



IMAGLS: INTERAURAL LEVEL DIFFERENCE WITH MAGNITUDE LEAST-SQUARES LOSS FOR OPTIMIZED FIRST-ORDER HEAD-RELATED TRANSFER FUNCTION

O. Berebi^{1*} Z. Ben-Hur² D. Lou Alon² B. Rafaely¹

¹ School of Electrical and Computer Engineering, Ben Gurion University of the Negev, Beer Sheva, 8410501, Israel

² Reality Labs Research, Meta, 1 Hacker Way, Menlo Park, 94025, USA

ABSTRACT

Binaural reproduction for headphone-based listening is an active research area due to its widespread use in evolving technologies such as augmented and virtual reality (AR and VR). On the one hand, these applications demand high quality spatial audio perception to preserve the sense of immersion. On the other hand, recording devices may only have a few microphones, leading to low-order representations such as first-order Ambisonics (FOA). However, first-order Ambisonics leads to limited externalization and spatial resolution. In this paper, a novel head-related transfer function (HRTF) preprocessing optimization loss is proposed, and is minimized using nonlinear programming. The new method, denoted iMagLS, involves the introduction of an interaural level difference (ILD) error term to the now widely used MagLS optimization loss for the lateral plane angles. Results indicate that the ILD error could be substantially reduced, while the HRTF magnitude error remains similar to that obtained with MagLS. These results could prove beneficial to the overall spatial quality of first-order Ambisonics, while other reproduction methods could also benefit from considering this modified loss.

Keywords: *binaural reproduction, ambisonics, MagLS, ILD, perceptually motivated loss*

*Corresponding author: berebio@post.bgu.ac.il.

Copyright: ©2023 O. Berebi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

In recent years, binaural sound reproduction has gained significant attention due to its ability to create an immersive and realistic listening experience. Binaural sound refers to the technique of capturing and reproducing sound in a way that mimics natural listening when using headphones. One format that has gained popularity and seen widespread adoption for binaural sound reproduction is Ambisonics [1]. Ambisonics signals are typically derived from recordings with spherical microphone arrays and filtered with measured or modeled head-related transfer function HRTF to produce binaural signals [2, 3].

First-order Ambisonics (FOA) is a commonly used format in Ambisonics due to the relatively simple microphone array required for recording (only four microphones), established theoretical and algorithmic literature, and availability and broad adoption in applications such as 360° video, surround sound, and virtual reality (VR) [2]. However, FOA has shortcomings with regard to spatial resolution and timbre degradation that limit its application for VR audio [4].

To improve the listening experience for signals processed with FOA, researchers have proposed various HRTF preprocessing methods. These methods aim to overcome some of the spatial resolution and timbre degradation issues of FOA by correcting the low-order HRTF spatial and spectral errors. Examples of these methods include global equalization, time alignment, ear alignment, and magnitude least squares (MagLS) [5–8].

While the MagLS method has been proven to be very beneficial in reducing spectral error in FOA, the resulting binaural signals still have significant spatial errors [2, 4].

The aim of this paper is to propose and investigate improvements over the current MagLS method. This is achieved by the use of the HRTF preprocessing with integrated MagLS and interaural level difference (ILD) optimized errors. This approach aims to improve FOA's spatial attributes while preserving low spectral errors, ultimately providing a more immersive and realistic audio experience for listeners.

2. PROBLEM FORMULATION

In Ambisonics, binaural signals can be rendered in the spherical harmonics (SH) domain by combining the Ambisonics signal with the left-ear and right-ear HRTFs [9]. This can be expressed as [10]:

$$p^{L/R}(f) = \sum_{n=0}^N \sum_{m=-n}^n [\tilde{a}_{nm}(f)]^* h_{nm}^{L/R}(f), \quad (1)$$

where $p^{L/R}(f)$ denotes the binaural signal for the left/right ear and f denotes the frequency. The N 'th order Ambisonics signal is denoted as $a_{nm}(f)$ and is modified to $\tilde{a}_{nm}(f) = (-1)^m [a_{n(-m)}(f)]^*$, encoding the sound-field information that can be captured using a spherical microphone array [2]. Henceforth we refer only to the left-ear for simplicity, while the same operations are conducted for the right ear. The $(N+1)^2$ left-ear HRTF SH coefficients are represented by $h_{nm}^L(f)$, which can be obtained by minimizing the following generalized problem:

