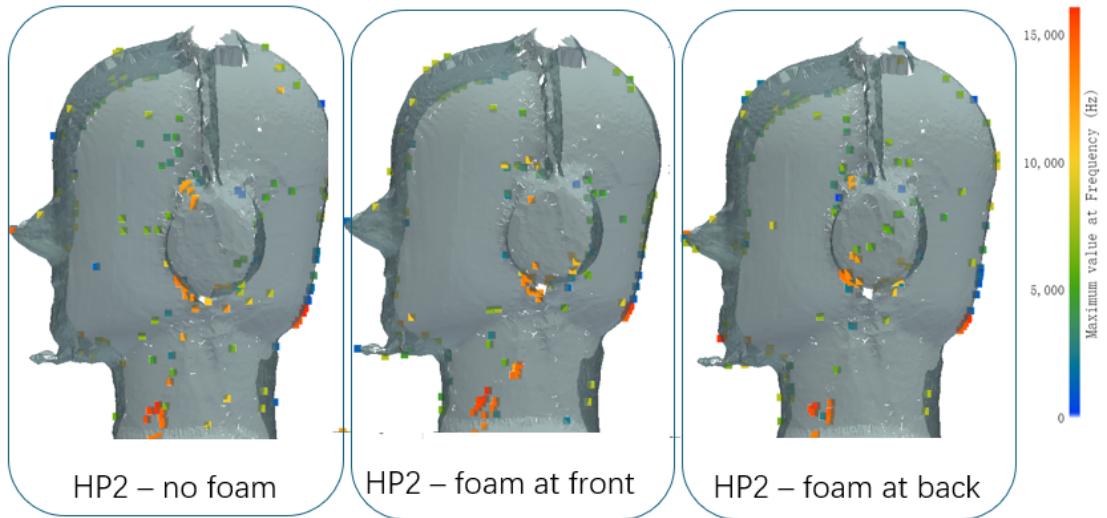


Figure 5.8: The three highest beamforming values obtained by CleanSC as a location response to frequencies. Due to the complexity of plotting, the order of point levels is not demonstrated in this plot, but we list them equally as a square. From this plot, we can understand how the sound leakage “travels” under different frequencies.

yellow, and frequencies above 10kHz in red. This color-coding facilitated an observation 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. 5.9, 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

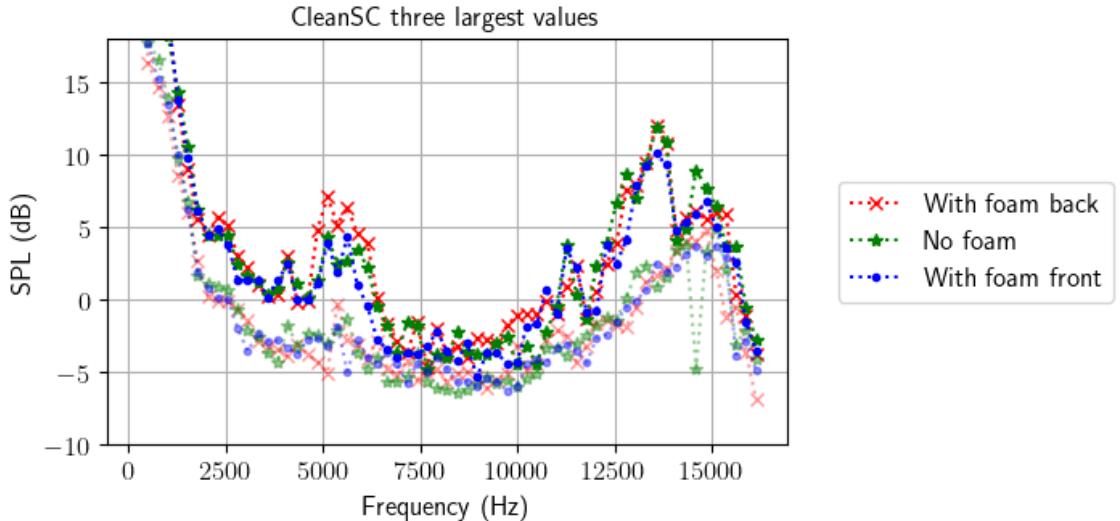


Figure 5.9: As illustrated in fig. 5.8, we present the SPL of the top three highest value points, as determined by the CleanSC beamforming, across various foam configurations and their frequency response. To facilitate easier identification within the figure, the highest value point is depicted with a solid line, while the second and third highest value points are represented with transparency in same color.

the front bottom and the upper back of the ear cup. This analysis suggests that the addition of foam at front can effectively reduce high-frequency sound leakage, especially for frequencies above 6000Hz.

However, when the foam is taped at the back, 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.11 and A.12, 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 energy release, leading to enhanced structural vibrations. This might be due to the same principle of the occlusion effect [3], that a well-sealed headphone will transfer the vibration to the ear cartilage, which might result in an additional vibration. 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

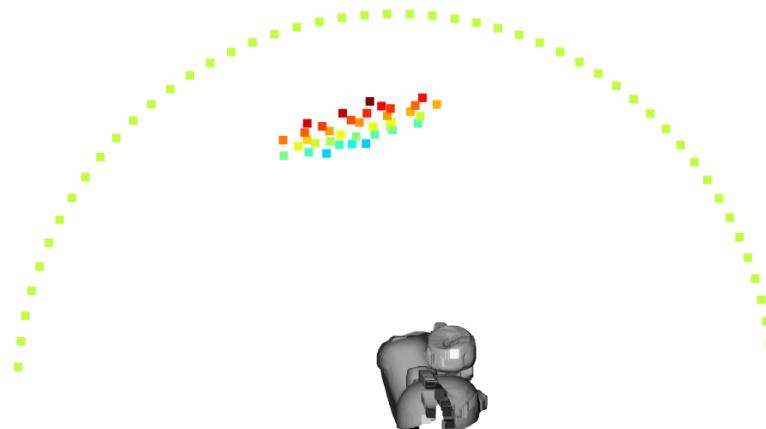


Figure 5.10: A total of 50 points are evenly spaced in a horizon plane perpendicular to the earcups of the headset under test, with the angle to the normal vector of the earcups arranged clockwise from -90 degrees to 90 degrees. The point set in between the circular gird and mesh model is the virtual microphone array.

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 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 through the structure.

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 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 fig. 5.10, 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 fig. 5.11. 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,

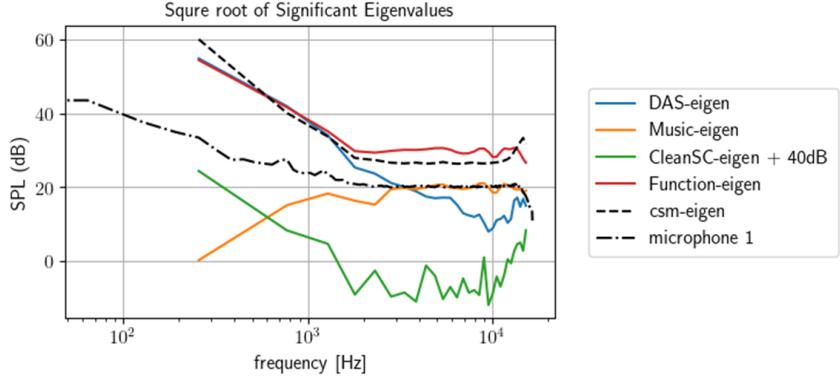


Figure 5.11: Significant eigenvalues from the square root of the cross-spectrum matrix of the array by actual recording and simulation of HP2 measurement. And the actual recorded frequency response at the microphone 1 position.

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 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.

As for the Music, it shows a very close alignment with the on-axis response at high frequency but deviates from both the eigenvalue and the microphone response under 1kHz. Considering its super-resolution properties, this might be due to its mainlobe under low frequency being too narrow to reflect the actual situation.

To further dissect these findings, we plotted the simulation and actual pressure frequency response for microphone 1 in the array, as shown in fig. 5.11. Here, the Music Beamformer's level matches the actual recordings remarkably well, particularly at higher frequencies, with negligible discrepancy, but deviates from both the eigenvalue and the microphone response under 1kHz. This might be due to its mainlobe under low frequency being too narrow to reflect the actual situation. Also, the plot shows that the Music Beamformer exhibits a frequency response at high frequencies almost identical to that of the Functional Beamformer.

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. 4.7, 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

As addressed in chapter 6, the simulation to 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. The resulting sound radiation patterns, with foam placed at the front and without foam, are illustrated in fig. 5.12 and fig. A.13, respectively. It's important to emphasize 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 results at all angles receive a constant value 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 beamformer surface reconstructions, albeit without affecting the directional characteristics of the sound radiation pattern.

In our investigation of sound radiation patterns, the grid setting and the object orientation are the same as in fig. 5.10, as the left headphone faces the grid, with the nose pointing to the +90 degree. Also, we simulated all four beamforming algorithms for comparative analysis. All results were equalized concerning 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, if the sound field is considered as the combination of multiple monopoles distributed on the surface of the geometry, 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 the 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, though there is no available directivity

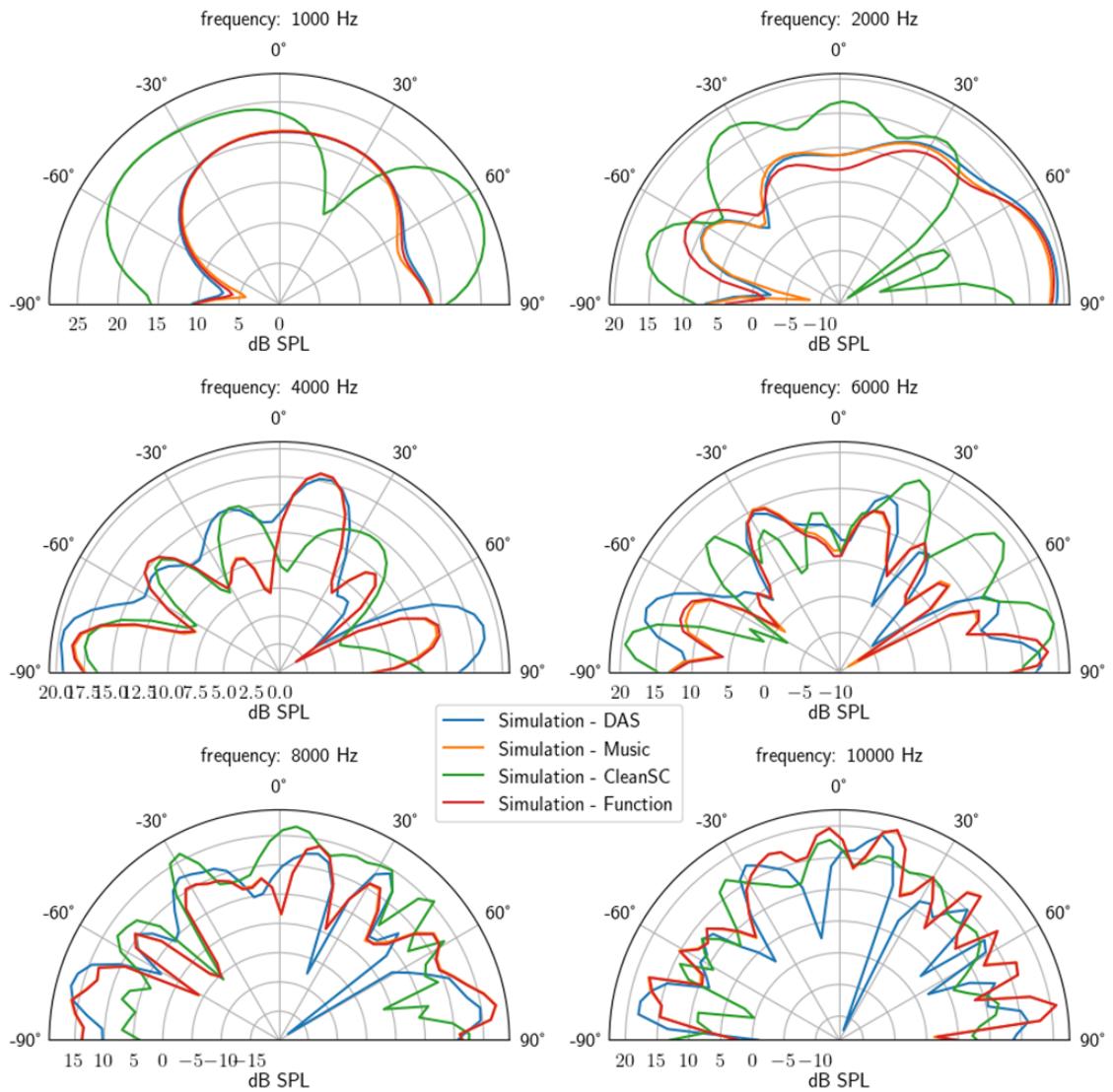


Figure 5.12: Simulated sound radiation pattern when the foam is attached at the front gap of HP2. All the directive response curves are equalized based on the response shown in fig. 5.11. Note that the yellow curve which represents the simulation result from Music almost overlaps with the simulation result from Function, especially under high frequency.

