Low-complexity computer simulation of multichannel room impulse responses

Low-complexity computer simulation of multichannel room impulse responses

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op vrijdag 22 november 2013 om 10:00 uur
door

Jorge Abraham MARTÍNEZ CASTAÑEDA

Master of Science in Advanced Informatics, Universitat Autònoma de Barcelona

geboren te México D.F., México

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. R.L. Lagendijk

Copromotor:
Dr. ir. R. Heusdens

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. R.L. Lagendijk,	Technische Universiteit Delft, promotor
Dr. ir. R. Heusdens,	Technische Universiteit Delft, copromotor
Prof. dr. ir. G.J.T. Leus,	Technische Universiteit Delft
Prof. dr. R.M. Aarts,	Technische Universiteit Eindhoven
Prof. dr. ir. D. de Vries,	Rheinisch-Westfaelische Technische Hochschule Aachen, Deutschland
Prof. Dr.-Ing. W. Kellerman,	Friedrich-Alexander-Universität Erlangen-Nürnberg, Deutschland
Prof. dr. ir. M. Moonen,	Katholieke Universiteit Leuven, België
Prof. dr. ir. A.J. van der Veen,	Technische Universiteit Delft, reservelid

The work described in this thesis was financially supported by STW and Google Inc.

In loving memory of my father, P. C. M.

Summary

The “telephone” model has been, for the last one hundred thirty years, the base of modern telecommunications with virtually no changes in its fundamental concept. The arise of smaller and more powerful computing devices have opened new possibilities. For example, to build systems able to give to the user the illusion of being talking to the remote party as if both where in the same place. To achieve this still many challenges have to be overcome. In this thesis, a part of the acoustical signal processing problem is treated.

To acoustically create the illusion of presence, fast and accurate control over the sound field in a room is required. The sound field given one or more sources is subject to different acoustical phenomena, such as reflection and diffraction. Because of these, to model or estimate the sound field in a room is in general a difficult task. In particular acoustical reflection poses an important challenge. The sound field reflects on the walls, ceiling and floor and a moment later those reflections reflect again, and later these reflect again. This recursive process makes the number of reflections as a function of time to increase, in general, at a geometric rate. To synthesize an artificial sound field in real time, one has to be able to model these reflections fast and accurately enough. In this thesis a fast algorithm to model the sound field in box-shaped rooms is proposed.

Part one of this thesis begins with an introduction to the topic, here the different acoustical phenomena of interest are explained, and the concept of room impulse response (RIR) is introduced. The RIR is defined as the time-domain signal sensed at a receiver position as generated by a point source that emits an impulse. Assuming a linear time-invariant (LTI) model, if the point source emits not an impulse but an arbitrary signal, the actual sound field at a given observation location can then be modeled as a convolution of the source signal with the RIR. Moreover, since we are assuming a linear model, the sound field generated by an arbitrary number of point sources emitting arbitrary signals can be easily computed once the RIRs from the locations of the sources to the observation locations are known. Efficient computation of the RIR is therefore of theoretical and practical interest. Consequently, this part concludes with a summary of the most prominent algorithms to simulate the RIR.

Part two of this thesis contains the relevant papers that make up this work. The analysis is given first for the case of fully reflective walls. It is noted that in this case all the acoustical reflections can be modeled by a set of virtual sources following a periodic structure over a lattice. The whole set of virtual sources is generated by the repetitions of a small set of sources called "the mother sources". On the other side, the Poisson summation formula establishes the relation between periodicity and discretization under the Fourier transform. Relating these concepts, it is shown that by carefully discretizing the spectral representation of the RIR of the mother sources in free-field, the exact periodic structure that makes up the sound field in a room can be obtained. This is the key idea behind the proposed method. Carefully discretizing all domains, and making use of the fast Fourier transform (FFT), a fast multichannel RIR simulation method is obtained. Unfortunately this idea only works for fully reflective walls. By allowing the walls to have constant complex-valued reflection coefficients (this is, to model absorption and phase shift at the walls) the sound field of the set of virtual sources is not anymore periodic. A generalization of the Fourier transform is then introduced. First, a generalized Poisson summation formula is derived. This formula relates discretization in the generalized Fourier domain to a geometrically weighted periodic summation in the reciprocal domain. Basic properties of this transform are derived, its application to non zero-padded linear convolution is derived, but moreover a fast implementation, called the generalized fast Fourier transform (GFFT), is given. The proposed method is then extended to account for walls with constant complex-valued reflection coefficients. It is shown that by separating the sound field of the mother sources into its orthant-sided parts (the analogue of the single-sided parts of a function of a scalar variable), the sound field inside a room can be expressed as a sum of geometrically weighted sound fields generated by the periodic set of virtual sources. This summation is then related to a sampling condition on the generalized spectrum of the orthant-sided parts of the sound field of the mother sources. Using the GFFT the method simulates the RIR given a source at a dense set of spatial positions with very low complexity. In the experiments a comparison with a model called the mirror image source method (MISM) is given. In one scenario, the time the MISM would take to compute the RIR at a dense set of positions is estimated to be about one and a half years. The newly proposed method computes the RIR at all positions in only forty-eight minutes. This shows the contrasting difference in computational complexity, making the new method an important step on the road to simulate realistic sound fields in real time.

Samenvatting

Stel je een telecommunicatie systeem voor waarmee je het gevoel hebt dat je echt met de andere partij in dezelfde ruimte aan het praten bent.
Dit proefschrift is een kleine stap in die richting.

Telecommunicatietechnologie nadert een cruciaal moment. Het honderddertig jaar oude “telefoon” paradigma begint te veranderen. Het doel voor komende technologieën is om realistischere systemen te ontwikkelen. Systemen waarmee de verbonden partijen het gevoel hebben dat ze samen in dezelfde ruimte zijn. Dat is zeker een uitdaging voor meer dan één tak van wetenschap. In dit proefschrift wordt alleen de geluidstechnische kant van het probleem besproken.

Nauwkeurige en snelle controle over het geluidsveld is het doel. Dit kan alleen bereikt worden als het geluidsveld van de zender snel genoeg kan worden ingeschat en realistisch worden nagebootst voor de ontvanger. Het probleem zijn de reflecties (reverberatie en echo) in een kamer. Wiskundige modellen om een geluidsveld in een kamer te kunnen bepalen zijn reeds sinds het begin van de twintigste eeuw bekend. De oplossingen van dit soort modellen zijn echter of niet bekend (behalve voor de eenvoudigste gevallen), of zeer ingewikkeld te berekenen met behulp van computers. In dit proefschift wordt een nieuw, snel algoritme om het geluidsveld in rechthoekige kamers te kunnen simuleren geintroduceerd.

Deel één van dit proefschrift verklaart het wiskundige concept van “room impulse response” (RIR), waar dit proefschrift om draait. In grote lijnen is de RIR het geluidsveld dat ontstaat door een puntvormige geluidsbron die een puls uitzendt. De RIR is een wiskundige afbeelding van de bron en de reverberatie in een kamer. Het is een essentiële werkwijze want daardoor kan elk ander geluidsveld gemakkelijk berekend worden. In hoofdstuk twee wordt, daaropvolgend, een samenvatting van gangbare computer gebaseerde methodes om de RIR te kunnen berekenen gegeven.

Deel twee van dit proefschrift volgt een andere stijl. Hier zijn kopieën van de gepubliceerde artikelen verbonden aan dit promotieonderzoek toegevoegd. In het

eerste hoofdstuk wordt het hoofdconcept achter de nieuw methode uitgelegd. Het idee is simpel. Elk geluidsreflectie in een kamer wordt als een virtuele bron gezien die zich buiten de kamerruimte bevindt. Reflecties van reflecties maken dus een (in principe oneindige) collectie van virtuele bronnen. In een rechthoekige kamer vormt de collectie van bronnen een periodiek patroon. Sommige computer algoritmes proberen direct de virtuele bronnen te modeleren, maar hoe meer reverberatie hoe meer bronnen gemaakt moeten worden. Zulke algoritmen vragen om zeer omvangrijke berekeningen. Aan de andere kant vertelt de Fourier theorie (de wiskundig theorie over de relatie tussen tijd en frequentie) dat als een functie in de tijd domein een periodiek patroon vertoont, dan wordt haar frequentie representatie door een discrete functie gegeven. Het geluidsveld is een functie van tijd en ruimte. Het domein van deze functie wordt dus tijdruimte genoemd. Je begint met de frequentie representatie van het geluidsveld (de RIR) zonder kamer, vervolgens bemonster je het op een exacte manier om een discrete functie te maken. Als je terug naar het tijdruimte domein gaat, krijg je automatisch het periodieke patroon van de virtuele bronnen, en dus de RIR in de kamer. Dat is het hoofdconcept. Bovendien is het via een bekend algoritme, "the fast Fourier transform" (FFT), snel om tussen de frequentie en tijdruimte te wisselen. De nieuwe methode maakt dus gebruik van de FFT om zeer snel het geluidsveld in de kamer te berekenen.

Helaas werkt dit idee alleen als het geluidsveld van alle virtuele bronnen perfect periodiek is, met andere woorden, alleen als de muren geen absorptie vertonen. In het praktijk is dit nooit het geval. De muren absorberen altijd een deel van de geluidsenergie. Om de methode te kunnen uitbreiden, wordt in hoofdstuk twee een uitbreiding van de Fourier theorie voorgesteld. De "generalized" Fourier theorie vertelt dat als je een discrete functie in dit domein hebt, dan toont het in het tijdruimte domein een periodiek patroon met een exponentiële demping aan. Verder is een snelle implementatie van de generalized Fourier transformatie (GFFT) ontwikkeld.

Hoofdstuk drie gaat over een diepere analyse van de generalized Fourier theorie en maakt de eerste verbindingen voor de toepassing van de nieuw RIR simulatie methode.

In hoofstuk vier is het concept van de snelle RIR simulatie methode gecombineerd met de generalized Fourier theorie om een algemene methode te ontwikkelen. Via de GFFT is een snelle simulatie van de RIR in rechthoekige kamers mogelijk. In de experimenten, met een computer algoritme dat direct de virtuele bronnen probeert te modelleren zou het anderhalf jaar kosten om het geluidsveld te berekenen. De nieuwe computer algoritme doet deze berekening in slechts achtenveertig minuten. Dit is een groot verschil in berekeningssnelheid. Hiermee is een belangrijke stap gezet naar nauwkeurigere en snellere controle over het geluidsveld.

Resumen

El paradigma telefónico, en el cual se han basado las telecomunicaciones en los últimos ciento treinta años, está cambiando. La nueva meta tecnológica es la creación de sistemas de telecomunicación que proporcionen al usuario la sensación de encontrarse en el mismo lugar que sus interlocutores, con un realismo que permita crear la ilusión de *telepresencia*. El trabajo descrito en esta tesis constituye un paso importante en esta dirección, centrándose en el aspecto acústico del problema.

Para crear la ilusión de telepresencia es necesario simular el campo de sonido en tiempo real. El principal problema al caracterizar el campo de sonido son las reflexiones acústicas; es decir, la reverberación que existe en cualquier espacio cerrado. En la actualidad, los métodos para calcular la reverberación son excesivamente complejos. El trabajo aquí descrito propone un nuevo algoritmo para simular el campo de sonido en habitaciones rectangulares en un tiempo muy reducido.

La tesis está estructurada en dos partes. La primera presenta una introducción al problema. En el primer capítulo se describen conceptos importantes, por ejemplo, el de respuesta al impulso de la habitación, *room impulse response* (RIR). De manera general, la RIR es una representación matemática de los fenómenos acústicos que ocurren en la habitación, incluida la reverberación. Es una herramienta muy útil, ya que el campo de sonido se modela muy fácilmente a partir de la RIR. El segundo capítulo presenta un análisis de los métodos más relevantes para simular la RIR en una habitación. Entre éstos, el algoritmo conocido como *Mirror Room Image Method* (MISM) se usa como punto de comparación, ya que es un algoritmo ampliamente estudiado y utilizado.

La segunda parte de la tesis sigue un estilo diferente. Los capítulos asociados son reproducciones de las publicaciones científicas que conforman este trabajo. El tercer capítulo expone la idea principal del nuevo algoritmo, que se resume a continuación. Cada una de las reflexiones acústicas que ocurren sobre las superficies de una habitación se expresa como una copia de la fuente de sonido. El proceso seguido es similar a la forma en la que se simula la imagen de un objeto frente a un

espejo, posicionando una copia del objeto al otro lado del espejo. Estas copias de la fuente original constituyen un conjunto de *fuentes virtuales* (reflexiones), y cada una de estas reflexiones genera a su vez otras. Para simular todos los posibles órdenes de reflexión se requiere un número infinito de fuentes virtuales. En una habitación rectangular, esta colección de fuentes virtuales forma un patrón periódico. Algunos algoritmos (por ejemplo el MISM) calculan la RIR usando un conjunto finito de fuentes virtuales, descartando aquellas cuyo efecto sobre el campo de sonido es poco apreciable. Sin embargo, cuanto mayor es la reverberación, mayor es el número de fuentes virtuales que deben considerarse. Por lo tanto, la complejidad de dichos algoritmos es alta.

La teoría de Fourier define que si una función en un dominio es *periódica*, su representación en el dominio transformado – el dominio de la frecuencia – viene dada por una función *discreta*. El campo de sonido – o de manera equivalente la RIR – es una función en el dominio espacio-tiempo. En ausencia de reflexiones ésta función no es periódica, pero si suponemos que las superficies de una habitación son totalmente reflejantes, en este caso el campo de sonido es periódico. Inicialmente, el algoritmo propuesto representa en frecuencia el campo de sonido sin reflexiones. El siguiente paso consiste en calcular un conjunto discreto de muestras de esta función. Esta discretización en el dominio de la frecuencia induce el patrón periódico que forman las fuentes virtuales. De esta manera se generan, en un solo paso, todas las reflexiones acústicas que constituyen el campo de sonido en la habitación. Esta es la idea fundamental del nuevo algoritmo.

Desafortunadamente, el método anterior sólo funciona si el campo de sonido es perfectamente periódico, es decir, si las superficies de la habitación son totalmente reflejantes. Pero en realidad nunca se da este caso; en el momento en que las ondas de sonido alcanzan las superficies de una habitación, éstas absorben parte de la energía sonora. Por este motivo, en los capítulos cuarto y quinto se introduce una generalización de la teoría de Fourier, que permite extender la idea fundamental del algoritmo. La RIR – o el campo de sonido – se calcula siguiendo el mismo procedimiento descrito en el tercer capítulo, pero en este caso, se usa la teoría generalizada de Fourier para calcular las reflexiones acústicas en una habitación con superficies absorbentes. Para avalar la rapidez del nuevo método, uno de los resultados finales de este capítulo compara el algoritmo MISM con el algoritmo propuesto. En el escenario analizado, se estima que el MISM tardaría más de un año y medio en simular el campo de sonido, mientras que el nuevo algoritmo lo calcula en solo cuarenta y ocho minutos, demostrandose así la impresionante reducción de complejidad. El trabajo en esta tesis constituye un paso importante en la búsqueda científica por simular el campo de sonido en tiempo real y proporcionar la ilusión de telepresencia.

Contents

PART I: INTRODUCTION

1

Preamble

1.1 Motivation

The invention of the telephone more than 130 years ago marked a major milestone in the history of human communication. After all these years, it is still changing the way people communicate and interact with each other. Telephonic technology has improved to such extent that today digital solutions not only provide instant and reliable connections, but also remote access from virtually any place on earth. The telephonic paradigm, however, has not changed much. Communication is still based on a central device that is either a static base station, or has to be carried or worn. Therefore it provides a different, unnatural communication experience compared to on-site live interaction [1].

Human technologies are close to achieve the next milestone in communication experience, closing the (audio-visual) gap between on-site interaction and telecommunication. These next-generation technologies will be capable of giving the illusion of being at a remote place with great fidelity. Users will be able to roam freely, having the experience of talking to the other peers in a normal conversation, without the need to carry or wear any "intrusive" devices [1–3].

In order to create the illusion of presence and the ability to roam freely through the place, accurate acquisition and control of the sound fields is necessary. Several digital signal processing (DSP) technologies have appeared [4–7] that aim to create and record 3-D acoustic scenes using sets of microphones/loudspeakers (input/output channels). These can be installed, for example, on the walls, floor and ceiling in order to be non-intrusive to the user. However, to achieve a realistic experience in the whole space, large amounts of microphones/loudspeakers are needed [8]. To construct such sound field rendering systems is becoming a possibility thanks to current research in wireless sensor network (WSN) technologies [9], which focus on the efficient deployment of large amounts of devices. The next step is then to enable real-time, true-to-life communication on those systems. This is fundamentally more difficult that the creation or recording of immersive sound fields, since not only *joint* input-output sound field control is necessary, but also other acoustic problems arise.

As an example, imagine a room where many, possibly thousands, of small loudspeakers and microphones have been installed on the walls, floor and ceiling. These devices have been smartly hidden, so the system is fully transparent to the users. In order to achieve a natural communication experience, the sound fields of the far-end party must be reproduced in the room and, at the same time, the sound fields of the users of the system have to be recorded. If both the loudspeakers and the microphones are operating at the same time, the far-end sound field would be fed back to each of the microphones, resulting in a very disturbing echo at the far-end party in the best case and, up to instability of the whole system in the worst case. The signals from each loudspeaker to each microphone have then to be cancelled. If thousands of devices are being used, this easily gives rise to millions of possible signal combinations that have to be cancelled. However, the amount of signals is actually not the difficult part. The problem is particularly challenging because of the *reverberation* that is introduced by the acoustic properties of the room; this is, the sound field created inside the room will be modified by all the possible sound reflections induced by the walls, floor, ceiling and objects present in the room. Moreover, any small changes in the positions of the objects, the temperature and density of the air and many other phenomena cause, in general, complex changes to the sound field. The modified sound field is what is captured by the microphones. This particular problem is known as acoustic echo. *Acoustic echo cancellation* (AEC) is therefore addressed using data-driven approaches (in the form of *adaptive filters*), since model-based methods are unsuitable to characterize and track the underlying acoustical dynamics to the required degree of precision. For the case of one microphone and one loudspeaker the acoustic echo problem has been solved satisfactorily [10, 11], but in the multichannel case, starting with two microphones and two loudspeakers, the problem becomes more challenging. Here the cross-correlations of the loudspeaker signals prevent the identifiability of the unique reverberation paths from the loudspeakers to the microphones [12], and the adaptive filters converge to solutions that are dependent on the characteristics of the loudspeakers signals. Therefore, any changes in the acoustics of the transmission room result in discontinuous (often quite audible) errors, and the filters have to reconverge [11]. Research addressing the multiple-input multiple-output (MIMO) AEC problem has, consequently, dealt with the decorrelation of the loudspeaker signals. The addition of non-linear distortions [12] and uncorrelated noise [13], were early approaches. These methods produce, however, either objectionable perceptual distortion or increase complexity and introduce delay, which makes them unattractive options. A more recent frequency-selective phase modulation approach, used in conjunction with a perceptual model, keeps the distortion below the human perceptual threshold and performs with low complexity in a 5-channel surround sound scenario [14]. However, systems that aim to render 3-D sound fields using large amounts of loudspeakers, are very sensitive to phase modulations [15]. State-of-the-art research on MIMO AEC has therefore started to investigate *model-motivated, data-driven* approaches. In this category, one prominent method employs a spatio-temporal decoupling of the degrees of freedom in the AEC system using *orthogonal* basis functions. The selected basis functions are solutions to the wave equation and the approach is therefore named *wave-domain* adaptive filtering (WDAF) [16–18]. This scheme is scalable, and allows for a trade-off between low complexity and high accuracy. Another approach exploits the capacity of the 3-D sound field rendering system to synthesize quiet zones in the enclosure (normally up to a certain low-frequency). By placing the microphones of the system inside these quiet zones, and performing phase modulation on frequencies above the spatial resolution of the system, scalable and accurate MIMO AEC is achieved [15]. The development of fast and scalable models to characterize the acoustical dynamics in a room, can therefore result in improved schemes to solve problems that are traditionally addressed using purely data-driven approaches such as AEC.

Another interesting application where the sound field in a room has to be modeled is immersive virtual gaming. So here we are again in our special room equipped with a large amount of loudspeakers that can create immersive sound fields. The users of the room start playing an immersive game where an audio-visual experience is given to them such that they have the feeling of really being in the game field. The game will take the users through different scenarios. For example, at one moment the users could be at an open location such as a park while at a next moment they could be inside a room. In order to create a realistic experience, the system has to reproduce different acoustic scenarios in real time with transparent fidelity in a large zone (ideally the whole room), since the users should be able to roam freely through the room. Therefore, not only real-time simulation of the full sound field in a room is needed, but also *acoustic room compensation* is necessary to equalize the acoustic characteristics of the room in order to render the desired (virtual) sound field [19].

The main motivation of this thesis is to advance the state-of-the art of current computer based acoustic modeling methods (for small-rooms acoustics), paving the way towards the creation of next-generation communication technologies, where large amounts of input-output acoustic channel configurations are common, and fast, highly scalable algorithms are a must.

1.2 The room impulse response

Our perception of sound is continuous. Continuous sound fields, however, cannot be modeled within a digital computing framework. Instead we approximate them using discrete samples, both in time and space. The more samples we use, the more accurate the approximation. Sampling the sound fields in time results in the well-known approach used in single-channel digital signal processing (DSP) [10, 20, 21]. Sampling in space is as simple as to measure the sound field at discrete locations in space. Positioning a microphone at a certain location is equivalent to taking one spatial sample. Therefore, every time a reference to the sound field in a room is made it actually means the (discrete-time) signals measured or reproduced at a set of receiving or source locations.

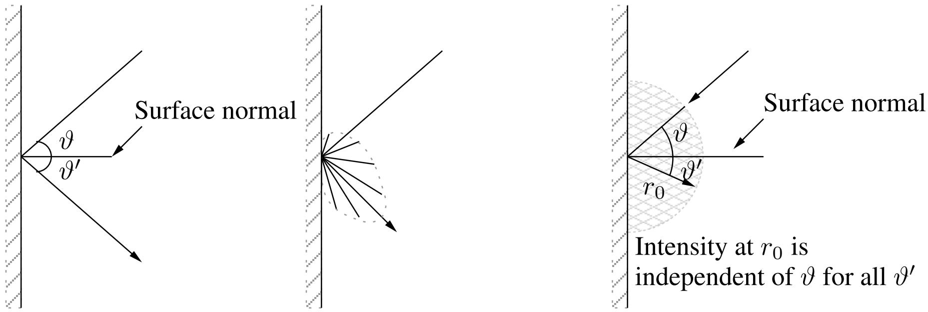
Given a source of sound and an observation point in a room, a mathematical description of all possible sound paths from the source to the receiver, which includes the reflections due to the walls, floor, ceiling and other obstacles, is given by the *room impulse response* (RIR) [11, 12, 22, 23]. This is, if the source is modeled as a point in space and emits an *impulse* (a mathematical idealization of an explosive, very short in time and loud sound), then what is measured at the receiver (e.g. a microphone) is the RIR. This idealized point source is referred to as a *monopole*.

The acoustical phenomena (of interest in room acoustics) that are implicitly carried by the RIR can be classified into the following categories, illustrated in Figs. 1.1 and 1.2 [22, 24, 25]:

- Specular reflection. Here the sound is reflected in a specular manner, following a simple geometric law [22]. Normally only part of the sound energy is reflected.

- Diffuse reflection. The sound is reflected in a scattered, not necessarily homogeneous, manner. Perfectly diffuse reflection occurs when the directional distribution of the scattered energy is independent of the direction of incident sound. Fully diffuse reflection in acoustics is described by Lambert's cosine law [22, 26].

- Diffraction. It is caused by the scattering effects that occur in some situations when sound waves encounter an obstacle [23]. The phenomenon is described as the apparent "bending" of the sound field around small obstacles, the spreading of the sound field around edges of large objects, or when passing through small openings [25].

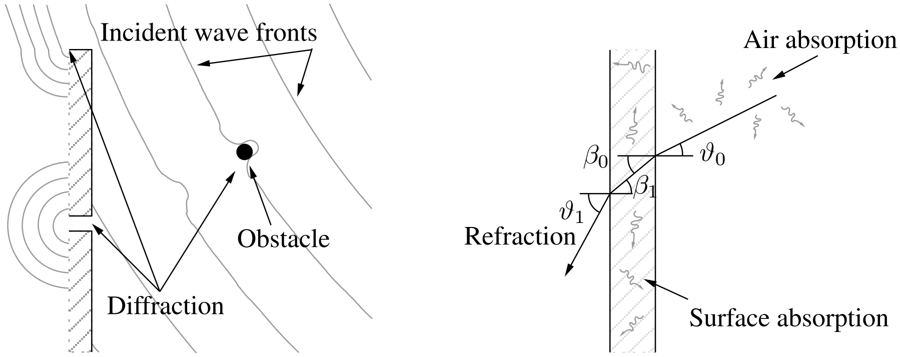


(a) *Specular reflection* (b) *Semi-diffuse reflection* (c) *Totally diffuse reflection*

Figure 1.1: Specular and diffuse reflection.

- Absorption, transmission and refraction. When a sound wave encounters a boundary, part of the energy is reflected while the remainder energy is transmitted to the other side of the boundary [23, 24]. When the "outer" side of the boundary is not part of the domain of interest the energy that is transmitted can be seen as lost and is characterized by an absoption factor. Transmission from a fluid (such as air) to a solid (such as a wall) involves not only longitudinal waves (as in air), but also shear (transverse) waves [24]. However, only longitudinal waves are transmitted from a solid to a fluid since shear waves cannot exist in fluids [24]. The characterization of a transmitted sound wave through a finite body (fluid-solid-fluid transmission) can thus be given in terms of absorption and refraction (i.e. changing of the angle of incidence at the boundaries) [23, 24]. In practice, air also absorbs part of the sound energy.

Consider a monopole source that emmits an impulse. The signal observed at a location in space is, by definition, an RIR. The generated sound field expands spherically in time and is modified by the different acoustic phenomena present along a given path until it reaches the receiver. The RIR is thus a signal made up of all possible (modified) copies of the impulse that arrive at the receiver after traveling their corresponding paths. In Fig. 1.3 a simple reflection example with two paths is given. One of the paths shown corresponds to the *direct* sound path (solid line), which is the shortest possible path (if it is not occluded by any objects) and therefore the first copy of the impulse to arrive at the receiver. The contribution to the RIR of this path is indicated in Fig. 1.4. Another path is depicted (dashed line) which corresponds to a specular reflection from two walls (hence a second order reflection). The con-



(a) *Diffraction phenomena* (b) *Transmitted wave. Absorption and refraction*

Figure 1.2: Other phenomena of sound dynamics.

tribution of this path is part of what is called the *early reflections*. In Fig. 1.4 an example of a recorded RIR is given. As it is seen the RIR is divided into three parts, representing the contributions of the arriving pulses at different times. These components represent different aspects of the acoustic perception from the point of view of a listener and therefore, the distinction is important in many applications. The exact number of RIR subdivisions and the time frames associated with them can vary according to the application. Some authors agree that exact time boundaries are not defendable [4, 27] and that overlap between zones must be allowed. The subdivision model given in Fig. 1.4 is described as follows. Perceptually the direct path is the main carrier of the information content of the source. Early reflections arriving normally between 0 and 20 ms after the direct path enhance the direct sound by the human hearing mechanism [7, 28], together with the direct path they contribute to the intelligibility and *definition* of speech and the *clarity* of music (see e.g. [29, 30] for the technical descriptions of *clarity* and *definition*). Early reflections arriving in a time window between 20 ms and 50 ms after the direct path can contribute to good perceptual music conditions, for example, the spaciousness, loudness and *clarity* in a concert hall. Reflections arriving later than 100 ms after the direct path create a diffuse reverberant effect, the so-called *late reverberation* zone. In specialized enclosures, such as concert halls, these contributions can determine qualifications such as the warmth of music, but in non-acoustically specialized enclosures, such as a swimming hall, this late reverberation field can be detrimental for speech intelligibility. The acoustic events in a room, under some assumptions, can be mathematically idealized to be *linear and time-invariant* (LTI) [22, 23, 31]. Under this mathemati-

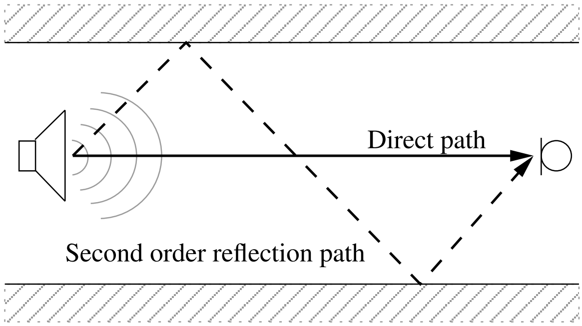


Figure 1.3: A reflection example with two paths.

cal model if the source emits an arbitrary sound, the sound as it would be measured at the receiver can be calculated directly by *convolution* of the RIR and the source signal. In digital signal processing the convolution is an operation that can be performed efficiently via the fast Fourier transform (FFT) [20, 21, 32–34], and thus the RIR constitutes a powerful signal processing model that characterizes the acoustic properties of a room.

Since it is generated by many different acoustic phenomena the RIR depends on many factors, such as the geometry of the room, the acoustic characteristics of the walls, the positions of the source and the receiver, objects present in the room and even the temperature and humidity levels of the air [22, 23]. An accurate estimation (modeling) of the RIR even for only one pair of source and receiver positions is in general a computationally expensive task [22]. Today the computational complexity of the fastest algorithms is still an issue (see Chapter. 2), especially if the goal is an implementation in next-generation acoustic technologies where real-time modeling of massive amounts of acoustic channels is needed.

This work follows an approach used in acoustical physics [23, 24, 35], and more recently also in signal processing [31, 36]. Instead of looking at the RIR as a particular function per source location and receiver location, the RIR is seen as a global space-time function. This multidimensional function characterizes the physics of sound propagation at any point inside the room, quantifying the complexity and richness of a room's sound field. Mathematically this spatio-temporal RIR model is known as the Green's function [7, 22, 23] (the fundamental solution to the wave equation [37]). This approach leads to a theoretical framework based on fast multidimensional Fourier techniques, to simultaneously model multichannel RIRs (i.e. the RIR at many spatial locaions) with low computational complexity.

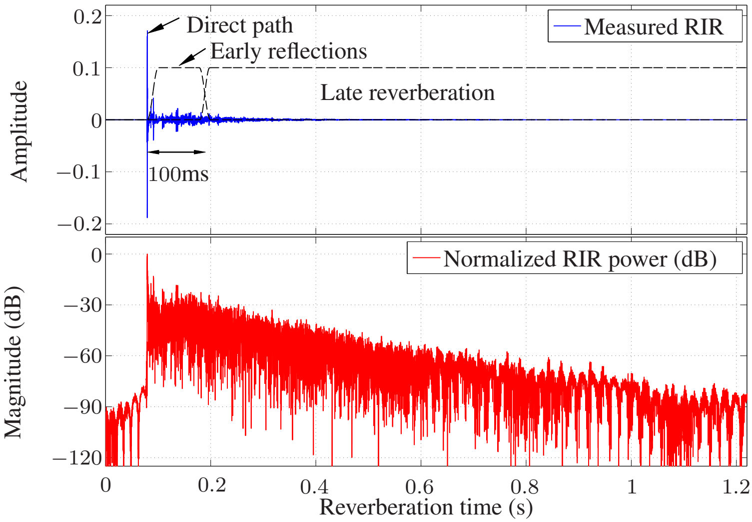


Figure 1.4: Recorded room impulse response (RIR).

The details of the framework proposed in this thesis are given in Part II. Before, a summary of contributions is given next, and a general overview of current RIR simulation approaches is given in the next chapter.

1.3 Contributions

The main contribution of this thesis is the introduction of an efficient computer algorithm to simulate the sound field in a box-shaped (empty) room. To have a better appreciation of this contribution relative to the algorithms proposed in the literature, a general overview of prominent RIR simulation approaches is given in Chapter 2. At the end of this chapter a short description of the newly proposed algorithm, called the generalized Fourier domain (GFD) method, is given together with a discussion on the approach (the technical analysis is found in Part II). The overview given in Chapter 2 also serves as an introduction to the topic, and to highlight the main challenges involved in simulating the RIR on a computer. The overview includes different methods subdivided into two major frameworks: wave theory based methods, Sec. 2.2, and geometrical acoustics based methods, Sec. 2.3, both cornerstones in the development of RIR simulation algorithms. Wave theory represents a formal mathematical description of the evolution of the sound field in a room and geometrical acoustics can be described as a (simpler) subset of wave theory.

Part II of this thesis presents the technical details of the newly proposed GFD method. Efficient simulation of the sound field in a box-shaped (empty) room is addressed using an innovative idea. The key observation is that the sound field in a room with perfectly reflective walls can be modeled using a periodic spatial structure. This is then mathematically related to a sampling condition in the Fourier domain. By carefully making all relevant quantities discrete, a computer algorithm to model the sound field is proposed. The FFT us used to compute the core components of the algorithm. This model has, however, no practical application since sound energy absorption at the walls cannot be simulated in this way. Consequently, a generalized Fourier domain is proposed. It is shown how the sound field in a room with walls with complex-valued reflection coefficients can be related to a sampling condition in this domain. Further, a fast implementation of the generalized discrete Fourier transform (the GFFT) is given. A low-complexity multichannel RIR simulation algorithm is then obtained. The newly proposed method is compared against one well stablished, important algorithm called the mirror image source method (MISM) (see Sec. 2.3.1). These contributions are given in Part II, where each chapter corresponds to a journal article in the list of publications in Sec. 1.4. The general flow of ideas is as follows:

1. *Chapter 3* proposes a low-complexity multichannel RIR simulation algorithm. Here, the fundamental idea behind the method is introduced in the context of box-shaped rooms with perfectly reflective walls using standard Fourier theory. The material presented in this chapter has been published in [i].

2. *Chapter 4* introduces a generalized Poisson summation formula. This formula relates the samples of a function in a generalized Fourier domain to a geometrically weighted periodic summation of a function in the reciprocal domain. A fast generalized discrete Fourier transform algorithm is proposed, and it is shown how this theory can be used to perform fast linear convolutions in the generalized Fourier domain without the need of zero-padding. The material presented in this chapter has been published in [ii].

3. *Chapter 5* presents a more detailed analysis of the generalized discrete Fourier transform, its relationship with the z-transform and analyticity. Important properties of the GFT are derived and the first connections with its application to multichannel RIR simulation are made. The material presented in this chapter has been published in [iv].

4. *Chapter 6* presents a low-complexity multichannel RIR simulation algorithm in box-shaped rooms with absorptive walls. The key idea behind the method in Chapter 3 is extended to include a model of absorptive walls using the main properties of the generalized Fourier transform. Comparisons with the MISM are given. The efficiency of the method is stressed. The material presented in this chapter has been submitted for publication in [v].

1.4 List of Publications

[i] J. Martinez and R. Heusdens. On low-complexity simulation of multichannel room impulse responses, *IEEE Signal Processing Letters*, vol. 17, no. 7, pp. 667 -670, 2010.

[ii] J. Martinez, R. Heusdens and R.C. Hendriks. A generalized Poisson summation formula and its application to fast linear convolution, *IEEE Signal Processing Letters*, vol. 18, no. 9, pp. 501-504, September 2011.

[iii] J. Martinez, R. Heusdens. and R.C. Hendriks. A spatio-temporal generalized Fourier domain framework to acoustical modeling in enclosed spaces, in: *Proc. ICASSP 2012*, pp. 529 - 532, March 2012.

[iv] J. Martinez, R. Heusdens and R.C. Hendriks. A generalized Fourier domain: signal processing framework and applications, *Elsevier Signal Processing*, vol. 93, no. 5, pp. 1259 - 1267, May 2013.

[v] J. Martinez, R. Heusdens. Fast modeling of multichannel room impulse responses, Submitted to *IEEE Transactions on Signal Processing*, July 2011 (Last iteration August 2013).

2

Computer simulation of room impulse responses

2.1 Preliminaries

For many people, daily life activity occurs mostly indoors. Living rooms, offices, restaurants, concert halls, factories, the cabin of a car, the passenger concourse of a train or a plane, to name just a few examples. These spaces have different acoustic properties. They could be quiet, noisy, good for having a conversation, very pleaseant for listening to music, or terrible to convey any acoustic activity. These properties, inherent to the enclosure, have been subject to scientific interest for many years [22, 38–41], for example, in the design of concert halls to provide a pleasant acoustic experience. Depending on the specific application one would not necessarily be interested in an accurate physical model of the RIR, but only in its subjective properties. In the literature one can find several measures and techniques to characterize and quantify the (psychoacoustic) listening conditions in a room [22, 42, 43]. In order to keep it focused, this thesis does not consider subjective qualities; the work is restricted to mathematically characterize and efficiently model the spatio-temporal RIR.

Since the introduction of the digital computer, acoustical design of halls and other enclosures has been changing from expensive physical models, either in natural or reduced scale, to the cheaper and more efficient method of computer simulation. Some authors credit the pioneering use of digital computers for the design of room acoustics to Manfred Schroeder and Heinrich Kuttruff [30, 40, 44], although the first published computer algorithm for calculating the RIR in three-dimensional rooms is due to Asbjørn Krokstad et al. [45] in 1968. This "early days" computer algorithm is based on a theoretical framework known as *geometrical acoustics*. It is derived from the analogous and much older theory of geometrical optics [22] which can be tracked back to the times of Newton and Fermat [46].

Twentieth century acoustic room theory saw the use of rigorous physics in the developing of accurate frameworks to model RIRs based on the wave nature of sound dynamics [22–24, 35]. Although of important theoretical value, *wave theory* solutions have been of relatively little use in digital room acoustic modeling. Wave theory can be used to model all of the acoustical phenomena mentioned in Sec. 1.2 that are relevant to room acoustics, but even today this degree of precision comes at the expense of highly complex models, both analytically and numerically. Geometrical acoustics on the other hand represents a simplified physical model, which allows faster but less accurate algorithms to be derived.

Although geometrical acoustics can be described by a simplified subset of wave theory [23], these two descriptions of sound dynamics are separately cornerstone in modern room acoustics. In trying to close the gap between accuracy and speed, hybrid methods have also appeared that combine key elements of both descriptions. These are often combined with stochastic models to achieve faster, more flexible and more accurate room acoustics simulation algorithms. In the following sections these frameworks are described. A graphical summary of the algorithms covered is given in Fig. 2.1.

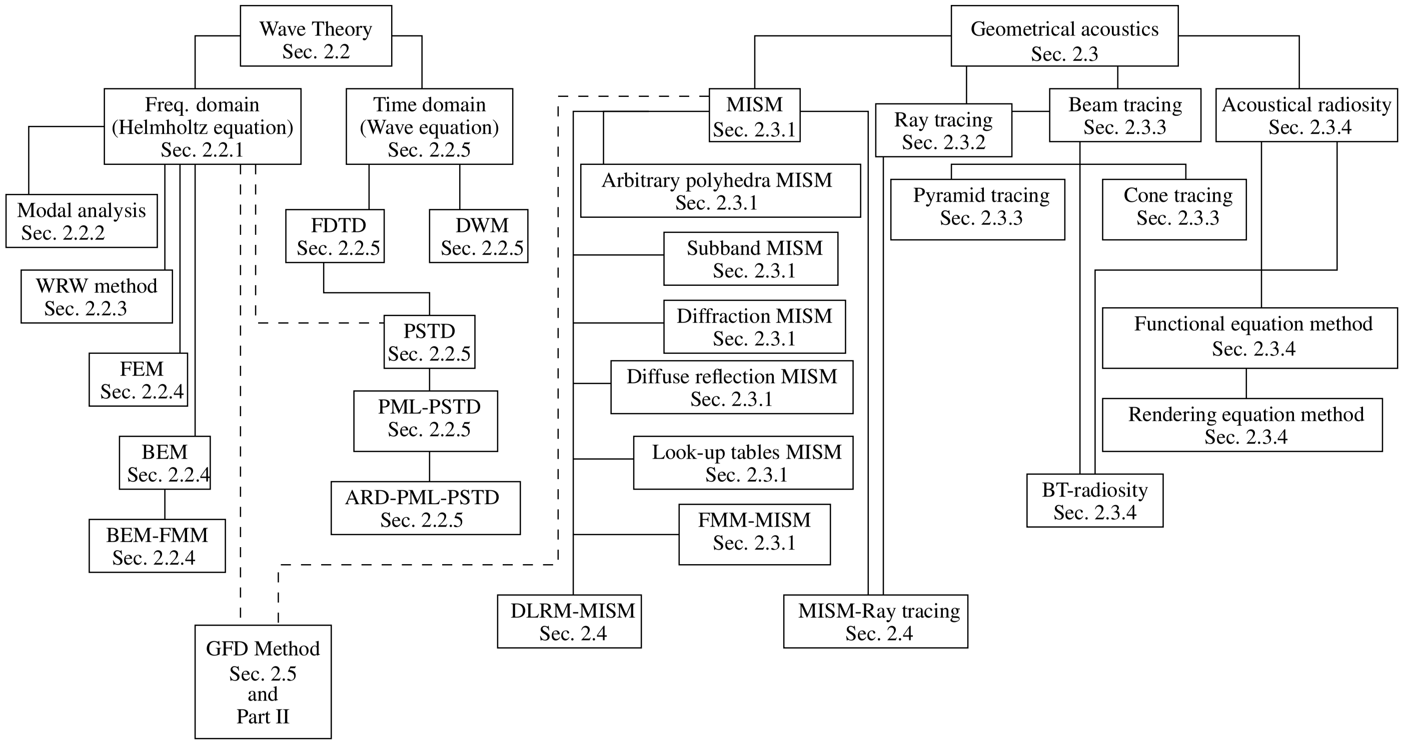


Figure 2.1: Summary of RIR simulation methods reviewed in this chapter. Solid lines denote a strong relationship between the algorithms, dashed lines denote an indirect or conceptual relationship between the algorithms.

2.2 Wave-theory based methods

The problem of computing the RIR is addressed in wave-based methods using mathematical models that describe the dynamics of wave propagation and wave reflection. In this section an introduction to these mathematical models is given, followed by a review of some prominent wave-based RIR simulation methods.

2.2.1 Introduction

The wave equation

A formal description, of the evolution of the sound field in a room is given by the wave theory of room acoustics [7, 22, 23]. Using this theory, the problem can be tackled by solving the *acoustic wave equation* together with some boundary conditions that describe the acoustical properties of the walls and objects in a room, and a driving function that represents the form and nature of the sound sources.

The sound field, as we perceive it, is given by the instantaneous variation of pressure that occurs in air. Let these small variations of pressure be given by a scalar function of space and time $p(\mathbf{x}, t)$, where $\mathbf{x} \in \mathbb{R}^3$ is the location in space, and $t \in \mathbb{R}$ is time. This function cannot be arbitrary, since any sound field must obey the laws of physics (conservation of momentum, continuity, etc [7, 22–24]). The acoustic wave equation together with the boundary conditions and the driving function, also referred to as *the acoustic boundary value problem*, characterizes these restrictions. The solution to the acoustic boundary value problem gives the exact value of $p(\mathbf{x}, t)$, this is, it tells us what the sound field is in the zone where the problem is defined. Without boundary conditions (e.g. in free-field), the wave equation is given by,

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = s(\mathbf{x}, t), \tag{2.1}$$

where $s(\mathbf{x}, t)$ is a scalar function that represents a distribution (continuous or discrete) of sound sources, c represents the speed of sound propagation. The *Laplacian*, denoted by ∇^2, is a second-order partial differential operator acting solely on the space variable; its form is dependent of the coordinate system. In Cartesian coordinates it is written as [22–24, 37, 47],

$$\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right).$$

In plain words, the wave equation says that for a function $p(\mathbf{x}, t)$ to be a valid sound field, the difference in the rate of variation of sound pressure in space and the rate of variation in time (proportional to the speed of sound) must be given by the distribution of sources.

Of equal importance is the *homogeneous* wave equation,

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = 0. \tag{2.2}$$

First, it is clear that valid sound fields in zones "free of sources" (or equivalently where $s(\mathbf{x}, t) = 0$) must fulfill this equation. Secondly, note that we can add any linear combination of solutions of (2.2) into (2.1) without altering the inhomogeneous equation (2.1). Moreover, sources can also be defined purely in terms of boundary conditions, so the acoustic boundary value problem can be expressed in terms of (2.2), [23].

When a sound field is enclosed in a room with objects in it, all relevant acoustic phenomena that can occur (see Sec. 1.2) can be expressed mathematically. Of all these effects, reflection is very important. Let us, therefore, analyze the theory underlying sound reflection phenomena inside a room.

Wave reflection and wall reflection coefficient

Any sound field can be mathematically modeled as a superposition of basis waves, e.g. *plane waves* [23, 24, 35]. To study and characterize the reflective properties of a wall, we can start our discussion with the analysis of an idealized case involving a single *harmonic* plane wave, – a wave with planar spatial geometry and with a purely sinusoidal spatio-temporal dependency,

$$p_{\mathrm{pw}}(\mathbf{x}, t) = A\mathrm{e}^{j(\mathbf{k}^T\mathbf{x} + \omega_0 t)}, \text{ for } \mathbf{k} = [k_x, k_y, k_z]^T \in \mathbb{R}^3, \text{ with } \|\mathbf{k}\| = \frac{|\omega_0|}{c}, \tag{2.3}$$

where the superscript ^{T} denotes vector or matrix transposition, A is an arbitrary complex-valued amplitude and ω_0 is a given temporal frequency. The *wave vector* is given by $\mathbf{k}$, where its Cartesian components k_x, k_y and k_z are known as the *trace wave numbers*, and its magnitude $k = \|\mathbf{k}\| = |\omega_0|/c$ is called the *wave number* [35]. The wave number represents the number of spatial cycles (*wavelengths*) per 2π units of distance and, therefore, it can be seen as a measure of spatial frequency. Wave movement is, in this case, *monochromatic* in time (i.e. it represents a single temporal frequency ω_0), the phase function of this wave, say $\Phi(\mathbf{x}, t)$, is therefore given by,

$$\Phi(\mathbf{x}, t) = \mathbf{k}^T\mathbf{x} + \omega_0 t = k_x x + k_y y + k_z z + \omega_0 t.$$

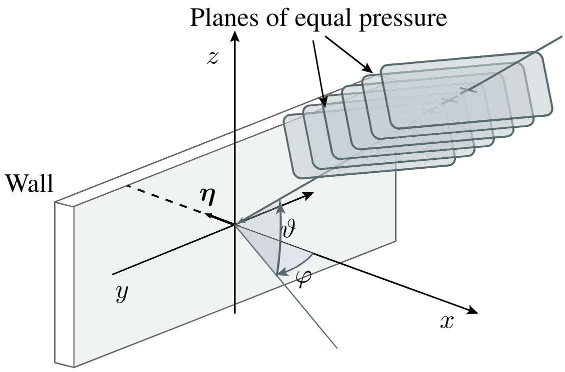


Figure 2.2: A plane wave traveling backwards in the x direction strikes a wall at oblique incidence. The dip and azimuth angles are given by ϑ and φ respectively. The outward wall normal is denoted $\boldsymbol{\eta}$.

