Room Impulse Response
Interpolation via Optimal Transport

Aaron Geldert

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.
Espoo 30.12.2022

Supervisor

Prof. Sebastian Jiro Schlecht

Advisor

Nils Meyer-Kahlen
Interpolation between multiple room impulse responses is often necessary for dynamic auralization of virtual acoustic environments, in which a listener can move with six degrees-of-freedom. The spatial room impulse response (SRIR) represents the combined effects of the surround room as sound propagates from a source to the listener, and varies as the source or listener positions change. The early portion of the SRIR contains sparse reflections, considered to be distinct sound events, that tend to be impaired with interpolation methods based on simple linear combinations. With parametric processing of SRIRs, corresponding sound events are able to be mapped to one another and produce a more physically accurate spatio-temporal interpolation of the early portion of the SRIR.

In this thesis, a novel method for parametric SRIR interpolation is proposed based on the principle of optimal transportation. First, SRIRs are represented as point clouds of sound pressure in a 3D virtual source space. Mappings between two point clouds are obtained by defining a partial optimal transport problem solvable with familiar linear programming techniques. The partial relaxation is implemented by permitting both point-to-point mappings and dummy mappings. The obtained optimal transport plan is used to compute the interpolated point cloud which is converted back to an SRIR.

Testing of the proposed method against three baseline comparison methods was done with SRIRs generated by geometrical acoustical modeling. An error metric based on the difference in energy between low-passed rendering of the omnidirectional room impulse response was used. Statistical results indicate that the proposed method consistently outperforms the baseline methods of interpolation. Qualitative examination of the mapping methods confirms that partial transport produces more physically accurate spatio-temporal mappings. For future work, it is suggested to consider different cost functions, interpolate between measured SRIRs, and to render the responses to allow perceptual tests.

**Keywords** impulse response, interpolation, optimal transport, early reflection
Preface

This master’s thesis work was completed at Aalto University between February and December 2022. However, my first explorations working towards a thesis topic began back in June 2021 at Bose corporation, in a role as a signal processing research intern. I started out investigating the individualization of head related transfer functions, through modeling the anthropometry of head and pinnae shapes. Though learning a lot and continuing further research at Aalto, I found myself struggling to hone in on a concise scope for the content of the master’s thesis.

After discussion with Sebastian Schlecht, he proposed a new research direction in HRTF interpolation using optimal transport, building off of my prior knowledge while hoping for a greater sense of novelty. Through the twists and turns of the research process, defining and redefining the scope of the thesis, and the typical distractions and frustrations, I am truly satisfied with how this thesis work represents nearly a year of my learning and personal growth.

In particular, I sincerely thank Sebastian for his supervision, patience, and brilliant guidance through the thesis process and my Master’s studies. I also extend enormous thanks to Nils Meyer-Kahlen for advising me with incredible enthusiasm, including his collaboration from three different continents! They both have always gone above and beyond to provide support and inspiration.

Special thanks go to David Dick for mentoring me between 2019 and 2021, along with my other colleagues at Bose who helped ease my way into spatial audio research. I also sincerely appreciate the wisdom and instruction of professors Ville Pulkki, Vesa Välimäki, and Tapio Lokki. Additional thanks go to Anton Vavilov for generously guiding me through the mathematical basis of transportation theory.

Perhaps most of all, I am deeply grateful for the support of my family and friends, even overseas; my encouraging and inspiring colleagues at the Acoustics Lab, my dear friends in the AAT program, Media Lab, Effective Altruism, Hydepakki, and broader Aalto student communities. My time in Finland has been greatly enriched by so many wonderful people.

December 15th, 2022
Espoo, Finland

Aaron Geldert
Contents

Abstract 3

Preface 4

Contents 5

Symbols and abbreviations 7

1 Introduction 9

2 Background on virtual acoustics 11
   2.1 Room acoustics 11
   2.2 The room impulse response (RIR) 12
   2.3 The spatial room impulse response (SRIR) 14
   2.4 Parametric processing of SRIRs 15
   2.5 Geometrical acoustics 16
      2.5.1 Model assumptions 16
      2.5.2 Image method 17
   2.6 SRIR interpolation 19
      2.6.1 Problem definition 19
      2.6.2 Linear methods 20
      2.6.3 Nonlinear methods 20
      2.6.4 Parametric methods 21

3 Background on optimal transport 22
   3.1 Optimal matching 23
   3.2 Monge formulation 24
   3.3 Kantorovich formulation 25
   3.4 Partial transport 26
   3.5 Applications 28

4 Methods 29
   4.1 Virtual source point clouds 29
   4.2 Interpolation problem definition 29
   4.3 Partial optimal transport of point clouds 30
   4.4 Constructing the interpolated SRIR 32
      4.4.1 Transported mass 34
      4.4.2 Vanishing/Appearing dummy mass 34
      4.4.3 Moving dummy mass 34
   4.5 Baseline interpolation methods 36
      4.5.1 Simple linear 36
      4.5.2 Aligned linear 36
      4.5.3 Greedy nearest-neighbor mapping 36
   4.6 Evaluation methods 38
5 Results and Discussion

5.1 Statistical results

5.1.1 Cuboid and canted rooms

5.1.2 Trapezoidal room

5.2 Discussion

5.2.1 Statistical results

5.2.2 Linear methods

5.2.3 Parametric methods

5.2.4 Limitations of partial OT interpolation

5.2.5 Future work and outlook

6 Conclusion

References
Symbols and abbreviations

Symbols

- $h$ impulse response
- $c(\cdot)$ cost function
- $T$ surjective transport plan (Monge map)
- $a, b$ distributions of mass
- $P, Q, R$ virtual source point cloud representation of SRIR
- $p, q, r$ effective pressure of a virtual source
- $x, y, z$ positions of virtual sources
- $C$ cost matrix
- $T$ transport matrix
- $\pi$ transport plan
- $\Pi$ set of all transport plans
- $\tau$ time of arrival (TOA)
- $\theta$ direction of arrival (DOA)
- $w$ spreading function (4 ms Hann window)
- $s$ receiver (or listener) position
- $\kappa$ interpolation parameter
- $u, v$ dummy mapping vectors
- $\xi$ cost of dummy mappings
- $\sigma$ total amount of transported mass in POT
- $\mathcal{E}$ error function
- $\gamma$ extended value, for partial optimal transport
- $\hat{\cdot}$ estimated value
- $\sharp$ push forward operator
- $\langle \cdot \rangle_F$ Frobenius inner product
- $1_N$ vector of ones, of dimensions $N \times 1$
- $\mathbb{R}$ set of real numbers
- $\| \cdot \|_1$ 1-norm (cardinality)
- $\| \cdot \|_2$ 2-norm (length)
Abbreviations

3DoF  3 degrees-of-freedom  
6DoF  6 degrees-of-freedom  
DOA  Direction of arrival  
GA  Geometrical acoustics  
IR  Impulse Response  
RIR  Room Impulse Response  
SRIR  Spatial Room Impulse Response  
OT  Optimal Transport  
POT  Partial Optimal Transport  
SDM  Spatial Decomposition Method  
TOA  Time of arrival  
VS  Virtual source
1 Introduction

Producing virtual acoustic environments in which a listener can navigate with six degrees-of-freedom has increasingly become a focus of spatial audio research. Rendering such spatial audio for playback over headphones or loudspeaker arrays requires being able to spatialize various sound sources as well as simulate the acoustical effects of the surrounding room. This is possible by making use of a measured or simulated Spatial Room Impulse Response (SRIR), which represents the combined effects of sound propagation from a source to a receiver in a room. With any changes to the configuration of the sound sources, room geometry, or listener, the SRIRs must be dynamically updated.

Updating SRIRs may be performed by selecting from and interpolating between a limited set of responses. Such a set of responses could be collected from real world measurements or acoustical simulations of the room in different source-receiver configurations. Numerous methods for interpolating between spatial audio signals and/or impulse responses have been developed, from simple linear combinations to various nonlinear or parametric algorithms. Defining an accurate and smoothly-changing interpolant between multiple SRIRs is a non-trivial task, in particular because the early portion of the SRIR contains sparse early reflections. As the source-listener configuration changes, these reflection sound events vary in level, time of arrival, and direction of arrival. They may also spontaneously vanish or appear as specular reflection paths are not guaranteed to exist between all regions of the room.

In the ‘blind interpolation’ problem, it is assumed that no explicit knowledge about the room geometry or source and listener coordinates is given. However, some spatial information can be readily extracted from the SRIRs. Parametric methods attempt to utilize this information by processing the SRIR with a method such as Plane Wave Decomposition (PWD), Spatial Impulse Response Rendering (SIRR), Spatial Decomposition Method (SDM) or Ambisonic Spatial Decomposition Method (ASDM). The processed SRIR can be used to distinguish prominent sound events from diffuse components. By modeling such sound events as virtual sources, corresponding sources shared between multiple responses can be mapped to one another, informing a more physically accurate parametric interpolation. Determining appropriate mappings in the blind interpolation problem is also not easily resolved, and state-of-the-art methods depend on carefully-tuned, iterative parametric algorithms.

It is also possible to define mappings based on an optimal transport (OT) formulation between sets of multidimensional data. A centuries-old field of mathematics that has found a wide range of contemporary applications, OT problems consider the transportation of matter between different distributions with the least total effort. By decomposing SRIRs into spatial distributions of acoustic energy, an optimal transport plan between SRIRs can be obtained with familiar linear programming methods and be used to define an interpolant. A partial transport formulation allows for virtual sources to change in level, appear, or disappear without generating unsuitable mappings. With OT-based interpolation, the advantages of parametric mapping methods may be able to be represented within a more general, simpler formulation.