$$h_{nm}^L(f) = \arg \min_{\hat{h}_{nm}^L(f)} \int_{\Omega \in \mathcal{S}^2} D \left(\sum_{n=0}^N \sum_{m=-n}^n \hat{h}_{nm}^L(f) Y_n^m(\Omega), H(f, \Omega) \right) d\Omega, \quad (2)$$

with $H(f, \Omega)$ corresponding to a reference measured or modeled HRTF, and the spherical harmonic function is represented by $Y_n^m(\Omega)$ with degree m and order n . The complex coefficients $h_{nm}^L(f)$ serve to minimize the distance function $D(\cdot, \cdot)$ across a collection of directions $\Omega = (\varphi, \theta) \in \mathcal{S}^2$, with $D(\cdot, \cdot)$ representing a measure of dissimilarity.

The choice of function $D(\cdot, \cdot)$ can greatly impact the spectral and spatial quality of $p^L(f)$. For example, using $D(x, y) = |x - y|^2$, the least squares (LS) distance yields a closed-form solution to Eqn. (2) given by the inverse spherical harmonic transform (ISHT) of $H(f, \Omega)$ [11]. However, when N is small, particularly with $N = 1$, the LS solution results in a $p^L(f)$ that is not perceptually comparable to its high-order counterpart, with high-

frequency rolloff and limited spatial resolution [2]. Another example is using $D(x, y) = ||x| - |y||^2$, the magnitude least squares (MagLS) distance. Solving Eqn. (2) for the MagLS distance for frequencies greater than a cut-off frequency f_c greatly improves the spectral quality of $p^L(f)$ compared to the LS method even for low N , as shown in [8]. Unfortunately, the spatial attributes of the MagLS solution are still poor when considering $N = 1$ reproduction [2, 4].

3. PROPOSED METHOD

The MagLS formulation has been shown to be effective in addressing rolloff issues, even with $N = 1$. However, it is important to acknowledge that the spatial resolution at $N = 1$ is still a concern [2]. To tackle this limitation, we propose a new $D(\cdot, \cdot)$ function that addresses the issue of binaural information while preserving the spectral benefits of the MagLS formulation. This function will be used in the problem formulation that we present in the following section. Note that for the rest of this paper, we omit the superscript L for abbreviation.

The ILD is known to be an important binaural cue for sound localization [12], and preserving ILD in an FOA HRTF could potentially enhance spatial perception. Consider the following order- N HRTF representation:

$$\mathcal{SFT}(h_{nm}(f)) \equiv \sum_{n=0}^N \sum_{m=-n}^n h_{nm}(f) Y_n^m(\Omega). \quad (3)$$

We begin by defining the ILD for a given HRTF as [9]:

$$\text{ILD}(\Omega_0, f_0) = 10 \log_{10} \frac{\int_{f_1}^{f_2} G(f_0, f) |p^L(\Omega_0, f)|^2 df}{\int_{f_1}^{f_2} G(f_0, f) |p^R(\Omega_0, f)|^2 df}. \quad (4)$$

Here, $G(f_0, f)$ denotes the Gammatone function centered at f_0 , and $p^{L,R}(\Omega_0, f)$ represents the left- and right-ear signals resulting from a single plane wave sound field at an incident angle Ω_0 . The ILD is evaluated over the horizontal plane $\Omega \in (\theta = 90^\circ, 0^\circ \leq \phi < 360^\circ)$ for frequencies $f_1 \leq f \leq f_2$. This computation facilitates perceptually-motivated smoothing of the ILD across frequencies. The ILD error as a function of both frequency and direction is defined as follows:

$$\epsilon_{\text{ILD}}(\Omega_0, f_0) = |\text{ILD}_{ref}(\Omega_0, f_0) - \text{ILD}(\Omega_0, f_0)|. \quad (5)$$

Here, $\text{ILD}_{ref}(\Omega_0, f_0)$ and $\text{ILD}(\Omega_0, f_0)$ refer to the ILD of the reference HRTF $H(\Omega_0, f)$ and the order N HRTF $\mathcal{SFT}(h_{nm}(f))$, respectively.

