國 立 清 華 大 學

碩 士 論 文

基於多聲道模型匹配和核插值方法的全域聲場聚焦控制

Extended acoustic zone control via multichannel model matching and kernel-based interpolation approaches

系級：動力機械工程學系碩士班

組別：電機控制組

學號姓名：110033529 賴柏儒 Bo-Ru Lai

指導教授：白明憲 博士 (Dr. Mingsian R. Bai)

中 華 民 國 一 一 二 年 八 月

# 摘要

本篇論文提出基於多通道模型匹配理論(multi-channel model matching principle)和再生核插值(reproducing kernel interpolation)的全域聲場控制技術(global acoustic zone control technique)。該方法旨在實現對廣域聆聽區域的有效聲壓控制，以使在亮區和暗區之間產生明顯的對比。首先，將聲場區域控制問題定義為模型匹配問題。損失函數被定義為具有加權聲學轉移函數的連續積分，並被用於最小化受控區域內的能量匹配誤差來實現聲場控制。通過使用再生核插值方法，實現了量測控制點的廣域聲場控制。基於高斯求積的數值積分用於評估損失函數。在所提出的方法的實施中，只需要揚聲器與採樣點之間的頻率響應。為了驗證所提出的方法，使用揚聲器陣列進行了模擬和實驗。與幾個基準方法相比，所提出的演算法在亮區和暗區之間展現出優秀的聲場差距效果。

***關鍵詞 ― 再生核內插技術，全域聲場控制，聲場差距控制***

# ABSTRACT

This thesis presents a global acoustic zone control technique based on the multi-channel model matching principle and reproducing kernel interpolation. The proposed method aims to achieve effective sound pressure control over an extended listening area to create a sharp contrast between the bright and dark zones. First, the acoustic zone control problem is formulated as a model matching problem. The cost function, posed as a continuous integral with a weighted acoustic transfer function, is exploited to minimize the power of the matching error in the intended controlled zones. Sound field control in an extended neighborhood of the measured control points is accomplished by using a reproducing kernel interpolation approach. Numerical integration based on Gaussian quadrature is used to evaluate the cost function. In the implementation of the proposed method, only the frequency responses between the loudspeakers and sampled points are required. Simulations and experiments were conducted using a loudspeaker array. The proposed system demonstrated excellent acoustic contrast between the bright zone and the dark zone, outperforming several baseline approaches.

***Index Terms — reproducing kernel interpolated method, global zone control, acoustic contrast control.***

**致謝**

時光悄悄流逝，轉眼間碩士生涯也即將進入尾聲，首先感謝指導教授—白明憲教授的耐心指導，在研究上給予許多建議，如今也才能順利進行口試，這些日子也在老師身上看到了對研究的熱愛，這也勉勵我不論是學業或是未來工作甚至是生活都該保有熱忱，此外。產學合作讓我增加了實務經驗，了解到如何將理論應用於實際情境下，這些日子的研究生活可謂是收穫滿滿。

非常感謝陳榮順教授及簡仁宗教授願意於百忙之中擔任口試的委員，且撥冗閱讀並指導學生的口試論文並使其更加完整，於此獻上無限的感激。

求學生涯中特別感謝父母在我遇到困難時給予精神上的最大的支持與鼓勵，也感謝身邊的朋友暖心陪伴，最後感謝同實驗室的陳佑祥學長、許逸誠學長、孔繁傑學長、馬丞鴻學長、蔡岳峰學長、陳籥東學長、張淳喻學長及張宏維學長在學業上的幫助，得以及時解決研究上的問題，且有幸與同學林邑軒、吳仲倫、張馨予及陳思涵互相討論，一起成長，承蒙一路上支持與幫助我的貴人，由衷感謝大家。

# CONTENTS

[摘要 ii](#_Toc144296558)

[ABSTRACT iii](#_Toc144296559)

[致謝 iv](#_Toc144296560)

[CONTENTS v](#_Toc144296561)

[LIST OF FIGURES vii](#_Toc144296562)

[LIST OF TABLES x](#_Toc144296563)

[Chapter 1. INTRODUCTION 1](#_Toc144296564)

[Chapter 2. MULTI-CHANNAL MODEL MATCHING PROBLEM 5](#_Toc144296565)

[2.1. The Underdetermined Tikhonov Regularization (TIKR) Method 7](#_Toc144296566)

[2.2. Frequency-domain Underdetermined Multi-channel Inverse Filtering System With Linearly Constrained Minimum Variance (FUMIF-LCMV) Approach 9](#_Toc144296567)

[2.3. The Weighted Acoustic Transfer Function Matching (WATFM) Approach 11](#_Toc144296568)

[2.3.1. Kernel Interpolation Method 12](#_Toc144296569)

[2.3.2. The Weighted Acoustic Transfer Function Matching (WATFM) Approach 14](#_Toc144296570)

[2.3.3. Gaussian Quadrature 18](#_Toc144296571)

[Chapter 3. SIMULATIONS 21](#_Toc144296572)

[3.1. Kernel Interpolation Method 21](#_Toc144296573)

[3.2. Multi-Channel Model Matching Zone Control 26](#_Toc144296574)

[Chapter 4. EXPERIMENTS 42](#_Toc144296575)

[4.1. Kernel Interpolation Method 45](#_Toc144296576)

[4.2. Multi-Channel Model Matching Zone Control 48](#_Toc144296577)

[Chapter 5. CONCLUSIONS AND FUTURE WORK 58](#_Toc144296578)

[5.1. Conclusions 58](#_Toc144296579)

[5.2. Future Work 59](#_Toc144296580)

[REFERENCES 59](#_Toc144296581)

[Appendix A 65](#_Toc144296582)

[Optimal solution of the weighted acoustic transfer function matching approach 65](#_Toc144296583)

# LIST OF FIGURES

[Fig. 1 The model matching problem of multi-channel room response equalization. 5](#_Toc144296604)

[Fig. 2 Schematic diagram for defining the interpolation problem. 12](#_Toc144296605)

[Fig. 3 The target model is synthesized inside the target region  using *Ns* loudspeakers. 14](#_Toc144296606)

[Fig. 4 Simulation setup. 21](#_Toc144296607)

[Fig. 5 Magnitude result of kernel interpolation method. The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. 24](#_Toc144296608)

[Fig. 6 Unwrapped phase result of kernel interpolation method. The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. 24](#_Toc144296609)

[Fig. 7 Matching error of kernel interpolation method. 25](#_Toc144296610)

[Fig. 8 The average matching error. The red dots represent the measured points. 25](#_Toc144296611)

[Fig. 9 Simulation setup. 26](#_Toc144296612)

[Fig. 10 FRs in the bright zone (0°) and dark zone (). The dotted line represents the unprocessed FRs. The solid line represents the processed FRs. The blue line represents the FRs in the bright zone. The red line represents FRs in the dark zone. 29](#_Toc144296613)

[Fig. 11 Resulting FRs of the measured and interpolated control points. The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. The dotted black line spatial aliasing frequency. 33](#_Toc144296614)

[Fig. 12 Normalized magnitude matching error. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. 34](#_Toc144296615)

[Fig. 13 Mean power The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. 36](#_Toc144296616)

[Fig. 14 Beampattern of three approaches. 38](#_Toc144296617)

[Fig. 15 (a) The side view of the experiment setting. (b) The front view of the experiment setting. (c) The abridged general view of the experiment setting. 45](#_Toc144296618)

[Fig. 16 Magnitude result of kernel interpolation method. The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. 47](#_Toc144296619)

[Fig. 17 Unwrapped phase result of kernel interpolation method. The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. 47](#_Toc144296620)

[Fig. 18 Matching error of kernel interpolation method. 48](#_Toc144296621)

[Fig. 19 FRs in the bright zone (0°) and dark zone (). The dotted line represents the unprocessed FRs. The solid line represents the processed FRs. The blue line represents the FRs in the bright zone. The red line represents FRs in the dark zone. 49](#_Toc144296622)

[Fig. 20. Resulting FRs of the measured and interpolated control points. The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. The dotted black line spatial aliasing frequency. 52](#_Toc144296623)

[Fig. 21 Normalize magnitude matching error. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. 54](#_Toc144296624)