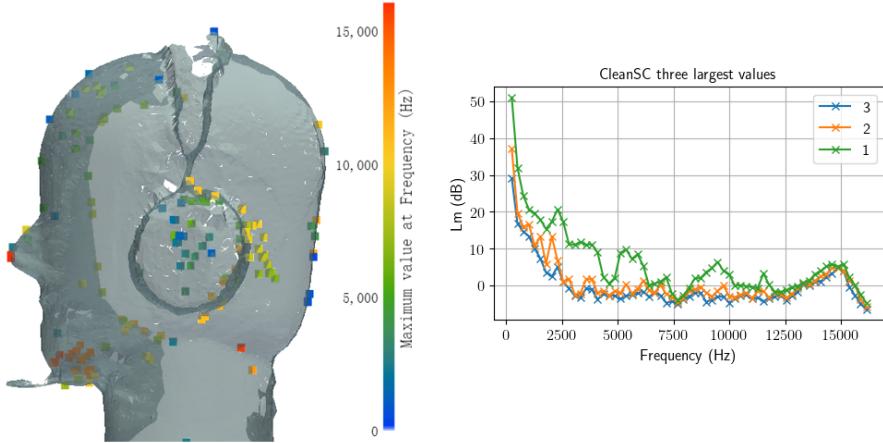


Figure 5.13: The 3 highest source strength SPL and it's location with response to the frequency.

test to HP2, the conventional radiation pattern measurements conducted using the Klipper system revealed similar trends overall [8]. However, the accuracy of the results remains to be verified.

The comparison between the sound radiation patterns in the normal wearing scenario and 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. 5.13, 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, fig. 5.14 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

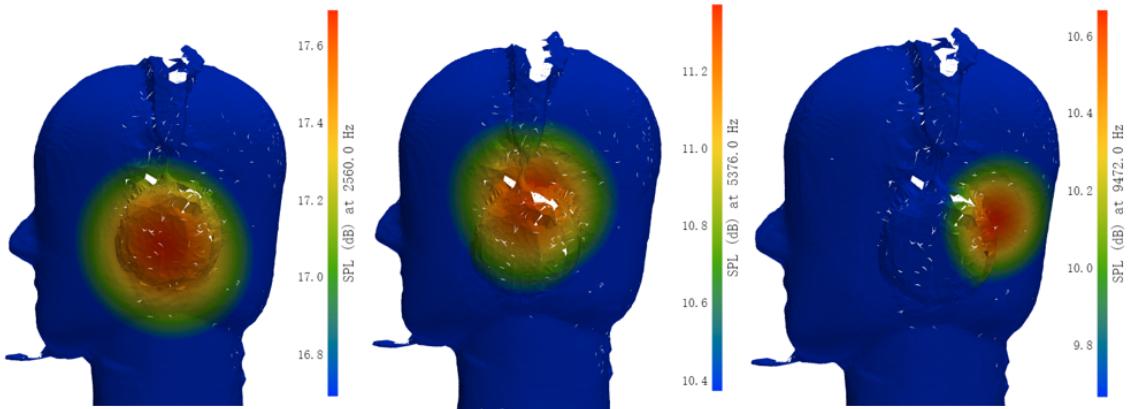


Figure 5.14: HP1 SPL map reconstructed by the Beamformer Functional. The dynamic range is limited to 1dB.

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 across frequencies have been relocated to appendix A.5. The simulated and measured eigenvalues, alongside the recorded array frequency response, are presented in fig. A.14, while the radiation pattern is detailed in fig. A.15. 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 glasses

5.6.1 Leakage localization

The Vue glasses, introduced to the public in 2016 through a crowdfunding campaign [34], represent an innovative approach to personal audio. Constructed from hard plastic, these glasses feature vibrators mounted on each glasses 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 glasses 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 illus-

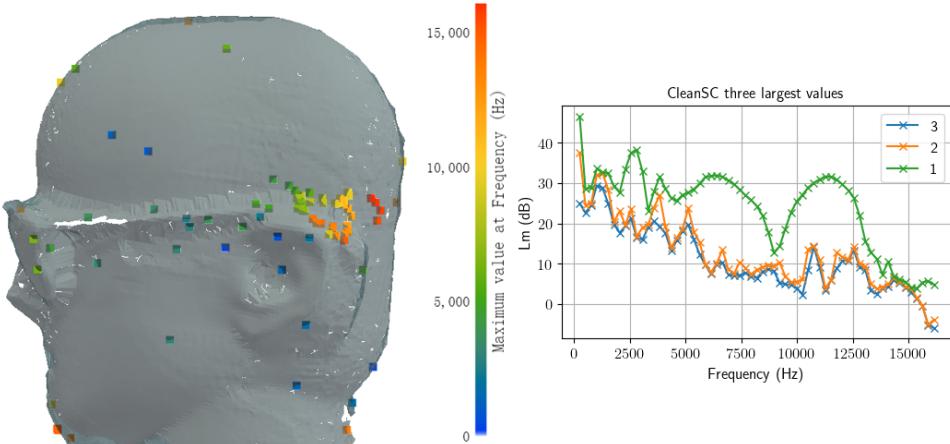


Figure 5.15: CleanSC highest value point travel plot of the Vue glasses, and the corresponding strength values. This time we only plot one highest points instead of three, which is due to the second and third points might also located close to the glasses frame and result confusion.

trated in fig. 5.15. Compared to traditional headphone leakage, the Vue glasseses 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 glasses temple tips, while low-frequency radiation is more central along the glasses 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 glasses 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 glasses have both structural vibration and direct radiation from the vibrator. Also, from the point strength we observed a significant decaying of the level around frequency 8500Hz, which might be due to the anti-resonance of the structure.

5.6.2 Simulation of Radiation directivity pattern

The radiation directivity of the Vue glasses and its verification process are comprehensively detailed in appendix A.6. In a departure from previous demonstrations, this analysis juxtaposes the simulated on-axis response directly with actual measurements taken at array microphone 1 position, as depicted in fig. A.16. Furthermore, significant eigenvalues derived from both simulation and measurement are showcased in fig. A.17, unveiling intriguing trends between these two configurations.

The comparison of on-axis responses reveals that the Functional Beamformer closely aligns with the measured curve at lower frequencies, up to 3000Hz. Beyond this point, although it continues to mimic the peaks and troughs of the measured frequency response, the deviation from the original curve becomes increasingly noticeable. In contrast, the extracted eigenvalues highlight divergent trends at higher frequencies, with the Music Beamformer, in particular, displaying a notable level discrepancy, indicating a potential misalignment with the actual recordings.

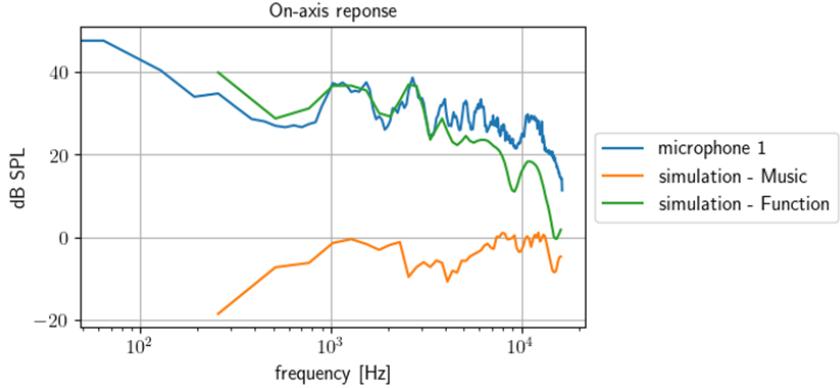


Figure 5.16: Simulated and measured frequency response of Vue smart glass.

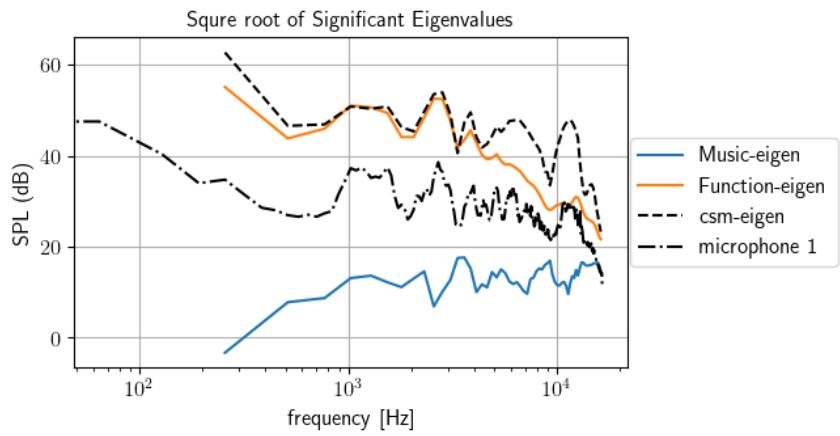


Figure 5.17: Simulated and measured eigenvalue with frequency response of Vue smart glass.

Despite previous findings indicating nearly identical radiation patterns between the Music and Functional Beamformers after equalization based on the recorded cross-spectrum matrix eigenvalue, the Vue glasses exhibit a significantly different simulated radiation pattern, as illustrated in fig. A.18. Beyond 2000Hz, the Music and Functional beamforming techniques present distinct patterns, which tend to align again around 8000Hz and 10000Hz. At the highest frequency of 10000Hz, the observed patterns between 30 to 60 degrees suggest that the Music Beamformer yields lower strength but more defined peak values at 30 degrees, while the Functional Beamformer offers higher values with a less precise directivity at this angle.

