PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

SPATIAL AUDIO: MAXIMIZING THE PSYCHO-ACOUSTIC SWEET SPOT

PEDRO IZQUIERDO LEHMANN

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
CARLOS A. SING LONG

Santiago de Chile, January 2022

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

SPATIAL AUDIO: MAXIMIZING THE PSYCHO-ACOUSTIC SWEET SPOT

PEDRO IZQUIERDO LEHMANN

Members of the Committee:

CARLOS A. SING LONG — DocuSigned by: Carlos Sing Long 3574ABD6A96E44F...
RODRIGO CÁDIZ — DocuSigned by: Rodrigo Cádiz E0AFA235D0AD428...
CARLOS PÉREZ — DocuSigned by: Carlos Perez Arancibia 757070840AA9493...
MARCELO GONZÁLEZ — DocuSigned by: Marcelo González 9C492EACB23C4A9...

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, January 2022

To my father Luis Izquierdo W.,

whom I admire.

ACKNOWLEDGEMENTS

My greatest gratitude goes to my advisor, professor Carlos Sing Long. He has been not only an exceptional teacher, but also a wise guide. I truly appreciate his approach to applied mathematics: understanding it not as an isolated logic-derived system nor as just a convenient "plug-and-play" tool to solve problems but as a language to pursue the creation of models that organize the real world. His constant encouragement to look further from the "easy way" and to face the research problems from their first principles has been fundamental to this work.

I want to thank our collaborator, professor Rodrigo Cádiz for his insights over the field of audio and acoustics; without his support this project would not even exist. Also, I want to thank to Julius O. Smith because of the invaluable conversation that we had in February 2020 at Stanford, a seed that latter would evolve into the questions that we tackle in this work.

Last but not least I want to thank to my friends and family, their infinite support and interest shown over my research. They have accompanied me along this journey.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

LIST OF FIGURES vii

LIST OF TABLES ix

ABSTRACT x

RESUMEN xi

1. INTRODUCTION 1
   1.1. Stereophony . . . 3
   1.2. Sound field reconstruction . . . 4
      1.2.1. The Wave Equation . . . 4
      1.2.2. Mode Matching Methods . . . 11
      1.2.3. Pressure Matching Methods . . . 14
      1.2.4. Wave Field Synthesis . . . 16
      1.2.5. Why sound field reconstruction is not the best approach? . . . 18
   1.3. Psycho-acoustic approaches . . . 19
   1.4. Contributions . . . 20
   1.5. Structure of the thesis . . . 21

2. MAXIMIZING THE PSYCHO-ACOUSTIC SWEET SPOT 22
   2.1. Mathematical model . . . 22
      2.1.1. Acoustic framework . . . 22
      2.1.2. Psycho-acoustic preliminaries . . . 23
   2.1.2.1. The hearing threshold . . . 23
   2.1.2.2. The damage and discomfort risk threshold . . . 25

2.1.3. Psycho-acoustic framework 26
2.2. Maximizing the sweet-spot 29
2.2.1. The layer-cake representation 31
2.2.2. The variational problem 33
2.3. DC Formulation 34
2.3.1. SWEET algorithm 35
2.3.2. SWEET-ReLU algorithm 39
2.4. Implementation 41
2.5. Experiments 45
2.5.1. Comparison with state-of-the-art methods 46
2.5.2. Applications 53
2.5.2.1. The effect of multiple frequencies 53
2.5.2.2. Multiple zone control 55
2.6. Discussion 56
2.7. Conclusion 58

REFERENCES 59

APPENDIX 69
A. Proof of Proposition 1 70
B. Proof of Proposition 8 72

LIST OF FIGURES

1.1 Sound scenes, auditory scenes, and sweet spot in spatial sound. 1

1.2 Sound field reconstruction settings: target region Ω, loudspeakers, and target sound wave u_0. 2

1.3 Cartesian and spherical coordinates. 9

2.1 Threshold of audibility. 24

2.2 Interpolation of the loud discomfort levels for sinusoidal signals given in (Knobel & Sanchez, 2006) by cubic splines. 44

2.3 Comparison with state-of-the-art methods. *Columns:* (i) Variation of the distance between the source and the center of the arrangement, (ii) Variation of the frequency of the source. *Rows:* (i) Sweet spot as a fraction of Ω, (ii) Average value for $\log(1 + Tu)$, (iii) Average IDE. 48

2.4 Experimental setup. (a) Near field setup. (b) Focus source setup. (c) Multiple zone setup. 49

2.5 u_0 for near field setup. 49

2.6 Near field instance. *Rows:* (i) u, (ii) $\log(1 + Tu)$, (iii) IDE. *Columns:* (i) SWEET-ReLU, (ii) NFC-HOA, (iii) WFS, (iv) L^2-PMM. 50

2.7 u_0 for focus source setup. 51

2.8 Focus source instance. *Rows:* (i) u, (ii) $\log(1 + Tu)$, (iii) IDE, (iv) Sound level (dB). *Columns:* (i) SWEET-ReLU, (ii) NFC-HOA, (iii) WFS, (iv) L^2-PMM. . 52

2.9 The effect of multiple frequencies: $\log(1 + Tu)$ multi frequencies. 53

2.10 The effect of multiple frequencies. *Rows:* (i) u_0, (ii) u multi frequencies, (iii) u single frequencies, (iv) $\log(1 + Tu)$ single frequencies. *Columns:* (i) 100 Hz, (ii) 200 Hz, (iii) 300 Hz, (iv) 400 Hz. 54

2.11 u_0 for multiple zone setup. 55

2.12 Multiple zone control. *Rows:* (i) SWEET, (ii) L^2- PMM. *Columns:* (i) u, (ii) $\log(1 + Tu)$, (iii) IDE. 56

LIST OF TABLES

2.1 Sweet spots as a fraction of Ω in Near field (NF) and Focus Source (FS) instances. 47

ABSTRACT

Sound field reconstruction (SFR) is a popular approach to recreate an auditory scene over a region of interest using an arrangement of a few loudspeakers. Methods following this approach typically aim to approximate the sound wave that created the desired auditory scene by minimizing physically inspired error metrics such as the L^2-norm. However, these metrics do not account for *psycho-acoustic effects*. Hence, the auditory artifacts generated by these methods may be physically small, but psycho-acoustically large. Although there are methods that incorporate psycho-acoustic principles, we believe there is still a gap between them and the SFR approaches: the link between the precise control of the reconstructed sound wave and the psycho-acoustic effectiveness of it has not been correctly developed yet.

In order to fill the gap, in this work we define a *sweet spot* that comprises the region where the generated sound wave is *psycho-acoustically close* to the desired auditory scene. Then, we develop a method that aims to generate a sound wave that directly maximizes this sweet spot. Our method incorporates psycho-acoustic principles from the onset and is flexible: while it imposes little to no constraints on the regions of interest, the arrangement of speakers, and the radiation pattern of the loudspeakers, it allows for a wide array of psycho-acoustic models that include state-of-the-art monaural psycho-acoustic models. Our method leverages tools from analysis and optimization that allow for its mathematical analysis and efficient implementation. Our numerical results show that our method yields larger sweet spots compared to state-of-the-art SFR methods for sinusoidal point sources using van de Par's psycho-acoustic model.

Keywords: Spatial sound, sweet spot, psycho-acoustics, non-convex DC optimization.

RESUMEN

Sound field reconstruction (SFR) es un enfoque popular para recrear una escena auditiva en una región de interés utilizando una cantidad finita de parlantes. Los métodos que siguen este enfoque suelen tener como objetivo aproximar a la onda sonora que creó la escena auditiva deseada minimizando métricas de error inspiradas físicamente, como la norma L^2. Sin embargo, estas métricas no tienen en cuenta *efectos psicoacústicos*. Por lo tanto, los artefactos auditivos generados por estos métodos pueden ser físicamente pequeños, pero psicoacústicamente grandes. Aunque existen métodos que incorporan principios psicoacústicos, creemos que aún existe una brecha entre ellos y los enfoques SFR: el vínculo entre el control preciso de la onda sonora reconstruida y la efectividad psicoacústica de la misma no se ha desarrollado correctamente todavía.

Para cerrar la brecha, en este trabajo definimos un *sweet spot* que comprende la región donde la onda sonora generada es *psicoacústicamente cercana* a la escena auditiva deseada. Luego, desarrollamos un método cuyo objetivo es generar una onda sonora que maximice directamente este *sweet spot*. Este es flexible: impone pocas o ninguna restricción en las regiones de interés, sobre la disposición espacial y el patrón de radiación de los parlantes, y permite la aplicación de una amplia familia de modelos psicoacústicos monoaurales, en particular de aquellos del estado del arte. Además, utiliza herramientas de análisis matemático y optimización que permiten su interpretación e implementación eficiente. Nuestros resultados numéricos muestran que nuestro método produce *sweet spots* más grandes que los métodos de SFR del estado del arte para fuentes puntuales sinusoidales bajo el modelo psicoacústico de van de Par.

Palabras Claves: Audio espacial, sweet spot, psicoacústica, optimización DC.

1. INTRODUCTION

> *"What could be meant by copying a fact would be hard to grasp even if there were any such things as facts."*
> Nelson Goodman

The field of spatial sound addresses the question: *how do we create a desired* auditory scene *over a spatial region of interest from a* sound scene *generated with a finite set of loudspeakers?* In this context, the *sound scene* represents the objective nature of a sound wave propagating in the physical world, whereas the *auditory scene* represents the imprint of the sound scene in our subjectivity, that is, the result of the auditory system perceiving and organizing sound into meaning (Spors et al., 2013; Blauert, 1997).

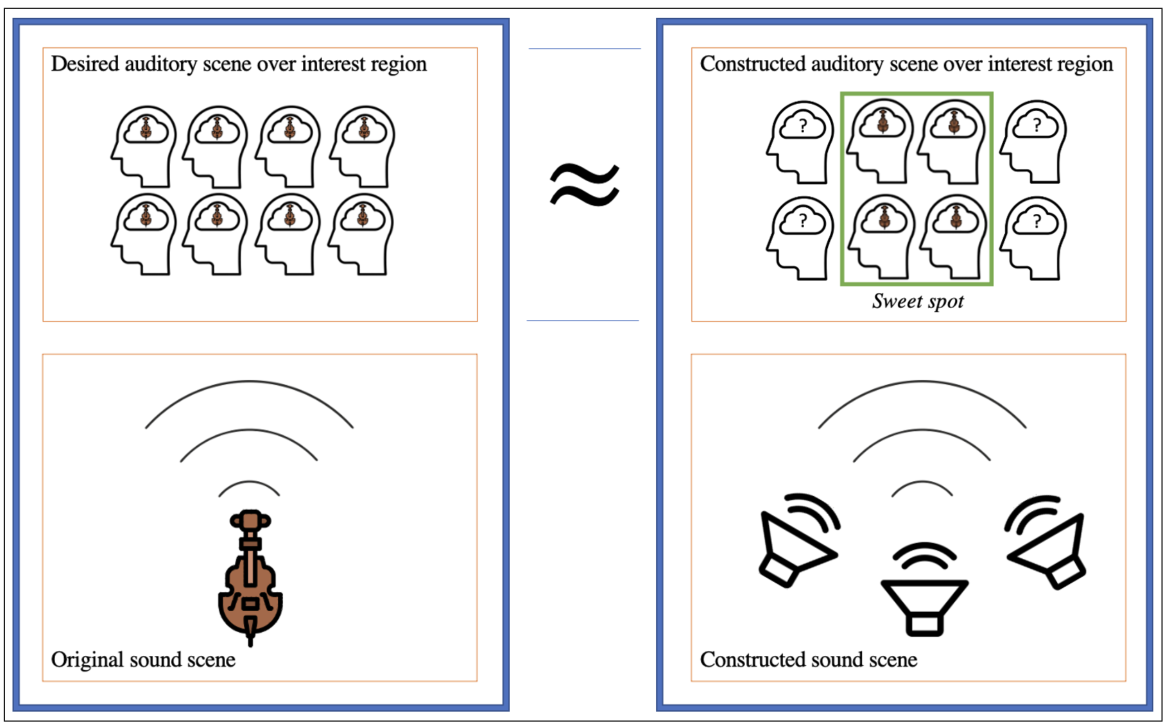


Figure 1.1. Sound scenes, auditory scenes, and sweet spot in spatial sound.

Over the last century, several methods have been proposed to answer this question. They can be divided in three groups: i) simple/heuristic methods that implicitly exploit psycho-acoustic features, e.g. stereophony, (Hacihabiboglu, De Sena, Cvetkovic, Johnston, & Smith III, 2017); ii) methods that pursue to reconstruct the sound scene, i.e. sound field reconstruction methods (also called *sound field synthesis*) (Spors et al., 2013); iii) methods that explicitly exploit psycho-acoustic features (Ziemer, 2020).

The performance of the methods can be compared in terms of the size of the region where the sound scene creates an auditory scene that most closely resembles the desired one. In this work, we call this region the *sweet spot*. It should be noted that in this last definition we considered the auditory scene as a feature of the points of the region of interest. However, our definition of an auditory scene is only properly clarified as a feature of the subjectivity of a person. To transfer the definition from the person to the region of interest, we must think the region of interest as a place for potential listeners (see Fig. 1.1); the auditory scene is correctly recreated in a point of the region of interest when the auditory scene of any potential listener lying over the point is correctly recreated.

For the description of the methods, we consider a region of interest Ω of arbitrary shape, a set of loudspeakers located at $x_1, \ldots, x_{n_s} \in \mathbb{R}^3$, and a target sound pressure wave u_0, as depicted in Fig. 1.2.

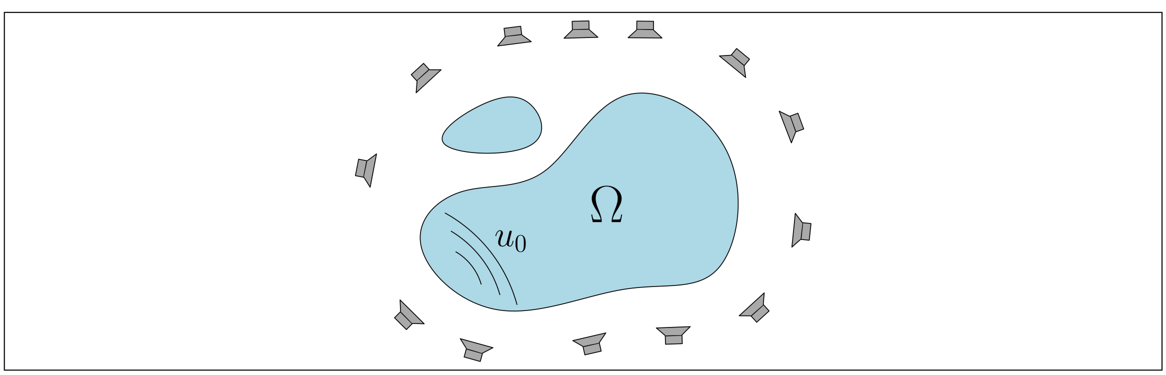


Figure 1.2. Sound field reconstruction settings: target region Ω, loudspeakers, and target sound wave u_0.

1.1. Stereophony

One of the earlier and most widespread spatial sound approaches is stereophony (Union, 2012; Lipshitz, 1986). Its origins go back to 1931, when Alan Blumein introduced the first two-loudspeaker sound system (Alexander, 2013). But its influence is still current: modern stereophonic systems as Dolby 5.1 are widely used in commercial products. These methods, also called *panning techniques*, adjust the level and time-delay of the audio signals for each speaker utilizing a *panning law* to simulate steering the perceived direction of the sound source.

For example, Vector Base Amplitude Panning (VBAP) (Pulkki, 1997, 2001) is a well known amplitude based panning system that can reproduce 3D spatial sounds on sphere-like settings. It is defined as follows: let the origin of the coordinate system be the center of the target region Ω, the loudspeakers $x_1, \ldots, x_{n_s} \in \partial B(0,1)$ and u_0 produced by a sound source located at $x_0 \in \partial B(0,1)$. Then, select the positions of the loudspeakers that optimally encloses x_0 as $x_1^\star, x_2^\star, x_3^\star$ and define $L = \begin{bmatrix} x_1^\star & x_2^\star & x_3^\star \end{bmatrix} \in \mathbb{R}^{3\times 3}$. Assuming that L is invertible, the VBAP system reproduces the time signal of the source at the loudspeakers located at $x_1^\star, x_2^\star, x_3^\star$ with gains equal to

$$g = (g_1, g_2, g_3)^T = \frac{L^{-1}x_0}{\|L^{-1}x_0\|_2}$$

Due to psycho-acoustic features of the auditory system such as the binaural decoloration mechanism, stereophonic systems work sufficiently well in some applications, even with few speakers (Spors et al., 2013). However, they can only simulate sound sources that lay approximately on the surface of the convex hull of the speakers. Furthermore, its quality degrades rapidly as the listener moves away from the center of the target region (Spors et al., 2013). Hence, the auditory scenes they can reproduce accurately are limited, and their sweet spot seems to be very localized.

1.2. Sound field reconstruction

A more ambitious strategy to recreate an auditory scene is to directly approximate the sound wave that created it. In the literature, this strategy is called *sound field reconstruction* (or *sound field synthesis* (Spors et al., 2013)) and, in this context, the sweet spot is assumed to be the same as the region where the generated sound wave closely resembles the target sound wave.

Following Huygens' principle, any sound scene can be approximated accurately with a sufficiently dense arrangement of loudspeakers. However, selecting the audio signals for the loudspeakers is an *ill-conditioned* problem (Fazi & Nelson, 2007a), that is, large changes in the loudspeaker signals do not necessarily produce large changes in the rendered sound wave and therefore small changes in the rendered sound wave may imply large changes in the loudspeaker signals. Moreover, when the problem is tackled from a continuous perspective (with a continuum of loudspeakers surrounding the target region), there might be multiple solutions, rendering the problem *ill-posed* (Fazi & Nelson, 2012). This means that the reconstruction problem may not be easy to solve; in some cases the reconstruction will be prone to numerical issues.

Three classes of commonly used methods for sound field reconstruction are *mode matching methods*, *pressure matching methods* and *wave field synthesis*. All of them rely on an ondulatory physical description of sound, and over the wave equation that can be derived from it, which for completeness we present before the exposition of the latter.

1.2.1. The Wave Equation

The simplest model of sound propagation over fluids rely over the linearization of three principles (Pierce, 2019, Chapter 1.6): the conservation of mass, Newton's force equation and a pressure-density relation.