[Fig. 22 Mean power The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. 56](#_Toc144296625)

# LIST OF TABLES

[Table I. 39](#_Toc143264629)

[Table II. 41](#_Toc143264630)

[Table III. 57](#_Toc143264631)

# INTRODUCTION

Over the last two decades, there has been considerable research interest in personal audio systems with multiple sound zones, allowing listeners to enjoy different music or audio content privately within a shared physical space. The approach aims to establish two zones: a bright zone, where the desired sound is reproduced, and a dark zone with lower power levels than the bright zone. Several techniques have been developed to produce personal sound zones, including Acoustic Contrast Control (ACC) [1], Pressure Matching (PM) [2], Mode Matching (MM) [3] and multiple-input/output inverse theorem (MINT) [4]. Among these techniques, MINT, which minimizes the matching error, is the most commonly used. Although the MINT theorem offers a unique and precise solution in an invertible square matrix, it may lead to high-gain filtration and limit the optimal parameters to the measured points. The problem was addressed by Bai et al. who proposed the time-domain underdetermined multi-channel inverse filtering (TUMIF) approach [5] to solve the model matching problem. The TUMIF approach utilizes the Tikhonov regularization (TIKR) method [6][7][8] to solve a regulated underdetermined inverse filtering problem. Since the TUMIF approach provides multiple solutions, the gain of the filters can be adjusted accordingly. Moreover, the L-curve method [9][10] can be used to determine the optimal regularization parameter. Alternatively, the TUMIF approach can be extended to the frequency domain, leading to the frequency-domain underdetermined multi-channel inverse filtering (FUMIF) approach. This extension aims to reduce the computational complexity associated with the TUMIF approach. The zone control methods introduced earlier are mainly confined to local or boundary control, necessitating either a significant number of loudspeakers or a microphone array. In order to achieve global control of the sound field, many methods incorporate interpolated acoustic transfer functions (ATFs). For instance, the FUMIF system with the linearly constrained minimum variance (FUMIF-LCMV) approach [11] considers the interpolated ATFs in the objective function, providing a general solution for adjusting the performance between local control and global control. The control points consist of both measured points and interpolated points, and they are distributed throughout the listening area. Only the frequency responses between the loudspeakers and the measured points need to be measured, while the unknown frequency responses between the loudspeakers and the fictitious points can be interpolated. However, the underdetermined constraint [12], which these methods need to comply with, can lead to poor control performance in regions where the measured control points are sparse. The performance of zone control in FUMIF-LCMV approach can be optimized by adjusting the two regularization terms using the particle swarm optimization (PSO) [13][14] algorithm.

This thesis presents the weighted acoustic transfer function matching (WATFM) approach, which is inspired by the weighted pressure matching method [15]. Unlike other methods, this approach allows for the inclusion of measured control points without being limited by the underdetermined constraint. Moreover, the proposed method achieves global control by utilizing kernel ridge regression [16] on the interpolated ATFs. Furthermore, the PSO [13][14] algorithm is also employed to optimize the two regularization terms in the WATFM approach. Additionally, the integral operation in the WATFM approach is approximated using Gaussian quadrature [17]. The sensor interpolation problem is also mentioned and is used in the WATFM approach. The sensor interpolation problem can be regarded as a manifold learning problem [18][19]. Traditionally, basis function approaches were employed to learn the lower-dimensional manifold of ATF by representing the sound field as a series expansion of plane-wave components [20][21][22][23] or equivalent simple sources [24][25][26]. A sparse representation is usually possible by using compressed sensing techniques [27][28]. In addition, a sensor interpolation approach that captures the low-frequency dynamics of a room was suggested from the perspective of state-apace acoustic modal analysis [29]. Recently, a promising approach based on kernel ridge regression in machine learning was suggested for sound field interpolation[16][30]. In light of the representer theorem [31], the lower-dimensional manifold can be represented by finite terms of “reproducing kernels” in the Hilbert space. The manifold learning problem can thus be reformulated into a system of linear equations. Unlike the aforementioned basis function approach, the kernel approach is exact and without truncation. Another advantage of the kernel approach lies in its simplicity because the determination of the basis number is unnecessary. e.g., Complicated two-staged procedure for trimming components and solving amplitudes in Plane Wave Decomposition (PWD) [20][21][22][23].

Simulations and experiments are conducted with the subarray of loudspeaker array for underdetermined TIKR, FUMIF-LCMV and WATFM approaches. In the simulation, the ATFs are generated using the image source method (ISM) [32]. The results of the WATFM approach are compared with those of the underdetermined TIKR approach and the FUMIF-LCMV approach. Objective experiments were conducted to evaluate the effectiveness of the proposed method in a realistic application. The word error rate (WER) and the acoustic contrast (AC) were adopted as the objective evaluation metric.

# MULTI-CHANNAL MODEL MATCHING PROBLEM

|  |
| --- |
|  |
| Fig. 1 The model matching problem of multi-channel room response equalization. |

A multi-channel model matching problem can be represented as shown in Fig. 1. In this problem, the mono input signal *x*(*f*) at frequency *f* is filtered by the pre-filters denoted as **h**(*f*). Each filtered signal is reproduced using the measured frequency response (FR) functions of the loudspeaker drivers, denoted as **G**(*f*), which represent the transfer functions between the loudspeakers and the measured control points. The objective of this problem is to minimize the error between the output signal **y**(*f*) and the desired signals **d**(*f*). The desired signals are generated based on a predefined target model of the measured points **m**(*f*), which represents the desired response at those points. The matching error refers to the difference between the output signal and the desired signals. The goal is to minimize this error, indicating a close match between the actual output and the desired signals. To ensure causality in the system, a modeling delay *e-j*2π*f*△ is introduced, where △ represents a time delay in seconds. This delay is added to the system to ensure that the output signal is causally related to the input signal and the desired signals. To ensure the causality of the system, a modeling delay of **m***m* = *e-j*2π*f*△**m** is incorporated. For simplicity, the frequency index *f* will be omitted in the subsequent notation and discussions. Therefore, the model-matching problem can be expressed as

|  |  |
| --- | --- |
|  | ( 1 ) |

Consider *Ns* loudspeakers and *Nm* measured control points. Equation ( 1 ) can be explicitly rewritten with its components as

|  |  |
| --- | --- |
|  | ( 2 ) |

where *Gij*, *i =* 1*,…, Nm*, *j =* 1*,…, Ns*, denotes the FR function between the *j*th loudspeaker and the *i*th measured control point, *hj* is the *j*th prefilter, and *mi* denotes the target model for the *i*th measured control point. The target model *mi*(*f*) is designed in the form of bandlimited equalization, which spans from a lower frequency limit *fl* to an upper frequency limit *fh*. Mathematically, it can be represented as

|  |  |
| --- | --- |
|  | ( 3 ) |

The absolute operation |∙| is used to denote the absolute value. Therefore, within the bandlimited interval, the target model exhibits a flat response that corresponds to the root-mean-square power of each frequency response from the source to the *i*th control point. To achieve acoustic contrast control, the model, *mi*(*f*), in the bright zone is set to be one, while the model in the dark zone is set to a small value.

## The Underdetermined Tikhonov Regularization (TIKR) Method

In the underdetermined TIKR [6][7][8] approach, a perfect matching error solution is achieved by enforcing the rank of the frequency responses matrix **G** to be at least *Nm*. This condition ensures that the system has enough degrees of freedom to perfectly match the desired signals with the output signals. As a result, an intentionally underdetermined system is formulated by selecting *Ns* > *Nm*. To mitigate the ill-posed nature of the matrix inversion problem, an additional regularization term is introduced to the solution in the underdetermined TIKR approach. This regularization term takes the form of a diagonal matrix and serves to stabilize the inversion process and enhance the robustness of the solution. By incorporating regularization, the underdetermined TIKR approach is able to effectively handle scenarios where the matrix inversion problem may be ill-conditioned or susceptible to noise in the measurement data. By this setting, the solution of Equation ( 1 ) can be obtained by using TIKR:

|  |  |
| --- | --- |
|  | ( 4 ) |

where "*H*" denotes the matrix conjugate transpose, and  represents the regularization parameter, which can be chosen using the L-curve method [9][10]. In the L-curve method, the regularization parameter, , is determined by considering the solution norm, , and the residual norm, . The incorporation of the regularization term in the underdetermined TIKR approach considers both the performance of zone control and the feasibility of implementation. By introducing regularization, the approach aims to strike a balance between achieving accurate model matching and ensuring the stability and robustness of the solution. This consideration enhances the practicality and reliability of implementing the zone control algorithm in real-world scenarios. By adjusting the value of  in the underdetermined TIKR approach, it is possible to assign different levels of importance to either the matching error or the filter gain. A lower value of  places more emphasis on reducing the matching error, which can result in improved model matching performance but potentially higher filter gain. On the other hand, a higher value of  prioritizes the regularization term, promoting stability and reducing the filter gain at the cost of slightly higher matching error. The selection of  depends on the specific requirements of the application and the desired balance between accuracy and robustness. By tuning the value of , the underdetermined TIKR approach can be tailored to suit different scenarios and achieve the desired trade-off between model matching performance and stability.