This observed discrepancy likely stems from the unique acoustic properties of the Vue glasses. Particularly beyond 8000Hz, the primary source of radiation seems to transition from the vibrator's location to the tip of the glasses' temple, hinting at a vibration pattern resembling that of a monopole. The orientation of the temple tip towards 60 to 90 degrees, with the temple's normal vector directed towards 0 degrees, may account for the differences noted: the Music Beamformer captures a narrower main lobe, accurately delineating the structure of primary and secondary sources, whereas the Functional Beamformer's broader main lobe tends to merge closely situated sources.

The acoustic characteristics of the Vue glasses diverge markedly from the sound leakage typically associated with headphones, which often emanates from more discrete and

closely spaced monopole-like sources. This variance poses significant challenges to the conventional assumptions used in constructing steering vectors for beamforming applications, potentially introducing inaccuracies in measurements, especially for devices with larger source areas. Such discrepancies are particularly evident during the verification process with loudspeakers, as elaborated in chapter 6.

According to our understanding of the steering vector, accurate sound field reconstruction via beamforming is ostensibly confined to scenarios that can be approximated as aggregations of multiple monopoles. Therefore, when these monopoles are situated in proximity to each other—closer than a given beamformer’s maximum resolution—the reconstructed sound field is likely to be imprecise. Given that the MUSIC algorithm typically exhibits a narrower main lobe compared to the Functional beamformer, it inherently possesses superior resolution for distinguishing between closely spaced monopoles. Consequently, MUSIC is capable of reconstructing sound fields composed of multiple closely situated monopoles, capturing acoustic features that the Functional Beamformer might fail to differentiate.

So, at lower frequencies, such as 2000Hz, 4000Hz, and 6000Hz, the Music Beamformer demonstrates more pronounced directivity than the Functional Beamformer, indicating a higher sensitivity to detected source variations on the surface. Conversely, while the Functional Beamformer exhibits a similar trend to the Music Beamformer at these frequencies, it struggles to keep pace with rapid directional changes, suggesting that the Music Beamformer, post-correction, may be better suited to accommodate surface sound pressure fluctuations.

6 Reverse sound field simulation by BEM

From previous sections, we utilize beamforming to project collected sound pressure recordings from an array to various field points, thereby creating a source map. This map indicates the potential distribution of sound sources, also estimates their strengths, and more over it is laying onto the surface of scanned mesh geometries.

On another hand, the Boundary Element Method (BEM) serves as a tool for simulation, grounded in the principles of the Helmholtz Integral Equation (HIE). BEM simulates sound pressure levels in a field, given sufficient surface information from the mesh geometry being analyzed. This method traditionally requires input on two out of three key aspects: sound pressure, volume velocity, or boundary conditions, to effectively model sound radiation from any given closed surfaces.

There are already some attempt to combine the BEM and Beamforming, which is the Inverse Boundary Element Method [35], which is a method that use the HIE to reconstruct the sound field from the far field recordings. This method has a proven precision, and we also have the ability to scan through the object surface geometry. However, since those more simple beamformers can achieve fairly accurate surface sound pressure reconstruction, there might be no need to imply this advanced method.

Despite these considerations, our interest in understanding the radiation characteristics of objects propels us further. By analyzing the surface sound pressure, we can provide the necessary information for BEM operations, facilitating the calculation of sound pressure in the far field through HIE. This development forms the core reason for this section of our research, aiming to employ the existing source strength map and scanned surface mesh to calculate the field pressure accurately.

6.1 Theory

The acoustic Helmholtz Integral Equation (HIE) in BEM most of the time was expressed as the following form, if not consider the incident wave,

$$C(P)p(P) = \int_S \left[p(Q) \frac{\partial G(P, Q)}{\partial n} + jk\rho c v_n(Q) G(P, Q) \right] dS \quad (6.1)$$

it is describing that the field pressure $p(P)$ is a result of the integration of surface pressure $P(Q)$ and volume velocity $v_m(Q)$ on the surface, Transferred by the Green's function G , which $G(P, Q) = \frac{e^{-jkR}}{R}$

However, in our application, we do not need to have the volume velocity of each discrete surface, another form of Green's function can be used instead

$$P(\vec{x}) = \int_S \left[G(\vec{x}, \vec{x}') \frac{\partial P}{\partial n}(\vec{x}') - P(\vec{x}') \frac{\partial G}{\partial n}(\vec{x}, \vec{x}') \right] dS' \quad (6.2)$$

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}' .

From this equation, it is clear that we need to estimate the pressure gradient and normal derivative of Green's function.

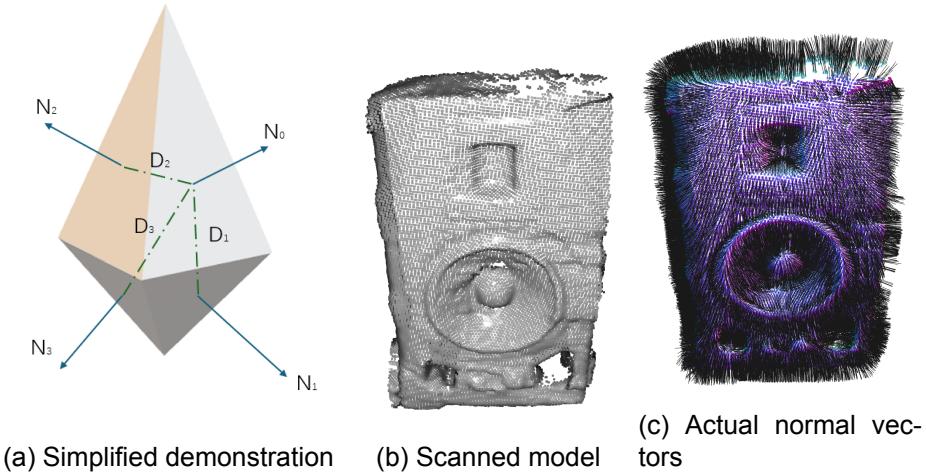


Figure 6.1: Gradient differential calculation relative to normal vectors in acoustic modeling. (a) Simplified representation with N_i as the normal vector and D_i denoting distance from Mesh 0 to the i th mesh, illustrating the basic principle of gradient calculation. (b) Actual scanned loudspeaker model used for gradient analysis. (c) Loudspeaker model with normal vectors, essential for calculating the sound field's interaction with the object surface through Green's function and its derivatives

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

The Green's function in 3d is

$$\nabla^2 G + k^2 G = -4\pi\delta(Q - P) \quad (6.3)$$

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

$$G = G(P, Q) = G(R), \quad R = |P - Q| \quad (6.4)$$

Expressions of the Green's function:

$$\text{In 3-D: } G(P, Q) = \frac{e^{-jkR}}{R} \quad (e^{iwt} \text{ convention is used}) \quad (6.5)$$

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

$$\frac{\partial G}{\partial n} = \left(-\frac{e^{-jkR}}{R^2} - jk \frac{e^{-jkR}}{R} \right) \left(\frac{x}{R} n_x + \frac{y}{R} n_y + \frac{z}{R} n_z \right) \quad (6.6)$$

The Beamforming results provide a spatial map of acoustic pressure P at the discrete mesh 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, also demonstrated as in fig. 6.1a

Identification of Nearest Neighbors

Utilize a KD-tree algorithm to identify the four nearest neighboring mesh points [36] to the point of interest. We had utilized the Python library [scipy.spatial.cKDTree][37] for this approach. The KD-tree efficiently searches for the nearest neighbors in a multidimensional space which will save the computational resources. This task can be also done by directly sorting and extracting but due to the size of our mesh matrix, it will takes far more computational times.

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

Gradient Estimation

The pressure gradient at the point of interest ∇P_0 can be approximated as an average rate of change of pressure concerning the distances to the neighboring points. Mathematically, this can be expressed as:

$$\nabla P_0 = \left(\frac{\partial P_0}{\partial x_0}, \frac{\partial P_0}{\partial y_0}, \frac{\partial P_0}{\partial z_0} \right) \approx \frac{\sum_{i=1}^3 (P_0 - P_i)}{D_i} \quad (6.7)$$

Once the pressure gradient ∇P_0 at the mesh point is estimated, the next step is to calculate the derivative of pressure concerning 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:

$$\frac{\partial P}{\partial n}(\vec{x}) = \nabla P_0 \cdot n_0 \quad (6.8)$$

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

$$P(x) = \sum_{i=0}^{i=N} \left[G_i \cdot \dot{P}_i - \dot{G}_i \cdot P_i \right] S_i \quad (6.9)$$

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 points.

6.2 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 Klipper Near-field Scanner (NFS): The NFS, a robotic measurement system capable of rotating the object and measuring from various angles (as referenced in fig. 6.2a), 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 Klipper NFS system in our own lab, so the data is from a

review to the Adam A5x in the "audiosciencereview.com" [29], 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.

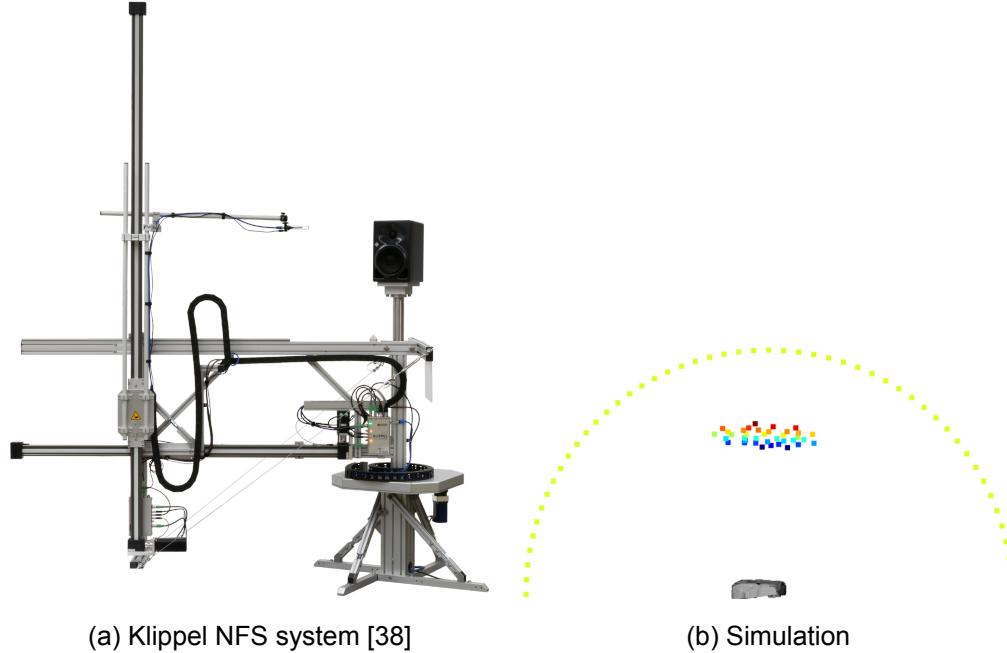


Figure 6.2: (a) The Klippel NFS measurement system. (b) The mesh model of Loudspeaker and the circular grid was colored green for simulation to replicate the similar grid set in the Klippel measurement. The rainbow-colored points are the location of the microphone array, which has the known sound field recordings

6.2.1 Significant eigenvalues and eigenvector

Building upon the discretized Boundary Element Method (BEM) framework outlined in equation eq. (6.9), our analysis progresses towards validating the accuracy of our simulations by directly comparing them with actual measurements. A critical step in this process involves back-projecting the simulated sound pressure to the locations of the individual microphones within our array. This procedure is pivotal for facilitating a direct comparison between the simulated sound pressures and those previously recorded. Ideally, if our simulations are highly accurate, there should be a close alignment between these two datasets. The recorded sound pressure is understood to be a composite of room reflections, ambient noise, and the direct sound pressure from the source, which our simulations aim to replicate.