(i) **The conservation of mass**: for a fixed volume Ω, the variation of net mass in a volume Ω at any time can be expressed as the integral over Ω of the time derivative of its density ρ:

$$\int_{\Omega} \frac{\partial}{\partial t}\rho(\mathrm{x}, t)\,\mathrm{dx}. \tag{1.1}$$

Also, the entrance and exit of mass through the boundary of Ω can be expressed as the integral over the boundary of the density times the velocity of the flow $\vec{v}$ through the boundary:

$$\int_{\partial\Omega} \rho(\mathrm{x}, t)\vec{v}(\mathrm{x}, t)\cdot n\,\mathrm{d}S(\mathrm{x}) \tag{1.2}$$

Then, the conservation of mass implies that the variation of the net mass over Ω needs to be equal to the entrance/exit of mass by its boundary:

$$\begin{aligned}\int_{\Omega} \frac{\partial}{\partial t}\rho(\mathrm{x}, t)\,\mathrm{dx} &= \int_{\partial\Omega} \rho(\mathrm{x}, t)\vec{v}(\mathrm{x}, t)\cdot n\,\mathrm{d}S(\mathrm{x}) \\ &= -\int_{\Omega} \nabla\cdot(\rho(\mathrm{x}, t)\vec{v}(\mathrm{x}, t))\,\mathrm{dx},\end{aligned}$$

where the last equation follows from Gauss' theorem (Apostol, 1969, Chapter 12.19). Then, by the linearity of the integral,

$$\int_{\Omega} \left(\frac{\partial}{\partial t}\rho(\mathrm{x}, t) + \nabla\cdot(\rho(\mathrm{x}, t)\vec{v}(\mathrm{x}, t))\right)\,\mathrm{dx} = 0. \tag{1.3}$$

Finally, since the latter is true for any arbitrary Ω, it implies the conservation of mass equation

$$\frac{\partial}{\partial t}\rho(\mathrm{x}, t) + \nabla\cdot(\rho(\mathrm{x}, t)\vec{v}(\mathrm{x}, t)) = 0 \tag{1.4}$$

(ii) **Newton's force equation**: the net force over over the fluid inside a fixed volume Ω can be expressed as the integral of the surface forces per unit area, denoted by

$\vec{f}_S$, exerted by the immediate surroundings of Ω:

$$\int_{\partial\Omega} \vec{f}_S(\mathrm{x}, t)\ \mathrm{d}S(\mathrm{x}), \tag{1.5}$$

plus the integral of the the body forces per unit volume, denoted by $\vec{f}_B$, exerted by non local forces such as gravity from the very outside of Ω:

$$\int_{\Omega} \vec{f}_B(\mathrm{x}, t)\ \mathrm{dx}. \tag{1.6}$$

Although gravity is a fundamental force that is always present, it can be taken as negligible in the context of acoustic disturbance phenomena at all but very low (and inaudible!) frequencies (Pierce, 2019, Chapter 1.3). Also, a classical assumption for $\vec{f}_S$ (Pierce, 2019, Chapter 1.3) is that it is directed normally into the surface $\partial\Omega$, becoming

$$\vec{f}_S = -u(\mathrm{x}, t)\vec{n}, \tag{1.7}$$

where u is called the *pressure* of the fluid and $\vec{n}(\mathrm{x})$ is the normal vector of Ω at x pointing outwardly. On the other hand, the time variation of the net momentum of the particles inside Ω can be expressed as the material derivative of the vectorized integral of the density times the velocity of the fluid:

$$\frac{d}{dt}\int_{\Omega} \rho(\mathrm{x}, t)\vec{v}(\mathrm{x}, t)\ \mathrm{dx}. \tag{1.8}$$

Then, Newton's force equation, i.e. the net force is equal to the derivative of the momentum, gives

$$\begin{aligned}
-\int_{\Omega} \nabla u(\mathrm{x},t)\ \mathrm{dx} &= \int_{\partial\Omega} \vec{f}_S(\mathrm{x},t)\ \mathrm{d}S(\mathrm{x}) && \text{(Gauss' theorem)} \\
&= \frac{d}{dt}\int_{\Omega} \rho(\mathrm{x}(t),t)\vec{v}(\mathrm{x}(t),t)\ \mathrm{dx} && \text{(Newton's equation)} \\
&= \int_{\Omega} \frac{d}{dt}\left(\rho(\mathrm{x}(t),t)\vec{v}(\mathrm{x}(t),t)\right)\ \mathrm{dx} && \text{(Regularity of } \rho\vec{v}) \\
&= \int_{\Omega} \rho(\mathrm{x},t)\frac{\partial}{\partial t}\vec{v} + (\vec{v}\cdot\nabla)\vec{v}\ \mathrm{dx} && \text{(Chain rule)}
\end{aligned} \tag{1.9}$$

Finally, since the latter is true for any arbitrary Ω, it implies Euler's equation for fluids

$$\nabla u(\mathrm{x},t) + \rho(\mathrm{x},t)\frac{\partial}{\partial t}\vec{v} + (\vec{v}\cdot\nabla)\vec{v} = 0 \tag{1.10}$$

(iii) **Pressure-density relation**: In modern physics, the pressure u of a fluid in a specific moment and position over a region Ω can be regarded as a function of the specific density ρ and the specific entropy s (Pierce, 2019, Chapter 1.4):

$$u = u(\mathrm{x}, t, \rho, s). \tag{1.11}$$

Over this, we will utilize the common assumption that the specific entropy of a fluid particle is considered as constant in time, i.e. the material derivative is equal to zero:

$$\frac{d}{dt}s(\mathrm{x},t) = 0. \tag{1.12}$$

The latter implies that the heat flow over Ω is negligible (Pierce, 2019, Chapter 1.4). Also, together with the assumptions that s is initially constant over Ω and that the state equations are all equal over Ω, we have $s(\mathrm{x},t) \equiv s_0$ and then

$$u = u(\mathrm{x}, t, \rho, s_0) = u(\mathrm{x}, t, \rho). \tag{1.13}$$

In practice, the pressure, velocity and density of a fluid in the acoustic disturbance processes can be expressed as the sum of an ambient state $(u_0, \vec{v}_0, \rho_0)$, related to the medium over which the acoustic disturbances occur, plus a perturbation $(u', \vec{v}', \rho')$:

$$u = u_0 + u', \quad \vec{v} = \vec{v}_0 + \vec{v}', \quad \rho = \rho_0 + \rho'. \tag{1.14}$$

We assume that the medium is *homogeneous*, i.e. the ambient state is independent of the position over Ω; and *quiescent*, i.e. the ambient state is independent of time and $\vec{v}_0 \equiv 0$. Then, equations (1.4), (1.10), and (1.13) become

$$0 = \frac{\partial}{\partial t}(\rho_0 + \rho') + \nabla \cdot ((\rho_0 + \rho')\vec{v}') \tag{1.15}$$

$$0 = \nabla(u_0 + u') + (\rho_0 + \rho')\frac{\partial}{\partial t}\vec{v}' + (\vec{v}' \cdot \nabla)\vec{v}' \tag{1.16}$$

$$u(\rho_0 + \rho', s_0) = u_0 + u' = u_0 + \left(\frac{\partial u}{\partial \rho}\right)_{\rho_0} \rho' + \frac{1}{2}\left(\frac{\partial^2 u}{\partial^2 \rho^2}\right)_{\rho_0} (u')^2 + \dots, \tag{1.17}$$

where in (1.17) we have expanded the last expression into its Taylor series around ρ_0. Then, neglecting all the non linear terms we get the *linear acoustic equations*

$$0 = \frac{\partial}{\partial t}\rho' + \rho_0 \nabla \cdot \vec{v}' \tag{1.18}$$

$$0 = \nabla u' + \rho_0 \frac{\partial}{\partial t}\vec{v}' \tag{1.19}$$

$$u' = c^2 \rho', \quad c^2 = \left(\frac{\partial u}{\partial \rho}\right)_{\rho_0}, \tag{1.20}$$

where $\left(\frac{\partial u}{\partial \rho}\right)_{\rho_0}$ is taken positive for thermodynamic considerations (Pierce, 2019, Chapter 1.5). In the following, for simplicity, we will omit the superscript for the perturbed variables. Then, substituting (1.20) into (1.18) we have

$$\begin{aligned}
& 0 = \frac{1}{c^2}\frac{\partial}{\partial t}u + \rho_0 \nabla \cdot \vec{v} \\
\Rightarrow\ & 0 = \frac{1}{c^2}\frac{\partial^2}{\partial^2 t}u + \rho_0 \frac{\partial}{\partial t}(\nabla \cdot \vec{v}) && \text{(Partial differentiation in time)} \\
\Leftrightarrow\ & 0 = \frac{1}{c^2}\frac{\partial^2}{\partial^2 t}u - \nabla \cdot \left(\rho_0 \frac{\partial}{\partial t}\vec{v}\right) && \text{(Regularity of } \vec{v}\text{)} \\
\Rightarrow\ & 0 = \frac{1}{c^2}\frac{\partial^2}{\partial^2 t}u - \nabla^2 u && \text{(Using (1.19)),}
\end{aligned} \tag{1.21}$$

which is the so called *wave equation*. Taking a time Fourier transform on it and fixing a specific angular frequency ω leads to the *Helmholtz equation*,

$$0 = k^2 \widehat{u} + \nabla^2 \widehat{u}, \tag{1.22}$$

where $k = \omega/c$ is called the *wave number* and $u = u(\mathrm{x}, \omega) = u(\mathrm{x})$ because ω is fixed. If we rewrite (1.21) in terms of the spherical coordinates, see Fig. 1.3,

$$\begin{aligned} x &= r \sin\theta \cos\phi \\ y &= r \sin\theta \sin\phi \\ z &= r \cos\theta \end{aligned} \tag{1.23}$$

then we obtain

$$0 = \frac{1}{r} \frac{\partial^2}{\partial^2 r} (ru) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial u}{\partial\theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 u}{\partial^2 \phi} - \frac{1}{c^2} \frac{\partial^2 u}{\partial^2 t}. \tag{1.24}$$

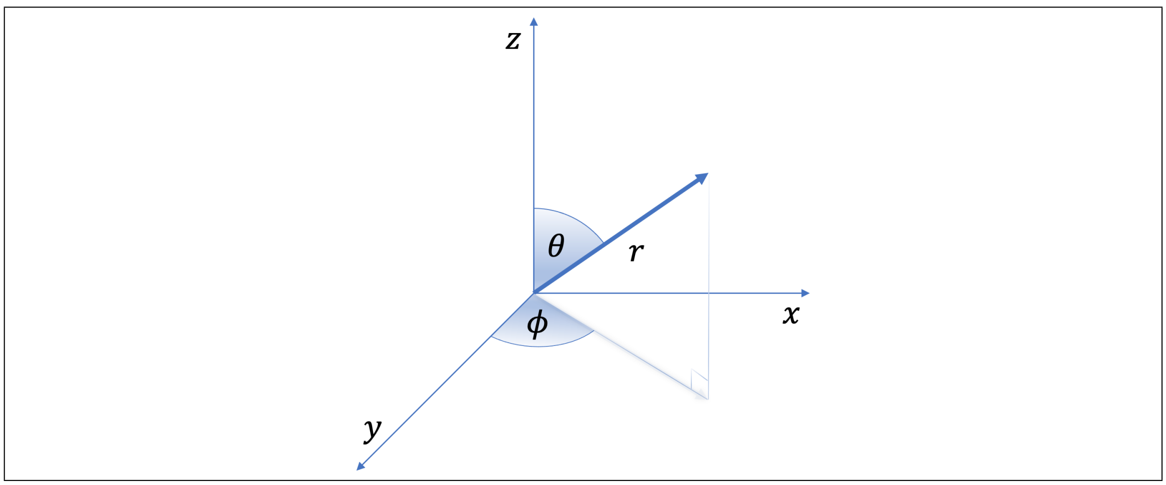


Figure 1.3. Cartesian and spherical coordinates.

If we suppose that u is radially symmetric, i.e. $u = u(r, \theta, \phi, t) = u(r, t)$, then (1.24) simplifies into

$$
\begin{aligned}
& 0 = \frac{1}{r}\frac{\partial^2}{\partial^2 r}(ru) - \frac{1}{c^2}\frac{\partial^2 u}{\partial^2 t} \\
\Leftrightarrow\ & 0 = \frac{\partial^2}{\partial^2 r}(ru) - \frac{1}{c^2}\frac{\partial^2}{\partial^2 t}(ru) && \text{(Multiplying by } r\text{)} \\
\Leftrightarrow\ & 0 = \left(\frac{\partial}{\partial r} - \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial r} + \frac{1}{c}\frac{\partial}{\partial t}\right)(ru) && (a^2 - b^2 = (a-b)(a+b)) \\
\Leftrightarrow\ & 0 = -\frac{4}{c^2}\frac{\partial}{\partial \xi}\frac{\partial}{\partial \eta} h && \left(\xi = t - \frac{r}{c},\ \eta = t + \frac{r}{c},\ h = ru\right) \\
\Leftrightarrow\ & h = g_1\left(t - \frac{r}{c}\right) + g_2\left(t + \frac{r}{c}\right) && \text{(Integrating)} \\
\Leftrightarrow\ & u = \frac{1}{r}g_1\left(t - \frac{r}{c}\right) + \frac{1}{r}g_2\left(t + \frac{r}{c}\right),
\end{aligned}
\tag{1.25}
$$

where g_1 and g_2 are arbitrary functions. This means that the solution to the wave equation in the radially symmetric case is the sum of an outgoing wave g_1 and an incoming wave g_2. But $g_2 \equiv 0$ because of *causality*, i.e. a wave cannot exist in the medium before the excitation that originated it at the origin of the coordinate system. Thus, we conclude that

$$u(t, r) = \frac{1}{r} g\left(t - \frac{r}{c}\right). \tag{1.26}$$

This establishes that a radially symmetric wave propagates outwardly at velocity c decreasing its pressure magnitude as the inverse of the distance from the origin. Then, we can call c the *speed of sound*. In the Fourier domain, (1.26) is equivalent to

$$\widehat{u}(\omega, r) = \widehat{g}(f)\frac{e^{-ikr}}{r}. \tag{1.27}$$

In the latter case, the frequency information of the signal is encapsulated in $\widehat{g}$ whereas the spatial propagation information is contained in e^{-ikr}/r, which is called a *monopole*. The Green function of the wave equation, actually not a function but a generalized function, satisfies by definition (Pierce, 2019, Chapter 4.3.3)

$$-4\pi\delta_0 = \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial^2 t}\right) G. \tag{1.28}$$

In an unbounded domain it turns out (Pierce, 2019, Chapter 4.3.3) that it also satisfies (1.26) with $g = \delta_0/4\pi$, $\widehat{g} = 1/4\pi$. Thus, in the frequency domain it becomes the normalized monopole

$$\widehat{u}(\omega, r) = \frac{e^{-ikr}}{4\pi r}. \tag{1.29}$$

1.2.2. Mode Matching Methods

Mode matching methods find an approximation to the desired sound wave u_0 by matching, usually in an L^2 sense, its oscillatory spatial modes with the ones of the sum of the loudspeakers. The theory for the latter starts by finding a solution to (1.24) using separation of variables, i.e.

$$u(r, \theta, \phi, t) = R(r)\Theta(\theta)\Phi(\phi)T(t). \tag{1.30}$$

Equation (1.30) leads to four ordinary differential equations (Skudrzyk, 2012, Chapter 19.2)

$$\frac{1}{c^2 T}\frac{d^2 T}{d^2 t} = k_1 \tag{1.31}$$

$$\frac{1}{\Phi}\frac{d^2\Phi}{d^2\phi} = k_2 \tag{1.32}$$

$$\frac{r^2}{R}\frac{d^2 R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} + k_1^2 r^2 = k_3 \tag{1.33}$$

$$\sin\theta\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \left(k_2 + k_3\sin^2\theta\right)\Theta = 0, \tag{1.34}$$

where, in principle, k_1, k_2 and k_3 are arbitrary constants. However, in view of (1.31) we choose k_1 as negative, i.e. $k_1 = -(k_1')^2$; in this way we obtain from (1.31) solutions of the form

$$T(t) = T_2 e^{-ik_1' ct} + T_1 e^{ik_1' ct}, \tag{1.35}$$

instead of exponentially time-decaying functions. Moreover, it is clear that the angular frequency of the solution corresponds to $\omega = k_1' c$, and thus $k_1 = k$ in the sense of the Helmholtz equation (1.22). Furthermore, we only keep the part of the solution that evolves *from the past to the future*, which by convention correspond to the first term, giving

$$T(t) = T_1 e^{-i\omega t}. \tag{1.36}$$

In view of (1.32), the positiveness of k_2 is assumed by the same considerations as for k_1: we take $m^2 = k_2$. Furthermore, since Φ informs the azimuthal angle we need $\Phi(\phi) = \Phi(\phi + 2\pi)$. For simultaneously guaranteeing this and the continuity of Φ, we need m to be an integer. In this way we obtain from (1.32) solutions of the form

$$\Phi(\phi) = \Phi_1 e^{-im\phi} + \Phi_2 e^{im\phi}. \tag{1.37}$$

Equation (1.34) can be solved in a procedure (Skudrzyk, 2012, Chapter 19.3) that ultimately leads to the need for $k_3 = n(n+1)$, where n is an integer and solutions of the form

$$\Theta(\theta) = \Theta_1 P_{n,m}(\cos\theta), \tag{1.38}$$

where $P_{n,m}$ is a *Legendre Polynomial* (also called *Legendre Function* (Williams, 1999, Chapter 6.3.2)). Equation (1.33) can be solved by a procedure related to the Bessel equation (Skudrzyk, 2012, Chapter 19.4), that leads to solutions of the form

$$R(r) = R_1 h_n^{(1)}(kr) + R_2 h_n^{(2)}(kr), \tag{1.39}$$

where $h_n^{(1)}(kr) \propto e^{ikr}$ and $h_n^{(2)}(kr) \propto e^{-ikr}$ are *spherical Hankel functions* of the first and second kind, representing incoming and outgoing waves, respectively. Gathering up (1.30), (1.36), (1.37), (1.38) and (1.39) we get that every solution to the 3D wave equation for a fixed angular frequency ω, i.e. to the Helmholtz equation (1.22) can be written as an infinite sum of terms of the form

$$u(\omega, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} D_{n,m}(r, k) Y_{n,m}(\theta, \phi) \tag{1.40}$$

where $D_{n,m}(r, k) = B_{n,m} h_n^{(1)}(kr) + C_{n,m} h_n^{(2)}(kr)$ is called the *harmonic coefficient* of *order* n and *mode* m (Ward & Abhayapala, 2001), and

$$Y_{n,m}(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_{n,m}(\cos\theta) e^{im\phi} \tag{1.41}$$

is called the *order* n and *mode* m *spherical harmonic*. The reason of the omission of first component of (1.37) in (1.41) is that it is superfluous for a complete description of the solution; the spherical harmonics are a complete orthonormal base of $L^2(\partial B(0, 1))$ (Kirsch & Hettlich, 2016, Theorem 2.19), which also explains the normalizing coefficient in (1.41) given by $\sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}}$. Thus,

$$\int_{\partial B(0,1)} Y_{n,m}(\theta, \phi)^* Y_{n',m'}(\theta, \phi)\, \mathrm{d}\theta \mathrm{d}\phi = \begin{cases} 0 & \text{if } n = n' \wedge m = m' \\ 1 & \text{i.o.c.} \end{cases} \tag{1.42}$$

Equations (1.40) and (1.42) imply that, for a fixed radius r, one can encode a known (measured or simulated) sound wave u by taking the inner products

$$D_{n,m}(r, k) = \int_{\partial B(0,1)} u(\omega, r, \theta, \phi) Y_{n,m}(\theta, \phi)\, \mathrm{d}\theta \mathrm{d}\phi \tag{1.43}$$

up to a given order n_0. Then, MMMs find an approximation to a target sound wave by matching the first n_o order harmonic coefficients of the target and generated sound waves (Daniel, 2000). This match is usually done by L^2 minimization as follows: let $\Omega = B(0, r)$, $\{D^{\ell}_{n,m}\}$ the harmonic coefficients of the loudspeaker ℓ, and $\{D^0_{n,m}\}$ the harmonic coefficients of the target sound wave u_0. Since any reconstructed sound wave is a linear combination of the sound waves generated by the loudspeakers and the harmonic representations converge uniformly on every compact set (Kirsch & Hettlich, 2016, Theorem 2.33), the harmonic coefficients of the reconstructed sound wave are equal to the linear combination of the harmonic coefficients of the loudspeakers. Then, a L^2-minimization MMM system of order n_0 searches for the optimal linear combination of the sound waves generated by the loudspeakers by solving the following convex problem

$$\underset{g\in\mathbb{C}^{n_s}}{\text{minimize}} \quad \sum_{n=0}^{n_0}\sum_{m=-n}^{n}\left|D_{m,n}^{0}-\sum_{\ell=1}^{n_s} g_\ell D_{m,n}^{\ell}\right|^2 \quad . \tag{1.44}$$

Some well-known MMMs are Ambisonics (Gerzon, 1973), Higher-Order Ambisonics (HOA), and Near-Field Compensated Ambisonics (NFC-HOA) (Daniel, Moreau, & Nicol, 2003; Poletti, 2005). Ambisonics assumes the loudspeakers emit plane waves and uses only the leading harmonic coefficient, whereas HOA uses a larger but fixed number of coefficients. In contrast, NFC-HOA assumes the loudspeakers are monopoles, or even higher order loudspeakers, (Samarasinghe, Poletti, Salehin, Abhayapala, & Fazi, 2013). Ambisonics, HOA and NFC-HOA, by their construction, are designed for circular or spherical regions of interest. When approximating a plane wave, NFC-HOA create a spherical sweet spot with a radius that is inversely proportional to the frequency of the source (Ward & Abhayapala, 2001). Some variations of these methods consider a weighted mode matching problem to prioritize certain spatial regions (Ueno, Koyama, & Saruwatari, 2019), or a mixed pressure-velocity mode matching problem, exploiting the modes of the velocity to achieve a better accuracy (Zuo, Abhayapala, & Samarasinghe, 2020). Furthermore, MMMs have been analyzed in the limit of a continuum of speakers (Ahrens & Spors, 2008; Fazi, Nelson, Christensen, & Seo, 2008; Wu & Abhayapala, 2009).

1.2.3. Pressure Matching Methods

Instead of using expansions in spatial spherical harmonics, Pressure Matching Methods (PMM) minimize the L^2 spatio-temporal error between the the target and generated sound waves (Kirkeby & Nelson, 1993; Kirkeby, Nelson, Orduna-Bustamante, & Hamada, 1996; Gauthier, Berry, & Woszczyk, 2005). Due to Parseval-Plancherel's identity, this problem can be tackled in the Fourier domain minimizing each frequency separately as follows: let $\{\mathrm{x}_n\}_{n=1}^{n_d} \subseteq \Omega$ a discretization scheme of Ω and $\widehat{u}_\ell(\mathrm{x}, \omega; g_\ell)$ the pressure sound wave of the loudspeaker ℓ at the Fourier domain. Then, a L^2-PMM system searches for the optimal linear combination of the sound waves generated by the loudspeakers by solving the convex problem

$$\underset{g \in \mathbb{C}^{n_s}}{\text{minimize}} \quad \sum_{n=1}^{n_d} \left| \widehat{u}_0(\mathrm{x}_n, \omega) - \sum_{\ell=1}^{n_s} \widehat{u}_\ell(\mathrm{x}_n, \omega; g_\ell) \right|^2, \tag{1.45}$$

where, typically the loudspeakers are modeled as monopoles, i.e.

$$\widehat{u}_\ell(\mathrm{x}, \omega; g_\ell) = g_\ell \frac{e^{-ik\|\mathrm{x}-\mathrm{x}_\ell\|}}{4\pi\|\mathrm{x}-\mathrm{x}_\ell\|}. \tag{1.46}$$

The magnitude of the loudspeaker gains g_ℓ are often penalized by their L^p-norm to mitigate the effects of ill-conditioning or to promote sparse representations of the reconstructed sound wave (Lilis, Angelosante, & Giannakis, 2010; Radmanesh & Burnett, 2013; Gauthier, Lecomte, & Berry, 2017; Jia, Zhang, Wu, & Wang, 2018; Feng, Yang, & Yang, 2018).