## Frequency-domain Underdetermined Multi-channel Inverse Filtering System With Linearly Constrained Minimum Variance (FUMIF-LCMV) Approach

Despite the ability of the underdetermined TIKR approach to achieve nearly perfect equalization at the measured control points, its performance is limited or even suffers from control spillover in other regions. To address this issue, the FUMIF-LCMV [11] method was developed to extend the effective rendering area beyond the measured control points and encompass a wider region that connects these points. This is achieved by introducing additional control points, referred to as "interpolated" control points, which are not directly measured. By incorporating these interpolated points, a constrained optimization problem can be formulated. The objective is to minimize the matching error at the interpolated points while ensuring that the rendering path aligns with the target model. The FUMIF-LCMV method expands the scope of control and enables accurate control over a larger region, bridging the gap between the measured control points and achieving a more comprehensive sound field reproduction and control performance. The objective function can be expressed as follows:

|  |  |
| --- | --- |
|  | ( 5 ) |

where  denotes the prefilter vector of FUMIF-LCMV. The matching error vector  associated with the interpolated points between the desired model,  and the rendering path, **G***f***h***LCMV* can be written as

|  |  |
| --- | --- |
|  | ( 6 ) |

where  is the room frequency response matrix between the loudspeakers and the interpolated control points and the *Nf* represents the number of interpolated control points. By this setting, the solution of Equation ( 6 ) can be obtain as

|  |  |
| --- | --- |
|  | ( 7 ) |

where  and  are regularization parameters and are inserted to mitigate the ill-posedness during matrix inversion. By considering both the matching error at the interpolated control points and the constraint imposed by the measured control points, the FUMIF-LCMV approach achieves a more comprehensive control over the sound field. This means that it not only achieves good performance at the measured control points but also extends effective control to the surrounding region. As a result, the FUMIF-LCMV approach offers a broader scope of control and improved rendering performance compared to the underdetermined TIKR approach. However, due to the reliance on the Lagrange multiplier method, the FUMIF-LCMV approach is constrained by the underdetermined condition. Consequently, it can only include a limited number of control points within the constraint [12]. This limitation becomes more evident in regions where control points are sparsely distributed, resulting in decreased performance at the interpolated control points. The effectiveness of the method is highly influenced by the placement and density of the measured control points within the underdetermined constraint.

## The Weighted Acoustic Transfer Function Matching (WATFM) Approach

Despite achieving global zone control, the FUMIF-LCMV method has limitations in terms of the number of control points within the constraint and their distribution strategy. For example, in regions where control points are sparsely distributed, leading to a decrease in performance at the interpolated control points. This limitation of the FUMIF-LCMV method highlights the importance of carefully considering the distribution of control points within the constraint. To overcome this limitation, we introduce the weighted acoustic transfer function matching approach. This method expands the effective rendering area beyond the locally measured control points, encompassing a broader region that connects these measured points. Importantly, the proposed method is not constrained by the underdetermined condition, enabling flexible placement of measured control points and achieving better global control performance. In the proposed approach, global zone control is achieved by incorporating a weighted matrix. This matrix is constructed using a kernel method [15] to interpolate to the ATFs and the model. By utilizing this weighted matrix, we are able to extend the control beyond the measured points and effectively control the entire region of interest, including the interpolation points.

### Kernel Interpolation Method

|  |
| --- |
|  |
| Fig. 2 Schematic diagram for defining the interpolation problem. |

Here, we provide a brief introduction to the interpolation problem and kernel interpolation method [16][30]. Fig. 2 depicts the ATF interpolation problem, which includes *Ns* loudspeaker and *Nm* measured control points. Let *sj*(*f*) represent the *j*th loudspeaker signal at frequency *f*. The sound pressure  captured at the *i*th measured control point can be expressed as

|  |  |
| --- | --- |
|  | ( 8 ) |

where  is the ATF between the *j*th loudspeaker and the *i*th measured control point at location **r***i*, and *ni*(*f*) denotes the associated sensor noise. The frequency variable *f* is omitted hereafter for simplicity. The goal of ATF interpolation is to find a lower-dimensional manifold to represent the ATF, *Gf,j*(**r**), between the *j*th loudspeaker and the interpolation control point at position **r**, based on a finite number of ATF measurements*.* On the other hand, by analogy, the objective of model interpolation is to find a lower-dimensional manifold that represents the model, *mf*(**r**), at the interpolation control point position **r**, based on a finite number of control point models, *m*(**r***i*), where *i =* 1, 2, …, *Nm.*

In the kernel ridge regression approach, we consider an ATF manifold *Gf,j*(**r**) as a function of position vector **r** in the Hilbert space  The regression problem can be formulated as follows:

|  |  |
| --- | --- |
|  | ( 9 ) |

where  is the interpolation ATF manifold at **r***i* and  is the regularization parameter. If the function space  is the reproducing kernel Hilbert space that has the form of ATF manifold:

|  |  |
| --- | --- |
|  | ( 10 ) |

where <∙, ∙>denotes the inner product in Hilbert spaceand *κ* is the reproducing kernel, the solution of the regression problem in ( 9 ) has a representation of the form:

|  |  |
| --- | --- |
|  | ( 11 ) |

where for *i* = 1, …, *Nm*. By substituting ( 11 ) to ( 9 ), a closed form of the ridge regression can be derived. In consideration of the Helmholtz equation, the ATF manifold  can be further derived as:

|  |  |
| --- | --- |
|  | ( 12 ) |

where  is the *Nm*×*Nm* identity matrix,  is the ATF vector between the *j*th loudspeaker and *Nm* measured control points,

|  |  |
| --- | --- |
|  | ( 13 ) |
|  | ( 14 ) |

in which *k* is the wave number, and *j*0(∙) = sinc(∙) denotes the zeroth-order spherical Bessel function. By analogy, through the same derivation, we can also derive the model manifold  as:

|  |  |
| --- | --- |
|  | ( 15 ) |

### **The** **Weighted Acoustic Transfer Function Matching (WATFM) Approach**

|  |
| --- |
|  |
| Fig. 3 The target model is synthesized inside the target region  using *Ns* loudspeakers. |

