Applied Acoustics 134 (2018) 138–144

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Technical note

Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction

Zamir Ben-Hur*, Jonathan Sheaffer, Boaz Rafaely

Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel

ARTICLE INFO

Keywords:
Microphone arrays
Spherical harmonics
Binaural signals
Binaural reproduction

ABSTRACT

With the recent proliferation of spherical microphone arrays for sound field recording, methods have been developed for rendering binaural signals from these recordings and free-field head related transfer functions (HRTFs). Employing spherical arrays naturally leads to methods that are formulated in the spherical harmonics (SH) domain, using order-limited SH representations. However, the incorporation of HRTFs and enclosed sound fields typically leads to methods that are formulated in the space domain using plane-wave (PW) representation. Although these two representations are widely used, the current literature does not offer a complete theoretical framework to derive sampled PW representation from the SH representation in the context of binaural reproduction and sound perception. This paper develops a mathematical framework showing that when specific conditions for the joint sampling of the sound field and the HRTFs are maintained, sampled PW representation can be derived from the SH representation without error, and the resulting binaural signals are independent of the employed spatial sampling scheme. Furthermore, analysis of the aliasing error shows that the sound field is more sensitive to aliasing than the HRTFs. The theoretical analysis is complemented by a listening experiment, in which both PW and SH representations are perceptually evaluated for different spatial sampling schemes, SH orders, and levels of aliasing when deviating from the joint sampling conditions.

1. Introduction

Rendering binaural signals from microphone array recordings and from computer simulations plays an important role in the auralization of architectural acoustics [1,2], in hearing research [3], and in virtual reality [4]. Traditionally, binaural signals are obtained using microphones placed at the left and right ear canals of a manikin, in which case the transfer functions of the room and the head are jointly captured. Alternatively, if an array of microphones is used for the recording, it is then possible to synthesize binaural signals in post processing [5,6].

Spherical microphone arrays have recently gained popularity for capturing and analyzing sound fields. One of the advantages of using spherical arrays is the ease of the representation of signals in the spherical harmonics (SH) domain, leading to the well known Ambisonics format in spatial audio [7], and to a full three-dimensional representation of the sound field [8]. However, for spherical arrays with a limited number of microphones, the information on the captured sound field in the SH domain is order limited [9], leading to sound reproduction with limited spatial resolution [10]. One approach to overcome this limitation is to estimate the true, high-order sound field from the limited measurements, using sparse recovery and compressed sensing [11,12], for example. These approaches, although appropriate for sound fields composed of a small number of plane waves, cannot guarantee accurate binaural reproduction in the general case. Therefore, most methods for binaural reproduction from spherical microphone arrays use the measured sound field in a more direct manner. These methods can be broadly divided into two categories, one based on plane-wave (PW) representation and one on SH representation. In addition, this categorization is also relevant for methods of binaural reproduction of simulated sound fields, although typically allowing for higher spatial resolution compared to measured fields [13,14].

In the SH approach, binaural signals are generated by using a summation of the product of the SH coefficients of the sound field [8] with SH coefficients of the free-field HRTFs [15,16]. With this approach, binaural reproduction can be readily integrated with other SH computations, such as head rotations [17] and beamforming [18]. While this representation is natural for spherical arrays [19], accurate computation of the SH coefficients of HRTFs requires suitable sampling grids, leading to errors if the grid does not cover the full sphere of directions [20].