Besides PMMs' flexibility, in most cases, its solution can only be found numerically, and the discretization of the region of interest plays an important role: with an uniform discretization only local sound field reconstruction works reasonably well, i.e. when the loudspeakers are sufficiently far from Ω. A modification to PMMs considers an additional term accounting for the error in the velocity of the target and generated sound waves. This modification yields robust results even when matching the pressure and velocity at points that lie only along the boundary of the region of interest (Buerger, Hofmann, & Kellermann, 2018; Buerger, Maas, Löllmann, & Kellermann, 2015; Shin, Nelson, Fazi, & Seo, 2016).

1.2.4. Wave Field Synthesis

Wave Field Synthesis (WFS) solves the reconstruction problem searching a *driving function* $D(\mathrm{x}, \omega)$ according to the single-layer potential (Spors et al., 2013)

$$P(\mathrm{x}, \omega) = \int_{\partial\Omega} D(\mathrm{x}_s, \omega) G(\mathrm{x} - \mathrm{x}_s, \omega)\, \mathrm{d}S(\mathrm{x}_s). \tag{1.47}$$

This representation models the problem as if Ω were covered with a continuum of monopole loudspeakers. Then, practical implementations with non continuum real loudspeakers are treated as a discretization of the potential. To find an explicit form to $D(\mathrm{x}, \omega)$, WFS utilizes the fact that every solution of the 3D wave equation over an open and bounded region Ω free of sources can be represented using Kirchoff-Helmholtz' formula (Williams, 1999, Chapter 8.3)

$$\int_{\partial\Omega} \left[P(\mathrm{x}_s, \omega) \frac{\partial G(\mathrm{x} - \mathrm{x}_s, \omega)}{\partial n} + \frac{\partial P(\mathrm{x}_s, \omega)}{\partial n} G(\mathrm{x} - \mathrm{x}_s, \omega) \right] \mathrm{d}S = \begin{cases} P(\mathrm{x}, \omega) & \text{if } \mathrm{x} \in \Omega \\ \frac{1}{2} P(\mathrm{x}) & \text{if } \mathrm{x} \in \partial\Omega \\ 0 & \text{if } \mathrm{x} \in \overline{\Omega}^c \end{cases} . \tag{1.48}$$

This means that the sound wave inside Ω is fully characterized by its pressure and the normal derivative of its pressure over $\partial\Omega$. If we take Ω as an unbounded region whose boundary is a plane, the Kirchoff-Helmoltz representation can be simplified to the Rayleigh formula (Williams, 1999, Chapter 2.10): for $\mathrm{x} \in \Omega$,

$$P(\mathrm{x}, \omega) = -2 \int_{\partial\Omega} \frac{\partial P(\mathrm{x}_s, \omega)}{\partial n} G(\mathrm{x} - \mathrm{x}_s, \omega)\, \mathrm{d}S(\mathrm{x}_s). \tag{1.49}$$

Then,

$$D(\mathrm{x}, \omega) = -2 \frac{\partial P(\mathrm{x}, \omega)}{\partial n}. \tag{1.50}$$

Unfortunately, for practical reasons it is preferable to have an horizontal *line* of loudspeakers more than *plane*. To transfer the Rayleigh formula to a line setting, WFS utilizes the *stationary phase approximation* (SPA) (Williams, 1999, Chapter 4.6.1), which for functions $f, \phi \in C^2(\mathbb{R}; \mathbb{R})$ such that $\phi' \gg f'$ yields the following approximation

$$\begin{aligned} \int_{\mathbb{R}} f(z) e^{i\phi(z)} \, \mathrm{d}z &\approx f(z^*) e^{i\phi(z^*)} \int_{\mathbb{R}} e^{i\frac{1}{2}\phi''(z^*)(z-z^*)^2} \, \mathrm{d}z \\ &= \sqrt{\frac{2\pi}{|\phi''(z^*)|}} f(z^*) e^{i\phi(z^*) + i\frac{\pi}{4}\operatorname{sgn}(\phi''(z^*))}, \end{aligned} \tag{1.51}$$

where z^* is such that $\phi(z^*)' = 0$. For simplicity, in the following we will assume that $\Omega = \{\mathrm{x} = (x, y, z) \in \mathbb{R}^3 \mid y > 0\}$, and then the Rayleigh formula can be written as

$$P(\mathrm{x}, \omega) = -2 \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\partial P(\mathrm{x}_s, \omega)}{\partial y} G(\mathrm{x} - \mathrm{x}_s, \omega) \, \mathrm{d}z_s \mathrm{d}x_s. \tag{1.52}$$

Traditional WFS, initially developed by Berkhout (Berkhout, de Vries, & Vogel, 1993; Berkhout, 1988), assumes that the desired sound wave is produced by a monopole source, lying at $\mathrm{x}_0 = [x_0, y_0 < 0, 0]$, and applies the SPA over the z-coordinate over $z^* = 0$. This gives (Start, 1996)

$$P(\mathrm{x}, \omega) \approx -2 \int_{-\infty}^{\infty} \sqrt{\frac{2\pi}{ik}} \sqrt{\frac{|\mathrm{x}_s - \mathrm{x}_0| \cdot |\mathrm{x} - \mathrm{x}_s|}{|\mathrm{x}_s - \mathrm{x}_0| + |\mathrm{x} - \mathrm{x}_s|}} \frac{\partial P(\mathrm{x}_s, \omega)}{\partial y} G(\mathrm{x} - \mathrm{x}_s, \omega) \, \mathrm{d}x_s. \tag{1.53}$$

Then,

$$D_{\text{trad}}(\mathrm{x}, \omega) = -2 \sqrt{\frac{2\pi}{ik}} \sqrt{\frac{|\mathrm{x} - \mathrm{x}_0| \cdot |\mathrm{x}_r - \mathrm{x}|}{|\mathrm{x} - \mathrm{x}_0| + |\mathrm{x}_r - \mathrm{x}|}} \frac{\partial P(\mathrm{x}, \omega)}{\partial y}, \tag{1.54}$$

where $\mathrm{x}_r \in \Omega$ is a reference point where the amplitude of the reconstruction is approximately exact when the stationary phase approximation is valid. *Revisited WFS* (Spors, Rabenstein, & Ahrens, 2008) assumes that the desired sound wave is essentially 2D, i.e. $P((x, y, z), \omega) = P((x, y, z'), \omega)$ for all $z, z' \in \mathbb{R}$, and applies the SPA over the z-coordinate and $z^* = 0$. This gives

$$\begin{aligned} P(\mathrm{x}, \omega) &= -2 \int_{\mathbb{R}} \frac{\partial P(\mathrm{x}_s, \omega)}{\partial y} \int_{\mathbb{R}} G(\mathrm{x} - \mathrm{x}_s, \omega) \mathrm{d}z_s \mathrm{d}x_s \\ &\approx -2 \int_{\mathbb{R}} \sqrt{\frac{2\pi |\mathrm{x} - \mathrm{x}_s|}{ik}} \frac{\partial P(\mathrm{x}_s, \omega)}{\partial y} G(\mathrm{x} - \mathrm{x}_s, \omega) \mathrm{d}x_s. \end{aligned} \tag{1.55}$$

Then,

$$D_{\text{rev}}(\mathrm{x}, \omega) = -2\sqrt{\frac{2\pi|\mathrm{x}_r - \mathrm{x}|}{ik}}\frac{\partial P(\mathrm{x}, \omega)}{\partial y}, \tag{1.56}$$

To deal with non planar regions of interest, it was proposed in Revisited WFS to apply the derived driving function for planar boundaries multiplied by a window function (Spors et al., 2008)

$$w(\mathrm{x}) = \begin{cases} 1 & \text{if } \langle \vec{Iu}_0(\mathrm{x}), n(\mathrm{x})\rangle < 0 \\ 0 & \text{elsewhere} \end{cases}, \tag{1.57}$$

where $\vec{Iu}_0(\mathrm{x})$ is the averaged acoustic intensity vector (Pierce, 2019, Chapter 1.11) of u_0 at x and $\vec{n}(\mathrm{x})$ is the normal vector of Ω at x pointing outwardly. The window function activates only the part of the boundary of the interest region that is 'illuminated' by the desired sound wave. When the surface is smooth, convex and its curvature is small compared with the wavelength of the desired sound wave, this planar approximation is considered valid.

Further generalizations and unifications of the theory of WFS, that encompass the *Traditional* and *Revisited* ones, are developed in (Firtha, Fiala, Schultz, & Spors, 2017; Firtha, 2019). It has been shown that the spatial properties of the auditory scene are correctly simulated by WFS and do not depend on the position of the listener over the region of interest (Wierstorf, Raake, & Spors, 2013). However, it suffers from coloration effects due to spatial aliasing artifacts (Wierstorf, Hohnerlein, Spors, & Raake, 2014).

1.2.5. Why sound field reconstruction is not the best approach?

There is extensive literature analyzing these methods and comparing their performance (Daniel et al., 2003; Spors & Ahrens, 2008; Fazi, Nelson, & Potthast, 2009; Franck, Wang, & Fazi, 2017; Firtha, Fiala, Schultz, & Spors, 2018). In fact, they become equivalent in the limit of a continuum of loudspeakers, differing only when using a finite number (Fazi & Nelson, 2007b). Although they are amenable to mathematical analysis and have computationally efficient implementations, there is no natural psycho-acoustic justification for minimizing the ℓ^2-error for the coefficients of the expansion in spherical harmonics, the L^2-error for the pressure or velocity, nor the solution to an integral equation, to produce a large sweet spot as we have defined it. As a consequence, the artifacts introduced by these methods, due to approximation errors, may produce noticeable, and possibly avoidable, psycho-acoustic artifacts.

1.3. Psycho-acoustic approaches

An alternative to reproduce better the auditory scene is to explicitly account for psycho-acoustic principles (Zwicker & Fastl, 2007; Blauert, 1997) in the methods. The first steps in this direction were taken in (Johnston & Lam, 2000) by proposing a simple model that aims to preserve the spatial properties of the desired auditory scene. They emphasized that a perception based system should recreate the *interaural level difference* (ILD) and *interaural time difference* (ITD) of the potential listeners. This way, it would at least correctly recreate the horizontal spatial properties of the desired auditory scene.

A method to reproduce an active intensity field, itself a proxy for the spatial properties, that is largely uniform in space was then proposed in (Sena, Hacihabiboglu, & Cvetkovic, 2013). It is based on an optimization problem yielding audio signals where at most two loudspeakers are active simultaneously. However, it makes the restrictive assumption that the target sound wave is a plane wave, and that the loudspeakers emit plane waves. In (Ziemer & Bader, 2017) the *radiation method* and the *precedence fade* are proposed. The former is equivalent to applying a PMM over a selection of frequencies that are most relevant psycho-acoustically, whereas the latter is a method to overcome the localization problems associated to the *precedence effect* (Blauert, 1997). Finally, in (Lee, Nielsen, & Christensen, 2020) a PMM is extended to account for psycho-acoustic effects by considering the L^2-norm of the differences in pressure convolved in time by a suitable filter, i.e. it is based on the convex problem

$$\underset{g \in \mathbb{C}^{n_s}}{\text{minimize}} \quad \sum_{n=1}^{n_d} \sum_{m=1}^{n_t} |\varepsilon(\mathrm{x}_n, t_m; g)|^2 \quad , \tag{1.58}$$

where ρ is the convolution filter and ε is the convolved error between the synthesized and the target sound waves, i.e.,

$$\varepsilon(\mathrm{x}, t; g) = \left(\rho * \left[u_0 - \sum_{\ell=1}^{n_s} u_\ell(g_\ell) \right] \right)(\mathrm{x}, t). \tag{1.59}$$

Note that the structure of (1.58) generalizes the radiation method, since selecting specific frequency bands is equal to the convolution with a sum of weighted sinc functions.

1.4. Contributions

We believe that there is a gap between methods that aim to directly approximate a sound wave to reproduce a desired auditory scene, and methods that leverage psycho-acoustic models to reproduce the same auditory scene. In this work, we develop a method from first principles that incorporates monaural psycho-acoustic models to generate a sound wave that directly maximizes the sweet spot. This method is amenable to mathematical analysis, has an efficient computational implementation, and incorporates psycho-acoustic principles from the onset. Our numerical results over sinusoidal instances and van de Par's spectral psycho-acoustic model (van de Par, Kohlrausch, Heusdens, Jensen, & Jensen, 2005) show our method outperforms the most common state-of-the-art methods for sound field reconstruction in three senses: (i) we generate the largest monaural sweet spots in every instance evaluated, (ii) we implicitly perform intensity direction reconstruction (a proxy for binaural cues) with consistently better results than the other methods, (iii) differently from other methods, we guarantee no discomfort for every potential listener lying on the target region. Also, the latter benefits of our method are preserved in multiple zones (with a zone of silence) instances.

1.5. Structure of the thesis

This work is organized as follows. In Section 2.1 we introduce the main physical and psycho-acoustic models that we use. In Section 2.2 we formulate the variational problem of maximizing the sweet spot, proposing an accurate approximation, and analyzing its properties. In Section 2.3 we show this approximation can be recast as a Difference-of-Convex (DC) program, and we introduce the SWEET algorithm as an efficient method to solve it approximately. In Section 2.4 we show a concrete implementation of our method based on van de Par's spectral psycho-acoustic model (van de Par et al., 2005). Finally, in Section 2.5 we perform several numerical experiments analyzing its performance, comparing its results with WFS, NFC-HOA and PMM over sinusoidal instances, and showing some concrete applications in multiple frequency and multiple zones (with a zone of silence) instances.

2. MAXIMIZING THE PSYCHO-ACOUSTIC SWEET SPOT

2.1. Mathematical model

2.1.1. Acoustic framework

Consider an array of n_s speakers located at $x_1, \ldots, x_{n_s} \in \mathbb{R}^3$. When the medium is assumed homogeneous and isotropic, and each loudspeaker is modeled as an isotropic point source, the sound wave they generate is (Evans, 2010, Section 2.5.2)

$$u(t,x) = \sum_{k=1}^{n_s} \frac{c_k(t - c_s^{-1}\|x - x_k\|)}{4\pi\|x - x_k\|}$$

where c_s is the speed of sound in the medium, and $c_1, \ldots, c_{n_s}$ are the audio signals of every loudspeaker. In the frequency domain, this is represented as

$$\widehat{u}(f,x) = \sum_{k=1}^{n_s} \widehat{c}_k(f) \frac{e^{-2\pi i c_s^{-1} f\|x-x_k\|}}{4\pi\|x - x_k\|} \tag{2.1}$$

where $\widehat{c}_k$ is the Fourier transform of c_k in time

$$\widehat{c}_k(f) := \int c_k(t) e^{-2\pi i f t}\, dt.$$

To model the spatial radiation pattern of each loudspeaker, along with time-invariant effects such as reverb (Gauthier et al., 2005; Betlehem & Abhayapala, 2005), the representation (2.1) can be replaced by

$$\widehat{u}(f,x) = \sum_{k=1}^{n_s} \widehat{c}_k(f) G_k(f,x). \tag{2.2}$$

where G_k are the corresponding Green's functions. In addition to this array, consider a bounded domain $\Omega \subset \mathbb{R}^3$ containing no loudspeakers, i.e. $x_k \notin \overline{\Omega}$, allowing us to avoid the singularities in (2.1) at each x_k. On this domain, we could attempt to approximate *as best as possible* a sound wave u_0 with the array of loudspeakers.

If we had a continuum of isotropic point sources on $\partial\Omega$ then, under suitable conditions, the *simple source formulation* (Williams, 1999, Section 8.7) shows we can reproduce u_0 *exactly*. However, when only a finite number of physical loudspeakers are available, we must find $\widehat{c}_1, \ldots, \widehat{c}_{n_s}$ such that

$$\widehat{u}_0(f, x) \approx \sum_{k=1}^{n_s} \widehat{c}_k(f) G_k(f, x), \tag{2.3}$$

in an suitable sense, for $x \in \Omega$. In many cases $\widehat{u}_0$ is real-analytic on its second argument over Ω. As a consequence, when the speakers are isotropic point sources or G_k is real-analytic on its 2nd argument, the approximation cannot be exact on *any* open set unless u_0 was actually generated by the speaker array (Krantz & Parks, 2002, Corollary 1.2.5). This suggests (2.3) can hold only in average.

From now on we let W_S be the set of acoustic waves that can be generated by the array, represented in the frequency domain as in (2.2). We formalize this set in Section 2.2 and we first turn our attention to the psycho-acoustic criteria that determine a suitable sense to interpret (2.3).

2.1.2. Psycho-acoustic preliminaries

To interpret (2.3) adequately, we consider two basic aspects of the human auditory system: the hearing threshold and the damage/discomfort risk level threshold. The former allows us to determine when the differences between u_0 and the approximating wave are negligible, whereas the latter ensures we do not harm listeners.

2.1.2.1. The hearing threshold

An important psycho-acoustic problem is to determine when the difference between two audio signals $v_0 = v_0(t)$ and $v = v(t)$ is audible. A key concept to address it is the

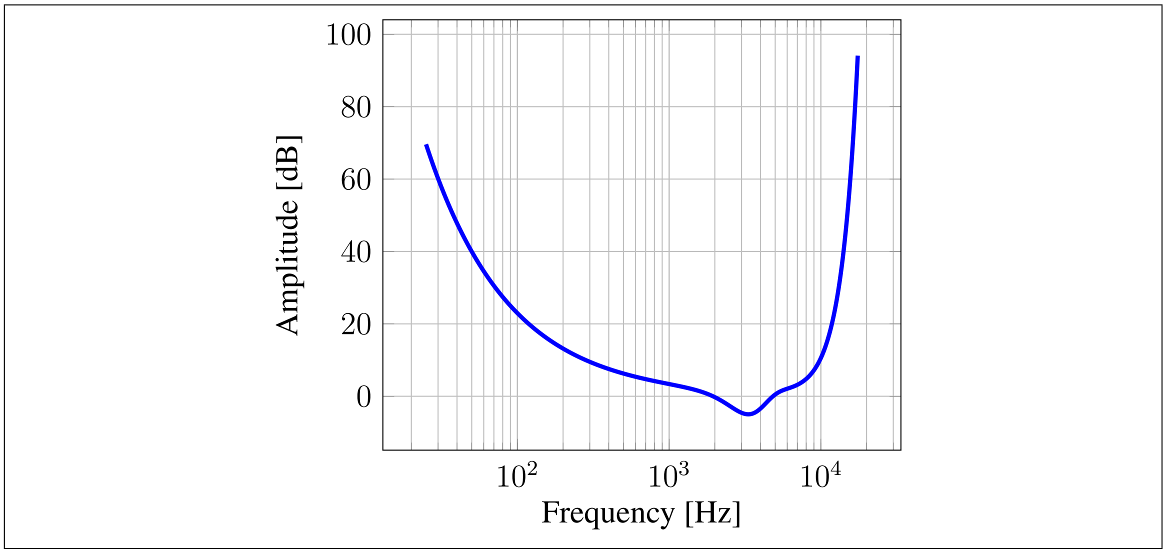


Figure 2.1. Threshold of audibility.

absolute threshold of hearing (Zwicker & Fastl, 2007, Section 2.1) (see Figure 2.1): when $v_0 \equiv 0$, a pure tone v is imperceptible if its intensity falls below it.

In complex audio signals other mechanisms come into play and the criteria for perception depend on the signal v_0 being approximated. It has been proposed that the human auditory system first computes an *internal representation* of the audio signal $v \mapsto \Phi(v)$ to then apply an *internal detector* $(\Phi(v), \Phi(v_0)) \mapsto D^\star(\Phi(v), \Phi(v_0))$. The difference is perceptible if this value exceeds a given threshold (Jepsen, Ewert, & Dau, 2008; Disch et al., 2018). These studies do not provide a tractable form for this representation nor for the internal detector. A simplification yielding a tractable model is given in (Plasberg & Kleijn, 2007). The model is simplified to a non-symmetric *distortion measure*

$$D(v, v_0) = \int_{\mathbb{R}} |L(v - v_0)(t)|^2 \, dt \tag{2.4}$$

where L is a transform modelling locally time-invariant filters that may depend on v_0. Another simplification in the literature is to consider a sum of convolved-weighted-squared errors (Taal, Hendriks, & Heusdens, 2012)

$$D(v, v_0) = \sum_k \int_{\mathbb{R}} |h_k * (v - v_0)(t) g_k(t)|^2 \, dt \tag{2.5}$$

where h_k and g_k represent a spectral and time weighting respectively. Together they model the difference over the k*-th auditory filter*. The filters may depend themselves on v_0. A further simplification introduced in (van de Par et al., 2005) consists in taking a constant g, i.e.,

$$D(v, v_0) = \sum_k \int_{\mathbb{R}} |\widehat{v}(f) - \widehat{v}_0(f)|^2 \rho_k(f) \, df. \tag{2.6}$$

This proposal works only with spectral information and thus it may not capture temporal *masking* effects accurately (Taal et al., 2012). The main reason to make these models dependent on v_0 is to account for the psycho-acoustic equivalence of v and v_0 when the approximation error is masked by v_0. This is a principle already used in audio coding (Painter & Spanias, 2000; Bosi & Goldberg, 2012).