With the underdetermined TIKR approach, effective control is limited to the vicinity of the measured control points. On the other hand, although the FUMIF-LCMV method allows for control over a broader region, its constraint is bound by the underdetermined condition [12]. Therefore, in order to overcome these limitations, we propose to approximate the cost function *J* by interpolating the ATF and model based on the measured ATF and model at the measured control points. This approach not only enables us to achieve more comprehensive control of the desired region without the explicit need for ATF interpolation but also provides the flexibility to include measured control points without being limited by the underdetermined condition like in FUMIF-LCMV. Suppose that *Ns* loudspeakers are placed around a target control region , as depicted in Fig. 3. The target model at  is denoted by *m*(**r**). So we can defined as

|  |  |
| --- | --- |
|  | ( 16 ) |

where , *j =* 1*,…, Ns*, denotes the FR function between the *j*th loudspeaker and the target model at **r**,  is the prefilter for *j*th loudspeaker, and  denotes the target model at **r**. According to Equation ( 16 ), the cost function for the zone control problem can be defined as

|  |  |
| --- | --- |
|  | ( 17 ) |

where  and . Our goal is to find the prefilter **h***kernel*that minimize the cost function *J.* By minimizing this cost function, we aim to optimize the performance of the system and achieve the desired control characteristics within the specified region. However, due to the time-consuming nature of measuring all the ATF, it may not be feasible to obtain ATF data for all control points. Therefore, we consider approximating the cost function *J* by interpolating the ATF and model from the available ATF and model at the measured control points. Through the kernel interpolation [16][30], the ATF manifold *Gf,j*(**r**) and the target model manifold *mf*(**r**) can be interpolated from their values at the measured control points. This interpolation process can be represented by the following equation:

|  |  |
| --- | --- |
|  | ( 18 ) |
|  | ( 19 ) |

Then, the cost function *J* can be approximated as

|  |  |
| --- | --- |
|  | ( 20 ) |

where **W** is defined as

|  |  |
| --- | --- |
|  | ( 21 ) |

with

|  |  |
| --- | --- |
|  | ( 22 ) |

The matrix **W** is referred to as the weighted matrix [15] because it assigns weights to the components of the resulting cost function. These weights determine the relative importance of each component in the optimization process. In the context of the proposed method, the weighted matrix **W** is constructed based on the interpolation weights, which are determined by the kernel function and the distances between the interpolation point and the measured control points. By incorporating these weights into the cost function, the resulting optimization problem becomes a weighted mean square error problem, where the weights reflect the importance of each component in achieving the desired control objectives. In practice, the weighted matrix can be approximated using numerical integration techniques, such as Gaussian quadrature [17].

Through kernel interpolation, the cost function can be approximated and estimated by utilizing the controlling zone region , the positions of the measured control points, the target model, and the measured control points' ATFs. The optimization problem can be approximated as

|  |  |
| --- | --- |
|  | ( 23 ) |

where  is the regularization parameter. It is shown in APPENDIX A that the optimal solution of Equation ( 23 ) is

|  |  |
| --- | --- |
|  | ( 24 ) |

The WATFM approach achieves both zone control and the desired effect by incorporating kernel interpolation and the weighted matrix **W**. Furthermore, unlike the FUMIF-LCMV approach, which is constrained by the underdetermined condition, the proposed system allows for the placement of additional measured control points beyond the number of loudspeakers.

### Gaussian Quadrature

One limitation of the WATFM approach is the need to perform integration of the weighted matrix, which may not have an analytical solution. To overcome this challenge, numerical integration techniques, such as Gaussian quadrature [17], can be employed to approximate the integral operation. Gaussian quadrature is a reliable and efficient method for numerical integration, allowing for accurate approximation of the weighted matrix. By utilizing this technique, the integration step can be efficiently carried out, enabling the practical implementation of the WATFM approach. In the Gaussian quadrature method, we can strategically choose the weights and abscissas to achieve exact integration for a particular class of integrands. Specifically, this class encompasses integrands that can be represented as polynomials multiplied by a known function *W*(*x*). By tailoring the selection of weights and abscissas to match the characteristics of these integrands, we can achieve precise and accurate integration results. This flexibility allows us to handle a wide range of functions beyond simple polynomials, enhancing the applicability and effectiveness of the Gaussian quadrature technique in numerical integration. In other words, given the function *W*(*x*) and an integer *N*, it is possible to determine a set of weights *wi* and abscissas *xi* that satisfy the following approximation:

|  |  |
| --- | --- |
|  | ( 25 ) |

We can choose different integration regions, [a, b], and weight functions *W*(*x*) to perform Gaussian quadrature. In our case, we chose the integration region to be to , [,], and the weight function *W*(*x*) is equal to 1. So in this case, the *wi* can be obtain through following equation:

|  |  |
| --- | --- |
|  | ( 26 ) |

where *PN*(*x*) is the Legendre polynomials. The *xi* values are the roots of the Legendre polynomials. By specifying the desired number of points *N*, we can calculate a set of weights *wi* and abscissas *xi* using Gaussian quadrature. These weights and abscissas enable us to approximate the integration of a given function using the Gaussian quadrature method. To perform Gaussian quadrature on an integration range [a, b], we need to scale the range of integration from [a, b] to [,]. This scaling can be achieved by applying the following transformation:

|  |  |
| --- | --- |
|  | ( 27 ) |

By applying this transformation, we can compute the weights and abscissas in the standard interval [,] and then use them to approximate the integration over the original range [a, b].

# SIMULATIONS

1. **Kernel Interpolation Method**

|  |
| --- |
|  |
| Fig. 4 Simulation setup. |

In this section, we perform a simulation to validate the effectiveness of the kernel interpolation approach [16][30]. The detailed setup of the simulation is illustrated in Fig. 4. The simulation makes use of room impulse responses generated using the image source method (ISM) [32]. This method is commonly used to model the acoustic behavior in enclosed spaces. The simulated room has dimensions of 4.6 m×5 m×2.6 m, and the reverberation time (T60) is set to be 300 ms. The sampling rate used in the simulation is 48 kHz, which determines the resolution of the audio signals and measurements. The sound source is located at the coordinates (2.3, 1.0, 1.1) m within the simulated room. To capture the acoustic characteristics of the room, a total of 36 measured points are evenly positioned throughout the space. The distance between each measured point is set to be 3.4 cm, ensuring a comprehensive sampling of the acoustic field. The 36 measured points are positioned at the corners of the room, covering various locations within the space. One of these points is selected as the reference microphone, positioned at coordinates (0.47, 2.83, 1.1) m. For evaluating the performance of the ATF interpolation technique, the interpolation point A is chosen at coordinates (0.38, 2.91, 1.1) m. It is noteworthy that the spatial aliasing frequency associated with the microphone spacing in this setup is 5 kHz, which means that frequencies beyond this limit may exhibit aliasing effects in the spatial domain. By applying Equation ( 12 ), we can estimate the interpolated ATF by considering the positions and ATF values of the measured points. We employ kernel interpolation with a regularization term  = 10-6.

The interpolated ATF of the ground truth and the kernel interpolation mothed are demonstrated in Fig. 5 and Fig. 6, respectively. The two figures illustrate the interpolated magnitude and phase, respectively, and demonstrate a relatively good agreement with the ground truth up to the spatial aliasing frequency of 5 kHz. To evaluate the interpolation performance, we define a performance measure to quantify the interpolation error. The matching error is calculated at a specific frequency and location to assess the accuracy of the interpolated values and is defined as

|  |  |
| --- | --- |
|  | ( 28 ) |

where (**r**, f) and *g*(**r**, f) denote the interpolated and ground truth ATF, respectively. As depicted Fig. 7, the matching error is minimal below the spatial aliasing frequency. This observation confirms the effectiveness of the kernel interpolation approach, especially below the spatial aliasing frequency. The interpolation technique exhibits satisfactory performance in accurately estimating the ATF at points that are not directly measured. To evaluate the overall frequency band interpolation performance, we define the averaged matching error as follows:

|  |  |
| --- | --- |
|  | ( 29 ) |

By utilizing Equation ( 29 ), we can evaluate the averaged matching error at different points and generate an error map that corresponds to the interpolated region. This map helps visualize the distribution of matching errors throughout the area. As shown in Fig. 8, the averaged matching error is calculated for the frequency band below the spatial aliasing frequency. The red points in the figure represent the measured points where the interpolation error is relatively small when the interpolation points are located close to them. This indicates that the interpolation performance is better in the vicinity of the measured points. This observation validates the effectiveness of the kernel interpolation approach within the region where measurements were taken.

|  |
| --- |
|  |
| Fig. 5 Magnitude result of kernel interpolation method.  The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. |
|  |
| Fig. 6 Unwrapped phase result of kernel interpolation method.  The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. |

|  |
| --- |
|  |
| Fig. 7 Matching error of kernel interpolation method. |

|  |
| --- |
|  |
| Fig. 8 The average matching error.  The red dots represent the measured points. |