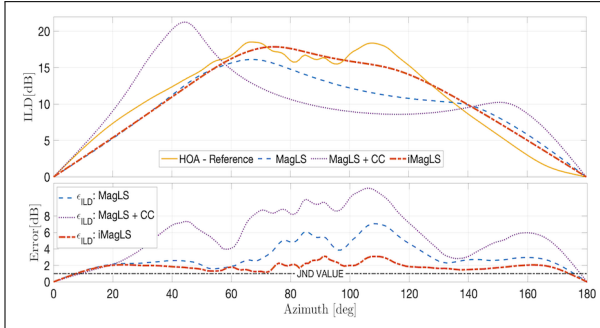


Figure 1. ILD curves (top) and ILD error curves (bottom) as a function of Ω_0 , averaged over frequency.

Next, the magnitude error between $H(\Omega, f)$ and $SFT(h_{nm}(f))$ is defined in a similar manner to the MagLS formulation:

$$\epsilon_{mag}(\Omega, f) = \left| |H(\Omega, f)| - |SFT(h_{nm}(f))| \right|^2. \quad (6)$$

Finally, combining both error terms leads to the following optimization problem:

$$h_{nm}(f) = \arg \min_{\hat{h}_{nm}(f)} \left[\sum_{\Omega} \epsilon_{mag}(\Omega) + \lambda \sum_{\Omega_0} \epsilon_{ILD}(\Omega_0) \right], \quad (7)$$

with $\epsilon_{mag}(\Omega)$ and $\epsilon_{ILD}(\Omega_0)$ both referring to error frequency averaging, and $\lambda \in \mathcal{R}$ is used as a regularization term. The formulation in Eqn. (7) is referred to as ILD-MagLS (iMagLS for short).

4. NUMERICAL EVALUATION

The Broyden–Fletcher–Goldfarb–Shannon Quasi-Newton algorithm [13] was used to minimize Eqn. (7) over $1.2 \leq f < 20$ kHz. Note that $\epsilon_{mag}(\Omega)$ and $\epsilon_{ILD}(\Omega_0)$ are both averaged over frequencies. Therefore, the numerical solver minimizes the average error over all frequencies, rather than individually for each frequency. It was observed that this leads to smoother error curves, which may be more perceptually favorable in terms of noticeable spectral artifacts and noise.

The simulated KEMAR HRTF [14] was used to evaluate the proposed method, with order $N = 35$ as the reference and $N = 1$ used to evaluate the MagLS and iMagLS solutions. The evaluation used a Lebedev sampling scheme of order 35 with 1730 nearly-uniformly-distributed directions to evaluate $\epsilon_{mag}(\Omega)$. As suggested

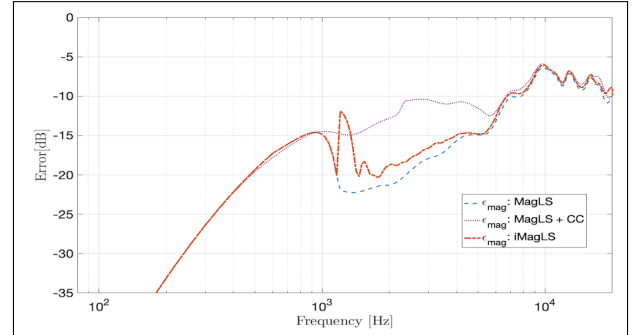


Figure 2. Magnitude error as a function of frequency, averaged over Ω directions.

in [2], the MagLS solution presented in this analysis also includes the covariance constraint (MagLS+CC) global EQ variant [15]. The MagLS solution served as the initial solution for the solver, and λ was chosen such that the error terms were equal at the first iteration. The results were evaluated in terms of the magnitude error $\epsilon_{mag}(f)$ averaged over all Ω directions and in terms of the ILD error $\epsilon_{ILD}(\Omega_0)$ averaged over $1.2 \leq f < 20$ kHz.