The objective of this thesis is to develop and present a novel method for SRIR
interpolation based on a partial optimal transport plan in a virtual source domain. In
doing so, this thesis will formulate an optimization problem suitable for interpolating
between sets of virtual sources corresponding to sound events in the early portion
of the room impulse response. To test the interpolation method, SRIRs generated
by geometrical acoustical modeling techniques will be used to perform interpolation
using the proposed method and three baseline interpolation methods: a simple linear
combination of impulse responses, a linear combination where the time-of-arrival of
the direct sound is aligned, and a nearest-neighbor spatial mapping method without
any consideration of partial transport. Objective evaluation of the temporal alignment
error in the monaural room impulse responses and qualitative analysis of the spatial
mappings will be used to assess the suitability of the interpolation methods. The
thesis will also demonstrate and discuss the utility of optimal transportation as a
versatile tool for comparing multidimensional data.

The scope of this thesis is limited to the development and objective evaluation of
interpolation between SRIRs generated using specular geometrical modeling. The
decomposition of measured SRIRs, rendering techniques for specific playback systems,
and real-time implementation considerations are left for future work. Perceptual
testing and auditory modeling are also excluded from this thesis.

The remainder of this thesis is organized as follows. Chapter 2 reviews key concepts
about room acoustics, with a focus on SRIR processing, geometrical acoustical
modeling, and the SRIR interpolation problem. Chapter 3 explains the mathematical
foundations of optimal transport problems and relevant computational methods.
Chapter 4 details a novel method for interpolation between several SRIRs using
optimal transport. Chapter 5 presents the results of the thesis, comparing the
proposed method against conventional techniques, and discusses the characteristics
of the different approaches. Chapter 6 concludes the thesis by summarizing the
capabilities of the proposed method, as well as its shortcomings and aspects needing
further study.
2 Background on virtual acoustics

The term ‘virtual acoustics’ encompasses a variety of topics related to the modeling of acoustical phenomena and spatial sound reproduction [1]. Such topics include the simulation of the physics of sound propagation in rooms, which enables designers to predict the acoustical behavior of imagined architectural spaces. Also of particular interest is the recording, processing, and reproduction of three-dimensional sound scenes in conjunction with a head-tracked virtual or augmented reality device. By incorporating dynamic head rotations and/or positional movements to update a binaural virtual acoustic rendering, listeners can experience a virtual acoustic scene different from their actual surroundings with an increased sense of immersion, externalization and realism [2, 3].

In this chapter, the main concepts and physical behaviors related to room acoustics and room impulse responses are summarized. Relevant methods for analyzing and modeling room impulse responses are then discussed. Finally, the ‘blind interpolation’ problem being studied is formally introduced, and state-of-the-art related works are reviewed.

2.1 Room acoustics

Much of virtual acoustics focuses on modeling room acoustics, where sound travels and interacts with an enclosed space. Fundamentally, airborne sound is local variations in air pressure that propagate and interact with the surrounding environment. Important interactions in room acoustics include the absorption of sound energy by materials (including dissipation in air), the diffraction of sound around boundaries of surfaces, the reflection of sound off of surfaces, and diffuse scattering of sound when reflecting surfaces are rough or non-ideal. These interactions are generally frequency-dependent and best described by wave behaviors, though simpler frequency-independent, ray-based modeling is used in most of this thesis work. Often, a model of an acoustical system is divided into three subsystems:

**Sound sources:** Sources such as loudspeakers, voices, musical instruments, or other vibrating objects that produce a sound signal such as an impulse, noise, speech, music, etc.

**Transmission medium:** The surrounding physical environment in which sound propagates in and interacts with.

**Receiver:** The receiver observes the local sound field as a signal, and may be a human listener, single microphone, or microphone array.

In practice, the source-medium-receiver paradigm is useful for making virtual acoustic models more understandable and modular. In this thesis work, parametric modeling of the transmission medium is of greatest interest, and models of sound sources or receiver are left out of scope.

It is worth pointing out several other typical simplifications of the physics of sound, also used in this thesis work. All systems are assumed to be causal, linear
and time-invariant; the medium for sound propagation is homogeneous air with a uniform speed of sound, \( c \); and only the frequencies relevant to human perception, at most from 20 to 20,000 Hz, are of interest.

### 2.2 The room impulse response (RIR)

The Room Impulse Response (RIR) is the complete time-domain representation of acoustic propagation from one source to one receiver inside a room.

![Figure 1: Left: example of a measured room impulse response, with sound pressure plotted over time. Right: an echogram depicting the energy in decibels of the impulse response.](image)

First, consider a sound source that generates an impulsive local change in air pressure inside a room. The impulse radiates outwards spherically as a wave until encountering the nearest surface. At the surface, sound can be reflected off of, absorbed by, and transmitted through the surface to different extents. As a simple approximation, a frequency-independent reflection coefficient describes the fraction of the effective sound pressure reflecting off of the surface, though the exact behavior of the wave at the surface is frequency-dependent and depends on material properties and dimensions.

Figure 2 shows a model of the traveling waves inside a room that generate an idealized RIR, in time intervals. In (a), the impulse has propagated spherically outwards from the sound source, and partially reflected off of the top wall. The impulse response (below) is empty as no wavefront has propagated to the receiver yet.

The sound arriving first at the receiver is the *direct sound*. It travels along the shortest path and is also expected to arrive with the greatest amplitude.

Figure 2 (b) shows a moment in which the direct sound has passed the receiver position before any subsequent reflections have arrived, and appears in the impulse response as a singular peak.

Sound reflecting off of one or more surfaces inherently travels along a longer distance, thus it arrives delayed in time and at a lower amplitude due to distance attenuation and loss in the reflection. The number of surfaces that the sound sequentially reflects off of is also referred to as the *order* of the reflection path. As the possible reflection paths between the source and receiver increases with increasing
Figure 2: Geometrical model of the room impulse response in a 2D rectangular room, showing the sound source, receiver, and ideal wave propagation in time steps a-f. The impulse response recorded by the receiver is displayed underneath, in dB.
reflection order, the echo density (number of reflections per interval of time) increases proportional to the square of time during the impulse response. For a short time interval (often 50 to 80 ms) following the direct sound, early reflections arrive at the receiver. During the very early response, reflections that are sparse enough to distinguish a specific time of arrival and direction of arrival. These reflections also contribute a significant amount of energy to signal obtained at the receiver. Figure 2 (c) and (d) show further time steps where early reflections have propagated to the receiver position and appear sparsely in the impulse response.

As time continues, the echo density continues to increase and the sound field becomes more diffuse, and transitions to late reverberation. The acoustic energy in the room decays exponentially until the impulse response disappears into the background noise floor. The rate of decay determines the reverberation time $RT_{60}$, a frequency-dependent measure of the time required for the impulse response’s sound level to decrease by 60 decibels (dB).

Figure 2 (e) and (f) show the continued response, as reflections become denser and energy decays. This type of idealized room model is not very appropriate for describing the sound field of late reverberation, as will be discussed further in section 2.5.

### 2.3 The spatial room impulse response (SRIR)

While a measurement of the pressure signal at a single position is able to reveal many of the key acoustical properties of the room, it is often desirable to gather spatial information about the local sound field, namely, to capture the different directions of arrival of different sound events. By sampling the sound field at multiple nearby positions, spatial information can be obtained in a Spatial Room Impulse Response (SRIR).

SRIR measurement therefore requires a microphone array, which could vary in dimensions and complexity. Typical designs either arrange four cardioid microphones in a compact tetrahedral shape, or mount capsules on a rigid sphere [4]. Increasing the number of microphones or modifying the dimensions of the array affects the spatial resolution and bandwidth obtainable.

The spatial representation of the SRIR can also vary. Though the original microphone array signals contain all the information collected about the SRIR, these signals are not directly very useful for sound field analysis or spatial audio reproduction.

SRIRs are often encoded into Ambisonics, a spatial format based on a set of spherical harmonic basis functions, up to a fixed order. The four channels of First Order Ambisonics (FOA) are naturally interpreted as the omnidirectional, y-axis, z-axis, and x-axis components. With Higher Order Ambisonics (HOA), spatial resolution increases and $(N + 1)^2$ channels are required, where $N$ is the order. Ambisonic audio offers several advantages that make it well suitable for SRIRs. Ambisonics is able to describe the surrounding sound field in a consistent manner that remains agnostic of the recording or playback setups.

Here, the SRIR has been broadly defined as any multichannel response in the
time domain that contains spatial information about the sound field. The related term Directional Room Impulse Response (DRIR) is sometimes used interchangeably with SRIR, but may also refer to a single channel response that corresponds to a particular orientation of a directional receiver [5]. These types of DRIRs therefore could be obtained using a microphone with a cardioid-type polar pattern, or through beamforming in a certain ‘look direction’ within an SRIR. The Binaural Room Impulse Response (BRIR) is a particular case in which a two-channel response specifically corresponds to the left and right ear signals for the listener, when oriented in a particular direction. BRIRs can be recorded with dummy-head or wearable binaural microphones, as well as generated through spatial processing of an SRIR [6].