1. **Multi-Channel Model Matching Zone Control**

|  |
| --- |
|  |
| Fig. 9 Simulation setup. |

The simulation setup of the anechoic environment is illustrated in Fig. 9. The simulation makes use of room impulse responses generated using the image source method (ISM) [32]. It consists of a twelve-loudspeaker subarray with an interelement spacing of 0.075 m. The subarrays are spaced 0.15 m apart from each other. The loudspeakers and the measured control points were positioned at the same height of 1.1 m. This specific configuration is utilized in the simulation to analyze and evaluate the performance of the three approaches. The control region of interest is defined as a 180° arc line that is centered around the loudspeaker array. This specific region is selected for the distribution of control points in order to focus on the study objectives. The loudspeaker array is positioned 1 m away from the listening area, and the sampled points are measured at 10° intervals. In our setup, we define the bright zone range as to . Conversely, the dark zone range is defined as to and 50° to 90°. In the dark zone, ten measured control points are selected, uniformly spaced at intervals of 10° within the angular range of 50° to 90° and to . Additionally, within the angular range of to 30°, control points are chosen at intervals of 10°. These measured points are considered as the bright zone control points specifically for the proposed method. On the other hand, only the control point at 0° is specifically chosen as the bright zone control point for both the underdetermined TIKR and FUMIF-LCMV approaches. This selection is in accordance with the requirements of these approaches, where the matrix **G** in underdetermined TIKR or the constraint in FUMIF-LCMV needs to be underdetermined. By designating the control point at 0° as the bright zone control point, the underdetermined nature of the problem is maintained, allowing for accurate estimation of the desired ATF while fulfilling the specific constraints of the underdetermined TIKR and FUMIF-LCMV methods. All the sampled points are utilized in the interpolation process to estimate the ATF at unmeasured points.

The weighted matrix is integrated over the range from to 90°, and the number of points *N* is set to 50. On the other hand, the interpolated points are specifically selected based on the abscissas obtained from the Gaussian quadrature [17] formula. These interpolated points are then used in the FUMIF-LCMV method to perform the desired calculations or evaluations. Moreover, this approach guarantees that the weighted matrix encompasses the complete control region, incorporating an ample number of integration points to accurately represent the desired function. The target model at the bright zone is set to 100, representing a power level of 1, while the target model at the dark zone is set to 10-0.5, representing a power level of approximately 0.316. Using Equation ( 3 ), we can calculate the target model and determine the power contrast between the bright zone and the dark zone to be approximately 16 dB. To ensure a consistent power level in the bright zone, we define all bright zone target models to match the power of the model at the 0° point. The sampling rate is 48 kHz.

We compare the underdetermined TIKR, FUMIF-LCMV, and the WATFM approaches. In the underdetermined TIKR approach, the prefilters are calculated using Equation ( 4 ), with a regularization parameter of =10-3.5.The value of  is determined using the L-curve [9][10] method. Similarly, in the FUMIF-LCMV approach, the prefilters are obtained using Equation ( 7 ), with regularization parameters of =10-2 and the =10-3. These values are determined using the PSO [13][14] method. On the other hand, in the WATFM approach, the prefilters are determined using Equation ( 24 ), with the regularization parameters of =10-3 and the =10-4. This value is also determined using the PSO [13][14] method.

First, in Fig. 10, the unprocessed and processed FRs of the bright zone and the dark zone are depicted. We have sampled points at 0° and . The power of the unprocessed FR is quite similar, especially in the low-frequency range. On the other hand, the processed FR shows that all three approaches achieve zone control. The power in the bright zone is preserved, while the power in the dark zone is reduced. This result confirms that all three approaches successfully achieve zone control.

|  |
| --- |
|  |
| Fig. 10 FRs in the bright zone (0°) and dark zone (). The dotted line represents the unprocessed FRs. The solid line represents the processed FRs. The blue line represents the FRs in the bright zone. The red line represents FRs in the dark zone. |

The resulting FRs of the measured and interpolated control points in bright zone or dark zone are depicted in Fig. 11. The measured control points we observe are at 0°, 50°, and 60°. Additionally, we have interpolated control points at 15°, 25°, and 55°. The comparison of FRs in Fig. 11(a) demonstrates that the underdetermined TIKR approach achieves a more accurate match for the measured control points compared to the FUMIF-LCMV and the WATFM approaches. The FRs obtained using the underdetermined TIKR method show a closer alignment with the target response at the measured control points, indicating its superior performance in accurately controlling those points. This validates the effectiveness of the underdetermined TIKR approach in achieving precise control in the measured control points. However, in contrast to Fig. 11(a), Fig. 11(b) shows a different result. It demonstrates that the underdetermined TIKR approach has limited control over the FRs at the interpolated points, suggesting that it may struggle to accurately control those points. Conversely, the WATFM and the FUMIF-LCMV approaches exhibit better model fitting accuracy compared to the underdetermined TIKR approach.

In order to visually represent the rendering performance within the control area, we define a performance metric based on the normalized magnitude matching error as follows:

|  |  |
| --- | --- |
|  | ( 30 ) |

where  is the FR vector between the loudspeakers and the preselected grid points at *f* Hz and position **r**,  is the prefilter vector designed using the underdetermined TIKR, the FUMIF-LCMV or the WATFM approaches, and *mgrid*(**r**, *f*) is the target model at the designated grid point and at the *f* Hz and position **r**. In this case, we calculate the average values of the frequencies ranging from 100 Hz to 1000 Hz, which roughly corresponds to the spatial aliasing frequency. A smaller value of the preceding metric indicates a better matching performance. A more detailed analysis of the matching performance is presented in Fig. 12, which depicts the matching performance along an arc connecting the control points. This plot enables us to observe the variation in performance throughout the interpolation region and offers further insights into the effectiveness of the proposed approaches. It is evident that the underdetermined TIKR approach achieves high matching performance specifically at the measured control points. However, the WATFM approach and FUMIF-LCMV demonstrate a more balanced distribution of performance across all control points in the reproduction area, including both measured and interpolated points. Furthermore, the proposed method exhibits a lower normalized magnitude matching error than both the FUMIF-LCMV and underdetermined TIKR approaches in the bright zone region. In the dark zone region, the proposed method achieves a comparable performance to the other methods. This indicates that the proposed method is effective in achieving accurate magnitude matching in the bright zone and maintains a similar level of performance in the dark zone. In Fig. 12, we observe that the normalized magnitude matching error at the interpolation points tends to increase in regions where the control points are sparse. For example, in the range from to 0° for the FUMIF-LCMV and underdetermined TIKR approaches, or in the range from to for the WATFM approach. However, in regions with dense control points, the interpolation points are able to achieve good matching performance. This suggests that the density and distribution of control points play a significant role in the effectiveness of interpolation-based methods. This highlights the importance of strategically selecting the placement of control points, especially in approaches like underdetermined TIKR and FUMIF-LCMV approaches, where the number of control points is limited due to the underdetermined nature of the problem. In contrast, the proposed method does not have this limitation and allows for the inclusion of more control points, enabling a more global control of the reproduction area. This flexibility in control point placement is advantageous in achieving better overall matching performance across the entire interpolation region.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |
| Fig. 11 Resulting FRs of the measured and interpolated control points. The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. The dotted black line spatial aliasing frequency. |
|  | | |
| Fig. 12 Normalized magnitude matching error. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. | | |

In order to visually represent the rendering performance within the control area, we define a performance metric based on the mean power as follows:

|  |  |
| --- | --- |
|  | ( 31 ) |
|  | ( 32 ) |

where  is the FR vector between the loudspeakers and the preselected grid points at *f* Hz and position **r**,  is the prefilter vector designed using the underdetermined TIKR, the FUMIF-LCMV or WATFM approaches. The frequency range defined from the lower limit *fl* to the upper frequency limit *fh* covers 100 Hz to 1000 Hz. And the mean power is expressed in dB scale. By focusing on this frequency range, we can evaluate the performance of the rendering system specifically in the context of spatial aliasing effects. This metric allows us to evaluate the average power level achieved in the area and provides a measure of the effectiveness of the rendering system. In Fig. 13, the target model in the bright zone maintains the same power as the model at 0°, while the power contrast between the bright zone and the dark zone is approximately 16 dB, as mentioned previously. As shown in Fig. 13, the underdetermined TIKR approach achieves good performance at the measured control points, indicating its effectiveness in accurately controlling those specific points. On the contrary, both the WATFM and the FUMIF-LCMV approaches demonstrate a more precise fit to the model compared to the underdetermined TIKR approach. This suggests that these approaches have a broader scope of control and can effectively match the desired response throughout the control region. However, the proposed method achieves the best performance among the three methods. This can be attributed to the dense distribution of control points, which allows for better model maintenance and improved fitting of the interpolation model. Unlike the underdetermined TIKR and FUMIF-LCMV approaches, the proposed method does not have the constraint of satisfying an underdetermined condition. Therefore, it can allocate a higher number of control points, resulting in superior performance.