Given the complexity introduced by our setup, which includes an array of 36 microphones, conducting a straightforward comparison poses significant challenges due to the high dimensionality of the data.

To render this analysis more manageable and meaningful, it is imperative to employ dimensionality reduction techniques. A particularly effective approach is the eigenvalue decomposition of the cross-spectrum matrix (CSM) derived from both the simulated and

recorded sound pressures. This decomposition process isolates eigenvalues and eigenvectors that represent the primary modes of sound interaction within the environment. By comparing the significant eigenvalues obtained from both the simulated and recorded configurations, we can simplify our analysis to concentrate on these dominant acoustic 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 an illustrative example, we examine the eigenvalue decomposition of the cross-spectrum matrix (CSM) for the ADAM A5x, comparing simulation results with actual measurements across different frequencies, as depicted in fig. 6.3. This analysis reveals a notable distinction between the actual measurements and the simulations. In the actual measurements, the CSM encompasses sound contributions from both the primary source (the loudspeaker) and ambient noise, resulting in a spectrum of eigenvalues that reflect the complexity of the acoustic environment. This complexity is manifested in multiple components within the CSM, indicative of the interplay between source sound and noise.

Conversely, the simulation, which inherently lacks noise contribution, yields a CSM characterized by a single dominant eigenvalue. This stark difference underscores the purity of the simulated sound field, devoid of the environmental noise that accompanies actual measurements. The presence of a singular significant component in the simulated CSM highlights the focused nature of the simulation, isolating the sound radiation from the loudspeaker without the confounding effects of ambient noise.

By isolating the significant eigenvalues from the CSM, we effectively narrow the scope of our comparison to the primary contributors of the sound field. This approach, while still susceptible to the influence of ambient noise—especially when external noise dominates—provides a focused framework for analysis. In scenarios where the primary sound source, rather than ambient noise, is the most significant contributor, this method proves particularly effective.

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.

6.3 Calculation Settings

We assess the performance of four distinct beamforming methods, as outlined in chapter 5: Delay-and-Sum (DAS), Multiple Signal Classification (MUSIC), Functional Beamforming, and CleanSC. These algorithms have shown varying levels of proficiency in source localization. Our analysis further explores their effectiveness in reconstructing the sound field based on previously obtained sound pressure results, thereby evaluating their suitability for the task of sound field reconstruction. The computational analysis is facilitated through a dedicated GitHub project [3d_beamforming] made for this thesis [33].

6.4 Results

The resulting simulation and real recording csm after the significant eigenvalue extraction as shown in fig. 6.4. This plot consists of 4 beamformer's surface reconstruction and the resulting simulation to the Array microphone position. This result shows case that a very different simulation results, in comparison with all the other simulation results in chapter 5.

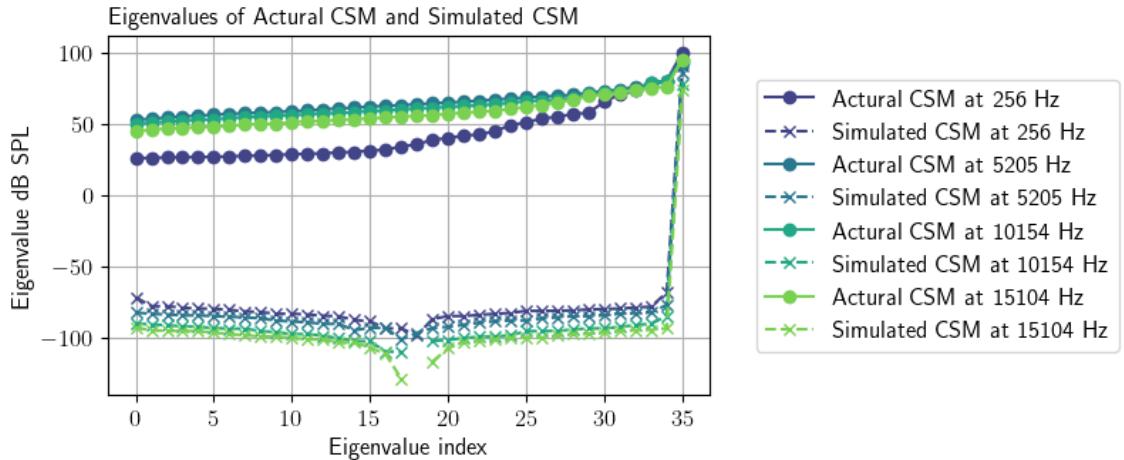


Figure 6.3: Eigenvalue comparison of the actual recording and simulate sound pressure at microphone array's position, under different frequency. This simulation is conducted from the sound pressure at the surface which is reconstructed by Beamforming Functional

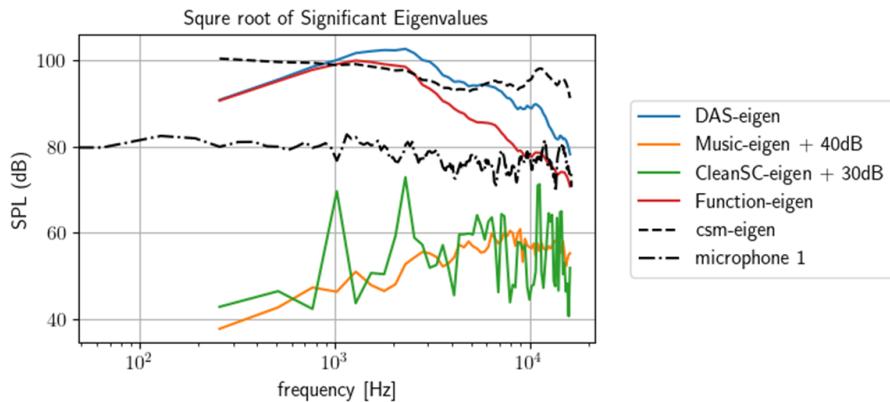


Figure 6.4: Significant eigenvalues comparison between different algorithms and the microphone array recordings. Eigenvalues are extracted from cross spectrum matrix so the square root of its eigenvalue magnitude will be in the unit in dB SPL

6.4.1 level correction

While beamforming is adept at source localization, its capacity to precisely reconstruct the sound field and accurately reflect source strength is limited. To address this, level correction is essential, aligning the simulated outcomes more closely with actual measurements. This correction is facilitated by comparing simulated eigenvalues with those derived from recorded data. This approach mirrors the equalization process described in section 4.4, with recorded cross-spectrum matrix (CSM) eigenvalues serving as the reference. The discrepancy between these eigenvalues and those from the simulation informs the correction filter applied to each beamformer's simulated results.

Given that this analysis does not incorporate time-domain data, constructing a Finite Impulse Response (FIR) filter is unnecessary. Instead, a compensation frequency response applied to the simulation suffices. As described in section 5.4.3, this correction uniformly adjusts the simulation levels in all directions. This uniformity results from using eigenvalues as the reference, as a result, it will apply a constant level adjustment in dB across the simulation grid.

6.4.2 Radiation directivity pattern

The analysis of radiation directivity patterns for headphone leakage and the Vue glasses' bone conduction vibrator radiation was previously introduced in section 5.4.3. This section extends our investigation to the Adam A5x loudspeaker, a single-box configuration, as illustrated in fig. 6.5. The figure juxtaposes simulation results with actual directivity pattern measurements performed using the Klippel NFS system [39]. The NFS measurements demonstrate remarkable consistency across all frequencies and most directions. The on-axis response is consistently maintained at 80dB, with observed side attenuation (around ± 90 degrees) ranging from a minimum of 10dB at lower frequencies (around 1000Hz) to a maximum of 20dB at higher frequencies (around 10000Hz). This measured response transitions smoothly from on-axis to side attenuation without any irregular fluctuations.

However, the simulation outcomes reveal discrepancies between the different beamforming techniques—Music, DAS, CleanSC, and Functional—and the NFS measurements across the frequency spectrum. Particularly at frequencies above 4000Hz, the simulation results diverge not only from the NFS measurements but also among the beamformers themselves. Interestingly, at lower frequencies of 1000Hz and 2000Hz, the results from all beamformers, except CleanSC, show similar responses in most directions, though they still do not align with the actual NFS measurements.

Given the loudspeaker's distinct vibration characteristics compared to a monopole source, such response curves are understandable. The simulation, which relies on surface information to model the sound field at the boundary, may not accurately represent the loudspeaker's behavior due to its reliance on a monopole assumption. This assumption, when misaligned with the actual complex radiation behavior of the loudspeaker, can introduce errors. Specifically, the loudspeaker's omnidirectional wavefront at lower frequencies resembles that of a vibrating monopole at the woofer's center. At higher frequencies, the tweeter becomes the dominant source, producing a more directive wavefront that, to the array, may appear as emanating from multiple monopoles or higher-order spherical waves. Thus, the reconstructed sound field might resemble a configuration of multiple monopoles, particularly at higher frequencies, thus influencing the directivity pattern simulation to reflect a higher-order spherical wave model.

Further insights into the sound pressure level map, reconstructed by the beamformers, are provided in appendix A.1. The findings suggest that the fidelity of surface reconstruction—and, by extension, the depiction of the actual sound field—is significantly influenced by the bandwidth of the main lobe. An inaccurate model of this bandwidth can lead to a complete divergence from the real scenario, underscoring the necessity for reconstructing the sound field accurately.