2.4 Parametric processing of SRIRs

Further processing of the SRIR can help reveal more information about the room response. Parametric approaches, in general, make use of a model of the sound field to extract features from the impulse response, which can assist in further analysis or rendering [7]. Two such approaches are Spatial Impulse Response Rendering (SIRR) [8], and its higher-order extension (HO-SIRR) [9], which primarily perform directional estimation and diffuseness estimation of sound events, over frequency bands in the short-time Fourier transform. Another is the Reverberant Spatial Audio Object (RSAO) method [10], which decomposes the impulse response into parameters expressing the direct sound, early reflection, and late reverberation in a compact form. Similar parametric techniques can be used for estimation of the geometry of image sources and reflectors corresponding to early reflection sound events [11].

The Spatial Decomposition Method (SDM), first introduced in [12], applies a simpler broadband sound field model to convert the samples of the SRIR into a set of ‘image sources’ in 3D space. It assumes that each sample of the room impulse response is attributed to exactly one sound event, with a corresponding direction of arrival (DOA), time of arrival (TOA), and amplitude. The DOA estimation is performed for each sample using a sliding analysis window, and can be done with either the time difference of arrival between microphone channels or through calculation of the pseudointensity vector [13]. TOA estimation is more precisely given by the time delay of the sample in the impulse response, provided that the $t = 0$ instant corresponds with the moment of the impulse being emitted by the source. This requires that any latency in the audio hardware I/O is removed, while the acoustic propagation time delay is kept. The image source of each sample is therefore located in the direction of the DOA, at the radial distance corresponding to the propagation time of the TOA, with an effective pressure according to the absolute amplitude of the response.

The result is that SDM can readily decompose any SRIR into point clouds of image sources that represent the temporal-spatial distribution of acoustic energy. It is useful for spatial analysis of diverse listening environments, from nightclubs [14] to car cabins [15], concert halls [16], and cinemas [17]. Responses decomposed with SDM can be to re-synthesize responses for virtual acoustic rendering over loudspeakers or headphones, in conjunction with roughness compensation and equalization [18]. The suitability of SDM for identifying and locating early reflections [19] makes it a
particularly relevant SRIR decomposition to compare with geometrical acoustical modeling as an input to blind interpolation between impulse responses, discussed in the following sections.

2.5 Geometrical acoustics

The following section discusses the principles of geometrical room acoustic modeling, known as geometrical acoustics (GA) [20]. By idealizing sound propagation as rays, and surface reflections to behave in a specular manner, it becomes possible to estimate the impulse response for sound sources and receivers in a given room geometry. In comparison to physical modeling methods based in numerical analysis, like finite-difference time-domain (FDTD) or boundary element method (BEM), GA models are often much simpler to define and compute at the expense of physical accuracy. In this section, the key assumptions underlying GA modeling are explained, and the generalized image method capable of modeling the early portion of SRIRs is detailed.

2.5.1 Model assumptions

The first key assumption made with GA modeling is that sound propagates as a ray rather than as a wave. This assumption is suitable at high frequencies, when the wavelengths are short in comparison to the dimensions of the room geometry. At lower frequencies, wave behaviors like diffraction and wave interference are more significant but poorly modeled with rays.

This assumption is appropriate if the surface is ideally flat, smooth, and rigid. Such a surface therefore would not produce scattering or lensing effects, extended reactions, or structure-borne sound transmission. Some methods for modeling diffuse scattering with GA have been proposed [21], most often, only a limited number of specular reflections are used in practice. The energy of the reflection is scaled by a reflection coefficient, which may be estimated from material properties such as the surface impedance or absorption coefficient.

Additionally, the impulse response can be modeled energetically, such that positive-valued units of sound energy propagate as rays [20]. Energetic modeling allows interference and phase shifts of the complex pressure signal to be disregarded. Despite some loss of information, the energetic impulse response can be converted back to an estimated pressure response for auralization [22].

By combining the assumptions of ray-like propagation and specular reflections off of surfaces, each reflected path between a sound source and receiver can be equivalently described by considering a virtual sound source located at the image point opposite the surface [23]. This principle extends to higher order reflections and with any planar surface geometry. The next section details a specific method for computing an energetic, ray-based, room acoustic model known as the image method.
Figure 3: A virtual source, located at the image point of the real source across the reflecting surface, generates equivalent paths of sound propagation in the free-field as the real source does with specular reflections.

2.5.2 Image method

The image method is an approach to modeling room reflections where virtual sources (VS) are positioned according to the TOA and DOA of each specular reflection [23]. By calculating the image points for all surface reflections occurring inside the room geometry, the early SRIR can be modeled instead by a set of virtual sources located in a free-field condition. Conveniently, this representation corresponds very similarly to the structure of the point clouds generated by SDM discussed in section 2.4. The image method works best for the early portion of the RIR, when reflections are sparse, and tends to poorly model diffuse the more sound field of late reverberation.

Calculating image points  The set of image points is found by recursively reflecting the sound sources (both real or virtual) across all surfaces. Each surface is defined by a boundary plane with a reflective interior face. For any given source and reflecting boundary, the resulting image point is located at twice the source-boundary distance.
Figure 5: General image method procedure: a) The image point of one source across a reflecting boundary, showing the reflected path of sound propagation. b) All valid first order image points of a source reflected across all defined boundaries, regardless of visibility. c) The valid first and second order image points, in which all sources are again reflected across all boundaries. d) The visible, valid virtual sources are located at image points where a true path of sound propagation is possible, for a particular position of the receiver.

away from the source in the direction of the boundary normal. The term image point is being used, because it is not guaranteed that valid virtual source belongs at that position. The reflection of the original sound source inside the geometry across all surfaces generates the 1st order image points, each of which is then reflected across all surfaces to generate 2nd order image points, and continuing on recursively [24]. The recursion can be ended either with a restriction on the order of sources or by setting a maximum allowable distance between image points and the listener position.

Determining virtual sources Though all possible image points are obtained recursively, not all possible image points can correspond to a virtual source. Both reflection validity and image visibility criteria must be fulfilled in order to determine that a reflection path from source to receiver indeed exists for the image point in
1. Validity: only image points produced by reflections on the interior side of boundaries are valid, and successive reflections must increase in distance from the listener. The reflection of an image point back across the surface that generated it is always invalid.

2. Visibility: only image points produced by reflections lying within the extent of the boundary are actually ‘visible’ for a given listener position. This requires checking if the line connecting the image point and the listener position intersects the reflecting boundary in question. If the intersection point is inside the boundary, then the image point may contain a visible virtual source.

In the special case of the cuboid, or ‘shoebox’ shaped room, a regular lattice of mirrored rooms is generated. Figure 4 shows the structure of such a lattice in two dimensions. Note that all image points are valid and visible for any listener position in the room.

An SRIR generated through geometrical acoustical modeling can be further processed or rendered for auralization like any other SRIR can. In order to apply the SRIR in a virtual acoustic room model, it must be decoded in order to ensure a match with the spatial audio playback system being used. In the case of out-loud spatial audio reproduction over loudspeakers, well-known methods like vector base amplitude panning (VBAP) and wavefield synthesis (WFS) may be used. For binaural reproduction over headphones, a set of head-related transfer functions (HRTFs) can be applied to generate the direction-dependent binaural signals at the listener’s left and right ears. Distinct sound objects or sound events with specific DOAs may be directly filtered with the nearest appropriate HRTFs, whereas diffuse sound field [1]

2.6 SRIR interpolation

2.6.1 Problem definition

In this thesis, the task of interpolation of the listener position between multiple SRIRs is studied. Much previous work [25, 26, 27, 28, 29] has investigated the dynamic navigation of the listener between multiple-position HOA audio streams. Some interpolation methods (such as a weighted linear combination, discussed further below) are possible solutions to either task. However, in the case of parametric methods, the information contained in spatial room impulse responses and time-varying sound scenes to be extracted will differ, and so the distinction must be made.

Here, the task of ‘blind interpolation’ between room impulse responses is defined. Each SRIR is assumed to represent a particular source and receiver configuration inside of a given room. Given any two known room impulse responses (which have a matching spatial encoding format), the blind interpolation attempts to estimate the RIR that corresponds to an intermediate source-receiver configuration. Tylka and Choueiri’s work [25] on virtual navigation of sound fields categorizes the interpolation
methods as being either **linear**, **nonlinear**, or **parametric**. Each approach addresses different aspects of the problem while suffering inherent limitations.

### 2.6.2 Linear methods

Linear methods, in general, use only the available measurement and interpolation positions and apply only linear operations (scalar multiplication, summation, or time delay) to the signals.

One example of a linear method of blind interpolation is calculating a trivial weighted combination of time-domain impulse responses. Weighted linear combination produces a smooth interpolation and can be easily performed between any number of matching time-domain spatial formats.

This method can be improved with direct sound alignment, where the IRS are first aligned by detecting the TOA of the direct sound in each, and removing it as a bulk delay. This prevents the weighted interpolation from duplicating the direct sound event or mishandling vastly different initial delay times. The TOA can be linearly interpolated and reintroduced as a fixed bulk delay.

Though the weighted linear combination method is simple, undesirable artifacts are found in the interpolated responses it generates. One undesirable effect is that the echo density doubles. Another is that clearly localizable sound events, in particular the DOA of the direct sound, is not well-interpolated in space. Instead of a smooth spatial motion, a fade between distinct DOAs occurs. While these artifacts may or may not be perceptible, they demonstrate the motivation to develop a more physically accurate method for interpolation.

Weighted linear interpolation of two measurements is suggested as a practical method for real-time processing in [30], not just along the connecting linear path but also in the surrounding area. This introduces substantial DOA error, which could be mitigated by using a three measurement mix.