Looking at regions in space of constant phase at a given time and setting, without loss of generality, $t=0$ and $\Phi(\mathbf{x},t)=0$, we obtain,

$$k_x x + k_y y + k_z z = 0,$$

which is clearly the equation of a plane. The origin $\mathbf{0}$ is included in the plane and $\mathbf{k}$ is perpendicular to any point $\mathbf{x}$ satisfying the equation, i.e. $\mathbf{k}^T(\mathbf{x}-\mathbf{0})=0$, where the direction of propagation is perpendicular to the orientation of the plane and therefore, given by the direction of the wave vector $\mathbf{k}$. However, for $\omega_0 > 0$ the wave would propagate backwards in space as time passes since $\mathbf{k}^T\mathbf{x} = -\omega_0 t$ for $\Phi(\mathbf{x},t) = 0$; for $\omega = 0$ the wave does not change in time. Note that if a negative time-domain harmonic dependency is selected in the definition of the plane wave, i.e. $p_{pw}(\mathbf{x},t) = A\mathrm{e}^{j(\mathbf{k}^T\mathbf{x}-\omega_0 t)}$, then the wave would be traveling forwards in space as time passes. For several authors this is the preferred behavior (e.g. [35, 48–50]), but it has a minor implication in the selection of a Fourier basis when the sound field is expanded in terms of plane waves via the Fourier transform [24, 35].

Let a wall of infinite extent be positioned on the y,z plane at $x=0$, so that its outward normal, say $\boldsymbol{\eta}$, is pointing in the negative x direction. Consider a plane wave arriving at the wall with dip ϑ and azimuth φ angles as depicted in Fig. 2.2 The trace wave numbers can be expressed in spherical coordinates as,

$$k_x = k\cos(\vartheta)\cos(\varphi),$$
$$k_y = k\cos(\vartheta)\sin(\varphi),$$
$$k_z = k\sin(\vartheta),$$

with $\vartheta, \varphi \in [-\pi/2, \pi/2]$, since the propagation direction is towards the wall. Note that when $\vartheta = |\pi/2|$ or $\varphi = |\pi/2|$, the wave travels perpendicular to the direction of the wall normal. Part of the energy is reflected in the form of another plane wave originating from the wall. The original and the reflected waves interfere with each other and form (at least partially) a *standing wave*.

The change in amplitude and phase taking place during a reflection is characterized by the complex-valued *wall reflection coefficient* or wall reflection factor,

$$\varrho = |\varrho| e^{j\angle\varrho}, \tag{2.4}$$

where $\angle\varrho$ quantifies the phase change and $|\varrho|$ the amplitude change. This quantity is a function of the wall surface, the arriving direction with respect to the wall normal, and the temporal frequency of the plane wave (and therefore, it is function of $\mathbf{k}$). The acoustical properties of the wall are completely characterized if the reflection coefficient is known for all points on its surface, for all incoming directions and for all frequencies (i.e. for all plane waves). In some (rather rare) cases, a wall can have reflective properties that are independent of the direction of arrival of the wave [22, 24], in such cases the wall is said to be *locally reactive*.

Another quantity of importance is the wall impedance, Z, which describes the opposition the wall presents to the force (in the form of the sound pressure at its surface) that makes it vibrate [22],

$$Z = \frac{p_{\mathrm{pw}}(\mathbf{x}, t)}{v_{\mathrm{n}}(\mathbf{x}, t)}, \quad \text{for } \mathbf{x} \in \mathcal{S}, \tag{2.5}$$

where $\mathcal{S}$ is the set of space points comprising the wall surface and $v_{\mathrm{n}}(\mathbf{x}, t)$ is the component of particle velocity normal to the wall. This quantity is a complex-valued function of the angle of incidence and the temporal frequency of the plane wave.

The wall impedance can be derived for a given reflection factor and a direction of arrival. Without loss of generality and for simplicity, we can set $\omega_0 \geq 0$ and take the wave normal (of the incident wave) to be in the x, y plane, so that $k_z = 0$ in (2.3) and $\vartheta = 0$. Referring again to Fig. 2.2, the incident wave arrives from the direction given by φ,

$$p_{\mathrm{in}}(\mathbf{x}, t) = A e^{j(\mathbf{k}^T\mathbf{x} + \omega_0 t)} = A e^{j(k_x x + k_y y + \omega_0 t)} = A e^{j(xk\cos(\varphi) + yk\sin(\varphi) + \omega_0 t)},$$

since by assumption $k_z = 0$, so that $k_x = k\cos(\varphi)$ and $k_y = k\sin(\varphi)$. Conservation of momentum (mass-velocity) is derived using Newton's second law [23, 24], and characterized by Euler's equation,

$$-\nabla p(\mathbf{x},t) = \rho_0 \frac{\partial \mathbf{v}(\mathbf{x},t)}{\partial t}, \tag{2.6}$$

where

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right),$$

is the gradient operator in Cartesian coordinates, ρ_0 is the medium (in our context air) density and $\mathbf{v}(\mathbf{x},t)$ is the particle velocity vector function. This equation allows us to obtain the component of particle velocity normal to the wall $v_{x_{\text{in}}}(\mathbf{x},t)$ (since the normal points in the x-direction), for the case of our plane wave $p_{\text{in}}(\mathbf{x},t)$,

$$\begin{aligned}
\frac{\partial v_{x_{\text{in}}}(\mathbf{x},t)}{\partial t} &= -\frac{1}{\rho_0}\frac{\partial p_{\text{in}}(\mathbf{x},t)}{\partial x}, \\
j\omega_0 v_{x_{\text{in}}}(\mathbf{x},t) &= -\frac{1}{\rho_0}\frac{\partial p_{\text{in}}(\mathbf{x},t)}{\partial x}, \\
v_{x_{\text{in}}}(\mathbf{x},t) &= -\frac{1}{j\rho_0\omega_0}\frac{\partial p_{\text{in}}(\mathbf{x},t)}{\partial x}, \\
v_{x_{\text{in}}}(\mathbf{x},t) &= -\frac{A}{\rho_0 c}\cos(\varphi)\mathrm{e}^{j(xk\cos(\varphi)+yk\sin(\varphi)+\omega_0 t)},
\end{aligned}$$

since $\omega_0 \geq 0$ and consequently $k/\omega_0 = 1/c$. When an incoming wave strikes the wall at $x = 0$ it is reflected, the direction of propagation in the x-coordinate is reversed and the sound pressure and particle velocity are multiplied by the corresponding reflection factor ϱ. Moreover the particle velocity gets multiplied by -1 since $\partial p_{\text{in}}(\mathbf{x},t)/\partial x$ has opposite sign for the reflected wave,

$$\begin{aligned}
p_{\text{rf}}(\mathbf{x},t) &= \varrho A\mathrm{e}^{j(-xk\cos(\varphi)+yk\sin(\varphi)+\omega_0 t)}, \\
v_{x_{\text{rf}}}(\mathbf{x},t) &= \frac{\varrho A}{\rho_0 c}\cos(\varphi)\mathrm{e}^{j(-xk\cos(\varphi)+yk\sin(\varphi)+\omega_0 t)}.
\end{aligned}$$

At the wall surface, both the incoming and the reflected waves are superimposed. Setting $x = 0$, and using $p_{\text{in}}+p_{\text{rf}}$ and $v_{x_{\text{in}}}+v_{x_{\text{rf}}}$ in (2.5), we obtain the wall impedance for this direction of arrival (for $\vartheta = 0$ and a given φ) and frequency ω_0,

$$Z = \frac{\rho_0 c}{\cos(\varphi)}\frac{\varrho+1}{\varrho-1}, \tag{2.7}$$

and from this equation,

$$\varrho = \frac{Z\cos(\varphi) + \rho_0 c}{Z\cos(\varphi) - \rho_0 c}. \tag{2.8}$$

The wall impedance is therefore useful for the derivation of the reflection coefficient and, as will be seen next, for evaluating the boundary conditions for a given enclosure.

Until now the analysis to characterize the reflective properties of a wall has been only for a special (idealized) kind of sound field, namely plane waves. As mentioned earlier, any sound field can be modeled as a superposition of basis waves (e.g. plane waves). We analyze next how a sound field can be decomposed into these basis waves.

Helmholtz equation, boundary conditions and plane wave decomposition

The evolution of the sound field in a room is determined by solving the *wave equation* together with some boundary conditions. These conditions mathematically define the sound field at the position of acoustical obstructions. In a room the walls, floor, and ceiling are arguably the most important boundary conditions. Since the boundary conditions are normally functions of temporal frequency, it is customary to use the time-independent form of the wave equation (2.2) obtained by applying the Fourier transform over the time variable,

$$\begin{aligned} 0 &= \mathcal{F}_t\left\{\nabla^2 p(\mathbf{x},t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2}\right\} \\ &= \nabla^2 P(\mathbf{x},\omega) - \frac{1}{c^2}(i\omega)^2 P(\mathbf{x},\omega) \\ &= \nabla^2 P(\mathbf{x},\omega) + k^2 P(\mathbf{x},\omega), \end{aligned} \tag{2.9}$$

since $k^2 = (\omega/c)^2$, where $\mathcal{F}_t\{\cdot\}$ is the Fourier transform operator over the specified variable, and P denotes the Fourier transform of p. The resulting equation is known as *Helmholtz equation*. To work with the Helmholtz equation (2.9) instead of with the wave equation (2.2) is preferred in many cases, since the time-domain derivative is not present in the former.

It has been shown (e.g. in [23]) that solutions to (2.9) take the form,

$$P(\mathbf{x},\omega) = C_1(\omega)\mathrm{e}^{j(\mathbf{k}^T\mathbf{x})} + C_2(\omega)\mathrm{e}^{-j(\mathbf{k}^T\mathbf{x})}, \tag{2.10}$$

for arbitrary constants C_1 and C_2, if and only if $\|\mathbf{k}\| = k = |\omega|/c$. The time-independent part of a plane wave (see (2.3)), is therefore a solution to the Helmholtz equation. To see this consider the temporal Fourier transform of a plane wave at a single frequency ω_0,

$$P_{\text{pw}}(\mathbf{x},\omega) = A_{\omega_0}\mathcal{F}_t\left\{\mathrm{e}^{j(\mathbf{k}^T\mathbf{x}+\omega_0 t)}\right\} = B_{\omega_0}\delta(\omega-\omega_0)\mathrm{e}^{j(\mathbf{k}^T\mathbf{x})}, \tag{2.11}$$

where $\delta(\cdot)$ is the Dirac's delta generalized function and B_{ω_0} is a different coefficient than A_{ω_0} to account for any possible normalization factors of the Fourier transform. Make $C_1(\omega) = B_{\omega_0}\delta(\omega-\omega_0)$ and $C_2 = 0$ in (2.10), then (2.11) is a solution to (2.9).

Let us now make the connection with the reflection coefficient. If the wall is not locally reactive, the reflection coefficient ϱ (and thus also the impedance Z) is dependent on the arriving direction of the sound field. Fortunately, we can decompose an incoming sound field at the wall surface in terms of plane waves via the Fourier transform [24, 35]. To see this let us go back to our plane wave example. The sound field pressure on a y, z plane at $x = x_0 \leq 0$ is given by,

$$p_{\text{in}}(\mathbf{x},t)\mid_{x=x_0} = p_{\text{in}}(x_0,y,z,t).$$

Here we assume that all sources are confined in the half space at $x > 0$. The Fourier synthesis of this function on the three remaining variables is,

$$p_{\text{in}}(x_0,y,z,t) = \frac{1}{8\pi^3}\int_{\mathbb{R}^3}\breve{P}_{\text{in}}(x_0,k_y,k_z,\omega)\mathrm{e}^{j(k_y y+k_z z+\omega t)}dk_y dk_z d\omega, \tag{2.12}$$

where the complex-valued coefficient function $\breve{P}_{\text{in}}(x_0,k_y,k_z,\omega)$, is given by the forward Fourier transform,

$$\breve{P}_{\text{in}}(x_0,k_y,k_z,\omega) = \int_{\mathbb{R}^3} p_{\text{in}}(x_0,y,z,t)\mathrm{e}^{-j(k_y y+k_z z+\omega t)}dydzdt. \tag{2.13}$$

Note that the complex exponential $\mathrm{e}^{j(k_y y+k_z z+\omega t)}$ has the form of a plane wave (as in (2.3)) measured in the y, z plane at $x = 0$. Let us then propose

$$\breve{P}_{\text{in}}(x_0,k_y,k_z,\omega) = \breve{P}_{\text{in}}(k_y,k_z,\omega)\mathrm{e}^{jk_x x_0}, \tag{2.14}$$

where $\breve{P}_{\text{in}}(k_y,k_z,\omega) \triangleq \breve{P}_{\text{in}}(0,k_y,k_z,\omega)$. The Fourier synthesis (2.12) now looks like a synthesis in terms of plane waves. Recall that we are assuming $p_{\text{in}}(x_0,y,z,t)$ to be a valid sound field in a zone free of sources. Therefore each of the weighted waves $\breve{P}_{\text{in}}(k_y,k_z,\omega)\mathrm{e}^{j(k_x x_0+k_y y+k_z z)}$ must fulfill the homogeneous Helmholtz equation (2.9).

For this to be true, the relation $k = \|\mathbf{k}\| = |\omega|/c$ must hold, which means that the trace wave numbers k_x, k_y and k_z are not independent of each other. We can choose a maximum of two independent wave numbers. Since k_x does not enter into the Fourier synthesis, it is the dependent variable. The following condition must then hold,

$$k_x = \pm\sqrt{\left(\frac{\omega}{c}\right)^2 - (k_y^2 + k_z^2)}.$$

The positive root is chosen since we want to decompose the sound field in terms of plane waves traveling into the negative x direction. The other trace wave numbers k_y and k_z take values from the entire real axis. The implication of this is thus,

$$p_{\text{in}}(x_0, y, z,t) = \frac{1}{8\pi^3}\times \begin{cases} \int_{\mathbb{R}^3} \breve{P}_{\text{in}}(k_y, k_z, \omega)\mathrm{e}^{j(k_x x_0 + k_y y + k_z z + \omega t)} dk_y dk_z d\omega, & k_y^2 + k_z^2 \leq \left(\frac{\omega}{c}\right)^2, \\ \int_{\mathbb{R}^3} \breve{P}_{\text{in}}(k_y, k_z, \omega)\mathrm{e}^{k'_x x_0}\mathrm{e}^{j(k_y y + k_z z + \omega t)} dk_y dk_z d\omega, & k_y^2 + k_z^2 > \left(\frac{\omega}{c}\right)^2. \end{cases}$$

where,

$$k'_x = \sqrt{k_y^2 + k_z^2 - \left(\frac{\omega}{c}\right)^2}.$$

Therefore, when $k_y^2 + k_z^2 > \omega/c$ the synthesis is made in terms of waves having an exponentially decaying component in the (negative) x direction. These waves are called *evanescent waves* [24, 35]. Notice that for evanescent waves the wave vector, $\mathbf{k} = [k_x, k_y, k_z]^T$, contains one non-real component ($k_x \in \mathbb{C}$ in the above example). The constraint $\mathbf{k} \in \mathbb{R}^3$, see (2.3) is only required for plane waves.

At $x_0 = 0$ the Fourier transform (2.12) guarantees the representation of any sound field $p_{\text{in}}(x_0, y, z, t)$ in terms of $\breve{P}_{\text{in}}(k_y, k_z, \omega)$. Moreover this is equivalent to a plane-evanescent wave decomposition. Given the dynamics of wave propagation, at any other plane $x_0 < 0$, plane-waves undergo only a phase shift $\mathrm{e}^{jk_x x_0}$ while evanescent waves only an exponential decay $\mathrm{e}^{k'_x x_0}$, but their associated complex-amplitudes given by $\breve{P}_{\text{in}}(k_y, k_z, \omega)$ do not further change, so that proposition (2.14) is well founded. A rigorous proof is given in [23]. In this domain, the sound field at any plane $x_0 < 0$ relative to the plane at $x = 0$ is obtained by a simple phase shift $\mathrm{e}^{jk_x x_0}$ (where k_x becomes purely imaginary for evanescent waves). This fact constitutes a powerful approach for wave field extrapolation [24, 35, 51].

Continuing with the discussion on boundary conditions, let us place a wall at $x_0 = 0$ and express the wall reflection coefficient ϱ in the wave-number domain as $\breve{\mathsf{P}}(k_y, k_z, \omega)$. This function modifies the complex amplitude of each of the waves computed in the expansion (2.12), accounting for both plane waves and evanescent waves. We write for the reflected sound field at $x_0 = 0$,

$$\breve{P}_{\mathrm{rf}}(k_y, k_z, \omega) = \breve{\mathsf{P}}(k_y, k_z, \omega)\breve{P}_{\mathrm{in}}(k_y, k_z, \omega), \tag{2.15}$$

or in the space-time domain,

$$p_{\mathrm{rf}}(x_0, y, z, t) = \int_{\mathbb{R}^3} \varrho(y - y', z - z', t - t')p_{\mathrm{in}}(x_0, y, z, t)dy'\, dz'\, dt'. \tag{2.16}$$

This last equation clearly shows that angle, surface and frequency dependent reflection can be expressed in terms of a spatio-temporal convolution of the incoming soundfield $p(\mathbf{x}, t)$ and the reflection function $\varrho(\mathbf{x}, t)$ at the wall surface $\mathcal{S}$.

Consider now an empty space enclosed by a wall surface denoted by $\mathcal{S}$. For a sound field given in the temporal frequency domain, we have from (2.5) that the normal component of the particle velocity at the surface can be expressed as

$$V_{\mathrm{n}}(\mathbf{x}, \omega) = \frac{P(\mathbf{x}, \omega)}{Z(\omega)}, \quad \text{for } \mathbf{x} \in \mathcal{S}.$$

According to Euler's equation (2.6) the boundary conditions can then be expressed as,

$$\nabla P(\mathbf{x}, \omega) \cdot \boldsymbol{\eta}(\mathbf{x}) = -j\omega\rho_0 \frac{P(\mathbf{x}, \omega)}{Z(\mathbf{x}, \omega)} \quad \text{for } \mathbf{x} \in \mathcal{S}, \tag{2.17}$$

where $\boldsymbol{\eta}(\mathbf{x})$ is the outward normal unit vector at every point of the boundary, $(\cdot)$ is the dot product and $Z(\mathbf{x}, \omega)$ is the impedance as a function of the wall surface and temporal frequency.

To summarize, the problem of finding the sound field in a room is addressed in wave-based methods using the theory of wave propagation and wave reflection. A model consisting of a partial differential equation, for example, the wave equation (2.1) or the Helmholtz equation (2.9), and a set of boundary conditions (e.g. (2.17)) are defined as the underlying problem to be solved. This model is commonly referred to as the *acoustic boundary value problem*. The model can be made more accurate (at the cost of higher complexity) by defining, e.g. non-linear versions of the wave equation (see for example [52]) to more realistically characterize inhomogeneous media (like air under nonuniform humidity and temperature conditions), or more complicated boundary conditions to model sound sources (e.g. the vibration of a loudspeaker membrane) or any other objects in the enclosure.

In the next subsections an analysis of prominent wave-based methods is given. The first method, modal analysis [22, 53], computes any sound field in the room

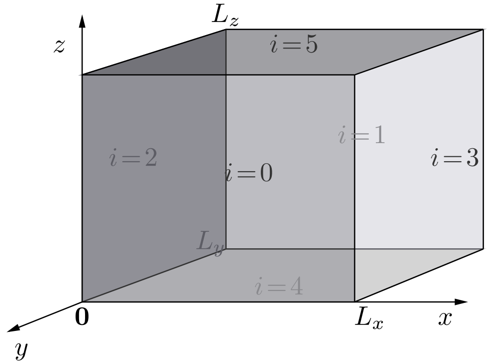


Figure 2.3: A box-shaped room with dimensions L_x, L_y and L_z. Its wall surfaces are denoted by $\mathcal{S}_i$ ($i=0,\ldots,5$).

by expressing it as a linear combination of a set of basis functions, called *modal functions*, that are independent solutions of the boundary value problem. The wave propagation-wave reflection-wave propagation (WRW) method [52], makes use of extrapolation theorems to model sound propagation. Following are the finite element method (FEM) [54], the boundary element method (BEM) [55, 56], the digital waveguide mesh (DWM) [57] and the finite-difference time-domain (FDTD) methods. These methods first represent the wave equation together with all the relevant conditions in different forms and then proceed with numerical evaluation of the equivalent problem. The DWM and FDTD are numerical approaches to solve the wave equation, while FEM and BEM attempt to solve the problem in the temporal frequency domain, where solutions to the Helmholtz equation are to be found.

2.2.2 Modal analysis

Any sound field can be considered a superposition of basic sound waves. Therefore, one way to find a solution to the problem is by expressing the sound field as the summation of a *complete set* of properly weighted *modal functions* (the basic waves) that independently fulfill the wave equation and the boundary conditions. The method that addresses the problem of determining the modal functions, their driving frequencies and the coefficients necessary to synthesize the sound field is known as *modal analysis* [22, 53]. Although mathematically rigorous, modal analysis turned out to be quite limited in practical applications due to its high complexity both analytical and numerical [22, 53, 58, 59].

To have a glimpse of the problem, consider the simple scenario of a box-shaped room enclosed by walls with direction and surface independent impedance. The room space is given by $\mathcal{V} = [0, L_x] \times [0, L_y] \times [0, L_z]$, so that one corner of the room is at the origin of the coordinate system and L_x, L_y, L_z, denote the dimensions of the room in the x, y and z directions, respectively. The scenario is depicted in Fig. 2.3. We would like to find the room's steady-state modal functions. As is done in (2.17), we set up the boundary value problem in the temporal frequency domain using (2.9) and (2.6), as,

$$\nabla^2 P(\mathbf{x}, \omega) + k^2 P(\mathbf{x}, \omega) = 0, \tag{2.18}$$

$$\nabla P(\mathbf{x}, \omega) \cdot \boldsymbol{\eta}_i = -j\rho_0\omega \frac{P(\mathbf{x}, \omega)}{Z_i(\omega)}, \quad \text{for } \mathbf{x} \in \mathcal{S}_i, \tag{2.19}$$

where $\boldsymbol{\eta}_i$ is the outward normal unit vector of the ith wall, $\mathcal{S}_i$ is the set of space points comprising the ith wall surface and $Z_i(\omega)$ is the impedance of the ith wall, which is constant for the wall surface. In the following, the dependence of Z_i on the frequency variable is not written explicitly for the sake of notational simplicity.

Normally, one starts with an educated guess on the form of the modal functions. In this example, we can directly proceed with the technique of separation of variables [22, 23]. Exponential functions arise naturally as solutions. Let a solution, say $\psi(\mathbf{x}, \omega)$, of (2.18) be decomposed into three functions, each dependent on only one spatial coordinate, i.e. $\psi(\mathbf{x}, \omega) = \psi_x(x, \omega)\psi_y(y, \omega)\psi_z(z, \omega)$. Inserting this solution into the boundary value problem, (2.18) and (2.19), results in three ordinary differential equations. For instance, $\psi_x(x, \omega)$ must satisfy

$$\frac{d}{dx}\, \psi_x(x, \omega) + k_x^2\, \psi_x(x, \omega) = 0, \tag{2.20}$$

together with

$$\frac{d}{dx}\psi_x(x, \omega) = -j\rho_0\omega \frac{\psi(x, \omega)}{Z_0} = -j\xi_0 k \psi(x, \omega) \quad \text{for } x = 0, \tag{2.21}$$

$$\frac{d}{dx}\psi_x(x, \omega) = -j\rho_0\omega \frac{\psi(x, \omega)}{Z_1} = -j\xi_1 k \psi(x, \omega) \quad \text{for } x = L_x. \tag{2.22}$$

where Z_0 is the impedance of the wall perpendicular to the x direction at $x = 0$ and Z_1 the impedance of the opposite wall (see Fig. 2.3). The *specific wall admittances* $\xi_i = \rho_0 c / Z_i$ are used to simplify the notation. Solutions of (2.20) take the form [22, 23, 59] (see also Eq. 2.10),

$$\psi_x(x, \omega) = A(\omega)\mathrm{e}^{jk_x x} + B(\omega)\mathrm{e}^{-jk_x x}, \tag{2.23}$$

where k_x can take on complex values. The constants $A(\omega)$ and $B(\omega)$ are used to adapt the solution to the boundary conditions (2.21) and (2.22). Substituting (2.23) into (2.21) and (2.22) we obtain an *acoustic eigenvalue equation* for the x-coordinate,

$$\frac{d}{dx}(\mathrm{e}^{jk_x x} + \mathrm{e}^{-jk_x x}) = -j\xi_i k(\mathrm{e}^{jk_x x} + \mathrm{e}^{-jk_x x}), \quad \text{for} \begin{cases} i = 0 & \text{and} \quad x = 0, \\ i = 1 & \text{and} \quad x = L_x, \end{cases}$$

so that,

$$jk_x(\mathrm{e}^{jk_x x} - \mathrm{e}^{-jk_x x}) = -j\xi_i k(\mathrm{e}^{jk_x x} + \mathrm{e}^{-jk_x x}),$$

or,

$$(k_x + \xi_i k)\mathrm{e}^{jk_x x} = (k_x - \xi_i k)\mathrm{e}^{-jk_x x}.$$

This implies

$$\begin{aligned} \frac{(k_x - \xi_0 k)}{(k_x + \xi_0 k)} &= 1, && \text{for } x = 0 \text{ (setting } i = 0), \\ \frac{(k_x - \xi_1 k)}{(k_x + \xi_1 k)} &= \frac{\mathrm{e}^{jk_x L_x}}{\mathrm{e}^{-jk_x L_x}}, && \text{for } x = L_x \text{ (setting } i = 1). \end{aligned} \tag{2.24}$$

Combining both conditions in (2.24) by multiplication we get,

$$\frac{(k_x - \xi_0 k)(k_x - \xi_1 k)}{(k_x + \xi_0 k)(k_x + \xi_1 k)} = \mathrm{e}^{jk_x(2L_x)}. \tag{2.25}$$

The (complex-valued) roots $\{k_{n_x}\}_{n_x \in \mathbb{Z}}$, of this eigenvalue equation are the eigenvalues in the x direction of the modal functions allowed in the room. Note that this is no longer an uncountable set as it is in the free-field case. Adapting the constants in (2.23), the n_xth modal function is found,

$$\psi_{n_x}(x, \omega) = \left(k_{n_x} + \xi_0 \frac{|\omega|}{c}\right) \mathrm{e}^{jk_{n_x} x} + \left(k_{n_x} - \xi_0 \frac{|\omega|}{c}\right) \mathrm{e}^{-jk_{n_x} x}. \tag{2.26}$$

The eigenfunctions $\psi_{n_y}(y, \omega)$ and $\psi_{n_z}(z, \omega)$ with respective eigenvalues k_{n_y} and k_{n_z} in the y and z directions are derived in the same way, and the complete eigenfunctions are given by $\psi_{\mathbf{n}}(\mathbf{x}, \omega) = \psi_{n_x}(x, \omega)\psi_{n_y}(y, \omega)\psi_{n_z}(z, \omega)$, with $\mathbf{n} = [n_x, n_y, n_z] \in \mathbb{Z}^3$, and eigenvalues $k_{\mathbf{n}}^2 = k_{n_x}^2 + k_{n_y}^2 + k_{n_z}^2$.

It has been proven that these eigenfunctions are *mutually orthogonal* [23]. Moreover, defining a suitable space of possible sound fields, it can be shown that the eigenfunctions form a complete set (with respect to some suitable definition of convergence) in this space (see e.g. [37, 60]) for all possible values of the impedances

$Z_i \in \mathbb{C}$. Once this is rigorously established (although this can be nontrivial), we are in principle able to evaluate any desired acoustical property of the room, such as the sound field inside the room given an arbitrary set of sources. To see this, consider the inhomogeneous Helmholtz equation,

$$\nabla^2 P(\mathbf{x},\omega) + k^2 P(\mathbf{x},\omega) = s(\mathbf{x},\omega). \tag{2.27}$$

Since the orthogonal eigenfunctions $\psi_{\mathbf{n}}(\mathbf{x},\omega)$ form a complete set, we can expand the source into a series of $\psi_{\mathbf{n}}(\mathbf{x},\omega)$,

$$s(\mathbf{x},\omega) = \sum_{\mathbf{n}\in\mathbb{Z}^3} C_{\mathbf{n}}(\omega)\psi_{\mathbf{n}}(\mathbf{x},\omega), \quad \text{with } C_{\mathbf{n}}(\omega) = \frac{1}{K_{\mathbf{n}}(\omega)} \int_{\mathcal{V}} s(\mathbf{x},\omega)\psi^*_{\mathbf{n}}(\mathbf{x},\omega) d\mathbf{x}, \tag{2.28}$$

for some constants $C_{\mathbf{n}}(\omega)$ and some normalization factors $K_{\mathbf{n}}(\omega)$, where the superscript * denotes complex conjugation. In this way, the solution $P(\mathbf{x},\omega)$ is expanded as,

$$P(\mathbf{x},\omega) = \sum_{\mathbf{n}\in\mathbb{Z}^3} D_{\mathbf{n}}(\omega)\psi_{\mathbf{n}}(\mathbf{x},\omega). \tag{2.29}$$

The problem is solved if the unknown coefficients $D_{\mathbf{n}}(\omega)$ can be found in terms of the known coefficients $C_{\mathbf{n}}(\omega)$. Inserting both expansions (2.28) and (2.29) into (2.27), and equating akin terms we obtain,

$$D_{\mathbf{n}}(\omega) = \frac{C_{\mathbf{n}}(\omega)}{(\omega/c)^2 - k^2_{\mathbf{n}}}, \tag{2.30}$$

since $\nabla^2\psi_{\mathbf{n}}(\mathbf{x},\omega) = -k^2_{\mathbf{n}}\psi_{\mathbf{n}}(\mathbf{x},\omega)$ and $k = |\omega|/c$.

Thus, in order to find the sound field in a room $p(\mathbf{x},t)$, first the problem is posed in the temporal frequency domain, with the Helmholtz equation (2.18) as the partial differential equation to be solved together with a set of boundary conditions. The boundary conditions characterize the reflective properties of the walls (and possibly any other reflective surfaces in the enclosure) and can be given, as in the example of (2.19), in terms of wall impedances. Then a set of basis functions, also called modal functions or eigenfunctions, that are (independently) solutions to the problem is given or proposed (in the example, exponential functions where used). In order to adapt these eigenfunctions to the boundary conditions, their eigenvalues are needed. An acoustic eigenvalue equation is then formed to find the eigenvalues for each of the eigenfunctions. Once the exact form of the basis functions is known, then any acoustic event in the room can be expressed in terms of these.

An unknown sound field is generated by an arbitrary but known set of sources. The unknown sound field and the known set of sources are then expanded in terms of the modal functions, resulting in a set of unknown sound field coefficients and a set known source coefficients (in the example, $D_{\mathbf{n}}$ and $C_{\mathbf{n}}$ respectively). The unknown sound field coefficients are then derived in terms of the known source coefficients ((2.30). The temporal frequency domain representation of the sound field, $P(\mathbf{x}, \omega)$, is then obtained by a linear combination of the basis functions weighted by the sound field coefficients (i.e. Eq. (2.29)). The actual sound field, $p(\mathbf{x}, t)$, is then found by applying an inverse Fourier transform over the temporal frequency variable. Recalling that the RIR is the sound field when the source is modeled as a point source (monopole) emitting an impulse, then modal analysis constitutes a RIR simulation method.

In order apply the method, one must first solve the acoustic eigenvalue equation for the set of modal functions. In the simple example given above, for a box-shaped room and walls with angle and surface-independent impedances, we already see from (2.25) that the resulting acoustic eigenvalue equation is non-linear and transcendental. Until today we do not know how to solve this kind of equations analytically [53, 59]. To find the roots (i.e. the eigenvalues) of the eigenvalue equation one must then turn to numerical methods, which are complicated and time consuming [53, 59]. Moreover, the number of modal functions in (2.29) (and therefore the number of eigenvalues to be found) is infinite. In practice one would be interested in band-limited sound fields up to a maximum temporal frequency ω_B. In this case, it is only necessary to find eigenvalues and to limit the summation in (2.29) such that

$$|k_{\mathbf{n}}| \leq \left(\frac{\omega_B}{c}\right).$$

Then, the number of modal functions needed grows as the cube of the temporal frequency. This is why modal analysis is normally used for the study of the room responses at low frequencies in cases where the room geometry and boundary conditions do not deviate too much from basic forms, like box-shaped rooms with fully reflective (rigid), or slightly damped walls [53].

2.2.3 WRW model

One important approach due to Berkhout et al. [52], called the wave propagation-wave reflection-wave propagation (WRW) model, can model realistic sound fields. This method is based on a technique called *wave field analysis/synthesis* developed in the 1980's by Berkhout in the context of acoustic seismic exploration [24], and

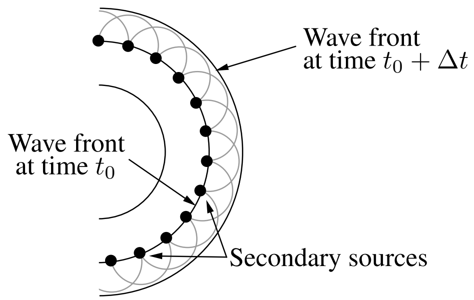


Figure 2.4: Huygens Principle. A propagating wave front can be modeled in terms of a set of "secondary sources" emanating from every point on the wavefront at a previous instant.

later extended to room acoustics [4]. At the core of the approach lie the concepts of wave field extrapolation and wave field reflection [24, 35], of which a glimpse is already given in the discussion following equations (2.14) and (2.15).

Huygens principle states that the fronts of any wave field can be modeled in terms of a set of "secondary" sources emanating from every point on the wavefront at a previous instant. In the case of sound fields this can be regarded intuitively as the air particles pushing each other while transmitting the sound. Every other particle can be seen as a generating sound source for the next particles in the chain (see Fig. 2.4). This principle is quantified and mathematically formalized by the Kirchhoff-Helmholtz integral [24, 35, 52],

$$P(\mathbf{x},\omega) = \frac{1}{4\pi} \oint_{\mathcal{S}} \left(P(\mathbf{x}',\omega) \frac{\partial}{\partial \boldsymbol{\eta}(\mathbf{x}')} \left(\frac{\mathrm{e}^{-jk\|\mathbf{x}-\mathbf{x}'\|}}{\|\mathbf{x}-\mathbf{x}'\|} \right) + \frac{\partial P(\mathbf{x}',\omega)}{\partial \boldsymbol{\eta}(\mathbf{x}')} \left(\frac{\mathrm{e}^{-jk\|\mathbf{x}-\mathbf{x}'\|}}{\|\mathbf{x}-\mathbf{x}'\|} \right) \right) d\mathcal{S}, \qquad \text{for } \mathbf{x} \in \mathcal{V}, \tag{2.31}$$

where $\mathbf{x}'$ is the integration variable that takes values on the surface $\mathcal{S}$ and $\boldsymbol{\eta}(\mathbf{x}')$ is the outward normal unit vector at surface point $\mathbf{x}'$. The integral states that the sound field in a source-free volume $\mathcal{V}$ is totally described by a distribution of secondary sources on the boundary $\mathcal{S}$ of $\mathcal{V}$, driven by the sound pressure function at the boundary $P(\mathbf{x}',\omega)$ (see Fig. 2.5). The surface integral in (2.31) is composed of two summing terms. The second term contains an expression involving the *free-field Green's* *function*,

$$\frac{\mathrm{e}^{-jk\|\mathbf{x}-\mathbf{x}'\|}}{\|\mathbf{x}-\mathbf{x}'\|},$$

which is the temporal frequency domain representation of the RIR (also known as the *room transfer function*) for unbounded space. The first term involves the directional gradient of the Green's function,

$$\frac{\partial}{\partial \boldsymbol{\eta}(\mathbf{x}')}\left(\frac{\mathrm{e}^{-jk\|\mathbf{x}-\mathbf{x}'\|}}{\|\mathbf{x}-\mathbf{x}'\|}\right).$$

These terms appear as extrapolation terms in the integral, in other words, they allow the sound field inside the whole volume to be expressed in terms of its value on the boundary. The Green's function is the response to a monopole (point source) emitting an impulse with its directional gradient being the response to a *dipole*. In other words, a pair of acoustic point sources separated by an infinitesimal distance emitting each an impulse, with both impulses being of equal magnitude but opposite sign (see e.g. [23]). The Kirchhoff-Helmholtz integral (2.31) then states that the sound field inside volume $\mathcal{V}$ can be reconstructed from a distribution of monopoles and dipoles on the boundary. The dipoles are weighted by the sound pressure at the boundary, and the monopoles are weighted proportionally to the local normal particle velocity at the boundary, since (by Euler's equation (2.6)) we have that,

$$\frac{\partial P(\mathbf{x}',\omega)}{\partial \boldsymbol{\eta}(\mathbf{x}')} = j\omega\rho_0 V_{\mathrm{n}}(\mathbf{x}',\omega).$$

When the closed surface degenerates to a plane of infinite extent between source and receiver domains, e.g. the y,z plane at $x = x'$, the Kirchhoff-Helmholtz integral can be simplified into the Rayleigh I and Rayleigh II integrals [4, 24, 52]. The Rayleigh integrals are important as they explain that if the sound pressure field or particle velocity at a certain infinite plane is known (all sources are assumed to be behind this plane), then it is possible to extrapolate the sound field or the particle velocity at a distant plane without error. The Rayleigh I integral extrapolates the sound field with a distribution of monopoles on the plane using only particle velocity information. The Rayleigh II integral describes the pressure field in terms of a distribution of dipoles on the plane using only pressure information. The Rayleigh II integral is the basis for the theory presented in [52]. Let the sound pressure $P(\mathbf{x}',\omega)$, be known

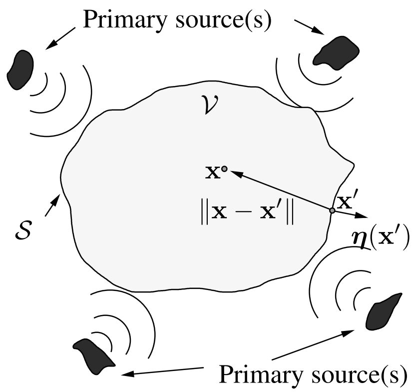


Figure 2.5: Kirchhoff-Helmholtz principle. The sound field inside a source-free volume, say $\mathcal{V}$, can be totally described if the sound field and its local normal particle velocity is known at a boundary surface, say $\mathcal{S}$, enclosing the volume.

on the y, z plane at $x = x'$, then the Rayleigh II integral is written as [24],

$$\begin{aligned} P(\mathbf{x},\omega) &= -\frac{1}{2\pi}\int_{\mathbb{R}^2} P(\mathbf{x}',\omega)\frac{\partial}{\partial x'}\left(\frac{\mathrm{e}^{-jk\|\mathbf{x}-\mathbf{x}'\|}}{\|\mathbf{x}-\mathbf{x}'\|}\right) dz'dy', \\ &= \frac{|x-x'|}{2\pi}\int_{\mathbb{R}^2} P(\mathbf{x}',\omega)\mathrm{e}^{-jk\|\mathbf{x}-\mathbf{x}'\|}\frac{1+jk\|\mathbf{x}-\mathbf{x}'\|}{\|\mathbf{x}-\mathbf{x}'\|^3}\; dz'dy', \end{aligned} \tag{2.32}$$

for $x < x'$, where $\mathbf{x}' = [x', y', z']^T$. This is a convolution integral of the sound field at the boundary $P(\mathbf{x}',\omega)$ with a *propagation* kernel,

$$G(\mathbf{x},\omega) = -\frac{1}{2\pi}\frac{\partial}{\partial x}\left(\frac{\mathrm{e}^{jk\|\mathbf{x}\|}}{\|\mathbf{x}\|}\right) = \frac{|x|}{2\pi}\mathrm{e}^{-jk\|\mathbf{x}\|}\frac{1+jk\|\mathbf{x}\|}{\|\mathbf{x}\|^3},$$

such that the sound field is propagated to a more distant plane. In order to be calculated on a computer, the integral has to be made proper and discrete such that it can be expressed in terms of matrix multiplications. When the integration limits are made finite, truncation errors appear. Remarkably, the truncation error is equivalent to a realistic diffraction effect at the boundary [61], and so the method accurately models diffraction effects at object boundaries [52]. Furthermore, the convolution operator can be diagonalized via the Fourier transform, increasing the efficiency of the computations. By performing a Fourier transform on $P(\mathbf{x}',\omega)$ over the two spatial variables,

y and z, we obtain

$$P(\mathbf{x}', \omega) \xleftrightarrow{\mathcal{F}_{y,z}} \breve{P}(x', k_y, k_z, \omega).$$

Moreover, it has been shown in [35, 52] that

$$G(\mathbf{x}, \omega) \xleftrightarrow{\mathcal{F}_{y,z}} \breve{G}(x, k_y, k_z, \omega) = \mathrm{e}^{jk_x x},$$

with $k_x = (k^2 - k_y^2 - k_z^2)^{1/2}$. Therefore, the Rayleigh II convolution integral (2.32) can be expressed as,

$$\begin{aligned} P(\mathbf{x}, \omega) &= \mathcal{F}^{-1}_{k_y,k_z} \left\{ \breve{P}(x', k_y, k_z, \omega) \breve{G}(x, k_y, k_z, \omega) \right\} \\ &= \mathcal{F}^{-1}_{k_y,k_z} \left\{ \breve{P}(x', k_y, k_z, \omega) \mathrm{e}^{jk_x x} \right\}, \end{aligned}$$

where $\mathcal{F}^{-1}_{k_y,k_z}\{\cdot\}$ is the inverse Fourier transform operator over the specified variables. Sound field propagation is described by a simple phase shift in this spatio-temporal Fourier domain, which is also known as the angular spectrum representation or the wave-domain [35]. Sound reflection is also expressed as spatial convolution, that is a function of the surface of the walls using a periodic (aliased) and discrete version of (2.15), so that the full WRW model can be expressed conveniently in terms of matrix multiplications.

The WRW method can model reflection and diffraction in rooms with arbitrary piece-wise planar geometry (including any planar objects that might be in the enclosure) and it is naturally a multichannel method, i.e. it simultaneously obtains the RIRs at discrete positions on a line (or a plane). However, its computational complexity is high and it heavily depends on the number (and size) of matrix multiplications it has to perform, especially in the 3-D case. Higher order reflections are modeled as recursive matrix multiplications. These are expressed as a Neumann series, and therefore equal to the inverse of a certain matrix. The efficient inversion of this matrix, however, remained a problem to solve [52]. Another problem that has to be controlled carefully is aliasing due to the discretization and truncation of the underlying integrals [24, 31, 35], this occurs in space and time, and can become expensive to overcome.

2.2.4 Finite element analysis and boundary element analysis

The finite element method (FEM) can be defined as "*a general discretization procedure of continuum problems posed by mathematically defined statements*" [62]. The FEM can be used to find numerically approximate solutions to, for example, partial differential equations with boundary conditions. Finite element analysis is by itself a very rich research area in engineering and mathematics, growing at an impressive pace since the 1960's [48, 62]. In acoustics, it is mostly used to model time-harmonic sound fields, i.e. to find solutions to the Helmholtz equation, or to model narrow-band sound fields (e.g. by synthesis of individual solutions at different frequencies). In this method, the room space is subdivided into small volumetric elements, and a large system of ordinary differential equations is formulated at the grid points. One procedure, known as the *Galerkin method* [62], first ensures that the individual solutions of these equations are the sound pressures at the grid points. These values are then used as weights in the interpolation of a (finite) set of basis functions (rigorously proven, numerically computed and optimized, or sometimes just conjectured) to satisfactorily approximate the sound field. When the values at the grid points are combined, an approximation of the solution to Helmholtz equation at any point can be obtained [48]. One of the advantages of the FEM is that the space can be partitioned using cells having different shapes, thus allowing complex room geometries to be modeled accurately. Yet, to find good quality partitions for arbitrary domains is a hard problem of central importance [63]. In practice, the number of unknowns is too large for the system of differential equations to be solved in a reasonable time. However, one can exploit the fact that the matrices that characterize the system are in general sparse. One then resorts to different methods to directly solve, or to further simplify the model. For example, transforming the system to a domain where it can be accurately represented by a few *eigenmodes*. The eigenvalues for these transformed modes are then found to solve the system. Still, the number of unknowns (and therefore the complexity of the method) increases as the cube of the temporal frequency [64], since the grid size must be smaller than the smallest wave-length to achieve accurate approximations. Additionally, broadband solutions must be synthesized from the individual results at the frequencies of interest. Because of this, FEM solutions are commonly used to model the acoustics of a room at low frequencies, or to simulate small enclosures.

Another related technique is the boundary element method (BEM). Like the WRW method, the BEM uses the Kirchhoff-Helmholtz integral to effectively reduce the dimension of the problem by one [7, 55, 56, 65]. Therefore, to solve a 3-D spatial problem, only a mesh of discrete surface (2-D) elements at the boundaries are required to extrapolate the sound field of the whole enclosure. The Kirchhoff-Helmholtz integral (2.31) is formulated in a discrete form and solved using matrix algebra. Unlike the FEM, the application of the BEM reduces the Helmholtz problem to a non-linear eigenvalue problem, the matrices that characterize the system are in general fully populated and have no particular structure, although they normally are much smaller than the ones generated by the application of the FEM (before reduction). To help ease the problem, techniques have been developed to formulate equivalent or approximate systems that can be expressed as a linear, algebraic eigenvalue problem [65]. More recent research has combined fast multipole methods (FMM) with BEM. The resulting BEM-FMM has shown promising results, yielding performances that scale as the square of the temporal frequency (and linearly as the surface of the boundary) [63], making them better suited than the FEM for larger enclosures, but they are still too complex to allow real-time simulation for broadband acoustic scenes.

2.2.5 Time-domain wave solvers

Numerical techniques can be applied to solve the wave equation directly in the time domain. Two common challenges of time-domain wave-based acoustic simulation are computational complexity (particularly at high frequencies) and *numerical dispersion*. The latter refers to phase errors that occur since higher frequencies tend to travel slower on the numerical mesh than lower frequencies, leading in virtually all cases to an unnatural effect and mistuning of key modal functions [57, 63].

The digital waveguide

The digital waveguide is based upon the discretization of the d'Alembert solution $p(x,t)$ of the 1-D wave equation [7, 23, 57]. The solution reads $p(x,t) = p_+(x - ct) + p_-(x + ct)$, which represents two traveling waves: $p_+(x - ct)$ in the positive x direction, and $p_-(x + ct)$ in the negative x direction. Assuming both functions to be bandlimited, one can sample them with sampling intervals in x and t, and the resulting system can be implemented with two parallel digital delay lines that represent the left and right traveling functions. After defining the spatial length in number of junctions (samples), the lines are terminated at both ends. Each junction, including the termination junctions, is coupled in order to obey the dynamics of the vibrating system. The junctions are therefore called *scattering* junctions. The chosen dynamics at the junctions represent a discrete version of the boundary value problem. More complex vibrating systems like the bridge of a guitar, wind instruments, or even the vocal tract can be simulated through appropriate junction coupling [57]. In this way, complex interconnections can be made via scattering junction conditions in order to form 2-D, 3-D, and higher hyperdimensional meshes, which can model the vibrating properties of multidimensional objects (and thus room acoustics). This technique is therefore known as the digital waveguide mesh (DWM). Although the DWM has been used to efficiently model 1-D (and to a lesser extent but still efficiently 2-D) spatial systems [57], its application to room acoustics comes not without challenges. Besides the numerical dispersion problem, one major challenge is to reduce the computational complexity of the model, since a direct application of the 1-D DWM to multidimensional spaces implies the need of calculations at every junction in space for every time-step, growing as the fourth power of the maximum temporal frequency. In order to reduce the complexity of the problem, research has been conducted to explore, as in BEM or WRW, the possibility of using extrapolation techniques in order to obtain the sound field using 2-D models [30, 66], and to efficiently model accurate frequency and angle-dependent wall reflection coefficients [67, 68].