2.1.2.2. The damage and discomfort risk threshold

Exposure to loud sound waves may be uncomfortable. Then, unrestricted spatial sound systems may reproduce undesirable sound scenes where some features prevail at the expense of the discomfort of some listeners. Empirical thresholds for loud discomforts levels for sinusoidal signals over a finite set of frequencies have been defined in the literature, e.g. in (Knobel & Sanchez, 2006; Sherlock & Formby, 2005). Naturally, these can be expressed as

$$\int_{\mathbb{R}} |\widehat{u}(f, x)|^2 \, \rho(f)^2 df, \tag{2.7}$$

where $\rho(f)$ is the multiplicative inverse of the threshold.

2.1.3. Psycho-acoustic framework

Although there is no definitive model for the hearing threshold, the literature supports the idea that the effects that must be taken into account depend on the sound wave u_0 itself. In this work we consider a general form for these models that includes some proposals in the literature. Inspired by (2.4), if u is a acoustic wave on Ω, a map of the form

$$Bu(x) = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K_B(t,t',x)(u-u_0)(t',x)\,dt' \right|^2 dt \tag{2.8}$$

where K_B is a suitable kernel, not necessarily time-invariant, quantifies the differences in perception between u and u_0 at a given x. A map of this form can account for time-variant effects, such as temporal masking, and also for time-invariant effects, such as spectral masking. Then, as the form in (2.4), K_B models a locally time-invariant filter. Also, it can be regarded as a factorization of a quadratic form that relates the difference of the signals at different times, i.e., formally we have,

$$\begin{aligned} Bu(x) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} K_B(t,t',x)(u-u_0)(t',x)\,dt' \right)^* \left(\int_{\mathbb{R}} K_B(t,t'',x)(u-u_0)(t'',x)\,dt'' \right) dt \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} (u-u_0)(t',x) \underbrace{\left(\int_{\mathbb{R}} K_B(t,t',x)^* K_B(t,t'',x)\,dt \right)}_{:=W(t',t'',x)} (u-u_0)(t'',x)\,dt'dt'' \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} (u-u_0)(t',x) W(t',t'',x)(u-u_0)(t'',x)\,dt'dt'', \end{aligned}$$

where W is the kernel of the quadratic form at each x.

By choosing suitable kernels we can represent the differences in perception over several auditory filters as a collection of functionals $B_1, \ldots, B_{n_b}$ of the form (2.8). Consequently, we define the *threshold map* as

$$Tu(x) := \Psi(B_1u(x), \ldots, B_{n_b}u(x)) \tag{2.9}$$

where $\Psi : \mathbb{R}_+^{n_b} \to \mathbb{R}$ is a continuous convex function that is non-decreasing on each one of its components. Without loss, we consider the difference between u and u_0 is *not audible* at x if $Tu(x) \leq 0$. Remark that by choosing a suitable function Ψ we may incorporate interactions between different auditory filters, e.g., when Ψ is a positive-definite quadratic function, representing the integration of the difference of perception between different frequency bands.

By choosing a suitable kernel K_B and integrating function Ψ, a functional of the form (2.9) reproduces the auditory metrics (2.4), (2.5) and (2.6). To reproduce (2.6) we choose $\sigma_k(t, x)$ as a function associated to the k-th auditory filter such that $\rho_k(f, x) = |\widehat{\sigma}_k(f, x)|^2$ and choose the kernel $K_{B_k}(t, t', x) = \sigma_k(t' - t, x)$. By Parseval's identity and the convolution theorem we have

$$\begin{aligned} B_k u(x) &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} \sigma_k(t' - t, x)(u - u_0)(t', x)\, dt' \right|^2 dt \\ &= \int_{\mathbb{R}} |\sigma_k * (u - u_0)(t, x)|^2 \, dt \\ &= \int_{\mathbb{R}} |(\widehat{u} - \widehat{u}_0)(f, x)|^2 \, |\widehat{\sigma}_k(f, x)|^2 \, df \\ &= \int_{\mathbb{R}} |(\widehat{u} - \widehat{u}_0)(f, x)|^2 \, \rho_k(f, x)\, df. \end{aligned}$$

By considering the integrating function $\Psi(z_1, \ldots, z_{n_b}) = \sum_{k=1}^{n_b} z_k$ we conclude that $Tu(x)$ has the same form as (2.6) at every $x \in \Omega$.

Similarly, to obtain (2.5) we choose the kernel $K_{B_k}(t, t', x) = h_k(t' - t, x) g_k(t, x)$. Then,

$$\begin{aligned} B_k u(x) &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} h_k(t' - t, x) g_k(t, x)(u - u_0)(t', x)\, dt' \right|^2 dt \\ &= \int_{\mathbb{R}} |h_k * (u - u_0)(t, x) g_k(t, x)|^2 \, dt. \end{aligned}$$

If we consider the same integrating function $\Psi(z_1, \ldots, z_{n_b}) = \sum_{k=1}^{n_b} z_k$ we conclude that $Tu(x)$ has the same form as (2.5) at every $x \in \Omega$.

Finally, to obtain (2.4) we first assume L can be represented by an integral kernel. We believe this is not a restrictive assumption in practical applications. Hence,

$$L(u - u_0)(t, x) = \int_{\mathbb{R}} K_B(t, t', x)(u - u_0)(t', x)\, dt'.$$

Note this is precisely the form

$$\int_{\mathbb{R}} |L(u - u_0)(t, x)|^2\, dt = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K_B(t, t', x)(u - u_0)(t', x)\, dt' \right|^2 dt.$$

With this, and considering $\Psi(z) = z$ for $n_b = 1$ we conclude that $Tu(x)$ has the same form as (2.4) at every $x \in \Omega$.

Therefore, given an approximating wave $u \in W_S$, we define its *sweet spot* as the set where u is psycho-acoustically equivalent to u_0, i.e.,

$$\mathcal{S}(u) = \{x \in \Omega : \; Tu(x) \leq 0\}. \tag{2.10}$$

Note the psycho-acoustic equivalence that defines the sweet spot is *monaural*. Although at each point the audio signal is in this sense equivalent to the original, this does not account *a priori* for *binaural* effects, e.g., whether the position of an audio source is perceived correctly.

Analogously, to model the discomfort level threshold we consider a collection of functionals $Q_1, \ldots, Q_{n_p}$ of the form (2.8) with $u_0 \equiv 0$. Note that those generalize (2.7) as they can account for time-variant effects. To enhance flexibility, we do not assume the same selection of auditory filters for the functionals B and Q, nor that $n_b = n_p$. Hence, we define the *discomfort map* P as in (2.9) for $u_0 \equiv 0$

$$Pu(x) := \Pi(Q_1 u(x), \ldots, Q_{n_p} u(x))$$

where Π is a function with the same properties as Ψ. Then,

$$\mathcal{P} := \{u : \ \forall x \in \Omega. \ Pu(x) \leq 0\} \tag{2.11}$$

is the collection of sound waves below the discomfort threshold at every x. The domain of T and P are sound waves, and thus are part of the *sound scene*. In contrast, their image are part of the *auditory scene*. Hence, T and P link the objective and subjective aspects of the problem.

Our goal is to find an acoustic wave $u \in W_S$ that maximizes the *weighted area* of the sweet spot $\mu(\mathcal{S}(u))$ while remaining comfortable, i.e., $u \in \mathcal{P}$. From now on, we assume u_0 is known and fixed. Particularly, all the parameters that we have introduced to define the threshold map (2.9) may depend on u_0.

2.2. Maximizing the sweet-spot

To formalize the problem of maximizing the sweet spot we make some critical assumptions. We consider the space

$$W := \left\{u : \ \sup_{x\in\Omega} \int_{\mathbb{R}} |u(t,x)|^2 \, dt < \infty \right\}$$

of sound waves that have finite energy at every $x \in \Omega$. Spaces of this form are called *mixed L^p-spaces* and were introduced in (Benedek & Panzone, 1961). The space W is complete under the norm

$$\|u\|_W := \sup_{x\in\Omega} \left(\int_{\mathbb{R}} |\, u(t,x)|^2 \, dt \right)^{1/2} .$$

An important feature of this norm is that the energy is preserved in time and frequency, i.e., $\|u\|_W = \|\widehat{u}\|_W$. From now on, we assume $u_0 \in W$. The following proposition summarizes the technical results that ensure that the methods we propose are well-posed. We defer its proof to Appendix A.

PROPOSITION 1. *Suppose that*

(i) The audio signals $\widehat{c}_k$ *in* (2.2) *are all bandlimited to an interval* I_c *and their* L^2*-norm is uniformly bounded.*

(ii) The functions G_k *in* (2.2) *are continuous and bounded on* $I_c \times \overline{\Omega}$.

(iii) For every $K \in \{K_{B_1}, \ldots, K_{B_{n_b}}\} \cup \{K_{Q_1}, \ldots, K_{Q_{n_p}}\}$ *there is a constant* C_K *such that*

$$\int_{\mathbb{R}} |K(t,t',x)|\, dt, \;\; \int_{\mathbb{R}} |K(t,t',x)|\, dt' \leq C_K$$

for a.e. $x \in \Omega$.

Then the following assertions are true.

(i) The map $T : W \to L^\infty(\Omega)$ *is continuous, and for almost every* $x \in \Omega$ *the map* $u \to Tu(x)$ *is convex.*

(ii) The set $\mathcal{S}(u)$ *is Borel measurable for any* $u \in W$.

(iii) The set W_S *is compact in* W.

(iv) The set $\mathcal{P}$ *is closed in* W.

We assume the hypotheses of the proposition hold throughout. This does not impose strong constraints on the threshold map (2.9). However, this implies the sound waves in W_S are continuous in space and time.

The weighted area of the sweet spot is measured with a Borel measure μ (Cohn, 2013, Section 1.2). The problem of maximizing the sweet spot becomes

$$(P_0) \quad \begin{cases} \underset{u \in W_S}{\text{maximize}} & \mu(\mathcal{S}(u)) \\ \text{subject to} & u \in \mathcal{P}. \end{cases}$$

In the above problem the feasible set is closed and bounded and, in fact, compact. To prove there exists a solution, we need to characterize the regularity of the objective function. However, this implies characterizing the behavior of the *set-valued function* $u \rightrightarrows \mathcal{S}(u)$.

This could be very difficult in practice. For this reason, we propose an approximation to (P_0) that can be analyzed with standard methods, and for which approximate solutions can be found efficiently.

2.2.1. The layer-cake representation

The *layer-cake representation* allows us to approximate the area of $\mathcal{S}(u)$ in terms of an integral over a function of u. Let φ be a bounded non-negative function of bounded variation such that $\varphi(t) = 0$ for $t < 0$ and $\|\varphi\|_{L^1} = 1$. Let φ_ε denote the function $\varphi_\varepsilon(t) = (1/\varepsilon)\varphi(t/\varepsilon)$ for $\varepsilon > 0$ and define

$$\Phi_\varepsilon(t) = \int_{-\infty}^{t} \varphi_\varepsilon(s)\, ds.$$

Suppose $v \in L^\infty(\Omega)$, $\alpha > 0$ and let $S_\alpha := \{x :\ v(x) > \alpha\}$. Since Ω is bounded, this implies $v \in L^1(\Omega)$. We claim the area $\mu(S_\alpha)$ can approximated by

$$A_\varepsilon^{(\alpha)}(v) = \int_\Omega \Phi_\varepsilon(v(x) - \alpha)\, d\mu(x).$$

PROPOSITION 2. *For every fixed $v \in L^\infty(\Omega)$ and $\alpha \in \mathbb{R}$ we have*

$$\lim_{\varepsilon \downarrow 0} A_\varepsilon^{(\alpha)}(v) = \mu(\{x \in \Omega :\ v(x) > \alpha\}).$$

PROOF OF PROPOSITION 2. Let $\{\varepsilon_n\}$ be monotone decreasing to zero and $V_{t,n} := \{x \in \Omega :\ v(x) \geq \alpha + \varepsilon_n t\}$. For every fixed $t \geq 0$ we have $V_{t,n} \subseteq V_{t,n+1}$. Define $V := \bigcup_{n>0} V_{t,n} = \{x \in \Omega :\ v(x) > \alpha\}$ and $h_n(t) = \varphi(t)\mu(V_{t,n})$. Note the latter are measurable as $t \mapsto \mu(V_{t,n})$ is monotone. Then $h_n(t) \uparrow \varphi(t)\mu(V)$ as $n \to \infty$ by continuity from below (Cohn, 2013, Proposition 1.2.5). By Fubini's theorem

$$\begin{aligned}
A^{(\alpha)}_{\varepsilon_n}(v) &= \int_\Omega \int_{-\infty}^{v(x)-\alpha} \varphi_{\varepsilon_n}(t)\, dt d\mu(x) \\
&= \int_{\mathbb{R}} \varphi_{\varepsilon_n}(t) \int_\Omega \chi_{\{v(x)-\alpha \geq t\}}(t,x)\, d\mu(x)\, dt \\
&= \int_{\mathbb{R}} \varphi_{\varepsilon_n}(t) \mu(\{x \in \Omega :\ v(x) \geq \alpha + t\})\, dt \\
&= \int_0^\infty \varphi(t)\mu(V_{t,n})\, dt \\
&\xrightarrow{n\to\infty} \mu(\{x \in \Omega :\ v(x) > \alpha\})
\end{aligned}$$

where we used the monotone convergence theorem (Cohn, 2013, Theorem 2.4.1). As $\{\varepsilon_n\}$ is arbitrary, the claim follows. $\square$

Therefore, writing $A_\varepsilon = A^{(0)}_\varepsilon$, we have for $u \in W$ and ε small that

$$\begin{aligned}
A_\varepsilon(Tu) &= \int_\Omega \Phi_\varepsilon(Tu(x))\, d\mu(x) \\
&\approx \mu(\{x \in \Omega :\ Tu(x) > 0\}) \\
&= \mu(\mathcal{S}(u)^c) \\
&= \mu(\Omega) - \mu(\mathcal{S}(u))
\end{aligned}$$

whence $\mu(\mathcal{S}(u)) \approx \mu(\Omega) - A_\varepsilon(Tu)$. This allows us to use directly an integral functional of a function of Tu thereby removing the need to use the set $\mathcal{S}(u)$ as an optimization variable.

2.2.2. The variational problem

We propose to solve the ε-approximated problem

$$(P_\varepsilon) \begin{cases} \underset{u \in W_S}{\text{minimize}} & A_\varepsilon(Tu) \\ \text{subject to} & u \in \mathcal{P}. \end{cases} \tag{2.12}$$

We can characterize the regularity of the objective function for this problem.

PROPOSITION 3. *The function $A_\varepsilon : L^\infty(\Omega) \to \mathbb{R}$ is continuous. Since W_S is compact, there exists at least one solution to* (2.12)*.*

PROOF OF PROPOSITION 3. Let $\delta > 0$ and $v_0, v \in L^\infty(\Omega)$ be such that $\|v - v_0\|_{L^\infty} < \delta/2$. Then $|v(x) - v_0(x)| < \delta/2$ for x on a set of full measure. Since $\varphi(t)$ is bounded for $-\|v_0\|_{L^\infty} - \delta/2 \leq t \leq \|v_0\|_{L^\infty} + \delta/2$ we have

$$|\Phi_\varepsilon(v(x)) - \Phi_\varepsilon(v_0(x))| \leq \int_{v_0(x)-\delta/2}^{v_0(x)+\delta/2} \varphi(t)\, dt \leq c_{v_0,\varphi}\delta$$

where $c_{v_0,\varphi} > 0$ depends only on φ and v_0. Thus, $|A_\varepsilon(v_2) - A_\varepsilon(v_1)| \leq c_{v_0,\varphi}\mu(\Omega)\delta$ whence A_ε is continuous. The existence of solutions follows from the compactness of $W_S \cap \mathcal{P}$.

□

Unfortunately, we cannot assert that the solution to (2.12) is unique and, in fact, several solutions may exist as two distinct sound waves may be the best psycho-acoustic approximation to u_0 on Ω. Consider the case $u_0 \equiv 0$: any sound wave $u \in W_S$ of sufficiently small magnitude falls below the pain and hearing thresholds, and is thus optimal for (2.12). In addition, although the feasible set is convex, the objective function is not. Therefore, in principle, there may not be efficient algorithms to solve (2.12), and several local minima may exist.

2.3. DC Formulation

To introduce a suitable algorithm to solve (2.12) we first rewrite it as

$$(P'_\varepsilon) \quad \begin{cases} \underset{u \in W_S, v \in L^\infty(\Omega)}{\text{minimize}} & A_\varepsilon(v) \\ \text{subject to} & Tu \leq v, \quad u \in \mathcal{P}. \end{cases} \tag{2.13}$$

We interpret the auxiliary variable v as an overestimate of the threshold map over Ω. The proof of the following proposition shows that for all practical purposes we can assume $Tu = v$.

PROPOSITION 4. *The following assertions are true.*

(i) The set $\{(u, v) : Tu \leq v\}$ *is closed and convex.*

(ii) If $u^\star$ *is an optimal solution to* (P_ε) *then* $(u^\star, Tu^\star)$ *is an optimal solution to* (P'_ε)*. In particular,* (P'_ε) *has a solution.*

(iii) If $(Tu^\star, v^\star)$ *is an optimal solution to* (P'_ε) *then* $(u^\star, Tu^\star)$ *is also an optimal solution, and* $u^\star$ *is an optimal solution to* (P'_ε)*.*

(iv) The problems (P_ε) *and* (P'_ε) *are equivalent.*

PROOF OF PROPOSITION 4. We omit details for brevity. (i) Convexity follows from (i) in Proposition 1. Similarly, the set is closed by the continuity of T. (ii)-(iv) Let u be an optimal solution to (P_ε). By choosing $v = Tu$ it is clear the optimal solution to (P'_ε) is less or equal than that of (P_ε). Let (u', v') be an optimal solution to (P'_ε). We claim that we can choose v' so that $v' = Tu'$. First, remark that $v_1 \leq v_2$ implies that $A_\varepsilon(v_1) \leq Ae(v_2)$. Therefore, we can define $v'' \in L^\infty(\Omega)$ as $v''(x) = \min\{v'(x), Tu'(x)\} = Tu'(x)$ whence $A_\varepsilon(v'') \leq A_\varepsilon(v')$. Since (u', v') is optimal, $A_\varepsilon(v'') = A_\varepsilon(v')$ and, without loss, we can assume $v' = Tu'$. Consequently, the optimal value of (P'_ε) is greater or equal to that of (P_ε). Hence the problems are equivalent and, by Proposition 3, they both have at least one solution. $\square$

From now on, we denote both (2.12) and (2.13) as (P_ε) and we omit the subscript ε when possible. Note that in (2.13) the objective is the *difference of convex functions*. Since φ is of bounded variation we can consider its Jordan decomposition (Royden & Fitzpatrick, 2010, Chapter 6, Jordan's Theorem) $\varphi = \varphi_+ - \varphi_-$ where $\varphi_+, \varphi_- : \mathbb{R} \to \mathbb{R}$ are non-decreasing functions. Define

$$\Phi_+(t) = \int_{-\infty}^{t} \varphi_+(s)ds, \quad \Phi_-(t) = \int_{-\infty}^{t} \varphi_-(s)ds$$

By construction, $\Phi = \Phi_+ - \Phi_-$. Hence, we can decompose A as $A = A_+ - A_-$ where

$$A_+(v) := \int_\Omega \Phi_+(v(x))\, d\mu(x), \quad A_-(v) := \int_\Omega \Phi_-(v(x))\, d\mu(x).$$

PROPOSITION 5. *The functions $A_+, A_- : L^\infty(\Omega) \to \mathbb{R}$ are convex and continuous.*

PROOF OF PROPOSITION 5. By construction, both Φ_+ and Φ_- have a derivative almost everywhere that is non-decreasing, hence monotone (Bauschke & Combettes, 2011, Proposition 17.10). Therefore Φ_+ and Φ_- are convex. With this, and the linearity and monotonicity of the integral, A_+ and A_- are convex. Moreover, they are continuous as the proof of Proposition 3 holds *mutatis mutandis*. $\square$

Hence, the formulation (2.13) is a Difference-of-Convex (DC) program (Tao & An, 1997; Horst & Thoai, 1999). For this type of problems, there are efficient algorithms that attempt to find a solution.