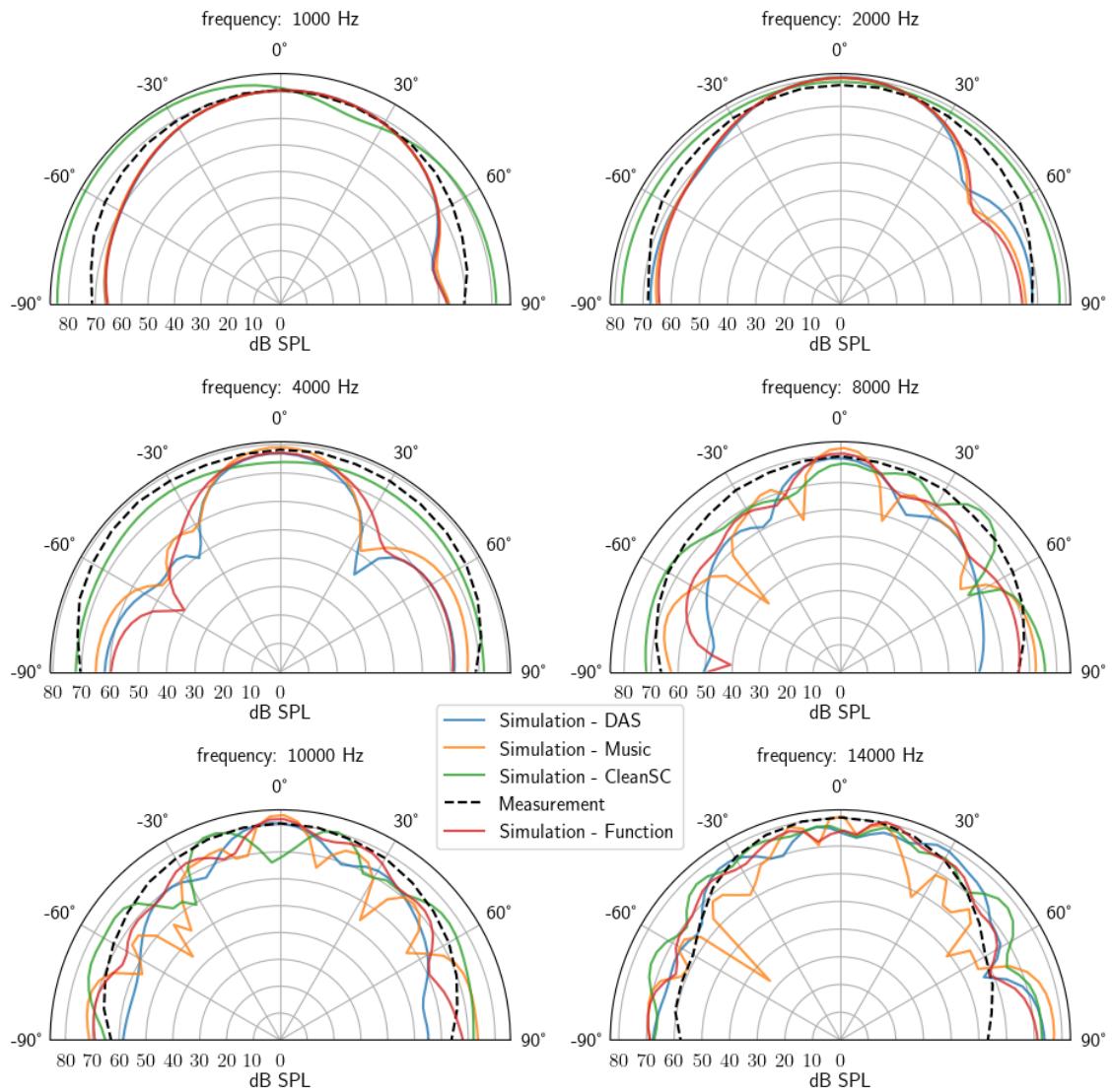


Figure 6.5: Directivity Patterns of Simulated Surface Pressure Reconstruction and Measured Results Across Frequencies

7 Conclusion

In conclusion, this thesis has presented a comprehensive exploration of the sound leakage problem in personal audio devices, emphasizing the complexities and multifaceted nature of this issue. Our findings demonstrate that 3D beamforming is possible to achieve a high degree of precision in localizing sound leakage sources. This methodology not only enhances the accuracy of sound leakage detection but also provides a foundation for understanding the mechanism of the leakage. Furthermore, the comparative analysis of beamforming algorithms revealed that certain techniques, previously underutilized in this context, offer substantial improvements in detecting and analyzing sound leakage patterns.

Beamformers

Our analysis unearthed nuanced insights into the capabilities and limitations of four beamformer algorithms. Notably, the MUSIC algorithm emerged as a standout, demonstrating exceptional accuracy and super-resolution performance at low frequencies. The enhanced steering vector accuracy achieved through our refined scanning and alignment process plays a crucial role in unlocking the full potential of the MUSIC algorithm. By providing a more precise representation of the device's surface geometry, we enable the algorithm to perform nearly to its best, ensuring to make it achieves a great low-frequency super-resolution ability.

Conversely, the Functional Beamformer was identified as the most robust across a broader range of frequencies, offering consistency and reliability in its results. This robustness is crucial in the practical application of our methodology, ensuring that our approach remains effective across diverse scenarios and device types. Additionally, the CleanSC Beamformer exhibited a capability to localize multiple incoherent sources and generate a shifting frequency response, a feature that could prove invaluable in refining our understanding of sound leakage pathways.

These distinctions among the beamforming algorithms highlight the complexity of sound leakage detection and the importance of selecting the appropriate tool for specific aspects of the problem. As we look to the future, the insights gained from this comparative analysis will guide the further development of our detection methodologies. Emphasizing the optimization of these algorithms for specific use cases, such as low-frequency sound leakage, will be paramount. Moreover, the exploration of how these algorithms can be combined or modified to leverage their individual strengths offers a promising avenue for enhancing the accuracy and applicability of our approach.

Experiments

An integral assumption underpinning our beamforming approach was the monopole model of sound radiation, a premise that proved effective in analyzing traditional headphones. However, the investigation into the VUE smart glasses and Adam A5x loudspeaker revealed limitations to this assumption. The unique interaction between bone conduction and structural vibrations in the VOE glass leads to a sound radiation pattern that diverges significantly from a monopole source. This deviation presents a challenge for our beamforming techniques, particularly in accurately reconstructing the sound field on the device's surface. Our analysis indicates that among the various beamforming algorithms evaluated, the MUSIC algorithm emerged as the most promising in addressing this complexity. Yet, this finding underscores a critical gap in our methodology—the absence of a robust verification mechanism to confirm the superiority of MUSIC in such nu-

anced scenarios. Identifying and developing methods to validate these analytical results is paramount, marking a vital area for future research.

As for the Headphone leakage, it involving two distinct models of Jabra headphones — differing primarily in size. The larger model exhibited more pronounced structural vibrations, while the smaller variant was more susceptible to high-frequency leakage. This differentiation underscores the complexity of sound leakage phenomena and the factors influencing it. Also, contrary to initial expectations, the larger model exhibited substantially more sound leakage. This phenomenon was not merely a function of its size but was intricately linked to increased structural vibrations and the potential for gaps between the cushion and the ear.

A particularly revealing aspect of our investigation was the counterintuitive effect observed when taped a porous foam to one headphone in an attempt to reduce sound leakage. Contrary to expectations, this intervention resulted in an enhancement of the leakage, a phenomenon we attributed to the occlusion effect. This effect, sometimes overlooked in the measurement, highlights the intricate balance between sealing to prevent sound escape and the unintended consequences such sealing may have on sound propagation.

Simulation of the radiation directivity pattern

Leveraging the surface sound pressure level by beamformer and geometry obtained through our 3d scanner, we embarked on an attempt to reconstruct the sound field. This endeavor aimed to derive the radiation directivity pattern, a critical information necessary for the definition and analysis of sound radiation. Our innovative approach holds the promise of drastically reducing the time and resources typically required for such measurements, offering a rapid simulation alternative. However, this simulation needs a fully validated assessment and further investigation, it also needs a further understanding of the performance and the hidden mechanism.

Since this study concentrates on beamforming estimate the sound pressure at the surface of the object, It is essential to acknowledge that various beamforming algorithms produce different source distributions and strength. As this technologe essentially is not designed for a sound field reconstruction but for the direction of arrival detection. It's ability is limited by the point source assumption. While Near-Field Acoustic Holography (NAH) may offer enhanced precision for sound field estimation, it is outside of the scope of this research. Instead, our objective is to evaluate the efficacy of the beamforming techniques in reconstructing surface pressure.

Bibliography

- [1] Rebecca. *What is Headphone Sound Leakage and How to Fix It?* <https://blog.taotronics.com/headphones/headphone-sound-leakage/>. 2020.
- [2] Jordan Jamieson-Mane. *Sound Leakage in Headphones: What It Is and What You Can Do About It | Headphonesty*. <https://www.headphonesty.com/2021/04/sound-leakage/>.
- [3] Stefan Stenfelt and Sabine Reinfeldt. "A model of the occlusion effect with bone-conducted stimulation". In: *International journal of audiology* 46.10 (2007), pp. 595–608.
- [4] Fanyu Meng, Anbo Yu, and Dani Fernandez. "Sound leakage investigation of ANC headphones using particle velocity sensors". In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Vol. 261. 6. Institute of Noise Control Engineering. 2020, pp. 698–708.
- [5] Jose J Lopez, Sergio Martinez-Sanchez, and Pablo Gutierrez-Parera. "Array processing for echo cancellation in the measurement of Head-Related Transfer Functions". In: *Euronoise: Crete, Greece* (2018).
- [6] C Renato C Nakagawa, Sven Nordholm, and Wei-Yong Yan. "Analysis of two microphone method for feedback cancellation". In: *IEEE Signal Processing Letters* 22.1 (2014), pp. 35–39.
- [7] Todd Welti. "Improved measurement of leakage effects for circum-aural and supra-aural headphones". In: *Audio Engineering Society Convention 138*. Audio Engineering Society. 2015.
- [8] Robert Werner. *Measuring 3D Directivity of Microphones and Radiated Sound in Headsets and Hearables*. 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).
- [9] 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>.
- [10] *Vehicle Interior Measurements: The Acoustic Camera as an Effective Tool for the NVH Analysis of Vehicles*. https://www.gfaitech.com/fileadmin/gfaitech/documents/application_notes/3d-vehicles-interior-measurement-the-acoustic-camera-as--tool-for-nvh-analysis-2022-11.pdf.
- [11] Andy Meyer and Dirk Döbler. „Noise source localization within a car interior using 3D-microphone arrays “. In: *Proceedings of the BeBeC* (2006), pp. 1–7.
- [12] Dirk Döbler, Jörg Ocker, and Christof Puhle. "On 3D-beamforming in the wind tunnel". In: *6th Berlin Beamforming Conference–6th BeBeC*. 2016.
- [13] 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.
- [14] Ralph Schmidt. "Multiple emitter location and signal parameter estimation". In: *IEEE transactions on antennas and propagation* 34.3 (1986), pp. 276–280.
- [15] Pieter Sijtsma. "CLEAN based on spatial source coherence". In: *International journal of aeroacoustics* 6.4 (2007), pp. 357–374.

- [16] Robert P Dougherty. "Functional beamforming". In: *5th Berlin beamforming conference*. GFaI, eV Berlin. 2014, pp. 19–20.
- [17] Prem K Kythe. *An introduction to boundary element methods*. CRC press, 2020.
- [18] 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>.
- [19] *Technical Review: No. 1 2004 Beamforming (BV0056)*. <https://www.bksv.com/media/doc/bv0056.pdf>. Accessed: 2024-03-19. 2004.
- [20] *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.
- [21] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).
- [22] Will Schroeder, Ken Martin, and Bill Lorensen. *The Visualization Toolkit (4th ed.)*. Kitware, 2006. ISBN: 978-1-930934-19-1.
- [23] 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.
- [24] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. "Open3D: A Modern Library for 3D Data Processing". In: *arXiv:1801.09847* (2018).
- [25] *Surface reconstruction - Open3D 0.18.0 documentation*. https://www.open3d.org/docs/release/tutorial/geometry/surface_reconstruction.html. Accessed: 2024-03-31. 2024.
- [26] F. Bernardini et al. "The ball-pivoting algorithm for surface reconstruction". In: *IEEE Transactions on Visualization and Computer Graphics* 5.4 (1999), pp. 349–359. DOI: 10.1109/2945.817351.
- [27] *Jabra Evolve2 75*. <https://www.jabra.dk/business/office-headsets/jabra-evolve/jabra-evolve2-75/buy#27599-999-999>. Accessed: 2024-03-24. 2024.
- [28] *Jabra Evolve2 65 – designet til at sikre din fleksibilitet. Arbejd alle steder med overlegen trådløs lyd*. <https://www.jabra.dk/business/office-headsets/jabra-evolve/jabra-evolve2-65>. Accessed: 2024-03-24. 2024.
- [29] *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.
- [30] *Your Everyday Smart Glasses*. <https://vueglasses.com/>. 2024.
- [31] *High-Frequency Head and Torso Simulator | Type 5128 | BK*. <https://www.bksv.com/en/transducers/simulators/head-and-torso/hats-type-5128>. Accessed: 2024-03-30. 2024.
- [32] Jussi Rämö. "Equalization Techniques for Headphone Listening". In: (). URL: <https://aaltdoc.aalto.fi/bitstreams/a5681c9a-ca63-49e5-84cd-648ecc87352a/download>.
- [33] *1ieng2en/3d_beamforming*. https://github.com/1ieng2en/3d_beamforming. 2024.
- [34] *Vue: Your Everyday Smart Glasses by Vue — Kickstarter*. <https://www.kickstarter.com/projects/vue/vue-your-everyday-smart-glasses>. Accessed: 2024-03-29. 2024.
- [35] Andreas Schuhmacher et al. "Sound source reconstruction using inverse boundary element calculations". In: *The Journal of the Acoustical Society of America* 113.1 (2003), pp. 114–127.
- [36] Piotr Kraus and Witold Dzwinel. "Nearest neighbor search by using Partial KD-tree method". In: *Theor. Appl. Genet* 20 (2008), pp. 149–165.