In [31], wave-based acoustical modeling was used to generate a sparse subset of RIRs, whose intermediate positions are rendered using a linearly interpolation of B-format signals. It is noted that while interpolating between large changes in receiver position, through the linear interpolated impulse response’s bulk delay and temporal fine structure are incorrect, the DOA of the sound source is still mostly retained. Patricio et al. also apply a weighted linear interpolation, but suggest a method for alleviating the distance-dependent ambiguity in source localization when at an intermediate position [27].

### 2.6.3 Nonlinear methods

Nonlinear methods make use of nonlinear operations, but are otherwise formulated from the same information as linear methods. Dynamic time warping (DTW) is one such method, that has been usable to align and interpolate different sound events in the SRIR [32]. However, DTW has limitations in handling sound events that overlap or interleave their times of arrival.

Statistical methods are capable of estimating an impulse response using compressed sensing [33, 34]. The accuracy of the RIR prediction has been shown to be
very good in both time and frequency domains. However, these methods demand larger datasets and computation requirements than are applicable in the given blind interpolation problem.

### 2.6.4 Parametric methods

The first step towards physically more accurate interpolation is to recognize that distinct early reflections generally move smoothly in the virtual source space. Incremental changes to the position of the receiver therefore correspond to the incremental changes to the position of the reflection’s virtual source on the same order. Some situations may not fit this assumption: the receiver position may move to a position where the reflection path becomes occluded, or the virtual source becomes invisible. Therefore, methods are needed that attempt to map corresponding virtual sources, while tolerating the possibility that a mapping is not feasible.

Müller demonstrated using an iterative virtual source matching process to interpolate the direct sound and detected matched peaks in [35]. This process requires consideration of the feasible time of flight based on the assumed virtual source position and known coordinates of the measurement positions. Additionally, accurately locating and matching virtual sources between multiple SRIRs is a challenging task, and has been studied in detail in [36]. In the case of real measurement data, estimating TOA of sound events can be done quite accurately given the time resolution of responses, but DOA estimation incurs significant error. In [37], the TOA alone is used to determine corresponding reflection events for these reasons.
3 Background on optimal transport

Transportation theory refers to a field of mathematics focused on optimizing the transportation or allocation of resources. Though the first formulation of the optimal transport (OT) problem was presented by Gaspard Monge in the 18th century, computational optimal transport only became practical for general use in the 20th century, after mathematical and technological advances made computational OT problems directly solvable. This chapter of the thesis overviews the core concepts of transport theory in general, the key formulations of optimal transport, and introduces a particular approach to performing partial transport. Instead of repeating mathematical derivations, this chapter is meant to present the theory most relevant towards solving practical problems of transport between discretized data points.

Figure 6: Example of an interpolated one-dimensional distribution (shaded in purple), calculated using optimal transport to be intermediate between two continuous distributions (shaded in red and blue).

The core of optimal transportation is to transport a source distribution $\alpha$ to a target distribution $\beta$ with the least total cost. Each distribution is, in fact, better generalized as a measure; though it is common to primarily focus on the case of discrete probability distributions or histograms. The discrete distribution considered here is a collection of points, each with a mass $a$, and at a defined position $x$. From here onward we will refer to each point’s weight using a mass analogy. Weights must be non-negative, a natural assumption for both the mass transport analog as well as probability distributions themselves.

Transportation between distributions therefore should only move the positions of points, conserving mass, rearranging the source distribution into the target. In the general case, symmetric transportation from the target to the source is not guaranteed, but may be feasible in some formulations. The optimization of transportation is
Figure 7: Example of an interpolated two-dimensional point cloud (shaded in purple), calculated using optimal transport to be intermediate between two discrete point clouds (outlined in red and blue).

meant to follow the principle of least effort, where some ground cost ‘effort’ is incurred whenever points are moved. The optimal transport plan is that incurs the least total transportation cost.

**Example:** In a classic example of the optimal transport problem, a certain number of $n$ bakeries and $m$ cafes distributed across a city. Every day, each bakery $i \in [n]$ produces a specific amount of bread, $\alpha_i$. Likewise, each cafe $j \in [m]$ needs to receive a specific amount of bread, $\beta_j$. The aim of the problem is to determine the best way to transport bread between bakeries and cafes, i.e., find the optimal transport plan $T$. Optimality here simply means that we aim to minimize the total cost of transportation, making optimal transport often described as following the principle of least effort. Transportation cost is given by a cost function $c(x, y)$, which can be found for any bakery located at $x_i$ and cafe located at $y_j$.

### 3.1 Optimal matching

One simple approach to solving this transportation problem is to create pairs of data points. In the example of bakeries and cafes, this approach assumes that each
bakery will produce and deliver the correct quantity of bread to exact one cafe. This pairing problem is known as the optimal matching or optimal assignment problem. One-to-one pairing is precisely a bijection $\gamma$ that maps $x_i \rightarrow y_{\gamma(i)}$. This requires that there is an equal number of points in the source and target distributions, $n = m$. Here we assume that all masses are equal, or otherwise disregarded, so that any source-target pairing is valid. The ground cost of transporting a unit of mass between each pair is also defined, likely based on a distance metric between their positions. The cost between all pairs of positions defines a cost matrix:

$$C_{i,j} \overset{\text{def.}}{=} c(x_i, y_j)$$

(1)

Therefore, there exist $n!$ permutations of valid transport plans that match source and target distributions. The optimal plan is that which minimizes the total transport cost of all pairings:

$$\min_{\gamma \in \text{Perm}(n)} \left\{ \sum_{i=1}^{n} C_{i,\gamma(i)} \right\}$$

(2)

Note that the optimal matching problem is inherently unable to handle unmatched numbers of points or creating multiple mappings from a single point.

### 3.2 Monge formulation

Gaspard Monge’s formulation generalizes the transportation problem as a surjective map, where every element of the source distribution has a deterministic mapping to an element in the target distribution [38]. More precisely, the Monge transport plan $T$ is a surjection:

$$T : \mathcal{X} \rightarrow \mathcal{Y}$$

$$T : \{x_1, \ldots, x_n\} \rightarrow \{y_1, \ldots, y_m\}$$

(3)

The movement of all the points in $\mathcal{X}$ is done via the transport plan’s corresponding push-forward operator, $T_i$. The push-forward operator can be intuitively be thought of as a function that transforms the entire measure of one space to another measure in another space. The pushforward of $\alpha$ is given by:

$$T_i\alpha \overset{\text{def.}}{=} \sum_{i=1}^{n} a_i \delta_{T(x_i)}$$

(4)

The optimal Monge map is that which minimizes the total cost of transportation:

$$\min_{T} \left\{ \sum_{i=1}^{n} c(x_i, T(x_i)) : T_i\alpha = \beta \right\}$$

(5)
Note that it is possible that a surjective mapping is not feasible. This is guaranteed to be true anytime that \( n < m \) since no mass is split during the pushforward operation. Another problem is that the set of \( T \) that fulfill the condition in 5 is non-convex, making the general Monge formulation difficult to solve. Finally, notice that Monge’s surjective mapping is asymmetrical, since \( \beta \) cannot be deterministically ‘pulled-back’ to obtain \( \alpha \), in general.

\[ \gamma(\{1, 2, 3, 4\}) = \{i, iii, iv, ii\} \]

Figure 8: Left: a bijection \( \gamma \) between discrete masses in the optimal matching problem. Right: a Monge map \( T \) that pushes forward \( \alpha \) onto \( \beta \); a surjection which conserves mass.

### 3.3 Kantorovich formulation

The constraints on the Monge formulation are a significant issue when applying them to many real optimal transport problems. In the example of bakeries transporting bread to cafes, it is reasonable that a single bakery may split up its quantity bread for delivery to several cafes - but such splitting is not possible in the surjective Monge transportation. In particular, when seeking an interpolant, it is also desirable to describe a symmetric transportation plan, so that the source and target are considered interchangeably.

Leonid Kantorovich formulated a relaxed transportation plan based on a probabilistic coupling of distributions. This allows for mass at any given source point to transport to one or more target points, a behavior known as mass splitting. The transport plan is now defined with a coupling matrix \( T \in \mathbb{R}^{n \times m}_+ \), where \( T_{ij} \) describes the amount of mass transported between point \( x_i \) and \( y_j \). The transport plan again must account for all of the mass present in each distribution and respect conservation of mass during transport, and that the total mass of each distribution is matched. The set of valid coupling matrices \( \Pi \) between \( a \) and \( b \) is defined in 6 by the marginal constraints:

\[ \Pi(a, b) \overset{\text{def.}}{=} \{ T \in \mathbb{R}^{n \times m}_+ : T1_m = a \text{ and } T^\top 1_n = b \} \]  

The optimal coupling matrix is that which minimizes the total cost of transportation:
\[
\min_{T \in \Pi(a, b)} \langle C, T \rangle_F \overset{\text{def.}}{=} \sum_{i=1}^{n} \sum_{j=1}^{m} C_{i,j} T_{i,j}
\]  

(7)

Figure 9: Left: two distributions, \(a\) and \(b\), and the Kantorovich optimal transport plan between them. Right: the transport matrix \(T\) defining the mappings, where the contents of the matrix are masses to be transported between \(a\) and \(b\). The distributions \(a\) and \(b\) are shown along the columns and rows to demonstrate that \(T\) fulfills the marginal constraints of (6).

Importantly, this probabilistic transport formulation is always symmetric:

\[
T \in \Pi(a, b) \iff T^\top \in \Pi(b, a)
\]

(8)

It also makes the problem more tractable in several ways: unlike the Monge constraint in 5, the optimal coupling is found within a convex polytope. The problem is now solvable as a linear program.

