Audio Engineering Society

Conference Paper

Presented at the AES 2026 International Conference on
Audio for Virtual and Augmented Reality and Immersive Games
2026 June 30 – July 3, Sorbonne University/d'Alembert & IRCAM/STMS, Paris, France

The Influence of Binauralizer and HRTF Preprocessing on Objective Loudness in Ambisonics

Nils Peters[1]

[1]*Trinity College, The University of Dublin, School of Engineering, Dublin, Ireland*

Correspondence should be addressed to Nils Peters (nils.peters@tcd.ie)

ABSTRACT

Accurate loudness estimation is essential for audio production, quality control, and loudness compliance, but no established recommendation exists for binaural playback over headphones. This paper investigates the influence of binauralizers and HRTF preprocessing on objective loudness estimation for binauralized Ambisonics content. Two experiments were conducted using 163 Ambisonics clips binauralized with two open-source renderers and three HRTF sets under three HRTF preprocessing conditions. Objective loudness metrics are compared against ground-truth loudness data derived from 7.1.4 loudspeaker feeds according to ITU-R BS.1770. Results reveal small to moderate differences in Integrated Loudness and larger differences in the True Peak values between the evaluated binauralizers, and that diffuse-field equalization can effectively eliminate loudness and True Peak differences across binauralizers and HRTFs. The findings can help to better predict and ensure loudness compliance in binauralized audio consumption in XR and gaming, especially when importing 3rd-party HRTFs is supported.

1 Introduction

Loudness is an essential aspect of sound perception. It pertains to the volume level perceived by the listener and does not linearly correspond to the signal amplitude. In streaming and broadcast systems, maintaining a consistent long-term loudness level is both a practical goal and a regulatory requirement ([1, 2]). Traditionally, program loudness is measured and estimated algorithmically prior to distribution, with program loudness values signaled as metadata. During reproduction, this metadata enables the receiver to normalize the loudness of the program to a consistent reference level.

The ITU-R BS.1770 recommendation [3] defines algorithms for objectively estimating the perceived loudness of audio program material. As depicted in Fig. 1, the audio signal of each channel passes through a K-weighting filter, which applies a high-shelf boost to account for the acoustic effects of the human head, followed by a high-pass filter to reduce the impact of low frequencies. Next, the mean square energy of each filtered channel is calculated over short, overlapping time blocks of 400 ms. These energy values are then scaled by G based on the channel-dependent loudspeaker direction, summed together, and converted to a logarithmic value. Finally, a two-stage gating process is applied in which blocks are discarded for with the energy is below -70 LUFS or below a relative threshold (-10 LU below the ungated average loudness). The remaining blocks are averaged to obtain the integrated loudness value.

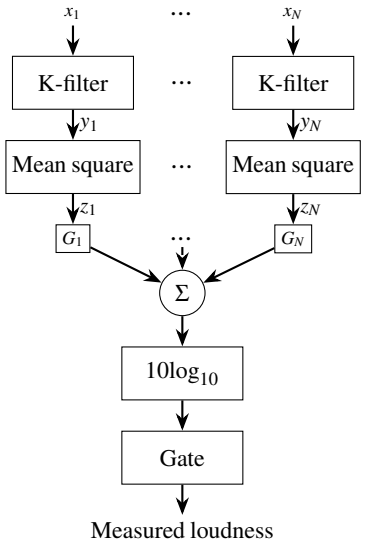

Fig. 1: Loudness measurement processing of N-channel program material, adapted from [3]

Initially developed for stereo and 5.1 surround formats, it was later extended to support additional channel-based configurations up to 22.2 [4]. With the growing diversity of reproduction scenarios and the advent of object-based and scene-based audio, predicting the program loudness from the audio material before content delivery becomes challenging. This is because both object-based and scene-based audio render the programme according to the individual downstream reproduction setup, which may be unknown at the time of content delivery. While ITU-R BS.1770 accommodates immersive channel-based material (e.g. 7.1.4) reproduced over loudspeakers, research in estimating loudness for object- and scene-based audio largely relies on rendering the program material to different loudspeaker configurations, from which objective loudness can be calculated and then compared with results from subjective loudness assessment tests [5, 6, 7].