|  |
| --- |
|  |
| Fig. 13 Mean power The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. |

By utilizing Equation ( 31 ) and Equation ( 32 ), the beampattern of the mean power can also be obtained. The range of average frequency is the same as mentioned above. As shown in Fig. 14, while the underdetermined TIKR approach can control the measured control points, it cannot achieve global control. On the other hand, the FUMIF-LCMV approach can achieve global control by considering the interpolated points. However, when the measured control points are sparse, the performance of global control decreases, as indicated by the bright zone in Fig. 14(b), which is more global than the underdetermined TIKR approach. Moreover, the WATFM approach achieve the best global control, as indicated by the bright zone in Fig. 14(c). This can be attributed to the fact that more measured control points are utilized.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |
|  |
| (c) |
| Fig. 14 Beampattern of three approaches. |

To analyze the performance of zone control using objective indicators, we define the acoustic contrast (AC) performance metric as follows:

|  |  |
| --- | --- |
|  | ( 33 ) |

where  is the number of samples in the bright zone,  is the number of samples in the dark zone, the  represents the mean power at the sample points within the bright zone,  represents the mean power at the sample points within the dark zone. The acoustic contrast (AC) metric provides a quantitative measure of the power contrast between the two zones, expressed in dB. A higher AC value indicates a greater differentiation and control of the sound field between the bright and dark zones, demonstrating better performance in achieving the desired acoustic contrast. In the simulation, we sample the entire region at intervals of 5°. The bright zone is defined as the range from to 30°, while the dark zone is the combined range of to and 50° to 90°, as previously mentioned. A more detailed analysis of the AC performance is presented in Table I. In Table I, the trend of the results aligns with the previous observations. Both WATFM approach and FUMIF-LCMV approach demonstrate better performance compared to underdetermined TIKR. This can be attributed to the fact that both methods take into account the interpolation points when designing the filters, leading to improved performance at those points. Furthermore, the proposed method, which allows for a greater number of control points and is not constrained by the underdetermined condition, outperforms FUMIF-LCMV in terms of performance. The results in Table I further emphasize the effectiveness of the proposed method in achieving accurate control within the control region and performing global control.

Table I.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Unprocessed | Underdetermined TIKR | FUMIF-LCMV | WATFM |
| AC (dB) | 5.35 | 13.54 | 14.59 | 14.64 |

To assess the voice performance of zone control using objective indicators, we employ the average word error rate (WER). The average WER can be calculated using the following formula:

|  |  |
| --- | --- |
|  | ( 34 ) |

where *S* is the number of substitutions, *D* is the number of deletions, *I* is the number of insertions, *Nw* is the number of words in the reference and *N* is the number of sample points. A lower average WER indicates clearer speech, whereas a higher average WER suggests lower clarity or reduced intelligibility. We also defined the WER contrast metric as follows:

|  |  |
| --- | --- |
|  | ( 35 ) |

A higher WER contrast value indicates a more pronounced differentiation and control of the sound field between the bright and dark zones, showcasing superior performance in achieving the intended acoustic contrast. In the simulation, we sample the entire region at intervals of 5°. The bright zone is defined as the range from to 30°, while the dark zone is the combined range of to and 50° to 90°, as previously mentioned. All ASR experiments are conducted using the SpeechBrain toolkit [33] and the test speech is selected from the LibriSpeech dataset [34]. A more detailed analysis of the AC performance is presented in Table II. In Table II, the trend of the results corresponds with the AC values. The FUMIF-LCMV approach achieves better voice zone control as it considers the interpolated points. On the other hand, the WATFM approach achieves the best performance, as observed in the AC values. Both Table I and Table II confirm that the proposed method, which takes into account more measured control points, can achieve superior global zone control.

Table II.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Unprocessed | Underdetermined TIKR | FUMIF-LCMV | WATFM |
| WERaverage,bright | 0 % | 0 % | 28 % | 0 % |
| WERaverage, dark | 17 % | 28 % | 94 % | 77 % |
| WERcontrast | 17 % | 28 % | 66 % | 77 % |

# EXPERIMENTS

To validate the proposed approach under realistic conditions, experiments were carried out in a listening room with dimensions of 4.6 m × 5 m × 2.6 m and a reverberation time (T60) of 250 ms. The experiments aimed to assess the performance of the proposed method in a real-world setting and evaluate its effectiveness in achieving accurate control and zone matching in a practical acoustic environment. The experimental arrangement consisted of a twelve-loudspeaker subarray, with each loudspeaker spaced at an interelement distance of 0.075 m. The subarrays themselves were positioned at a distance of 0.15 m apart from each other. This setup, as depicted in Fig. 15, allowed for the generation of a controlled sound field and facilitated the evaluation of the proposed method's performance in the physical listening environment. The intended listening area was positioned at a distance of 1 m from the loudspeaker array. The control region of interest was defined as a 180° arc line centered around the loudspeaker array, and sampled points were measured at intervals of 10°. Within this control region, the bright zone was specified as the angular range from to 30°, where specific control points were selected. On the other hand, the dark zone encompassed the angular ranges from to and 50° to 90°, representing areas outside the bright zone where additional control points were chosen. In the dark zone, a total of ten measured control points were selected, uniformly spaced at intervals of 10° within the angular range of 50° to 90° and to . Moreover, control points are selected at intervals of 10° within the angular range of to 30° and these measured points are specifically designated as the bright zone control points for the proposed method. However, for both the underdetermined TIKR and FUMIF-LCMV approaches, only the control point at 0° is specifically chosen as the bright zone control point. The loudspeakers and the measured control points were positioned at the same height of 1.1 m. The positioning of the microphones was done to capture the acoustic characteristics and evaluate the performance of the interpolation technique in the desired control region. FRs between loudspeakers and microphones are measured by a signal analyzer, PreSonus®, at the sampling rate 48 kHz.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |
|  |
| (c) |
| Fig. 15 (a) The side view of the experiment setting. (b) The front view of the experiment setting. (c) The abridged general view of the experiment setting. |

1. **Kernel Interpolation Method**

As mentioned previously, we utilize kernel interpolation method [16][30] to approximate the FRs before designing the FUMIF-LCMV filters, and we use the kernel interpolation method to drive WATFM approach. We employ kernel interpolation with a regularization term  = 10-6. The interpolation degrees are set to , 5°, 25°, and 45°. The experimental results of the interpolated FRs for the first loudspeaker and the interpolated points are presented in Fig. 16, Fig. 17 and Fig. 18. These results demonstrate the effectiveness of the kernel interpolation method in accurately estimating the FRs at the interpolated points in a real-world setting. The magnitude and phase of the interpolated FRs are compared with the measured FRs at the corresponding interpolated points, as illustrated in Fig. 16 and Fig. 17. The results illustrate that the interpolated FRs effectively capture the magnitude and phase of the FRs below the spatial aliasing frequency (1000 Hz). On the other hand, as shown in Fig. 18, the matching error, which is estimated through Equation ( 28 ), is minimal below the spatial aliasing frequency, confirming the effectiveness of the kernel interpolation approach, particularly in this frequency range. The results validate the effectiveness of the kernel interpolation method in real-world scenarios, demonstrating its ability to accurately interpolate the FRs at the interpolated points. Moreover, these results further confirm the validity of WATFM approach and support its applicability in real-world applications.

|  |
| --- |
|  |
| Fig. 16 Magnitude result of kernel interpolation method. The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. |
|  |
| Fig. 17 Unwrapped phase result of kernel interpolation method.  The blue line represents ground truth. The red line represents estimated FRs. The black line represents spatial aliasing frequency. |
|  |
| Fig. 18 Matching error of kernel interpolation method. |

1. **Multi-Channel Model Matching Zone Control**

In the conducted experiments, the proposed approach for global zone control is validated. The regularization parameters for the approach, namely  and , are determined using the PSO [13][14] method. The regularization parameters,  and , are determined as =10-3 and =10-4. Additionally, the weighted matrix is integrated over the range from to 90°, and the number of points *N* is set to 50. This ensures that the weighted matrix covers the entire control region adequately, allowing for precise control over the desired area. On the other hand, the selection of interpolated points is based on the abscissas obtained from the Gaussian quadrature [17] formula. These chosen interpolated points are then employed in the FUMIF-LCMV method for performing the desired calculations or evaluations. The underdetermined TIKR approach utilizes a regularization parameter of = 0.0004, which is determined using the L-curve [9][10] method. In the case of FUMIF-LCMV approach, the regularization parameters are set to =0.005and the =10-2, and these values are determined through the PSO [13][14] method.