Finite-difference methods

Another time domain solver is the finite-difference time-domain (FDTD) method, a technique originally developed to solve electromagnetic boundary value problems, in an attempt to find solutions to the wave equation in the time domain [69]. In this method, all the (partial) derivatives that occur in the problem are replaced by finite differences. The discretization implies a point-wise evaluation of the sound field at each volume element. In approximating the derivatives with time differences, one must also restrict the *order* of the approximation to a finite value (see, e.g. [63]), with larger values effectively reducing the numerical dispersion, but increasing complexity. In order to better tackle the problem, in [50] the Laplacian operator (the spatial partial derivatives) is approximated by a finite step size in the spatial discrete Fourier domain. The result is then transformed back to the space domain. This is the pseudo-spectral time-domain (PSTD) method. The technique allows to reduce the dispersion errors even for meshes with sample rates approaching two times the maximum frequency (the Nyquist rate). The main drawback is that such approximation only holds for spatially periodic sound fields (because of the periodicity-discretization relation of the discrete Fourier transform [21]), and therefore errors can be seen, especially at the boundaries where the aperiodic transition of the waves is more prominent. Although these errors can be controlled using "perfectly matched layers" schemes [70], time update is performed using standard time-stepping (normally of the second-order), and therefore errors due to temporal approximation still occur.

An important method that reduces numerical dispersion at low sampling rates (potentially improving both speed and accuracy) is the adaptive rectangular decomposition (ARD) method, originally proposed in [71] and further extended to allow highly parallel execution on GPUs in [63]. In ARD the space is first partitioned into small rectangular volumetric elements at a resolution that guarantees minimum or no dispersion errors and other numerical inaccuracies, and is close to the Nyquist frequency. Afterwards it groups the different elements into two types of (also rectangular) zones. The so called “air” zones are destined to model sound propagation in air, whereas the “perfectly matched layer” (PML) zones are used to model (partially or totally) absorptive surfaces. These two types of zones are 3-D. Besides these zones, 2-D interfaces are created between adjacent air-air and air-PML zones. The boundaries for all zones are assumed rigid (fully reflective boundaries), so that the sound field naturally displays a spatially periodic behavior [22, 31, 72]. The wave equation is therefore effectively solved in the spatial discrete Fourier domain using a cosine basis, and can be efficiently implemented via 3-D FFTs and IFFTs, avoiding numerical dispersion at minimum spatial sampling rates. Absorption is implemented in PML zones, with the sound field also decomposed into a cosine basis, but modeled by a highly dissipative (non-physical) wave equation. At all interfaces, a purely time-domain set of finite-difference equations is formulated, which effectively controls the behavior at the boundaries. This still introduces a certain amount of numerical dispersion, but since the interfaces are only 2-D, one can afford higher order representations to minimize the error. The actual computation is then performed on a per-time-frame basis, updating the values of the zones and the interfaces at each time-step. Although the method does not attempt to model exact boundary conditions (i.e. the real frequency and angle dependent wall reflection properties), in [71, 73] the performance of the method is shown to be of enough perceptual quality to be used in broadband (massive) multichannel RIR simulation for computer games or sound field visualization purposes. Moreover, the computation times for single core machines are in the order of hours ($\sim$4 - 6), and for massive multicore (GPU) implementations [63], in the order of minutes ($\sim$18).

2.2.6 Conclusion wave-based methods

In room acoustics, wave-based methods play a very important role. As a mathematical and physical framework, wave theory provides an insightful and rigorous treatment of the underlying dynamics. This description can be made as rigorous as our comprehension of the underlying physical problem goes, increasing the accuracy and realism of the model with every added detail. Unfortunately, the complexity of the problem often increases to a point where no solutions can be found, either analytically or numerically, with reasonable effort. In order to keep a balance between accuracy and complexity, linear versions of the wave equation, e.g. (2.2) or (2.9), are used (see [24] for an example of a non-linear version). The speed of sound is most of the time assumed to be constant, which directly implies that the inhomogeneities due to air humidity and temperature are ignored. In some cases, the reflective properties of the walls and objects in the room are assumed surface, angle and sometimes even frequency independent. Despite these simplifications such models are accepted to be accurate enough for, e.g. predicting the subjective acoustical characteristics of spaces such as concert halls, for testing the absorptive parameters of the walls in a room, or for creating virtual acoustic scenarios for applications like video games (see e.g. [74]). Numerical methods based on wave-theory are thus characterized as being able to model complex scenarios across different room geometries and reverberant conditions, generating realistic sound fields. Another advantage of wave methods is that in many cases the RIR at many positions can be easily obtained. This is because wave motion and wave reflection are modeled as spatio-temporal phenomena and solutions are then normally obtained on a dense grid of spatial points; each of these solutions represents an individual RIR. As previously noted, the degree of accuracy and flexibility of these methods is, however, paid with a high level of complexity. The numerical burden increases particularly rapidly as a function of the desired temporal bandwidth. Because of this, wave-based methods have earned fame of being prohibitively complex, limiting a widespread use. This may change in the near future as latest research in wave methods, such as the ARD method discussed above [63], is starting to approach real-time simulation of multichannel RIRs.

2.3 Geometrical acoustics

To obtain practical methods for modelling RIRs, often simplified models are used. One greatly simplified theoretical description, called geometrical acoustics [22], which has its origins in geometrical optics [75], replaces the concept of a sound wave with the concept of a sound ray. In the limit of high frequency, the wave equation reduces to a geometric description [23]. This is, wave propagation is modeled as *rays* of sound energy quanta, that specularly reflect when encountering obstacles. The description thus assumes that the sound waves have significantly smaller wavelengths than the size of the reflecting objects or surfaces. In its original description, geometrical room acoustics neglects the effects of diffusive reflection, refraction, diffaction and interference [22]. Because of these simplifications and the high frequency assumption, algorithms based on this description achieve less realistic solutions, but at least they aim to provide good quantitative/qualitative results. One particular drawback of this framework is that as a linear sound trajectory is assumed, if a receiver moves from an area "visible" for the sound rays to an area occluded by an object, a sudden discontinuous change in acoustic pressure is observed. This phenomenon

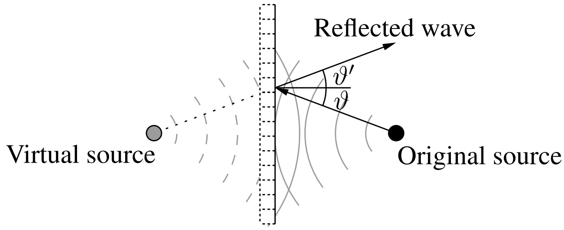


Figure 2.6: Specular reflection is equivalently modeled with a virtual image (copy) of the original source positioned at the far side of the reflective boundary.

is physically impossible, since the sound field in a zone free of sources must be a continuous function (actually a smooth function [23, 37]) of space and time. To conceal this error, techniques have been developed to account for sound diffraction (e.g. [76]), at the cost of added complexity. Most geometrical acoustics methods are then faced with the challenge of fast and accurate computation of propagation rays, especially in the case of complex room geometries. In many cases, geometrical methods allow to model single RIRs at one spatial point independently of other spatial points, contrary to the majority of wave-based methods where spatial grid discretization of the waves imply calculations at several spatial points at the same time. This property might be seen as a complexity advantage if only small sets of RIRs are needed, but might impose a computational disadvantage if large amounts of RIRs are required to, for example, model entire virtual acoustic spaces. There exist however, some more advanced geometrical methods (e.g. [77]) that allow fast recalculation of the RIR at different spatial positions, a very useful property in virtual walkthrough applications where e.g. a moving receiving position is to be modeled.

2.3.1 The mirror image source method

Introduced by Allen and Berkey [78] as a digital computer algorithm, and later improved by Peterson [79], the mirror image source model (MISM) or image method models reflections using (virtual) image sources at the far-side of the boundaries emulating the law of optical specular reflection [22, 75], see Fig. 2.6. When the sound field is enclosed between two parallel reflective boundaries, the reflections involved can be described in terms of the recursive mirroring of the original source, generating an (in principle) infinite set of virtual sources outside the boundaries in the direction perpendicular to the planes. Allen and Berkey considered the case of box-shaped rooms with parallel walls positioned perpendicular to each Cartesian co-

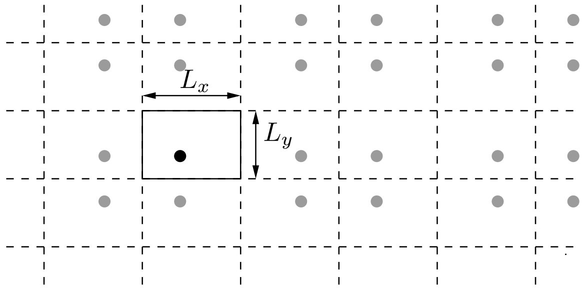


Figure 2.7: A (2D) top-view of the set of virtual source images (gray circles) generated by the Allen and Berkley's image method inside a box-shaped room [78, 79]. The solid line rectangle denotes the room with dimensions L_x, L_y and L_z. The real source is depicted by the black circle inside the room.

ordinate. The walls are assumed to have constant reflection factors. In this case, virtual images of the original source are created in 3-D space, forming a periodic set over a lattice [22, 72]. In Fig. 2.7 a (2D) top-view from the z direction of this set is depicted. Their algorithm calculates the position of any virtual source and the amplitude and time-delay its sound field has at the measuring position (inside the room), and then adds the contribution to the total sound field. The contributions of the virtual sources can arrive at arbitrary times, the RIR, however, is calculated in discrete-time. To solve this problem Allen and Berkey associate the contributions to the nearest sample time. This introduces interaural (spatially related) phase errors in the computed RIR. In [79], Peterson proposes to filter the delta-pulse contribution of each virtual source using a Hamming-windowed, ideal low-pass filter with cut-off frequency set to the Nyquist rate, and to use the original arrival times of the contributions. This preserves the phase coherence between RIRs at different measuring possitions. In this way, when all possible virtual sources are considered, all possible reflection paths from the source to a receiver are covered and an RIR is formed. Using this scheme it is not possible to model an infinite amount of sources, and one stops calculating virtual sources after the total sound field decays beyond a certain threshold. One popular threshold is the reverberation time (T_{60}), defined as the time it takes for the sound field energy to decay 60 dB from its original value. The T_{60} was proposed by Wallace C. Sabine [38] in the early 20$^{\text{th}}$ century following his subjective experiments. Expressing the reverberation time in samples N_{60}, the MISM simulates the RIR from one source to one measurement point in $\mathcal{O}(N_{60}^3)$ operations, making it a relatively fast algorithm of special interest in simple scenarios when faster compu-

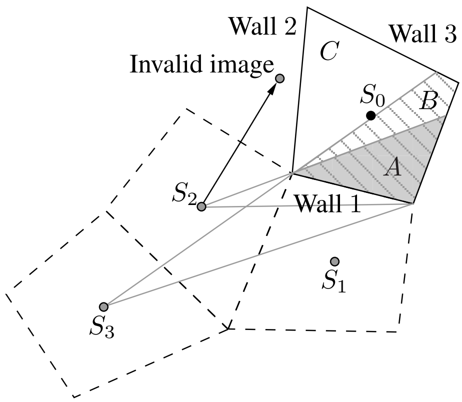


Figure 2.8: A particular reflection sequence in the extended image method algorithm by Borish [84]. The source S_0 is mirrored over wall 1 to create source S_1, S_1 is reflected over wall 2 (of the first virtual room) to create S_2, and S_2 is reflected over wall 3 (of the second virtual room) to create S_3. The visibility zone of S_2 in the real room is given by zone A. The visibility zone of S_3 is given by the union of zones $A \cup B$. For a receiver positioned in zone C both the virtual sources S_2 and S_3 are invisible. An example of an invalid virtual image is depicted.

tations take precedence over accuracy of the simulation. Although the MISM is an old algorithm, the approach is of theoretical and practical importance. Many methods have been introduced in the past that extend the MISM to, e.g., model diffraction [80] and diffuse reflection [81], to synthesize the RIR per frequency band [82] allowing for frequency-dependent (complex-valued) reflection coefficients, to model interference through complex superposition [76], and to improve the speed of the method by the use of efficient virtual source identification and search through look-up tables [83]. Moreover, an extension of the method to model RIRs for arbitrarily shaped (piecewise-planar) rooms has been derived [84].

In the extended MISM proposed by Borish et al. [84], the sound sources are mirrored over each piece-wise linear surface in the enclosure. Virtual sources are considered individually and are iteratively mirrored up to the desired reflection order. Unlike the case of box-shaped rooms, a check for validity and visibility is needed for each newly created mirrored source at the cost of extra complexity. For example, a virtual source generated by reflection over the outer side of a wall cannot be valid. Further, when a virtual source's path is obscured by another surface, or when the sound path from the virtual source to a receiver position intersects a point on a wall plane outside the boundaries, the source is deemed invisible. After detection, invalid sources can be discarded and their recursive mirroring stopped. On the other hand, descendant generation of invisible sources has to be continued, since higher-order mirrored descendants of invisible sources may have visible paths to the receiver(s). In Fig. 2.8 an example reflection sequence is given. The source S_0 (and the room) are mirrored over wall 1 to create source S_1, S_1 is then reflected over wall 2 (of the first virtual room) to create S_2, and then S_2 is reflected over wall 3 (of the second virtual room) to create S_3. The visibility zone of S_2 in the real room is given by the gray-shadowed zone A. The visibility zone of S_3 in the real room is given by the union of the gray-shadowed zones $A \cup B$. For a receiver positioned in zone C both the virtual sources S_2 and S_3 are invisible. In zone B the sound field of S_2 does not contribute to the RIR, but the generation of S_2 is important as this virtual source is the progenitor of S_3, which does contribute to the RIR in zone B. An example of an invalid virtual image is also depicted. In this extendend MISM, the complexity involved in the visibility and validity tests is higher than the generation of the response itself [84–86]. More strikingly, with a fixed number of (planar) walls, say M, and a given number of desired reflections proportional to N_{60}, the number of virtual sources that have to be generated in the general case is upper bounded to $\mathcal{O}(M^{N_{60}})$, which is now exponential in N_{60}. The symmetry of the rectangular case is lost and different descendant paths in the recursion may lead to different virtual sources. Even when holding the number of walls constant, distorting the shape of the room space would affect the total number of virtual sources that have to be generated for a given N_{60}. Moreover, all possible reflection paths across all boundaries per virtual source must be tested for validity and visibility. In the limit, as the number of planar walls goes to infinity (which could be made to converge to a room with curved boundaries), the algorithm would accordingly need to model an infinite amount of virtual point sources even at low-order reflections. As a consequence, this extended MISM is not suitable for modeling reflections due to curved surfaces.

The basic MISM guarantees that all specular reflection paths (up to the desired order) from the source(s) to the receiver(s) are found. In non-convex rooms, or when objects are present in the environment, diffraction (and to a given extent transmission) cannot be neglected and therefore a diffraction model has to be used at the cost of increased complexity. Because of the added overall complexity of the extended MISM approach, it is mainly used to model lower orders of specular reflections in complex geometries/environments.

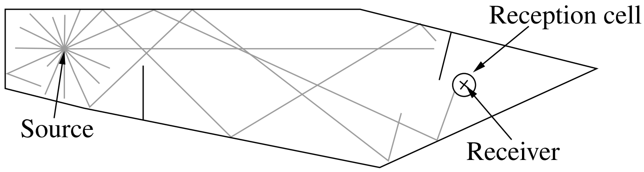


Figure 2.9: Ray tracing method in 2D. A finite set of rays emanating from the source are extended linearly and via specular reflection (traced). The tracing process finish when the ray reaches the receiver cell (a volumetric cell in 3D) or when e.g., the maximum reflection order has been reached. The method can be extended to trace ray paths of diffuse reflection, diffraction and refraction.

2.3.2 Ray tracing

The first credited computer algorithm to simulate RIRs (as mentioned at the beginning of Sec. 2) was based on ray tracing [45], which was during many years preferred over other methods due to its algorithmic simplicity, particularly for complex room geometries [84]. In ray tracing a set of sound rays is generated at the source of sound (or the receiver position) emanating uniformly in all directions. Each ray is extended linearly and via specular reflection until it reaches the zone of space surrounding the receiver (or source) position. In this way, more realistic simulations can be obtained as not only paths of specular reflection can be modeled, but also paths of diffuse reflection, diffraction and refraction can be generated and traced even for surfaces with curved geometries [77, 87]. A RIR is thus formed by adding properly attenuated contributions of the calculated paths. Since it is impossible to generate an infinite number of rays (covering the 4π steradians around the point source), the method will always find a certain number of paths (for the generated rays), but these are not necessarily a complete set of paths for the desired simulation time slot (e.g. until T_{60} is reached).

This can be seen (and it is defined) as an *aliasing* problem caused by the sampling of a continuum of 4π steradians space of possible rays. Important propagation paths might be missed by all samples. The countable amount of rays used and the fact that receiver (or source) positions are approximated by volumes in space lead to fundamental problems [88]. For instance, if the receiver is modeled with a large volume, a ray that was supposed to pass close, but not yet exactly at the receiver position could have been taken into account at the wrong reflection order. On the other hand, if the receiver is modeled by a small enough volume, chances are that the chosen rays would never hit the receiver at all. This happens because, in the limit,

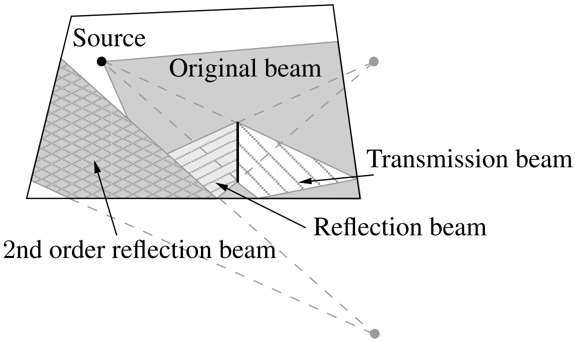


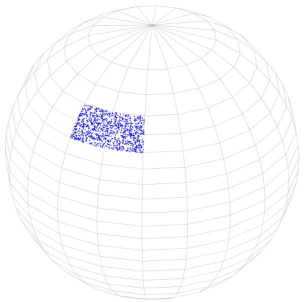
Figure 2.10: Beam tracing.

when a volumetric receiver zone becomes a point receiver, an uncountable amount or rays (at a continuum of emanating directions) is needed to guarantee that the receiver is reached by the rays at their correct reflection orders. Although one would think that the volumetric representation of the receiver is actually a more realistic model than a point receiver, the actual problem lies in the fact that there is no way to know whether a given path has been missed. These anomalies and errors have been observed in real-life experiments [88, 89]. Still, ray tracing can give accurate results (up to any desired threshold) if a good ray generator is used and enough rays are employed. An example of a good ray generator model is random generation using a Monte Carlo approach. For the generation to be accurate and to ensure convergence as the number of rays tends to infinity, a good quality pseudo-random generator must be employed. Uniform (or close to uniform) ray generation on the surface of the unit sphere must be ensured [90]. Using three independent random generators for the three Cartesian spatial coordinates would not give the required results, as this produces a "cube of rays" instead of a sphere. Another possibility, as used in e.g. [90], is to subdivide the sphere surface into a large number of polygons of equal area, and to cover each of them with random rays (see Fig. 2.11a). This ray generator method is reported to be accurate and reliable provided a very large number of rays is generated (usually more than 10^5) [90].

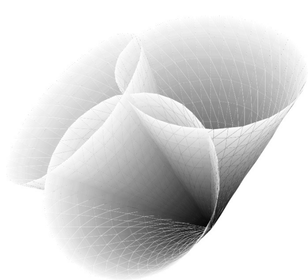
2.3.3 Beam tracing

Despite the flexibility and relative simplicity of ray tracing, aliasing due to sampling the continuum space of possible rays remains an important problem. In order to solve the problem, beam tracing was proposed as a natural generalization [91, 92], first in the context of computer graphics [77, 90, 93], and later applied to acoustic simulations. Instead of using infinitely thin rays, in beam tracing volumetric "beams", i.e. solid-angles with finite polygonal cross-sections, are traced through the environment to find propagation paths from a source to a receiver. In this way, the full 4π steradians around a source can be covered exactly with a finite amount of non-overlapping beams. Fig. 2.11c shows an example using pyramidal beams [77, 90, 94]. Intersections with polygons in the scene are evaluated in backward order, so that no propagation-reflection on a polygon is included in the propagation-reflection sequence until all others that at least partially occlude the polygon have been first considered. After intersections with polygons are identified, the original beam is clipped and (optionally) a transmission-refraction beam is created that exactly matches the occluded region. A reflection beam is then constructed by mirroring the transmission beam over the polygon's plane see, Fig. 2.10. Using the same approach, sets of diffraction beams can be added following a given diffraction model to obtain more realistic results. After the propagation-reflection sequence of each beam that intersects the receiver is found, an algorithm to find the exact paths from the source to the receiver is used (e.g. [90, 91, 94]). In this way, beam tracing eliminates the problems related to sampled rays and allows intersections with infinitesimal (point) sources, solving the problem of false path detection. The main disadvantage of beam tracing is the complexity of the geometric operations (i.e. intersection and clipping) to trace the beams in 3D spaces, as each beam is to be reflected and/or obstructed many times by the surfaces present. Therefore, the main challenge of beam tracing is not only to develop techniques to trace the beams fast and robustly but also to generate the actual propagation paths as efficiently as possible.

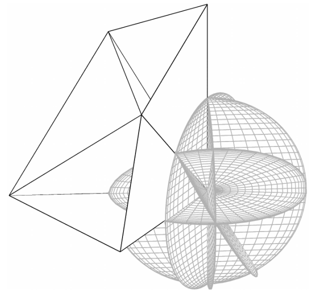
Research in beam tracing methods has shown progress in reducing the complexity of the algorithms (e.g. [77]), allowing RIR modeling at interactive rates for e.g., virtual modeling of a moving receiver in large and complex room geometries. To achieve this, Funkhouser et al. [77] separate the problem into a computationally intensive off-line preprocessing step and a sufficiently lightweight on-line stage. During the off-line process the space is carefully partitioned, different propagation-reflection-transmission-scattering sequences of beams are constructed for each source and stored in special beam-tree structures optimized for fast look-up of critical parameters. In the on-line stage, the propagation paths are efficiently calculated from the *a priori* gathered information and the RIR is formed. In this stage, the precomputed tables are first queried for beam sequences of transmissions, reflections and diffractions potentially reaching the receiver position. Particular paths are then computed by finding the shortest among all possible piecewise linear paths from the source to the receiver inside the beam path. This is done by considering points of



(a) *A ray tracing approach. The sphere surface is subdivided in polygons of equal area. Each area is uniformly covered by a set of randomly generated rays. Here four faces are covered with* 250 *rays each (blue dots).*



(b) *Cone tracing beams. The space of possible rays around the source cannot be totally covered without overlapping of the beams.*



(c) *An example of a non-overlapping partition of the space of rays as used in pyramid tracing [90]. The resulting beams are triangular pyramids.*

Figure 2.11: Examples of sphere surface covering approaches in different beam tracing methods.

intersection of the path with every face and edge in the beam sequence. Additionally, fast convolution of the RIR with an arbitrary time-domain signal can be performed for interactive auralization purposes.

Ray tracing can be seen as a special case of beam tracing (using beams with a point cross-section) see Fig. 2.11a. Another technique called cone tracing [91] uses beams with circular cross-sections, i.e. cones. This technique is therefore also a special case of beam tracing. Using cones as beams, however, it is impossible to cover the entire space around the source without gaps or overlap, see Fig. 2.11b. If gaps are allowed, it is likely that important paths will be missed. On the other hand, if overlap is allowed, computer power is wasted as some paths are calculated multiple times. Some techniques have been proposed to address this problem by, for example, using windowing schemes (spatial weighting) across the cone faces to better account for the overlap [92]. Redundancy is, however, not totally eliminated.

2.3.4 Acoustical Radiosity

Although the theory of radiosity has been historically derived and used in other fields of science, e.g. optics [95, 96], thermal engineering [97–100] and computer graphics [101–105], the first formulation of acoustic radiosity is given by Kuttruff [22], where he studied the reverberation properties in rectangular rooms with totally diffusive (Lambertian) walls [22, 26]. Totally diffusive reflection is independent of the direction of the incoming ray (see Fig 1.1c). This has the advantage that reflections do not need to be followed (traced) explicitly. The acoustic response is found in terms of the total radiation density at the receiver given the energy exchanges between the source and the enclosure. Neglecting (as in all geometrical methods) the effects of phase coherence of the sound waves, Kuttruff derived an integral equation that characterizes this exchange in acoustic energy due to a source and all diffusive reflections in the room. Because of this, the approach is also referred to as “the integral equation method”. A numerical solution of the integral equation for rectangular rooms is given in [106]. There, an exponential time-decay of the radiation density is proven. Moreover, a solution is derived for both time-varying sources and steady-state sound fields. In its original form, radiosity does not track specular reflections (not even partially), and therefore simulations using this method where found to be similar to those using ray tracing with totally diffuse reflections [107]. Although not realistic [26], the assumption of Lambertian reflection has been suggested to be less restrictive than the assumption of purely specular reflection (as in the other geometric approaches) [41, 108–110]. Besides, it is known that the fine structure of the RIR becomes more diffusive as a function of time [22, 26] (it is in fact this property that motivates some other hybrid and/or statistical methods). The accuracy of radiosity at modeling the late-reverberation part of the RIR is explored in [41, 110, 111], where it is indeed confirmed that the method fails to model the fine and sharp detail of the early reflections of the RIR, but matches more accurately the late reverberation part. Improvements on the basic algorithm have been proposed in e.g. [41]. In [112], Le Bot proposes a modification of the integral equation (in the form of a functional equation) with the aim to include specular reflection in the model. In [94], Lewers proposes a combined beam tracing-radiosity method. In general, the main challenge of radiosity is the efficient solution of the underlying integral equation at every time step. A heavy off-line, fast on-line, algorithmic partition is suggested in [26] to allow interactive modeling rates for e.g. of moving receivers. More recently, the underlying integral equation of the radiosity method has been generalized, giving rise to a theoretical framework called *the room acoustics rendering equation* [87] that encompasses several geometrical acoustic algorithms, being able to naturally model arbitrary levels of diffuse and specular reflection.

2.3.5 Conclusion geometrical methods

The geometrical acoustics description is in its original basic form constitutes a simplified model that in many cases leads to fast broadband simulations and is known to give good results under rather stringent assumptions. Many methods include approaches to model diffractions, transmissions, and to a lesser extent diffuse reflections (with the exception of radiosity which by construction only models diffuse reflections), increasing realism and accuracy at the expense of the corresponding penalty in complexity. One of the major challenges of geometrical methods is efficient and accurate computation of the propagation paths. In complex scenarios, the number of reflection, transmission and diffraction sequences increases very rapidly as a function of time [77], and becomes expensive to compute.

2.4 Hybrid methods

To be useful in practical applications, the simulation of the RIR ought to be as realistic and fast as possible. From the previous discussion we can summarize that wave-based methods provide a mathematically rigorous description, but the proposed numerical solutions are complex, their complexity increasing most drastically as a function of temporal frequency. On the other hand, the geometrical framework provides a high-frequency approximation that increases its complexity mainly as a function of the desired reverberation time.

It is then not surprising to see that from the above described methods, state-of-the-art candidates (e.g. [73, 77]) are not purely based on a single approach, but combine aspects of different models in order to bring together the contradictory goals of speed and accuracy. Moreover, commercial acoustic simulation software such as ODEON [113–116], CATT [81, 117], EASE [118], RAMSETE [90, 119], Bose Modeler [120], and research projects such as MCRoomSIM [121] or DIVA [122], use hybrid approaches or combine methods using spectral or temporal thresholds in order to generate broadband results over long simulation times.

Some hybrid methods have been proposed that combine only geometrical approaches. As mentioned in Sec. 1.2, the density of reflections in the RIR grows quadratically with time, giving rise to the distinction between early reflections and late reverberation. With the exception of radiosity, this clearly has a direct impact on all geometrical methods. Hybrid methods are then proposed to ease the computations needed, and/or to increase the accuracy of the solutions. For example, exhaustive search for valid and visible virtual sources in the extended MISM is avoided by

Vorländer [85] using the more efficient reflection path calculation of ray tracing to identify the virtual sources. Lehmann and Johansson propose in [123] a combination of an image model to model only early reflections of the RIR and to use a diffuse reverberation model for the late reverberation part of the RIR. In this way the computational complexity of the creation of higher order virtual images is avoided, while still providing a perceptually accurate simulation performance.

2.5 The newly proposed GFD method

As described in Sec. 1.3, the main contribution of this thesis is the introduction of an efficient method to simulate very large amounts of RIRs covering the entire room space, i.e. to simulate the sound field given a set of sources in the room evaluated on a dense grid of listening positions. The method is motivated by the observation that the sound field in a box-shaped room can be modeled using a periodic spatial structure of virtual sources (see Fig. 2.7). This is the same structure that models reflections in the MISM. In principle the set of virtual sources is infinite, and therefore, in all algorithms based on the MISM a finite subset of virtual sources must be chosen and its contribution to the sound field computed in one way or another. The more the virtual sources considered, the more accurate the computed sound field becomes. As it is explained in Sec. 2.3.1, the density of virtual sources in box-shaped rooms increases as the cube of the reverberation time. This translates to a high computational complexity in the computation of the virtual sources, especially for large reverberation times.

If the room has perfectly reflective walls, the sound field generated by the virtual sources is also periodic [22, 72]. Moreover, in the original paper by Allen and Berkley describing the MISM [78] it is shown that in this case the periodic summation of the individual sound fields is equal to the synthesis of the room modes as given in modal analysis (see Sec. 2.2.2). The sound field generated by the virtual source model is then an exact solution to the wave equation with the given boundary conditions of perfectly reflective walls.

It is well known from Fourier analysis that sampling in one domain results is an (infinite) periodic summation in the reciprocal domain [21, 32, 37, 124]. The key idea behind the newly proposed method exploits this property. The Fourier transform of a free-field sound field is first calculated analytically. This spectrum is then carefully sampled so that this sampling corresponds to the periodic sum of virtual sources in the reciprocal domain. In this way, the contribution of all virtual sources is considered at once, and no truncation of the number of virtual sources as in MISM related methods is needed. Further, all relevant quantities are made discrete resulting in a DFT-based algorithm. The FFT is then used to compute the core components of the algorithm.

This model, however, has no practical application since it is only valid for perfectly reflective walls. Consequently, a generalized Fourier domain (GFD) is proposed. The sound field in a room with walls with complex-valued reflection coefficients can be related to a sampling condition is this domain. A fast implementation of the generalized discrete Fourier transform (the generalized fast Fourier transform, GFFT) is also described. A low-complexity multichannel RIR simulation algorithm is then derived that can model frequency and angle dependent, complex-valued (and thus can model absorption and phase changes) reflections. The method computes the sound field at a dense grid of receiver positions and it is thus best suited for the computation of large amounts of RIRs covering the entire room space. The GFD method is compared against MISM to validate that the model indeed corresponds to the exact periodic summation of virtual sources. The results of these comparisons and the detailed technical analysis are given in Part II.

In this thesis, the GFD method has been implemented for box-shaped rooms. The algorithm in its current form can model walls with frequency-dependent, angle-dependent, constant (as a function of the wall surface) reflection coefficients. Although the method has been implemented for box-shaped rooms only, the theory is derived for any convex room geometry that tessellates the Euclidean 3-D space in a periodic packing. It has been shown that some convex polyhedra, such as the equilateral triangular prism or the hexagonal prism, have this property [125]. Therefore, the method can be readily applied on those geometries.

For the GFD method to be of widespread use in addressing acoustic problems for next-generation communication technologies, are still many challenges to overcome. To model rooms with arbitrary geometries, to include reflection coefficients that are function of the wall surface, and to model objects and changing conditions in the enclosure are important extensions to the model, and the topic of of current research.

2.6 Conclusion

The combined research efforts in different fields, such as seismology [24], underwater and holographic acoustics [35], signal processing [31, 82, 83], computer vision [64, 71, 73, 77], digital musical-instrument simulation [57], room acoustics [22, 75], and the new possibilities open by massive multicore parallelization [34, 63], have led to a plethora of algorithms for room acoustics simulation, – the generation of audio signals such that when played, they mimic as accurately as possible the impression of actually being in the room space. This directly or indirectly involves modeling of the RIR at different spatial positions. Historically, wave methods where used only as theoretical formulations that provided insight into the problem [22, 23], and although wave theory can be used to model all possible room sound fields, the lack of efficient computational methods has confined the theory to the simulation of low-frequency sound fields. On the other hand, the first computational methods to model RIRs where based on the simpler geometrical acoustics [45, 78, 79], but this framework is only realistic enough in the high-frequency range. Nowadays, this binary distinction is not valid. In the quest to derive realistic and fast numerical methods, the newest algorithms combine concepts, techniques and representations of wave and geometrical acoustics alike. A real-time solution that offers flexibility and realistic accuracy is still a major challenge.

As a final summary a comparison between all the algorithms reviewed in this chapter is given in the table below. The algorithms are evaluated with respect to some key properties discussed in this chapter using a discrete scale with seven steps: $(+++)$, $(++)$, $(+)$, (0), $(-)$, $(--)$ and $(---)$, where $(+++)$ denotes a very positive score and $(---)$ denotes a very negative score. The properties evaluated are the following:

- *Computational complexity.* It refers to how efficient an algorithm is in practice. It does not represent a purely objective score since the practical complexity of an algorithm is determined by many factors. For example, some algorithms can be used for real-time RIR simulation by factoring the computation into a complex off-line stage followed by efficient on-line computations. The scores are thus given based on total average complexity.

- *Accuracy*. It refers to how accurate an algorithm models the acoustical phenomena described in Sec. 1.2. In other words, how realistic the method is.

- *Arbitrary geometry*. It is an indication on how flexible an algorithm is with respect to the boundary conditions. It quantifies how arbitrary the room geometry is allowed to be and also if objects can be included in the model.

- *Massive multichannel*. It gives a score on how suitable an algorithm is to model massive amounts of RIRs. This is also related to the low-complexity score since the less complex an algorithm is, the more apt it is to model many RIRs.

Properties / *Method*	Low Complexity	Accuracy	Arbitrary Geometry	Massive Multichannel	Comments
Modal Analysis (Sec.2.2.2)	− − −	+ + +	+ + +	+ + +	General but impractically complex. Used to model narrowband RIRs for simple room shapes.
WRW (Sec. 2.2.3)	−−	+ + +	++	++	Flexible. Suitable to model RIRs at planar arrays of receiver positions.
FEM (Sec. 2.2.4)	− − −	+ + +	+ + +	+ + +	Apt for intricate room shapes but complex. Used to model RIRs in small rooms or narrowband RIRs.
BEM (Sec. 2.2.4)	− − −	+ + +	+ + +	+ + +	Better to model larger rooms than FEM, but still complex.
BEM-FMM (Sec. 2.2.4)	−	++	++	+ + +	The use of fast multipole methods (FMM) yields less complexity than the simple BEM.
DWM (Sec. 2.2.5)	− − −	++	+ + +	+ + +	A problem to overcome in DWM is the error caused by *numerical dispersion*.

Continued on next page.

Table 2.1: Comparison of RIR simulation methods (Part 1 of 5).

Properties *Method*	Low Complexity	Accuracy	Arbitrary Geometry	Massive Multichannel	Comments
FDTD (Sec. 2.2.5)	− − −	++	+ + +	+ + +	As in DWM numerical dispersion is a challenge.
PSTD (Sec. 2.2.5)	+	++	++	+ + +	Same principle as FDTD, numerical dispersion is reduced by approximating the derivatives via the DFT.
PML-PSTD (Sec. 2.2.5)	+	++	++	+ + +	Perfectly matched layers (PML) are used to better emulate absorption and reflection at the boundaries.
ARD-PSTD (Sec. 2.2.5)	++	++	++	+ + +	Adaptive rectangular decomposition (ARD) is used on top of PML-PSTD to added flexibility and efficiency.
MISM (Sec. 2.3.1)	0	−−	− − −	− − −	Simple algorithm. Suitable for simulations of individual RIRs in empty box-shaped rooms and for estimation of acoustic room qualities.
Polyhedra MISM (Sec. 2.3.1)	−−	−	++	− − −	Generalization of the MISM to rooms shaped as arbitrary polyhedra.

Continued on next page.

Table 2.1: Comparison of RIR simulation methods (Part 2 of 5).

Properties / *Method*	Low Complexity	Accuracy	Arbitrary Geometry	Massive Multichannel	Comments
Sub-band MISM (Sec. 2.3.1)	−	+	− − −	− − −	Extension of the MISM to simulate frequency-dependent wall reflection/absorption.
Diffraction MISM (Sec. 2.3.1)	−	+	− − −	− − −	Generalization of the MISM to also simulate diffraction.
Diffuse MISM (Sec. 2.3.1)	−	+	− − −	− − −	Extension of the MISM to simulate diffuse reflection.
Look-up tables MISM (Sec. 2.3.1)	+	−−	− − −	0	The use of look-up tables reduces redundancy in the operations needed by the MISM, increasing efficiency.
FMM-MISM (Sec. 2.3.1)	++	+	−−	+ + +	Efficient algorithm that uses fast multipole methods (FMM). Apt for multichannel RIR simulation.

Continued on next page.

Table 2.1: Comparison of RIR simulation methods (Part 3 of 5).

Properties *Method*	Low Complexity	Accuracy	Arbitrary Geometry	Massive Multichannel	Comments
DLRM-MISM (Sec. 2.4)	++	− − −	0	+	The MISM is combined with a diffuse late reverberation model (DLRM) to increase efficiency. Some accuracy is lost but subjective quality is retained.
Ray tracing (Sec. 2.3.2)	+	− − −	++	0	Historically the first RIR computer algorithm. Suffers from discretization errors. Suitable for estimation of acoustic room qualities.
MISM-RT (Sec. 2.4)	+	0	+	0	Ray tracing (RT) is combined with the MISM to reduce complexity and decrease the discretization error in RT.
Beam tracing (Sec. 2.3.3)	+	++	+ + +	++	A generalization of ray tracing. It solves the ray discretization problem. Accuracy depends on the chosen beam geometry.
Cone tracing (Sec. 2.3.3)	+	+	++	++	A special case of beam tracing. The radiating space around a source cannot be totally covered without beam overlap.

Continued on next page.

Table 2.1: Comparison of RIR simulation methods (Part 4 of 5).

Properties / *Method*	Low Complexity	Accuracy	Arbitrary Geometry	Massive Multichannel	Comments
Pyramid tracing (Sec. 2.3.3)	+	++	++	++	The beams are triangular pyramids. It solves the beam discretization problem without beam overlap.
Acoustical radiosity (Sec. 2.3.4)	+	− − −	+ + +	++	It models totally diffuse reflection. Suitable to model the late reverberation part of the RIR.
Functional equation (Sec. 2.3.4)	+	+	+ + +	++	A generalization of acoustical radiosity to also model specular reflection.
Rendering equation (Sec. 2.3.4)	0	+ + +	+ + +	+	Generalized theoretical framework that encompasses several geometrical acoustics methods. It models arbitrary levels of diffuse and specular reflection.
BT-radiosity (Sec. 2.4)	+	+	++	++	Beam tracing is combined with radiosity to better model specular and diffuse reflections.
GFD method (Sec. 2.5) and (Part II)	+ + +	++	− − −	+ + +	Newly proposed method. Computes massive amounts of RIRs. Low complexity order per computed RIR. Limited to box-shaped rooms.

Table 2.1: Comparison of RIR simulation methods (Part 5 of 5).

Bibliography

[1] J. Emparanza, P. Dadlani, B. de Ruyter, and A. Härmä, "Ambient telephony: Designing a communication system for enhancing social presence in home mediated communication," in *Proc. ACII*, Sept. 2009, pp. 1 –8.

[2] A. Härmä, "Ambient telephony: scenarios and research challenges," in *Proc. Interspeech*, Antwerp, Belgium, Aug. 2007, pp. 562–565.

[3] A. Härmä, S. van de Par, and W. Bruijn, "On the use of directional loudspeakers to create a sound source close to the listener." in *AES 124th Conv.*, May 2008, paper 7328.

[4] A. J. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Amer.*, vol. 93, pp. 2764–2778, 1993.

[5] J. Ahrens and S. Spors, "Analytical driving functions for higher order ambisonics," in *ICASSP*. Las Vegas, Nevada, USA: IEEE Signal Processing Society, 31 March - April 4 2008, pp. 373 –376.

[6] M. Kolundzija, C. Faller, and M. Vetterli, "Sound field reconstruction: An improved approach for wave field synthesis," in *AES 126th Conv.*, May 2009.

[7] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, 1st ed., ser. RWTHedition Series. Berlin, Germany: Springer, 2008.

[8] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 53, no. 11, pp. 1004–1025, 2005.

[9] A. Swami, Q. Zhao, and Y.-W. Hong, *Wireless Sensor Networks: Signal Processing and Communications*. England, UK: Wiley, Dec. 2007.

[10] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*. West Sussex, England: John Wiley & Sons, 2006.

[11] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing (Signals and Communication Technology)*. New York: Springer-Verlag, 2006.

[12] J. Benesty, D. Morgan, and M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[13] T. Gänsler and P. Eneroth, "Influence of audio coding on stereophonic acoustic echo cancellation," in *ICASSP*. Seattle, Washington, USA: IEEE Signal Processing Society, May 1998, pp. 3649–3652.

[14] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *ICASSP*. Honolulu, Hawaii, USA: IEEE Signal Processing Society, Apr. 2007, pp. I17–I20.

[15] K. Helwani, S. Spors, and H. Buchner, "Spatio-temporal signal preprocessing for multichannel acoustic echo cancellation," in *ICASSP*. Prague, Czech Republic: IEEE Signal Processing Society, May 2011, pp. 93–96.

[16] H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: acoustic echo cancellation for full-duplex systems based on wave-field synthesis," in *ICASSP*. Montreal, Quebec, Canada: IEEE Signal Processing Society, May 2004, pp. IV117–IV120.

[17] H. Buchner and S. Spors, "A general derivation of wave-domain adaptive filtering and application to acoustic echo cancellation," in *IEEE Asilomar conf. on signals, syst. and computers*. Pacific Grove, California, USA: IEEE Signal Processing Society, Oct. 2008, pp. 816–823.

[18] M. Schneider and W. Kellermann, "A wave-domain model for acoustic mimo systems with reduced complexity," in *HSCMA*. Edinburgh, Scotland: IEEE Signal Processing Society, May–Jun. 2011, pp. 133–138.

[19] S. Spors, H. Buchner, and W. Kellermann, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J. Acoust. Soc. Amer.*, vol. 122, no. 1, pp. 354–369, Jul. 2007.

[20] P. P. Vaidyanathan, *Multirate Systems And Filter Banks*. New Jersey: Prentice Hall PTR, 1992.

[21] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms and Applications*, 3rd ed. New Jersey: Prentice Hall, Inc., 1996.

[22] H. Kuttruff, *Room Acoustics*. London, U.K.: Taylor and Francis, 2000.

[23] P. Morse and K. Ingard, *Theoretical Acoustics*. New York: McGraw-Hill, 1968.

[24] A. Berkhout, *Applied seismic wave theory*. Elsevier Science Publ. Co., Inc, 1987.

[25] B. Kapralos, "The sonel mapping acoustical modeling method," Ph.D. dissertation, York University, Canada, 2006.

[26] E.-M. Nosal, "Room sound field prediction by acoustical radiosity," Master's thesis, University of British Columbia, Canada, 2002.

[27] L. Cremer and H. Müller, *Principles and applications of room acoustics*, ser. Principles and Applications of Room Acoustics. Applied Science, 1982, vol. 1.

[28] I. Arweiler and J. M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 130, no. 2, pp. 996–1005, 2011.

[29] International Standard 3382, *Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms*. ISO, Geneva, Switzerland, 2008.

[30] D. T. Murphy, "Digital waveguide mesh topologies in room acoustics modelling," Ph.D. dissertation, The University of York, United Kingdom, 2000.

[31] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling." *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

[32] K. B. Wolf, *Integral Transforms in Science and Engineering*. New York: Plenum Press, 1979.

[33] C. Van Loan, *Computational Frameworks for the Fast Fourier Transform.*, ser. Frontiers in Applied Mathematics. Philadelphia, PA: SIAM, 1992.

[34] F. Franchetti, M. Puschel, Y. Voronenko, S. Chellappa, and J. Moura, "Discrete Fourier transform on multicore," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 90 – 102, 2009.

[35] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. London, UK: Academic Press, 1999.

[36] F. Pinto and M. Vetterli, "Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications," *Signal Processing, IEEE Transactions on DOI - 10.1109/TSP.2010.2052045*, vol. 58, no. 9, pp. 4608–4620, 2010.

[37] R. Kanwal, *Generalized Functions, Theory and Technique*, 2nd ed. Boston: Birkhauser, 1998.

[38] W. C. Sabine, *Collected papers on acoustics*. Cambridge: Harvard University Press, 1922.

[39] C. F. Eyring, "Reverberation time in "dead" rooms," *J. Acoust. Soc. Amer.*, vol. 1, no. 2A, pp. 168–168, 1930.

[40] P. Svensson and A. Krokstad, *Reflections on sound: in honour of Professor Emeritus Asbjørn Krokstad*. Trondheim, Norway: NTNU, 2008.

[41] E.-M. Nosal, M. Hodgson, and I. Ashdown, "Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms," *J. Acoust. Soc. Amer.*, vol. 116, no. 2, pp. 970–980, 2004.

[42] W. J. Cavanaugh, G. C. Tocci, and J. A. Wilkes, *Architectural Acoustics: Principles and Practice*, 2nd ed. Wiley, 1999.

[43] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, 2nd ed. Springer, 2003.

[44] M. R. Schroeder and B. F. Logan, "Colorless artificial reverberation," *J. Audio Eng. Soc*, vol. 9, no. 3, pp. 192–197, 1961.

[45] A. Krokstad, S. Strom, and S. Sorsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound Vib.*, vol. 8, no. 1, pp. 118 – 125, Jul. 1968.

[46] A. Schuster, *An introduction to the theory of optics*, 2nd ed. London, UK: E. Arnold, 1909.

[47] E. Jordan, *Electromagnetic waves and radiating systems*. Constable and Company Ltd, 1950.

[48] L. L. Thompson, "A review of finite-element methods for time-harmonic acoustics," *J. Acoust. Soc. Amer.*, vol. 119, no. 3, pp. 1315–1330, 2006.

[49] Y. Wu and T. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 1, pp. 107 –116, Jan. 2009.

[50] Q. Liu, "The pseudospectral time-domain (PSTD) method: a new algorithm for solutions of Maxwell's equations," in *IEEE Symp. Antennas Propag.*, vol. 1, Jul. 1997, pp. 122 –125 vol.1.

[51] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Design of transform filter for reproducing arbitrarily shifted sound field using phase-shift of spatio-temporal frequency," in *ICASSP*. Kyoto, Japan: IEEE Signal Processing Society, Mar. 2012, pp. 381–384.