### 3.4 Partial transport

In practice, many transportation problems would involve unbalanced distributions, i.e., \(\|a\|_1 \neq \|b\|_1\), as data may exist on different scales, or be noisy or unknown [39, 40]. Unbalanced distributions can be handled by first normalizing the distributions to ensure mass balance before obtaining the transport plan, but some distributions may not be suitable for normalization or rescaling. Other problems may involve differently sampled data points or prohibitively large quantities of data, making numerical solutions impractical. Various relaxations, splitting schemes, and regularizations have been proposed to handle some of these challenges [41].

One approach well-suited for handling unbalanced distributions is **partial transport**, in which the mass conservation constraint is relaxed, so that the transport plan need not account for all mass. Partial optimal transport (POT) describes transportation where only a portion of mass is described in the transport plan.
Chapel et al. [42] presented a method for performing POT by adding a dummy point to each distribution.

\[ 0 \leq \sigma \leq \min (\|a\|_1, \|b\|_1). \]  

(9)

Each dummy point has a mass set in order to balance the total mass of the distribution. The dummy points are appended to the vectors of masses, as in:

\[ \begin{bmatrix} a \\ \|b\|_1 - \sigma \end{bmatrix}, \quad \begin{bmatrix} b \\ \|a\|_1 - \sigma \end{bmatrix}. \]  

(10)

The cost of transportation to or from a dummy point is defined parametrically, as the positions of dummy points are left undefined. The cost and transport matrices are extended to \( \mathbf{C}, \mathbf{T} \in \mathbb{R}^{n+1 \times m+1} \), with the inclusion of the dummy points:

\[ \mathbf{C} = \begin{bmatrix} \mathbf{C} \\ \xi \mathbf{1}_n \\ \mathbf{1}_m \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{T} \\ \mathbf{u} \\ \mathbf{v}^{\top} \end{bmatrix}. \]  

(11)

The parameter \( \xi \) is directly interpreted as the cost of transportation to or from any dummy point. It is therefore the maximum allowable distance for transportation between real points, beyond which transportation is suboptimal. The entry \( \mathbf{C}_{n+1,m+1} \) is set to infinite cost to guarantee that an optimal solution transports no mass between dummy points, such that \( \mathbf{T}_{n+1,m+1} = 0 \). The vectors \( \mathbf{u} \) and \( \mathbf{v} \) denote the mass transport from and to a dummy point.

A partial optimal transport formulation is now defined for the extended problem:

\[
\begin{aligned}
\min_{\mathbf{T}} & \quad \langle \mathbf{C}, \mathbf{T} \rangle_F \\
\text{s.t.} & \quad \mathbf{T}_{m+1} = \mathbf{a} \quad \text{and} \quad \mathbf{T}^{\top} \mathbf{1}_{n+1} = \mathbf{b}, \\
& \quad \mathbf{1}_{n+1}^{\top} \mathbf{T}_{m+1} = \sigma
\end{aligned}
\]  

(12)

Figure 10: Left: two distributions, \( \mathbf{a} \) and \( \mathbf{b} \), and the optimal transport plan between them with dummy mappings denoted separately. Right: the extended transport matrix \( \mathbf{T} \) with dummy mappings appended to the last row and column.
The parameter $\sigma$ used in Chapel’s formulation [42] is useful if a specific amount of shared mass transport is desired, but it may be preferable to instead include this parameter in the optimization problem. By doing so, the total transportation cost can be minimized, determining the optimal amount of transported mass and the optimal transport plan.

This can be done by modifying the constraints of (12) to omit $\sigma$, $\overline{a}$, and $\overline{b}$, and instead enforce the marginal sum constraints of the original distributions:

$$\mathbf{T} \in \Pi(\mathbf{a}, \mathbf{b}) \overset{\text{def}}{=} \left\{ \begin{array}{c} \mathbf{T} \in \mathbb{R}_+^{n+1 \times m+1} ; \\ [\mathbf{T} \mathbf{u}]_m = \mathbf{a} \\ [\mathbf{T}^\top \mathbf{v}]_n = \mathbf{b} \end{array} \right\}$$

(13)

3.5 Applications

Computational methods have made solving optimization problems related to OT simpler, leading to practical applications ranging from literal resource allocation problems [39] to color transfer [41] and 2D shape modification [40]. Computing an optimal transport plan can be useful in a variety of ways, as it can be applied as a transformation of one distribution to another, or be used to generate an interpolant (as this thesis work will focus on), as well as to define a distance metric based on the total cost of transportation. In the Kantorovich OT formulation, such a metric is known as the Wasserstein distance.

A variety of efficient computational solvers have been developed to solve different OT problems, including the Sinkhorn algorithm for entropic regularized problems, and the Network Simplex method [39, 38]. The complexity of solving OT problems inherently depends on size and the dimensionality of the data being used. For certain 1-dimensional cases, solutions can be found by morphing the cumulative distribution function using direct algorithms [39]. In higher dimensions, the problems become substantially more challenging to solve. As the broader field of convex optimization has been well-studied, many solvers are now widely available. Throughout this thesis work, CVX, a package for specifying and solving convex programs in MATLAB [43], has been used for obtaining numerical solutions to optimal transport problems.

In recent years, optimal transport has been introduced for applications related to audio technology, including a frequency-domain interpolation described as a ‘generalized portamento’, usable as a real-time audio effect [44]. Computation of the optimal transport plan is performed between two 1-dimensional magnitude spectra in each buffer, using the explicit ‘Northwest corner’ algorithm. Other audio applications have also focused on OT of the audio spectrum for the purposes of morphing [45] or obtaining a spectral distance measure [46].
4 Methods

This chapter defines the proposed method for parametric interpolation of the sparse sound events found in the early portion of the spatial room impulse response.

4.1 Virtual source point clouds

The parametric interpolation method being proposed represents each SRIR as a virtual source point cloud $P$. Each distinct, localizable sound event in the early portion of the SRIR is represented by virtual source. Each sound event is associated the following attributes:

- **Effective sound pressure** $a \in \mathbb{R}_+$ is the root-mean-square sound pressure associated with the sound event, and is analogous to the ‘mass’ described in optimal transport formulations.

- **Time of arrival (TOA)** $\tau$ is the total propagation time, between when the source generates the impulse at $(t = 0)$ and when it arrives at the receiver’s position (at the origin).

- **Direction of arrival (DOA)** $\theta$ is the spherical angle from which the sound event arrives at the receiver, which may be determined from parametric decomposition of SRIR measurements or directly obtained from a GA model.

- **Position** $x \in \mathbb{R}^3$ is the cartesian vector relative to the receiver attributed to the sound event, found in the direction of the DOA, at a distance corresponding to the time of flight calculation $\tau = \frac{\|x\|}{c}$.

The set of sound events in virtual source space for a given SRIR is then referred to as a virtual source point cloud.

$$
P = \{(p_i, x_i) \in \mathbb{R}_+ \times \mathbb{R}^3, 1 \leq i \leq N\}.
$$

In the case of in-situ measured SRIRs, determining the subset of sound events and their attributes can be done through methods from literature [47, 36]. As the detection and localization of early reflections is a challenging task to perform reliably, we leave the necessary parametric processing of SRIRs out of the scope of this thesis and instead assume that a set of prominent sound events can be detected and reasonably well-localized.

This type of well-localized point cloud can instead be directly generated through geometrical acoustical modeling methods as explained in Section 2.5, and so the proposed method is tested instead with deterministic GA-based datasets.

4.2 Interpolation problem definition

The main objective of this thesis is to evaluate the physical accuracy of a novel method for SRIR interpolation.
Figure 11: Example of a configuration with a sound source inside a cuboid room, where two receiver positions are defined, 2 meters apart.

The novel interpolation method was tested using SRIRs generated by the generalized image method as discussed in 2.5.2. The interpolation task was defined by having two known SRIRs, corresponding to two receiver positions in a room, and estimating the SRIR corresponding to an intermediate receiver position along a straight line between the given pair.

The proposed method was compared against three other simpler methods: simple linear combination, aligned linear combination, and greedy nearest-neighbor mapping. The estimated SRIRs were compared against the ‘ground truth’ output from the geometrical acoustical model in time and spatial domains separately.

4.3 Partial optimal transport of point clouds

In this section, we discuss how the partial optimal transport plan can be used to interpolate between several point clouds.

In order to implement an optimal transport-based blind interpolation method between point clouds, it is expected that the transport plan defines a symmetrical coupling, where all virtual sources exhibit smoothly changing pressures and positions.
POT is used to determine the optimal transport plan $\mathbf{T}$ between two point clouds $\mathcal{P}$ and $\mathcal{Q}$, with the same formulation as presented in Section 3.4. The constraints on the extended transport plan are given by (13), which require non-negativity and marginal sum constraints. Recall that this POT formulation does not require specifying the amount of transported mass, $\sigma$, to compute the optimal extended transport plan.

The extended cost matrix $\overline{\mathbf{C}}$ is generated according to (1) and (11), against squared Euclidean distance between real points. The optimization problem (12) was implemented in MATLAB 2020b [48] using the SeDuMi solver in CVX [43].

The entries in the optimal $\mathbf{T}$ with a negligible effective pressure were rounded down to zero, to ensure that only a sparse set of prominent sound events are considered in the interpolation.