2.3.1. SWEET algorithm

The Convex-Concave Procedure (CCCP) (Lipp & Boyd, 2016) is an efficient method, which can be thought as a primal version of the DCA algorithm (Tao & An, 1997), to find a solution to (2.13). Although it can be shown that if it converges, then its limit is a stationary point (Tao & An, 1997, Theorem 3), our results in Section 2.5 suggest that in practice we are able to find local minima for (2.13). The CCCP is an iterative method

Algorithm 1: SWEET

input: A decreasing sequence $\{\varepsilon_k\}$ of positive numbers with $\varepsilon_0 \gg 1$, $N_\varepsilon, N_u \in \mathbb{N}$ and $v_0 \in L^\infty(\Omega)$
set : $N = N_\varepsilon N_u$
for $i = 0, \ldots, N_\varepsilon - 1$ **do**
  for $j = 0, \ldots, N_u - 1$ **do**
    $k = iN_u + j$
    (u_{k+1}, v_{k+1}) solution to (P_{ε_i, v_k})
  end
end
return u_N

that uses an affine majorant for the concave part, e.g., using subgradients, to then majorize the objective function in (2.13) with a convex function. The resulting convex problem can then be solved efficiently.

Since A_- is continuous and convex, it has a subdifferential $\partial A_-(v_0) \subset L^\infty(\Omega)'$ at every $v_0 \in L^\infty(\Omega)$ (Barbu & Precupanu, 2012, Proposition 2.36). If $g \in \partial A_-(v_0)$ then

$$A_-(v) \geq A_-(v_0) + g(v - v_0)$$

for any $v \in L^\infty(\Omega)$. The functional g is called a subgradient. Therefore, we can use the convex majorizer

$$A(v) = A_+(v) + A_-(v) \leq A_+(v) - A_-(v_0) - g(v - v_0).$$

Although it may be difficult to characterize the subdifferential of a convex function on a Banach space, in our case we can always find at least a subgradient at any v_0.

PROPOSITION 6. *Let* $v_0 \in L^\infty(\Omega)$. *Then*

$$g_{v_0}(v) = \int_\Omega \varphi_-(v_0(x))v(x)\, d\mu(x) \tag{2.14}$$

is a subgradient for A_- *at* v_0.

PROOF OF PROPOSITION 6. Let $v \in L^\infty(\Omega)$, $t, t_0 \in \mathbb{R}$. By the monotonicity of φ_- we have

$$\begin{aligned}\Phi_-(t_0) + \varphi_-(t_0)(t - t_0) &= \int_{-\infty}^{t_0} \varphi_-(s)ds + \int_{t_0}^{t} \varphi_-(t_0)ds \\ &\leq \int_{-\infty}^{t} \varphi_-(s)ds \\ &= \Phi_-(t).\end{aligned}$$

Since $t \in \mathbb{R}$ is arbitrary, $\varphi_-(t_0)$ is a subgradient of Φ_- at t_0. Moreover, since $t_0 \in \mathbb{R}$ is arbitrary, we have that

$$(\Phi_- \circ v_0) + (\varphi_- \circ v) \cdot (v - v_0) \leq (\Phi_- \circ v).$$

Whence, by the monotonicity of the integral, integrating over Ω yields

$$A_-(v) \geq A_-(v_0) + \int_\Omega \varphi_-(v_0(x))(v(x) - v_0(x))\, d\mu(x).$$

Since $v \in L^\infty(\Omega)$ is arbitrary, g_{v_0} is a subgradient of A_- at v_0. $\square$

The CCCP solves at each iteration the convex problem

$$(P_{\varepsilon,v_0}) \quad \begin{cases} \underset{u \in W_S, v \in L^\infty(\Omega)}{\text{minimize}} & A_+(v) - A_-(v_0) - g_{v_0}(v - v_0) \\ \text{subject to} & Tu \leq v, \quad u \in \mathcal{P}. \end{cases}$$

PROPOSITION 7. *There exists at least one solution to* (2.3.1).

PROOF OF PROPOSITION 7. We first construct a candidate for an unconstrained minimizer of the objective. Let $\tilde{v}_0 : \Omega \to \mathbb{R}$ be any representative of the equivalence class v_0 and, for $x \in \Omega$, define $f_x(t) = \Phi_+(t) - \varphi_-(\tilde{v}_0(x))t$. Then define the multivalued map $F : \Omega \rightrightarrows \mathbb{R}$ as

$$F(x) := \arg\min\{f_x(t) :\ t \in \mathbb{R}\}.$$

Note that f_x is convex; similarly to the proof of Proposition 6, we have that $g_x(t) := \varphi_+(t) - \varphi_-(\tilde{v}_0(x)) \in \partial f_x(t)$. Let $\Omega_b = \{x \in \Omega : g_x(-\|v_0\|_{L^\infty(\Omega)}) < 0\}$. Since $(\varphi_+ - \varphi_-)_{|[0,\infty)} \geq 0$ and φ_+ is non decreasing, for $t \geq \|v_0\|_{L^\infty(\Omega)}$ and x a.e. in Ω we have that $g_x(t) \geq 0$. Similarly, since $(\varphi_+ - \varphi_-)_{|(-\infty,0]} \equiv 0$ and φ_+ is non decreasing, for $t \leq -\|v_0\|_{L^\infty(\Omega)}$ and x a.e. in Ω we have that $g_x(t) \leq 0$. Thus, considering that f_x is continuous and convex, F takes non-empty closed convex values and

(i) $F(x) \subseteq [-\|v_0\|_{L^\infty(\Omega)}, \|v_0\|_{L^\infty(\Omega)}]$ for x a.e. in Ω_b,

(ii) $-\|v_0\|_{L^\infty(\Omega)} \in F(x)$ for $x \in \Omega_b^c$.

Therefore, F admits a measurable selection $\tilde{v}^{\star\star}$ (Aubin & Frankowska, 1990, Theorem 8.2.2 and Theorem 8.1.13). Because of (i), $|\tilde{v}^{\star\star}(x)| \leq |\tilde{v}_0(x)|$ for $x \in \Omega_b$. Let $\tilde{v}^\star$ be such that $\tilde{v}^\star_{|\Omega_b} = \tilde{v}^{\star\star}_{|\Omega_b}$ and $\tilde{v}^\star_{|\Omega_b^c} = -\|v_0\|_{L^\infty(\Omega)}$. Because of (ii), and considering that Ω_b is a measurable set, $\tilde{v}^\star$ is still a measurable selection for F. If we let $v^\star$ denote its equivalence class, we deduce that $\|v^\star\|_{L^\infty} \leq \|v_0\|_{L^\infty}$ whence $v^\star \in L^\infty(\Omega)$. By construction, for every $v \in L^\infty(\Omega)$,

$$\Phi_+(v^\star(x)) - \varphi_-(\tilde{v}_0(x))v^\star(x) \leq \Phi_+(v(x)) - \varphi_-(\tilde{v}_0(x))v(x)$$

for x a.e. in Ω. Therefore, by the monotonicity of the integral,

$$A_+(v) - g_{v_0}(v) \geq A_+(v^\star) - g_{v_0}(v^\star).$$

Thus, $v^\star$ is an unconstrained minimizer of the objective. However, it might not exist $u \in W_S$ such that $(u, v^\star)$ is feasible. But we can define a a minimizing sequence $\{(u_k, v_k)\}$, and then let $w_k = \max\{v^\star, Tu_k\}$ and $\Omega_k := \{x \in \Omega : Tu_k(x) = w_k(x)\}$. Since v_k is feasible, $v^\star(x) \leq Tu_k(x) \leq v_k(x)$ a.e over Ω_k. Moreover, since f_x is convex, it is non-decreasing over the set $[\tilde{v}^\star(x), \infty)$. Therefore we have

$$\begin{aligned}\Phi_+(v_k(x)) - \varphi_-(v_0(x))v_k(x) &\geq \Phi_+(Tu_k(x)) - \varphi_-(v_0(x))Tu_k(x)\\ &= \Phi_+(w_k(x)) - \varphi_-(v_0(x))w_k(x).\end{aligned}$$

for x a.e. in Ω_k. Analogoloustly, since $\Omega_k^c \subseteq \{x \in \Omega : w_k = v^\star\}$,

$$\Phi_+(v_k(x)) - \varphi_-(v_0(x))v_k(x) = \Phi_+(w_k(x)) - \varphi_-(v_0(x))w_k(x).$$

for x a.e. in Ω_k^c. Then, by the monotonicity of the integral,

$$A_+(v_k) - g_{v_0}(v_k) \geq A_+(w_k) - g_{v_0}(w_k).$$

Since W_S is compact, there exist a convergent subsequence $\{u_{k(\ell)}\} \subset \{u_k\}$ such that $\lim_{\ell\to\infty} u_{k(l)} =: u_\infty \in W_S$. Since the sequence is minimizing

$$p^\star = \lim \inf_{k\to\infty} (A_+(v_k) - g_{v_0}(v_k)) \geq \lim \inf_{k\to\infty} (A_+(w_k) - g_{v_0}(w_k)) \geq p^\star$$

where $p^\star \in \mathbb{R}$ is the optimal value. Hence, $\{(u_k, w_k)\}$ is also minimizing. Since T is continuous, we have $\lim_{\ell\to\infty} w_{k(\ell)} = \max\{v^\star, Tu_\infty\}$. Hence, $(u_\infty, \max\{v^\star, Tu_\infty\})$ is a minimizer. □

By solving a sequence of problems of the form $(P_{\varepsilon,v_{k+1}})$, where (u_{k+1}, v_{k+1}) is an optimal solution to (P_{ε,v_k}), we can attempt to find a solution to (P_ε).

Assuming the CCCP converges to a local minimizer to (P_ε), we can then solve a sequence of problems of the form (P_{ε_k}) for a decreasing sequence $\{\varepsilon_k\}$ to approximate a solution to (P_0). In this case, we initialize the CCCP to solve $(P_{\varepsilon_{k+1}})$ with the solution found for (P_{ε_k}). We call this the SWEET algorithm and is shown in Algorithm 1.

Finally, remark we could apply the decomposition $A \circ T = A_+ \circ T - A_- \circ T$ in (2.12). Although the term $A_- \circ T$ is convex when $\phi_+ \geq 0$, majorizing $-A_- \circ T$ would be more involved than the approach we have taken here.

2.3.2. SWEET-ReLU algorithm

When φ is a step function, the function Φ is the difference of two *Rectified Linear Units* (ReLUs). The resulting instance of Algorithm 1 is simple and interpretable. Let $\varepsilon > 0$

and $\varphi = \varepsilon^{-1}\chi_{[0,\varepsilon]}$. Choosing $\varphi_+ = \varepsilon^{-1}\chi_{[0,\infty)}$ and $\varphi_- = \varepsilon^{-1}\chi_{[\varepsilon,\infty)}$, the decomposition $\Phi = \Phi_+ - \Phi_-$ becomes

$$\Phi(x) = \frac{1}{\varepsilon}(x_+ - (x-\varepsilon)_+)$$

whence Φ_+ and Φ_- are ReLUs. Moreover, the subgradient (2.14) becomes

$$g_{v_0}(v) = \frac{1}{\varepsilon}\int_{\{x:\, v_0(x)>\varepsilon\}} v(x)\, d\mu(x).$$

Let $\Omega_{\varepsilon,v_0} := \{x :\ v_0(x) \leq \varepsilon\}$. Since $A_-(v_0)$ and $g_{v_0}(v_0)$ in (P_{ε,v_0}) the terms are constant, it suffices to compute

$$\begin{aligned} A_+(v) - g_{v_0}(v) &= \frac{1}{\varepsilon}\int_\Omega v(x)_+\, d\mu(x) - \frac{1}{\varepsilon}\int_{\Omega^c_{\varepsilon,v_0}} v(x)\, d\mu(x) \\ &= \frac{1}{\varepsilon}\int_{\Omega_{\varepsilon,v_0}} v(x)_+\, d\mu(x) + \frac{1}{\varepsilon}\int_{\Omega^c_{\varepsilon,v_0}} (-v(x))_+\, d\mu(x) \end{aligned}$$

where we used the fact that $t_+ - t = (-t)_+$. The second term is non-negative, and it is positive only when v takes negative values. The restriction $Tu \leq v$ in (P_{ε,v_0}) allows us to choose v arbitrarily large over $\Omega^c_{\varepsilon,v_0}$, decreasing the objective value, and allowing us to neglect the second integral. Therefore, only the first term contributes to the objective in (P_{ε,v_0}). Hence, for this choice of φ, φ_+ and φ_- we obtain

$$(P_{\varepsilon,v_0}) \begin{cases} \underset{u\in W_S, v\in L^\infty(\Omega)}{\text{minimize}} & \displaystyle\int_{\Omega_{\varepsilon,v_0}} v(x)_+\, d\mu(x) \\ \text{subject to} & u \in \mathcal{P} \\ & Tu \leq v,\ 0 \leq v|_{\Omega^c_{\varepsilon,v_0}}. \end{cases} \tag{2.15}$$

Because of the monotonicity of the positive-part function we can eliminate the auxiliary variable v to obtain the problem

$$(P^{ReLU}_{\Omega_{\varepsilon,v_0}}) \begin{cases} \underset{u\in W_S}{\text{minimize}} & \displaystyle\int_{\Omega_{\varepsilon,v_0}} (Tu(x))_+\, d\mu(x) \\ \text{subject to} & u \in \mathcal{P}. \end{cases} \tag{2.16}$$

Note it depends on v_0 only through the set Ω_{ε,v_0}. With this in mind, notice that at each iteration of Algorithm 1 we need an optimal solution (u_{k+1}, v_{k+1}) to (P_{ε,v_k}). However, solving (2.16) only yields an optimal solution u_{k+1}. Fortunately, from a given solution u_{k+1} to (2.16) we can choose v_{k+1} such that (u_{k+1}, v_{k+1}) is an optimal solution to (2.15) as follows: let $v_{k+1}|_{\Omega_{\varepsilon,v_k}} = Tu_{k+1}|_{\Omega_{\varepsilon,v_k}}$ and $v_{k+1}|_{\Omega^c_{\varepsilon,v_k}} = \max\{\varepsilon, Tu_{k+1}|_{\Omega^c_{\varepsilon,v_k}}\}$. Using this choice, note that

$$\begin{aligned}\Omega_{\varepsilon,v_{k+1}} &= \{x : v_{k+1}(x) \leq \varepsilon\}\\ &= \{x \in \Omega_{\varepsilon,v_k} : Tu_{k+1}(x) \leq \varepsilon\}\\ &= \Omega_{\varepsilon,v_k} \cap \{x \in \Omega : Tu_{k+1}(x) \leq \varepsilon\}.\end{aligned}$$

We call this simplification the SWEET-ReLU algorithm. It is shown in Algorithm 2. Due to compactness, the iterates $\{u_k\}$ have at least one accumulation point, which must be a stationary point for (2.12) (Tao & An, 1997, Theorem 3). SWEET-ReLU can be interpreted as a *greedy algorithm* that improves at each step the approximation over the set Ω_k while neglecting the approximation outside Ω_k. Intuitively, a point in Ω is neglected by the algorithm as soon as it determines that it cannot belong to the sweet spot. Furthermore, the sequence of sets generated by the algorithm are precisely an approximation for the sweet spot as, in fact, $\mathcal{S}(u_N) \approx \Omega_N$. Additionally, initializing the algorithm with ε_0 sufficiently large we have $\Omega_1 = \Omega$, making the choice of u_0 irrelevant [here u_0 stands for the first element of the sequence of solutions, not for the desired sound wave]. Finally, the choice of $\{\varepsilon_i\}$ can be adaptive. For instance, ε_i can be selected as the p-th percentile of Tu_{iN_u-1}.

2.4. Implementation

We provide an implementation of SWEET-ReLU for approximating a sound wave generated by a (pseudo) sinusoidal isotropic point source emitting at frequencies $f_1^\star, \ldots, f_{n_f}^\star$.

Algorithm 2: SWEET-ReLU

input: A decreasing sequence $\{\varepsilon_i\}$ of positive numbers with $\varepsilon_0 \gg 1$, $N_\varepsilon, N_u \in \mathbb{N}$, $u_0 \in W$ and $\Omega_0 = \Omega$

set : $N = N_\varepsilon N_u$

for $i = 0, \ldots, N_\varepsilon - 1$ **do**

  for $j = 0, \ldots, N_u - 1$ **do**

    $k = iN_u + j$

    $\Omega_{k+1} = \Omega_k \cap \{x \in \Omega : Tu_k(x) \leq \varepsilon_i\}$

    u_{k+1} solution to $(P^{ReLU}_{\Omega_{k+1}})$

  end

end

return u_N

The loudspeakers are modeled as equivalent (pseudo) sinusoidal point sources, i.e. we use

$$\widehat{c}_k(f) = \sum_{\ell=1}^{n_f} a_{k,\ell}\, e^{-(f-f_\ell^\star)^2/2\sigma^2}$$

in (2.1) for coefficients $a_{k,\ell} \in \mathbb{C}$ and a fixed spectral localization parameter $\sigma \ll 1$. Since the signals are almost stationary, temporal masking is almost non-existent. This allows us to define the threshold map T using van de Par's spectral psycho-acoustic model (van de Par et al., 2005). In this case, the filters in (2.8) are time-invariant. Thus, for van de Par's model we have

$$B_j u(x) = \int |(\widehat{u} - \widehat{u}_0)(f,x)|^2 \rho_{B_j}(f,x)\, df$$

for

$$\rho_{B_j}(f,x) = \frac{w_{B_j}(f)}{C_A + \int |\widehat{u}_0(f,x)|^2 w_{B_j}(f) df}.$$

The constant $C_A > 0$ limits the perception of very weak signals in silence. The weight w_{B_j} is defined as $w_{B_j} := |\eta\gamma_j|^2$ where

$$\log_{10} \eta(f) = C_{\eta,0} - C_{\eta,1} f^{-0.8} - C_{\eta,2}(f - 3.3 \times 10^3)^2 + C_{\eta,3} f^4$$

with $C_{\eta,0} = 4.69$, $C_{\eta,1} = 18.2 \times 10^{1.4}$, $C_{\eta,2} = 32.5 \times 10^{-7}$ and $C_{\eta,3} = 5 \times 10^{-16}$ models the outer and middle ear as proposed by Terhardt (Terhardt, 1979), and

$$\gamma_j(f) = \left(1 + \left(\frac{945\pi(f - f_j)}{48\mathrm{ERB}(f_j)}\right)^2\right)^{-2}$$

models the filtering property of the basilar membrane in the inner ear at the center frequency f_j, where the Equivalent Rectangular Bandwidth (ERB) of the auditory filter centered at f_j is $\mathrm{ERB}(f_j) = 24.7(1 + 4.37 \times 10^{-3} f_j)^{-1}$ as suggested by Glasberg and Moore (Glasberg & Moore, 1990). The center frequencies f_j are uniformly spaced on the ERB-rate scale $\mathrm{ERBS}(f) = 21.4 \log(1 + 4.37 \times 10^{-3} f)$. For n_b center frequencies f_j we obtain n_b maps B_j that are combined with the integrating function $\Psi(b_1, \ldots, b_{n_b}) = -1 + C_\Psi b_1 + \ldots + C_\Psi b_{n_b}$ for a suitable constant $C_\Psi > 0$. The threshold map becomes

$$\begin{aligned} Tu(x) &= -1 + C_\Psi \sum_{j=1}^{n_b} \frac{\int |\widehat{u}(f,x) - \widehat{u}_0(f,x)|^2 w_{B_j}(f) df}{C_A + \int |\widehat{u}_0(f,x)|^2 w_{B_j}(f) df} \\ &\approx -1 + C'_\Psi \sum_{\ell=1}^{n_f} \sum_{j=1}^{n_b} \frac{|\widehat{u}(f_\ell^\star, x) - \widehat{u}_0(f_\ell^\star, x)|^2 w_{B_j}(f_\ell^\star)}{C_A + w_{B_j}(f_\ell^\star) |\widehat{u}_0(f_\ell^\star, x)|^2} \end{aligned}$$

where $C'_\Psi = 2^{1/4} \pi^{1/2} \sigma C_\Psi$ and we used the approximation for (pseudo) sinusoidal signals

$$\int \varphi(f) |\widehat{u}_0(f,x)|^2 \, df \approx 2^{1/4} \pi^{1/2} \sigma \sum_{\ell=1}^{n_f} \varphi(f_\ell^\star) |\widehat{u}_0(f_\ell^\star, x)|^2$$

when $\sigma \ll 1$. The constants C'_Ψ and C_A are defined as suggested in (van de Par et al., 2005). This considers the absolute threshold of hearing and the just-noticeable difference in level for sinusoidal signals, which gives, $C'_\Psi \approx 1.555$ and $C_A \approx 4.481$ when considering $n_b = 100$ as the number of center frequencies, and $f_1 = 20$, $f_{n_b} = 10^3$ as the first and last center frequency.