[52] A. Berkhout, D. de Vries, J. Baan, and B. van den Oetelaar., "A wave field extrapolation approach to acoustical modeling in enclosed spaces." *J. Acoust. Soc. Amer.*, vol. 105, pp. 1725–1733, Mar. 1999.

[53] B. Xu and S. D. Sommerfeldt, "A hybrid modal analysis for enclosed sound fields," *J. Acoust. Soc. Amer.*, vol. 128, no. 5, pp. 2857–2867, Nov. 2010.

[54] J. R. Wright, "An exact model of acoustic radiation in enclosed spaces," in *AES 96th Conv.*, Feb. 1994.

[55] S. Kirkup, *The boundary element method in acoustics: a development in Fortran*, ser. Integral equation methods in engineering. Stephen Kirkup, 1998, vol. 1.

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

[57] D. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley, "Acoustic modeling using the digital waveguide mesh," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 55 –66, Mar. 2007.

[58] S. R. Bistafa and J. W. Morrissey, "Numerical solutions of the acoustic eigenvalue equation in the rectangular room with arbitrary (uniform) wall impedances," *J. Sound Vib.*, vol. 263, no. 1, pp. 205 – 218, May 2003.

[59] Y. Naka, A. A. Oberai, and B. G. Shinn-Cunningham, "Acoustic eigenvalues of rectangular rooms with arbitrary wall impedances using the interval newton/generalized bisection method," *J. Acoust. Soc. Amer.*, vol. 118, no. 6, pp. 3662–3671, 2005.

[60] W. Rudin, *Functional Analysis*. New York: McGraw-Hill, 1973.

[61] B. W. v. d. Oetelaar, "A new wavefield extrapolation model for the simulation of impulse responses in enclosures," Master's thesis, Delft University of Technology, Delft, The Netherlands, Sept. 1997.

[62] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 5th ed. Oxford, U.K.: Butterworth-Heinemann, 2000, vol. 1.

[63] R. Mehra, N. Raghuvanshi, L. Savioja, M. C. Lin, and D. Manocha, "An efficient GPU-based time domain solver for the acoustic wave equation," *Appl. Acoust.*, vol. 73, no. 2, pp. 83–94, Feb. 2012.

[64] E. Deines, M. Bertram, J. Mohring, J. Jegorovs, F. Michel, H. Hagen, and G. Nielson, "Comparative visualization for wave-based and geometric acoustics," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 1173 –1180, Sept. - Oct. 2006.

[65] A. Ali, C. Rajakumar, and S. M. Yunus, "Advances in acoustic eigenvalue analysis using boundary element method," *Computers & Structures*, vol. 56, no. 5, pp. 837 – 847, 1995.

[66] A. Kelloniemi, V. Valimaki, and L. Savioja, "Simulation of room acoustics using 2-d digital waveguide meshes," in *ICASSP*, vol. 5. Toulouse, France: IEEE Signal Processing Society, May 2006, pp. V313–V316.

[67] J. Huopaniemi, L. Savioja, and M. Karjalainen, "Modeling of reflections and air absorption in acoustical spaces: a digital filter design approach," in *WASPAA*, Oct. 1997, p. 4.

[68] A. Kelloniemi, "Improved adjustable boundary condition for the 3-d digital waveguide mesh," in *WASPAA*, Oct. 2005, pp. 191 – 194.

[69] K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas Propag.*, vol. 14, no. 3, pp. 302 –307, May 1966.

[70] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *J. Comput. Physics*, vol. 114, no. 2, pp. 185 – 200, 1994.

[71] N. Raghuvanshi, R. Narain, and M. Lin, "Efficient and accurate sound propagation using adaptive rectangular decomposition," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 5, pp. 789 –801, Sept.-Oct. 2009.

[72] J. Martinez and R. Heusdens, "On low-complexity simulation of multichannel room impulse responses," *IEEE Signal Process. Lett.*, vol. 17, no. 7, pp. 667 –670, Jul. 2010.

[73] N. Raghuvanshi, J. Snyder, R. Mehra, M. Lin, and N. Govindaraju, "Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes," in *ACM proceedings of SIGGRAPH*, ser. SIGGRAPH '10. NY, USA: ACM, 2010, pp. 68:1–68:11.

[74] Z. Chen and R. C. Maher, "Addressing the discrepancy between measured and modelled impulse responses for small rooms," in *AES 123rd Conv.*, October. 2007.

[75] E. Dereniak and T. Dereniak, *Geometrical and trigonometric optics*. Cambridge University Press, 2008.

[76] P. Calamia, "Advances in edge-diffraction modeling for virtual-acoustic simulations," Ph.D. dissertation, Princeton University, Jun. 2009.

[77] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan., "A beam tracing method for interactive architectural acoustics," *J. Acoust. Soc. Amer.*, vol. 115, no. 2, pp. 739–756, Feb. 2004.

[78] J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[79] P. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, pp. 1527–1529, Nov. 1986.

[80] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications*. Acoustical Society of America, 1994.

[81] B. L. Dalenbäck, "Room acoustic prediction based on a unified treatment of diffuse and specular reflection," *J. Acoust. Soc. Amer.*, vol. 100, no. 2, pp. 899–909, Aug. 1996.

[82] F. Jabloun and B. Champagne, "A fast subband room response simulator," in *ICASSP*. Istanbul, Turkey: IEEE Signal Processing Society, Jun. 2000, pp. II925–II928.

[83] S. G. McGovern, "Fast image method for impulse response calculations in box-shaped rooms," *Appl. Acoust.*, vol. 70, pp. 182–189, Jan. 2009.

[84] J. Borish, "Extension of the image method to arbitrary polyhedra," *J. Acoust. Soc. Amer.*, vol. 75, no. 6, pp. 827–1836, Jun. 1985.

[85] M. Vorländer, "Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm," *J. Acoust. Soc. Amer.*, vol. 86, no. 1, pp. 172–178, 1989.

[86] H. Lee and B.-H. Lee, "An efficient algorithm for the image model technique," *Appl. Acoust.*, vol. 24, no. 2, pp. 87–115, 1988.

[87] S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja, "The room acoustic rendering equation," *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. 1624–1635, Jul. 2007.

[88] H. Lehnert, "Systematic errors of the ray-tracing algorithm," *Appl. Acoust.*, vol. 38, no. 2-4, pp. 207–221, 1993.

[89] M. R. Schroeder, "Digital simulation of sound transmission in reverberant spaces," *J. Acoust. Soc. Amer.*, vol. 47, no. 2A, pp. 424–431, 1970.

[90] A. Farina, "Pyramid tracing vs. ray tracing for the simulation of sound propagation in large rooms." in *Proceedings of the International Conference on Computational Acoustics and its Enviromental Applications*, C. A. Brebbia, Ed., vol. 11. Southampton, U.K.: Wessex Institute, 1995.

[91] J. Amanatides, "Ray tracing with cones," in *ACM proceedings of SIGGRAPH*, ser. SIGGRAPH '84. New York, USA: ACM, 1984, pp. 129–135.

[92] D. van Maercke and J. Martin, "The prediction of echograms and impulse responses within the epidaure software," *Appl. Acoust.*, vol. 38, no. 2-4, pp. 93–114, 1993.

[93] P. S. Heckbert and P. Hanrahan, "Beam tracing polygonal objects," in *ACM proceedings SIGGRAPH*, ser. SIGGRAPH '84. New York, USA: ACM, 1984, pp. 119–127.

[94] T. Lewers, "A combined beam tracing and radiatn exchange computer model of room acoustics," *Appl. Acoust.*, vol. 38, no. 2-4, pp. 161–178, 1993.

[95] Z. Yamauti, "The light flux distribution of a system of interreflecting surfaces," *Journal of the Optic Society of America*, vol. 13, no. 5, pp. 561–561, Nov 1926.

[96] P. Moon and D. E. Spencer, "Light distributions in rooms," *Journal of the Franklin Institute*, vol. 242, no. 2, pp. 111–141, 1946.

[97] H. Buckley, "On the radiation from the inside of a circular cylinder. part iii," *Philosophical Magazine Series 7*, vol. 17, no. 113, pp. 576–581, 1934.

[98] E. Sparrow and R. Cess, *Radiation heat transfer*, ser. Thermal science series. Brooks Pub. Co., 1966.

[99] H. Hottel and A. Sarofim, *Radiative transfer*, ser. McGraw-Hill series in mechanical engineering. McGraw-Hill, 1967.

[100] J. Howell, R. Siegel, and P. Mengüç, *Thermal Radiation Heat Transfer*. CRC Press, 2010.

[101] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, "Modeling the interaction of light between diffuse surfaces," in *ACM proceedings SIGGRAPH*, ser. SIGGRAPH '84. New York, USA: ACM, 1984, pp. 213–222.

[102] D. S. Immel, M. F. Cohen, and D. P. Greenberg, "A radiosity method for non-diffuse environments," in *ACM proceedings of SIGGRAPH*, ser. SIGGRAPH '86. New York, USA: ACM, 1986, pp. 133–142.

[103] M. F. Cohen, J. Wallace, and P. Hanrahan, *Radiosity and realistic image synthesis*. San Diego, USA: Academic Press Professional, Inc., 1993.

[104] I. Ashdown and E. Haines, "Radiosity on-line: a bibliography," *SIGGRAPH Comput. Graph.*, vol. 28, no. 4, pp. 277–279, Nov. 1994.

[105] F. Sillion and C. Puech, *Radiosity and Global Illumination*, ser. The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling Series. Morgan Kaufmann Publishers, 1994.

[106] R. Miles, "Sound field in a rectangular enclosure with diffusely reflecting boundaries," *J. Sound Vib.*, vol. 92, no. 2, pp. 203–226, 1984.

[107] A. Le Bot and A. Bocquillet, "Comparison of an integral equation on energy and the ray-tracing technique in room acoustics," *J. Acoust. Soc. Amer.*, vol. 108, no. 4, pp. 1732–1740, 2000.

[108] H. Kuttruff, "A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries," *J. Acoust. Soc. Amer.*, vol. 98, no. 1, pp. 288–293, 1995.

[109] ——, "Energetic sound propagation in rooms," *Acta Acustica united with Acustica*, vol. 83, no. 4, pp. 622–628, 1997.

[110] E.-M. Nosal, M. Hodgson, and I. Ashdown, "Investigation of the validity of radiosity for sound-field prediction in cubic rooms," *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3505–3514, 2004.

[111] M. Hodgson and E.-M. Nosal, "Experimental evaluation of radiosity for room sound-field prediction," *J. Acoust. Soc. Amer.*, vol. 120, no. 2, pp. 808–819, 2006.

[112] A. Le Bot, "A functional equation for the specular reflection of rays," *J. Acoust. Soc. Amer.*, vol. 112, no. 4, pp. 1276–1287, 2002.

[113] G. Naylor, "Odeon-another hybrid room acoustical model," *Appl. Acoust.*, vol. 38, no. 2-4, pp. 131–143, 1993.

[114] Y. W. Lam, "A comparison of three diffuse reflection modeling methods used in room acoustics computer models," *J. Acoust. Soc. Amer.*, vol. 100, no. 4, pp. 2181–2192, 1996.

[115] C. L. Christensen, "A new scattering method that combines roughness and diffraction effects," in *Proc. Forum Acousticum*. Budapest, Hungary: ASA/EEA, 2005.

[116] ODEON A/S. (2012, May) ODEON room acoustics program. [Online]. Available: http://www.odeon.dk

[117] CATT. (2012, May) CATT-Acoustic. [Online]. Available: http://www.catt.se

[118] AFMG Technologies GmbH. (2012, May) Enhanced acoustic simulator for engineers (EASE). [Online]. Available: http://ease.afmg.eu

[119] LAE group. (2012, May) RAMSETE: Room acoustics modeling on pc. [Online]. Available: http://www.ramsete.com

[120] BOSE Professional systems division. (2012, May) BOSE modeler software. [Online]. Available: http://worldwide.bose.com/pro/en/web/modeler_software/page.html

[121] A. Wabnitz, N. Epain, C. T. Jin, and A. Van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *International Symposium on Room Acoustics*, 2010, pp. CDROM: 1–6.

[122] L. Savioja, T. Lokki, and J. Huopaniemi, "Auralization applying the parametric room acoustic modeling technique - the DIVA auralization system," in *Proc. 8th Int. Conf. Auditory Display*, ser. ICAD 2002, Kyoto, Japan, 2002, pp. 219–224.

[123] E. Lehmann and A. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1429 –1439, Aug. 2009.

[124] R. Bracewell, *The Fourier Transform & Its Applications*, 3rd ed. New York: McGraw-Hill, 1999.

[125] I. Dokmanic, Y. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-d polygonal case," in *ICASSP*. Prague, Czech Republic: IEEE Signal Processing Society, 2011, pp. 321–324.

Part II: Included Papers

3

On low-complexity simulation of multichannel room impulse responses

Jorge Martínez and Richard Heusdens

Abstract

In this letter we present a method for low-complexity simulation of multi-channel room impulse responses (RIRs). Low-complexity RIR methods will become inevitable in next generation communication systems having massive amounts of microphones/loudspeakers. For a room with rigid boundaries, we show that proper sampling of the free-field plenacoustic spectrum results in the solution of the wave equation at any position in the room. We show that the spatial aliasing introduced by spectral sampling represents the wall reflections. These wall reflections are usually modelled, at least in low-complexity simulation algorithms, by the creation of virtual free-field sources outside the room, an image source model commonly referred to as the image method. The image method requires $\mathcal{O}(N^3)$ operations per receiver position, whereas the newly proposed method requires only $\mathcal{O}(N \log N)$ operations per receiver position.

3.1 Introduction

In more than 30 years of research on digital modeling of room impulse responses (RIRs) [1], methods have appeared that can achieve very realistic results. However, the complexity of these methods increases rapidly with the desired reverberation time, which makes them unusable in the case where a very large amount of channels are to be simulated.

It is expected that the future of human communication technology will be based on systems having massive multi-input, multi-output (MMIMO) channel configurations. Such systems will offer never before achieved experiences in a variety of new applications, like ambient telephony [2] or virtual teleconferencing. Going multi-channel, however, has shown to be fundamentally different than its single-channel counterpart in several acoustic signal processing problems. As an example, in the multi-channel acoustic echo cancellation (AEC) problem, the optimal filter coefficients are, in general, not unique since instead of estimating the individual RIRs, we only estimate the sum of them. As a consequence, drastic errors can occur with slight changes of the acoustical environment (the problem is ill conditioned) [3]. This can be overcome if we can accurately compute the individual RIRs as well and use these estimates to construct the sum signal. Clearly, specially in the case of MMIMO AEC, this is only feasible if the computation of the individual RIRs is of sufficiently low complexity. In this letter we present a very low-complexity method for modeling multi-channel RIRs. The method is based on proper sampling of the free-field plenacoustic spectrum. The theory and experimental results presented here are for the case of rigid walls only. Efficient inclusion of wall reflection coefficients is the topic of current research.

3.2 The evolution of the sound field in a room

The plenacoustic function (PAF), first introduced in [4], completely characterizes the sound field in space. Given an omni-directional point source (monopole) S at a certain position $\mathbf{s} \in \mathbb{R}^3$, the PAF is given by $p(\mathbf{x}, t) = (h_{\mathbf{s}} * s)(\mathbf{x}, t)$, where s is the signal emitted by S, $h_{\mathbf{s}}$ is the room impulse response (RIR) at location $\mathbf{x} \in \mathbb{R}^3$ with respect to the source location $\mathbf{s}$ and $*$ denotes the (temporal) convolution operator. Multiple sources are considered as the superposition of single sources [4]. The plenacoustic spectrum (PAS) is given by the 4-D Fourier transform of the PAF, that is, $P(\boldsymbol{\phi}, \omega) = \mathcal{F}_{\mathbf{x},t}\left\{p(\mathbf{x}, t)\right\}$, where $\mathcal{F}_{\mathbf{x},t}\left\{\cdot\right\}$ is the Fourier transform operator over the specified dimensions $\mathbf{x}$ and t, $\boldsymbol{\phi} \in \mathbb{R}^3$ is the spatial-frequency vector and ω the temporal frequency. Consider a monopole S in free-field, emitting a Dirac pulse at position $\mathbf{s}$ and time $t = 0$, and let c denote the speed of sound. The PAF (RIR in this case) at position $\mathbf{x}$, corresponds to the free space Green function given by [5]

$$p(\mathbf{x}, t) = \frac{\delta\left(t - \frac{r}{c}\right)}{4\pi r}, \qquad r = \|\mathbf{x} - \mathbf{s}\|,$$

and its corresponding spectrum by [4]

$$P(\boldsymbol{\phi}, \omega) = \frac{e^{j\boldsymbol{\phi}^T\mathbf{s}}}{\|\boldsymbol{\phi}\|^2 - \left(\frac{\omega}{c}\right)^2}, \tag{3.1}$$

where the superscript ^{T} denotes matrix transposition. In order to study the sound field in a box-shaped room we use the classical image source model proposed in [6], which is generally referred to as the image method. The image method reconstructs the sound field within a box-shaped room by creation of virtual free-field sources (outside the room) which represent the reflections introduced by the walls. Fig. 3.1 shows an example of the constellation of virtual sources in the image method for a 2-D scenario. The solid rectangular box denotes the room with dimensions L_x and L_y. The source S_0 denotes the direct-path contribution of S, whereas the sources S_1, S_2 and S_3 denote the contributions due to reflections at the left, bottom and combination of left and bottom wall, respectively. All other reflection contributions are obtained by a periodic repetition of these four sources throughout the whole space.

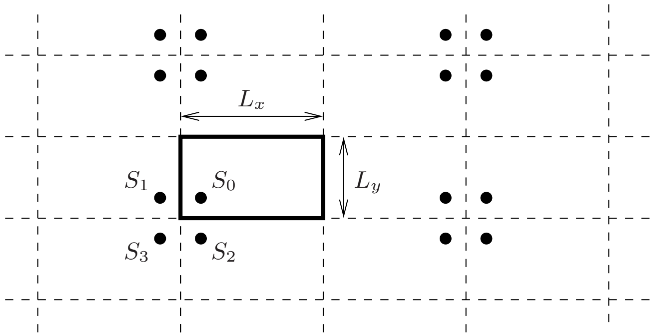


Figure 3.1: Example of the constellation of virtual sources in the image method [6] for a 2-D scenario.

The set of (in this case four) virtual sources that generate the periodic constellation will be referred to as the set of mother sources [4]. More general, in the ν-D scenario, there are 2^ν mother sources.

In order to derive an efficient algorithm for computing the complete sound field in a real (3-D) scenario, we will formalize the above discussion. Given the set of mother sources, the complete set of virtual sources is obtained by a 3-D periodic packing [7] over a lattice, say Λ, given by

$$\Lambda = \{\lambda \in \mathbb{R}^3 : \lambda = \mathbf{\Lambda n},\ \mathbf{n} \in \mathbb{Z}^3\}.$$

The matrix $\mathbf{\Lambda}$ is called the generator matrix of the lattice of which the columns are the basis vectors of the lattice. In the case of a box-shaped room with dimensions L_x, L_y and L_z, $\mathbf{\Lambda}$ is given by

$$\mathbf{\Lambda} = \mathrm{diag}(2L_x, 2L_y, 2L_z).$$

Using this notation, the total sound field at any position $\mathbf{x}$ is given by

$$p(\mathbf{x}, t) = \sum_{\mathbf{n} \in \mathbb{Z}^3} p_{\mathrm{m}}(\mathbf{x} + \mathbf{\Lambda n}, t), \tag{3.2}$$

with

$$p_{\mathrm{m}}(\mathbf{x}, t) = \sum_{l=0}^{7} \frac{\delta\left(t - \frac{r_l}{c}\right)}{4\pi r_l}, \quad r_l = \|\mathbf{x} - \mathbf{s}_l\|,$$

the contribution of the set of mother sources S_l, where $l = 0, \ldots, 7$. Clearly, direct computation of the sound field involves the contribution of infinite many virtual sources, which is practically unfeasible. Therefore, the number of periodic repetitions taken into account in practical situations is made finite, which will result in a computational complexity of $\mathcal{O}(N^3)$, where N is the number of repetitions per dimension which is proportional to the reverberation time. Over the past years, several improvements of the image method have been proposed to reduce the computational complexity [8], [9], [10]. However, none of these methods has led to a total complexity less than $\mathcal{O}(N^3)$ per channel. Obviously, in real-time scenarios or large-scale sensor network applications this complexity is far too high and alternative algorithms for computing the sound field are necessary.

3.3 Fast computation of multichannel RIRs

We propose a new method to compute the complete sound field in a room by spatial sampling of the PAS generated by the mother sources. For a 3-D scenario, this PAS is given by

$$P_{\mathrm{m}}(\boldsymbol{\phi}, \omega) = \sum_{l=0}^{7} \frac{e^{j\boldsymbol{\phi}^T \mathbf{s}_l}}{\|\boldsymbol{\phi}\|^2 - \left(\frac{\omega}{c}\right)^2}. \tag{3.3}$$

We have the following result.

Proposition 3.1. *Let* $\mathbf{\Lambda}$ *be the generating matrix of the lattice specifying the periodic packing of the mother sources in space, and let* $\mathbf{\Phi}$ *denote the generating matrix of the lattice* Φ *specifying the spectral sampling points. Then* $P_{\mathrm{m}}(\mathbf{\Phi}\mathbf{k}, \omega)$, $\mathbf{k} \in \mathbb{Z}^3$, *is the (spatial) Fourier series expansion of*

$$q(\mathbf{x}, t) = |\mathbf{\Lambda}| \sum_{\mathbf{n} \in \mathbb{Z}^3} p_{\mathrm{m}}(\mathbf{x} + \mathbf{\Lambda}\mathbf{n}, t),$$

if and only if $\mathbf{\Phi} = 2\pi\mathbf{\Lambda}^{-T}$.

The proof is given in appendix 3.A. Proposition 3.1 gives us a recipe for constructing the complete sound field in a room, which is given by (3.2). We sample the PAS of the set of mother sources given by (3.3) using the sampling lattice generated by $\mathbf{\Phi} = 2\pi\mathbf{\Lambda}^{-T}$, scale the result by $|\mathbf{\Lambda}|^{-1}$ and inverse Fourier transform the coefficients thus obtained. For a box-shaped room with dimensions L_x, L_y and L_z, the generator matrix for the sampling lattice is simply given by

$$\mathbf{\Phi} = \text{diag}\left(\frac{\pi}{L_x}, \frac{\pi}{L_y}, \frac{\pi}{L_z}\right).$$

In order to implement the proposed algorithm on a computer, we need to sample ω as well. However, since $p(\mathbf{x}, t)$ is not time limited (it has infinite support), the process of sampling ω will introduce (undesired) time-domain aliasing. Let $\mathbf{\Psi}$ denote the generating matrix of the lattice specifying the spectral sampling points of both the spatial and temporal frequencies, defined by

$$\mathbf{\Psi} = \text{diag}(\mathbf{\Phi}, \Omega_s),$$

where Ω_s denotes the temporal-frequency sampling interval and where we assumed separate sampling of spatial and temporal frequencies. Using the same arguments as used in the proof of Proposition 3.1, it follows that sampling the PAS generated by the mother sources using a sampling lattice Ψ generated by $\mathbf{\Psi}$ yields the Fourier series expansion of

$$q(\mathbf{x}, t) = |\mathbf{\Delta}| \sum_{\mathbf{n}\in\mathbb{Z}^3} \sum_{n\in\mathbb{Z}} p_{\text{m}}(\mathbf{x} + \mathbf{\Lambda n}, t + T_p n), \tag{3.4}$$

if and only if

$$\mathbf{\Psi} = 2\pi \mathbf{\Delta}^{-T} = 2\pi \text{diag}(\mathbf{\Lambda}, T_p)^{-T}.$$

By inspection of (3.2) and (3.4), we conclude that

$$|\mathbf{\Delta}|^{-1} q(\mathbf{x}, t) = p(\mathbf{x}, t) + \sum_{\mathbf{n}\in\mathbb{Z}^3} \sum_{\substack{n\in\mathbb{Z} \\ n\neq 0}} p_{\text{m}}(\mathbf{x} + \mathbf{\Lambda n}, t + T_p n),$$

where the last term of the right-hand side represents the time-domain aliasing contribution introduced by sampling the temporal frequency ω. Although $p(\mathbf{x}, t)$ is not time limited, for practical applications we will have that $\lim_{t\to\infty} p(\mathbf{x}, t) = 0$, so that by making T_p sufficiently large (i.e. making Ω_s sufficiently small), we can make the error due to time-domain aliasing arbitrarily small. Clearly, ignoring time-domain aliasing effects, (3.4) can be used to compute the PAF in any position in the room. However, the reconstruction of $p(\mathbf{x}, t)$ out of its Fourier coefficients involves an infinite summation. In addition, if we want to know the PAF for another location in space, we have to recompute the complete inverse Fourier transform. Thus, in applications involving many microphones or situations where the microphone is being moved through the room, this computational complexity is too high.

In the remainder of this section we will show that, by approximating the PAF using the fast Fourier transform (FFT), we can significantly reduce the computational complexity. To do so, we limit the summation of the inverse Fourier transform to a finite number of elements and periodically extend this finite set over the (4-D) spatio-temporal frequency space. Making the PAS periodic, however, implies a discretization of the PAF. That is, using this approach we can only compute the PAF in a finite number of positions. As we will see below, we can make these sample points in space arbitrarily dense at the expense of taking more Fourier coefficients into account when reconstructing the PAF.

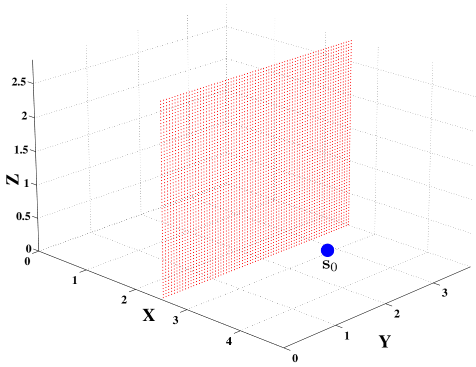
Let Σ be a sublattice of Ψ ($\Sigma \subseteq \Psi$ is a subset of Ψ which itself is a lattice) denoting the lattice for generating the 4-D periodic packing of the frequency space. Moreover, let Γ denote the spatio-temporal sampling lattice imposed by making the PAS periodic and assume $\Delta \subseteq \Gamma$. Clearly, we have $\mathbf{\Gamma} = 2\pi\mathbf{\Sigma}^{-T}$. With this, the PAF can be approximated by

$$p(\mathbf{\Gamma n}) \approx \frac{1}{\mathrm{N}(\Delta/\Gamma)} \sum_{\mathbf{k}\in V_{\Sigma}(\mathbf{0})} |\mathbf{\Gamma}|^{-1} P_{\mathrm{m}}(\mathbf{\Psi k}) e^{j(\mathbf{k}^T\mathbf{\Psi}^T\mathbf{\Gamma n})}, \tag{3.5}$$

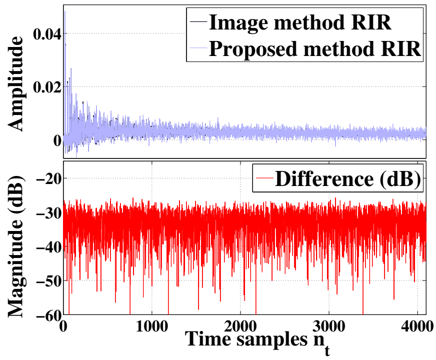
where $\mathrm{N}(\Delta/\Gamma)$ denotes the number of lattice points of Γ that lie inside $V_{\Delta}(\mathbf{0})$, the Voronoi region of Δ around the origin. Obviously, the more frequency samples we take into account (by making $V_{\Sigma}(\mathbf{0})$ larger), the finer we sample the PAF because of the relation $\mathbf{\Gamma} = 2\pi\mathbf{\Sigma}^{-T}$. Indeed, we have

$$\mathrm{N}(\Delta/\Gamma) = \frac{|\mathbf{\Delta}|}{|\mathbf{\Gamma}|} = \frac{(2\pi)^4|\mathbf{\Psi}^{-T}|}{(2\pi)^4|\mathbf{\Sigma}^{-T}|} = \frac{|\mathbf{\Sigma}|}{|\mathbf{\Psi}|} = \mathrm{N}(\Sigma/\Psi),$$

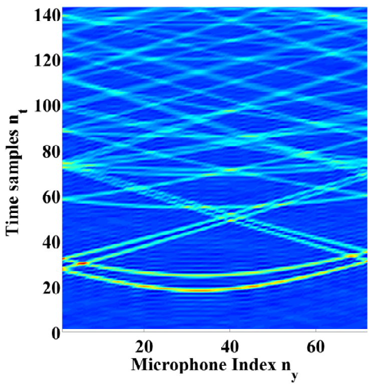
and we conclude that the number of spectral evaluation points $\mathrm{N}(\Sigma/\Psi)$, equals the number of spatio-temporal samples $\mathrm{N}(\Delta/\Gamma)$ of the PAF. Since in this case the PAF is periodic as well, we only need to evaluate it in $V_{\Delta}(\mathbf{0})$. The computational complexity of this approach is given by the evaluation of (3.5) limited to $\{\mathbf{n} \in \mathbb{Z}^4 : \mathbf{\Gamma n} \in V_{\Delta}(\mathbf{0})\}$. Moreover, we have that most of the energy of the PAS (3.3) is localized in the region $\|\boldsymbol{\phi}\| \leq |\omega/c|$ [4]. For a maximum temporal frequency ω_b, this fact is used to determine $\mathbf{\Sigma}$ and consequently $\mathrm{N}(\Sigma/\Psi)$, allowing us for a trade between speed and accuracy of reconstruction. The evaluation of (3.5) will take $\mathcal{O}(N^4 \log N)$ operations, with N proportional to ω_b for computing the PAF in $\mathrm{N}(\Delta/\Gamma)$ spatio-temporal positions. Since $\mathrm{N}(\Sigma/\Psi) = \mathrm{N}(\Delta/\Gamma)$ the method is of complexity $\mathcal{O}(N \log N)$ per receiver position (although all positions are calculated at once). The image method, on the other hand, is of complexity $\mathcal{O}(N^3)$ per receiver position, with N proportional to the reverberation time. Therefore when large amounts of receiver positions are to be simulated, the newly proposed method significantly outperforms the image method.



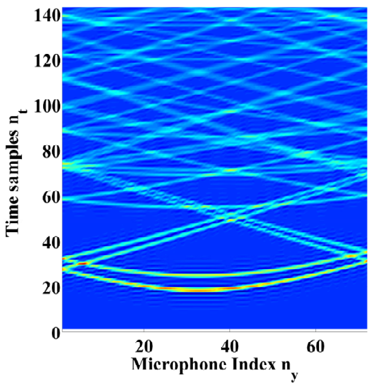
(a) *Simulated Room.*



(b) *1 sec. RIR length.*



(c) *Proposed: 18 seconds.*



(d) *Image method: 15 hours.*

Figure 3.2: (a) 4608 receivers (red dots) positioned in the (y, z) plane at $x = 2.51$. The sound source is s_0. (b) Comparison of both methods only for receiver $n_y = 44$ of one line in the y direction at $n_z = 16$. (c) Experimental results for all receivers at the same line, RIRs simulated with the proposed method. (d) RIRs simulated with the image method [6].

3.4 Numerical Experiments

In this section we compare the RIRs generated by the standard image method (3.2) to the ones generated by the newly proposed method (3.5). The presence of totally reflective walls requires the PAF to satisfy a boundary condition which results in a non-decaying sound field [1], [5]. Therefore, in order to present an illustrative example of the theory, we will set fully reflective walls in two dimensions and no walls (or fully absorptive walls) in one dimension. To overcome the non-bandlimited representation of the delta pulses in the image method, we follow Peterson's approach [11]. We replace each pulse by the impulse response of a Hanning-windowed ideal low-pass filter of length 8 ms. For the experiment we will consider temporal signals bandlimited to 2 kHz. The room dimensions are $(L_x, L_y, L_z) = (4.78, 3.79, 2.83)$. The temporal sampling-frequency is set to $\omega_s = (4000)2\pi$. We will consider a simulation time of $T_h = 1.024$ s, which accounts for $N_t = \lfloor (T_h)(\omega_s/2\pi) \rfloor = 4096$ temporal samples. Further, by choosing the temporal-frequency sampling interval to be $\Omega_s = \omega_s/4N_t$, the spectral sampling matrix is given by $\mathbf{\Psi} = \mathrm{diag}\{\frac{\pi}{L_y}, \frac{\pi}{Lz}, \Omega_s\}$, since no reflections are considered in the x direction. The spectral periodicity matrix is chosen to be $\mathbf{\Sigma} = (\mathbf{\Psi})(\mathrm{diag}\{144, 128, 4N_t\})$, which for the purpose of our experiment gives a good compromise between spectral aliasing and speed of reconstruction. This directly defines the sampling lattice Γ with matrix $\mathbf{\Gamma} = \mathrm{diag}\{0.052, 0.044, 0.00025\}$. Since $\mathbf{\Gamma}$ is rectangular, the receiver positions (spatial samples of the PAF) are arranged in a plane orthogonal to the x direction. Its position is fixed at $x = 2.51$. This gives us a total of 4608 receiver positions in the room, with $(72, 64)$, receivers in the y and z directions. The source S is at position $\mathbf{s}_0 = (3.98, 1.70, 0.63)$. Fig. 3.2a shows a view of the scenario.

First we evaluate the sampled free-field PAS of the mother sources $P_{\mathrm{m}}(\mathbf{\Psi k})$, using the parameters already given. Small damping constants are introduced in the spatial frequencies close to the singularities of the PAS. This is justified since the finiteness of the experiment forces the function to be bounded [1]. After FFT synthesis, we obtain our approximation of the PAF at all microphone positions. For the image method algorithm, we evaluate individually the RIRs from the source to each microphone. We use a triplet of integers n_y, n_z, n_t to index the PAF samples. The colormap plots in 3.2c and 3.2d display a top view of the results only for one line of microphones in the y direction at $n_z = 16$ and time samples $0 \leq n_t \leq 143$, for the proposed method and the image method respectively. Additionally in 3.2b, a comparison plot for microphone $n_y = 44$ of the same line is given over the full time span of the RIRs. It is important to note that the spatio-temporal "locations" of the reflections are perfectly modeled by the proposed method. However, discrepancies between these approaches can be observed. These are caused by temporal aliasing that can be made arbitrarily small by decreasing the spectral sampling interval Ω_s. The experiment was implemented in a x86 PC system at 2.3 Ghz. with MATLAB®, using C++ mex functions for both methods. With this configuration, the image method took 15 hours to complete, whereas the proposed method only took 18 seconds. Clearly showing the huge saving in computational complexity.

3.5 Conclusions

For a room with rigid walls, and a geometrical shape that allows the sound field to be mathematically modeled in terms of a spatially periodic structure, a very powerful representation is obtained via its Fourier series expansion. It was shown that this expansion is obtained by spatial sampling of the plenacoustic spectrum of a set of mother sources. By making the temporal-frequency space discrete, a very low complexity algorithm for multi-channel RIRs computation was derived. It was also shown how speed of reconstruction can be further decreased at the expense of inaccuracies due to temporal aliasing. The theory and experiments were presented for the case of rigid walls, efficient inclusion of wall reflection coefficients is the topic of current research.

Appendix 3.A Proof of Proposition 3.1

Proof. Sampling the PAS of the mother sources at spectral sampling points $\mathbf{\Phi k}$, $\mathbf{k} \in \mathbb{Z}^3$, yields

$$\begin{aligned} P_{\mathrm{m}}(\mathbf{\Phi k}, \omega) &= \iint p_{\mathrm{m}}(\mathbf{x}, t) e^{-j(\mathbf{k}^T \mathbf{\Phi}^T \mathbf{x} + \omega t)} d\mathbf{x} dt \\ &= \int \sum_{\mathbf{n} \in \mathbb{Z}^3} \int_{V_\Lambda(\mathbf{n})} p_{\mathrm{m}}(\mathbf{x}, t) e^{-j(\mathbf{k}^T \mathbf{\Phi}^T \mathbf{x} + \omega t)} d\mathbf{x} dt, \end{aligned} \tag{3.6}$$

where $V_\Lambda(\mathbf{n})$ is called a Voronoi region of the lattice Λ and is defined by

$$V_\Lambda(\mathbf{n}) \triangleq \left\{ \mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x} - \mathbf{\Lambda n}\| \le \|\mathbf{x} - \mathbf{\Lambda n}'\|, \ \forall \mathbf{n}' \in \mathbb{Z}^3 \right\}.$$

Making the substitution $\mathbf{x} \to \mathbf{x} + \mathbf{\Lambda n}$, (3.6) becomes

$$\begin{aligned} & P_{\mathrm{m}}(\mathbf{\Phi k}, \omega) \\ &= \iint_{V_\Lambda(\mathbf{0})} \sum_{\mathbf{n} \in \mathbb{Z}^3} p_{\mathrm{m}}(\mathbf{x} + \mathbf{\Lambda n}, t) e^{-j(\mathbf{k}^T \mathbf{\Phi}^T (\mathbf{x} + \mathbf{\Lambda n}) + \omega t)} d\mathbf{x} dt \\ &= \iint_{V_\Lambda(\mathbf{0})} \sum_{\mathbf{n} \in \mathbb{Z}^3} p_{\mathrm{m}}(\mathbf{x} + \mathbf{\Lambda n}, t) \beta e^{-j(\mathbf{k}^T \mathbf{\Phi}^T \mathbf{x} + wt)} d\mathbf{x} dt, \end{aligned}$$

where $\beta = e^{-j\mathbf{k}^T\mathbf{\Phi}^T\mathbf{\Lambda}\mathbf{n}}$, and $V_\Lambda(\mathbf{0})$ is the Voronoi region around the origin. This is the (spatial) Fourier series expansion of

$$|\mathbf{\Lambda}| \sum_{\mathbf{n}\in\mathbb{Z}^3} p_{\mathrm{m}}(\mathbf{x} + \mathbf{\Lambda}\mathbf{n}, t)\beta.$$

The term $\beta = e^{-j\mathbf{k}^T\mathbf{\Phi}^T\mathbf{\Lambda}\mathbf{n}} = 1$ for all $\mathbf{k}, \mathbf{n} \in \mathbb{Z}^3$ if and only if $\mathbf{\Phi}^T\mathbf{\Lambda} = 2\pi\mathbf{I}$, $\mathbf{I}$ the identity matrix, or equivalently $\mathbf{\Phi} = 2\pi\mathbf{\Lambda}^{-T}$, which completes the proof. □

Acknowledgment

The authors would like to thank Dr. ir. Richard C. Hendriks of Delft University of Technology for his valuable remarks and comments.

3.6 Bibliography

[1] H. Kuttruff, *Room Acoustics*. London: Taylor and Francis, 2000.

[2] A. Härmä, “Ambient telephony: scenarios and research challenges,” in *Proc. Interspeech*, Antwerp, Belgium, August 2007, pp. 562–565.

[3] Y. Huang, J. Benesty, and J. Cheng, *Acoustic MIMO Signal Processing*. Berlin Heidelberg New York: Springer-Verlag, 2006.

[4] T. Ajdler, L. Sbaiz, and M. Vetterli, “The plenacoustic function and its sampling.” *IEEE Trans Signal Process.*, vol. 54, no. 10, pp. 3790–3804, October 2006.

[5] P. Morse and K. Ingard, *Theoretical Acoustics*. New York: McGraw-Hill, 1968.

[6] J. Allen and D. Berkley, “Image method for efficiently simulating small room acoustics,” *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, April 1979.

[7] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*. New York, USA: Springer, 1993.

[8] S. G. McGovern, “Fast image method for impulse response calculations in box-shaped rooms.” vol. 70, pp. 182–189, Jan. 2009.

[9] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, “Fast evaluation of the room transfer function using multipole expansion.” *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 2, pp. 565–576, February 2007.

[10] F. Jabloun and B. Champagne, “A fast subband room response simulator,” in *Proceedings ICASSP’00.* Washington, DC, USA: IEEE Computer Society, June 2000, pp. II925–II928.

[11] P. Peterson, “Simulating the response of multiple microphones to a single acoustic source in a reverberant room,” *J. Acoust. Soc. Am.*, vol. 80, pp. 1527–1529, 1986.

4

A generalized Poisson summation formula and its application to fast linear convolution

Jorge Martínez, Richard Heusdens and Richard C. Hendriks

Abstract

In this letter, a generalized Fourier transform is introduced and its corresponding generalized Poisson summation formula is derived. For discrete, Fourier based, signal processing, this formula shows that a special form of control on the periodic repetitions that occur due to sampling in the reciprocal domain is possible. The present paper is focused on the derivation and analysis of a *weighted* circular convolution theorem. We use this specific result to compute linear convolutions in the generalized Fourier domain, without the need of zero-padding. This results in faster, more resource-efficient computations. Other techniques that achieve this have been introduced in the past using different approaches. The newly proposed theory however, constitutes a unifying framework to the methods previously published.

4.1 Introduction

The classical Poisson summation formula, expresses the fact that discretization in one domain implies periodicity in the reciprocal domain [1]. This periodicity comes in the form of a periodic summation of the signal values. In Fourier based digital signal processing (DSP), we have to deal with the periodic repetitions that appear due to the discrete nature of the domains where our signals are defined. In virtually all applications, the effect of overlapped repetitions is an issue that must be avoided, or at least controlled. In many cases the only way to achieve this is by increasing the sampling rate in the reciprocal domain [1], a costly operation in terms of memory and computational resources. In this letter we introduce a generalized Fourier domain (GFD) and derive its generalized Poisson summation formula. The newly proposed equation, relates the samples of the continuous generalized spectrum of a signal, with a geometrically *weighted* periodic extension of the signal. As it will be explained, this formula shows that a parametric form of control on the periodic repetitions that occur due to sampling in the GFD is possible, without the need to increase the sampling rate. This result has in principle many potential applications, the work presented here however, will be focused on the derivation and analysis of the *weighted* circular convolution theorem for the generalized discrete Fourier transform (GDFT).

For finite discrete-time signals of length $N \in \mathbb{Z}$, point-wise multiplication of the discrete spectra of the signals corresponds to circular convolution [1]. In order to perform a linear convolution (i.e. LTI filtering), the input signals are first zero-padded to at least length $2N-1$ and then transformed, increasing the number of computations needed [1]. In this letter we show how the newly proposed theorem can be used to perform linear convolutions in the GFD without the need of zero-padding, implementing the GDFT by means of the FFT. As we show, this results in a more efficient computation in terms of memory locations and operations needed. Although techniques to obtain a linear convolution without zero-padding have been developed in different contexts e.g [2] and [3], we show that the proposed theory constitutes a unifying framework to these approaches.

4.2 Poisson summation formula and circular convolution

Consider the Fourier transform for a signal $f(t) \in L^2(\mathbb{R})$ given by,

$$F(\Omega) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-j\Omega t}\,\mathrm{d}t, \tag{4.1}$$

where $\Omega \in \mathbb{R}$ is the angular frequency variable, and $t \in \mathbb{R}$ represents time. If (4.1) exists, then we call F the spectrum of the signal f. The inverse transformation is given by [1]

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\Omega)\mathrm{e}^{j\Omega t}\,\mathrm{d}\Omega. \tag{4.2}$$

The Poisson summation formula then links the signal f to the samples of its spectrum F [1], i.e,

$$\sum_{p\in\mathbb{Z}} f(t+pT_p) = \frac{1}{T_p}\sum_{k\in\mathbb{Z}} F\left(\frac{2\pi k}{T_p}\right)\mathrm{e}^{j\frac{2\pi kt}{T_p}}. \tag{4.3}$$

From this equation we see that the repetitions of the periodic summation at the right-hand side will overlap if the length of the support of signal f is larger than T_p. Thus, given a fixed signal with finite-length (compact) support, the only way to avoid overlapping using (4.3) is to increase the value of T_p. This means a smaller spectral sampling interval $2\pi k/T_p$, which implies increased spectral sampling rate, and consequently, more memory and computational resources. This formula also allows us to easily understand the effect that frequency discretization has on the convolution product of two signals, say $f(t)$ and $h(t)$, when performed in the frequency-domain. That is,

$$\sum_{p\in\mathbb{Z}} (f*h)(t+pT_p) = \frac{1}{T_p}\sum_{k\in\mathbb{Z}} (FH)\left(\frac{2\pi k}{T_p}\right)\mathrm{e}^{j\frac{2\pi kt}{T_p}}, \tag{4.4}$$

where $*$ denotes the linear convolution operator, and where we have made use of the convolution theorem [1]. The periodic repetitions that appear in the time domain will overlap when the support of the convolution between f and h is larger than T_p, and consequently, it will be impossible to obtain the linear convolution result using (4.4). This is exactly how circular convolution is defined, as a periodic or cyclic version of linear convolution [1]. We will show that, by introducing a generalized Poisson summation formula, a special form of control on these repetitions can be obtained.

4.3 A Generalized Poisson summation formula

We now define a generalized Fourier transform for $f \in L^2(\mathbb{R})$, and $\alpha \in \mathbb{C}\backslash\{0\}$ as follows,

$$F_\alpha(\Omega) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{\beta t}\mathrm{e}^{-j\Omega t}\,\mathrm{d}t, \tag{4.5}$$

where $\beta = \log(\alpha)/T_p$. This is equivalent to the ordinary Fourier transform (if it can be defined) of the modulated signal $f(t)\mathrm{e}^{\beta t}$. Note that (4.5) can be seen as a particular case of the Laplace transform, therefore for $|\Re\{\beta\}| > 0$, with $\Re\{\beta\}$ the real part of β, we have that (4.5) would not be defined on all $L^2(\mathbb{R})$, but only on the dense subset of all causal and anticausal (single-sided) functions [4]. Hence, for finite-length signals, operation (4.5) can always be defined, since these signals are special cases of single-sided functions. The inverse transformation follows as,

$$f(t) = \frac{\mathrm{e}^{-\beta t}}{2\pi}\int_{-\infty}^{\infty} F_\alpha(\Omega)\mathrm{e}^{j\Omega t}\,\mathrm{d}\Omega. \tag{4.6}$$

Evaluating F_α in (4.3), we obtain a generalization of the Poisson summation formula,

$$\sum_{p\in\mathbb{Z}} \mathrm{e}^{\beta(t+pT_p)} f(t+pT_p) = \frac{1}{T_p}\sum_{k\in\mathbb{Z}} F_\alpha\left(\frac{2\pi k}{T_p}\right)\mathrm{e}^{j\frac{2\pi k t}{T_p}}$$

so that,

$$\sum_{p\in\mathbb{Z}} \mathrm{e}^{\beta p T_p} f(t+pT_p) = \frac{\mathrm{e}^{-\beta t}}{T_p}\sum_{k\in\mathbb{Z}} F_\alpha\left(\frac{2\pi k}{T_p}\right)\mathrm{e}^{j\frac{2\pi k t}{T_p}}$$

or,

$$\sum_{p\in\mathbb{Z}} \alpha^p f(t+pT_p) = \frac{\mathrm{e}^{-\beta t}}{T_p}\sum_{k\in\mathbb{Z}} F_\alpha\left(\frac{2\pi k}{T_p}\right)\mathrm{e}^{j\frac{2\pi k t}{T_p}} \tag{4.7}$$

since $\alpha = e^{\beta T_p}$. This equation, relates the samples of the continuous generalized spectrum of a signal, with a geometrically *weighted* periodic extension of the signal. Therefore, an extra form of control can be obtained over the repetitions via the parameter α. In analogy, let us now define the generalized discrete Fourier transform (GDFT) for finite length signals $x(n)$, $n = \{0, \ldots, N-1\}$, $k = \{0, \ldots, N-1\}$, as follows,

$$X_\alpha(k) = \sum_{n=0}^{N-1} x(n) e^{\beta n} e^{-j\frac{2\pi}{N}kn} \tag{4.8}$$

where, $\beta = \log(\alpha)/N$. The inverse GDFT is given by,

$$x(n) = \frac{e^{-\beta n}}{N} \sum_{k=0}^{N-1} X_\alpha(k) e^{j\frac{2\pi}{N}kn}. \tag{4.9}$$

In this case the generalized Poisson summation formula takes the following form,

$$\sum_{p\in\mathbb{Z}} \alpha^p x(n+pN) = \frac{e^{-\beta n}}{N} \sum_{k=0}^{N-1} X_\alpha(k) e^{j\frac{2\pi kn}{N}}, \tag{4.10}$$

since $\alpha = e^{\beta N}$. Taking $Z_\alpha(k) = (X_\alpha Y_\alpha)(k)$ in (4.10), and applying the convolution theorem we obtain,

$$z(n) = \sum_{p\in Z} \alpha^p (x * y)(n+pN). \tag{4.11}$$

For $\alpha = 1$, this corresponds to the standard Fourier case as already shown for the continuous time case in (4.4). From (4.11) we see how circular convolution is related to linear convolution. Although the original sequences x and y are of length N, its convolution product is however of length $2N-1$. Then it follows that $z(n)$ for $n = \{0, \ldots, N-1\}$, represents the linear convolution of x and y plus the last $N-1$ terms of one overlapped repetition. The classical approach of zero-padding avoids this situation by making N large enough, so that the periodic repetitions in (4.11) are sufficiently appart to avoid overlapping. In the next section we will derive the *weighted* circular convolution theorem for the GDFT pair (4.8) and (4.9). This theorem will be used to perform linear convolutions in the GFD without the need of zero padding, taking advantage of the weighting effect that factor α^p has on the repetitions of the signal as expressed by the generalized Poisson summation formula (4.10).