**Dummy mapping cost parameter**

The parameter $\xi$ is used to define the extended cost matrix, directly representing the cost of transportation between a dummy point and real point, regardless of the position of the real point. It is necessary to parametrize this ground cost as the position of the dummy point is left undefined. Therefore, $\xi$ also corresponds to the maximum tolerated distance of interpolation between real points. Optimal transportation favors any point-to-point mappings feasible below this distance, whereas above this distance,
Figure 13: Projection map of the directions of arrival and amplitude of two virtual source point clouds, as the receiver position smoothly moves from $s_P$ to $s_Q$. In the case of the cuboid room, all virtual sources continuously move in space without any sources appearing or disappearing.

a dummy mapping offers a lower ground cost. Though the amount of allowable point-to-point mappings is also relate to the partial transport parameter $\sigma$, we will prescribe a value for $\xi$ based on the receiver position, which allows for the iterative optimization of $\sigma$.

In the case of a moving receiver with a static sound source and fixed room geometry, all virtual sources (disregarding their visibility) inherently remain stationary relative to the room coordinate system. Therefore each appropriate virtual source mapping is expected to be tolerated up to this distance, but not above it. By setting $\xi = c(s_P, s_Q)$, this constraint is enforced and large distance mappings are not permitted. Increasing $\xi$ by a factor of 1.1 to 1.5 times is suggested to provide tolerance for imprecision involved in directional estimation.

### 4.4 Constructing the interpolated SRIR

After the optimal extended transport matrix $T$ is calculated, it can be interpreted as a prescriptive method for interpolation. Each unit of mass being interpolated therefore must have defined start and end positions in 3D space. Each intermediate
Algorithm 1 Constructing the interpolated point cloud

Require: $\mathcal{P}$, $\mathcal{Q}$, $0 \leq \kappa \leq 1$, $\mathbf{T}$

$k \leftarrow 1$
$\mathbf{T}' \leftarrow \mathbf{T}$

for $i$ where $\mathbf{u}_i > 0$ do
    if $\|\mathbf{T}_{i,:}\|_1 = 0$ then
        $r_k \leftarrow (1 - \kappa)\mathbf{u}_i$
        $z_k \leftarrow \mathbf{x}_i$
        $k \leftarrow k + 1$
    else
        $\mathbf{T}'_{i,:} \leftarrow \mathbf{T}'_{i,:} + (1 - \kappa)\mathbf{u}_i \frac{\mathbf{T}_{i,:}}{\|\mathbf{T}_{i,:}\|_1}$

for $j$ where $\mathbf{v}_j > 0$ do
    if $\|\mathbf{T}_{,:j}\|_1 = 0$ then
        $r_k \leftarrow \kappa\mathbf{v}_j$
        $z_k \leftarrow \mathbf{y}_j$
        $k \leftarrow k + 1$
    else
        $\mathbf{T}'_{,:j} \leftarrow \mathbf{T}'_{,:j} + \kappa\mathbf{v}_j \frac{\mathbf{T}_{,:j}}{\|\mathbf{T}_{,:j}\|_1}$

for $i, j$ where $\mathbf{T}'_{i,j} > 0$ do
    $r_k \leftarrow \mathbf{T}'_{i,j}$
    $z_k \leftarrow (1 - \kappa)\mathbf{x}_i + \kappa\mathbf{y}_j$
    $k \leftarrow k + 1$

return $\mathcal{R} = \{(r_k, z_k), 1 \leq k \leq K\}$

position $\mathbf{z}$ can therefore be found using the interpolation parameter, $0 \leq \kappa \leq 1$:

$$z_k \leftarrow (1 - \kappa)\mathbf{x}_i + \kappa\mathbf{y}_j \quad (15)$$

It is important to recall that the distribution dummy points lack a defined position vector $\mathbf{x}$ in this formulation of a partial optimal transport plan.

Since masses are permitted to split, it is possible that mass in the point clouds can transport in three possible manners: from real point to real point (transported mass), from real point entirely to a dummy point (vanishing/apppearing mass), or from real point partially to a dummy point (moving dummy mass). Determining the effective pressure and position of the virtual source in each situation must be considered separately before the complete interpolated point cloud can be determined.

Algorithm 1 defines the calculation of each interpolated point’s mass and position, which requires handling these possible mappings in $\mathbf{T}$:

Figure 14 shows an example of the mappings generated between two SRIRs corresponding to a moving receiver in a cuboid room. In a simple geometry like the cuboid room, the mappings generated appear similar to a bijection, and the moving dummy mass accounts for changes in the amplitudes of virtual sources.
4.4.1 Transported mass

The first type of mass is that which moves between two defined points in space, described simply as transported mass. The intermediate positions are computed directly as a linear combination of the source and target positions, weighted by the interpolation parameter. Masses found in $T$ move along a linear path between positions $x_i$ and $y_j$ with no scaling, thus fully following the principle of conservation of mass.

4.4.2 Vanishing/Appearing dummy mass

When a real point maps entirely to the dummy point $u_i = a_i$, positional interpolation cannot be defined. Therefore, the mass remains in place at the only defined position $z_k = x_i$, but linearly scales to 'vanish' as it is moved to the dummy point: $r_k(\kappa) = (1 - \kappa)a_i$. The reciprocal case of 'appearing' mass, for mappings $T_{n+1,j} = b_j$, places mass $r_k(\kappa) = \kappa b_j$ at position $z_k = y_j$.

This behavior is equivalent to a simple linear weighted combination of SRIRs, and would be the only behavior in the case where the POT parameter $\sigma = 0$ and no mass is transported.

4.4.3 Moving dummy mass

When a real point $i$ maps to both the dummy point and one or more real points, the dummy mapping does not need to 'vanish'. Instead, its mass can be linearly scaled similar to the vanishing/appearing case, while its position can be determined by the transported mass(es) originating from the same starting point. The mass $u_i$ mapped to the dummy point is scaled by $(1 - \kappa)$ and proportionally redistributed to $T'_{i,:}$ to move along with the transported masses. The reciprocal case redistributes $\kappa v_j$ to $T'_{:,j}$. 
Figure 14: Projection map of continuous interpolation performed using partial optimal transport between two receiver positions, 2 meters apart, in a cuboid room geometry.
4.5 Baseline interpolation methods

The proposed POT-based method was compared against three simpler baseline methods for SRIR interpolation.

4.5.1 Simple linear

The simplest method of interpolation uses a weighted linear combination of the two input SRIRs. The combination of effective sound pressures (without any normalization) is performed in the virtual source point cloud representation, proportional to the interpolation parameter $\kappa$:

$$\hat{R} = \{((1 - \kappa)p_i, x_i), (\kappa q_j, y_j)\}$$  \hspace{1cm} (16)

Note that no change to the DOA or TOA of any virtual source occurs during simple linear interpolation.

4.5.2 Aligned linear

In aligned linear interpolation, each input SRIR is first time-aligned by removing the bulk initial onset delay, which corresponds to the TOA of the direct sound event. The time-aligned SRIRs are then combined with a weighted linear combination just as in the simple linear method. Finally, the TOA is linearly interpolated using $\kappa$ and re-applied as a bulk delay to the RIR:

$$\tau_R = (1 - \kappa)\tau_P + \kappa\tau_Q$$  \hspace{1cm} (17)

The aligned linear method only affects the amplitude and TOA of sound events, and like the simple linear cannot generate any smooth interpolation of the DOA of virtual sources.

4.5.3 Greedy nearest-neighbor mapping

A simple, iterative, parametric mapping algorithm referred to as the ‘greedy mapping method’ is used as a baseline parametric interpolation method for comparison. As observed in figure 14, the solution to the optimization problem often appears to generate mappings between the nearest neighboring points. Therefore, it is worth testing if a different, simpler method could produce a similar result to the POT method. The greedy mapping method is an ad-hoc method, designed based on three simple goals: 1) to determine a complete mapping between any two point clouds; 2) to prioritize the lowest cost mappings available; and 3) to operate deterministically, without an optimization problem formulation.

The first step of the method is normalization of the total effective pressure of each input SRIR, which ensures mass balance. A cost matrix, which is defined is based on the same cost matrix defined in (7) without the partial OT extensions.

Next, it generates a transport plan by iteratively adding mappings to the transport plan at the lowest cost entries in the cost matrix (i.e., ‘nearest-neighbors’). This
process continues until no more unmatched mass from each distribution remains. The transport plan is then applied as in the case of the \textit{transported masses} in section 4.4 and rescaled by interpolating the original total masses.

The greedy mapping method makes no use of a partial mass transport relaxation, instead using normalization to handle mass imbalance. It also does not consider any constraints on the permissible mappings to be made. Though this exact greedy mapping algorithm does not exactly correspond to any known method for SRIR interpolation in practice, it is useful as a baseline method to demonstrate the characteristics of a naïve parametric interpolation method, without the features essential to the proposed POT method.
4.6 Evaluation methods

4.6.1 SRIR configurations

The physical accuracy of the interpolation method is tested statistically using geometrical acoustical models of room impulse responses. The generalized image method, as described in section 2.5.2, is used to calculate point clouds of virtual sources with up to third order reflections. Three example rooms are simulated; a cuboid (room in the shape of a rectangular prism, or a ‘shoebox’ room), and a general polyhedral room with trapezoidal faces in a convex shape.

![Figure 15: Geometry of the three rooms used for geometrical acoustic modeling.](image)

A fixed frequency-independent reflection coefficient of 0.707 is applied for all surface reflections, ensuring that significant reflections are modeled. The broadband reverberation time $T_{60}$ was estimated by (18), the well-known Sabine [49] equation:

$$T_{60} = 0.161 \frac{V}{S(1 - R)}$$

where $V$ is the room volume, $S$ is the total surface area, and $R$ is the average reflection coefficient.

Figure 15 shows the modeled room shapes and Table 1 lists the attributes of the example rooms.