The ILD evaluation is presented in Fig. 1. The ILD of Eqn. (4) was averaged over frequency; the yellow solid curve represents the KEMAR reference, the dashed blue curve represents the ILD of an $N = 1$ MagLS solution, the purple dotted curve represents the ILD of an $N = 1$ MagLS+CC solution, and the red dash-dotted curve represents the proposed iMagLS of a similar order. The graph below represents the ILD error of Eqn. (5) averaged over frequency for MagLS (dash blue), MagLS+CC (dotted purple) and iMagLS (dash-dotted red). As shown in Fig. 1, the improvement in terms of ILD is noticeable both in terms of averaged ILD and averaged ILD error. An ILD error of below 2 dB was preserved by iMagLS for most incident angles, which is close to the Just Noticeable Differences (JND) (~ 1 dB [16, 17]), while the MagLS and MagLS+CC error was significantly higher than the JND value for most angles.

The magnitude error of Eqn. (6) averaged over the Ω directions is presented in Fig. 2. The dash blue, dotted purple, and dash-dotted red curves represent MagLS, MagLS+CC, iMagLS error, respectively. Although the iMagLS error was higher for all $f > 1200$ Hz, compared to the optimal MagLS solution, it was only slightly higher with an averaged absolute difference of 1.72 dB over these frequencies.

The results suggest that improving the ILD error while maintaining a relatively low binaural magnitude error is possible. We argue that when searching for a low-order h_{nm} , one should consider the relation between both ears in addition to the magnitude accuracy of each ear individually. While the rationale of the MagLS+CC variant align with this argument, results show that iMagLS outperformed MagLS+CC both in terms of ILD and magnitude error.

5. CONCLUSIONS

This paper proposed a novel optimization method, iMagLS, to address the limitations of FOA by improving its spatial information shortcomings. The method optimizes HRTF preprocessing with an ILD error term incorporated into the widely used MagLS method. The results show that the proposed method can significantly reduce ILD errors, while maintaining similar HRTF magnitude errors compared to the MagLS method. Based on these results, we propose that selecting a low-order h_{nm} should consider not only the individual magnitude accuracy of each ear but also the interaural relationship between them. Additionally, a more comprehensive study involving a broader range of HRTFs and listening tests is suggested for future work to validate these claims.

6. REFERENCES

- [1] M. A. Gerzon, "Periphery: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.
- [2] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.
- [3] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *The Journal of the Acoustical Society of America (JASA)*, vol. 133, no. 5, pp. 2711–2721, 2013.
- [4] B. Rafaely, V. Tourbabin, E. Habets, Z. Ben-Hur, H. Lee, H. Gamper, L. Arbel, L. Birnie, T. Abhayapala, and P. Samarasinghe, "Spatial audio signal processing for binaural reproduction of recorded acoustic scenes—review and challenges," *Acta Acustica*, vol. 6, p. 47, 2022.
- [5] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *JASA*, vol. 141, no. 6, pp. 4087–4096, 2017.
- [6] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *JASA*, vol. 143, no. 6, pp. 3616–3627, 2018.
- [7] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Efficient representation and sparse sampling of head-related transfer functions using phase-correction based on ear alignment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2249–2262, 2019.
- [8] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proceedings of the DAGA*, vol. 44, pp. 339–342, 2018.
- [9] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.
- [10] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *JASA*, vol. 127, no. 2, pp. 823–828, 2010.
- [11] B. Rafaely, *Fundamentals of spherical array processing*, vol. 8. Springer, 2015.
- [12] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.
- [13] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. general considerations," *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.
- [14] M. Burkhard and R. Sachs, "Anthropometric manikin for acoustic research," *JASA*, vol. 58, no. 1, pp. 214–222, 1975.
- [15] J. Vilkamo, T. Bäckström, and A. Kuntz, "Optimized covariance domain framework for time–frequency processing of spatial audio," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 403–411, 2013.
- [16] A. W. Mills, "Lateralization of high-frequency tones," *JASA*, vol. 32, no. 1, pp. 132–134, 1960.
- [17] W. A. Yost and R. H. Dye Jr, "Discrimination of interaural differences of level as a function of frequency," *JASA*, vol. 83, no. 5, pp. 1846–1851, 1988.