4.4 Weighted circular convolution theorem

We have the following result.

Property 4.1. *Let* $x(m)$, $m \in \{0, \dots, N-1\}$ *and* $y(l)$, $l \in \{0, \dots, N-1\}$ *be the signals to be convolved. For* $n \in \{0, \dots, N-1\}$, *we have,*

$$\frac{\mathrm{e}^{-\beta n}}{N} \sum_{k=0}^{N-1} X_\alpha(k) Y_\alpha(k) \mathrm{e}^{j\frac{2\pi}{N}nk} = \tag{4.12}$$

$$\sum_{m=0}^{n} x(m)y(n-m) + \alpha \sum_{m=n+1}^{N-1} x(m)y(N+n-m), \tag{4.13}$$

where the left hand summation in (4.13) represents the contribution of N linear convolution terms, and the right hand summation the contribution of N circular convolution terms (which are in fact the last terms of the linear convolution). The factor α effectively weights the amount of circular convolution that is obtained. Thus, given two discrete-time, finite length signals, $x(n)$ and $y(n)$, point-wise multiplication of their generalized discrete spectra, $X_\alpha(k)$ and $Y_\alpha(k)$, corresponds to a *weighted* circular convolution in the time-domain. The proof is given in appendix 4.A.

4.5 Linear convolution using the GDFT

To compute the weighted circular convolution operation given by (4.12), the generalized spectrum of the signals to be convolved is obtained using the GDFT (4.8). Note that this operation can be implemented, by taking the FFT of the modulated signal, $x(n)\mathrm{e}^{\beta n}$ with $\beta = \log(\alpha)/N$. The inverse transform (4.9) can be obtained by multiplying the IFFT of the generalized spectrum with the inverse function $\mathrm{e}^{-\beta n}$. For finite energy signals, the theory of Laplace integrals [4], ensures the existence of the GDFT for any value of $\alpha \in \mathbb{C} \backslash \{0\}$, as previously mentioned in Sec. 4.3. Let us now analyze a special case of the weighted convolution theorem, setting $\alpha = j$, with j the imaginary unit. For real signals x and y, the resulting signal obtained after applying (4.12) becomes complex (its spectrum not being Hermitian symmetric). By a careful inspection of (4.13), it is noted that as a result, the first N values of the linear convolution are stored in the real part of the signal, and the remaining $N-1$ terms are perfectly preserved in its imaginary part. Hence, concatenating the real and imaginary parts, a $2N-1$ point, error-free convolution can be obtained in the GFD without the need of zero-padding. If x and y are complex, then two transforms are necessary to obtain a linear convolution, i.e. for $\alpha = \pm j$. Two show this, let us denote by z_j the result of (4.12) setting $\alpha = j$, and by z_{-j} the result setting $\alpha = -j$. Further denoting by $z_{\Re}$ and $z_{\Im}$ the real and imaginary parts respectively, of the linear convolution between x and y, then by (4.13) we obtain,

$$z_j(n) = z_{\Re}(n) + jz_{\Im}(n) + j(z_{\Re}(N+n) + jz_{\Im}(N+n)),$$

$$z_{-j}(n) = z_{\Re}(n) + jz_{\Im}(n) - j(z_{\Re}(N+n) + jz_{\Im}(N+n)).$$

Clearly, the first N samples of the linear convolution can be obtained by $(z_j(n) + z_{-j}(n))/2$, and the remaining samples by $(z_j(n) - z_{-j}(n))/(2j)$.

If one takes $\alpha \in \mathbb{R}$, with $0 < \alpha << 1$ in (4.12), then not an exact, but an approximation of linear convolution is obtained. In fact, only the first N values of the linear convolution are approximated, since the circular convolution terms in (4.13) are attenuated by a very small factor. The approximation can be made as accurate as possible (making α as small as possible), up to the limits imposed by finite word-length arithmetic. This holds for both complex or real inputs. In this case for real time-domain signals, Hermitian symmetry still holds. Therefore, the point-wise multiplication of the spectra can be performed over the first $N/2 + 1$ DFT coefficients if N is even, and $(N-1)/2 + 1$ coefficients if N is odd.

Two techniques to obtain a linear convolution without the need of zero-padding have been previously introduced in [2] and [3], following different approaches. By setting $\alpha = j$ we directly obtain what the authors in [2] call the right-angle circular convolution (RCC), and for $\alpha = -j$ the left-angle circular convolution (LCC). Further in [3], the authors propose to multiply the input sequences by a scaling factor $s^n \in \mathbb{R}$. This is equivalent to use GDFTs with parameter $\alpha = s^N$ in our proposed derivation. Hence, these techniques can be seen as particular cases of the weighted circular convolution theorem. The newly proposed theory constitutes therefore a generalization that not only provides a unifying framework for previous methods, but also allows for a deeper insight into the problem. This leads to the formulation of new applications, such as the approximation here proposed setting $0 < \alpha << 1$.

4.6 Computational complexity analysis and experiments

In real applications, one would like to use the convolution theorem together with the FFT in order to perform linear convolutions in the frequency-domain with reduced complexity [1]. In some of these applications, the goal is to convolve two finite length signals without worrying about causality or delay (like in the multiplication of two long polynomials). In many real-time DSP problems however, the goal is to filter a long signal, using a finite length (in general much shorter) LTI filter. This is achieved in a block-by-block basis, performing shorter convolutions at each step using well known approaches [1]. In all cases, the operation at hand can be seen as the linear convolution of a length-N signal, say $x(n)$ with a length-M filter, say $h(n)$, where without loss of generality the assumption $M \leq N$ is made. The result of the convolution is therefore of size $N + M - 1$.

An exact complexity analysis is a lengthy and complex task. State-of-the-art FFT algorithms deliver quite inhomogeneous (although asymptotically "equivalent") performance [5]. In each specific case, advantageous conditions can be exploited by the algorithms. The complexity then, becomes a function of the FFT length, the signals class, symmetries present in the input and output signals, hardware architecture, etc [5]. Let us now analyze the total number of multiplications needed as a reasonable basis for comparing the computational complexity. For simplicity, real-valued signals are considered. Thus, given $x(n)$ and $h(n)$, the classical frequency-domain approach requires to pad both signals to at least a size of $N + M - 1$. Moreover, we have that $1 < M \leq N$, hence the complexity can be expressed as a function of N and the ratio, $\lambda = (M-1)/N$, where $1/N \leq \lambda < 1$. Then, we have that three (I)FFTs are needed to transform the signals to the frequency-domain and back to the time-domain. For each of these, the FFT (or IFFT) requires approximately $N(\lambda+1)\log_2(N(\lambda+1))$ real multiplications [5]. The point-wise multiplication of the spectra requires $2N(\lambda+1)+1$ multiplications, since Hermitian symmetry can be exploited [1]. Hence, the total number of multiplications performed is $3N(\lambda+1)\log_2(N(\lambda+1)) + 2N(\lambda+1)+1$.

For the GFD method setting $\alpha = j$, only the filter, $h(n)$ must be zero-padded to a size N. To implement a GFFT or IGFFT as proposed in Sec. 4.5, the (de)modulation of the signals require $2N$ multiplications each. The FFTs require approximately $N\log_2(N)$ multiplications each. The point-wise multiplication of the generalized spectra requires $8N$ multiplications. The total number of multiplications needed is thus $3N\log_2(N) + 14N$. From here, it is clear that the complexity ratio for the GDFT based convolution to the frequency-domain based convolution is approximately, $(1/(1+\lambda))(\log_2(N) + 14/3)/(\log_2(N(1+\lambda)) + 2/3)$. For example setting $N = 256$, in the limiting case $\lambda = 1/N$, which implies $M = 2$, the frequency-domain implementation shows to be about 1.4 times faster than the GDFT approach. This is caused by the very short length of the filter. The overhead produced by the extra multiplications needed in the GDFT method is not compensated. On the other hand, for $M = N$, we have that the complexity ratio can be expressed as $(1/2)(\log_2(N)+14/3)/(\log_2(N)+5/3)$. In this case the GDFT based convolution is

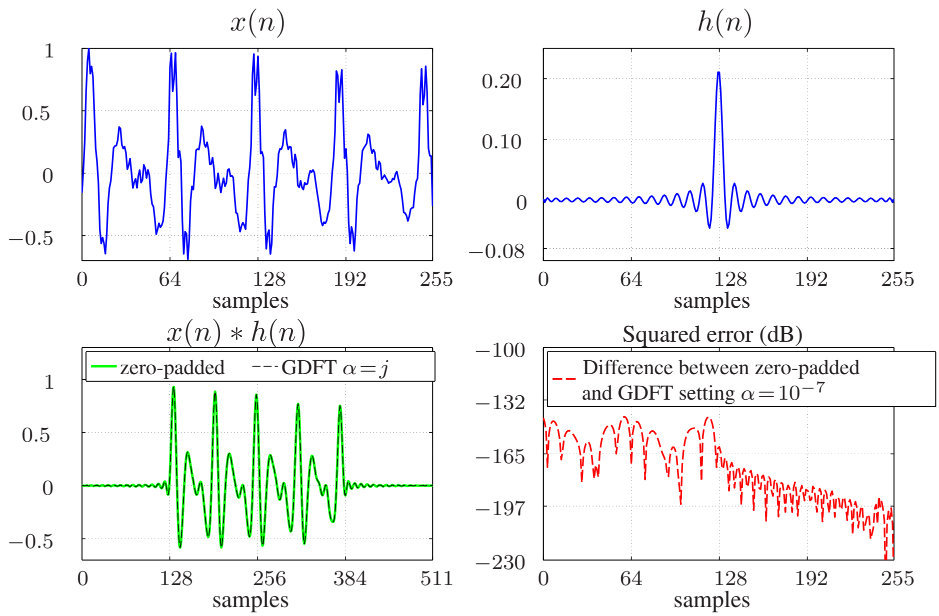


Figure 4.1: LTI filtering of a speech signal frame $x(n)$ and a low-pass filter $h(n)$, performed in the spectral domain for $N = 256$. The standard frequency-domain approach using zero-padding to $2N$ samples and the novel GDFT method are compared.

about 1.5 times faster than the classical frequency-domain approach. For the GDFT method setting $\alpha << 1$, the complexity ratio is approximately $(1/(1+\lambda))(\log_2(N)+2)/(\log_2(N(1+\lambda))+2/3)$, since Hermitian symmetry can be exploited and the (de)modulation of the signals require a total of $3N$ multiplications. In this case, for $N = 256$ the limiting case $\lambda = 1/N$ shows that the GDFT method is still slightly more complex than the frequency-domain algorithm. For $M = N$ on the other hand, the GDFT method is about 2 times faster. From these results we see that for the GDFT approach, the most efficient choice for the filter size is $M = N$. This avoids trivial operations on the zeros added to the filter to be computed, giving the maximum performance increase with respect to the standard frequency-domain method, since the signal and the filter do not need to be zero-padded. In this case, it is also easy to see that the new approach can be implemented using roughly half the amount of memory that it is needed for the standard frequency-domain based convolution. In Fig. 4.1 an example of LTI filtering of a speech signal frame $x(n)$, using a low-pass filter $h(n)$ for $N = 256$ is given. Both the classical frequency-domain approach using zero-padding to $2N$ samples and the novel GDFT approach (4.12) for $\alpha = j$ and $\alpha = 10^{-7}$ are compared. After repeating the experiment 10^6 times, the classical approach took 62.7s to complete. The GDFT approach took 45.4s to complete for the case $\alpha = j$, and 32.2s for the case $\alpha = 10^{-7}$, showing the advantage of the GDFT method in agreement with the analysis conducted above. For the case $\alpha = 10^{-7}$ the result is not exact, and only the first N values of the operation can be compared. A plot in dB of the square error between both approaches shows that a very accurate approximation of LTI filtering is obtained. All tests were performed using double floating-point precision in MATLAB®on a standard desktop PC.

4.7 Conclusions

In this work a generalized Poisson summation formula has been proposed. It conceptually allows us to obtain a special form of control on the periodic repetitions that occur due to sampling in the reciprocal domain. Using this result, a *weighted* circular convolution theorem for the GDFT is derived, which is used to perform efficient, non zero-padded linear convolutions. Altogether, these results have applications which range from simple multiplication of long polynomials, to Wiener filtering, adaptive filtering, near-field beamforming, and many more.

Appendix 4.A Proof of Property 4.1

Proof. Let

$$\begin{aligned} z(n) &= \frac{e^{-\beta n}}{N} \sum_{k=0}^{N-1} X_\alpha(k) Y_\alpha(k) e^{j\frac{2\pi}{N}nk} . \\ &= \frac{e^{-\beta n}}{N} \sum_{k=0}^{N-1} \left(\sum_{m=0}^{N-1} x(m) e^{\beta m} e^{-j\frac{2\pi}{N}mk} \right) \\ &\qquad \times \left(\sum_{l=0}^{N-1} y(l) e^{\beta l} e^{-j\frac{2\pi}{N}lk} \right) e^{j\frac{2\pi}{N}nk} \end{aligned}$$

$$z(n) = \frac{\mathrm{e}^{-\beta n}}{N} \sum_{m=0}^{N-1} x(m)\mathrm{e}^{\beta m} \sum_{l=0}^{N-1} y(l)\mathrm{e}^{\beta l} \times \left(\sum_{k=0}^{N-1} \mathrm{e}^{j\left(\frac{2\pi}{N}\right)k(n-m-l)} \right).$$

For $p \in \mathbb{Z}$ we have [1],

$$\sum_{k=0}^{N-1} \mathrm{e}^{j\left(\frac{2\pi}{N}\right)k(n-m-l)} = \begin{cases} N, & l = n - m + pN \\ 0, & \text{otherwise.} \end{cases} \tag{4.14}$$

Therefore we have that $l = (n - m + pN) \in \{0, \ldots, N-1\}$, and $p = -\lfloor (n-m)/N \rfloor$, where $\lfloor x \rfloor$ is the nearest integer $\leq x$. Further using (4.14) we obtain

$$\begin{aligned} z(n) &= \sum_{m=0}^{N-1} x(m)y(n-m+pN)\mathrm{e}^{\beta m}\mathrm{e}^{\beta(n-m+pN)}\mathrm{e}^{-\beta n} \\ &= \sum_{m=0}^{N-1} x(m)y(n-m+pN)\mathrm{e}^{\beta pN} \\ &= \sum_{m=0}^{N-1} x(m)\alpha^p y(n-m+pN). \end{aligned} \tag{4.15}$$

Since the output signal is of length N, we have that $(n-m) \in \{-N+1, \ldots, N-1\}$, and thus $p \in \{0, 1\}$, so that (4.15) can be rewritten as

$$z(n) = \sum_{m=0}^{n} x(m)y(n-m) + \alpha \sum_{m=n+1}^{N-1} x(m)y(N+n-m)$$

□

4.8 Bibliography

[1] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms and Applications*, 4th ed. New York: Prentice Hall, 2007.

[2] C. Radhakrishnan and W. Jenkins, "Modified discrete Fourier transforms for fast convolution and adaptive filtering," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, May -June 2010, pp. 1611 – 1614.

[3] Z. Babic and D. Mandic, "A fast algorithm for linear convolution of discrete time signals," in *Telecommunications in Modern Satellite, Cable and Broadcasting Service, 2001. TELSIKS 2001. 5th International Conference on*, vol. 2, Sept. 2001, pp. 595 –598 vol.2.

[4] K. B. Wolf, *Integral Transforms in Science and Engineering*. New York: Plenum Press, 1979.

[5] M. Frigo and S. Johnson, "The design and implementation of fftw3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, Feb. 2005.

5

A generalized Fourier domain: signal processing framework and applications

Jorge Martínez, Richard Heusdens and Richard C. Hendriks

Abstract

In this paper, a signal processing framework in a generalized Fourier domain (GFD) is introduced. In this newly proposed domain, a special form of control on the periodic repetitions that occur due to sampling in the reciprocal domain is possible, without the need to increase the sampling rate. Important properties of the generalized discrete Fourier transform (GDFT), such as a weighted circular correlation property and Parseval's relation are derived. A non-zero padded overlap-add algorithm for linear filtering of long signals is derived, both for stationary and non-stationary filter conditions. The novel framework opens possibilities for signal processing applications. Examples of these applications are given and discussed.

5.1 Introduction

In this paper we introduce a framework for signal processing in a generalized Fourier domain (GFD). In this domain a special form of control on the periodic repetitions that occur due to sampling in the reciprocal domain is possible, without the need to increase the sampling rate. First in Sec. 5.2 we review the definition of the generalized discrete Fourier transform (GDFT) and its associated generalized Poisson summation formula (GPSF), both previously introduced in [1]. Analogous to the periodic extension of a finite-length signal that occurs in standard Fourier theory [2, 3], here we introduce the concept of "weighted periodic signal extension" that naturally occurs when working in the GFD. Next we study the connections of the presented theory to spectral sampling, analyticity and the z-transform. This analysis also serves as a discussion of the generalized Fourier transform and its relationship to the standard Fourier transform. In Sec. 5.3 important properties of the GDFT are derived such as the *weighted* circular correlation property and Parseval's energy relation for the GFD that, together with the previously introduced *weighted* circular convolution theorem for the GDFT, are fundamental to build a general-purpose GFD-based signal processing framework. To finalize our discussion in Sec. 5.4 we show how the novel framework can be used in spatial-audio applications such as the simulation of multichannel room impulse responses for auralization purposes in e.g. virtual reality and telegaming systems.

5.2 A generalized Fourier domain

Let us define the generalized discrete Fourier transform for finite-length signals $x(n)$, $n=\{0,\ldots,N-1\}$, with parameter $\alpha\in\mathbb{C}\backslash\{0\}$ as,

$$\mathcal{G}_\alpha\{x(n)\}\triangleq X_\alpha(k)=\sum_{n=0}^{N-1}x(n)\mathrm{e}^{\beta n}\mathrm{e}^{-j\frac{2\pi}{N}kn}, \tag{5.1}$$

for $k=\{0,\ldots,N-1\}$, where $\beta=\log(\alpha)/N$. The inverse GDFT is given by [1],

$$\mathcal{G}_\alpha^{-1}\{X_\alpha(k)\}\triangleq x(n)=\frac{\mathrm{e}^{-\beta n}}{N}\sum_{k=0}^{N-1}X_\alpha(k)\mathrm{e}^{j\frac{2\pi}{N}kn}. \tag{5.2}$$

The GDFT (5.1) is equivalent to the ordinary discrete Fourier transform of the modulated signal $x(n)\mathrm{e}^{\beta n}$. The finite-length signal $\mathrm{e}^{\beta n}$ for $n=\{0,\ldots,N-1\}$, is of finite energy for all $\alpha\in\mathbb{C}\backslash\{0\}$, therefore for $x(n)$ a signal of finite energy, the GDFT can be properly defined [1, 2]. Note that when $\alpha=1$, the transform pair correspond to the standard DFT pair.

Let us denote the periodic extension of $X_\alpha(k)$ by $\tilde{X}_\alpha(k)$ for $k\in\mathbb{Z}$. Clearly we have that $\tilde{X}_\alpha(k)=X_\alpha((k))_N$, where $X_\alpha((k))_N\triangleq X_\alpha(k \mod N)$, i.e. the circular shift of the sequence is represented as the index modulo N. On the other hand (5.1) and (5.2) imply a geometrically *weighted* periodic extension of the signal $x(n)$ when evaluated outside $\{0,\ldots,N-1\}$. This is stated by the generalized Poisson summation formula (GPSF) associated with the transform [1],

$$\tilde{x}_\alpha(n)\triangleq\sum_{p\in\mathbb{Z}}\alpha^p x((n))_N=\frac{\mathrm{e}^{-\beta n}}{N}\sum_{k=0}^{N-1}X_\alpha(k)\mathrm{e}^{j\frac{2\pi}{N}kn}, \tag{5.3}$$

where $n\in\mathbb{Z}$, $p=-\lfloor n/N\rfloor$ and $((n))_N=n+pN$. We can regard $\tilde{x}_\alpha(n)$ as a superposition of infinitely many translated and geometrically *weighted* "replicas" of $x(n)$. The replicas outside the support of $x(n)$ are weighted by α^p and $\tilde{x}_\alpha(n)=x(n)$ for $n=\{0,\ldots,N-1\}$. This is illustrated in Fig. 5.1, where a finite signal and (a part of) its geometrically weighted extension are depicted for $\alpha=0.5$ and $N=10$. Therefore to work in the generalized Fourier domain implies a manipulation of the signals involved via their geometrically weighted extensions. This is an important fact as we will see through the rest of the paper.

Signals of the form $\tilde{x}_\alpha(n)$ although infinitely long and not being of finite energy can be decomposed into its generalized Fourier transform components by means of (5.1), evaluating the transform over a signal interval ("period") of length N. This

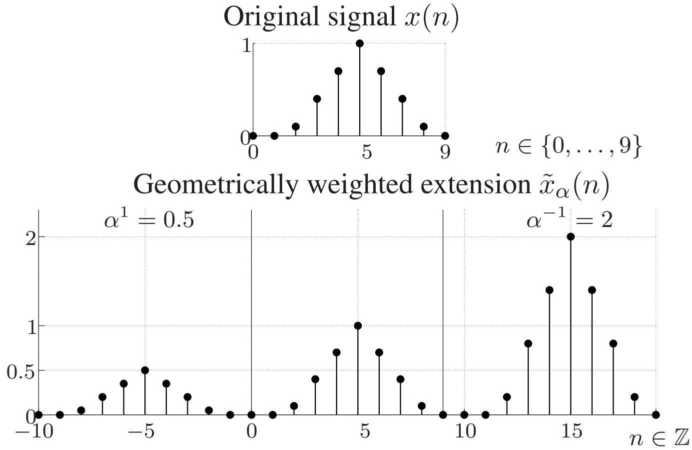


Figure 5.1: *Geometrically weighted extension of a finite length signal when evaluated outside its original domain, for $\alpha = 0.5$ and $N = 10$.*

fact follows directly from the generalized Poisson summation formula (5.3) which shows, that the inverse transform $\big(\exp(-\beta n)/N\big) \sum_{k=0}^{N-1} X_\alpha(k) \exp\big(j(2\pi/N)kn\big)$ is of the form $\tilde{x}_\alpha(n)$ when evaluated over $n = \mathbb{Z}$.

5.2.1 Connection to sampling, analyticity and the z-transform

The summation formula (5.3) has an important relationship to spectral sampling. The connection of the (discrete-time) generalized Fourier transform to analyticity and to the z-transform follows as part of the analysis. These relationships are used in a practical application of the theory in Sec. 5.4.

Let us begin with the connection to analyticity. Define the standard spectrum of a discrete-time signal by $S(\omega)$, where $\omega \in \mathbb{R}$ represents angular frequency and let $S(\omega) \in L^2[-\pi, \pi]$. Since the original signal $s(n)$, $n \in \mathbb{Z}$, is defined for discrete-time it is clear that $S(\omega)$ is a periodic function of ω with period equal to 2π. The signal $s(n)$ could represent the samples of a continuous-time signal, but without the sampling interval that information is lost and no particular analog representation is to be inferred. Let us assume for a moment that $S(\omega)$ can be analytically continued into the complex angular-frequency plane. This is $S(\omega) \to S(\omega_z)$, where $\omega_z \in \mathbb{C}$ is the complex-valued angular frequency. Let ω_r and ω_i denote the real and imaginary parts respectively of ω_z. Then from the definition of the discrete-time Fourier transform we have,

$$
\begin{aligned}
S(\omega_z) &= \sum_{n=-\infty}^{\infty} s(n)\mathrm{e}^{-j\omega_z n}, \qquad (5.4)\\
S(\omega_r + j\omega_i) &= \sum_{n=-\infty}^{\infty} s(n)\mathrm{e}^{-j(\omega_r + j\omega_i)n},\\
S(\omega_r + j\omega_i) &= \sum_{n=-\infty}^{\infty} s(n)\mathrm{e}^{\omega_i n}\mathrm{e}^{-j\omega_r n}.
\end{aligned}
$$

From here we see that if $s(n)$ is a causal sequence and the analytic continuation of $S(\omega)$ is done on the lower half of the complex plane then $\omega_i < 0$,

$$
S(\omega_r + j\omega_i) = \sum_{n=0}^{\infty} s(n)\mathrm{e}^{\omega_i n}\mathrm{e}^{-j\omega_r n},
$$

and the extra factor $\mathrm{e}^{\omega_i n}$ can only improve the convergence rate of the series. Now,

$$
\begin{aligned}
\lim_{\omega_i \to -0} S(\omega_r + j\omega_i) &= \lim_{\omega_i \to -0} \sum_{n=0}^{\infty} s(n)\mathrm{e}^{-j\omega_r n + \omega_i n},\\
&= \sum_{n=0}^{\infty} \lim_{\omega_i \to -0} s(n)\mathrm{e}^{-j\omega_r n + \omega_i n},\\
&= S(\omega).
\end{aligned}
$$

On the other hand we have that

$$
\begin{aligned}
\int_{-\pi}^{\pi} |S(\omega_r + j\omega_i)|^2 d\omega_r &= \int_{-\pi}^{\pi} S(\omega_r + j\omega_i) S^*(\omega_r + j\omega_i) d\omega_r,\\
&= \int_{-\pi}^{\pi} \left(\sum_{n=0}^{\infty} s(n)\mathrm{e}^{-j\omega_r n + \omega_i n} \sum_{m=0}^{\infty} s^*(m)\mathrm{e}^{j\omega_r m + \omega_i m} \right) d\omega_r,\\
&= \sum_{n=0}^{\infty} \left(s(n)\mathrm{e}^{\omega_i n} \sum_{m=0}^{\infty} s^*(m)\mathrm{e}^{\omega_i m} \int_{-\pi}^{\pi} \mathrm{e}^{j\omega_r (m-n)} d\omega_r \right),\\
&= 2\pi \sum_{n=0}^{\infty} s(n)\mathrm{e}^{\omega_i n} s^*(n)\mathrm{e}^{\omega_i n},\\
&= 2\pi \sum_{n=0}^{\infty} |s(n)|^2 \mathrm{e}^{2\omega_i n} < 2\pi \sum_{n=0}^{\infty} |s(n)|^2,
\end{aligned}
$$

where * denotes complex conjugation. Recalling Parseval's relation and noting that $S(\omega) \in L^2[-\pi, \pi]$ implies $s(n) \in l^2(\mathbb{Z})$ [2, 3], then for a positive constant C,

$$\int_{-\pi}^{\pi} |S(\omega_r + j\omega_i)|^2 d\omega_r < C.$$

So that $S(\omega_r + j\omega_i)$ is the analytic continuation from the real line into the lower half of the complex plane of the spectrum of the causal signal $s(n)$. By the same arguments if $s(n) = 0$ for $n > 0$ (i.e. is an anticausal signal), then its spectrum admits analytic continuation into the upper half of the complex angular-frequency plane.

Define now a discrete-time generalized Fourier transform as,

$$S_\alpha(\omega) = \sum_{n=-\infty}^{\infty} s(n)\mathrm{e}^{\beta n}\mathrm{e}^{-j\omega n}, \tag{5.5}$$

with inverse transformation,

$$s(n) = \frac{\mathrm{e}^{-\beta n}}{2\pi} = \int_{-\pi}^{\pi} S_\alpha(\omega)\mathrm{e}^{j\omega n} d\omega, \tag{5.6}$$

where, as in (5.1), $\beta = \log(\alpha)/N$ and $\alpha \in \mathbb{C}\backslash\{0\}$. By our previous discussion we can write

$$S_\alpha(\omega) = S(\omega + j\beta) = S(\omega - \beta_i + j\beta_r),$$

where β_r and β_i are the real and imaginary parts respectively of parameter β. Clearly if $\beta_r < 0$ (or equivalently $|\alpha| < 1$) the transform is well defined for causal signals. In the same way if $\beta_r > 0$ (i.e. $|\alpha| > 1$) the transform is well defined for anticausal signals. In both cases the generalized spectrum can be obtained via analytic continuation (into the proper half of the complex plane) of the standard Fourier spectrum. Also note that β_i implies a frequency shift. If the principal value of $\log(\alpha)$ is to be taken this shift is limited from $-\pi/N$ to π/N.

Recall now the definition of the z-transform,

$$S(z) = \sum_{n=-\infty}^{\infty} s(n)z^{-n}$$

Since parameter α is constant, we find the transform (5.5) to be a particular case of the z-transform, with $z = \mathrm{e}^{-\beta_r - j(\beta_i - \omega)}$. This implies $|z| = |\alpha|^{-1/N}$, with the real positive Nth-root being the only root satisfying the equation. The discrete-time GFT can be viewed as the z-transform of the signal evaluated on a circle of radius $|\alpha|^{-1/N}$. When $|\alpha|=1$ the evaluation is done on the unit circle and the GFT is equivalent to the standard Fourier transform shifted in frequency this is, $S_\alpha(\omega)=S(\omega-\theta/N)$, with $\alpha=\mathrm{e}^{j\theta}$. We can now extend the definition of the GFT to signals other than single-sided (causal or anticausal). If the z-transform of the signal has a region of convergence that includes the circle of radius $|\alpha|^{-1/N}$, then the GFT exists. Note that finite-length signals have as region of convergence the whole z-plane with exception of the points $z=0$ and/or $z=\infty$. Since $\alpha\in\mathbb{C}\backslash\{0\}$ the GFD always exists for finite-length signals of finite energy.

The relationship of (5.3) to spectral sampling is now explored. Since $S_\alpha(\omega)$ is a periodic function of ω the integral in (5.6) can be taken over any interval of length 2π. To make the derivation simpler let

$$s(n)=\frac{\mathrm{e}^{-\beta n}}{2\pi}=\int_0^{2\pi} S_\alpha(\omega)\mathrm{e}^{j\omega n}d\omega.$$

The integral can be approximated using a rectangular quadrature rule, dividing the integration interval uniformly into N subintervals and using the samples of the integrand at the subinterval points. Let $2\pi/N$ be the sampling interval, then

$$\int_0^{2\pi} S_\alpha(\omega)\mathrm{e}^{j\omega n}d\omega \approx \frac{2\pi}{N}\sum_{k=0}^{N-1} S_\alpha\left(\frac{2\pi}{N}k\right)\mathrm{e}^{j\frac{2\pi}{N}nk}.$$

An approximation of $s(n)$, call it $\tilde{s}_\alpha(n)$, is thus obtained as

$$\tilde{s}_\alpha(n)=\frac{\mathrm{e}^{-\beta n}}{N}\sum_{k=0}^{N-1} S_\alpha\left(\frac{2\pi}{N}k\right)\mathrm{e}^{j\frac{2\pi}{N}nk}.$$

Substituting (5.5) into this last expression reveals the connection of $\tilde{s}_\alpha(n)$ to the original signal $s(n)$,

$$\begin{aligned}\tilde{s}_\alpha(n)&=\frac{\mathrm{e}^{-\beta n}}{N}\sum_{k=0}^{N-1}\left(\sum_{m=-\infty}^{\infty} s(m)\mathrm{e}^{\beta m}\mathrm{e}^{-j\frac{2\pi}{N}km}\right)\mathrm{e}^{j\frac{2\pi}{N}kn}\\&=\frac{\mathrm{e}^{-\beta n}}{N}\sum_{m=-\infty}^{\infty} s(m)\mathrm{e}^{\beta m}\left(\sum_{k=0}^{N-1}\mathrm{e}^{j\frac{2\pi}{N}k(n-m)}\right).\end{aligned}$$

For $p\in\mathbb{Z}$, $p{=}{-}\lfloor n/N\rfloor$ we have [3],

$$\sum_{k=0}^{N-1}\mathrm{e}^{j\frac{2\pi}{N}k(n-m)}=\begin{cases} N, & m=n+pN\\ 0, & \text{otherwise.}\end{cases}$$

Then

$$\tilde{s}_\alpha(n) = \mathrm{e}^{-\beta n} \sum_{p=-\infty}^{\infty} s(n+pN)\mathrm{e}^{\beta(n+pN)}$$
$$= \sum_{p=-\infty}^{\infty} s(n+pN)\mathrm{e}^{\beta pN} = \sum_{p=-\infty}^{\infty} \alpha^p s(n+pN).$$

So that

$$\tilde{s}_\alpha(n) = \sum_{p=-\infty}^{\infty} \alpha^p s(n+pN) = \frac{\mathrm{e}^{-\beta n}}{N} \sum_{k=0}^{N-1} S_\alpha\left(\frac{2\pi}{N}k\right) \mathrm{e}^{j\frac{2\pi}{N}nk}.$$

We have arrived to the generalized Poisson summation formula (5.3), which states that uniform spectral sampling of the generalized Fourier spectrum $S_\alpha(\omega)$ implies a geometrically weighted periodic summation of the original discrete-time signal $s(n)$. This property is used in a spatial-audio application in Sec. 5.4, but first we analyze some properties of the GDFT.

5.3 Properties of the GDFT

In this section we present some important properties of the GDFT, these are a fundamental part in any GFD-based signal processing framework. In the following let $x(n)$ and $y(n)$ for $n = \{0, \ldots, N-1\}$ be two finite-duration and in general complex signals of length N, and $X_\alpha(k)$ and $Y_\alpha(k)$ for $k = \{0, \ldots, N-1\}$ their respective GDFTs.

Proposition 5.1. *Weighted circular convolution. Point-wise multiplication of $X_\alpha(k)$ and $Y_\alpha(k)$ in the GFD corresponds to the weighted circular convolution of $x(n)$ and $y(n)$ in the time domain, i.e.*

$$\begin{aligned}\mathcal{G}_\alpha^{-1}\{X_\alpha(k)Y_\alpha(k)\} &= \frac{\mathrm{e}^{-\beta n}}{N} \sum_{k=0}^{N-1} X_\alpha(k)Y_\alpha(k)\mathrm{e}^{j\frac{2\pi}{N}nk} \\ &= \sum_{m=0}^{N-1} x(m)\tilde{y}_\alpha(n-m),\end{aligned}$$

$$\begin{aligned}\mathcal{G}_\alpha^{-1}\left\{X_\alpha(k)Y_\alpha(k)\right\} &= \sum_{m=0}^{N-1} x(m)\alpha^p y(n-m+pN) \\ &= \sum_{m=0}^{N-1} x(m)\alpha^p y((n-m))_N, \end{aligned} \tag{5.7}$$

or

$$(x * \tilde{y}_\alpha)(n) \overset{\mathcal{G}_\alpha}{\longleftrightarrow} X_\alpha(k)Y_\alpha(k),$$

where $*$ is the linear convolution operator. The operation can thus be seen as the linear convolution of one of the signals e.g. $x(n)$, with the respective signal extension of the other, $\tilde{y}_\alpha(n)$. Note that for $n = \{0, \ldots, N-1\}$ we have that,

$$(n-m) \in \{-N+1, \ldots, N-1\},$$

and thus $p \in \{0, 1\}$, so that (5.7) can be rewritten as

$$\sum_{m=0}^{n} x(m)y(n-m) + \alpha \sum_{m=n+1}^{N-1} x(m)y(N+n-m), \tag{5.8}$$

where the left hand summation represents the contribution of N linear convolution terms, and the right hand summation the contribution of N circular convolution terms (which are in fact the last terms of the linear convolution). The factor α effectively weights the amount of circular convolution that is obtained. The proof is given in [1]. Property 5.1 can be exploited to compute linear convolutions in the GFD without the need of zero-padding, using GDFTs with e.g. parameter $\alpha = \pm j$ or $\alpha << 1 \in \mathbb{R}$, [1].

Next we present the shifting properties for the GDFT, the first accounts for a shift in the time-domain the second for a shift in the GFD.

Proposition 5.2. *Time-domain shift.*
For $n_0 \in \mathbb{Z}$*,*

$$\tilde{x}_\alpha(n-n_0) \overset{\mathcal{G}_\alpha}{\longleftrightarrow} \mathrm{e}^{\beta n_0} \mathrm{e}^{-j\frac{2\pi}{N}kn_0} X_\alpha(k).$$

The GDFT with parameter α of the shifted signal $\tilde{x}_\alpha(n-n_0)$ is equal to the modulated generalized spectrum of the original signal $x(n)$. The proof is given in 5.A.

Proposition 5.3. *GFD shift.*
For $k_0 \in \mathbb{Z}$*,*

$$x(n)\mathrm{e}^{j\frac{2\pi}{N}k_0 n} \overset{\mathcal{G}_\alpha}{\longleftrightarrow} \tilde{X}_\alpha(k-k_0) = X_\alpha((k-k_0))_N.$$

To circularly shift the generalized spectrum $X_\alpha(k)$ of a signal is equivalent in the time-domain to modulate the signal with the function $e^{j\frac{2\pi}{N}k_0 n}$. The proof is given in 5.B.

Proposition 5.4. *Time reversal.*

$$\tilde{x}_\alpha(-n) \stackrel{\mathcal{G}_{\alpha^{-1}}}{\longleftrightarrow} \tilde{X}_\alpha(-k).$$

The GDFT with parameter α^{-1} of the time-reversed extension of x with parameter α, $\tilde{x}_\alpha(-n)$, is thus equivalent to reversing (modulo N) the GDFT of $x(n)$ with parameter α. This result is a direct consequence of the reciprocal-symmetric structure of the signal extension $\tilde{x}_\alpha(n)$ with respect to α. Notice that if the extension is time reversed the geometrically weighted "replicas" outside the support of $x(-n)$ no longer correspond to a weight α^p but to α^{-p}. Therefore to obtain property 5.4 a GDFT with parameter α^{-1} has to be applied to the time reversed extension, $\tilde{x}_\alpha(-n)$. The proof is given in 5.C.

Proposition 5.5. *Time domain complex-conjugate.*

$$x^*(n) \stackrel{\mathcal{G}_{\alpha^*}}{\longleftrightarrow} \tilde{X}^*_\alpha(-k).$$

To take the inverse GDFT with parameter α^* of $\tilde{X}^*_\alpha(-k)$ is equivalent to take the complex conjugate of the time domain signal $x^*(n)$. The proof is given in 5.D.

Consider that the real part of a complex signal is given by $\Re\{x(n)\} = (1/2)(x(n) + x^*(n))$, and its imaginary part is given by $\Im\{x(n)\} = (1/2j)(x(n) - x^*(n))$. The last property (making $\alpha \rightarrow \alpha^*$) can then be used to derive the following results,

$$\Re\{x(n)\} \stackrel{\mathcal{G}_{\alpha}}{\longleftrightarrow} \frac{1}{2}\left(X_\alpha + \tilde{X}^*_{\alpha^*}(-k)\right).$$

The real part of a complex signal $x(n)$ can be obtained by taking the inverse GDFT with parameter α of a linear combination of the GDFT of $x(n)$ with parameter α, and the GDFT of $x(n)$ with parameter α^* conjugated and reversed (modulo N). Correspondingly we also have that,

$$\Im\{x(n)\} \stackrel{\mathcal{G}_{\alpha}}{\longleftrightarrow} \frac{1}{2j}\left(X_\alpha - \tilde{X}^*_{\alpha^*}(-k)\right).$$

Proposition 5.6. *GFD complex-conjugate.*

$$\tilde{x}^*_\alpha(-n) \stackrel{\mathcal{G}_{\alpha^{-*}}}{\longleftrightarrow} X^*_\alpha(k),$$

where $\alpha^{-*} = (\alpha^*)^{-1}$. To take the complex conjugate of the spectrum, $X^*_\alpha(k)$ is equivalent to take the GDFT with parameter α^{-*} of $\tilde{x}^*_\alpha(-n)$. The proof is given in 5.E.

Before we proceed with the next property let us define the weighted circular correlation of two in general complex length N signals, $x(n)$ and $y(n)$ by the linear (deterministic) correlation function of $x(n)$ and $\tilde{y}_{\alpha^{-*}}(n)$, i.e,

$$\begin{aligned}
\tilde{r}_{\alpha,xy}(n) &= x(n) * \tilde{y}^*_{\alpha^{-*}}(-n) \\
&= \sum_{m=0}^{N-1} x(m)\tilde{y}^*_{\alpha^{-*}}(m-n) \\
&= \sum_{m=0}^{N-1} x(m)\left(\alpha^{-p} y^*(m-n+pN)\right) \\
&= \sum_{m=0}^{N-1} x(m)\Big(\alpha^{-p} y^*((m-n))_N\Big),
\end{aligned} \tag{5.9}$$

where we have used the fact that linear correlation can be expressed in terms of the linear convolution of the signals with one of them time-reversed and conjugated. Note that for $n = \{0, \ldots, N-1\}$ we have,

$$\tilde{r}_{\alpha,xy}(n) = \begin{cases} \sum_{m=0}^{N-1} x(m)y^*(m), & n=0 \\ \alpha^{-1} \sum_{m=0}^{n-1} x(m)y^*(N+m-n) & \\ \quad + \sum_{m=n}^{N-1} x(m)y^*(m-n), & \text{otherwise.} \end{cases} \tag{5.10}$$

Let us now state the following property.

Proposition 5.7. *Weighted circular correlation.*

$$\tilde{r}_{\alpha,xy}(n) \xleftrightarrow{\mathcal{G}_\alpha} R_{\alpha,xy}(k) = X_\alpha(k)Y^*_{\alpha^{-*}}(k),$$

where by definition $R_{\alpha,xy}(k)$ is the GDFT of $\tilde{r}_{\alpha,xy}(n) = x(n) * \tilde{y}^*_{\alpha^{-*}}(-n)$. The proof of this property follows directly from the complex-conjugate property (property 5.6) with $\alpha \rightarrow \alpha^{-*}$, the weighted circular convolution property (property 5.1), and Eq. (5.9).

In analogy to the weighted circular convolution property a weighted circular correlation can be obtained by point-wise multiplication of the spectra in the GFD. However in this case one of the two spectra corresponds to the complex conjugate of the GDFT of the signal with parameter α^{-*}. Notice that when $|\alpha|=1$ we have,

$$\tilde{r}_{\alpha,xy}(n) \overset{\mathcal{G}_\alpha}{\longleftrightarrow} R_{\alpha,xy}(k) = X_\alpha(k)Y_\alpha^*(k),$$

this includes the standard DFT correlation theorem ($\alpha=1$).

The following property is a direct consequence of the weighted circular correlation theorem.

Proposition 5.8. *Parseval's energy relation.*

$$\sum_{n=0}^{N-1} x(n)y^*(n) = \frac{1}{N}\sum_{k=0}^{N-1} X_\alpha(k)Y_{\alpha^{-*}}^*(k). \tag{5.11}$$

This equality represents Parseval's theorem for the GDFT. It follows by evaluating $\mathcal{G}_\alpha^{-1}\{R_{\alpha,xy}\}=\tilde{r}_{\alpha,xy}(n)$ at $n=0$. For the case $y(n)=x(n)$ we further have,

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N}\sum_{k=0}^{N-1} X_\alpha(k)X_{\alpha^{-*}}^*(k). \tag{5.12}$$

The energy in the finite duration signal $x(n)$ is expressed in terms of the frequency components $\{X_\alpha(k)X_{\alpha^{-*}}^*(k)\}_{k=0}^{N-1}$. From here we see that if $|\alpha|=1$ the energy in $x(n)$ equals $1/N$ times the energy in $X_\alpha(k)$ i.e.,

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |X_\alpha(k)|^2, \quad \text{for all } \alpha : |\alpha| = 1. \tag{5.13}$$

5.4 A spatial-audio signal processing application

In this section we consider the simulation of room impulse responses as an application for the GFD-framework just presented. Let us start with the sound field inside a box-shaped room which always contains reverberation (at least in the vast majority of real-life cases). If the source of sound is perfectly omnidirectional (a monopole) and produces a perfect delta pulse at a certain time, then the resulting sound field measured at a single point in space is called the room impulse response (RIR) [4, 5]. Current approaches to model the sound field in a room although accurate are computationally complex [4, 6–9]. In the context of spatial-audio applications like virtual reality systems, real-time or interactive simulation of RIRs at all positions in a room becomes a challenging problem.

Consider now a room with fully reflective walls. In this case, the sound field inside the room is given by a *periodic summation* of the sound field of the source [10]. Intuitively this summation represents the effect of reverberation, since the reflections of the sound field produced by the source(s) on the walls can be modeled by spatial copies of the sources outside the room. If a room could have fully reflective walls, these copies would be perfect and the summation would be perfectly periodic. A key observation to derive a fast algorithm to model the sound field in a room is then the following, *sampling of a function results in a periodic summation of its Fourier transform*. This relation is given by the Poisson summation formula [1, 2] and it is a well known property in digital signal processing (see, e.g. [3]). If we carefully sample the spatio-temporal spectrum of the sound field produced by the source and apply an inverse Fourier transform on this sampled spectrum we can obtain the required periodic summation that constitutes the sound field in the whole room [10]. Using this method we dramatically reduce the complexity needed to compute individual impulse responses from $\mathcal{O}(N_t^3)$ per recever position (with N_t proportional to the desired reverberation time T_{60}) of approaches related or based on the mirror image source method [7], to $\mathcal{O}(N_\omega \log(N_\omega))$ (with N_ω proportional to the maximum desired temporal bandwidth say, ω_b) taking advantage of the FFT. On the other hand in virtually all real-life cases the walls in a room are at least partially absorptive, the summation defining the sound field in a room is therefore never perfectly periodic. Using standard Fourier theory is however impossible to obtain something different than a periodicity-sampling relation in reciprocal domains. And therefore although of theoretical importance, the method in [10] has no direct practical use.