To model the pain threshold we consider the experiments in (Knobel & Sanchez, 2006) about the discomfort caused by sinusoidal signals. We interpolated the data in this study using cubic splines with natural boundary (Quarteroni, Sacco, & Saleri, 2010, Section 8.6)

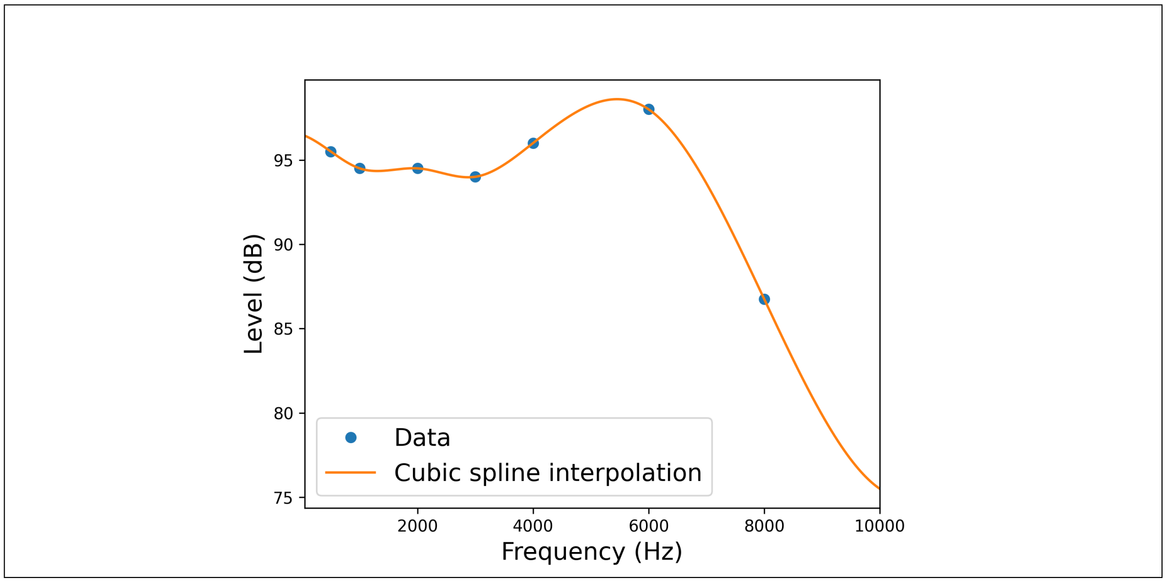


Figure 2.2. Interpolation of the loud discomfort levels for sinusoidal signals given in (Knobel & Sanchez, 2006) by cubic splines.

to obtain a function η_P, as shown in Fig. 2.2. For the auditory filter associated to the j-th frequency we define

$$\rho_{Q_j}(f,x) = |w_{Q_j}(f)/\eta_P(f)|^2$$

To our knowledge, there is no standard reference for the spectral integration that determines the levels of discomfort or pain. For simplicity, we consider, as in the van de Par model, a summing integrating function, but now with the center frequencies of the discomfort auditory filters equal to the sound frequencies $f_1^\star, \dots, f_{n_b}^\star$. Then, $\Pi(q_1, \dots, q_{n_f}) = -1 + C_\Pi q_1 + \dots + C_\Pi q_{n_f}$. This is actually a conservative choice of Π as this controls the sum of the contributions of every frequency, instead of each one separately. Consequently, we obtain

$$\begin{aligned} Pu(x) &= -1 + C_\Pi \sum_{\ell=1}^{n_f} \int |\widehat{u}(f,x)|^2 \rho_{Q_\ell}(f,x) df \\ &\approx -1 + C'_\Pi \sum_{\ell=1}^{n_f} |\widehat{u}(f_\ell^\star,x)|^2 \rho_{Q_\ell}(f_\ell^\star,x) \end{aligned}$$

where the same approximation holds by the same arguments as before. Naturally, $C'_{\Pi} = 1$.

To solve $(P^{ReLU}_{\Omega_\varepsilon,v_0})$ we discretize the integrals over Ω. The following proposition ensures that this approximation to the integral converges to the desired one under mild assumptions. We defer its proof to Appendix B.

PROPOSITION 8. *Suppose that that the statements (i) and (ii) in Proposition 1 hold, and that every $K \in \{K_{B_1}, \dots, K_{B_{n_b}}\}$ satisfies*

$$f_K(t,t') := \sup_{z\in\Omega} |K(t,t',z)| \in L^2(\mathbb{R}^2).$$

Then, for $u \in W_S$, $Tu \in C(\Omega)$. Furthermore, if Ω is compact, Tu is uniformly continuous over Ω.

Specifically, we discretize Ω using n_d disjoint squares or cubes of side $(|\Omega|/n_d)^{1/d}$ for $d \in \{2,3\}$. To avoid spatial aliasing, we need at least 2 points per spatial wavelength $\lambda_f = c_s/f$ for each frequency f of the source. This implies $(|\Omega|/n_d)^{1/d} < \lambda_f/2$ whence $n_d > (2/\lambda)^d|\Omega|$. To ensure the method performs well, we typically consider a denser discretization with at least 5 points per spatial wavelength.

2.5. Experiments

We perform two types of numerical experiments. First, we compare the performance of our method with the state-of-the-art methods WFS, NFC-HOA and L^2-PMM in terms of the size of the sweet spot they produce. Second, we explore other applications of our method related to sound field reconstruction. The setup for the numerical experiments consists of an equispaced arrangement of 20 loudspeakers lying on a circle of radius 2.5 m and at $\pi/4 \approx 0.785$ m from each other. The region of interest Ω is a concentric circle of radius 2.4975 m (Fig. 2.4). The speed of sound is $c_s = 343$ m/s.

The SWEET-ReLU algorithm and the L^2-PMM method were implemented in Python 3.8 using the CVXPY package (Diamond & Boyd, 2016; Agrawal, Verschueren, Diamond, & Boyd, 2018) and MOSEK (ApS, 2019). The simulations of 2.5D NFC-HOA and 2.5D WFS were done with the SFS Toolbox (Wierstorf & Spors, 2012). To compare the results of these methods, we compare the size of the sweet spot as a fraction of the area $|\Omega|$ of Ω. To compare the values of the threshold map Tu for u we use $\log(1+Tu)$. Hence, the sweet spot is the region where $\log(1+Tu) \leq 0$. Finally, we compare the Intensity Direction Error (IDE), defined as

$$\text{IDE}u(x) = \frac{1}{\pi} \arccos\left(\frac{\vec{I}u(x)}{|\vec{I}u(x)|} \cdot \frac{\vec{I}u_0(x)}{|\vec{I}u_0(x)|} \right),$$

where $\vec{I}$ is the time averaged acoustic intensity. For sinusoidal signals of frequency $f^\star$ it is given by (Williams, 1999, Section 2.3)

$$\vec{I}u(x) = \frac{1}{2}\text{Re}(u(f^\star, x)\vec{v}(f^\star, x)^*)$$

where $\vec{v}$ is the velocity vector field of u.

2.5.1. Comparison with state-of-the-art methods

To compare our method with state-of-the-art methods, we perform two types of numerical experiments. The first type consists of a sequence of instances where the source moves progressively away from the center of the loudspeaker array, starting at 0 m and ending at 15 m. Following the model in Section 2.4, the source is isotropic, and (pseudo) sinusoidal with $f_1^\star = 343$ Hz. Hence, its wavelength is 1 m. When the source is inside Ω, its intensity selected so that the wave has an amplitude of 60 dB at 1 m of the source. When the source is outside Ω we adjust the intensity so that the amplitude at the point where the segment joining the center of the arrangement and the source intersects the arrangement is 60 dB. This mitigates the effect of attenuation as the source moves away from the arrangement. A uniform discretization of 901 points was used for Ω at a distance of at most 0.145 m, achieving more than 6 points per wavelength.

The second type considers the same source at a distance of 5 m from the center of the array emitting a (pseudo) sinusouidal wave at different frequencies ranging from 50 Hz to 2000 Hz. To mitigate the issues due to non-convexity, we initialize SWEET-ReLU with the optimal solution obtained for the previous frequency value. A uniform discretization of 20848 points was used for Ω at a distance of at most 0.03 m, achieving more than 5 points per wavelength in the worst case. For both types of experiments we have chosen ε_i adaptively with percentile $p = 90$. The results are shown in Fig. 2.3. We see our method generates a larger sweet spot than that generated by every other method over the entire range of source locations and frequencies (Fig. 2.3a and Fig. 2.3b). When the source is at 2.5 m, lying over the arrangement, the sweet spot equals Ω, as expected (Fig. 2.3a). Furthermore, our method successfully attains the lowest average threshold value in most of the instances. Although the performance degrades at very low frequencies compared to other methods, it remains below the audible threshold (Fig. 2.3c and and Fig. 2.3d). This shows that on average the SWEET-ReLU algorithm does not produce large values of the threshold map outside the sweet spot.

Table 2.1. Sweet spots as a fraction of Ω in Near field (NF) and Focus Source (FS) instances.

	SWEET	NFC-HOA	WFS	L^2-PMM
NF	64%	32%	35%	0.7%
FS	48%	5%	0.02%	0.5%

To perform a finer analysis, we consider two additional instances: the *near-field instance*, where the source outside the arrangement at 5 m of its center, and the *focus-source instance*, where the source is inside the arrangement at 0.82 m of its center (Fig. 2.4). For these experiments we have chosen ε_i adaptively with percentile $p = 99$. The sweet spots generated by each method for each instance are shown in Fig. 2.6, 2.8, and their size is

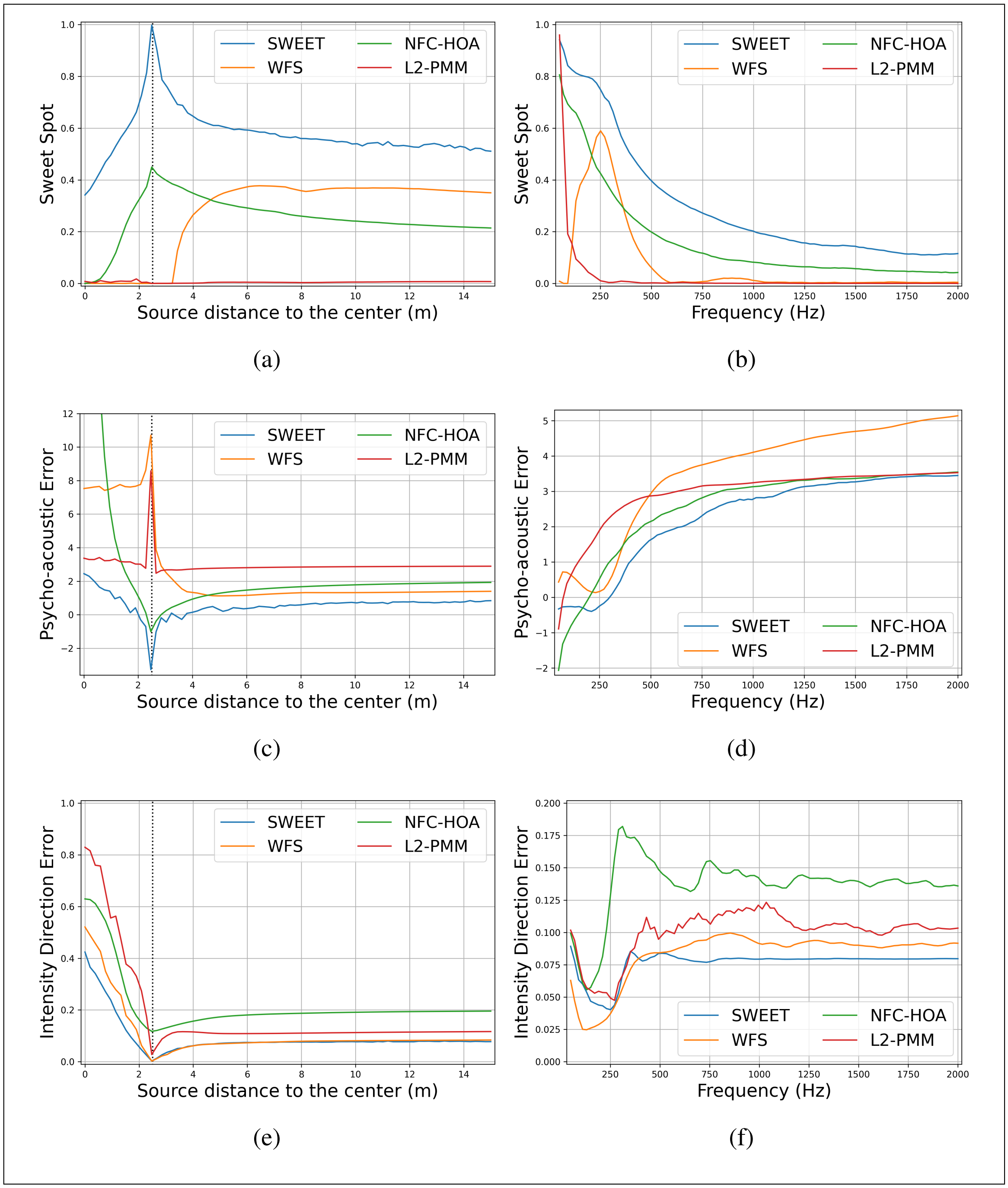
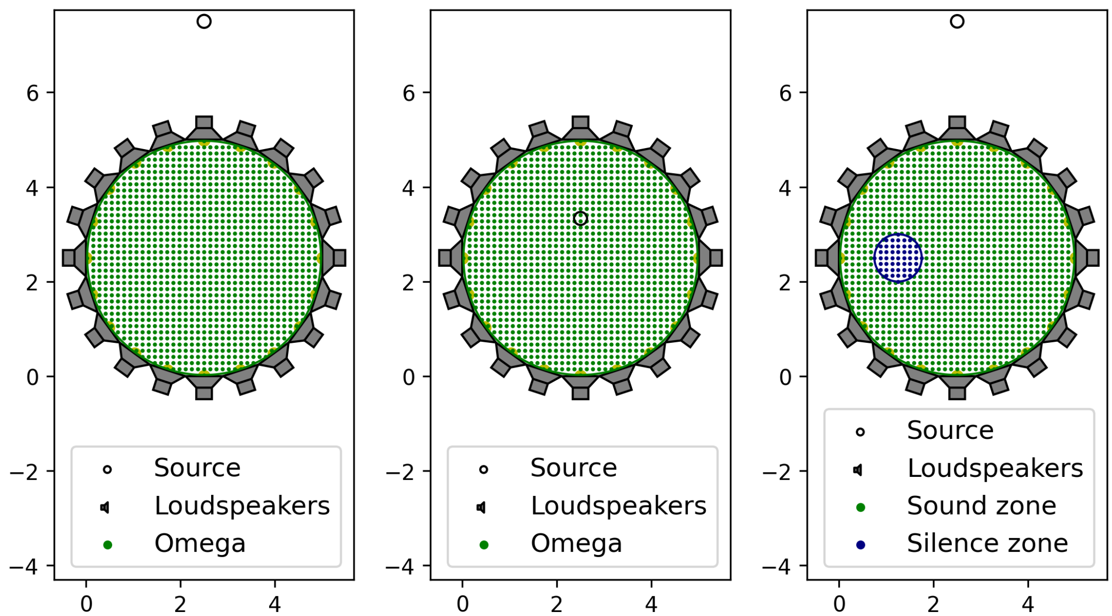


Figure 2.3. Comparison with state-of-the-art methods. *Columns:* (i) Variation of the distance between the source and the center of the arrangement, (ii) Variation of the frequency of the source. *Rows:* (i) Sweet spot as a fraction of Ω, (ii) Average value for $\log(1 + Tu)$, (iii) Average IDE.



(a) (b) (c)

Figure 2.4. Experimental setup. (a) Near field setup. (b) Focus source setup. (c) Multiple zone setup.

shown in Table 2.1. For the near field instance, the sweet spot generated by our method is almost twice as large as that of the other methods. The sweet spot generated by NFC-HOA (Fig. 2.6f) is centered, whereas that generated by WFS (Fig. 2.6f) is localized farther away from the source. This is consistent with the analysis in (Daniel et al., 2003). In contrast, the sweet spot generated by our method (Fig. 2.6e) behaves like that generated by WFS,

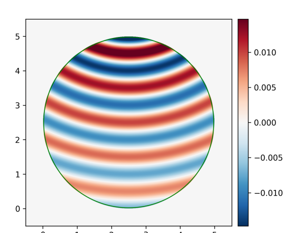


Figure 2.5. u_0 for near field setup.

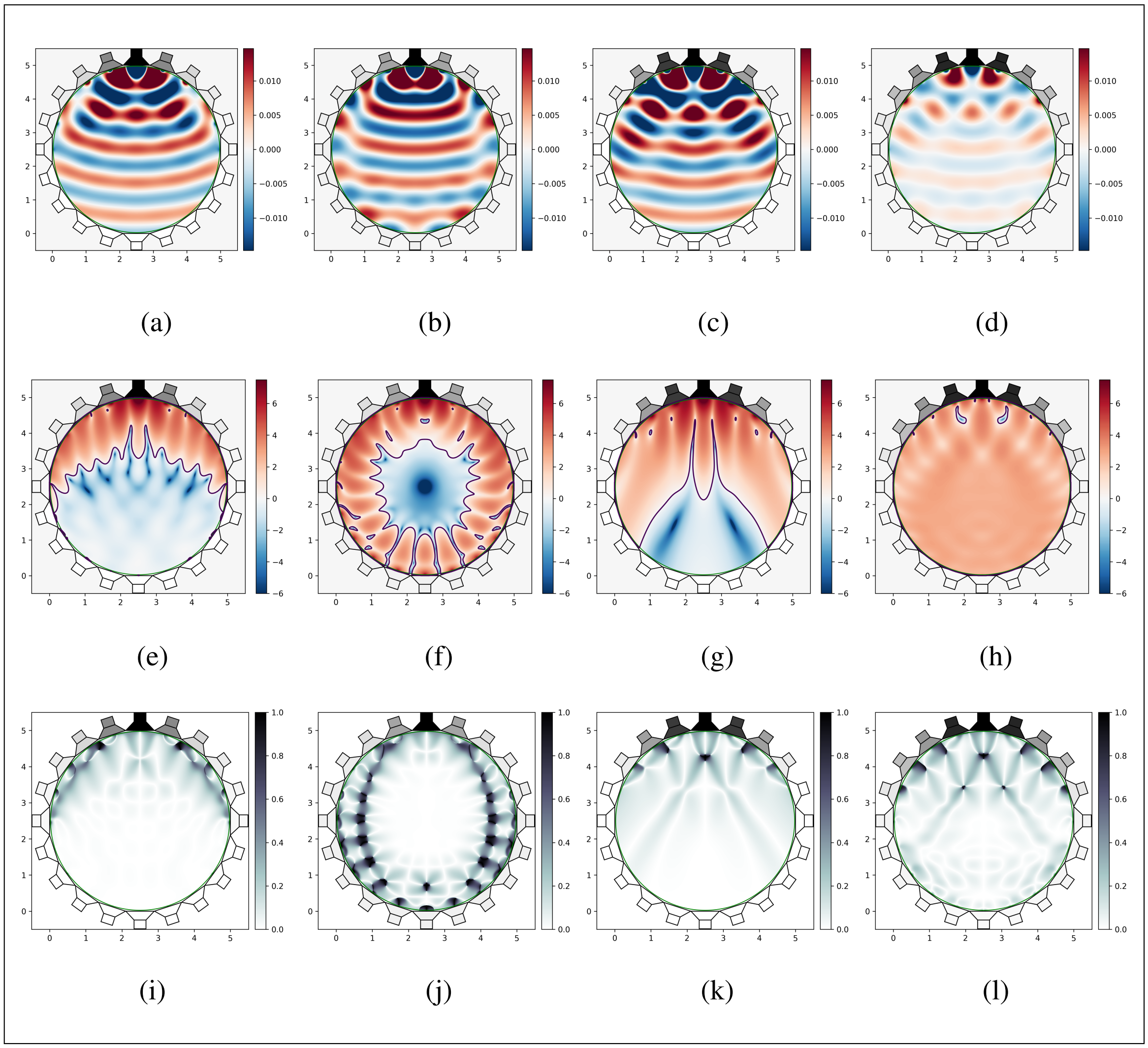


Figure 2.6. Near field instance. *Rows:* (i) u, (ii) $\log(1 + Tu)$, (iii) IDE. *Columns:* (i) SWEET-ReLU, (ii) NFC-HOA, (iii) WFS, (iv) L^2-PMM.

but almost encompasses the one generated by NFC-HOA. In all cases the aliasing artifacts appear roughly near the boundary of the sweet spot. This suggests that the principle behind sound field reconstruction, i.e., to avoid physically noticeable artifacts, does ensure a good monaural auditory scene. Our method exhibits less aliasing artifacts than the others. This may explain the low average IDE values and small psycho-acoustic errors in Fig. 2.3.