With the aim of relaxing the constraints on the HRTF sampling grid,

* Corresponding author.
E-mail address: zami@post.bgu.ac.il (Z. Ben-Hur).

https://doi.org/10.1016/j.apacoust.2018.01.016
Received 14 October 2017; Received in revised form 13 December 2017; Accepted 17 January 2018

researchers have developed methods that use PW representation of both the sound field and the HRTFs. While this representation is natural for HRTFs, as it can employ directly the measured transfer functions, various novel methods have been developed for the more challenging task of producing PW representation for the sound field. Duraiswami et al. [6] employed direct sampling of the order-limited PW expansion of the captured sound field at the measured HRTF directions, which is equivalent to using maximum-directivity beamforming to estimate the PW amplitudes [19]. Other beamforming methods were also investigated within the same context [21–23], including a comparison between modal and delay-and-sum beamforming [24,25]. While the PW approach is attractive, the relation between the various approaches for PW representation of the sound field, the role of the sampling grid in this representation, the effect of the SH orders of the sound field and HRTF functions, and the relation between the PW and the SH representations, has not been clearly established in the current literature. For example, while Bernschütz et al. [26] noted in a recent study that the spatial sampling scheme of the sound field affected spatial perception, no theoretical framework that can explain this observation was provided.

This paper presents the derivation of sampling conditions for the joint sampling of sound fields and HRTFs for binaural reproduction. It extends the existing theory for sampling individual functions on the sphere [19]. The joint sampling conditions involve the upper and lower bounds of the SH orders of the sound field and the HRTFs, and the attributes of the joint sampling grid. Furthermore, the sampling grids maintaining these conditions, which are not unique, guarantee equivalence between the PW and the SH representations for binaural reproduction.

The contributions of this paper are as follows:

1. A mathematical formulation of exact conditions for the joint sampling of the sound field and the HRTFs to obtain equivalence of the binaural signals when reproduced using PW and SH representations (Section 4).
2. Objective validation of the joint sampling conditions, providing insight into the effect of the SH order of the PW and HRTF representations on the binaural reproduction error (Section 5).
3. Validation of the joint sampling conditions and objective evaluations by a listening experiment (Section 6).

In addition, Section 2 reviews the basic theory of binaural reproduction from spherical microphone array recordings, Section 3 presents the theory for sampling an individual function on the sphere and Section 7 outlines the conclusions of the research.

2. Binaural reproduction from spherical microphone array recordings

This section provides a brief overview of current formulations for binaural reproduction using both SH and PW representations. Consider a sound field composed of a continuum of PWs, such that it could be described by a PW density function, $a(k,\Omega)$. The pressure observed at the left ear of a listener can be represented by [6,27]:

$$p^l(k) = \int_{\Omega\in S^2} a(k,\Omega)h^l(k,\Omega)\mathrm{d}\Omega, \tag{1}$$

where $k = 2\pi f/c$ is the wave number, f is the frequency, c is the speed of sound, $\Omega \equiv (\theta,\phi) \in \mathbb{S}^2$ is the spatial angle, and $h(k,\Omega)$ is the HRTF [28]. The superscript l denotes the left ear (the computation for the right ear can be done in a similar fashion), $p(k)$ is the pressure at the ear and $\int_{\Omega\in S^2}(\cdot)\mathrm{d}\Omega \equiv \int_0^{2\pi}\int_0^{\pi}(\cdot)\sin(\theta)\mathrm{d}\theta\mathrm{d}\phi$.

Alternatively, by applying Parseval's relation to Eq. (1), the pressure at the ear can be represented in the SH domain [16] by:

$$p^l(k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \tilde{a}_{nm}^*(k)h_{nm}^l(k), \tag{2}$$

where $\tilde{a}_{nm}(k)$ is the spherical Fourier transform (SFT) of $a^*(k,\Omega)$,$(\cdot)^*$ denotes the complex conjugate, and $h_{nm}^l(k)$ is the SFT of $h^l(k,\Omega)$. For spatially band-limited functions, with SH orders corresponding to N_h and N_a for the HRTF and PW density function, respectively, the infinite summation in Eq. (2) will be truncated to an order that matches the lower order function, $\underline{N} = \min(N_h,N_a)$,

$$p^l(k) = \sum_{n=0}^{\underline{N}}\sum_{m=-n}^{n} \tilde{a}_{nm}^*(k)h_{nm}^l(k). \tag{3}$$