If the walls are no longer fully reflective in [11] is shown that the sound field in the room can be modeled by a *weighted periodic summation*, of the form given by the generalized Poisson summation formula [1] (in this paper a discrete version of the formula is given by (5.3)). Every time the sound field is reflected on a wall, part of its energy is absorbed and its frequency components might experience a phase change. Higher order reflections can then be seen as geometrically weighted copies of the sound field of the source. A GFD based method for the simulation of RIRs is then derived as follows. For every source in the room the first-order reflections on orthogonal walls [10, 11] are considered first. To model these, another seven virtual sources are added at positions outside the room. A total of eight *mother sources* are then considered. The sound field of each mother source is then factored into waves traveling only in the direction of each of the eight space octants. This gives rise to a total of $2^{(3\times 2)}$ spatio-temporal functions to be considered. Let $p_{lq}(\mathbf{x}, t)$ represent these functions where $l=\{0,\ldots,7\}$ is the mother source index and $q=\{0,\ldots,7\}$ is an enumeration of the octants, $\mathbf{x}=[x,y,z]^T$ is the space variable vector where the superscript T denotes matrix transposition, and $t\in\mathbb{R}$ denotes time. The reverberated sound field in the room is then modeled by [11],

$$p(\mathbf{x},t)=\sum_{l=0}^{7}\sum_{q=0}^{7}\varrho_{lq}\sum_{\mathbf{n}\in\mathbb{Z}^3}\left(\prod_{i\in\{x,y,z\}}\varrho_i^{\pm n_i}\right)p_{lq}(\mathbf{x}+\mathbf{\Lambda n},t), \tag{5.14}$$

where $\mathbf{\Lambda}$ is the generator matrix of the periodicity lattice Λ (i.e. the multidimensional signal "period"), ϱ_{lq} are constants required, $\mathbf{n}=[n_x,n_y,n_y]$ is a triplet of integers, ϱ_i, for $i\in\{x,y,z\}$ are the reflection factors of the walls, e.g. $\varrho_x=\varrho_{x0}\varrho_{x1}$ where ϱ_{x0} is the reflection factor of the wall perpendicular to the x direction at the origin of coordinates and ϱ_{x1} the reflection factor of the opposite wall. The sign of the exponent in the product over $i\in\{x,y,z\}$ depends on the particular octant in the definition of the function ϱ_{lq}. The reader is referred to [11] for the details of this derivation. The important result behind (5.14) is that the infinite summation over $\mathbf{n}$ is a *weighted periodic summation*, of the form given by the generalized Poisson summation formula, which for multidimensional signals takes the form,

$$\sum_{\mathbf{n}\in\mathbb{Z}^\nu}\left(\prod_{i=0}^{\nu-1}\alpha_i^{n_i}\right)p(\mathbf{x}+\mathbf{\Lambda n})=\frac{\mathrm{e}^{-\boldsymbol{\beta}^T\mathbf{x}}}{|\mathbf{\Lambda}|}\sum_{\mathbf{k}\in\mathbb{Z}^\nu}P_{\boldsymbol{\alpha}}(\mathbf{\Phi k})\mathrm{e}^{j(\mathbf{k}^T\mathbf{\Phi}^T\mathbf{x})}, \tag{5.15}$$

where $\nu\in\mathbb{N}$ is the dimension of the space, $|\mathbf{\Lambda}|$ is the absolute value of the determinant of $\mathbf{\Lambda}$, $\mathbf{\Phi}=2\pi\mathbf{\Lambda}^{-T}$ is the generator matrix of the spectral sampling lattice Φ (the (scaled) reciprocal lattice of the periodicity lattice Λ), $\boldsymbol{\beta}=\mathbf{\Lambda}^{-T}\log(\boldsymbol{\alpha})$, $\boldsymbol{\alpha}\in\mathbb{C}^\nu:\alpha_i\neq 0\ \forall i=0,\ldots,\nu-1$, is the parameter of the multidimensional generalized Fourier transform $P_{\boldsymbol{\alpha}}$, and $\log(\boldsymbol{\alpha})\triangleq[\log(\alpha_0),\ldots,\log(\alpha_{\nu-1})]^T$.

The main result in [11] relates the sound field in a room with a sampling condition on the generalized Fourier spectrum. This is, if $\mathbf{\Lambda}$ denotes the generator matrix of the lattice specifying the spatial periodic packing of the sound fields $p_{lq}(\mathbf{x},t)$, and $\mathbf{\Phi}$ denotes the generator matrix of the lattice specifying the sampling points of the spatial-generalized spectra, then making $\mathbf{\Phi}=2\pi\mathbf{\Lambda}^{-T}$, the functions $P_{\boldsymbol{\alpha}lq}(\mathbf{\Phi k},\omega)$, $\mathbf{k}\in\mathbb{Z}^3$ are the generalized Fourier coefficients of

$$\sum_{\mathbf{n}\in\mathbb{Z}^3}\left(\prod_{i\in\{x,y,z\}}\varrho_i^{\varsigma_q(i)n_i}\right)p_{lq}(\mathbf{x}+\mathbf{\Lambda n},t), \tag{5.16}$$

with $\boldsymbol{\alpha}=[\varrho_x^{\varsigma_q(x)},\varrho_y^{\varsigma_q(y)},\varrho_z^{\varsigma_q(z)},1]^T$, where ω is the temporal frequency variable and $\varsigma_q(\cdot)=\pm1$, depending on the coordinates defining the qth octant of the space.

To apply the method on a computer, all frequency variables (not only the spatial-frequency variable) must be sampled. Sampling the temporal-frequency variable ω introduces temporal aliasing. Let $\boldsymbol{\Psi}$ be the matrix of the spectral sampling lattice $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\Phi}, \Omega_s)$, Ω_s is the temporal-frequency sampling interval. Define $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\Lambda}, T_p)$, so that $\boldsymbol{\Psi} = 2\pi\boldsymbol{\Delta}^{-T} = 2\pi\text{diag}(\boldsymbol{\Lambda}, T_p)^{-T}$, where $T_p = 2\pi/\Omega_s$ is the interval of temporal periodicity. Then

$$\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{x}, t) \triangleq \sum_{\mathbf{n}\in\mathbb{Z}^3}\sum_{n\in\mathbb{Z}}\left(\prod_{i\in\{x,y,z\}} \varrho_i^{\varsigma_q(i)n_i}\right) p_{lq}(\mathbf{x} + \boldsymbol{\Lambda}\mathbf{n}, t + T_p n). \tag{5.17}$$

The summation over $n \in \mathbb{Z}$, $n \neq 0$, is the temporal aliasing. We can neglect it making $\Omega_s << 1$, (T_p becomes large), this increases computational complexity since a smaller sampling interval implies more samples needed in the reconstruction. Taking advantage of the GFD framework, temporal aliasing can be further reduced using a temporal component different than 1 in the $\boldsymbol{\alpha}$ parameter of (5.16), without the need to change the sampling rate. We derive this below, after the current discussion.

The sound field is therefore approximated,

$$p(\mathbf{x}, t) \approx \sum_{l=0}^{7}\sum_{q=0}^{7} \frac{\mathrm{e}^{-\boldsymbol{\beta}_q^T\mathbf{x}-\beta t}}{|\boldsymbol{\Delta}|} \sum_{\mathbf{k}\in\mathbb{Z}^4} \varrho_{lq} P_{\boldsymbol{\alpha}lq}(\boldsymbol{\Psi}\mathbf{k})\mathrm{e}^{j(\mathbf{k}^T\boldsymbol{\Psi}^T[\mathbf{x}^T,t]^T)}, \tag{5.18}$$

Further, the infinite summation over $\mathbf{k}$ in (5.18), must be limited to a finite number of elements. Extending periodically this finite set of spectral coefficients, we impose a discretization of the space-time function, so that a sampled (in both space and time) sound field is approximated (which can then be handled on a computer). The spectral set must be big enough to cover the support of the spectrum if corruption due to aliasing is to be avoided.

Let $\Sigma \subseteq \Psi$, be the spectral periodicity lattice, and Γ the spatio-temporal sampling lattice, assume $\Delta \subseteq \Gamma$. Then $\boldsymbol{\Gamma} = 2\pi\boldsymbol{\Sigma}^{-T}$, so that,

$$\tilde{p}_{\boldsymbol{\alpha}lq}(\boldsymbol{\Gamma}\mathbf{n}) = \frac{\mathrm{e}^{-\boldsymbol{\beta}^T\boldsymbol{\Gamma}\mathbf{n}}}{\mathrm{N}(\Delta/\Gamma)} \sum_{\mathbf{k}\in V_{\boldsymbol{\Sigma}}(\mathbf{0})} |\boldsymbol{\Gamma}|^{-1} P_{\boldsymbol{\alpha}lq}(\boldsymbol{\Psi}\mathbf{k})\mathrm{e}^{j(\mathbf{k}^T\boldsymbol{\Psi}^T\boldsymbol{\Gamma}\mathbf{n})}, \tag{5.19}$$

where $V_\Sigma(\mathbf{0})$ is the (central) Voronoi region around the origin of lattice Σ, $\mathrm{N}(\Delta/\Gamma)$ is the number of lattice points of Γ that lie in $V_\Delta(0)$ (the central Voronoi region of lattice Δ). Making $V_{\boldsymbol{\Sigma}}(\mathbf{0})$ larger implies a finer sampling of $\tilde{p}_{\boldsymbol{\alpha}lq}$. Further we have that,

$$\mathrm{N}(\Delta/\Gamma) = \frac{|\boldsymbol{\Delta}|}{|\boldsymbol{\Gamma}|} = \frac{(2\pi)^4|\boldsymbol{\Psi}^{-T}|}{(2\pi)^4|\boldsymbol{\Sigma}^{-T}|} = \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Psi}|} = \mathrm{N}(\Sigma/\Psi).$$

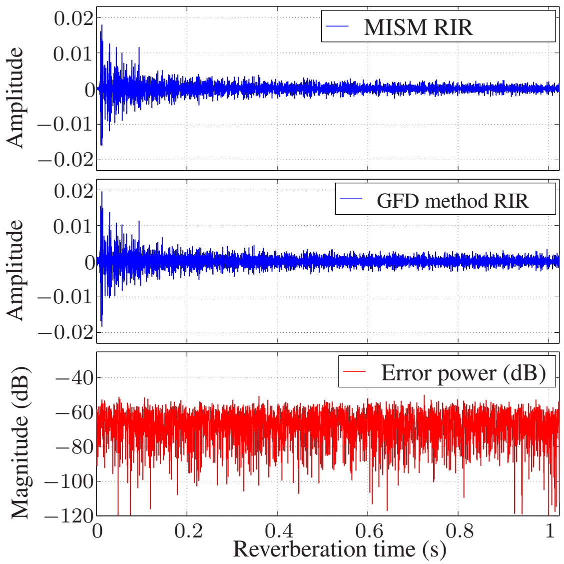


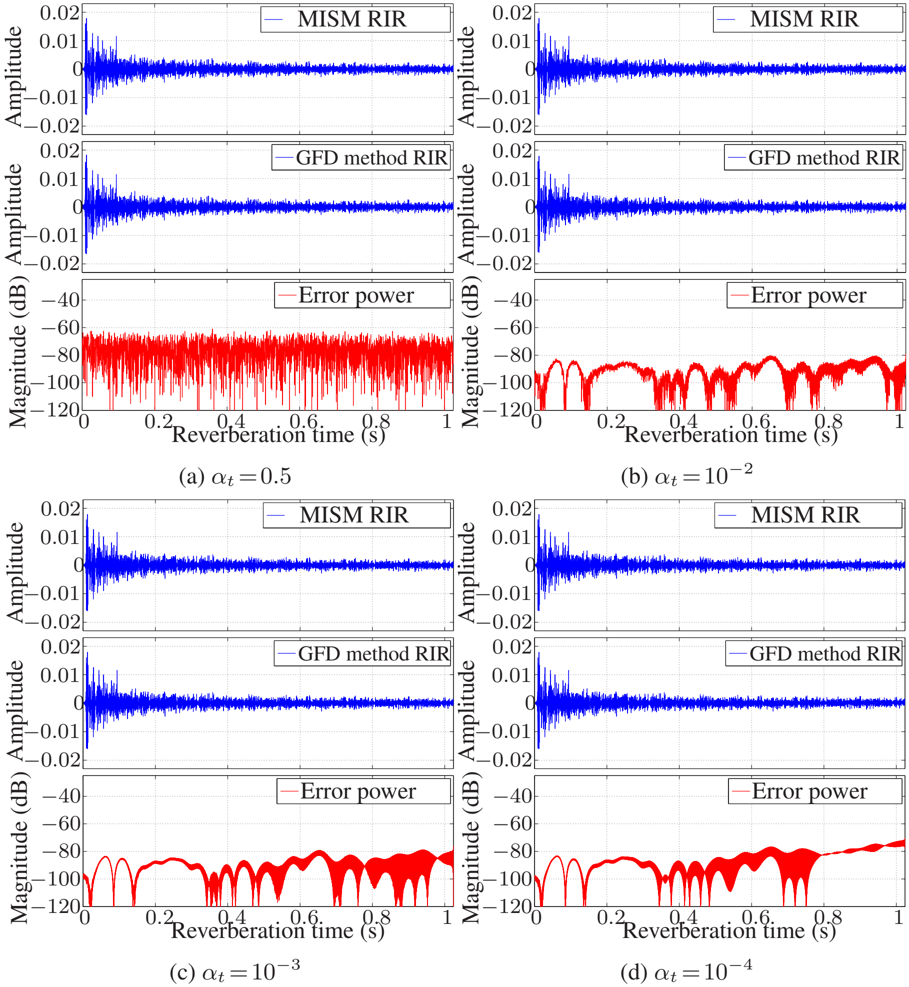
Figure 5.2: Simulated Room impulse responses.

The sampled sound field is thus obtained by,

$$p(\mathbf{\Gamma n}) \approx \sum_{l=0}^{7}\sum_{q=0}^{7} \varrho_{lq}\tilde{p}_{\boldsymbol{\alpha} lq}(\mathbf{\Gamma n}), \qquad \text{for } \mathbf{\Gamma n} \in V_{\Delta}(\mathbf{0}). \tag{5.20}$$

Considering that the spectrum energy is concentrated in $||\phi|| \leq |\omega/c|$ [12] we can evaluate up to a given ω_b. Note that (5.19) has the form of a generalized Poisson summation formula (the multidimensional extension of 5.3), the right hand term is thus a multidimensional GDFT and the inner summation over $\mathbf{k}$ corresponds to a DFT (this comes from the fact that the GDFT is equivalent to the DFT of the modulated input signal). Using the FFT, the operation will take only $\mathcal{O}(N_\omega^4 \log N_\omega)$ operations for computing $\mathrm{N}(\Delta/\Gamma)$ spatio-temporal positions, with N_ω proportional to ω_b. Since $\mathrm{N}(\Delta/\Gamma) = \mathrm{N}(\Sigma/\Psi)$, the method is of complexity $\mathcal{O}(N_\omega \log N_\omega)$ per receiver position. Again the reader is referred to [11] for detailed experimental results.

Returning to Eq. (5.17), we see that temporal aliasing is introduced due to spectral sampling of ω. Clearly, by making the spectral sampling period Ω_s smaller, the aliasing components in the time dimension appear further apart from each other, reducing the error. In time the RIR does not have compact support, but it is a *causal*



(a) $\alpha_t = 0.5$ (b) $\alpha_t = 10^{-2}$

(c) $\alpha_t = 10^{-3}$ (d) $\alpha_t = 10^{-4}$

Figure 5.3: Comparison of RIRs simulated with the GFD method setting parameter α_t to different values, and the same RIR simulated with the MISM depicted in Fig. 5.2.

(single sided) function, having a starting point in time and an exponential decay afterwards (see e.g. [5] for a parametric characterization of this decay). A simulated example RIR using the Mirror Image Source Method (MISM) [7], is depicted in Fig. 5.2, together with a RIR simulated using the GFD approach described above and in [11], the error power between both signals is plotted in dB. The bandwidth frequency is

$\omega_b = 2\pi(2\text{kHz})$, so that the temporal sampling frequency is $f_s = 4\text{kHz}$. The length of the RIR is $T_h = 1.02\text{s}$ or 4096 taps. The GFD method RIR shown in Fig. 5.2 is obtained according to (5.19) and (5.20) using a parameter $\boldsymbol{\alpha} = [\varrho_x^{\varsigma_q(x)}, \varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}, 1]^T$, so that the spatial part of the $\boldsymbol{\alpha}$ parameter applied in the generalized Fourier synthesis gives the required spatial weighted periodicity, and in time a standard Fourier synthesis is applied. The RIR is thus one spatial sample of the set $p(\boldsymbol{\Gamma}\mathbf{n})$. The spectral sampling period is set to $\Omega_s = 2\pi/(2T_h)$, so that the interval of temporal periodicity (and thus temporal aliasing) T_p is 2 times the reverberation time.

Using a temporal component $\alpha_t \neq 1$ in the $\boldsymbol{\alpha}$ parameter, we can further reduce the temporal aliasing. This is, since the RIR is a causal function of time, the repeated terms to the right of the temporal support of the RIR (for $n < 0$), do not contribute to the time-domain aliasing. Therefore we can rewrite (5.17) as,

$$\begin{aligned}\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{x}, t) = &\sum_{\mathbf{n}\in\mathbb{Z}^3} \left(\prod_{i\in\{x,y,z\}} \varrho_i^{\varsigma_q(i)n_i} \right) p_{lq}(\mathbf{x} + \boldsymbol{\Lambda}\mathbf{n}, t) \\ &+ \sum_{n=1}^{\infty} \sum_{\mathbf{n}\in\mathbb{Z}^3} (\alpha_t^n) \left(\prod_{i\in\{x,y,z\}} \varrho_i^{\varsigma_q(i)n_i} \right) p_{lq}(\mathbf{x} + \boldsymbol{\Lambda}\mathbf{n}, t + T_p n). \qquad (5.21)\end{aligned}$$

In principle by making e.g. $\alpha_t << 1$ we can further reduce the temporal aliasing without the need to increase the spectral sampling period Ω_s. In practice however, the accuracy of the computations is limited by the arithmetic precision used, moreover the causality of the RIR in time is only strictly valid in non-bandlimited scenarios. This is, by limiting the summation over $\mathbf{k}$ in (5.18) to a finite number of elements, we are effectively multiplying the discrete spatio-temporal spectrum by a multidimensional rectangular window. In space-time this has the effect of a convolution with a multidimensional sinc function, making the resulting band-limited RIR non-causal. In this case (5.17) for $n \neq 0$ defines the aliasing terms, but still the terms for $n < 0$ have less corruptive influence. Despite these practical issues it is still possible to reduce the temporal aliasing using a temporal component e.g. $\alpha_t << 1$ in the generalized Fourier synthesis (5.19). In Fig. 5.4a, Fig. 5.4b, Fig. 5.4c and Fig. 5.4d results of the generalized Fourier synthesis are given setting $\alpha_t = 0.5$, $\alpha_t = 10^{-2}$, $\alpha_t = 10^{-3}$ and $\alpha_t = 10^{-4}$ respectively. Indeed aliasing corruption decreases for the first two cases, but for $\alpha_t = 10^{-3}$ and $\alpha_t = 10^{-4}$ the repeated terms to the right of the temporal support of the RIR become too large, having a negative impact in the reconstruction. Clearly the corruption is more pronounced to the right of the support of the RIR. In this case setting the parameter $\alpha_t = 10^{-2}$ gives a good reconstruction, specially when compared with the result obtained setting $\alpha_t = 1$ depicted in Fig. 5.2. Working in the

GFD on the temporal dimension allows a non neglectable gain in accuracy.

The method for multichannel simulation of RIRs has an important application in immersive virtual gaming (using for example stereo headphones). In this case many RIRs for different (virtual) room conditions need to be computed and later fast convolved with a given audio signal (i.e. auralization) to give the users an audio experience such that they have the impression of being in the game field. For example, at one moment the users could be at an open location such as a park, and at another moment they could be inside a room. To create a satisfactory experience, the system have to reproduce the acoustic characteristics of different scenarios for moving sources/receivers. The computation of all the necessary RIRs can be done with low-complexity using the GFD method presented in [11]. A GDFT in the temporal dimension can be applied to reduce aliasing corruption as explained above. The novel GFD framework presented in this paper can then be used to perform fast convolution for auralization or other signal processing tasks in the GFD.

5.5 Concluding remarks

In this work, a generalized Fourier domain (GFD) signal processing framework is introduced. The proposed framework allows a special form of control on the periodic repetitions that occur due to sampling in the reciprocal domain. We show that this property can be expressed in terms of a *weighted periodic extension* of a signal. We demonstrate that the (discrete-time) generalized Fourier transform can be seen as a special case of the z-transform, and relate the analytic continuation of the standard Fourier spectrum to the generalized Fourier spectrum. Core properties of the generalized discrete Fourier transform (GDFT) are given. These allow to concisely work in the GFD. The close relationship of the GDFT to the DFT allows a generalized fast Fourier transform (GFFT) to be directly obtained via the FFT.

The novel framework opens possibilities for signal processing applications where working on the GFD results in a computational or analytical advantage. As an example, we review a method for low-complexity simulation of room impulse responses (RIR) [10, 11] based on the GFD. The framework presented in this paper can then be used to perform e.g. auralization, adaptive filtering or other acoustic signal processing operations in the GFD.

MATLAB® code to implement the GDFT is available on-line for educational and non-profit purposes at the webpage of TuDelft SIPLAB (http://siplab.tudelft.nl).

Appendix 5.A Proof of property 5.2

Proof. Time-domain shift property.

For $n_0 \in \mathbb{Z}$, we have that

$$\tilde{x}_\alpha(n-n_0) = \alpha^p x((n-n_0))_N = \alpha^p x(n-n_0+pN),$$

where $p = -\lfloor (n-n_0)/N \rfloor$. Then,

$$\begin{aligned}\mathcal{G}_\alpha\left\{\tilde{x}_\alpha(n-n_0)\right\} &= \sum_{n=0}^{N-1} \tilde{x}_\alpha(n-n_0)\mathrm{e}^{\beta n}\mathrm{e}^{-j\frac{2\pi}{N}kn} \\ &= \sum_{n=0}^{N-1} \alpha^p x(n-n_0+pN)\mathrm{e}^{\beta n}\mathrm{e}^{-j\frac{2\pi}{N}kn},\end{aligned}$$

make $m = n - n_0 + pN$, then

$$\begin{aligned}\mathcal{G}_\alpha\left\{\tilde{x}_\alpha(n-n_0)\right\} &= \sum_{m=0}^{N-1} \alpha^p x(m)\mathrm{e}^{\beta(m+n_0-pN)}\mathrm{e}^{-j\frac{2\pi}{N}k(m+n_0-pN)} \\ &= \mathrm{e}^{\beta n_0}\mathrm{e}^{-j\frac{2\pi}{N}kn_0}\sum_{m=0}^{N-1} x(m)\mathrm{e}^{\beta m}\mathrm{e}^{-j\frac{2\pi}{N}km} \\ &= \mathrm{e}^{\beta n_0}\mathrm{e}^{-j\frac{2\pi}{N}kn_0}X_\alpha(k).\end{aligned}$$

□

Appendix 5.B Proof of property 5.3

Proof. GFD shift property.

For $k_0 \in \mathbb{Z}$,

$$\begin{aligned}
\mathcal{G}_\alpha\left\{x(n)\mathrm{e}^{j\frac{2\pi}{N}k_0 n}\right\} &= \sum_{n=0}^{N-1} x(n)\mathrm{e}^{j\frac{2\pi}{N}k_0 n}\mathrm{e}^{\beta n}\mathrm{e}^{-j\left(\frac{2\pi}{N}kn\right)} \\
&= \sum_{n=0}^{N-1} x(n)\mathrm{e}^{\beta n}\mathrm{e}^{-j\left(\frac{2\pi}{N}n(k-k_0)\right)} \\
&= \sum_{n=0}^{N-1} x(n)\mathrm{e}^{\beta n}\mathrm{e}^{-j\left(\frac{2\pi}{N}n((k-k_0))_N\right)} \\
&= X_\alpha((k-k_0))_N = \tilde{X}_\alpha\left(k-k_0\right).
\end{aligned}$$

□

Appendix 5.C Proof of property 5.4

Proof. Time reversal.

For

$$\tilde{x}_\alpha(-n) = \begin{cases} x(n) & \text{for} \quad n=0 \\ \alpha x(N-n) & \text{for} \quad n=\{1,\cdots,N-1\}, \end{cases}$$

we have that,

$$\mathcal{G}_{\alpha^{-1}}\left\{\tilde{x}_\alpha(-n)\right\} = x(0) + \sum_{n=1}^{N-1}\alpha x(N-n)\mathrm{e}^{-\beta n}\mathrm{e}^{-j\frac{2\pi}{N}kn},$$

set $m = N - n$, then,

$$\begin{aligned}
\mathcal{G}_{\alpha^{-1}}\left\{\tilde{x}_\alpha(-n)\right\} &= x(0) + \sum_{m=1}^{N-1}\alpha x(m)\mathrm{e}^{-\beta(N-m)}\mathrm{e}^{-j\frac{2\pi}{N}k(N-m)} \\
&= x(0) + \sum_{m=1}^{N-1} x(m)\mathrm{e}^{\beta m}\mathrm{e}^{-j\frac{2\pi}{N}(N-k)m} \\
&= \sum_{m=0}^{N-1} x(m)\mathrm{e}^{\beta m}\mathrm{e}^{-j\frac{2\pi}{N}(N-k)m} \\
&= X_\alpha(N-k) = \tilde{X}_\alpha(-k).
\end{aligned}$$

since $\beta = \log(\alpha)/N$. □

Appendix 5.D Proof of property 5.5

Proof. Time domain complex-conjugate.

$$\begin{aligned}
\mathcal{G}_{\alpha^*}\left\{x^*(n)\right\} &= \sum_{n=0}^{N-1} x^*(n)\mathrm{e}^{\beta^* n}\mathrm{e}^{-j\frac{2\pi}{N}kn} \\
&= \left(\sum_{n=0}^{N-1} x(n)\mathrm{e}^{\beta n}\mathrm{e}^{j\frac{2\pi}{N}kn}\right)^* \\
&= \left(\sum_{n=0}^{N-1} x(n)\mathrm{e}^{\beta n}\mathrm{e}^{-j\frac{2\pi}{N}(N-k)n}\right)^* \\
&= \tilde{X}^*_{\alpha}(-k).
\end{aligned}$$

□

Appendix 5.E Proof of property 5.6

Proof. GFD complex-conjugate.

$$\begin{aligned}
\mathcal{G}^{-1}_{(\alpha^*)^{-1}}\left\{X^*_{\alpha}(k)\right\} &= \frac{\mathrm{e}^{\beta^* n}}{N}\sum_{k=0}^{N-1} X^*_{\alpha}(k)\mathrm{e}^{j\frac{2\pi}{N}kn} \\
&= \frac{\mathrm{e}^{\beta^* n}}{N}\left(\sum_{k=0}^{N-1} X_{\alpha}(k)\mathrm{e}^{-j\frac{2\pi}{N}kn}\right)^* \\
&= \begin{cases} x^*(0) & , \quad n=0 \\ \mathrm{e}^{\beta^* n}x^*(N-n)\mathrm{e}^{\beta^*(N-n)} & , \quad \text{otherwise} \end{cases} \\
&= \begin{cases} x^*(0) & , \quad n=0 \\ \big(\alpha x(N-n)\big)^* & , \quad \text{otherwise} \end{cases} \\
&= \tilde{x}^*_{\alpha}(-n).
\end{aligned}$$

since $\beta = \log(\alpha)/N$. □

5.8 Bibliography

[1] J. Martinez, R. Heusdens, and R. C. Hendriks, "A generalized Poisson summation formula and its application to fast linear convolution," *IEEE Signal Process. Lett.*, vol. 18, no. 9, pp. 501 –504, Sept. 2011.

[2] R. Bracewell, *The Fourier Transform & Its Applications*, 3rd ed. New York: McGraw-Hill, June 1999.

[3] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms and Applications*, 4th ed. New York: Prentice-Hall, 2007.

[4] H. Kuttruff, *Room Acoustics*. London, U.K.: Taylor and Francis, 2000.

[5] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing (Signals and Communication Technology)*. Secaucus, NJ: Springer-Verlag, 2006.

[6] B. Xu and S. D. Sommerfeldt, "A hybrid modal analysis for enclosed sound fields," *J. Acoust. Soc. Amer.*, vol. 128, no. 5, pp. 2857–2867, Nov. 2010.

[7] J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[8] A. Berkhout, D. de Vries, J. Baan, and B. van den Oetelaar., "A wave field extrapolation approach to acoustical modeling in enclosed spaces." *J. Acoust. Soc. Amer.*, vol. 105, pp. 1725–1733, Mar. 1999.

[9] S. R. Bistafa and J. W. Morrissey, "Numerical solutions of the acoustic eigenvalue equation in the rectangular room with arbitrary (uniform) wall impedances," *J. Sound Vib.*, vol. 263, no. 1, pp. 205 – 218, May 2003.

[10] J. Martinez and R. Heusdens, "On low-complexity simulation of multichannel room impulse responses," *IEEE Signal Process. Lett.*, vol. 17, no. 7, pp. 667 –670, Jul. 2010.

[11] J. Martinez, R. Heusdens, and R. C. Hendriks, "A spatio-temporal generalized Fourier domain framework to acoustical modeling in enclosed spaces," in *Proc. ICASSP*, Mar. 2012, pp. 529–532.

[12] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling." *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

6

Fast modeling of multichannel room impulse responses

Jorge Martínez, Richard Heusdens and Richard C. Hendriks

Abstract

In this paper a fast method to model multichannel room impulse responses (RIRs) is presented. The method exploits a spectral sampling condition in a generalized Fourier domain (GFD). The Poisson summation formula associated with the generalized Fourier transform relates a geometrically weighted summation of a signal over a lattice to the samples of its generalized spectrum over the reciprocal lattice. For box-shaped rooms with constant wall absorption coefficients we show that the spatial aliasing introduced by spectral sampling in the GFD represents the wall reflections. The approach is proven to be very fast, of order $\mathcal{O}(N \log N)$ on the reverberation time N per receiver position, when a close form expression for the generalized spectrum is known.

6.1 Introduction

The room impulse response (RIR) plays an important role, directly or indirectly, on the design and implementation of human telecommunication technologies. In several acoustic and speech signal processing applications an estimation, approximation or consideration of the RIR is necessary, examples of these are acoustic echo cancellation (AEC), derreverberation, blind source separation and microphone beamforming [1–3]. Moreover, with the advent of 3-D sound field control technologies like wave field synthesis (WFS) [4], higher order ambisonics (HOA) [5], and more recently sound field reconstruction (SFR) [6], a consideration of the RIRs between massive amounts of transmitters and receivers is needed to allow full-duplex telecommunication using these technologies.

State-of-the-art RIR modeling methods provide efficient multichannel solutions [7–11]. However, the total algorithmic complexity of RIR modeling methods is still an issue, specially when considering real-time systems, large-scale (MMIMO) microphone-loudspeaker networks, or in applications that call for full-bandwidth 3-D sound field modeling.

In this work we propose an efficient method to compute the sound field in a room (i.e. the spatio-temporal RIR). To present our approach we have organized the paper as follows. In Sec. 6.2 and Sec. 6.3 we introduce preliminary concepts behind the basic idea of our method. These concepts have been presented in our previous work [12] and are outlined next. We note that the construction of virtual sources in the the mirror image source model (MISM) [13] for box-shaped rooms has spatial periodicity. Assume a room could have perfectly reflective walls (i.e. no absorption), then the sound field generated by all the virtual sources can be represented by a periodic summation. We then recall the key fact that “sampling of a function results in a periodic summation of its Fourier transform” and construct a continuous spectral kernel function using the spatio-temporal Fourier representation of the free-field Green’s function. We carefully sample this spectral kernel function to induce the desired spatial periodicity in the reciprocal domain. Fourier synthesis of the coefficients obtained by sampling is implemented using the FFT resulting in a very efficient method to compute the full sound field in a room.

In practice, however, a room never has perfectly reflective (rigid) walls. In Sec. 6.4 we show that when absorptive boundaries are included in the model, the sound field can be expressed as a geometrically weighted periodic summation. In Sec. 6.5 we present a generalized Fourier domain (GFD), first introduced in our previous work [14]. We show that in this domain “sampling of a function results in a geometrically weighted periodic summation of its generalized Fourier transform (GFT)”. Besides the standard Fourier conditions, an extra sufficient condition for the GFT of a function to exist is that the function to be “single-sided” (causal or anti-causal), or when the domain is extended to Euclidean space to be “single-orthant”, therefore we cannot directly apply the same scheme as in the case of rigid walls to extend the method. In Sec. 6.6 the main contribution of the paper is given. The algorithm is outlined as follows. First, we factor the free-field Green’s function into its spatial “single-orthant” parts, and calculate the continuous spatio-temporal GFD representation of each part. Next, by sampling these continuous kernel functions we induce a geometrically weighted periodic summation in the reciprocal domain, equivalent to the reverberation effect in a room with absorptive walls. Finally, after generalized Fourier synthesis of the coefficients obtained by sampling, a representation of each of the single-orthant parts of the reverberated sound field is obtained. The full sound-field in the room is recovered by summing the parts. A fast discrete implementation of the GFT is obtained via the FFT [14], so that the generalized Fourier synthesis required is efficiently computed. In Sec. 6.7 the results of our experiments are given. We compare our method with the MISM in terms of performance and accuracy. Finally in Sec. 6.8 conclusions are given.

6.2 Preliminaries

A spatio-temporal characterization of the sound field is given by the plenacoustic function (PAF) [15]. Let an omni-directional point source (monopole) S, be placed in space at $\mathbf{s} \in \mathbb{R}^3$ and emit a signal $s(t)$. The PAF in this case is the sound field as it is registered at position $\mathbf{x} = [x, y, z]^T$, and at time t, and it is given by $p(\mathbf{x}, t) = (h_\mathbf{s} * s)(\mathbf{x}, t)$, where $*$ represents the (temporal) convolution operator. The function $h_\mathbf{s}(\mathbf{x}, t)$ is the spatio-temporal room impulse response (RIR) from point $\mathbf{s}$ to point $\mathbf{x}$, and therefore, it satisfies the inhomogeneous wave equation with given boundary conditions. Under this linear model, multiple sound sources are considered as a superposition of single sources. Assume now that the monopole at position $\mathbf{s}$ is in free-field, emitting a Dirac pulse at $t = 0$. The PAF becomes the RIR in this case, equivalent to the free-field Green's function given by [16],

$$p(\mathbf{x}, t) = \frac{\delta\left(t - \frac{r}{c}\right)}{4\pi r} \tag{6.1}$$

where δ is Dirac's delta function, $r = \|\mathbf{x} - \mathbf{s}\|$, is the distance from measurement point $\mathbf{x}$ to the source position $\mathbf{s}$, and c represents the velocity of sound propagation. The spatio-temporal (4-D) Fourier representation of the free-field Green's function, i.e. $P(\boldsymbol{\phi}, \omega) = \mathcal{F}\{p(\mathbf{x}, t)\}$, is given by [15],

$$P(\boldsymbol{\phi}, \omega) = \frac{\mathrm{e}^{j\boldsymbol{\phi}^T\mathbf{s}}}{\|\boldsymbol{\phi}\|^2 - \left(\frac{\omega}{c}\right)^2} \tag{6.2}$$

where $\boldsymbol{\phi} \in \mathbb{R}^3$ is the spatial frequency variable and ω denotes the temporal frequency variable. Conceptually the free-field Green's function (6.1) represents the impulsive spherical sound wave produced by the point source S, that propagates and decays forever. Inside a room with at least partially reflecting walls however, the sound field produced by the same point source is far more complicated. We analyze next the case of box-shaped rooms with perfectly reflective (rigid) walls [12].

6.3 RIR modeling for an enclosure with rigid walls

We analyze the sound field in box-shaped rooms with rigid walls using the mirror image source method (MISM), introduced by Allen and Berkey [13], as a starting theoretical framework. The MISM models the sound field inside the room by the creation of virtual free-field sources (outside the room), which represent the reflections introduced by the walls.

A 2-D example of such a constellation of virtual sources is shown in Fig. 6.1. The solid-line rectangle represents the actual room with dimensions L_x and L_y. The source S_0 denotes the direct-path contribution of S. The rest of the labeled sources $S_l,\ l = 1, \ldots, 3$, denote the contributions due to reflections on two orthogonal walls.

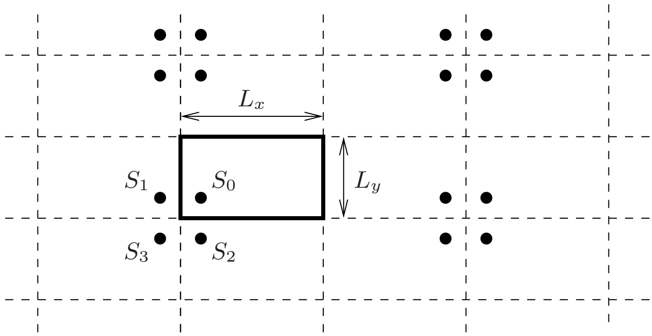


Figure 6.1: Example of a 2-D constellation of virtual sources created by the MISM [13].

A careful examination reveals that the full set of virtual sources is obtained by a periodic repetition of these fundamental sources. This set of fundamental sources is referred to as the set of *mother sources* [15]. In general there are 2^ν mother sources in the ν-D case [12].

The set of virtual sources is thus obtained by a ν-D periodic packing of the mother sources over a lattice [12], say Λ, defined as $\Lambda \triangleq \{\boldsymbol{\lambda} \in \mathbb{R}^\nu : \boldsymbol{\lambda} = \boldsymbol{\Lambda}\mathbf{n}, \mathbf{n} \in \mathbb{Z}^\nu\}$, where $\boldsymbol{\Lambda}$ is a non-singular matrix called the generator matrix of Λ. The columns of $\boldsymbol{\Lambda}$ are the basis vectors of the lattice. In a box-shaped room with dimensions L_x, L_y, L_z, with all of its walls being at least partially reflective, the periodic packing is 3-D and $\boldsymbol{\Lambda}$ takes the form [12],

$$\boldsymbol{\Lambda} = \mathrm{diag}(2L_x, 2L_y, 2L_z). \tag{6.3}$$

The sound field at any position $\mathbf{x}$ is thus given by a *periodic summation* [12],

$$p(\mathbf{x}, t) = \sum_{\mathbf{n} \in \mathbb{Z}^3} p_{\mathrm{m}}(\mathbf{x} + \boldsymbol{\Lambda}\mathbf{n}, t) \tag{6.4}$$

where p_{m} is the sound field generated by the set of mother sources $S_l,\ l = 0, \ldots, 7$, i.e.,

$$p_{\mathrm{m}}(\mathbf{x}, t) = \sum_{l=0}^{7} p_l(\mathbf{x}, t) = \sum_{l=0}^{7} \frac{\delta\left(t - \frac{r_l}{c}\right)}{4\pi r_l} \tag{6.5}$$

with $r_l = \|\mathbf{x} - \mathbf{s}_l\|$, where $\mathbf{s}_l \in \mathbb{R}^3$ is the position in space of source S_l.

Using this model, a direct computation of the reverberated sound field involves the contribution of infinite many virtual sources. In practice, a finite number, say N,

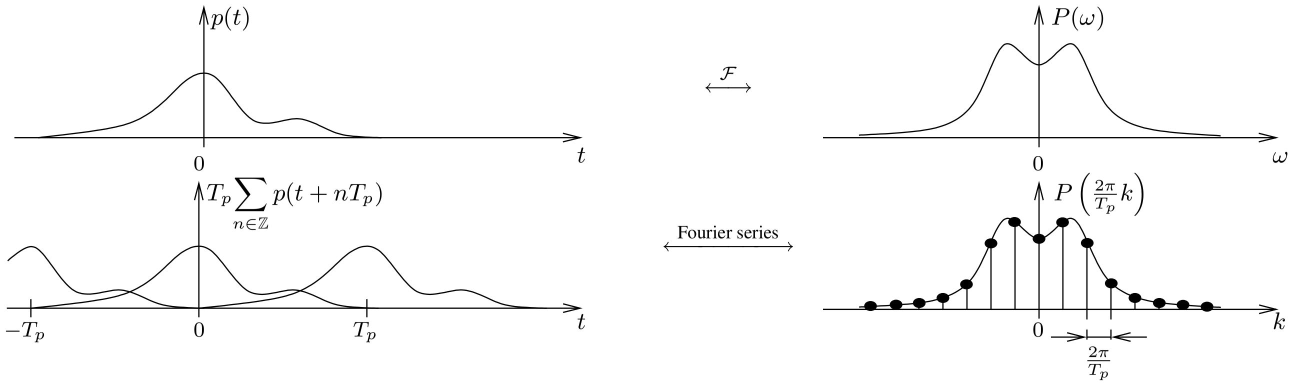


Figure 6.2: Poisson summation formula in 1-D. Sampling of the spectrum with period $2\pi/T_p$, results in a periodic summation with period T_p, of the time domain function. $\mathcal{F}$ denotes the Fourier transform operator.

of repetitions per dimension is taken into account. This number is proportional to the desired reverberation time [13, 17]. Therefore, an algorithm to compute the sound field using (6.4) directly is of complexity $\mathcal{O}(N^3)$ [12, 13, 17].

To derive a more efficient algorithm a key observation is given [12], i.e. *Sampling of a function results in a periodic summation of its Fourier transform.* The relation between the samples of the continuous Fourier transform of a function and its periodic summation is given by *Poisson's summation formula* [18–20]. In 1-D the formula takes the following form,

$$\sum_{n\in\mathbb{Z}} p(t+T_p n) = \frac{1}{T_p}\sum_{k\in\mathbb{Z}} P\left(\frac{2\pi}{T_p}k\right)\mathrm{e}^{j(2\pi/T_p)kt}$$

where $P(\omega)$ is the (continuous) Fourier transform of $p(t)$, T_p is the periodicity interval in the time-domain, and $2\pi/T_p$ is the sampling interval in the frequency domain. Clearly we have that $P(2\pi k/T_p) = T_p c_k$, where c_k are the Fourier series coefficients of the periodic function $\sum_{n\in\mathbb{Z}} p(t+T_p n)$. This construction is depicted in Fig. 6.2.

Consider now the Fourier representation of the sound field generated by the set of mother sources $P_{\mathrm{m}}(\boldsymbol{\phi},\omega)$, which follows directly from (6.2) and (6.5),

$$P_{\mathrm{m}}(\boldsymbol{\phi},\omega) = \sum_{l=0}^{7} \frac{\mathrm{e}^{j\boldsymbol{\phi}^T\mathbf{s}_l}}{\|\boldsymbol{\phi}\|^2 - \left(\frac{\omega}{c}\right)^2}. \tag{6.6}$$

Using the Poisson summation formula, the main result in [12] states that the sound field in the room, given by the periodic summation in (6.4), can be obtained by the Fourier synthesis of the (spatial) samples of the spectrum given by (6.6), this is,

$$p(\mathbf{x},t) = (2\pi|\boldsymbol{\Lambda}|)^{-1}\sum_{\mathbf{k}\in\mathbb{Z}^3}\int_{\mathbb{R}} P_{\mathrm{m}}(\boldsymbol{\Phi}\mathbf{k},\omega)\mathrm{e}^{j(\mathbf{k}^T\boldsymbol{\Phi}^T\mathbf{x}+\omega t)}\mathrm{d}\omega \tag{6.7}$$

where $\boldsymbol{\Phi} = 2\pi\boldsymbol{\Lambda}^{-T}$, is the generator matrix of the spectral sampling lattice Φ and $|\boldsymbol{\Lambda}|$ is the absolute value of the determinant of $\boldsymbol{\Lambda}$. The calculus of the sound field (6.7) involves an integral and an infinite summation. Discretizing all quantities, and using the FFT to perform the Fourier synthesis, an efficient method to compute the sound field is obtained. In the next sections we extend this model to rooms with constant wall reflection coefficients.

6.4 The MISM for walls with constant reflection coefficients

If walls with constant reflection coefficients are allowed, the sound field $p(\mathbf{x}, t)$ no longer shows a periodic structure as in (6.4) since the absorption induced by the walls modifies the contribution of each virtual source (reflection) following a geometric law [13]. In this section we show that the sound field inside the enclosure can be expressed as a geometrically weighted periodic summation. Each of the walls that enclose the room space is allowed to have its own particular reflection coefficient. For the walls perpendicular to the ith coordinate, $i \in \{x, y, z\}$, let us denote by $\varrho_{i0} \in \mathbb{C}$, $|\varrho_{i0}| \leq 1$, the reflection coefficient of the wall adjacent to the origin, by $\varrho_{i1} \in \mathbb{C}$, $|\varrho_{i1}| \leq 1$, the reflection coefficient of the opposing wall, and make $\varrho_i = \varrho_{i0}\varrho_{i1}$. The sound field produced by a particular mother source S_l is given by $p_l(\mathbf{x}, t)$ (as defined in (6.5)).

To simplify the discussion we first analyze the MISM in 1-D. Let the walls perpendicular to the y and z axis be totally absorptive or equivalently nonexistent. The sound field is then measured on a line in the x direction. The reverberated sound field on the line can be written as a weighted periodic summation of waves generated by the mother sources, this is

$$p(x,t) = \sum_{n_x=0}^{\infty} \varrho_x^{n_x} \Big(p_0(x+(2L_x)n_x, t) + \varrho_{x0}\, p_1(x+(2L_x)n_x, t) \Big) + \sum_{n_x=-\infty}^{-1} \varrho_x^{-n_x} \Big(p_0(x+(2L_x)n_x, t) + \varrho_{x0}^{-1} p_1(x+(2L_x)n_x, t) \Big). \tag{6.8}$$

The sound field $p_0(x, t)$ of the real mother source S_0, is periodically extended and damped. The repetitions to the left of the room (for $n_x \geq 0$) get an attenuation equal to $\varrho_x^{n_x}$, and the ones to the right (for $n_x < 0$) an attenuation equal to $\varrho_x^{-n_x}$. For the mother source S_1 the repetitions to the left of the room (for $n_x \geq 0$) get an attenuation equal to $\varrho_{x0}\varrho_x^{n_x}$ and the repetitions to the right (for $n_x < 0$) an attenuation equal to $\varrho_{x0}^{-1}\varrho_x^{-n_x}$. We call to ϱ_{x0} and ϱ_{x0}^{-1}, the “alignment factors”. An example of this construction is given in Fig. 6.3, where s_{x_l} are the x components of the position $\mathbf{s}_l$, of the mother sources.