For the focus-source instance we strengthen the intensity of the source so that the wave has an amplitude of 72 dB at 1 m of the source. The sweet spot generated by our method (Fig. 2.8e) is almost 10 times larger than those generated by other methods. The sweet spot generated by NFC-HOA (Fig. 2.8f) is contained in a circle with a radius equal to the distance of the source to the center of the room. This is also consistent with (Daniel et al., 2003). The sweet spot generated by WFS (Fig. 2.8g) is almost empty as the resulting u has large amplitude. This suggest that focus source formulation for WFS needs an amplitude factor normalization. In contrast, the sweet spot generated by our method almost comprises the half of Ω that faces the source. Furthermore, the artifacts are noticeable only behind the source. This shows the advantages of the greedy strategy of the SWEET-ReLU algorithm: during its first iterations it is capable to detect the direction of u_0 over Ω to then prioritize the part of Ω where a good fit to u_0 can be obtained. This is a possible explanation for the almost empty sweet spot generated by L^2-PMM both in the near field (Fig. 2.6h) and the focus source (Fig. 2.8h) instances. This, together with the proximity of the speakers, completely degrades its performance: the method attempts to minimize the L^2-error where it is largest, i.e., near the speakers. As a consequence, the resulting u is small over Ω. Finally, our method is efficient in the usage of the loudspeakers: the acoustic wave u resulting from WFS is uncomfortably loud around the source and near

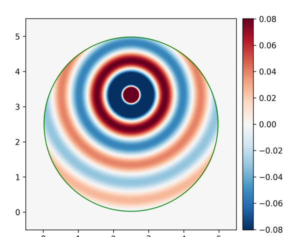


Figure 2.7. u_0 for focus source setup.

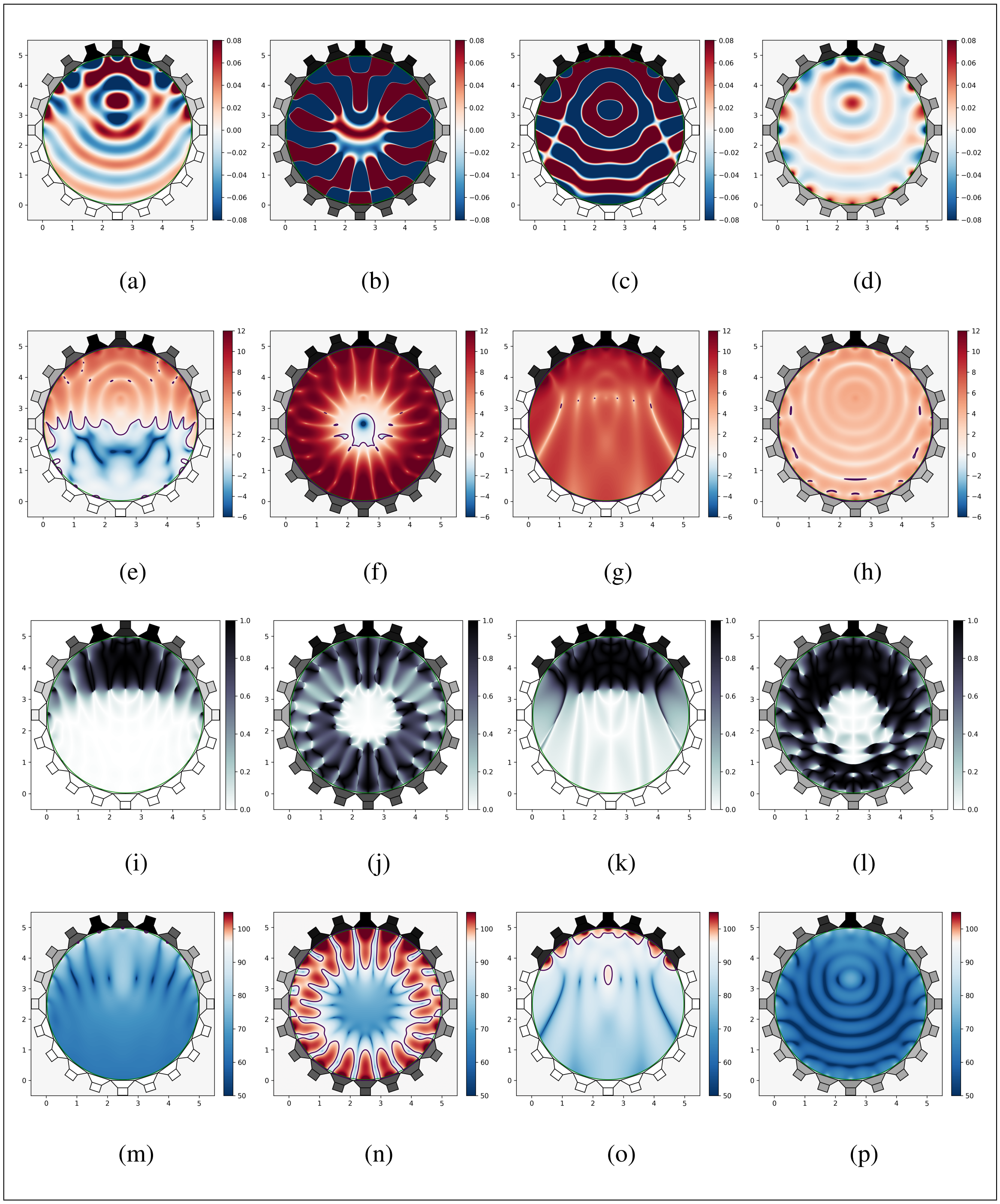


Figure 2.8. Focus source instance. *Rows:* (i) u, (ii) $\log(1 + Tu)$, (iii) IDE, (iv) Sound level (dB). *Columns:* (i) SWEET-ReLU, (ii) NFC-HOA, (iii) WFS, (iv) L^2-PMM.

the active loudspeakers in the array, whereas that obtained with NFC-HOA u_0 is uncomfortably loud in a large region outside a circumference concentric to Ω. Our method, in contrast, produces a negligible discomfort region by construction.

2.5.2. Applications

2.5.2.1. The effect of multiple frequencies

We now study the effect of a source generating a superposition of (pseudo) sinusoidal waves at $n_f = 4$ frequencies $f_1^\star = 400$ Hz, $f_2^\star = 300$ Hz, $f_3^\star = 200$ Hz, and $f_4^\star = 100$ Hz. Our goal is to study non-linear effects and their consequences on the sweet spot found for each frequency separately, and that found by solving the problem for a multi-frequency source. A uniform discretization of 9660 points was used for Ω. Contiguous points are at a distance of 0.04 m, achieving more than 19 points per wavelength in the worst case. The results are shown in the Fig. 2.10. Observe the sweet spots generated over Ω cover 54.3% of Ω for 400 Hz, 73.3% for 300 Hz, 85.5% for 200 Hz and 91% for 100 Hz. The sweet spot for the multi-frequency source covered 52% of Ω. In our standard setup it is easier to generate larger sweet spots at low frequencies, and these decrease as the frequency of the source increases. Furthermore, the sweet spots seem to be roughly nested as the frequency increases. Interestingly, the sweet spot generated for the multi-frequency source is comparable to that obtained at the highest frequency. This suggests

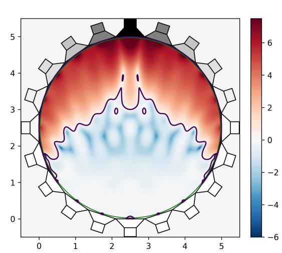


Figure 2.9. The effect of multiple frequencies: $\log(1 + Tu)$ multi frequencies.

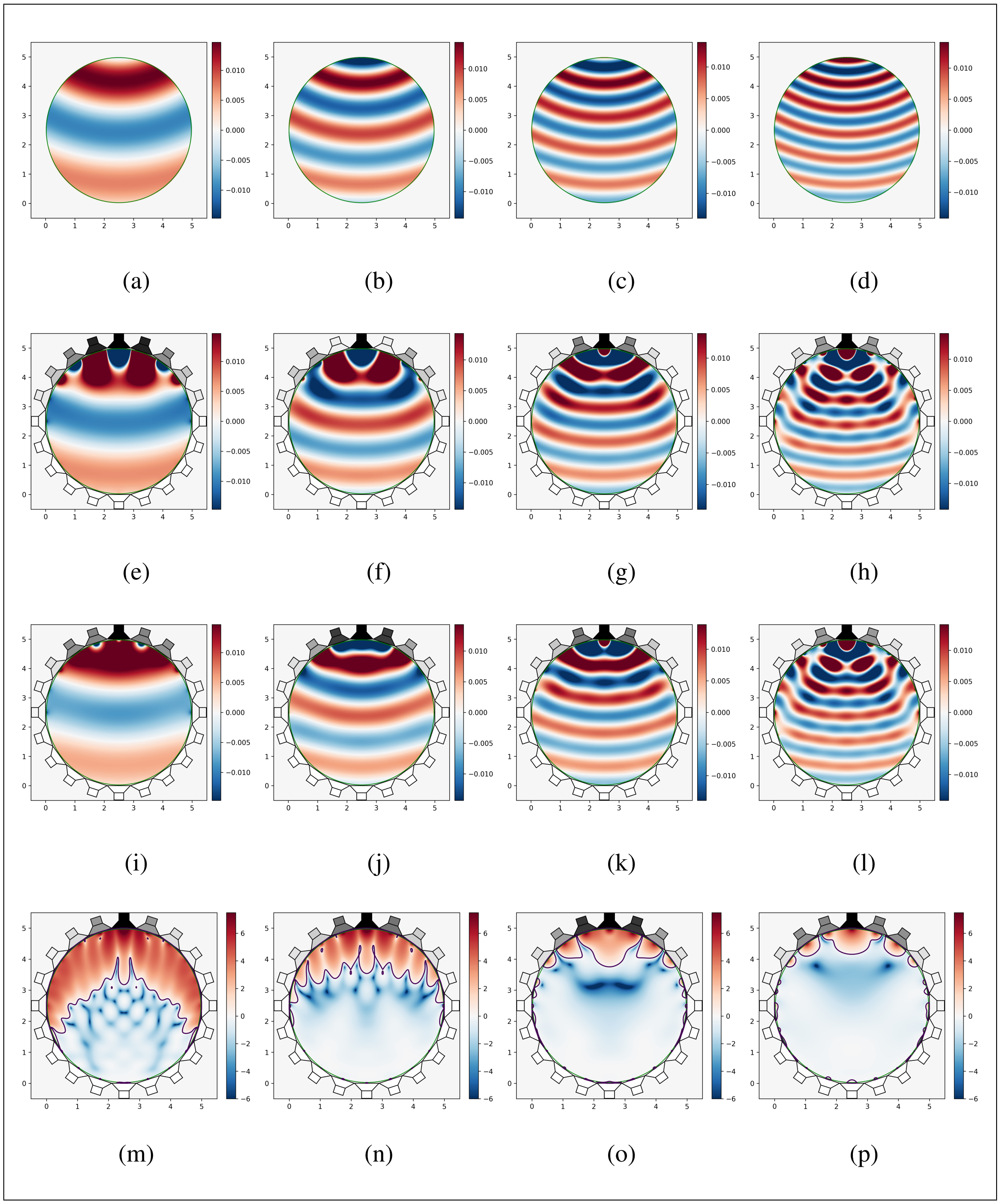


Figure 2.10. The effect of multiple frequencies. *Rows:* (i) u_0, (ii) u multi frequencies, (iii) u single frequencies, (iv) $\log(1 + Tu)$ single frequencies. *Columns:* (i) 100 Hz, (ii) 200 Hz, (iii) 300 Hz, (iv) 400 Hz.

that, in general, the sweet spot generated by our method for a multi-frequency source will be dominated by the frequency that is harder to approximate. This also yields insight into the setups for which a large sweet spot may be generated for a multi-frequency source.

2.5.2.2. Multiple zone control

The problem of creating a sound scene in a zone while keeping another silent has been extensively studied in the spatial sound literature, e.g. (Poletti, 2008; Wu & Abhayapala, 2010). Here we show our methods provide a solution to this problem. We consider the instance shown in Fig. 2.4c where u_0 is equal to 0 over the silent zone as shown in Fig. 2.11. In the silent zone we fix a psycho-acoustic tolerance of 20 dB above the absolute threshold of hearing, whereas in the zone for the sound scene, i.e., the sound zone, we keep the van de Par model as before. Since the silence zone is 24 times smaller than the sound zone, we balance the problem by choosing a non-uniform measure μ that takes the value 24 over the silent zone and 1 over the sound zone. A uniform discretization of 3274 points was used for the sound zone and 332 for the silence zone. Contiguous points are at a maximum distance of 0.075 m, achieving more than 13 points per wavelength. The results are shown in Figs. 2.12. Our method generates a sweet spot covering 32% of the sound zone, and 97.5% of the silent zone. Also, Fig. 2.12c shows that the direction of the source is correctly reproduced inside the sweet spot in the sound zone. Inside the silent zone the IDE

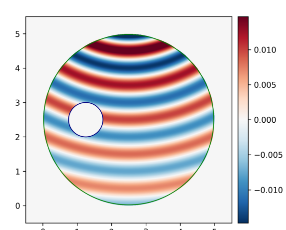


Figure 2.11. u_0 for multiple zone setup.

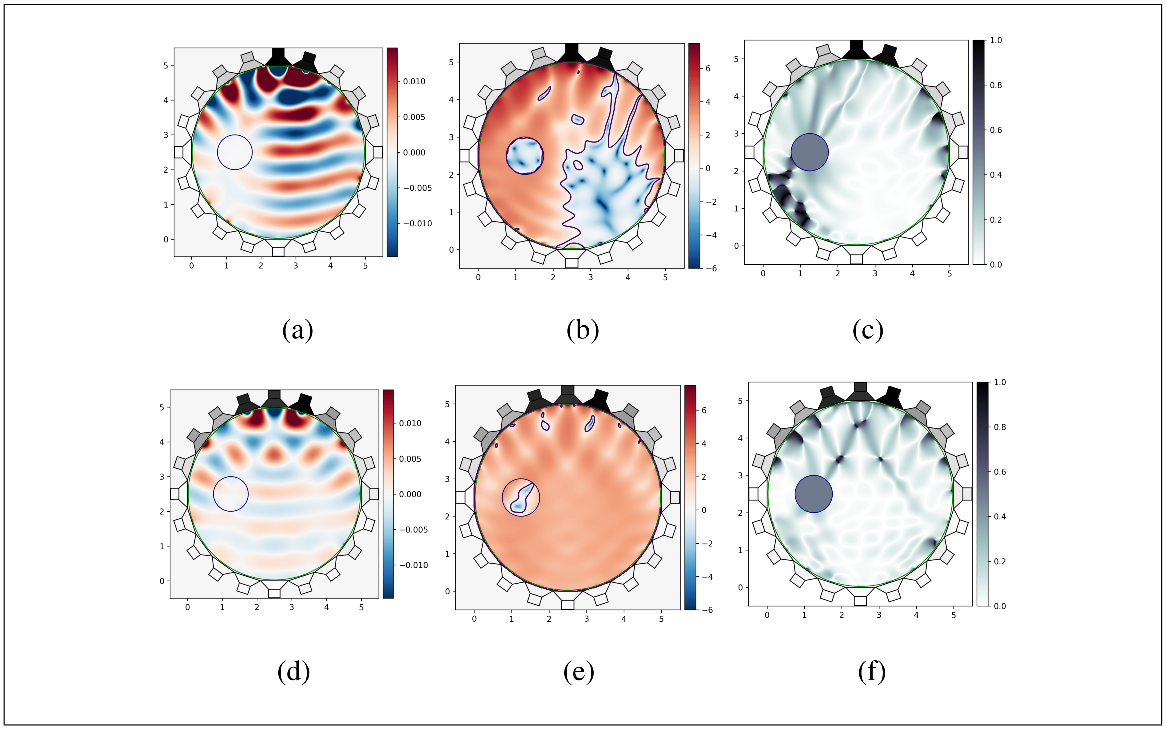


Figure 2.12. Multiple zone control. *Rows:* (i) SWEET, (ii) L^2- PMM. *Columns:* (i) u, (ii) $\log(1 + Tu)$, (iii) IDE.

is not incorrect but undefined. This indicates that the localization properties of the auditory scene may be correctly reproduced as well. In contrast, weighted L^2-PMM performs poorly in this *global* multi-zone instance for the same reasons explained in Section 2.5.1 (Figs. 2.12d-f). It generates a sweet spot covering 1.7% of the sound zone and 27% of the silent zone. This shows our method is flexible and can be used for *global* multi-zone instances.

2.6. Discussion

Our results show the SWEET-ReLU algorithm yields state-of-the-art results in standard numerical experiments. We believe the performance in these experiments is representative of what we would observe when using more complex pyscho-acoustic models for the hearing threshold and the loud discomfort level. A key component of our method is the threshold map T. Although its form is quite flexible, it does not account for spatialization and other binaural effects. Extending the form of T to account for these effects is the subject of future research. However, as it is shown in (Rumsey, Zieliński, Kassier, & Bech, 2005), the overall quality of a spatial sound system can be explained to 70% by coloration or timbral fidelity, which can be characterized by monoaural effects, and 30% by spatial fidelity, which needs to be characterized by binaural effects. Furthermore, our experiments show that in some settings our method achieves a lower intensity direction error, which is a proxy for the localization error, than state-of-the-art methods. Hence, it correctly simulates the spatial properties of the auditory scene, even though we are not explicitly enforcing it.

Although we have presented numerical results modeling the loudspeakers and the virtual sources as isotropic pseudo-sinusoudal monopoles, we believe our method can be readily implemented in real settings with non-trivial sound sources. For instance, reverberation, different radiation patterns for the loudspeakers, and other time-invariant effects can be incorporated by modifying the Green's functions G_k. For the representation of the sound scene, due to the fine discretization of the region of interest required, it may be also convenient to use an *object based* approach (Spors et al., 2013). In this case, the target sound wave u_0 is not measured with microphones, but instead is simulated when the location of the sources and their audio signals are known. Our method may be computationally expensive, as we need to solve a sequence of convex problems, precluding its use in real-time applications. Nevertheless, our multi-frequency experiments show the sweet spots nest as the frequency of a sinusoidal source increases. This suggests that an heuristic could be developed to improve the performance for multi-frequency sources. Furthermore, over a fixed instance, i.e. fixed room and loudspeaker arrangement, we may be able to approximate the map $u_0 \mapsto u$ from several simulated instances of pairs (u_0, u). Once approximated, the computational cost becomes negligible.

Finally, although we have not fully developed a theory for the convergence of SWEET-ReLU, our experiments show that it converges in practice. Further analysis will be the subject of future work.

2.7. Conclusion

In this work, we considered the sweet spot as the region where the a sound scene is psycho-acoustically close to a desired auditory scene. Furthermore, we developed a method that generates a sound scene that maximizes this sweet spot while guaranteeing no discomfort over a spatial region of interest. In this method, the sweet spot and the discomfort tolerance can be modeled within a flexible monaural psycho-acoustic framework. We provided a theoretical analysis of the method, and an efficient algorithm, the SWEET-ReLU algorithm, for its numerical implementation. Over isotropic pseudo-sinusoidal monopole instances our method successfully generates a larger sweet spot than the most common state-of-the-art sound field reconstruction methods. We believe our method is a step towards a new paradigm for spatial sound reconstruction, bridging a gap between methods based on psycho-acoustic principles, and sound field reconstruction methods.

REFERENCES

Agrawal, A., Verschueren, R., Diamond, S., & Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, *5*(1), 42–60.

Ahrens, J., & Spors, S. (2008). An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions. *Acta Acustica united with Acustica*, *94*(6), 988–999.

Alexander, R. (2013). *The inventor of stereo: the life and works of alan dower blumlein*. CRC Press.

Apostol, T. M. (1969). *Calculus, volume ii: multi-variable calculus and linear algebra, with applications to differential equations and probability*. John Wiley & Sons.

ApS, M. (2019). The mosek optimization toolbox for python manual. version 9.2.44 [Computer software manual].

Aubin, J.-P., & Frankowska, H. (1990). *Set-valued analysis*. Birkhauser.

Barbu, V., & Precupanu, T. (2012). *Convexity and Optimization in Banach Spaces* (4th ed.). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-007-2247-7

Bauschke, H. H., & Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York, NY: Springer New York. doi: 10.1007/978-1-4419-9467-7

Benedek, A., & Panzone, R. (1961). The space LP, with mixed norm. *Duke Mathematical Journal*, *28*(3), 301–324. doi: 10.1215/S0012-7094-61-02828-9

Berkhout, A. J. (1988). A holographic approach to acoustic control. *AES: Journal of the Audio Engineering Society*, *36*, 977-995.

Berkhout, A. J., de Vries, D., & Vogel, P. (1993). Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, *93*(5), 2764–2778.

Betlehem, T., & Abhayapala, T. D. (2005). Theory and design of sound field reproduction in reverberant rooms. *The Journal of the Acoustical Society of America*, *117*(4), 2100–2111.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Bosi, M., & Goldberg, R. E. (2012). *Introduction to digital audio coding and standards* (Vol. 721). Springer Science & Business Media.

Buerger, M., Hofmann, C., & Kellermann, W. (2018). Broadband multizone sound rendering by jointly optimizing the sound pressure and particle velocity. *The Journal of the Acoustical Society of America*, *143*(3), 1477–1490.

Buerger, M., Maas, R., Löllmann, H. W., & Kellermann, W. (2015). Multizone sound field synthesis based on the joint optimization of the sound pressure and particle velocity vector on closed contours. In *2015 ieee workshop on applications of signal processing to audio and acoustics (waspaa)* (pp. 1–5).