- [37] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.
- [38] KLIPPEL GmbH. *Near Field Scanner System (NFS)*. <https://www.klippel.de/products/rd-system/modules/nfs-near-field-scanner.html>. Accessed: 2024-03-13. 2024.
- [39] Jordan Jamieson-Mane. *Sound Leakage in Headphones: What It Is and What You Can Do About It | Headphonesty*. <https://www.headphonesty.com/2021/04/sound-leakage/>. Accessed: 2024-03-13. 2024.

A Appendix

A.1 Adam A5x additional Beamforming Plots

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

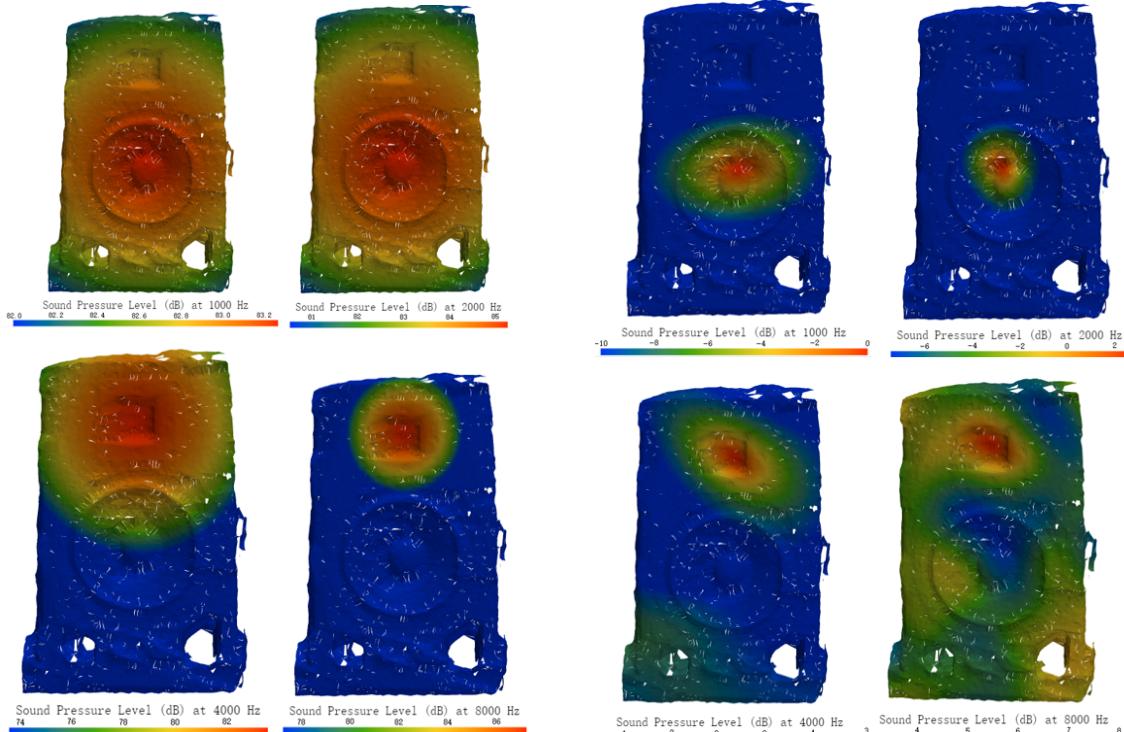


Figure A.1: DAS

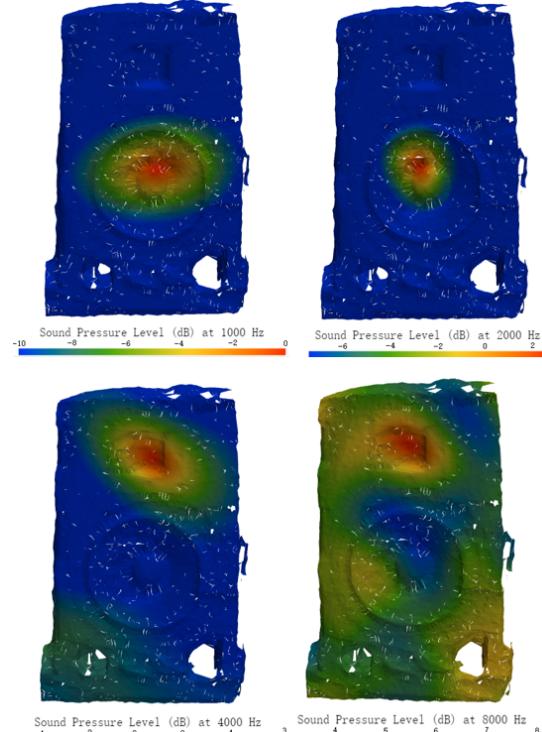
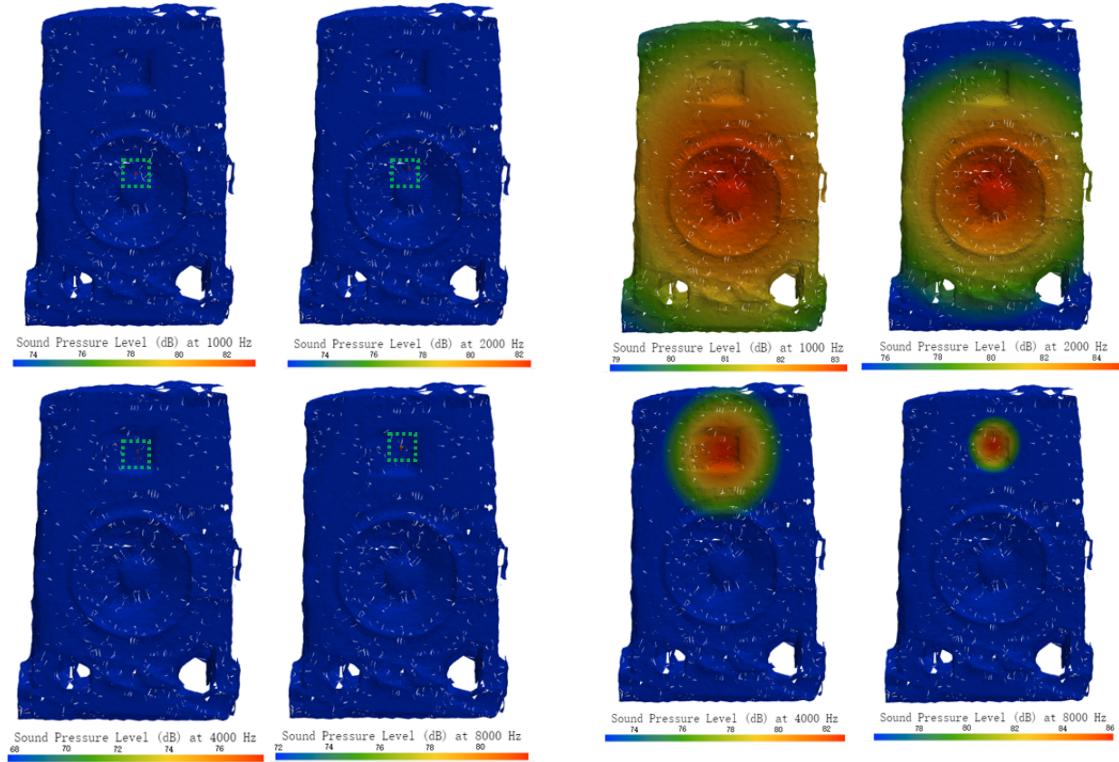


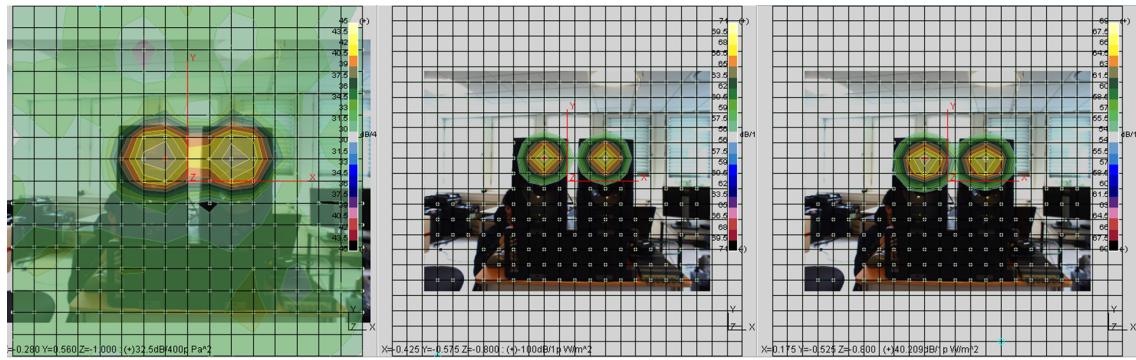
Figure A.2: Music

Figure A.3: Beamforming reconstructed sound pressure level to the surface of the single Adam A5x box. It's worth noting that the sound pressure level in the Music Beamforming plot is not representing the actual level, but only the reciprocal level of reconstructed noise map strength - the pseudo sound pressure level for Music



A.2 Comparison test for NSI and 3d beamforming on Adam A5x

A.2.1 From NSI Acoustic array postprocessing



A.2.2 From 3d beamforming

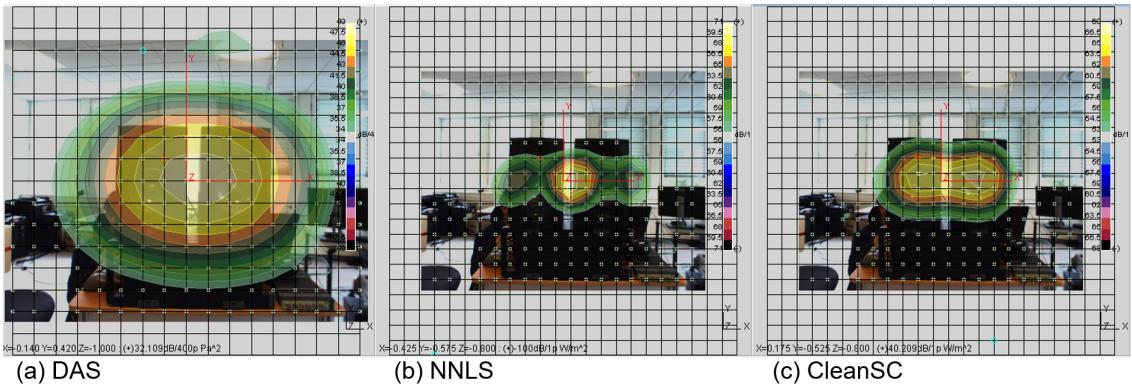


Figure A.8: NSI-Array acoustic postprocessing result under 2500Hz, 15dB dynamic range