In 1-D the sound field is factored into two summations involving weighted repetitions of the mother sources to the left and to the right of the room. Generalizing to 3-D, the sound field is factored into summations that involve weighted repetitions in the direction of a given octant of the space. In order to proceed, we introduce the

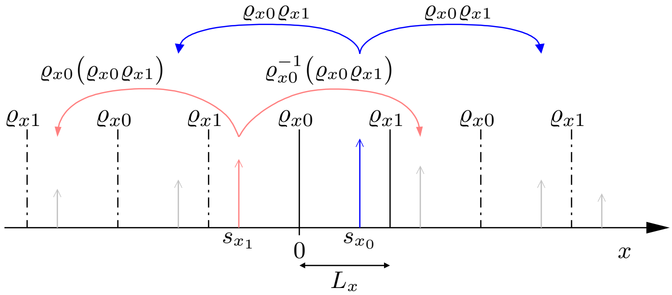


Figure 6.3: Example of the damped repetitions of the sound field created by the MISM in 1-D, to construct the reverberated sound field.

following definitions. Let O_q, $q = 0, \ldots, 7$, denote the qth octant of the 3-D space, and use a binary ordering scheme based on the signs of the coordinates to enumerate the octants. This is, if $(z \geq 0,\, y \geq 0,\, x \geq 0)$, or equivalently if the signs of the coordinates are $(+, +, +)$, then $q = 0$. If $(z \geq 0, y \geq 0, x < 0)$ or $(+, +, -)$, then $q = 1$, and so on. For example $\mathrm{O}_3 \triangleq \{\mathbf{x} \in \mathbb{R}^3 : z \geq 0, y < 0, x < 0\}$. Define $\varsigma_q(x) = \mathrm{sign}(x)$, $\varsigma_q(y) = \mathrm{sign}(y)$, and $\varsigma_q(z) = \mathrm{sign}(z)$ for $(x, y, z) \in \mathrm{O}_q$. Then for $i \in \{x, y, z\}$ each $\varsigma_q(i) = \pm 1$, depending on the signs of the coordinates defining the octant O_q. The sound field is obtained by extending (6.8) to 3-D space, this is,

$$p(\mathbf{x}, t) = \sum_{q=0}^{7} \sum_{l=0}^{7} \sum_{\mathbf{n} \in \mathbb{Z}^3 : \boldsymbol{\Lambda}\mathbf{n} \in \mathrm{O}_q} \varrho_{lq} \Big(\prod_i \varrho_i^{\varsigma_q(i) n_i} \Big) p_l(\mathbf{x} + \boldsymbol{\Lambda}\mathbf{n}, t) \tag{6.9}$$

where the product runs over $i \in \{x, y, z\}$, n_i represent the components of $\mathbf{n}$, the inner summation runs over those $\mathbf{n} \in \mathbb{Z}^3$ such that $\boldsymbol{\Lambda}\mathbf{n} \in \mathrm{O}_q$, and ϱ_{lq} are the corresponding alignment factors for each mother source S_l and each space octant O_q. These can be derived extending the 1-D analysis given above. In Table 6.1 a means to calculate the alignment factors is given.

6.5 A generalized Fourier domain

Given a signal $p(\mathbf{x}) \in L^2(\mathbb{R}^\nu)$, the Poisson summation formula relates the signal to the samples of its spectrum $P(\boldsymbol{\phi})$ [18, 19] as follows,

$$\sum_{\mathbf{n} \in \mathbb{Z}^\nu} p(\mathbf{x} + \boldsymbol{\Lambda}\mathbf{n}) = \frac{1}{|\boldsymbol{\Lambda}|} \sum_{\mathbf{k} \in \mathbb{Z}^\nu} P(\boldsymbol{\Phi}\mathbf{k}) \mathrm{e}^{j(\mathbf{k}^T \boldsymbol{\Phi}^T \mathbf{x})} \tag{6.10}$$

$l \backslash q$	0	1	2	3	4	5	6	7
	$A,\ B,\ C$	$A,\ B,\ C$	$A,\ B,\ C$	$A,\ B,\ C$	$A,\ B,\ C$	$A,\ B,\ C$	$A,\ B,\ C$	$A,\ B,\ C$
0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
1	1, 0, 0	−1, 0, 0	1, 0, 0	−1, 0, 0	1, 0, 0	−1, 0, 0	1, 0, 0	−1, 0, 0
2	0, 1, 0	0, 1, 0	0, −1, 0	0, −1, 0	0, 1, 0	0, 1, 0	0, −1, 0	0, −1, 0
3	1, 1, 0	−1, 1, 0	1, −1, 0	−1, −1, 0	1, 1, 0	−1, 1, 0	1, −1, 0	−1, −1, 0
4	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, −1	0, 0, −1	0, 0, −1	0, 0, −1
5	1, 0, 1	−1, 0, 1	1, 0, 1	−1, 0, 1	1, 0, −1	−1, 0, −1	1, 0, −1	−1, 0, −1
6	0, 1, 1	0, 1, 1	0, −1, 1	0, −1, 1	0, 1, −1	0, 1, −1	0, −1, −1	0, −1, −1
7	1, 1, 1	−1, 1, 1	1, −1, 1	−1, −1, 1	1, 1, −1	−1, 1, −1	1, −1, −1	−1, −1, −1

Table 6.1: Power coefficients required to calculate the alignment factors ϱ_{lq}. These are obtained by $\varrho_{lq} = \varrho_{x0}^{A} \varrho_{y0}^{B} \varrho_{z0}^{C}$.

where $\mathbf{\Lambda}$ is a generator matrix for the periodicity lattice Λ, and $\mathbf{\Phi} = 2\pi\mathbf{\Lambda}^{-T}$ is the generator matrix of the spectral sampling lattice Φ. As already mentioned in Sec. 6.3, this equation asserts that periodicity in one domain implies discretization in the reciprocal domain.

In [14] a *generalized* Poisson summation formula is introduced. The proposed equation associates the samples of a generalized spectrum of a signal, with a *geometrically weighted* periodic extension of the signal.

Let us define, for a signal $p(\mathbf{x}) \in L^2(\mathbb{R}^\nu)$, a generalized Fourier transform with parameter $\boldsymbol{\alpha} \in \mathbb{C}^\nu : \alpha_i \neq 0$, for all $i = 0, \ldots, \nu - 1$, as,

$$\mathcal{G}_{\boldsymbol{\alpha}}\left\{p(\mathbf{x})\right\} \triangleq P_{\boldsymbol{\alpha}}(\boldsymbol{\phi}) = \int_{\mathbb{R}^\nu} p(\mathbf{x}) \mathrm{e}^{\boldsymbol{\beta}^T\mathbf{x}} \mathrm{e}^{-j(\boldsymbol{\phi}^T\mathbf{x})} \mathrm{d}\mathbf{x} \tag{6.11}$$

where $\boldsymbol{\beta} = \mathbf{\Lambda}^{-T}\log(\boldsymbol{\alpha})$, $\mathbf{\Lambda}^{-T}$ is the generator matrix of the reciprocal lattice of Λ, and $\log(\boldsymbol{\alpha}) \triangleq [\log(\alpha_0), \ldots, \log(\alpha_{\nu-1})]^T$. The inverse transformation is given by,

$$\mathcal{G}_{\boldsymbol{\alpha}}^{-1}\left\{P_{\boldsymbol{\alpha}}(\boldsymbol{\phi})\right\} \triangleq p(\mathbf{x}) = \frac{\mathrm{e}^{-\boldsymbol{\beta}^T\mathbf{x}}}{(2\pi)^\nu} \int_{\mathbb{R}^\nu} P_{\boldsymbol{\alpha}}(\boldsymbol{\phi}) \mathrm{e}^{j(\mathbf{x}^T\boldsymbol{\phi})} \mathrm{d}\boldsymbol{\phi}. \tag{6.12}$$

The transform given in (6.11) is equivalent to the ordinary ν-dimensional Fourier transform (if it can be defined) of the modulated signal $p(\mathbf{x})\mathrm{e}^{\boldsymbol{\beta}^T\mathbf{x}}$.

Recall the definition of the ν-dimensional bilateral Laplace transform (BLT) [21],

$$P_{\text{BLT}}(\mathbf{s}) = \int_{\mathbb{R}^\nu} p(\mathbf{x}) \mathrm{e}^{-\mathbf{s}^T\mathbf{x}} \mathrm{d}\mathbf{x}. \tag{6.13}$$

where $\mathbf{s} \in \mathbb{C}^\nu$. The region of convergence (ROC) of the BLT depends on $p(\mathbf{x})$. Denote the components of parameter $\boldsymbol{\alpha}$ as $\alpha_i = |\alpha_i|\mathrm{e}^{j\angle\alpha_i}$ and compare (6.11) with (6.13). We find the GFT to be a particular case of the BLT with

$$\mathbf{s} = -\boldsymbol{\beta}_{\Re} + j(\boldsymbol{\beta}_{\Im} - \boldsymbol{\phi}),$$

where $\boldsymbol{\beta}_{\Re} = \mathbf{\Lambda}^{-T}[\log(|\alpha_0|), \ldots, \log(|\alpha_{\nu-1}|)]^T$ and $\boldsymbol{\beta}_{\Im} = \mathbf{\Lambda}^{-T}[\angle\alpha_0, \ldots, \angle\alpha_{\nu-1}]^T$. The GFT is equivalent to the BLT evaluated at a ν-dimensional linear manifold of the $\mathbf{s}$-space. Therefore if the BLT of the function has a ROC that includes the linear manifold at $-\boldsymbol{\beta}_{\Re}$, then the GFT for that particular value of $\boldsymbol{\alpha}$ exists. For $p(\mathbf{x}) \in L^2(\mathbb{R}^\nu)$ a function with support on a single orthant of $\mathbb{R}^\nu$ (i.e. a single-orthant function), the ROC of the BLT is also given by a single orthant region in $\mathbf{s}$-space and we can select a $\boldsymbol{\alpha}$ parameter such that the GFT is included in that region. Conversely we have that for $\boldsymbol{\beta} : |\Re\{\beta_i\}| > 0$, for a given i, the GFT is well defined on the subset of $L^2(\mathbb{R}^\nu)$ functions single-sided on that particular i-dimension [21]. Consistently, when $\boldsymbol{\alpha}$=$[1, 1, \ldots, 1]^T$, the transform pair (6.11) and (6.12) corresponds to the standard Fourier pair.

A generalized Poisson summation formula (GPSF) follows by evaluating (6.11) in (6.10), i.e.,

$$\sum_{\mathbf{n}\in\mathbb{Z}^\nu} \mathrm{e}^{\boldsymbol{\beta}^T\boldsymbol{\Lambda}\mathbf{n}} p(\mathbf{x}+\boldsymbol{\Lambda}\mathbf{n}) = \frac{\mathrm{e}^{-\boldsymbol{\beta}^T\mathbf{x}}}{|\boldsymbol{\Lambda}|} \sum_{\mathbf{k}\in\mathbb{Z}^\nu} P_{\boldsymbol{\alpha}}(\boldsymbol{\Phi}\mathbf{k}) \mathrm{e}^{j(\mathbf{k}^T\boldsymbol{\Phi}^T\mathbf{x})}. \tag{6.14}$$

Now $\boldsymbol{\beta}^T\boldsymbol{\Lambda}\mathbf{n}$=$\log(\boldsymbol{\alpha})^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}\mathbf{n}$=$\log(\boldsymbol{\alpha})^T\mathbf{n}$=$\sum_{i=0}^{\nu-1} n_i \log(\alpha_i)$,

$$\mathrm{e}^{\log(\boldsymbol{\alpha})^T\mathbf{n}} = \exp\left(\sum_{i=0}^{\nu-1} n_i \log(\alpha_i)\right) = \prod_{i=0}^{\nu-1} \alpha_i^{n_i},$$

so that (6.14) can be rewritten as,

$$\sum_{\mathbf{n}\in\mathbb{Z}^\nu} \Big(\prod_{i=0}^{\nu-1} \alpha_i^{n_i}\Big) p(\mathbf{x}+\boldsymbol{\Lambda}\mathbf{n}) = \frac{\mathrm{e}^{-\boldsymbol{\beta}^T\mathbf{x}}}{|\boldsymbol{\Lambda}|} \sum_{\mathbf{k}\in\mathbb{Z}^\nu} P_{\boldsymbol{\alpha}}(\boldsymbol{\Phi}\mathbf{k}) \mathrm{e}^{j(\mathbf{k}^T\boldsymbol{\Phi}^T\mathbf{x})}. \tag{6.15}$$

This equation relates a geometrically weighted periodic summation of a signal over a lattice Λ to the samples of its generalized spectrum over the 2π-scaled reciprocal lattice Φ.

Let us define $\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x}) \triangleq \sum_{\mathbf{n}\in\mathbb{Z}^\nu} \big(\prod_{i=0}^{\nu-1} \alpha_i^{n_i}\big) p(\mathbf{x}+\boldsymbol{\Lambda}\mathbf{n})$ as the geometrically weighted periodic extension of $p(\mathbf{x})$. We can regard $\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x})$ as a superposition of infinitely many translated "replicas" of $p(\mathbf{x})$ weighted by $\big(\prod_{i=0}^{\nu-1} \alpha_i^{n_i}\big)$.

In analogy to the Fourier series expansion of a periodic signal [18, 21, 22], the definition of the generalized Fourier series expansion of a geometrically weighted periodic signal of the form $\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x})$, follows. First we define a Voronoi region around a point of lattice Λ as,

$$V_\Lambda(\mathbf{n}) \triangleq \{\mathbf{x} \in \mathbb{R}^\nu : \|\mathbf{x}-\boldsymbol{\Lambda}\mathbf{n}\| \le \|\mathbf{x}-\boldsymbol{\Lambda}\mathbf{n}'\|, \forall \mathbf{n}' \in \mathbb{Z}^\nu\}.$$

The Voronoi region around the origin $V_\Lambda(\mathbf{0})$, is called the central Voronoi region. If the restriction $\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x}) \in L^2(V_\Lambda(\mathbf{0}))$, is a square integrable function over the central Voronoi region of lattice Λ, then $\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x})$ for $\mathbf{x}\in\mathbb{R}^\nu$, can be represented by (6.16) as follows,

$$\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x}) = \mathrm{e}^{-\boldsymbol{\beta}^T\mathbf{x}} \sum_{\mathbf{k}\in\mathbb{Z}^\nu} c_{\boldsymbol{\alpha},\mathbf{k}} \mathrm{e}^{j(\mathbf{k}^T\boldsymbol{\Phi}^T\mathbf{x})}. \tag{6.16}$$

The generalized Fourier series coefficients $c_{\boldsymbol{\alpha},\mathbf{k}}$, are given by,

$$c_{\boldsymbol{\alpha},\mathbf{k}} = |\mathbf{\Lambda}|^{-1} \int_{V_\Lambda(\mathbf{0})} \tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x}) \mathrm{e}^{\boldsymbol{\beta}^T \mathbf{x}} \mathrm{e}^{-j(\mathbf{k}^T \mathbf{\Phi}^T \mathbf{x})} \mathrm{d}\mathbf{x}. \tag{6.17}$$

Comparing (6.16) and (6.15) we have that $P_{\boldsymbol{\alpha}}(\mathbf{\Phi k}) = |\mathbf{\Lambda}| c_{\boldsymbol{\alpha},\mathbf{k}}$. The existence of both $P_{\boldsymbol{\alpha}}(\boldsymbol{\phi})$ and $c_{\boldsymbol{\alpha},\mathbf{k}}$ guarantees convergence "almost everywhere" at both sides of (6.15).

6.6 Fast modeling of multichannel RIRs

6.6.1 Derivation

In this section we derive an efficient algorithm to compute the complete spatio-temporal sound field in a room given a source. In order to proceed, we first establish a relation between the samples of a function in the generalized Fourier domain, and the sound field in a room given by (6.9). As before, we begin the analysis in 1-D.

The 1-D generalized Poisson summation formula, relates the samples of the generalized spectrum of a function with a weighted periodic summation of the function. Thus, for $\alpha \in \mathbb{C} \backslash \{0\}$, we have that

$$\begin{aligned} \sum_{n \in \mathbb{Z}} \alpha^n p(t + T_p n) &= \frac{1}{T_p} \sum_{k \in \mathbb{Z}} P_\alpha \left(\frac{2\pi}{T_p} k \right) \mathrm{e}^{j(2\pi/T_p)kt} \\ &= \sum_{k \in \mathbb{Z}} c_{\alpha,k} \mathrm{e}^{j(2\pi/T_p)kt} \end{aligned}$$

with $P_\alpha(\omega) = T_p c_{\alpha,k}$. As it is seen from Fig. 6.4, depending on the absolute value of α, the repetitions show an exponentially growing behavior to the left, or to the right of the support of $p(t)$. Consider sampling the generalized spectrum with parameter $\alpha = \varrho_x = \varrho_{x0}\varrho_{x1}$ of $p_0(x,t)$, the sound field of mother source S_0. The result of the generalized Fourier synthesis is depicted in Fig. 6.5 (in blue) on top of the damped effect as modeled by the MISM (in gray). In this case, the weighted repetitions do not correspond to the desired effect at one side of the x direction.

In facts, for a parameter $\alpha : |\alpha| \neq 1$, the spatio-temporal generalized spectrum of $p_0(x,t)$ cannot be defined since the the sound field expands in all directions as a function of t, hence it is not single-orthant. The problem is solved if we factor the sound field into its positive-side and negative-side parts and then process these parts separately. This is, if we separate the waves generated by the mother sources

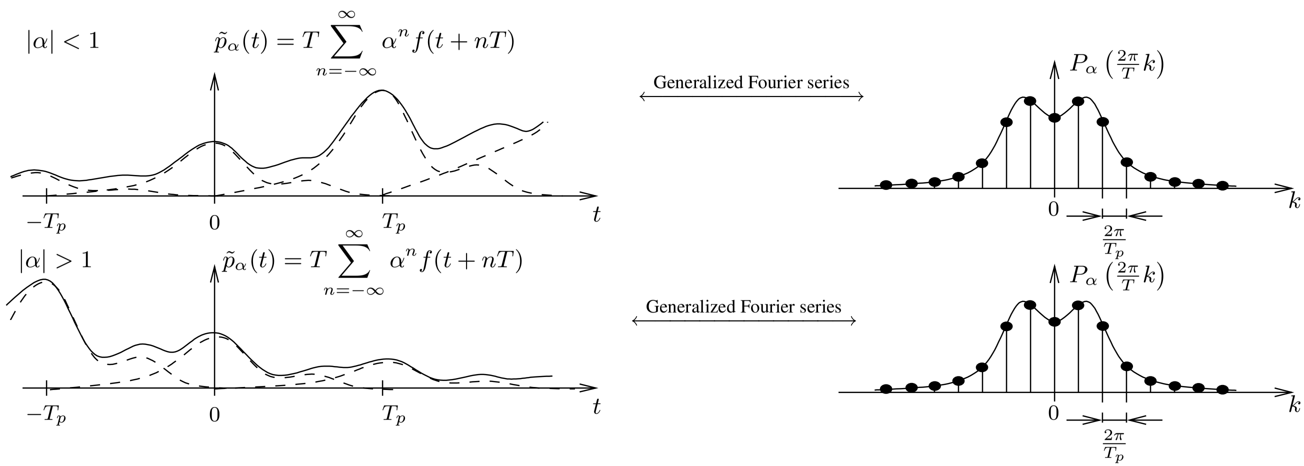


Figure 6.4: Generalized Poisson summation formula in 1-D. The growing behavior of the geometrically weighted repetitions, depends on the value of the parameter α.

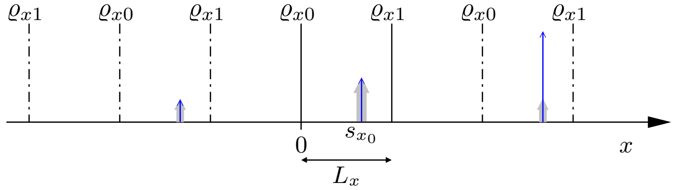


Figure 6.5: Generalized Fourier synthesis (in blue) of the samples of the GFT with parameter $\alpha = \varrho_x$, of the sound field of mother source S_0. Damped repetitions of S_0 as modeled by the MISM (in gray).

into waves traveling in positive and negative directions only. We thus define (in this case for $l = 0, 1$), $p_{l_+}(x,t) = p_l(x,t)H(x - s_{x_l})$ as the positive-side part, and $p_{l_-}(x,t) = p_l(x,t)H(-x + s_{x_l})$, as the negative-side part of the sound field of the mother sources. Here $H(x)$ is the 1-D Heaviside function (unit step function). The subscripts $+$ and $-$ in the notation are used to denote the (only) traveling direction of the resulting sound field. Then from (6.8), we have that the reverberated sound field in the zone $V_\Lambda(0) = \{-L_x \leq x \leq L_x\}$, can be expressed as,

$$p(x,t) = \sum_{n_x=0}^{\infty} \varrho_x^{n_x}\Big(p_{0_+}(x+(2L_x)n_x,t) + \varrho_{x0}\, p_{1_+}(x+(2L_x)n_x,t)\Big) + \sum_{n_x=-\infty}^{0} \varrho_x^{-n_x}\Big(p_{0_-}(x+(2L_x)n_x,t) + \varrho_{x0}^{-1} p_{1_-}(x+(2L_x)n_x,t)\Big). \quad (6.18)$$

Compare this equation with (6.8) and note that the second summation runs up to 0 (instead of up to -1). This is because at $n_x = 0$, $p_{l_+}(x,t) + p_{l_-}(x,t) = p_l(x,t)$, per definition of the Heaviside functions. Conceptually, we have that the repetitions to the left of $V_\Lambda(0)$ are waves traveling to the right (positive direction) and the repetitions to the right are waves traveling to the left (negative direction). The correct sound field inside $V_\Lambda(0)$ (an thus inside the room) is obtained for all t.

Define $\varrho_{0_+} = \varrho_{0_-} = 1$, and $\varrho_{1_+} = \varrho_{x0}$, $\varrho_{1_-} = \varrho_{x0}^{-1}$. Then we can rewrite (6.18) as,

$$p(x,t) = \sum_{l=0}^{1} \sum_{q\in\{+,-\}} \sum_{n_x\in\mathbb{Z}} \varrho_x^{(q)n_x}\, \varrho_{lq}\, p_{lq}(x+(2L_x)n_x,t), \; x \in V_\Lambda(0). \quad (6.19)$$

Besides the change in notation, the difference between (6.19) and the previous equation (6.18), is that we allow the summations over n_x to run over all integers. Conceptually we are now allowing exponentially growing repetitions to appear outside

$V_\Lambda(0)$ (like the example in Fig. 6.5), however these never form part of the sound field in the room, since by construction they travel in a direction away from the room. To prove (6.19), notice that in the zone $|x - s_{x_l}| \leq 2L_x$, or equivalently in $V_\Lambda(0) = \{|x| \leq L_x\}$, we have that,

$$p_{lq}(x + 2L_x n_x, t) = p_l(x + 2L_x n_x, t) H(x - s_{x_l} + 2L_x n_x) = 0 \tag{6.20}$$

for $n_x \leq 0$ and $q = +$, and for $n_x > 0$ and $q = -$, per definition of the Heaviside functions. Therefore, (6.19) equals (6.18) and (6.8) for $x \in V_\Lambda(0)$. The reverberated sound field is thus obtained by adding up the contributions of four geometrically weighted periodic summations of the sound fields $p_{lq}(x, t)$. In the ν-D spatial case, $2^{2\nu}$ contributions are necessary.

Generalizing to 3-D space, we consider waves traveling exclusively in the direction of a given octant of the space. To proceed with the analysis let $H_q(\mathbf{x})$ denote the 3-D Heaviside function given by,

$$H_q(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \mathbf{O}_q \\ 0, & \text{otherwise} \end{cases},$$

then we define,

$$p_{lq}(\mathbf{x}, t) = p_l(\mathbf{x}, t) H_q(\mathbf{x} - \mathbf{s}_l), \tag{6.21}$$

this is, the single-octant parts of the sound field of the mother sources. For a particular S_l, the functions $p_{lq}(\mathbf{x}, t)$ represent each of the eight parts of the sound field $p_l(\mathbf{x}, t)$, generated by the l-th mother source, traveling exclusively in the direction of a given octant.

Recapitulating from Sec. 6.4 and given the analysis above the reverberated sound field can be then expressed as,

$$p(\mathbf{x}, t) = \sum_{l=0}^{7} \sum_{q=0}^{7} \varrho_{lq} \sum_{\mathbf{n} \in \mathbb{Z}^3} \left(\prod_i \varrho_i^{\varsigma_q(i) n_i} \right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t), \quad \mathbf{x} \in V_\Lambda(\mathbf{0}) \tag{6.22}$$

for $i \in \{x, y, z\}$. In this case the alignment factors ϱ_{lq} are, as in Sec. 6.4, obtained from Table. 6.1. Referring to (6.20), we have that an equivalent situation occurs in the y and z directions, so that (6.22) equals (6.9) for $\mathbf{x} \in V_\Lambda(\mathbf{0})$ (which includes the room space).

The following important result relates the samples of the 4-D generalized spectra of the functions $p_{lq}(\mathbf{x}, t)$, to the geometrically weighted periodic summation required to construct the sound field in a room.

Proposition 6.1. *Let* $\mathbf{\Lambda}$ *be the generator matrix of the lattice specifying the spatial periodic packing of the sound fields* $p_{lq}(\mathbf{x}, t)$*, and let* $\mathbf{\Phi}$ *be the matrix basis of the lattice specifying the sampling points of the generalized spectra. Then the functions* $P_{\boldsymbol{\alpha} lq}(\mathbf{\Phi k}, \omega)$*,* $\mathbf{k} \in \mathbb{Z}^3$*, are the (spatial) generalized Fourier series expansions of*

$$|\mathbf{\Lambda}| \sum_{\mathbf{n} \in \mathbb{Z}^3} \left(\prod_i \varrho_i^{\varsigma_q(i) n_i} \right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t) \tag{6.23}$$

if and only if $\mathbf{\Phi} = 2\pi \mathbf{\Lambda}^{-T}$*, and* $\boldsymbol{\alpha} = [\varrho_x^{\varsigma_q(x)}, \varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}, \alpha_t]^T$*, with* $\alpha_t \in \mathbb{C} \backslash \{0\}$.

The proof is given in Appendix 6.A.

This result gives us a formula to reconstruct the full sound field in a room. Sample the spatial-generalized spectra $P_{\boldsymbol{\alpha} lq}(\boldsymbol{\phi}, \omega)$ using the sampling lattice generated by $\mathbf{\Phi} = 2\pi \mathbf{\Lambda}^{-T}$, apply an inverse generalized Fourier transform with parameter α_t over the temporal frequency variable, and use the generalized Poisson summation formula (6.15) on the coefficients just obtained to synthesize the functions,

$$\sum_{\mathbf{n} \in \mathbb{Z}^3} \left(\prod_i \varrho_i^{\varsigma_q(i) n_i} \right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t),$$

using these, the whole spatio-temporal sound field in the room is calculated as in (6.22). For a box-shaped room with dimensions L_x, L_y and L_z, $\mathbf{\Lambda}$ is given by (6.3). The generator matrix for the spectral sampling lattice is thus given by,

$$\mathbf{\Phi} = \text{diag}(\pi/L_x, \pi/L_y, \pi/L_z).$$

The continuous sound field given by (6.9), cannot be calculated on a digital computer. To implement the proposed algorithm we sample the temporal frequency ω as well. This will introduce undesired temporal aliasing, since $p(\mathbf{x}, t)$ is clearly not time limited (it has infinite support). Let $\mathbf{\Psi}$ denote the generator matrix for the lattice specifying the sampling points of both the spatial and temporal frequency variables, defined by, $\mathbf{\Psi} = \text{diag}(\mathbf{\Phi}, \Omega_s)$, where Ω_s denotes the temporal frequency sampling interval. The diagonal form of matrix $\mathbf{\Psi}$ implies an independent sampling of spatial and temporal frequencies. Following the same arguments and the value of $\boldsymbol{\alpha}$ as used in Proposition 6.1, it follows that the sampled spectra $P_{\boldsymbol{\alpha} lq}(\mathbf{\Psi k})$, $\mathbf{k} \in \mathbb{Z}^4$ yields the generalized Fourier series expansions of

$$|\mathbf{\Delta}| \sum_{\mathbf{n} \in \mathbb{Z}^3} \sum_{n \in \mathbb{Z}} (\alpha_t^n) \left(\prod_i \varrho_i^{\varsigma_q(i) n_i} \right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t + T_p n)$$

if and only if $\mathbf{\Psi} = 2\pi\mathbf{\Delta}^{-T}$, where $\mathbf{\Delta} = \mathrm{diag}(\mathbf{\Lambda}, T_p)$, and $T_p = 2\pi/\Omega_s$ is the interval of temporal periodicity. It follows that the evaluation of $P_{\boldsymbol{\alpha}lq}(\mathbf{\Psi k})$ in the (4-D) generalized Poisson summation formula (6.15) yields,

$$\begin{aligned}\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{x},t) &= \sum_{\mathbf{n}\in\mathbb{Z}^3}\sum_{n\in\mathbb{Z}}(\alpha_t^n)\left(\prod_i \varrho_i^{\varsigma_q(i)n_i}\right)p_{lq}(\mathbf{x}+\mathbf{\Lambda n}, t+T_p n)\\ &= \sum_{\mathbf{n}\in\mathbb{Z}^3}\left(\prod_i \varrho_i^{\varsigma_q(i)n_i}\right)p_{lq}(\mathbf{x}+\mathbf{\Lambda n}, t)\\ &\quad+\sum_{\substack{n\in\mathbb{Z}\\ n\neq 0}}\sum_{\mathbf{n}\in\mathbb{Z}^3}(\alpha_t^n)\left(\prod_i \varrho_i^{\varsigma_q(i)n_i}\right)p_{lq}(\mathbf{x}+\mathbf{\Lambda n}, t+T_p n)\end{aligned}$$

where the last summation term represents the time-domain aliasing introduced by sampling the temporal frequency ω. For each point in space $\mathbf{x}$, the RIR is a causal function of time, so that the repeated terms to the right of the temporal support of the RIR (for $n < 0$), do not contribute to the time-domain aliasing. Although the sound field $p(\mathbf{x},t)$ given by (6.9) has infinite temporal support, we have that $\lim_{t\to\infty} p(\mathbf{x},t) = 0$ for $\mathbf{x} \in V_\Lambda(\mathbf{0})$. Moreover we are interested in obtaining a good approximation of the sound field over the finite time span given by the reverberation time T_{60}. We have then two different ways to make the error due to time-domain aliasing negligible. The first approach is to make T_p sufficiently large (i.e. making Ω_s sufficiently small). The second approach is to take advantage of the generalized Poisson summation formula, making $\alpha_t << 1$ as small as possible (in practice this is principally limited by the non-causality of the bandlimited sound-field [23]). Recalling (6.9), the total sound field can thus be approximated as,

$$\begin{aligned}p(\mathbf{x},t) &\approx \sum_{l=0}^{7}\sum_{q=0}^{7}\varrho_{lq}\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{x},t)\\ &= \sum_{l=0}^{7}\sum_{q=0}^{7}\frac{\mathrm{e}^{-\boldsymbol{\beta}_q^T\mathbf{x}-\beta_t t}}{|\mathbf{\Delta}|}\sum_{\mathbf{k}\in\mathbb{Z}^4}\varrho_{lq}P_{\boldsymbol{\alpha}lq}(\mathbf{\Psi k})\mathrm{e}^{j(\mathbf{k}^T\mathbf{\Psi}^T[\mathbf{x}^T,t]^T)}\end{aligned}\tag{6.24}$$

where $\boldsymbol{\beta}_q = \mathbf{\Lambda}^{-T}\log\left(\left[\varrho_x^{\varsigma_q(x)}, \varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}\right]^T\right)$ and $\beta_t = \log(\alpha_t)/T_p$. To be calculated on a computer the continuous sound field $p(\mathbf{x},t)$ is made discrete. This induces aliasing in the 4-D spatio-temporal frequency space that has to be taken into account. The sampling points of the sound field can be made arbitrarily dense at the expense of taking more generalized Fourier coefficients $P_{\boldsymbol{\alpha}lq}(\mathbf{\Psi k})$ at reconstruction. We formalize the above discussion. Let Σ be a sublattice of Ψ, denoting the lattice for generating the periodic packing of the frequency space. Next, let Γ denote the spatio-temporal sampling lattice imposed by making the generalized spectra periodic and assume $\Delta \subseteq \Gamma$. Then clearly we have that $\mathbf{\Gamma} = 2\pi \mathbf{\Sigma}^{-T}$. Using these, the (sampled) functions $\tilde{p}_{\boldsymbol{\alpha}lq}$ can be calculated by,

$$\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{\Gamma n}) = \frac{\mathrm{e}^{-\boldsymbol{\beta}^T \mathbf{\Gamma n}}}{\mathrm{N}(\Delta/\Gamma)} \sum_{\mathbf{k} \in V_\Sigma(\mathbf{0})} |\mathbf{\Gamma}|^{-1} P_{\boldsymbol{\alpha}lq}(\mathbf{\Psi k}) \mathrm{e}^{j(\mathbf{k}^T \mathbf{\Psi}^T \mathbf{\Gamma n})} \tag{6.25}$$

where $\boldsymbol{\beta} = \mathbf{\Delta}^{-T} \log(\boldsymbol{\alpha})$, and $\mathrm{N}(\Delta/\Gamma)$ is the number of lattice points of Γ that lie inside $V_\Delta(\mathbf{0})$, the central Voronoi region of Δ. From here we see that by making $V_\Sigma(\mathbf{0})$ larger (i.e. taking more frequency samples), the finer we sample the functions $\tilde{p}_{\boldsymbol{\alpha}lq}$ (and consequently the sound field in the room), because of the relation $\mathbf{\Gamma} = 2\pi \mathbf{\Sigma}^{-T}$. Further we have that $\mathrm{N}(\Delta/\Gamma) = \mathrm{N}(\Sigma/\Psi)$ [12], this is, the number of evaluation points, $\mathrm{N}(\Sigma/\Psi)$, of the generalized spectra equals the number of spatio-temporal samples, $\mathrm{N}(\Sigma/\Psi)$, of the functions $\tilde{p}_{\boldsymbol{\alpha}lq}$. Since the PAF is of a geometrically weighted periodic form, we only need to evaluate it in $V_\Delta(\mathbf{0})$, this is

$$p(\mathbf{\Gamma n}) \approx \sum_{l=0}^{7} \sum_{q=0}^{7} \varrho_{lq} \tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{\Gamma n}), \qquad \text{for } \mathbf{\Gamma n} \in V_\Delta(\mathbf{0}). \tag{6.26}$$

The computational complexity of this approach is given by the evaluation of (6.25), limited to $\{\mathbf{n} \in \mathbb{Z}^4 : \mathbf{\Gamma n} \in V_\Delta(\mathbf{0})\}$. On the other hand, it is important to take into account the support of $P_{\boldsymbol{\alpha}lq}(\boldsymbol{\phi}, \omega)$, so that the region $V_\Sigma(\mathbf{0})$ is made large enough to not introduce (too much) undesired spectral aliasing. The support of $P_{\boldsymbol{\alpha}lq}(\boldsymbol{\phi}, \omega)$ although not being of compact support it has much of its energy concentrated in the region $\|\boldsymbol{\phi}\| \leq |\omega/c|$. We can then bound $V_\Sigma(\mathbf{0})$ and determine $\mathbf{\Sigma}$ and $\mathrm{N}(\Sigma/\Psi)$, allowing us for a trade between speed and accuracy of reconstruction. A direct evaluation of (6.25) (and consequently of (6.26)) is of order $\mathcal{O}(N_\omega^8)$ (a 4-D summation per spatio-temporal point), with N_ω proportional to ω_b. The summation over $\mathbf{k}$ in (6.25) represents a discrete Fourier transform (DFT). If the fast Fourier transform (FFT) is used instead, the operation takes only $\mathcal{O}(N_\omega^4 \log N_\omega)$ operations for computing (6.25) in $\mathrm{N}(\Delta/\Gamma)$ spatio-temporal positions. Since $\mathrm{N}(\Delta/\Gamma) = \mathrm{N}(\Sigma/\Psi)$, the method is of complexity $\mathcal{O}(N_\omega \log N_\omega)$ per receiver position (although all positions get calculated at once). This is far more efficient than the MISM, which has complexity $\mathcal{O}(N_t^3)$ per receiver position, with N_t proportional to the reverberation time T_{60}.

6.6.2 On the calculus of the generalized spectra

To implement the proposed method, the generalized Fourier coefficients $P_{\boldsymbol{\alpha}lq}(\boldsymbol{\Psi}\mathbf{k})$ are required. To obtain the generalized spectra one can compute the integral (6.11) of the single-octant parts of the sound field (6.21). However, a direct computation of these integrals is proven to be very difficult due to the absence of spherical symmetry and a different approach is needed. The approach we take here is based on Titchmarsh's theorem [24, 25], which proves that all of the following statements hold. Consider the 1-D case and let the spectrum of a single-sided function be denoted by $P(\phi)\in L^2(\mathbb{R})$. Then

1. $P(\phi)$ admits analytic continuation into one half of the complex plane. If the function $p(x) = 0$ for $x < 0$ (is a right-sided function), then $P(\phi_\Re + j\phi_\Im)$ is analytic for $\phi_\Im < 0$, and if $p(x)=0$ for $x > 0$ (is a left-sided function) then the spectrum is analytic for $\phi_\Im > 0$. In both cases we have that $\lim_{\phi_\Im\to 0} P(\phi_\Re + j\phi_\Im) = P(\phi)$.
2. $\int_{\mathbb{R}} |P(\phi_\Re + j\phi_\Im)|^2 \mathrm{d}\phi_\Re < \infty$ for $\phi_\Im < 0$ or $\phi_\Im > 0$, if $p(x)$ is right-sided or left-sided respectively.
3. The real and imaginary parts of $P(\phi)$ form a Hilbert transform pair.

Coming back to the problem at hand, let $p(x)$ represent a double-sided function and $P(\phi)$ its Fourier spectrum. As proposed in Section 6.6.1, we factor the function into its right-sided and left-sided parts $p_+(x)=p(x)H(x)$ and $p_-(x)=p(x)H(-x)$, respectively. The spectra of these single-sided parts are given by [22, 25, 26],

$$\begin{aligned} P_\pm(\phi) &= P(\phi) * \frac{1}{2}\left(\delta(\phi) \mp j\frac{1}{\phi}\right) \\ &= \frac{1}{2}\left(P(\phi) \mp j\mathcal{H}\{P\}(\phi)\right), \end{aligned}$$

where $\mathcal{H}\{\cdot\}$ is the Hilbert transform operator, and it follows that statement 3 in Titchmarsh's theorem holds. As a consequence, $P_\pm(\phi)$ admits an analytic continuation into one half of the complex plane.

Next consider the GFT integral (6.11) of the right/left-sided part of $p(x)$. We have

$$\begin{aligned} P_{\pm\alpha}(\phi) &= \int_{\mathbb{R}} p_\pm(x)\mathrm{e}^{\beta x}\mathrm{e}^{-j\phi x}\mathrm{d}x = \int_{\mathbb{R}} p_\pm(x)\mathrm{e}^{-j(\phi-\beta_\Im+j\beta_\Re)x}\mathrm{d}x \\ &= P_\pm(\phi - \beta_\Im + j\beta_\Re), \end{aligned}$$

with $\beta = \log(\alpha)/T_p$ and $\alpha \in \mathbb{C}\backslash\{0\}$. Writing $\alpha = |\alpha| e^{j\angle\alpha}$, then $\beta_{\Re} = \log(|\alpha|)/T_p$ and $\beta_{\Im} = \angle\alpha/T_p$ are the real and imaginary parts of β, respectively. Hence, we conclude that the GFT of $p_{\pm}(x)$ can be obtained by evaluating the analytic continuation of its Fourier spectrum $P_{\pm}(\phi)$, which exists by Titchmarsh's theorem, and is given by a linear combination of $P(\phi)$ and its Hilbert transform.

In order to generalize the 1-D results presented above, note that the Fourier transform of the orthant-sided Heaviside functions $\mathcal{F}\{H_q(\mathbf{x})\}$ is given by the Cartesian product of the unidimensional spectra. That is,

$$\mathcal{F}\{H_q(\mathbf{x})\} = \frac{1}{2^\nu} \bigotimes_{i=0}^{\nu-1} \left(\delta(\phi_i) - \varsigma_q(i) j \frac{1}{\phi_i} \right)$$

where $\bigotimes$ is the Cartesian product operator, ϕ_i are the components of the frequency vector $\boldsymbol{\phi}$, and as before, $\varsigma_q(i)$ denotes the sign of the coordinates of a given orthant with q an enumeration of the orthants. Recalling $p_{lq}(\mathbf{x}, t)$ (6.21) as the single-octact parts of the sound field of the mother sources, where l denotes the mother source number and q the octant number, and defining $\ast$ as the multidimensional convolution operator, we then write

$$P_{lq}(\boldsymbol{\phi}, \omega) = P_l(\boldsymbol{\phi}, \omega) \ast \frac{1}{8} \left(\bigotimes_i \left(\delta(\phi_i) - \varsigma_q(i) j \frac{1}{\phi_i} \right) \right),$$

where $i \in \{x, y, z\}$. This can be rewritten as,

$$\begin{aligned} P_{lq}(\boldsymbol{\phi}, \omega) = \frac{1}{8} \Big(& P_l(\boldsymbol{\phi}, \omega) - j \big(\varsigma_q(x) \mathcal{H}_x\{P_l\}(\boldsymbol{\phi}, \omega) + \varsigma_q(y) \mathcal{H}_y\{P_l\}(\boldsymbol{\phi}, \omega) \\ & + \varsigma_q(z) \mathcal{H}_z\{P_l\}(\boldsymbol{\phi}, \omega) \big) \\ & + \big(\varsigma_q(x)\varsigma_q(y) \mathcal{H}_{xy}\{P_l\}(\boldsymbol{\phi}, \omega) + \varsigma_q(x)\varsigma_q(z) \mathcal{H}_{xz}\{P_l\}(\boldsymbol{\phi}, \omega) \\ & + \varsigma_q(y)\varsigma_q(z) \mathcal{H}_{yz}\{P_l\}(\boldsymbol{\phi}, \omega) \big) \\ & - \varsigma_q(x)\varsigma_q(y)\varsigma_q(z) \mathcal{H}_{xyz}\{P_l\}(\boldsymbol{\phi}, \omega) \Big), \end{aligned}$$

where the Hilbert transform operator is taken on the specified (sub-indexed) dimensions. As it is seen, the Fourier spectrum of the single-octant sound field is given by a linear combination of the spectrum $P_l(\boldsymbol{\phi}, \omega)$ and its partial and full multidimensional Hilbert transforms. From Titchmarch's theorem we conclude that the resulting function is orthant-analytic in the 3-D complex space. The problem of finding the required GFT coefficients is thus reduced to computing the Hilbert transforms of the sound field spectra.

The spectrum of the sound field (centered at the origin of the coordinates) is given by (6.2) where we can ignore the term $\exp(j\boldsymbol{\phi}^T\mathbf{s})$ because of the shift-invariance property of the Hilbert transform. The partial Hilbert transform of the spectrum over one spatial variable is given by

$$\mathcal{H}_x\{P\}\left(\boldsymbol{\phi},\omega\right) = \frac{\phi_x}{\left(\phi_y^2+\phi_z^2-(\omega/c)^2\right)^{1/2}}\frac{1}{\|\boldsymbol{\phi}\|^2-(\omega/c)^2}, \tag{6.27}$$

equivalent expressions are obtained for the y or z direction. The partial Hilbert transform over two variables is given by

$$\begin{aligned}\mathcal{H}_{xy}\{P\}\left(\boldsymbol{\phi},\omega\right) = \frac{2}{\pi}\Bigg(\mathcal{H}_x\{P\}\left(\boldsymbol{\phi},\omega\right)\Bigg(&\log\left(\phi_y+\left(\phi_x^2+\phi_z^2-(\omega/c)^2\right)^{1/2}\right)\\ &-\log\left(\left(-(\omega/c)^2\right)^{1/2}\right)\Bigg)\\ +\mathcal{H}_y\{P\}\left(\boldsymbol{\phi},\omega\right)\Bigg(&\log\left(\phi_x+\left(\phi_y^2+\phi_z^2-(\omega/c)^2\right)^{1/2}\right)\\ &-\log\left(\left(-(\omega/c)^2\right)^{1/2}\right)\Bigg)\Bigg),\end{aligned} \tag{6.28}$$

with similar expressions for the xz or yz directions. See Appendix 6.B for the derivations. In a similar way, the Hilbert transform over the xyz-variables can be found.

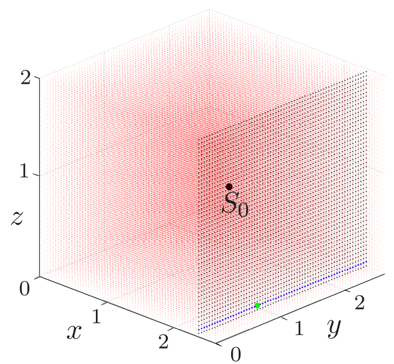
With these results, we can validate the theory presented in this paper. Note that knowledge of the partial Hilbert transform over one variable allows us to compute the spectrum of a single-sided (over the specified dimension) sound field. Knowledge of this transform and the partial transform over two variables allows us to obtain the spectrum of a quadrant-sided sound field. Using this "quadrant-sided" spectrum we can readily obtain (via analytic continuation) the GFT for general values of $\boldsymbol{\alpha}$ over two spatial dimensions, leaving the third dimension with a parameter equal to $\alpha : |\alpha| = 1$. In other words, we can specify reflection coefficients with absorption on four parallel walls in two dimensions leaving two parallel walls with reflection coefficients that only perform a phase shift but no absorption. The simulation experiments presented in Sec. 6.7 are thus based on the following expression,

$$P_{lq}(\boldsymbol{\phi},\omega) = \frac{\exp(j\boldsymbol{\phi}^T\mathbf{s}_l)}{4}\Big(P(\boldsymbol{\phi},\omega) - j\big(\varsigma_q(x)\mathcal{H}_x\{P\}(\boldsymbol{\phi},\omega) + \varsigma_q(y)\mathcal{H}_y\{P\}\,(\boldsymbol{\phi},\omega)\big) + \big(\varsigma_q(x)\varsigma_q(y)\mathcal{H}_{xy}\{P\}\,(\boldsymbol{\phi},\omega)\big)\Big), \tag{6.29}$$

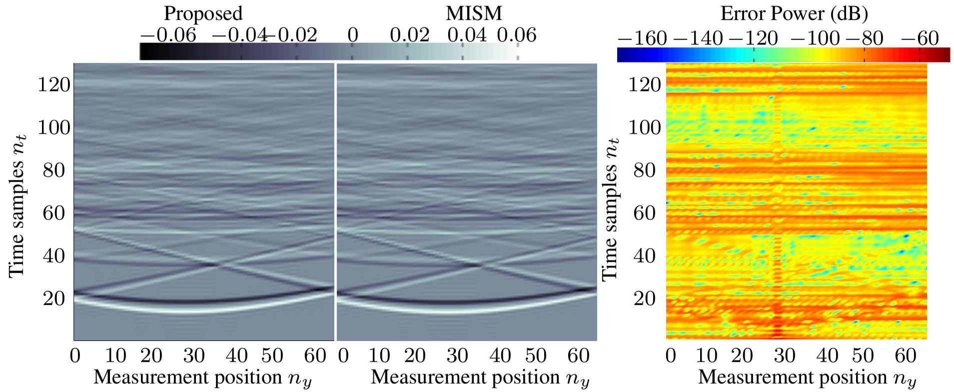
with equivalent expressions if absorption is to be modeled on parallel walls in the xz or yz direction.