Cohn, D. L. (2013). *Measure Theory* (2nd ed.). New York, NY: Springer New York. doi: 10.1007/978-1-4614-6956-8

Daniel, J. (2000). *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia* (Unpublished doctoral dissertation). University of Paris VI.

Daniel, J., Moreau, S., & Nicol, R. (2003). Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging. In *Audio engineering society convention 114*.

Diamond, S., & Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, *17*(83), 1–5.

Disch, S., van de Par, S., Niedermeier, A., Burdiel Pérez, E., Berasategui Ceberio, A., & Edler, B. (2018). Improved psychoacoustic model for efficient perceptual audio codecs. In *Audio engineering society convention 145.*

Evans, L. C. (2010). *Partial Differential Equations* (2nd ed.). American Mathematical Society.

Fazi, F. M., & Nelson, P. A. (2007a). The ill-conditioning problem in sound field reconstruction. In *Audio engineering society convention 123.*

Fazi, F. M., & Nelson, P. A. (2007b). A theoretical study of sound field reconstruction techniques. In *19th international congress on acoustics.*

Fazi, F. M., & Nelson, P. A. (2012). Nonuniqueness of the solution of the sound field reproduction problem with boundary pressure control. *Acta Acustica united with Acustica*, *98*(1), 1-14.

Fazi, F. M., Nelson, P. A., Christensen, J. E., & Seo, J. (2008). Surround system based on three-dimensional sound field reconstruction. In *Audio engineering society convention 125.*

Fazi, F. M., Nelson, P. A., & Potthast, R. (2009). Analogies and differences between three methods for sound field reproduction. *Relation*, *10*, 3.

Feng, Q., Yang, F., & Yang, J. (2018). Time-domain sound field reproduction using the group lasso. *The Journal of the Acoustical Society of America*, *143*, EL55-EL60.

Firtha, G. (2019). *A generalized wave field synthesis framework with application for moving virtual sources* (Unpublished doctoral dissertation). Budapest University of Technology and Economics.

Firtha, G., Fiala, P., Schultz, F., & Spors, S. (2017). Improved referencing schemes for 2.5 d wave field synthesis driving functions. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(5), 1117–1127.

Firtha, G., Fiala, P., Schultz, F., & Spors, S. (2018). On the general relation of wave field synthesis and spectral division method for linear arrays. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, *26*, 2393-2403.

Franck, A., Wang, W., & Fazi, F. M. (2017). Sparse 1-optimal multiloudspeaker panning and its relation to vector base amplitude panning. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, *25*.

Gauthier, P.-A., Berry, A., & Woszczyk, W. (2005). Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain. *The Journal of the Acoustical Society of America*, *117*, 662-678.

Gauthier, P.-A., Lecomte, P., & Berry, A. (2017). Source sparsity control of sound field reproduction using the elastic-net and the lasso minimizers. *The Journal of the Acoustical Society of America*, *141*, 2315-2326.

Gerzon, M. A. (1973). Periphony: With-height sound reproduction. *AES: Journal of the Audio Engineering Society*, *21*(1), 2–10.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, *47*(1-2), 103–138.

Hacihabiboglu, H., De Sena, E., Cvetkovic, Z., Johnston, J., & Smith III, J. O. (2017). Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-audio techniques based on psychoacoustics. *IEEE Signal Processing Magazine*, *34*(3), 36–54.

Horst, R., & Thoai, N. V. (1999). DC Programming: Overview. *Journal of Optimization Theory and Applications*, *103*(1), 1–43. doi: 10.1023/A:1021765131316

Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America*, *124*, 422-438.

Jia, M., Zhang, J., Wu, Y., & Wang, J. (2018). Sound field reproduction via the alternating direction method of multipliers based lasso plus regularized least-square. *IEEE Access*, *6*, 54550-54563.

Johnston, J. D., & Lam, Y. H. V. (2000). Perceptual soundfield reconstruction. In *Audio engineering society convention 109.*

Kirkeby, O., & Nelson, P. A. (1993). Reproduction of plane wave sound fields. *The Journal of the Acoustical Society of America*, *94*(5), 2992–3000.

Kirkeby, O., Nelson, P. A., Orduna-Bustamante, F., & Hamada, H. (1996). Local sound field reproduction using digital signal processing. *The Journal of the Acoustical Society of America*, *100*(3), 1584–1593.

Kirsch, A., & Hettlich, F. (2016). *Mathematical theory of time-harmonic maxwell's equations*. Springer.

Knobel, K. A. B., & Sanchez, T. G. (2006). Nível de desconforto para sensação de intensidade em indivíduos com audição normal. *Pró-Fono Revista de Atualização Científica*, *18*(1), 31–40.

Krantz, S. G., & Parks, H. R. (2002). *A Primer of Real Analytic Functions*. Boston, MA: Birkhäuser Boston. doi: 10.1007/978-0-8176-8134-0

Lee, T., Nielsen, J. K., & Christensen, M. G. (2020). Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, *28*, 2412-2426.

Lilis, G. N., Angelosante, D., & Giannakis, G. B. (2010). Sound field reproduction using the lasso. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *18*, 1902-1912.

Lipp, T., & Boyd, S. (2016). Variations and extension of the convex–concave procedure. *Optimization and Engineering*, *17*(2), 263–287. doi: 10.1007/s11081-015-9294-x

Lipshitz, S. P. (1986). Stereo microphone techniques: Are the purists wrong? *J. Audio Eng. Soc*, *34*, 716, 717, 719-722, 724-726, 728-738, 740, 742-744.

Painter, T., & Spanias, A. (2000). Perceptual coding of digital audio. *Proceedings of the IEEE*, *88*(4), 451–515.

Pierce, A. D. (2019). *Acoustics: an introduction to its physical principles and applications*. Springer.

Plasberg, J. H., & Kleijn, W. B. (2007). The sensitivity matrix: Using advanced auditory models in speech and audio processing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *15*, 310-319.

Poletti, M. (2005). Three-dimensional surround sound systems based on spherical harmonics. *AES: Journal of the Audio Engineering Society*, *53*.

Poletti, M. (2008). An investigation of 2-d multizone surround sound systems. In *Audio engineering society convention 125.*

Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *AES: Journal of the Audio Engineering Society*, *45*(6), 456–466.

Pulkki, V. (2001). *Spatial sound generation and perception by amplitude panning techniques*. Helsinki University of Technology.

Quarteroni, A., Sacco, R., & Saleri, F. (2010). *Numerical mathematics* (Vol. 37). Springer Science & Business Media.

Radmanesh, N., & Burnett, I. S. (2013). Generation of isolated wideband sound fields using a combined two-stage lasso-ls algorithm. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *21*, 378-387.

Royden, H. L., & Fitzpatrick, P. (2010). *Real analysis* (4th ed.). Pearson.

Rudin, W. (1986). *Real and Complex Analysis* (3rd ed.). McGraw-Hill Education.

Rumsey, F., Zieliński, S., Kassier, R., & Bech, S. (2005). On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America*, *118*(2), 968–976.

Samarasinghe, P. N., Poletti, M., Salehin, S. A., Abhayapala, T. D., & Fazi, F. M. (2013). 3d soundfield reproduction using higher order loudspeakers. In *2013 ieee international conference on acoustics, speech and signal processing* (pp. 306–310).

Sena, E. D., Hacihabiboglu, H., & Cvetkovic, Z. (2013). Analysis and design of multichannel systems for perceptual sound field reconstruction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *21*, 1653-1665.

Sherlock, L. P., & Formby, C. (2005). Estimates of loudness, loudness discomfort, and the auditory dynamic range: normative estimates, comparison of procedures, and test-retest reliability. *Journal of the American Academy of Audiology*, *16*(02), 085–100.

Shin, M., Nelson, P. A., Fazi, F. M., & Seo, J. (2016). Velocity controlled sound field reproduction by non-uniformly spaced loudspeakers. *Journal of Sound and Vibration*, *370*, 444-464.

Skudrzyk, E. (2012). *The foundations of acoustics: basic mathematics and basic acoustics*. Springer Science & Business Media.

Sogge, C. D. (2017). *Fourier Integrals in Classical Analysis* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/9781316341186

Spors, S., & Ahrens, J. (2008). A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling. In *Audio engineering society convention 125.*

Spors, S., Rabenstein, R., & Ahrens, J. (2008). The theory of wave field synthesis revisited. In *In 124th convention of the aes.*

Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., & Zotter, F. (2013). Spatial sound with loudspeakers and its perception: A review of the current state. *Proceedings of the IEEE*, *101*(9), 1920–1938.

Start, E. W. (1996). Application of curved arrays in wave field synthesis. In *Audio engineering society convention 100.*

Taal, C. H., Hendriks, R. C., & Heusdens, R. (2012). A low-complexity spectro-temporal distortion measure for audio processing applications. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *20*, 1553-1564.

Tao, P. D., & An, L. T. H. (1997). Convex Analysis Approach to D.C. Programming: Theory, Algorithms and Applications. *Acta Mathematica Vietnamica*, *22*(1), 289–355.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing research*, *1*(2), 155–182.

Ueno, N., Koyama, S., & Saruwatari, H. (2019). Three-dimensional sound field reproduction based on weighted mode-matching method. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, *27*, 1852-1867.

Union, I. T. (2012). Multichannel stereophonic sound system with and without accompanying picture. *ITU-R Recommendations*(BS.775-3).

van de Par, S., Kohlrausch, A., Heusdens, R., Jensen, J., & Jensen, S. H. (2005). A perceptual model for sinusoidal audio coding based on spectral integration. *EURASIP Journal on Advances in Signal Processing*, *2005*(9), 1–13.

Ward, D. B., & Abhayapala, T. D. (2001). Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, *9*.

Wierstorf, H., Hohnerlein, C., Spors, S., & Raake, A. (2014). Coloration in wave field synthesis. In *Audio engineering society conference: 55th international conference: Spatial audio.*

Wierstorf, H., Raake, A., & Spors, S. (2013). Localization in wave field synthesis and higher order ambisonics at different positions within the listening area. In *Proceeding of german annual conference on acoustics (daga).*

Wierstorf, H., & Spors, S. (2012). Sound field synthesis toolbox. In *Audio engineering society convention 132.*

Williams, E. G. (1999). *Fourier acoustics: sound radiation and nearfield acoustical holography*. Academic press.

Wu, Y. J., & Abhayapala, T. D. (2009). Theory and design of soundfield reproduction using continuous loudspeaker concept. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *17*, 107-116.

Wu, Y. J., & Abhayapala, T. D. (2010). Spatial multizone soundfield reproduction: Theory and design. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *19*(6), 1711–1720.

Ziemer, T. (2020). *Psychoacoustic music sound field synthesis*. Springer.

Ziemer, T., & Bader, R. (2017). Psychoacoustic sound field synthesis for musical instrument radiation characteristics. *AES: Journal of the Audio Engineering Society*, *65*, 482-496.

Zuo, H., Abhayapala, T. D., & Samarasinghe, P. N. (2020). Particle velocity assisted three dimensional sound field reproduction using a modal-domain approach. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, *28*, 2119-2133.

Zwicker, E., & Fastl, H. (2007). *Psychoacoustics: Facts and models* (Vol. 22). Springer Science & Business Media.

APPENDIX

A. PROOF OF PROPOSITION 1

We prove some auxiliary results. First, we claim the map

$$\mathcal{K}_B u(t,x) = \int K_B(t,t',x)u(t',x)\,dt',$$

where K_B satisfies the hypotheses, is continuous from W into W. To prove this, fix $x \in \Omega$ and apply Young's inequality for integral operators (Sogge, 2017, Theorem 0.3.1) to obtain

$$\int |\mathcal{K}_B u(t,x)|^2\,dt = \int \left| \int K_B(t,t',x)u(t',x)\,dt' \right|^2 dt \leq C_B^2 \int |u(t,x)|^2\,dt,$$

from where it follows that $\|\mathcal{K}_B u\|_W \leq C_B \|u\|_W$ and, in particular, $\mathcal{K}_B u \in W$. Second, a functional B of the form (2.8) is bounded. This is clear from the fact that

$$|Bu(x)| = \int |\mathcal{K}_B(u-u_0)(t,x)|^2\,dt \leq C_B^2 \|u-u_0\|_W^2.$$

Third, for any $\theta \in [0,1]$ it is apparent that

$$B(\theta u_1 + (1-\theta)u_2)(x) \leq \theta B(u_1)(x) + (1-\theta)B(u_2)(x).$$

whence for almost every x the map $u \mapsto Bu(x)$ is convex. Fourth, Bu is a measurable function by Fubini's theorem (Cohn, 2013, Theorem 5.2.2). Fifth, B is continuous on u. To prove this, let $v = |\mathcal{K}_B u_2| + |\mathcal{K}_B u_1| + 2|\mathcal{K}_B u_0|$ and $w = \mathcal{K}_B u_2 - \mathcal{K}_B u_1$ and note that

$$|Bu_2(x) - Bu_1(x)|^2 \leq \int |v(t,x)|^2\,dt \int |w(t,x)|^2\,dt$$

where we used the identity $|a^2 - b^2| = |a+b||a-b|$, the Cauchy-Schwarz inequality and the triangle inequality. The first term is bounded, as

$$\begin{aligned}\int |v(t,x)|^2\,dt &\leq 3\|\mathcal{K}_B u_1\|_W^2 + 3\|\mathcal{K}_B u_2\|_W^2 + 6\|\mathcal{K}_B u_0\|_W^2 \\ &\leq 3C_B^2(\|u_1\|_W^2 + \|u_2\|_W^2 + 2\|u_0\|_W^2),\end{aligned}$$

where we used the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$. For the second, we have

$$\int |w(t,x)|^2\,dt \leq \|\mathcal{K}_B(u_1-u_2)\|_W^2 \leq C_B^2\|u_1-u_2\|_W^2.$$

It follows that $u_1 \to u_2$ in W implies $Bu_1 \to Bu_2$ in $L^\infty(\Omega)$ whence $B: W \to L^\infty(\Omega)$ is continuous.

Proof of (i): We prove the result for $n_b = 2$ for simplicity. Since Ψ in (2.9) is continuous and each $B_k : W \to L^\infty(\Omega)$ is continuous by our auxiliary results, $T : W \to L^\infty(\Omega)$ is continuous. Similarly, since Ψ is convex and non-decreasing on every component, for every $\theta \in [0,1]$ we have

$$\begin{aligned} T(\theta u_1 + (1-\theta)u_2)(x) &\leq \Psi(\theta v_{1,1}(x) + (1-\theta)v_{1,2}(x), \theta v_{2,1}(x) + (1-\theta)v_{2,2}(x)) \\ &\leq \theta T u_1(x) + (1-\theta)Tu_2(x) \end{aligned}$$

where $v_{k,j} = B_k u_j$, proving the convexity of $u \to Tu(x)$ for almost every x. Analogously, P is continuous and $u \to Pu(x)$ is convex for almost every x.

Proof of (ii): Note that Ψ is measurable, because it is continuous, and so is $B_k u$. Therefore, T in (2.9) is measurable, and the set $\mathcal{S}(u)$ is measurable for any $u \in W$.

Proof of (iii): We show that W_S is a family functions $\Omega \mapsto L^2(\Omega)$ that is bounded, equicontinuous, and defined on a separable metric space, whence, by Arzelà-Ascoli's theorem (Rudin, 1986, Theorem 11.28), is compact with respect to the uniform norm, which coincides with the W-norm. Let $\gamma_{\max}$ be the uniform bound on $\|c_k\|_{L^2}$. Consider the map $\Omega \to L^2(\Omega)$ given by $x \to u_x$ where u_x denotes the function $t \mapsto u(t,x)$. Since each G_k is bounded on $I_c \times \overline{\Omega}$ then

$$\begin{aligned} \int |u(t,x)|^2\,dt &\leq n_s \sum_{k=1}^{n_s} \int_{I_c} |\widehat{c}_k(f)|^2 |G_k(f,x)|^2\,df \\ &\leq n_s \sum_{k=1}^{n_s} \sup_{(f,x)\in I_c\times\bar{\Omega}} |G_k(f,x)|^2 \|c_k\|_{L^2}^2 \end{aligned}$$

whence $x \mapsto u_x$ is uniformly bounded. To prove equicontinuity, fix $\varepsilon > 0$. Since each G_k is continuous on the compact set $I_c \times \bar{\Omega}$, we can find $\delta > 0$ such that for any $|x - y| < \delta$ and $f \in I_c$ we have $|G_k(f, x) - G_k(f, y)| < \varepsilon / n_s^2 \gamma_{\max}^2$. Then

$$\begin{aligned}\int |u(t,x) - u(t,y)|^2 \, dt &\leq n_s \sum_{k=1}^{n_s} \int_{I_c} |\widehat{c}_k(f)|^2 |G_k(f,y) - G_k(f,x)|^2 \, df \\ &< n_s \gamma_{\max}^2 \left(n_s \frac{\varepsilon}{n_s^2 \gamma_{\max}^2} \right) \\ &= \varepsilon\end{aligned}$$

showing not only that $x \mapsto u_x$ is continuous, but equicontinuous. We conclude W_S is compact in W.

Proof of (iv): We omit details for brevity. The map $P : W \to L^\infty(\Omega)$ is continuous by the same arguments we used in the proof of (i). Then, since non-strict inequalities are preserved under limits, $\mathcal{P}$ is closed.

B. PROOF OF PROPOSITION 8

Define the auxiliary variables

$$\begin{aligned} a_t &= \left| \int_{\mathbb{R}} K_B(t, t', x)(u(t, t', x) - u_0(t, t', x)) dt' \right| \\ b_t &= \left| \int_{\mathbb{R}} K_B(t, t', y)(u(t, t', y) - u_0(t, t', y)) dt' \right|. \end{aligned}$$

Then,

$$|Bu(x) - Bu(y)|^2 \leq \left(\int (a_t + b_t)^2 dt \right) \left(\int (a_t - b_t)^2 dt \right)$$

where we used the identity $|a^2 - b^2| = |a+b||a-b|$ and the Cauchy-Schwarz inequality. For the first term, because of Cauchy Schwarz and Young's inequalities we have

$$\begin{aligned}\int (a_t + b_t)^2 dt &= \int a_t^2 dt + \int b_t^2 dt + 2\int a_t b_t dt \\ &\int a_t^2 dt + \int b_t^2 dt + 2\left(\int a_t^2 dt\right)^{\frac{1}{2}} \left(\int b_t^2 dt\right)^{\frac{1}{2}} \\ &\leq C_B^2 \|u - u_0\|_W^2 + C_B^2 \|u - u_0\|_W^2 + 2C_B^2 \|u - u_0\|_W^2 \\ &= (2C_B \|u - u_0\|_W)^2\end{aligned}$$

For the second term we have

$$\begin{aligned}\int (a_t - b_t)^2 dt &\leq \int \left| \int |K_B(t,t',x)(u-u_0)(x,t') - K_B(t,t',y)(u-u_0)(y,t')| dt' \right|^2 dt \\ &\leq \int \left| \int |K_B(t,t',x)((u-u_0)(x,t') - (u-u_0)(y,t'))| dt' \right|^2 dt \\ &\quad + \int |(K_B(t,t',x) - K_B(t,t',y))(u-u_0)(y,t')| dt'|^2 \, dt \\ &\leq \underbrace{\int \left| \int |K_B(t,t',x)((u-u_0)(x,t') - (u-u_0)(y,t'))| dt' \right|^2 dt}_{(*)} \\ &\quad + \underbrace{\int \left| \int |(K_B(t,t',x) - K_B(t,t',y))(u-u_0)(y,t')| dt' \right|^2 dt}_{(**)},\end{aligned}$$

where we used the the triangle inequality, identity $ab - cd = a(b-d) + d(a-c)$, and Minkowski's inequality. Now, because of Young's inequality,

$$(*) \leq 2C_B^2 \int |(u-u_0)(x,t') - (u-u_0)(y,t')|^2 dt.$$

Since $u - u_0 \in W$, as shown in the proof of Proposition 1, the latter converges to 0 in the limit $x \to y$ uniformly in the choice of u. Also, because of Cauchy-Schwarz inequality,

$$
\begin{aligned}
(**) &\leq \int \left(\int |K_B(t,t',x) - K_B(t,t',y)|^2 dt' \right) \left(\int |(u-u_0)(y,t')|^2 dt' \right) dt \\
&\leq \|u - u_0\|_W^2 \left(\int \int |K_B(t,t',x) - K_B(t,t',y)|^2 dt' dt \right)^{\frac{1}{2}}.
\end{aligned}
$$

Now, note that

$$
\begin{aligned}
|K_B(t,t',x) - K_B(t,t',y)|^2 &\leq |K_B(t,t',x)|^2 + 2|K_B(t,t',x)K_B(t,t',y)| + |K_B(t,t',y)|^2 \\
&\leq 4 \sup_{z \in \Omega} |K_B(t,t',z)|^2.
\end{aligned}
$$

which means, by hypothesis, that the last integrand is dominated. Therefore, by Lebesgue's dominated convergence theorem we have that $(**)$ converges to 0 in the limit $x \to y$. It follows that $Bu \in C(\Omega)$. Since Ψ is continuous we have that $Tu \in C(\Omega)$.