Both functions are indeed expected to be order-limited in practice. $a_{nm}(k)$, derived from spherical microphone array recordings, can be considered order-limited with a maximum order that depends on the operating frequency and array radius [19]. For example, by capturing a sound field using the eigenmike [29] with a radius of 4.2 cm, the maximum order at the speech frequency range is about $N_a = 4$. Furthermore, as demonstrated by Zhang el al. [30], the HRTF can also be considered to be order limited in practice, where the maximum order is also frequency dependent [16,30]. For example, Rafaely and Avni [16] showed that in order to achieve less than -10 dB error in the representation of the HRTF up to 8 kHz, the order should be $N_h = 12$.

The sampling conditions for computing the order-limited functions from samples are well understood, and have been previously presented [19] and applied to compute both the PW density function from microphone signals [7] and the SH coefficients of the HRTFs from measured transfer functions [15]. These are briefly presented in the next section. However, the joint sampling of these two functions, leading to discretization of Eq. (1), and its relation to Eq. (3), have not been studied in great detail and are now examined and discussed in Section 4.

3. Sampling of a single function on the sphere

Consider a spatially band-limited function, $f(\Omega)$, of order N_f, which can be represented in the SH domain by [31]:

$$f(\Omega) = \sum_{n=0}^{N_f}\sum_{m=-n}^{n} f_{nm}Y_n^m(\Omega), \tag{4}$$

where f_{nm} is the SFT of $f(\Omega)$ and $Y_n^m(\cdot)$ are the complex SH functions of order n and degree m. The function is sampled over the sphere at $Q_{\mathscr{S}}$ sample points denoted $\{\Omega_q\}_{q=1}^{Q_{\mathscr{S}}}$. This sampling scheme, denoted by $\mathscr{S}$, is designed to be aliasing-free up to order $N_{\mathscr{S}}$. The SH coefficients of $f(\Omega)$, f_{nm}, can be estimated from its samples using the discrete SFT [9]:

$$\hat{f}_{nm} = \sum_{q=1}^{Q_{\mathscr{S}}} \alpha_q f(\Omega_q)[Y_n^m(\Omega_q)]^*, \tag{5}$$

where $\{\alpha_q\}_{q=1}^{Q_{\mathscr{S}}}$ are the sampling weights. Substituting Eq. (4) and applying the orthogonality property of the SH functions leads to [32]

$$\begin{aligned}\hat{f}_{nm} &= \sum_{n'=0}^{N_f}\sum_{m'=-n'}^{n'} f_{n'm'}\sum_{q=1}^{Q_{\mathscr{S}}}\alpha_q[Y_n^m(\Omega_q)]^*Y_{n'}^{m'}(\Omega_q)\\ &= f_{nm} + \sum_{n'=N_{\mathscr{S}}+1}^{N_f}\sum_{m'=-n'}^{n'} f_{n'm'}\,\epsilon(n,m,n',m'),\end{aligned} \tag{6}$$

where,

$$\sum_{q=1}^{Q_{\mathscr{S}}}\alpha_q[Y_n^m(\Omega_q)]^*Y_{n'}^{m'}(\Omega_q) = \begin{cases}\delta_{n-n'}\delta_{m-m'} & n,n' \leqslant N_{\mathscr{S}}\\ \epsilon(n,m,n',m') & n' > N_{\mathscr{S}}\end{cases}. \tag{7}$$

It is clear that $\epsilon = 0$ leads to $\hat{f}_{nm} = f_{nm}$. Therefore, ϵ represents the aliasing error. To demonstrate the behavior of the aliasing error, the elements $\epsilon(n,m,n',m')$ are plotted in Fig. 1. The figure shows the error as a function of n and n' for different sampling schemes of order $N_{\mathscr{S}} = 3$ and with $N_f = 9$, in the range $0 \leqslant n \leqslant N_{\mathscr{S}}$ and $0 \leqslant n' \leqslant N_f$. As reported