<table>
<thead>
<tr>
<th></th>
<th>Cuboid</th>
<th>Canted</th>
<th>Trapezoidal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. dimension (m)</td>
<td>7.85 × 5.35 × 3.15</td>
<td>7.85 × 5.7 × 3.15</td>
<td>7 × 10 × 4.5</td>
</tr>
<tr>
<td>Volume (m³)</td>
<td>132</td>
<td>132</td>
<td>231</td>
</tr>
<tr>
<td>Reflection coefficient</td>
<td>0.707</td>
<td>0.707</td>
<td>0.707</td>
</tr>
<tr>
<td>$T_{60}$ (s)</td>
<td>0.44</td>
<td>0.43</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Table 1: Properties of the example rooms used to define geometrical acoustic models.

In each of the simulated interpolation tasks, the sound source was placed at a fixed random position inside the volume of the room, at least 0.5 m away from any surface. A receiver movement path inside the room was also chosen randomly,
with the constraint that the endpoints must also be at least 0.5 m away from any surface. This test method means that a general multiple-position interpolation task was evaluated without any consideration of typical receiver heights or movements.

\[
\kappa = \{0.05, 0.10, \ldots, 0.25, \ldots, 0.95\}
\]

Figure 16: Diagram of the moving receiver interpolation problem, where estimation is performed over 19 positions of \(s_\kappa\) corresponding to the interpolation parameter \(\kappa\).

The source-receiver configurations were randomly generated 100 times for each combination of room geometry and movement distance, resulting in 1800 total configurations. In each configuration, the interpolated SRIR was estimated for 19 values of the interpolation parameter, \(\kappa = \{0.05, 0.10, \ldots, 0.95\}\), using each of the 4 interpolation methods described in Section 4.5.

<table>
<thead>
<tr>
<th>Property</th>
<th>Number</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room geometry</td>
<td>3</td>
<td>{cuboid, canted, trapezoidal}</td>
</tr>
<tr>
<td>Movement distance, (|s_P - s_Q|_2)</td>
<td>6</td>
<td>{0.125, 0.25, 0.5, 1, 2, 4} m</td>
</tr>
<tr>
<td>Source / receiver setup</td>
<td>100</td>
<td>random positions</td>
</tr>
<tr>
<td>Interpolation parameter, (\kappa)</td>
<td>19</td>
<td>{0.05, 0.10, ..., 0.95}</td>
</tr>
<tr>
<td>Interpolation method</td>
<td>4</td>
<td>{simple linear, aligned linear, greedy mapping, partial OT}</td>
</tr>
</tbody>
</table>

Table 2: Tested configurations for the multi-position impulse response interpolation problem under study.

### 4.6.2 Impulse response rendering

Each resulting virtual source point cloud was then rendered to a single-channel (mono) RIR. Each virtual source contributed its effective pressure to the impulse response at nearest sample corresponding to its time-of-flight.

The resulting rendered RIRs consist only of sparse set of positive-valued, discrete impulses. Though simple to generate from geometrical acoustical modeling, these are not very suitable for direct comparison (or auralization) in this form. For example, any error metric that disregards proximity along the time dimension (e.g. the \(L_2\) norm) requires exact sound events to have exact sample alignment to avoid incurring large cost. The application of temporal smearing to the impulse response can mitigate
Figure 17: Rendered mono RIRs of three receiver positions in a cuboid room geometry, plotted in dB. The shaded regions depict the low-passed responses after effective pressure responses were convolved with a 4ms Hann smoothing function.

this issue. The next section specifies the error metric used to compare the difference between several rendered RIRs.

4.6.3 Temporal alignment error metric

A simple error metric for comparison between multiple RIRs is necessary for evaluating their similarity. The temporal alignment error is defined as the error $E$ of the RIRs was measured as a sum of the squared differences between two RIRs, after low-pass filtering them using a 4ms long Hann window $w$.

$$E(\hat{h}, h) = \frac{\int_{0}^{T} [(\hat{h} * w)(t) - (h * w)(t)]^2 \, dt}{\int_{0}^{T} [(h * w)(t)]^2 \, dt},$$

where $h$ is a ground truth RIR and $\hat{h}$ is an interpolated response and $T$ is the length of the early response.

The low-pass filtering was used to convert the sparse time series of reflection into a smoother energy envelope, which enables reductions in the alignment error without needing exact coincidence of sound events.

The choice of a 4ms long Hann window was generally motivated in part by perceptual effects of temporal masking and human sensitivity to differences in TOA, but other types of filters could just as easily be applied. Nonetheless, it is clear that the more accurately an interpolation method can estimate the TOAs and amplitudes of sound events, the lower the alignment error.
5 Results and Discussion

In this chapter, the results of the test methodology are presented and discussed.

5.1 Statistical results

The statistical performance of each interpolation method in each of the room geometries is displayed for the six receiver movement distances tested. The properties of each interpolation method are discussed with focus on temporal fine structure accuracy and qualitative features of spatial mappings.

Figure 18 shows the frequency of each method producing the most accurately interpolated omnidirectional impulse response. Each bar represents the 1900 tested configurations of a particular distance between modeled SRIRs in a particular room. In each case, the proposed POT method most frequently produced the lowest error. For distances under 4 m in the cuboid and canted rooms, the proposed method produced the lowest error in over 95% of the tested configurations.

Figure 18: Bar plot displaying the frequency that each of 4 tested interpolation methods produced the lowest RIR temporal alignment error, in 1900 tests (corresponding to 19 values of \( \kappa \) between 0.05 to 0.95 tested in each of 100 randomized setups) per each of 6 interpolation distances and 3 room geometries.
Figure 19: Temporal alignment error $E$ plotted against of interpolation parameter for a receiver distance of 0.5 m. Of 100 randomized configurations in each of the three tested room geometries, the median error is plotted in the solid line, with the 25th and 75th percentile range shown in the hatched area. The proposed partial optimal transport method is compared with three baseline methods.
Figure 20: Temporal alignment error $E$ plotted against of interpolation parameter for a receiver distance of 2 m. Of 100 randomized configurations in each of the three tested room geometries, the median error is plotted in the solid line, with the 25th and 75th percentile range shown in the hatched area. The proposed partial optimal transport method is compared with three baseline methods.
5.1.1 Cuboid and canted rooms

The results of the temporal alignment error for all of the tested cuboid and canted room configurations show that the partial OT interpolation method produces dramatically lower error that each of the three baseline methods. The median error of the partial OT method is at least an order of magnitude lower than the each of the baseline methods until the error metrics slightly converge near a distance of 4 meters.

The cuboid and canted room results are very similar for all tested distances and interpolation methods, and so they are presented together. Both linear methods perform similarly, and the parametric greedy mapping method is capable of outperforming the linear methods in the lower distance tasks. The median error of the partial OT method in tasks below 1 meter of receiver movement is negligible.

Figure 21: Median temporal alignment error of the four interpolation methods plotted against the distance of the interpolation setup in a cuboid room. The interpolation task is evaluated at the midpoint of the receiver movement ($\kappa = 0.5$). Logarithmic axes.
Figure 22: Median temporal alignment error of the four interpolation methods plotted against the distance of the interpolation setup in a canted room. The interpolation task is evaluated at the midpoint of the receiver movement ($\kappa = 0.5$). Logarithmic axes.
5.1.2 Trapezoidal room

The trapezoidal room results differ substantially from the cuboid and canted room results. The statistical results of the temporal alignment error with the trapezoidal room shows that the proposed partial OT method now shows substantial error, but is consistently lower than the baseline methods for receiver movement distances over 0.5 m. The error for all methods at distances under 0.5 m is greatly increased compared to the cuboid and canted room configurations.

Figure 23: Median temporal alignment error of the four interpolation methods plotted against the distance of the interpolation setup in a trapezoidal room. The interpolation task is evaluated at the midpoint of the receiver movement ($\kappa = 0.5$). Logarithmic axes.
5.2 Discussion

5.2.1 Statistical results

The cuboid room task is somewhat simple to solve, as all virtual sources maintain the exact same relative configuration with no appearing or disappearing sources. For a completely deterministic geometrical acoustic model, this means that every virtual source could be accurately spatially interpolated using a bijective mapping, and the necessary rescaling of the spherical wave distance attenuation for each virtual source. Thus, the parametric methods were found to be capable of outperforming the linear methods. The greedy mapping method can generate the correct bijective mapping at smaller distances. When the distance is great enough to move virtual sources such that they are no longer closest to their bijective pair, then the nearest-neighbor assignment strategy breaks down. The partial OT performs better, even in the greater distance conditions, because the optimization problem inherently considers reducing the distance cost of the entire set of virtual sources.

The canted and trapezoidal room tasks are inherently more difficult to resolve than the cuboid room, as there is the possibility that virtual sources may appear or disappear for the receiver’s position along the interpolation path, due to the visibility conditions of geometrical acoustic modeling. Though the canted room is capable of producing situations in which the receiver moves in and out of the regions of visibility of some virtual sources, these occurrences are fairly uncommon give that only one wall was canted with a low angle.

5.2.2 Linear methods

The main limitation of using a simple weighted linear combination of virtual sources (or pressure samples) to perform interpolation is that features are only smoothly change in amplitude, but not in space or time. Failing to smoothly interpolate the TOA of the sound event is especially problematic. Crossfading between doubled direct sound events may degrade the quality of the time-domain transient response, generate strong comb-filter effects in the frequency domain, unusual spatial localization due to the precedence effect.