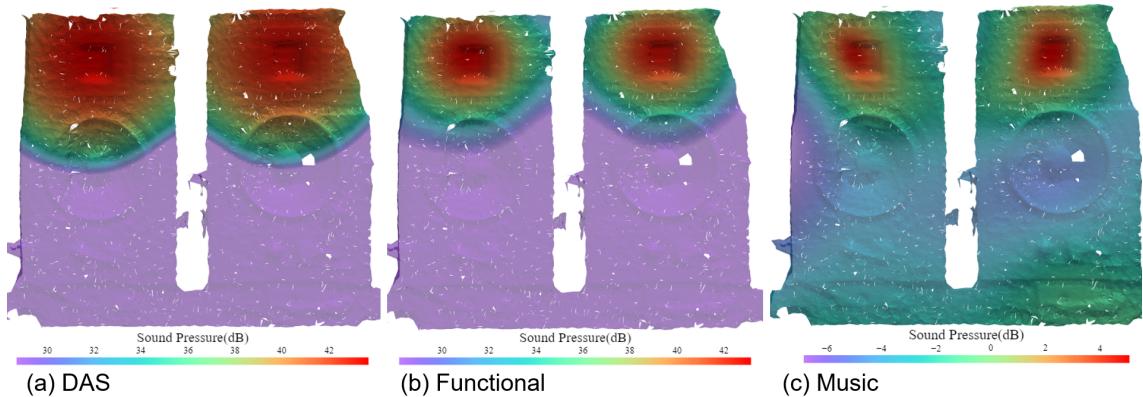


Figure A.9: 3d beamforming result under 5000Hz, 15dB dynamic range.

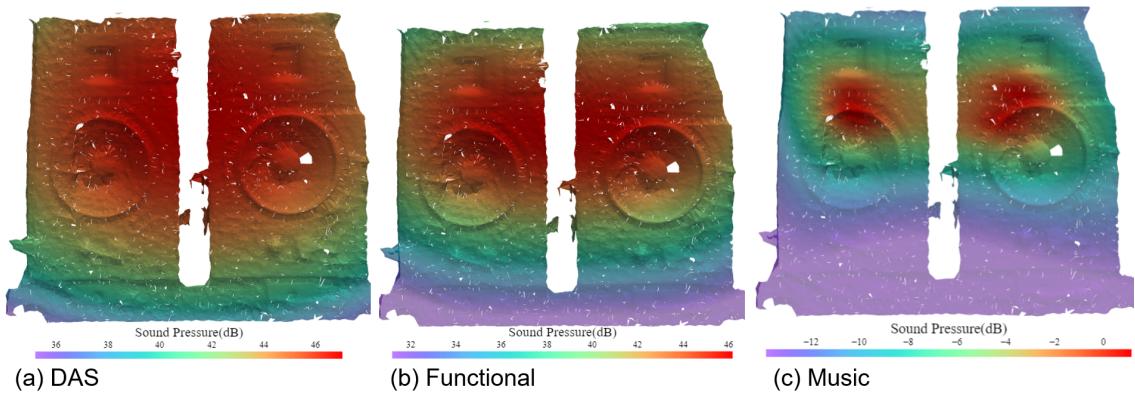


Figure A.10: 3d beamforming result under 2500Hz, 15dB dynamic range.

A.3 Additional HP2 functional beamforming map

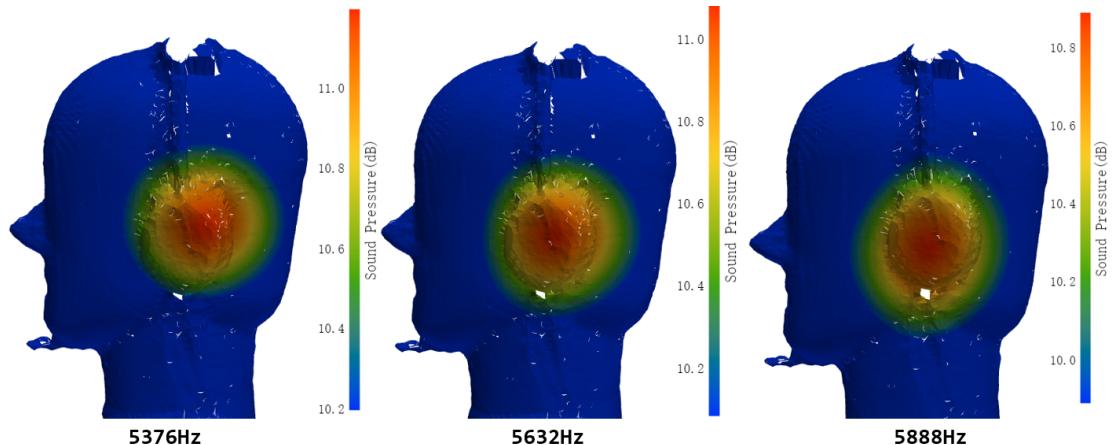


Figure A.11: Functional Beamforming result for the HP2 when foam placed at the back, The dynamic range is limited to be 1dB.

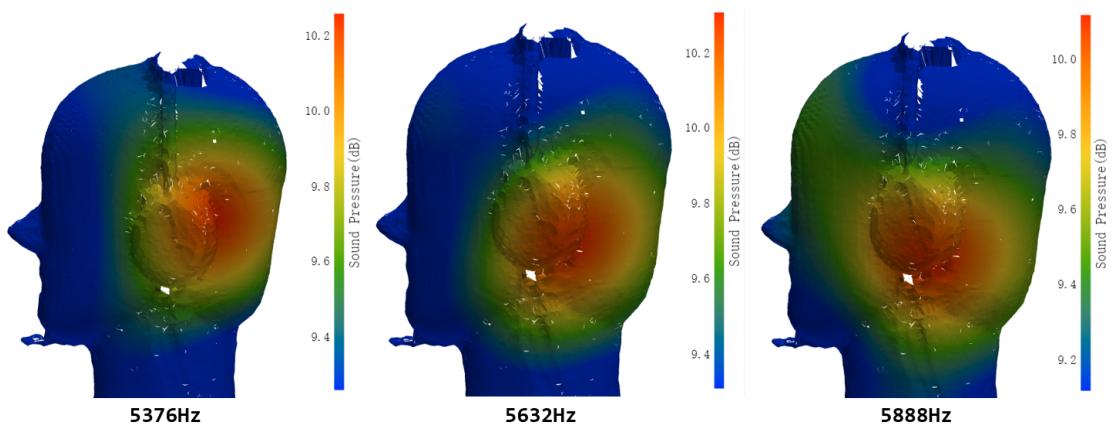


Figure A.12: Functional Beamforming results for the HP2 when the HP2 normally wore without foam, The dynamic range is limited to be 1dB.

A.4 Radiation Pattern simulation for HP2

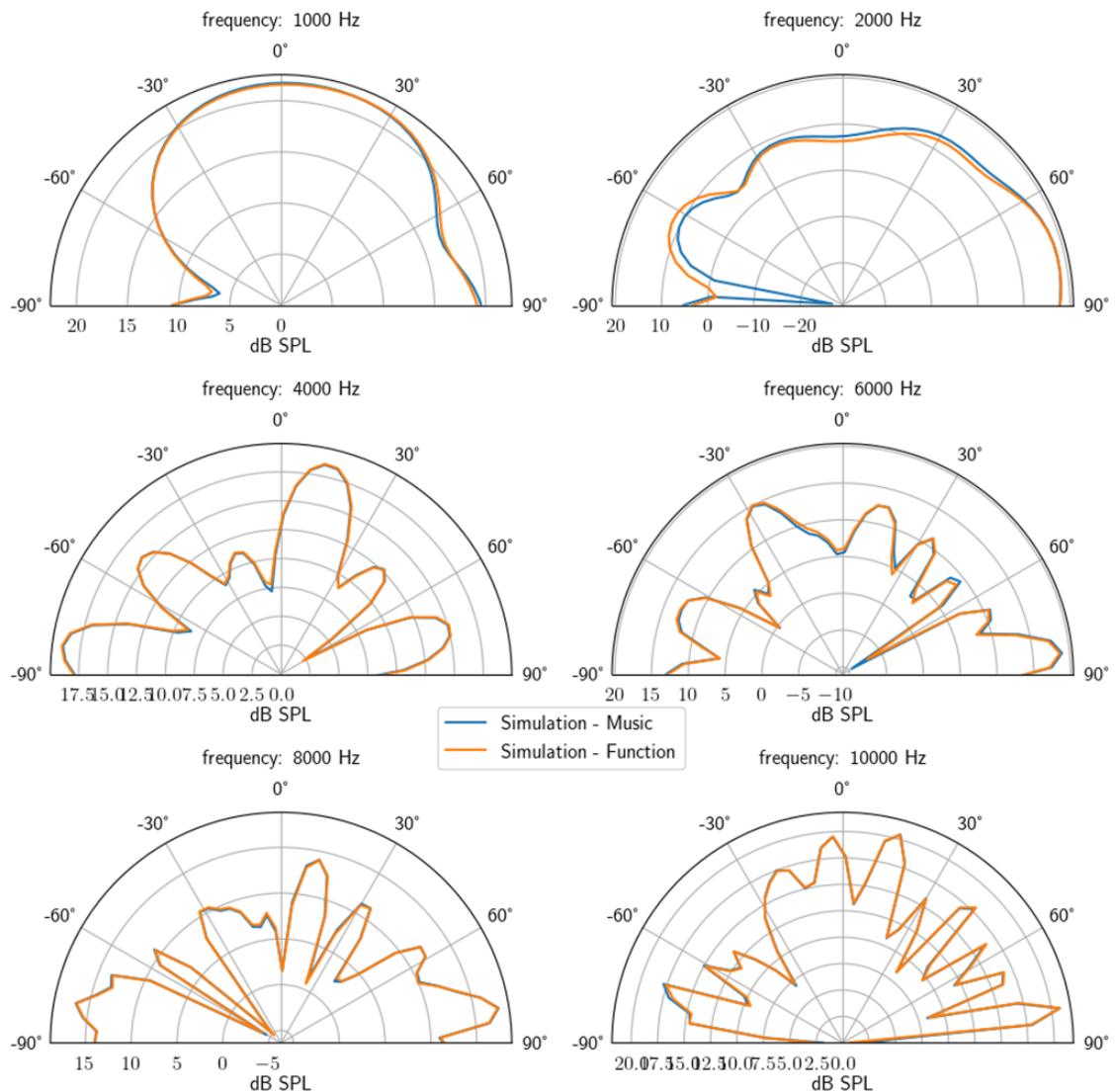


Figure A.13: Simulated sound radiation pattern when no foam is attached on HP2.