Fig. 19 illustrates the unprocessed and processed FRs of both the bright zone and the dark zone. Sampling was done at points 0° and . The outcome mirrors that of the simulation. The power of the unprocessed FRs are notably similar, particularly within the low-frequency range. Conversely, the processed FRs demonstrate that all three approaches effectively achieve zone control. The power within the bright zone is maintained, while that within the dark zone is diminished. This result reaffirms the successful zone control achieved by all three approaches in real-world applications.

|  |
| --- |
|  |
| Fig. 19 FRs in the bright zone (0°) and dark zone (). The dotted line represents the unprocessed FRs. The solid line represents the processed FRs. The blue line represents the FRs in the bright zone. The red line represents FRs in the dark zone. |

The measured control points we observe are at 0°, 60°, and 70°. Additionally, we have interpolated control points at 25°, 55°, and 65°. The FRs of the measured and interpolated control points in the bright zone and dark zone are shown in Fig. 20. Upon comparing the FRs in Fig. 20(a), it is evident that the underdetermined TIKR approach achieves a higher level of accuracy in matching the measured control points compared to the FUMIF-LCMV approach and weighted acoustic transfer function matching approach. The experimental results not only confirm the findings from the simulation but also provide further validation of the underdetermined TIKR approach's superior performance in accurately controlling the measured control points. In contrast to the superior performance of the underdetermined TIKR approach in accurately controlling the measured control points, Fig. 20(b) reveals a different outcome for the interpolated points. In Fig. 20(b), the FRs of the FUMIF-LCMV and WATFM approaches at the interpolated points closely match the target model. Additionally, the WATFM approach performs the best among the interpolated points in the bright zone. It indicates that the underdetermined TIKR approach has limited control over the FRs at the interpolated points, suggesting challenges in accurately controlling those points. On the other hand, the WATFM approach and the FUMIF-LCMV approach demonstrate a more balanced model matching performance, not only at the measured points but also at the interpolated points. This indicates their effectiveness in achieving accurate control throughout the entire control region, including both measured and interpolated points. These findings highlight the advantage of WATFM approach and the FUMIF-LCMV approach in achieving precise control in a broader range of locations within the control region.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |
| Fig. 20. Resulting FRs of the measured and interpolated control points.  The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. The dotted black line spatial aliasing frequency. |

Fig. 21 presents the normalized magnitude matching error, estimated through Equation ( 30 ), where the average values are calculated for the frequencies ranging from 100 Hz to 1000 Hz, which approximately corresponding to the spatial aliasing frequency. The results reveal that the underdetermined TIKR approach achieves high matching performance specifically at the measured control points. However, within the bright zone, the WATFM approach and the FUMIF-LCMV approach outperform the underdetermined TIKR approach at the interpolation points. The WATFM approach and the FUMIF-LCMV approach exhibit a more even distribution of performance across all control points within the reproduction area, encompassing both the measured and interpolated points. Furthermore, the proposed method demonstrates a lower normalized magnitude matching error magnitude compared to both the FUMIF-LCMV and underdetermined TIKR approaches in the bright zone region. In the dark zone region, the proposed method achieves comparable performance to the other methods. In both the simulation and the experiment, it is observed that the normalized magnitude matching error at the interpolation points tends to increase in regions where the control points are sparse. Conversely, in regions with dense control points, the interpolation points are able to achieve good matching performance. This observation is consistent with the behavior of the FUMIF-LCMV and underdetermined TIKR approaches, particularly in the range from to 0°. However, the proposed method overcomes this limitation by allowing for the inclusion of more control points, resulting in better performance in these areas and enabling a more global control of the reproduction area. The flexibility in control point placement offered by the proposed method is advantageous in achieving better overall matching performance across the entire interpolation region.

|  |
| --- |
|  |
| Fig. 21 Normalize magnitude matching error. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. |

In Fig. 22, the mean power, estimated through Equation ( 32 ), is shown for the frequency range from a lower limit *fl* to an upper limit *fh*, which is set to encompass frequencies from 100 Hz to 1000 Hz. The target model in the bright zone is designed to maintain the same power as the model at 0°, while the power contrast between the bright zone and the dark zone is approximately 16 dB, as previously mentioned. As depicted in Fig. 22, the underdetermined TIKR approach exhibits good performance at the measured control points, indicating its effectiveness in accurately controlling those specific points. In the context of the bright zone, the WATFM approach and the FUMIF-LCMV approach outperform the underdetermined TIKR approach at the interpolation points. This suggests that WATFM approach and FUMIF-LCMV approach have a broader scope of control and can effectively match the desired response throughout the control region, including both measured and interpolated points. Remarkably, the proposed method outperforms the other two methods, showcasing its superiority in achieving precise control across the entire control region and attaining the most effective acoustic contrast control. This can be attributed to the dense distribution of control points in the proposed method, which allows for better model maintenance and improved fitting of the interpolation model. Unlike the underdetermined TIKR and FUMIF-LCMV approaches, the proposed method does not have the constraint of satisfying an underdetermined condition. Therefore, it can allocate a higher number of control points, resulting in superior performance. The increased number of control points enhances the global control capabilities of the proposed method, leading to improved matching performance throughout the control region.

|  |
| --- |
|  |
| Fig. 22 Mean power The black line represents model FRs. The blue line represents unprocessed FRs. The magenta line represents Underdetermined TIKR approach. The green line represents FUMIF-LCMV approach. The red line represents WATFM approach. |

Table III presents the AC values, estimated through Equation ( 33 ), obtained by sampling the entire region at intervals of 5°. The bright zone is defined as the angular range from to 30°, while the dark zone encompasses the combined angular ranges of to and 50° to 90°, as mentioned earlier. Both WATFM approach and FUMIF-LCMV approach demonstrate better performance compared to underdetermined TIKR. This can be attributed to the fact that both methods consider the interpolation points during the filter design process, leading to improved performance at those points. Additionally, the proposed method, which allows for a greater number of control points and is not constrained by the underdetermined condition, surpasses FUMIF-LCMV approach in terms of overall performance. This suggests that the proposed method has a greater capacity for achieving accurate control throughout the entire control region. The results in Table III further emphasize the effectiveness of the proposed method in achieving accurate control within the control region and performing global control. By incorporating the weighted matrix and utilizing interpolation techniques, the proposed method demonstrates superior performance in terms of matching error, power contrast, and overall control capability. The ability to distribute control points more widely and consider the interpolation points allows for better control over the entire reproduction area, resulting in enhanced performance and improved rendering quality. These findings validate the efficacy of the proposed method in practical applications and highlight its potential for acoustic contrast control in real-world scenarios.

Table III.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Unprocessed | Underdetermined TIKR | FUMIF-LCMV | WATFM |
| AC (dB) | 0.13 | 5.39 | 6.15 | 6.33 |

# CONCLUSIONS AND FUTURE WORK

## 5.1. Conclusions

In this thesis, multi-channel model matching approaches for global zone control and acoustic contrast control have been presented. In the simulation, WATFM approach achieved the best performance with an AC value of 14.64 dB. Furthermore, in terms of the WER contrast value, WATFM approach also demonstrated the best performance with a value of 77%. This performance is attributed to the WATFM approach is not limited by the underdetermined condition. Conversely, the FUMIF-LCMV approach, with a simulated AC value of 14.59 dB and a WER contrast value of 66 %, achieves a better global zone control performance than the underdetermined TIKR approach, which has a simulated AC value of 13.54 dB and a WER contrast value of 28%, due to the consideration of interpolated points. However, the FUMIF-LCMV approach remains constrained by the underdetermined condition, leading to a relatively inferior global zone control performance compared to the WATFM approach.