6.7 Experimental results

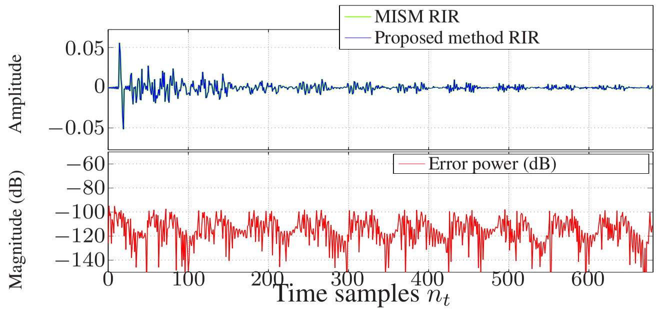
In this section we present simulation results. We compare the newly proposed method against the MISM [13] for a box-shaped room with constant wall reflection coefficients. To avoid the non-banlimited representation of the delta pulses in the MISM, we follow Peterson's approach [27]. Each delta pulse is replaced by the impulse response of a Hanning-windowed ideal low-pass filter with cut-off frequency set to the Nyquist frequency. The high-pass filter, suggested as post-processing operation in the original paper from Allen and Berkley [13] is disabled in the experiments to avoid biased results. All the experiments are performed in MATLAB®, on a PC computer with 16 cores running at 2 Ghz. The newly proposed method is implemented as a mixed m-file and C++ mex function. For the MISM implementation we modified the efficient C++ mex-function algorithm by E. Habets [28] making it multithreaded, this in order to take advantage of all the computational cores. The computed scenario is displayed in Fig. 6.7a. The room dimensions are $[L_x, L_y, L_z]^T = [2.6, 2.6, 2]^T$. The wall reflection coefficients are $\varrho_{x0} = 1$ and $\varrho_{x1} = -1$, $\varrho_{y0} = 0.5$ and $\varrho_{y1} = -0.6$, $\varrho_{z0} = 0.7$ and $\varrho_{x1} = -0.8$. Notice that the wall perpendicular to the x direction adjacent to the origin is rigid (i.e. fully reflective) and the opposite wall is soft (i.e. fully reflective with a π phase shift). We consider temporal signals bandlimited to 2Khz. The temporal sampling frequency is thus set to $f_s = 40000$ Hz, or in radians per second $\omega_s = f_s 2\pi$. The simulation length is $T_h = 1.024$s, or equivalently in samples, $N_t = 4096$. We choose the temporal frequency sampling interval to be $\Omega_s = \omega_s/2N_t$, or equivalently $T_p = 2T_h$. The spectral-sampling matrix is thus given by $\boldsymbol{\Psi} = \text{diag}(\pi/L_x, \pi/L_y, \pi/L_z, \Omega_s)$. The spectral-periodicity matrix is chosen to be $\boldsymbol{\Sigma} = 2\boldsymbol{\Psi}\text{diag}(N_x, N_y, N_z, N_t)$, with $N_y = 64$, $N_y = 64$ and $N_y = 48$. This directly defines a sound field sampling lattice Γ, with generator matrix $\boldsymbol{\Gamma} = 2\pi\boldsymbol{\Sigma}^{-T} = \text{diag}(L_x/N_x, L_y/N_y, L_z/N_z, 2\pi/\omega_s)$. The diagonal form of $\boldsymbol{\Gamma}$



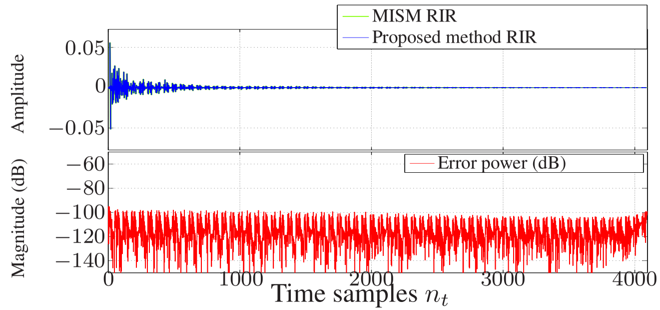
(a) *Computed scenario,* 196608 receiver positions ($64x$, $64y$, $48z$ red dots).



(b) Sound field measured at line $n_z = 2$ of plane $n_x = 58$ (blue line in (a)), up to time 0.032s or 128 samples. The proposed method and the MISM (left plot). Error power in dB (right plot).



(c) Comparison of both methods for measurement position $n_x = 58$, $n_z = 2$, $n_y = 23$ (green dot in (a)) and $0.17s$ or 680 samples.



(d) Comparison of both methods for measurement position $n_x = 58$, $n_z = 2$, $n_y = 23$ (green dot in (a)) for the whole simulation time.

Figure 6.7: Simulation for a room with dimensions $2.6x \times 2.6y \times 2z$ m. The source is given by S_0 in (a). The reflection coefficients are $\varrho_{x0} = 1$, $\varrho_{x1} = -1$, $\varrho_{y0} = 0.5$, $\varrho_{y1} = -0.6$, $\varrho_{z0} = 0.7$, $\varrho_{x1} = -0.8$. The simulation time is $T_h = 1.024$s, or $N_t = 4096$ samples.

imposes a rectangular arrangement of the sound field measurement positions. This can be seen in Fig. 6.7a, where the $N_x = 65$ $N_y = 64$ and $N_z = 48$ measurement positions are arranged in a rectangular lattice, giving a total of 196608 spatial positions to be calculated. The source S_0 is positioned at $\mathbf{s}_0 = [1.71, 1.14, 1.02]^T$. We set $\alpha_t = 0.9$, in this way we obtain a mild temporal aliasing suppression. The spatial components of the GFT parameter per octant are set to $\varrho_{qi} = (\varrho_{i0}\varrho_{i1})^{\varsigma_q(i)}$ for $q = 0, \ldots, 7$ and $i \in \{x, y, z\}$. The GFT parameter per octant is thus given by $\boldsymbol{\alpha}_q = [-1, \varrho_{qy}, \varrho_{qz}, \alpha_t]^T = [-1, -(0.3)^{\varsigma_q(y)}, -(0.56)^{\varsigma_q(z)}, 0.9]^T$. The generalized Fourier spectrum is obtained evaluating (6.29) at the complex frequency variables given by,

$$P_{\boldsymbol{\alpha}lq}(\boldsymbol{\phi}, \omega) = P_{lq}(\boldsymbol{\phi} - \boldsymbol{\beta}_{\Im} + j\boldsymbol{\beta}_{q\Re}, \omega - \beta_{\omega\Im} + j\beta_{\omega\Re})$$

with $\boldsymbol{\beta}_{q\Re} = [0, \varsigma_q(y)\log(0.3)/(2L_y), \varsigma_q(z)\log(0.56)/(2L_z)]^T$, $\boldsymbol{\beta}_{\Im} = ([\pm\pi/(2L_i)]_i)^T$, $\beta_{\omega\Im} = 0$ and $\beta_{\omega\Re} = \log(0.9)/T_p$. Next we evaluate generalized spectra at sampling points given by $\boldsymbol{\Psi}$ to obtain the coefficients $P_{\boldsymbol{\alpha}lq}(\boldsymbol{\Psi}\mathbf{k})$. To perform the generalized Fourier synthesis given by (6.25), first a multidimensional inverse FFT is applied on the coefficients, followed by a modulation by $\exp(-\boldsymbol{\beta}_q^T \boldsymbol{\Gamma}\mathbf{n})$, $\mathbf{n} \in \mathbb{R}^3$. The scaling factor $(|\boldsymbol{\Delta}|\mathrm{N}(\Delta/\Gamma))^{-1}$ can be applied in any order. The final result is obtained by (6.26) using the alignment factors given in Table 6.1. The synthesis gives us all the RIRs inside the room. For the MISM we evaluate individually the RIRs from the source to each measuring position. We use a triplet of integers n_y, n_z, n_t, to index the sound field samples. The colormap plots in Fig. 6.7b show the results only for a measurement line in the y direction at $n_z = 2$, of plane $n_x = 58$ (dark plane in in Fig. 6.7b) and time samples $0 \leq n_t \leq 128$. The measurement line is displayed in blue in Fig. 6.7a. In Fig. 6.7c, a plot is given where we compare the RIRs only for measurement position $n_y = 23$ of the same line (the green dot in Fig. 6.7a denotes this position) for time samples $0 \leq n_t \leq 680$. Additionally in Fig. 6.7d a comparison of the full 1.024s long RIRs is given. As it is seen, the spatio-temporal "locations" and amplitudes of the reflections are perfectly modeled by the newly proposed method. However, still small discrepancies are observed between both approaches, these are caused by two factors. First, temporal aliasing is still present. This can be further made arbitrarily small by decreasing the spectral sampling interval Ω_s. More importantly, the (finite-length) Hanning-windowed low-pass filter used in the MISM produces a misalignment between both methods. As the window length of the filter is increased the error is accordingly decreased. For this particular experiment the window length is set to 2s. The proposed algorithm took 4.61 hours to calculate the 196608 spatial positions. The MISM took 591days or 1.61 years. We did not compute the MISM on the whole set of 196608 positions. Instead we compute 3072 positions arranged on a plane (depicted by the dark plane in Fig. 6.7b) and extrapolate the result. This experiment shows the drastic difference in computational complexity for the given scenario.

6.8 Conclusions

A generalized Fourier domain framework to acoustic modeling in box-shaped rooms is presented in this paper. For rooms with constant wall reflection coefficients, we show that the sound field inside the enclosure can be expressed as a geometrically weighted periodic summation. We relate the samples of the generalized spectrum of the free-field sound field, to the weighted periodic summation that describes the spatio-temporal sound field in the room. Discretizing all domains and periodically extending the generalized Fourier series space over a lattice, we obtained the generalized discrete Fourier transform (GDFT). The GDFT can be implemented using the FFT, which allowed us to drastically reduce the complexity of the generalized Fourier synthesis. Using these results, we derived a method to compute massive amounts of RIRs over the full room space with very low complexity, of order $\mathcal{O}(N\log N)$ per measuring position, which significantly outperforms the $\mathcal{O}(N^3)$ complexity of the MISM [13].

Appendix 6.A Proof of Proposition 6.1

Proof. If $\boldsymbol{\phi}=2\pi\boldsymbol{\Lambda}^{-T}$, Proposition 6.1 follows directly by evaluating $P_{\boldsymbol{\alpha}lq}(\boldsymbol{\Phi}\mathbf{k},\omega)$, $\mathbf{k}\in\mathbb{Z}^3$ on the generalized Poisson summation formula (6.15). The converse is proven as follows. Given $\boldsymbol{\alpha}=[\varrho_x^{S_q(x)},\varrho_y^{S_q(y)},\varrho_z^{S_q(z)},\alpha_t]^T$, let $\beta_t=\log(\alpha_t)/T_p$ and $\boldsymbol{\beta}_q=\boldsymbol{\Lambda}^{-T}\log([\varrho_x^{S_q(x)},\varrho_y^{S_q(y)},\varrho_z^{S_q(z)}]^T)$, then we have that,

$$P_{\boldsymbol{\alpha}lq}(\boldsymbol{\Phi}\mathbf{k},\omega)=\iint_{\mathbb{R}^3\,\mathbb{R}} p_{lq}(\mathbf{x},t)\mathrm{e}^{\boldsymbol{\beta}_q^T\mathbf{x}+\beta_t t}\mathrm{e}^{-j(\mathbf{k}^T\boldsymbol{\Phi}^T\mathbf{x}+\omega t)}\mathrm{d}t\mathrm{d}\mathbf{x}$$

$$=\int_{\mathbb{R}}\sum_{\mathbf{n}\in\mathbb{Z}^3}\int_{V_\Lambda(\mathbf{n})} p_{lq}(\mathbf{x},t)\mathrm{e}^{\boldsymbol{\beta}_q^T\mathbf{x}+\beta_t t}\mathrm{e}^{-j(\mathbf{k}^T\boldsymbol{\Phi}^T\mathbf{x}+\omega t)}\mathrm{d}\mathbf{x}\mathrm{d}t. \tag{6.30}$$

Making the substitution, $\mathbf{x} \rightarrow \mathbf{x} + \mathbf{\Lambda n}$, (6.30) becomes

$$P_{\boldsymbol{\alpha}lq}(\mathbf{\Phi k}, \omega) = \int_{\mathbb{R}} \int_{V_\Lambda(\mathbf{0})} \sum_{\mathbf{n} \in \mathbb{Z}^3} p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t) \mathrm{e}^{\boldsymbol{\beta}_q^T (\mathbf{x} + \mathbf{\Lambda n})} \mathrm{e}^{\beta_t t} \mathrm{e}^{-j(\mathbf{k}^T \mathbf{\Phi}^T (\mathbf{x} + \mathbf{\Lambda n}) + \omega t)} \mathrm{d}\mathbf{x}\mathrm{d}t. \tag{6.31}$$

Since

$$\begin{aligned} \boldsymbol{\beta}_q^T \mathbf{\Lambda n} &= \log([\varrho_x^{\varsigma_q(x)}, \varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}]) \mathbf{\Lambda}^{-1} \mathbf{\Lambda n} \\ &= \log([\varrho_x^{\varsigma_q(x)}, \varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}])^T \mathbf{n}, \end{aligned}$$

we then have that $\mathrm{e}^{\boldsymbol{\beta}_q^T \mathbf{\Lambda n}} = \prod_{i \in \{x,y,z\}} \varrho_i^{\varsigma_q(i)}$, so that (6.31) can be rewritten as,

$$\begin{aligned} P_{\boldsymbol{\alpha}lq}(\mathbf{\Phi k}, \omega) = \int_{\mathbb{R}} \int_{V_\Lambda(\mathbf{0})} \sum_{\mathbf{n} \in \mathbb{Z}^3} \left(\prod_i \varrho_i^{\varsigma_q(i) n_i} \right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t) \mathrm{e}^{-j \mathbf{k}^T \mathbf{\Phi}^T \mathbf{\Lambda n}} \\ \times \mathrm{e}^{\boldsymbol{\beta}_q^T \mathbf{x} + \beta_t t} \mathrm{e}^{-j(\mathbf{k}^T \mathbf{\Phi}^T \mathbf{x} + \omega t)} \mathrm{d}\mathbf{x}\mathrm{d}t. \end{aligned} \tag{6.32}$$

If $\mathbf{\Phi}^T \mathbf{\Lambda} = 2\pi \mathbf{U}$, with $\mathbf{U}$ an integer matrix, the term $\mathrm{e}^{-j \mathbf{k}^T \mathbf{\Phi}^T \mathbf{\Lambda n}} = 1$ for all $\mathbf{k}, \mathbf{n} \in \mathbb{Z}^3$, and only if $\mathbf{\Phi}^T \mathbf{\Lambda} = 2\pi \mathbf{I}$, with $\mathbf{I}$ the identity matrix, or equivalently $\mathbf{\Phi} = 2\pi \mathbf{\Lambda}^{-T}$, then the coefficient functions $P_{\boldsymbol{\alpha}lq}(\mathbf{\Phi k}, \omega)$ correspond to the exact generalized Fourier series expansions (defined by (6.17)) of

$$|\mathbf{\Lambda}| \sum_{\mathbf{n} \in \mathbb{Z}^3} \left(\prod_i \varrho_i^{\varsigma_q(i) n_i} \right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t) \mathrm{e}^{-j \mathbf{k}^T \mathbf{\Phi}^T \mathbf{\Lambda n}},$$

which completes the proof. □

Appendix 6.B Hilbert transforms of the spectrum of the sound field

The partial Hilbert transform of $P(\boldsymbol{\phi}, \omega) = (\|\boldsymbol{\phi}\|^2 + (\omega/c)^2)^{-1}$ over one specified spatial dimension is first derived. Let us take ϕ_x as the transform variable, we then rewrite $P(\boldsymbol{\phi}, \omega)$ in terms of ϕ_x as, $P(\boldsymbol{\phi}, \omega) = (\phi_x^2 + a^2)^{-1}$, with $a^2 = \phi_y^2 + \phi_z^2 - (\omega/c)^2$. Combining equations (2.2), (2.3) and (2.4), pp. 459 in [29], we obtain,

$$\mathcal{H}_x\{P\}(\boldsymbol{\phi}, \omega) = \frac{\phi_x}{a} \frac{1}{\phi_x^2 + a^2}.$$

Substituting for the value of a we get the result in (6.27). The partial Hilbert transform of $P(\boldsymbol{\phi},\omega)$ over two spatial variables is denoted as,

$$\mathcal{H}_{xy}\{P\}(\boldsymbol{\phi},\omega) = \mathcal{H}_y\{\mathcal{H}_x\{P\}\}(\boldsymbol{\phi},\omega) = \mathcal{H}_x\{\mathcal{H}_y\{P\}\}(\boldsymbol{\phi},\omega).$$

From (6.27) we have,

$$\mathcal{H}_x\{\mathcal{H}_y\{P\}\}(\boldsymbol{\phi},\omega) = y\mathcal{H}_x\left\{\frac{1}{\left(\phi_x^2+\phi_z^2-(\omega/c)^2\right)^{1/2}}\frac{1}{\|\boldsymbol{\phi}\|^2-(\omega/c)^2}\right\}.$$

Applying Bedrosian's theorem [25] pp. 184, we obtain

$$\mathcal{H}_x\left\{\frac{1}{\left(\phi_x^2+\phi_z^2-(\omega/c)^2\right)^{1/2}}\frac{1}{\|\boldsymbol{\phi}\|^2-(\omega/c)^2}\right\} = \frac{y}{\|\boldsymbol{\phi}\|^2-(\omega/c)^2}\mathcal{H}_x\left\{\frac{1}{\left(\phi_x^2+\phi_z^2-(\omega/c)^2\right)^{1/2}}\right\}$$

and respectively,

$$\mathcal{H}_y\{\mathcal{H}_x\{P\}\}(\boldsymbol{\phi},\omega) = \frac{x}{\|\boldsymbol{\phi}\|^2-(\omega/c)^2}\mathcal{H}_y\left\{\frac{1}{\left(\phi_y^2+\phi_z^2-(\omega/c)^2\right)^{1/2}}\right\}.$$

Following the procedure to obtain result (4.3.81) in [25] pp. 203, we obtain,

$$\mathcal{H}_x\left\{\frac{1}{\left(\phi_x^2+\phi_z^2-(\omega/c)^2\right)^{1/2}}\right\} = \frac{2}{\pi}\times \frac{\log\left(\phi_x+(\phi_y^2+\phi_z^2-(\omega/c)^2)^{1/2}\right)-\log\left((-(\omega/c)^2)^{1/2}\right)}{\left(\phi_x^2+\phi_y^2-(\omega/c)^2\right)^{1/2}}$$

with a similar expression for $\mathcal{H}_y\left\{1/\left(\phi_y^2+\phi_z^2-(\omega/c)^2\right)^{1/2}\right\}$. Combining these results we arrive to (6.28).

Acknowledgment

The authors gratefully acknowledge Prof. dr. ir. A.J.E.M. Janssen of Eindhoven University of Technology for his valuable help and discussions on the calculus of the Fourier integral of the single-orthant free-field sound field, and Dr. ir. Richard Hendriks from Delft University of Technology for his remarks and proofreading of the manuscript.

6.9 Bibliography

[1] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing (Signals and Communication Technology)*. New York: Springer-Verlag, 2006.

[2] J. Benesty, M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*, 1st ed. Berlin, Germany.: Springer-Verlag, Dec. 2007.

[3] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, 1st ed. New York: Springer, Jun. 2001.

[4] A. J. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Amer.*, vol. 93, pp. 2764–2778, 1993.

[5] J. Ahrens and S. Spors, "Analytical driving functions for higher order ambisonics," in *ICASSP*, 31 March - April 4 2008, pp. 373 –376.

[6] M. Kolundzija, C. Faller, and M. Vetterli, "Sound field reconstruction: An improved approach for wave field synthesis," in *AES 126th Conv.*, May 2009.

[7] S. G. McGovern, "Fast image method for impulse response calculations in box-shaped rooms," *Appl. Acoust.*, vol. 70, pp. 182–189, Jan. 2009.

[8] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Fast evaluation of the room transfer function using multipole expansion." *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 565–576, Feb. 2007.

[9] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan., "A beam tracing method for interactive architectural acoustics," *J. Acoust. Soc. Amer.*, vol. 115, no. 2, pp. 739–756, Feb. 2004.

[10] E. Lehmann and A. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1429 –1439, Aug. 2009.

[11] R. Mehra, N. Raghuvanshi, L. Savioja, M. C. Lin, and D. Manocha, "An efficient gpu-based time domain solver for the acoustic wave equation," *Appl. Acoust.*, vol. 73, no. 2, pp. 83 –94, Feb. 2012.

[12] J. Martinez and R. Heusdens, "On low-complexity simulation of multichannel room impulse responses," *IEEE Signal Process. Lett.*, vol. 17, no. 7, pp. 667 –670, Jul. 2010.

[13] J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[14] J. Martinez, R. Heusdens, and R. C. Hendriks, "A generalized poisson summation formula and its application to fast linear convolution," *IEEE Signal Process. Lett.*, vol. 18, no. 9, pp. 501 –504, Sept. 2011.

[15] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling." *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

[16] P. Morse and K. Ingard, *Theoretical Acoustics*. New York: McGraw-Hill, 1968.

[17] H. Kuttruff, *Room Acoustics*. London, U.K.: Taylor and Francis, 2000.

[18] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms and Applications*, 3rd ed. New Jersey: Prentice Hall, Inc., 1996.

[19] P. P. Vaidyanathan, *Multirate Systems And Filter Banks*. New Jersey: Prentice Hall PTR, 1992.

[20] F. Marvasti, Ed., *Nonuniform Sampling: Theory and Practice*, ser. Information Technology: Transmission, Processing and Storage. New York: Kluwer Academic/Plenum Publishers, Jan. 2001, no. 2.

[21] K. B. Wolf, *Integral Transforms in Science and Engineering*. New York: Plenum Press, 1979.

[22] R. Bracewell, *The Fourier Transform & Its Applications*, 3rd ed. New York: McGraw-Hill, 1999.

[23] J. Martinez, R. Heusdens, and R. C. Hendriks, "A generalized Fourier domain: Signal processing framework and applications," vol. 93, no. 5, pp. 1259–1267, May 2013.

[24] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed. London: Oxford University Press, 1948.

[25] F. W. King, *Hilbert Transforms*. Cambridge: Cambridge University Press, 2009, vol. 1.

[26] R. Kanwal, *Generalized Functions, Theory and Technique*, 2nd ed. Boston: Birkhauser, 1998.

[27] P. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, pp. 1527–1529, Nov. 1986.

[28] E. Habets. (2010, Sept.) Room impulse response generator for MATLAB. RIR_Generator_v2.0_20100920.zip. [Online]. Available: http://home.tiscali.nl/ehabets/rir_generator

[29] F. W. King, *Hilbert Transforms*. Cambridge: Cambridge University Press, 2009, vol. 2.

7

Concluding comments and future work

7.1 The channel identification problem

Simulation of room impulse responses is part of a bigger and richer problem in acoustic DSP, acoustic communication and, in general, in any wireless communication system based on wave propagation. This is the modeling or estimation of the *medium channel*. In physics the medium channel is given by the Green's function, a concept used in Part I of this thesis. In mathematics an abstraction of this concept is given by the *fundamental solution* of a functional operator [1]. In room acoustics it is called the RIR and in signal processing just *impuse response* [2]. It is then not surprising to find that fundamental mathematical and conceptual relationships exist between RIR simulation algorithms and the methods and solutions found in other branches of science.

The RIR, i.e. the medium channel, plays a major role in many acoustical problems, e.g. multichannel acoustic echo cancellation [3], dereverberation and channel equalization [3, 4], beamforming [5], sound field synthesis [6, 7], acoustic underwater communication [8, 9]. The distinctive challenge is that in practice the RIR is a wideband, slowly decaying, highly space-time-varying function. The RIR is a function of the positions and directivities of sources and receivers, temperature, inhomogeneous medium density, moving objects in the enclosure and rapidly changing boundary conditions, just to name a few. Complex statistical methods are then used to address the levels of dynamism and complexity of the channel identification problem, but still solutions remain difficult to obtain in the most challenging scenarios.

The deterministic RIR models reviewed in Sec. 2 constitute a set of algorithms that attempt to mimic (with different levels of accuracy) the underlying physics of wave dynamics. Novel lines of research might be devoted to exploit the information given by these physical models as a *prior* or *constraint* in the statistical process in order to ease the channel identification problem. Some early steps in this direction have been taken in the context of *field estimation* and *localization* [10], showing the potential of using physical deterministic models to address the channel identification problem in the most challenging conditions. The spatio-temporal RIR model with low-complexity presented in this thesis might then prove of valuable practical use for applications beyond room acoustics simulation.

7.2 The generalized Fourier transform

The generalized Fourier transform (GFT), introduced in Chapters 4, 5 and 6, is originally derived to address an important restriction of the RIR simulation model presented in Chapter 3, namely that only perfectly reflective walls are simulated. In practice that assumption is not valid, and an extension of the model was needed. The method relays on the key property that *sampling* of a function results in a *periodic summation* of its Fourier transform [2, 11, 12]. This property is formalized by the Poisson summation formula [1, 13]. Using this idea, first the sound field in the room is modeled by a periodic summation and then it is associated with a sampling condition in the Fourier domain. From the samples of the spectrum the reverberated sound field is synthesized.

When sound absorption is allowed at the walls the sound field cannot anymore be expressed by a periodic summation. Every reflection gets damped by a factor, reflections of reflections then decay at a geometric rate. These reflections can be seen as geometrically damped periodic repetitions of the original sound field. The question was then: “Does a Poisson summation-like formula exist that relates the samples of a transformed function to a geometrically weighted summation of the function?” The search concluded in the derivation of a *generalized Poisson summation formula* (GPSF) that formalized this property. The associated transform is closely related to the Fourier transform (FT). It can be seen as the FT of a function modulated by an exponential window. The transform was then named generalized Fourier transform (GFT). Just as the FT, the GFT is a special case of the bilateral Laplace transform (BLT) [14], it can be seen as the BLT evaluated at a linear manifold in the s-space. If that linear manifold is included in the region of converge (ROC) of the BLT, then the GFT exists. What makes the GFT distinct from just a modulated FT or another special case of the BLT is the generalized Poisson summation formula associated with it.

The connections of the GFT with sampling, analyticity, the z-transform [2] and the Hilbert transform [15–17] are briefly analysed in Chapters 5 and 6. This analysis is still far from being complete, and just as special properties like the *weighted circular convolution theorem* where derived, a signal processing framework based on the GFT might still have some interesting and helpful properties to be discovered.

7.3 Bibliography

[1] R. Kanwal, *Generalized Functions, Theory and Technique*, 2nd ed. Boston: Birkhauser, 1998.

[2] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms and Applications*, 3rd ed. New Jersey: Prentice Hall, Inc., 1996.

[3] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing (Signals and Communication Technology)*. New York: Springer-Verlag, 2006.

[4] J. Benesty, M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*, 1st ed. Berlin, Germany.: Springer-Verlag, Dec. 2007.

[5] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, 1st ed. New York: Springer, Jun. 2001.

[6] A. J. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Amer.*, vol. 93, pp. 2764–2778, 1993.

[7] M. Kolundzija, C. Faller, and M. Vetterli, "Sound field reconstruction: An improved approach for wave field synthesis," in *AES 126th Conv.*, May 2009.

[8] M. Stojanovic, *Underwater Acoustic Communication*. John Wiley & Sons, Inc., 2001.

[9] M. Chitre, S. Shahabudeen, and M. Stojanovic, "Underwater acoustic communications and networking: Recent advances and future challenges," *Marine Technology Society Journal*, vol. 42, no. 1, pp. 103–116, 2008.

[10] T. van Waterschoot and G. Leus, "Distributed estimation of static fields in wireless sensor networks using the finite element method," in *ICASSP*. Kyoto, Japan: IEEE Signal Processing Society, May 2012, pp. 2853–2856.

[11] P. P. Vaidyanathan, *Multirate Systems And Filter Banks*. New Jersey: Prentice Hall PTR, 1992.

[12] R. Bracewell, *The Fourier Transform & Its Applications*, 3rd ed. New York: McGraw-Hill, 1999.

[13] F. Marvasti, Ed., *Nonuniform Sampling: Theory and Practice*, ser. Information Technology: Transmission, Processing and Storage. New York: Kluwer Academic/Plenum Publishers, Jan. 2001, no. 2.

[14] K. B. Wolf, *Integral Transforms in Science and Engineering*. New York: Plenum Press, 1979.

[15] F. W. King, *Hilbert Transforms*. Cambridge: Cambridge University Press, 2009, vol. 1.

[16] ——, *Hilbert Transforms*. Cambridge: Cambridge University Press, 2009, vol. 2.

[17] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed. London: Oxford University Press, 1948.

Acknowledgements

No person can claim personal success exclusively to her or his own merits. The opportunities we are offered during our lives are the result of the hard work, effort and sacrifice that people around us and before us have made.

The work described in this thesis was performed at the Signal & Information Processing (SIP) Lab, part of the faculty of Electrical Engineering, Computer Science and Mathematics (EWI), of Delft University of Technology. The research was conducted in the period 2008 – 2012. The financial support from STW and Google to perform the research is gratefully acknowledged.

My time at the SIP lab was very rewarding. I had the privilege to be supervised by one of the smartest scientists I know. Most importantly, I am happy to be able to regard Richard Heusdens as a close friend. Without his excellent supervision this work would have not been possible.

My recognition goes also to my colleagues Cees Taal and Richard Hendriks. Cees, thank you for all the fruitful discussions we had, for the cheerful environment you brought to our room, and most of all, for your friendship. Richard, I really appreciate your support and encouragement not only research wise, but also in other important aspects of life.

I am grateful to Inald Lagendijk for being my promotor, and to Jan Biemond for originally accepting me as part of the group. I would like to acknowledge Nikolay Gaubitch and Jan van der Lubbe. Nick, it has been a very rewarding experience working with you so far. Thank you for all your support. Jan, it was a pleasure to discuss with you the propositions accompanying this thesis, thank you for your advice. Of course I cannot forget my old and current colleagues of the Multimedia Signal Processing (MSP) Group: Christian, Zeki, Yuan, Guoqiang, João, Marco, Zhijie, Jan Erkelens, David, Emile, Laurens, Jeroen, Marc, Ahmed, Sephide, Behnaz, Maarten, Stevan, Feifei, Jan Bot, Wynand, Alessandro, Veronika, Carsten, Christoph, Cynthia, Raynor, Robbert, Stefanie. Thank you all for sharing with me all those moments that made work enjoyable and meaningful. A special mention goes to Saskia Peters. Saskia you excel at your work as the secretary of our group. Thank you for always being reachable and supportive.

I owe a lot to Bastiaan Kleijn, to Geert Leus and to Alle-Jan van der Veen. Bastiaan, you are an admirable scientist. Thank you for giving me the opportunity of being part of your group at Victoria University, and part of your Google project reseach team. Most importantly, thank you for your guidance, advice and friendship. Geert, I'm sincerely grateful to you for giving me the opportunity to work with you and to learn from you. The scientific level you have is remarkable. Alle-Jan, thank you for accepting me as part of the Circuits and Systems (CAS) group.

I would like to extend my appreciation to my colleagues of the CAS group. Hamid, Hadi, Sina, Seyran, Shahzad, Dony, Sundeep, Nihan, Venkat, Rob, Rosario, Nick, Millad, Yan, Yongchang, Minaksie. It has been an amazing experience to be part of the group. Always keep the good mood and passion for research and work that you have. Moreover, I would like to thank Raj Thilak Rajan, Andrea Simonetto and Rocío Arroyo Valles for helping me polishing the details of this thesis. I also extend an affectionate greeting to Wenyu, Petko and the rest of the people at Victoria for the nice time we had together.

My gratitude also goes to my friends and family-in-law in the Netherlands. The next lines are dedicated to them.

Ik zou de gelegenheid willen gebruiken om mijn vrienden en schoonfamilie in Nederland te bedanken. De groeten aan Anna en Bastiaan, Petra en Aiden, Janneke en Peter, Suzanne en Erwin, Ronni en Gal, Mariette en Jordy, Fabian en Nairobi. Bedankt jongens voor alle lol die wij samen hebben gehad, en de hulp dat ik van jullie heb gekregen.

Een speciale dank aan Anouk, Lieneke en Jos, en Klaas en Anita. Jullie zijn meer dan vrienden of schoonfamilie voor mij, jullie zijn echte familie van mij geworden, bedankt voor jullie steun, hulp en liefde.

I would like to take the opportunity to thank my friends in Mexico for all their support and trust.

Mi más cordial reconocimiento a Alejandro Valdés, Alejandro Villers, David Ribeiro, Jack Fermon y Octavio Capilla. Gracias por su apoyo, solidaridad y aprecio a lo largo de los años, tanto en los momentos de alegría como de tristeza. Contar con su amistad es un privilegio. Tampoco puedo olvidar a los demás miembros de la banda de la Anáhuac: Iván, Ricardo Sandoval, Ricardo Farrera, Rafa, Adolfo, Jackie, Adriana, Guille, Alfredo Gargari, Alfredo Serrano, Alberto Maya y David Mestas. Gracias amigos. Un saludo especial a Victor Cruz. Victor, gracias por compartir conmigo el cariño por las matemáticas. Finalmente, quiero extender mis agradecimientos a Caleb Navarro, Jarumi Álvarez y Sergio Hernández. Compartimos juntos momentos tanto agradables como difíciles durante nuestro tiempo de adolescencia. Gracias por su apoyo y cariño, aún los tengo muy presentes.

These lines would have never come to exist without the love and support that my family gave me. The following words are dedicated to them.

Estas líneas nunca hubieran tomado forma de no ser por el apoyo, sacrificio y amor que ustedes: Salus, Mely, Susi y Pepe, me han brindado durante toda mi existencia. Madre y Mamá Abuela, ustedes son los pilares en los que mi temple, mis sueños y mi pasión por el estudio se fundamentan. Es además gracias al cariño, la formación y la sabiduría de mi padre, Pedro Castañeda Martínez – que en paz descanse – que he adquirido los valores fundamentales para conducirme con rectitud por la vida. Con orgullo y profundo agradecimiento les entrego los resultados que he cosechado.

Finally, these last words are dedicated to Hester, my partner in life, best friend and love.

Het leven is een prachtige avontuur. Het is vol met blije momenten, gelegenheden en uitdagingen, maar ook met verdriet en pijn. Hester, je ben de bron van mijn kracht, mijn passie en liefde voor het leven, en mijn hoop voor de toekomst. Voor al jouw zorg, warmte en liefde, zeg ik met de hand op mijn hart: *bedankt*.

Jorge Abraham Martínez Castañeda,
Delft, November 2013.

Curriculum Vitae

Jorge Abraham Martínez Castañeda was born in 1981 in Mexico City, Mexico. In 2003 he received the Bachelor degree (cum-laude) in Information Technology and Telecommunications Engineering from Anáhuac University, Mexico.

From 2004 to 2006, he worked as information technology administrator and developer engineer in ACTIA de México S.A. de C.V. (ACTIA group). An international group with multimedia technology manufacturing and distribution as core business.

In 2006 he spent a month in Montreal, Canada, studying French, afterwards he moved to Spain to follow a master program at the Autonomous University of Barcelona under the supervision of Prof. Dr. Joan Serra Sagristà. As part of his experience in Barcelona, he signed up for Catalan language curses. In 2007 Jorge is granted the M.Sc. degree in Advanced Informatics, after presenting the thesis with title "Hyperspectral data coding with JPEG 2000 implementing 3D wavelet transforms", for which he received a cum-laude note.

In 2008, Jorge moves to the Netherlands, where he got a position as Ph.D. student at the Signal & Information Processing Lab of the Faculty of Electrical Engineering Computer Science and Mathematics (EWI) in Delft University of Technology. Here he conducted scientific research under the supervision of Dr. ir. Richard Heusdens, on the field of massive multichannel multiple-input, multiple-output acoustic systems, multidimensional Fourier processing of room transfer functions, and adaptive algorithms for multichannel acoustic echo cancellation. He was also appointed as teaching assistant for practical lessons on graduate signal processing curses. During his stay in Delft he took Dutch language curses.

After his Ph.D. student time, in 2012 Jorge is appointed research fellow at the Signal & Information Processing Lab to work on topics related to multichannel acoustic echo cancellation sponsored by a "Google Faculty Research Awards" grant. From February to April of the same year he engages in an internship at Victoria University of Wellington, New Zealand under the supervision of Prof. Dr. Bastiaan Kleijn to work on the topic "Sound field reconstruction in enclosed spaces".

From December 2012 Jorge joins the Circuits and Systems (CAS) group as a reseach fellow under the supervision of Prof. dr. ir. Geert Leus to work on topics related to field estimation in wireless sensor networks in the context of Leus's VICI project entitled "Self organizing wireless networks".

Jorge likes to read novels, among his favorite authors are Gabriel García Márquez, Paulo Cohelo, Herta Müller and Khaled Hosseini. Singing and drawing are his favorite hobbies (although he does this only occasionally). He likes very much cycling and running. For a time he practiced tae-kwon-do and enjoyed surfing in Cozumel, his home town island in Mexico.

Glossary

Acronyms

AEC	Acoustic echo cancellation
ARD	Adaptive rectangular decomposition
BEM	Boundary element method
BT	Beam tracing
DFT	Discrete Fourier transform
DLRM	Diffuse late-reverberation model
DSP	Digital signal processing
DWM	Digital waveguide mesh
FD	Fourier domain
FDTD	Finite-difference time-domain
FEM	Finite element method
FFT	Fast Fourier transform
FMM	Fast multipole method
FT	Fourier transform

GDFT	Generalized discrete Fourier transform
GFD	Generalized Fourier domain
GFFT	Generalized fast Fourier transform
GFT	Generalized Fourier transform
GPSF	Generalized Poisson summation formula
GPU	Graphics processing unit
HOA	Higher order ambisonics
IFFT	Inverse fast Fourier transform
IGFFT	Inverse generalized fast Fourier transform
LCC	Left-angle circular covolution
LTI	Linear time-invariant
MISM	Mirror image source method
MMIMO	Massive multiple-input multiple-output
PAF	Plenacoustic function
PAS	Plenacoustic spectrum
PC	Personal computer
PML	Perfectly matched layers
PSTD	Pseudo-spectral time-domain
RCC	Right-angle circular covolution
RIR	Room impulse response
ROC	Region of convergence

SRF Sound field reconstruction

WDAF Wave-domain adaptive filtering

WFA Wave field analysis

WFS Wave field synthesis

WRW Wave propagation wave Reflection Wave propagation

WSN Wireless sensor network

Notation

$\mathbb{R}$ The set of real numbers

$\mathbb{Z}$ The set of integer numbers

$\mathbb{C}\backslash\{0\}$ The set of complex numbers not including 0

$\mathbb{C}$ The set of complex numbers

$\mathbb{R}^\nu$ The ν-dimensional real space

$\mathbb{Z}^\nu$ The ν-dimensional integer lattice

j Imaginary number. $j = \sqrt{-1}$

$\mathbb{C}^\nu$ The ν-dimensional complex space

$(\cdot)^T$ Vector or matrix transposition

$\mathbf{x} \in \mathbb{R}^\nu$ Spatial position in ν dimensions. When $\nu = 3$, then $\mathbf{x} \in [x, y, z]^T$

$||\cdot||$ Euclidean norm.

$\cdot$ Dot product.

$|x|$ Absolute value of scalar x

$\lvert\mathbf{\Lambda}\rvert$	Absolute value of determinant of matrix $\mathbf{\Lambda}$
c	Speed of sound in meters per second
$t \in \mathbb{R}$	Time in seconds
$\omega \in \mathbb{R}$	Temporal angular frequency. It denotes frequency in radians per second. In Chap. 4 and Chap. 5, it denotes discrete-time frequency in radians per sample
$\boldsymbol{\phi} \in \mathbb{R}^\nu$	Spatial angular frequency vector. When $\nu = 3$ then $\boldsymbol{\phi} = [\phi_x, \phi_y, \phi_z]^T$
$\Omega \in \mathbb{R}$	Temporal angular frequency in radians per second (Chap. 4)
$\mathbf{k}$	Wave vector. $\mathbf{k} = [k_x, k_y, k_z]^T$
k_x, k_y, k_z	Trace wave numbers in the x, y and z directions respectively
$k \in \mathbb{R}$	Wave number. $k = \lvert\lvert\mathbf{k}\rvert\rvert = \lvert\omega\rvert/c$ (Chap. 1)
$k \in \mathbb{Z}$	Discrete temporal-frequency index (Chap. 4 and Chap. 5)
$n \in \mathbb{Z}$	Discrete time index
$\mathbf{n} \in \mathbb{Z}^\nu$	ν-dimensional discrete index
$\mathbf{k} \in \mathbb{Z}^\nu$	ν-dimensional spectral discrete index
$p(\mathbf{x}, t)$	Scalar field (scalar function of several variables)
$\mathbf{v}(\mathbf{x}, t)$	Vector field (vector function of several variables)
$x(n)$	Discrete-time scalar function or signal
$\boldsymbol{\eta} \in \mathbb{R}^3$	Outward normal vector with respect to a surface
$\boldsymbol{\eta}(\mathbf{x}) \in \mathbb{R}^3$	Outward normal vector function with respect to a surface
∇^2	Laplacian operator
∇	Gradient operator (vector derivative of a scalar field)
$\mathbf{0} \in \mathbb{R}^\nu$	Zero vector. $\mathbf{0} = [0, \ldots, 0]^T$
ϑ	Dip (elevation) angle

φ	Azimuth angle
$\varrho \in \mathbb{C}$	Wall reflection coefficient
$Z \in \mathbb{C}$	Wall impedance
$Z(\mathbf{x}, \omega) \in \mathbb{C}$	Wall impedance as a function of wall surface and temporal-frequency
$\xi \in \mathbb{C}$	Specific wall admittance. $\xi = \varrho_0 c/Z$
ρ_0	Medium density in $\mathrm{kg/m}^3$
$\mathcal{S}$	Set of space points comprising a surface
$\mathcal{V}$	Set of space points comprising a volume
$\frac{d^n}{dt^n}$	n-th order derivative of a function of t
$\frac{\partial^n}{\partial x^n}$	n-th order partial derivative of a scalar field with respect to variable x
$\delta(\cdot)$	Dirac's delta generalized function
$\psi(\cdot)$	Basis function or eigen function
$\mathcal{O}(N)$	Complexity order as a function of parameter N
T_{60}	Reverberation time in seconds. Time it takes for the reverberation energy to decay 60dB below its initial value
$\mathbf{\Lambda}$	Generator matrix of lattice Λ
Λ	Lattice Λ
$\mathbf{\Lambda}^{-1}$	Inverse of matrix $\mathbf{\Lambda}$
$\mathbf{\Lambda}^{-T}$	Inverse transpose of matrix $\mathbf{\Lambda}$. $\mathbf{\Lambda}^{-T} = (\mathbf{\Lambda}^{-1})^T = (\mathbf{\Lambda}^T)^{-1}$
$V_\Delta(\mathbf{x})$	Voronoi region of lattice Δ around point $\mathbf{x}$
$\mathrm{N}(\Delta/\Gamma)$	Number of lattice points of lattice Γ that lie inside $V_\Delta(\mathbf{0})$ the voronoi region of lattice Δ around the origin
$\mathbf{I}$	Identity matrix

$\mathrm{diag}(\mathbf{x})$	Diagonal matrix with diagonal elements equal to the elements of vector $\mathbf{x}$
$L^2(\mathbb{R}^\nu)$	Lebesgue 2-norm function space over the ν-dimensional real space. The space of (multidimensional) functions of finite energy
$L^2[-\pi,\pi]$	Lebesgue 2-norm function space on the interval $[-\pi,\pi]$. The space of functions of finite energy defined on the interval $[-\pi,\pi]$
$*$	Linear convolution operator
$\circledast$	Multidimensional linear convolution operator
$\Re\{\cdot\}$	Real part operator
$\Im\{\cdot\}$	Imaginary part operator
$x_\Re$	Real part of x
$x_\Im$	Imaginary part of x
$(\cdot)^*$	Complex conjugate
O_q	qth octact of the 3-D space $\mathbb{R}^3$
$H(x)$	Heaviside function (unit-step function)
$H_q(\mathbf{x})$	3-D Heaviside function. $H_q(\mathbf{x}) = 1$ if $\mathbf{x} \in \mathrm{O}_q$, and 0 otherwise
$\otimes$	Cartesian product operator
$x \mathrm{mod} N$	x modulo N
$(\cdot)_N$	Modulo N operator
$\mathcal{F}\{\cdot\}$	Fourier transform operator
$\mathcal{F}^{-1}\{\cdot\}$	Inverse Fourier transform operator
$\mathcal{F}_t\{\cdot\}$	Fourier transform operator over the specified subscripted dimension(s)
$\mathcal{F}_\omega^{-1}\{\cdot\}$	Inverse Fourier transform operator over the specified subscripted dimension(s)

P	Fourier transform of p
$\xleftrightarrow{\mathcal{F}}$	Fourier transform relation
$\xleftrightarrow{\mathcal{F}_{\mathbf{x}}}$	Fourier transform relation over the specified subscripted dimension(s)
$\breve{P}$	Angular spectrum or wave-domain representation of scalar field p
$\mathcal{G}_\alpha\{\cdot\}$	Generalized Fourier transform operator with parameter α
$\mathcal{G}_\alpha^{-1}\{\cdot\}$	Inverse generalized Fourier transform operator with parameter α
P_α	Generalized Fourier transform with parameter α of p
$\xleftrightarrow{\mathcal{G}_\alpha}$	Generalized Fourier transform relation
$\tilde{p}_\alpha$	Geometrically weighted periodic extension of p
$\alpha \in \mathbb{C}\backslash\{0\}$	Parameter of the generalized Fourier transform
$\beta \in \mathbb{C}$	Parameter of the generalized Fourier trasform. For a specified periodicity period $T_p \in \mathbb{R}$, or $N \in \mathbb{N}$ for discrete-time signals, then $\beta = \log(\alpha)/T_p$ or $\beta = \log(\alpha)/N$
$\boldsymbol{\alpha} \in \mathbb{C}^\nu$	Parameter of the generalized Fourier transform for multidimensional functions. $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{\nu-1}]^T$ with $\alpha_i \neq 0$ for all i
$\boldsymbol{\beta} \in \mathbb{C}^\nu$	Parameter of the generalized Fourier transform for multidimensional functions. For a specified periodicity lattice Λ with generator matrix $\mathbf{\Lambda}$, then $\boldsymbol{\beta} = \mathbf{\Lambda}^{-T}[\log(\alpha_0), \ldots, \log(\alpha_{\nu-1})]^T$.
$\mathcal{H}_x\{\cdot\}$	Hilbert transform operator over the specified subscripted dimension(s)

Always keep Ithaca in your mind.
To arrive there is your ultimate goal.
But do not hurry the voyage at all.
It is better to let it last for many years;
and to anchor at the island when you are old,
rich with all you have gained on the way,
not expecting that Ithaca will offer you riches.

Ithaca has given you the beautiful voyage.
Without her you would never have set out on the road.
She has nothing more to give you.
And if you find her poor, Ithaca has not deceived you.
Wise as you have become, with so much experience,
you must already have understood what Ithaca means.

Fragment from "Ithaca"
CONSTANTINE CAVAFY (1863–1933)
translated by Rae Dalven.