A.5 Radiation pattern simulation and corresponding eigenvalues for HP1

A.5.1 Simulation of the sound radiation pattern

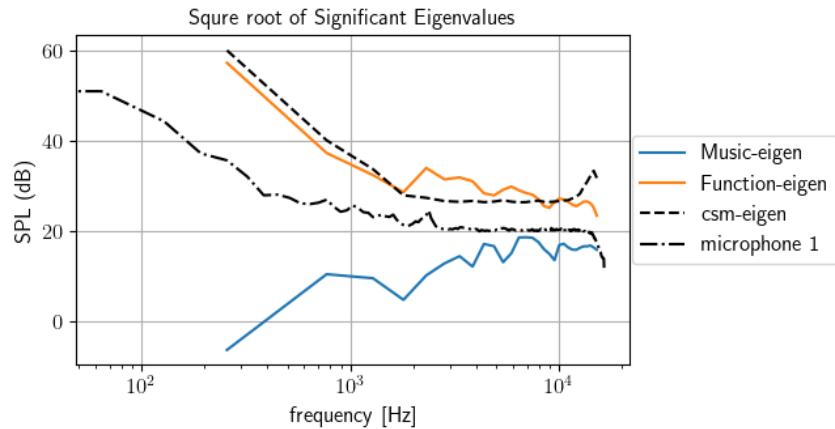


Figure A.14: Significant eigenvalues from the cross-spectrum matrix of the array by actual recording or simulation of HP1 measurement.

A.5.2 Radiation directivity pattern

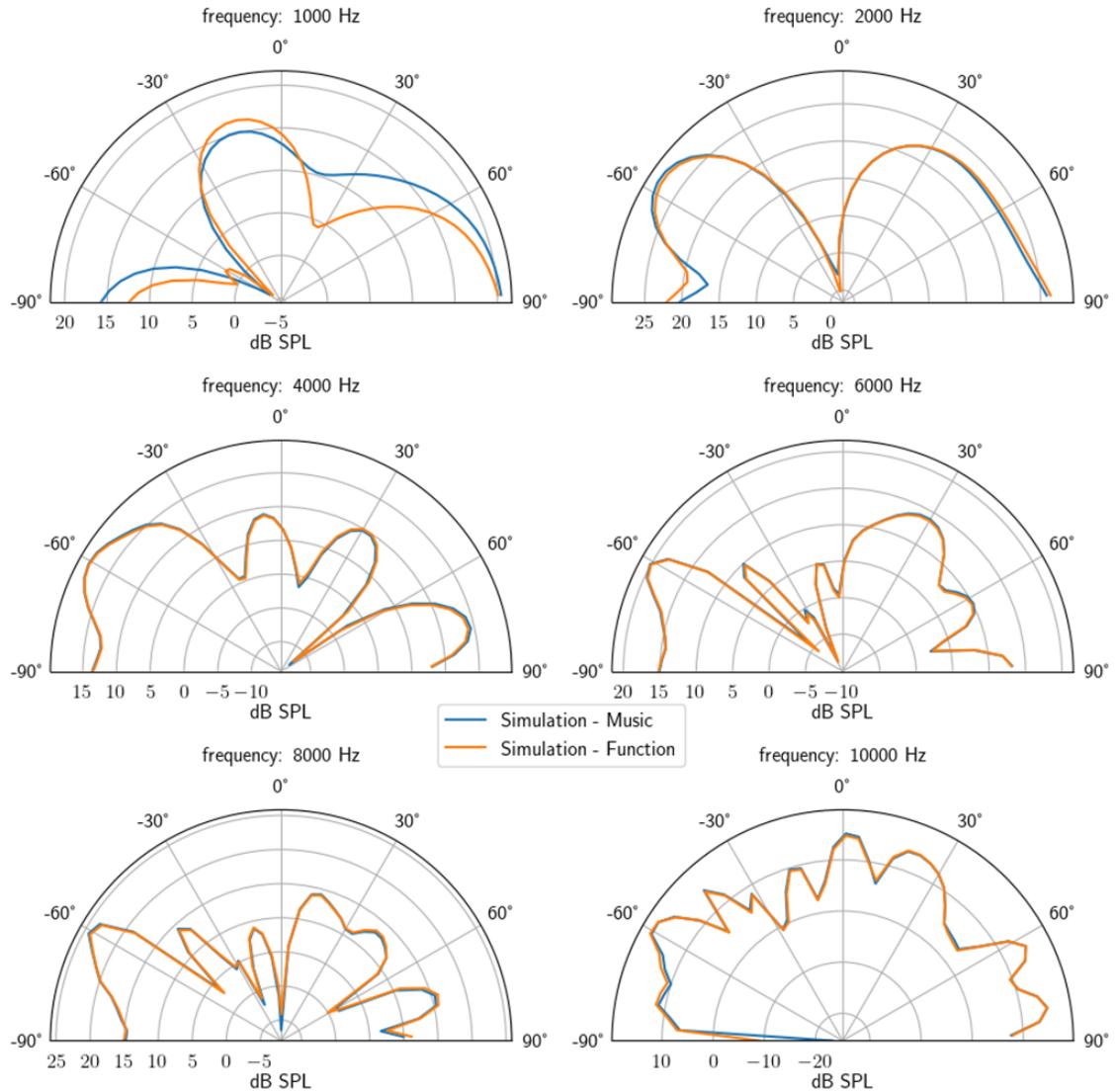


Figure A.15: Radiation directivity pattern of HP1. Both Music and Functional simulation result is equalized based on the eigenvalue at the origional array position.

A.6 Radiation pattern simulation and corresponding eigenvalues for Vue glass

A.6.1 Simulation of the sound radiation pattern

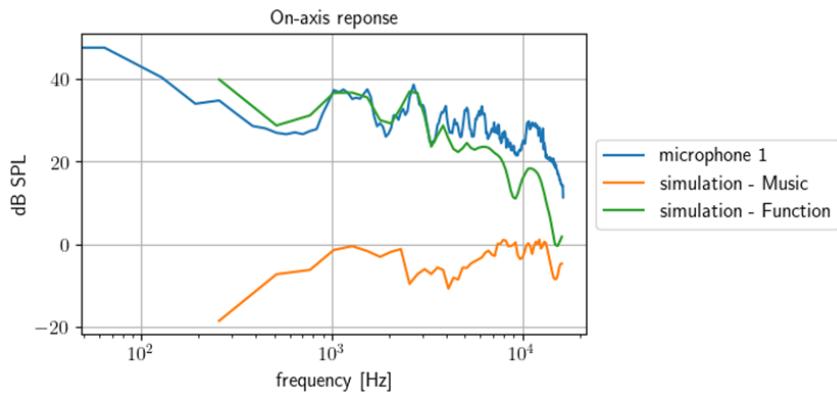


Figure A.16: Simulated and measured frequency response of Vue smart glass.

A.6.2 Radiation directivity pattern

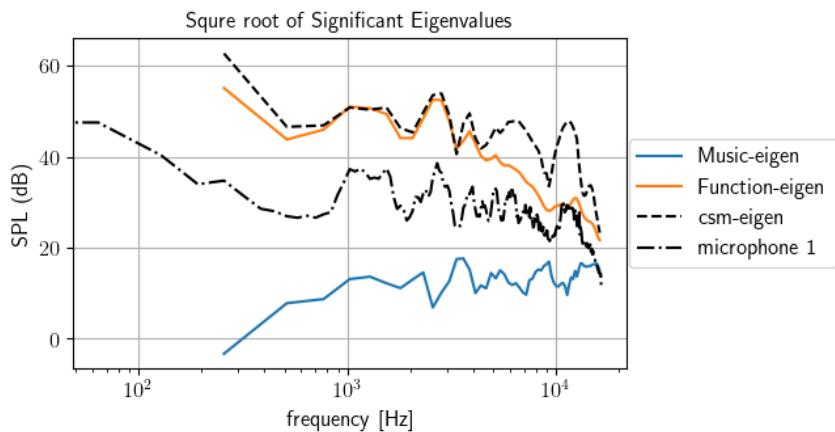


Figure A.17: Simulated and measured eigenvalue with frequency response of Vue smart glass.

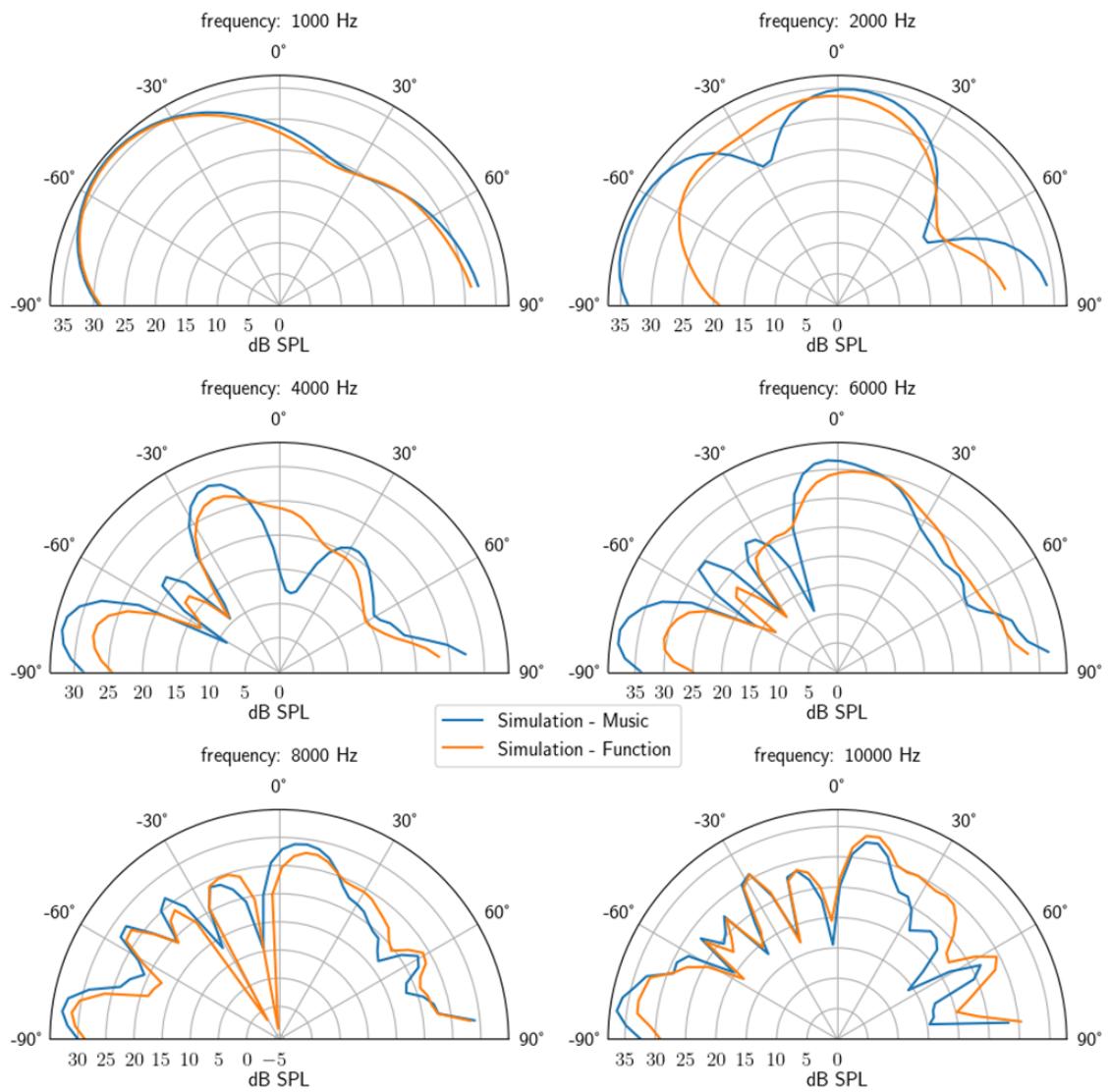


Figure A.18: Radiation directivity pattern of the Vue glass.

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