The trend remained consistent in the experiment, where WATFM approach maintained its superiority with an AC value of 6.33 dB. Conversely, the results from the FUMIF-LCMV approach, with an AC value of 6.15 dB, and the underdetermined TIKR approach, with an AC value of 5.39 dB, further confirmed the same conclusion as observed in the simulation.

## Future Work

The inclusion of stereo signal conditions in WATFM approach is a promising direction for future research. By utilizing stereo signals, it becomes possible to not only control the zone but also provide a more immersive auditory experience for the user. Integrating stereo signal conditions into WATFM approach has the potential to further improve the accuracy and realism of global zone control, resulting in a more immersive and enjoyable listening experience. Furthermore, employing neural networks to replace filters also holds significant potential for future advancements.

**REFERENCES**

[1] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources", *J. Acoust. Soc. Amer.*, vol. 111, no. 4, pp. 1695-1700, 2002.

[2]M. Poletti, "An investigation of 2D multizone surround sound system", *Proc. AES 125th Conv. Audio Eng. Society*, 2008.

[3] Y. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design", *IEEE Trans. Audio Speech Lang. Processing*, vol. 19, no. 6, pp. 1711-1720, 2011.

[4] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics", *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 36, pp. 145-152, 1988.

[5] M. R. Bai, Y. W. Chen, Y.-C. Hsu, and T. Y. Wu, "Robust binaural rendering with the time-domain underdetermined multichannel inverse prefilters," *J. Acoust. Soc. Amer.*, vol. 146, no. 2, pp. 1302-1313, Aug. 2019.

[6] C. W. Groetsch, "The theory of Tikhonov regularization for Fredholm equations of the first kind" in Research Notes in Mathematics, MA, Boston:Pitman, vol. 105, 1984.

[7] A. N. Tikhonov, "Solution of nonlinear integral equations of the first kind", *Soviet Math. Dokl.*, vol. 5, pp. 835-838, 1964.

[8] M. Bertero, T. A. Poggio and V. Torre, "Ill-posed problems in early vision", *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869-889, Aug. 1988.

[9] P.C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems", *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487-1503, Nov. 1993.

[10] P. R. Johnston and R. M. Gulrajani, "Selecting the corner in the L-curve approach to Tikhonov regularization", *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1293-1296, 2000.

[11] W.-L. Lin, Y.-S. Chen, B.-R. Lai, and M. R. Bai, "Multichannel room response equalization with a broadened control region using a linearly constrained approach and sensor interpolation," *J. Acoust. Soc. Amer.*, vol. 153, no. 4, pp. 1960, Apr. 2023

[12] C.-Y. Chi, W.-C. Li and C.-H. Lin, Convex Optimization for Signal Processing and Communications, Boca Raton, FL, USA:CRC Press, 2017.

[13] J. Kennedy and R. Eberhart, "Particle swarm optimization", *Proc. ICNN'95–Int. Conf. Neural Networks*, pp. 1942-1948, 1995.

[14] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory", *Proc. 6th Int. Symp. Micro Mach. Hum. Sci.*, pp. 39-43, 1955.

[15] S. Koyama, K. Kimura, and N. Ueno, "Weighted Pressure and Mode Matching for Sound Field Reproduction: Theoretical and Experimental Comparisons," arXiv preprint arXiv:2303.13027, 2023.

[16] N. Ueno, S. Koyama and H. Saruwatari, "Kernel ridge regression with constraint of helmholtz equation for sound field interpolation", *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, pp. 436-440, 2018.

[17] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Numerical Recipes, New York, NY, USA:Cambridge Univ. Press, 1986.

[18] B. Laufer-Goldshtein, R. Talmon and S. Gannot, "A study on manifolds of acoustic responses", *Proc. 12th Int. Conf. Latent Variable Analysis Independent Component Analysis*, pp. 203-210, Aug. 2015.

[19] M. Hahmann, S. A. Verburg and E. Fernandez-Grande, "Analysis of the sound field in a room using dictionary learning", *Proceedings of 23rd International Congress on Acoustics. German Acoustical Society (DEGA)*, 2019.

[20] M. Nolan, M. Berzborn and E. Fernandez-Grande, "Isotropy in decaying reverberant sound fields", *J. Acoust. Soc. Am*, vol. 148, pp. 1077-1088, Aug. 2020.

[21] M. Nolan, E. Fernandez-Grande, J. Brunskog and C.-H. Jeong, "A wavenumber approach to quantifying the isotropy of the sound field in reverberant spaces", *J. Acoust. Soc. Am*, vol. 143, pp. 2514-2526, Apr. 2018.

[22] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution", *J. Acoust. Soc. Amer.*, vol. 116, no. 4, pp. 2149-2157, 2004.

[23] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array", *J. Acoust. Soc. Amer.*, vol. 118, no. 5, pp. 3094-3103, 2005.

[24] A. Sarkissian, "Method of superposition applied to patch near-field acoustic holography", *J. Acoust. Soc. Amer.*, vol. 118, no. 2, pp. 671-678, 2005.

[25] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms", *J. Acoust. Soc. Amer.*, vol. 117, no. 4, pp. 2100-2111, 2005.

[26] C.-X. Bi, Y. Liu, L. Xu and Y.-B. Zhang, "Sound field reconstruction using compressed modal equivalent point source method", *J. Acoust. Soc. Amer.*, vol. 141, no. 1, pp. 73-79, Jan. 2017.

[27] Samuel A Verburg and Efren Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing", *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3770-3779, 2018.

[28] R. Mignot, L. Daudet and F. Ollivier, "Room reverberation reconstruction: Interpolation of the early part using compressed sensing", *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 11, pp. 2013-2312, Nov. 2013.

[29] H.-M. Jian, Y.-S. Chen, and M. R. Bai, "Acoustic modal analysis of room responses from the perspective of state-space balanced realization with application to field interpolation," *J. Acoust. Soc. Amer.*, vol. 152, no. 1, pp. 240-250, 2022.

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

[31] B. Schölkopf, R. Herbrich and A. J. Smola, "A generalized representer theorem", *Proc. Int. Conf. Comput. Learn. Theory*, pp. 416-426, Jul. 2001.

[32] J. B. Allen and D. A. Berkley, "Image method for efficiently simulation small-room acoustics", *Journal of Acoustical Society of America*, vol. 64, no. 4, pp. 943-950, 1979.

[33] Mirco Ravanelli, Titouan Parcollet, Peter Plantinga, Aku Rouhe, Samuele Cornell, Loren Lugosch, Cem Subakan, Nauman Dawalatabad, Abdelwahab Heba, Jianyuan Zhong et al., "SpeechBrain: A general-purpose speech toolkit", *arXiv preprint*, 2021.

[34]V. Panayotov, G. Chen, D. Povey and S. Khudanpur, "LibriSpeech: An ASR corpus based on public domain audio books", Proc. Int. Conf. Acoust. Speech Signal Process., pp. 5206-5210, 2015.

# Appendix A

## Optimal solution of the weighted acoustic transfer function matching approach

The cost function of the weighted acoustic transfer function matching approach can be written as

|  |  |
| --- | --- |
|  | ( A1 ) |

On the basis of the kernel interpolation method in section 4.3.1, the interpolated ATFs and interpolated model can be written as

|  |  |
| --- | --- |
|  | ( A2 ) |
|  | ( A3 ) |

Then, the cost function *J* can be approximated as

|  |  |
| --- | --- |
|  | ( A4 ) |

where **W** is defined as

|  |  |
| --- | --- |
|  | ( A5 ) |

with

|  |  |
| --- | --- |
|  | ( A6 ) |

The optimization problem of WATFM approach is formulated using the (A4) as:

|  |  |
| --- | --- |
|  | ( A5 ) |

where  is a vector of Lagrange multiplier. Taking the complex gradient of the Lagrangian and setting it to zero, we obtain

|  |  |
| --- | --- |
|  | ( A6 ) |

Solving (A6) for the optimal prefilter vector leads to

|  |  |
| --- | --- |
|  | ( A7 ) |

The regularization optimization in (A7) is required to avoid the singularity of the matrix inverse.