The aligned linear method avoids the doubled direct sound issues, but both linear methods still approximately double the echo density of the impulse response, as seen in Figure 28 (b). This effect causes interpolated responses to have a notably different temporal fine structure than the sparse input RIRs. These differences are not alleviated by temporal smoothing, and so the linear methods were consistently outperformed by the partial OT method at distances 0.25 m or greater. Though the accuracy of the direct sound TOA is generally improved over the simple linear method, the target TOA is not necessarily well estimated by the linear combination of multiple delay. As the VS of the direct sound event moves relative to the receiver, the it distance may be minimized at an intermediate position, causing the TOA of the intermediate position to be shorter than either of the endpoint TOAs. Figures 27 (a) and 28 (a) both show examples where the intermediate position has a direct sound TOA that is visibly different from the averaged TOA. Improving temporal
alignment further is therefore possible by considering three-dimensional movement, which the parametric methods are better suited for.

5.2.3 Parametric methods

The main improvement offered by the parametric mappings methods tested here is the continuous movement of virtual sources in space and time. A qualitative discussion on the characteristics of the virtual source mappings of each method are presented next, with consideration of both temporal alignment and the assignment of spatial mappings. Figures 27 (c) and 28 (c) show examples of omnidirectional impulse response renderings where the two parametric methods are displayed. Both parametric interpolation methods are able to better estimate the direct sound TOA than the linear methods.

The greedy mapping algorithm of the nearest-neighbor method is able to successfully couple many physically corresponding pairs of virtual sources, but also produces mappings across large distances, with no physical meaning. These large mappings are needed to fully define the transportation plan without a partial transport relaxation, and to handle the amplitude changes of corresponding virtual sources due distance attenuation. Though additional constraints and relaxations can easily be included in such a parametric mapping algorithms, the proposed method of partial OT is partially motivated to avoid the need for extensive parametrization.

The partial OT method creates many appropriate mappings as well as some confusions, as shown by comparison of Figures 25 and 24. The confused mappings usually involve a mapping of a virtual source that either appears or disappears along the ground truth interpolation path. Also, when a source is near a wall or corner, its first order reflections produce virtual sources closeby, which may cause lower-cost mappings to be created between virtual sources not meant to be matched. These physically inaccurate mappings are, however, constrained to shorter distances, and are assumed to have a minor perceptual impact.

Achieving perfect physical accuracy of the spatial positions of virtual sources is a reasonable target, but should not be regarded as necessary for effective multi-position interpolation. In practice, virtual sources estimated from real SRIR measurements should be assumed to include substantial DOA error, particularly for higher order reflections. Though the perceptual relevance of inaccurately mapped virtual sources of either parametric method is not evaluated in this thesis work, humans’ sense of spatial hearing will not sharply localize the virtual sources apart from the direct sound event [50]. Thus, conclusions about the suitability of an SRIR interpolation method should consider not just physical accuracy, but also the smoothness that features change in the time, space, and pressure dimensions.

5.2.4 Limitations of partial OT interpolation

The partial OT method is shown to generally outperform all of the baseline methods tested. However, the method may encounter computational limits not faced by other methods.
First, the three baseline interpolation methods tested were all far simpler in terms of computational cost, making direct comparisons of the interpolation method accuracies somewhat ambiguous to interpret. Quantifying the complexity of the presented algorithms remains out of the scope of this thesis. Similarly, discussion about practical implementation of the partial method OT for real-time 6DoF applications are not considered here.

Finally, solving the optimal transport problem requires that the size of the transport matrix is not excessively large. While the third order image method implementation used in this thesis typically required fewer than 100 virtual sources (and as few as six salient early reflections can produce effective spatial audio rendering [51]), it is impractical to considerably increase the number of virtual sources. This makes the direct implementation of partial OT interpolation on large numbers of virtual sources, such as those produced by scattering models [21] or obtained with SDM, impractical without further parametric processing or reflection identification [36].

5.2.5 Future work and outlook

The primary work of this thesis was to develop and test a novel method for performing multiple-position SRIR interpolation, leaving many avenues for further work in this
One topic in need of further study is the choice of the cost function $c$ used in the POT formulation. The current method uses a squared Euclidean distance metric, implying that mappings between (real) virtual sources across smaller distances are more optimal. However, in the static source, moving receiver interpolation problem studied, all matched virtual sources should optimally move a distance equal to the distance between receiver positions. This suggests that a cost function of $c(x, y) = (\|x - y\|_2 - \|s_P - s_Q\|_2)^2$ may be better suited for the physically accurate solution. Other modifications to the cost function to other interpolation problems may seek to reduce the number mappings, or differentiate between the unique and non-unique dummy mappings that affect how the interpolation algorithm from Section 4.4 behaves.

Other geometrical acoustical model setups not yet tested include cases where the direct or reflected sound paths may be occluded, such as in a non-convex shaped room geometry. Transitions between coupled rooms are a particularly challenging task for SRIR interpolation [52], in which a POT-inspired method might stand to offer improvement.

Other types of interpolation problems may also be of interest, such as the case of a moving sound source with a static receiver position. The movement of the sound source is in some ways reciprocal to the moving receiver problem, but also
causes shifts in the relative positions of virtual sources that cannot be resolved with a translation in a single direction. The interpolation between changes in the room geometry, or the SRIRs of entirely different spaces may also be of interest for future study.

Another key area for future work is to apply the method on real measured SRIRs instead of geometrical acoustic modeling. Doing so should make use of existing parametric spatial audio techniques to determine likely positions and amplitudes of virtual sources while separately handling the late reverberation and other diffuse components. In this case, the interpolation method should also be tested for robustness to localization errors, noise, and different measurement methods.

Further work should also seek spatial renderings of the interpolated SRIRs more suitable for auralization. Producing such impulse responses would certainly require modeling the temporal extent of virtual sources beyond the current single sample impulse. This could involve the addition of frequency-dependent modeling of surface reflection filters, scattering effects, the source and receiver responses. Work towards perceptual evaluation of SRIR interpolation methods may also involve a real-time implementation of the interpolation method from 4.4 in conjunction with binaural rendering for a 6DoF head-tracked device.

More generally, the utility of optimal transport is extremely broad, not only in calculating couplings or interpolants, but also in defining distance measures and
Figure 27: Comparison of omnidirectional impulse responses in the cuboid room, showing the ground truth RIRs of the blind interpolation problem (a), the target compared against the linear methods (b), and the target compared against the parametric methods (c) for an example where the given SRIRs are 2 meters apart and the target receiver position is located at the midpoint ($\kappa = 0.5$).
Figure 28: Comparison of omnidirectional impulse responses in the trapezoidal room, showing the ground truth RIRs of the blind interpolation problem (a), the target compared against the linear methods (b), and the target compared against the parametric methods (c) for an example where the given SRIRs are 2 meters apart and the target receiver position is located at the midpoint ($\kappa = 0.5$).
loss functions. The possible applications of optimal transport in audio technology ranges from inharmonic pitch detection [53], to source localization [54], to spectral morphing [45, 44], and further applications are anticipated to continue emerging.
6 Conclusion

This thesis work examined the task of ‘blind interpolation’ between two spatial room impulse responses (SRIRs), in which information about the room geometry is not explicitly known. A review of existing SRIR interpolation methods demonstrated that simple linear and nonlinear methods are practical for use in many cases, but encounter limitations in achieving physically accurate interpolations. Parametric methods, in which a model of the acoustic scene is utilized, enable more physically accurate results, though often requiring highly parameterized, iterative algorithms to determine appropriate mappings between sound events. The main aims of this thesis work were to introduce a novel parametric interpolation method, and demonstrate that it can interpolate the spatial and temporal features of early reflection sound events in a more physically accurate manner than linear baseline methods, while avoiding overly complex model parametrization.

The proposed method operates on a spatially distributed virtual source representation of the early portion of the room impulse response. A partial optimal transport (POT) formulation is used determine appropriate mappings between sparse, localizable sound events that correspond to the direct sound and early reflections. A transportation cost given by the squared euclidean distance between virtual sources, and a dummy cost set according to the squared distance of the largest tolerable mapping is defined. The optimization problem is solvable as a linear program, and the resulting optimal transport plan was used interpolate between SRIRs.

The performance of the proposed method was tested statistically using SRIRs produced by a geometrical acoustical model, in three model room geometries. The SRIRs were generated with an omnidirectional sound source at a fixed position, captured at two different listener positions inside the room. The intermediate SRIRs along the straight line path between the two positions were estimated from the given endpoint SRIRs by testing four interpolation methods: a direct weighted linear combination of impulse responses, a weighted linear combination of impulse responses after separately time-aligning and interpolating the direct sound events, a greedy mapping algorithm to iteratively couple virtual sources in order of spatial distance, and the proposed method where couplings are defined by partial optimal transport. Estimated RIRs were compared against the model simulated RIR, and a temporal alignment error metric was computed over the rendered, time-smeared omnidirectional RIRs.

Statistical results of randomized source-receiver configurations showed that the partial OT method produced the lowest median temporal alignment error in all tested room geometries and distances. In simple room geometries, where reflection paths rarely appear or disappear, the partial OT method consistently outperformed the baseline methods by an order of magnitude. The partial optimal transport plan produced interpolations that were more physically accurate and introduced fewer notable temporal and spatial artifacts than the baseline methods. It was also found that the characteristics of any interpolant based on partial optimal transport are mainly dependent on the definition of the cost matrix. Further studies are suggested where different POT problem formulations are used to test more challenging
configurations of the room geometry, source and receiver. Future work should also aim to extend the method to include real measured SRIRs, rendering suitable for auralization, and perceptual testing.
References