For object-based audio, the authors of [6] found that the loudspeaker configuration contributes more than the actual object-based rendering algorithm to the differences between objective loudness and subjective loudness. The study [5] found the loudness of object-based audio rendered to loudspeakers is reasonably well correlated with ITU-R BS.1770, but that some test items were measured significantly louder than they were perceived. In [7] the loudness of Scene-based audio material (i.e. Higher-Order Ambisonics) rendered to different loudspeaker configurations was objectively and subjectively tested. Also here, the authors found that the loudspeaker configuration affects the objective loudness. Furthermore, the subjectively assessed loudness was not always well estimated by ITU-R BS.1770.

Loudness estimation for (binauralized) headphone playback is currently not addressed in ITU-R BS.1770, and research on this topic is limited. In [8], the perceived loudness stability of Head-Related Transfer Functions (HRTFs) in the spherical harmonics (SH) domain was studied as a function of aliasing and SH truncation errors. Results suggest a substantial effect of the truncation error, while aliasing errors can improve the loudness stability in certain cases.

1.1 Research Questions

As loudness has predominantly been investigated in the context of loudspeaker reproduction, its applicability to headphone-based playback, common in gaming and XR environments, remains insufficiently understood. It is uncertain how the binauralization process changes the loudness level of the binauralized program material, e.g., an Ambisonics sound scene reproduced over headphones rather than loudspeakers. Thus, this study focuses on binauralized Ambisonics and investigates the influence of different binauralization methods and head-related transfer functions (HRTFs) on the objective loudness metrics as specified in ITU-R BS.1770. The main research questions are:

RQ1: To what extent do state-of-the-art Ambisonics binauralizers preserve objective loudness relative to loudspeaker-based measurements as defined in ITU-R BS.1770?

RQ2: What is the impact of varying HRTFs and their respective preprocessing methods on the objective loudness?

To answer these questions, this study presents two experiments. Experiment A addresses RQ1 and is described in Section 2. Section 3 describes Experiment B to address RQ2. All findings will be further discussed in Section 4.

2 Experiment A: Loudness Differences across Binauralizer

2.1 Methodology

For this experiment, two different binauralizers in two different binauralization modes are compared. 163 Ambisonics test items are binauralized and various loudness metrics are computed for each condition. These loudness values are then statistically evaluated with ground-truth loudness metrics that were derived from loudspeaker feeds generated with the EBU EAR[1] renderer. The 7.1.4 configuration, being the de facto standard for immersive audio production [9], serves as the ground-truth setup for which valid loudness metrics can be calculated according to ITU-R BS.1770 [3].

2.1.1 Binauralizers

The two evaluated binarualizers are Google's OBR[2] [10] and EBU's BEAR[3] [11]. These two renderers were selected for their ability to handle both up to 4$^{\text{th}}$-order HOA scene-based and 7.1.4 channel-based audio material. In addition, both binauralizers are open source, which facilitates reproducibility of the results. The binauralizers were used in their default configurations, including their respective HRIRs.

Figure 2 illustrates the process of generating the test signals used to evaluate different Ambisonics binauralization methods, as previously described e.g., in [12]. Primarily, the "direct" binauralization processing of the two renderers is labeled *OBR HOA* and *BEAR HOA*, respectively. Secondly, pre-rendered 7.1.4 ground-truth signals are binauralized using the channel-based binauralization modes of OBR and BEAR, resulting in the conditions *OBR pre-rendered* and *BEAR pre-rendered*, respectively. All processing is performed at a sampling rate of 48 kHz.

2.1.2 Dataset

To draw from a wide variety of well-produced Ambisonics content, the IEM HOAST Ambisonics library[4] was utilized [13]. This library consists of a wide variety of VR content (i.e., 360° Video and Ambisonics Audio), such as recorded music performances, radio plays, soundscapes, or studio-produced contemporary music, ranging from 1$^{\text{st}}$-order to 4$^{\text{th}}$-order Ambisonics at 48 kHz, all compressed with Opus at about 60 kbps per channel which suffices good quality according to [14]. From each audio asset we extracted two minutes from the middle as test material. To increase the number of low-order test items, higher-order Ambisonics content is order-truncated, resulting in a total of 163 tested audio scenes (bottom row in Table 1).

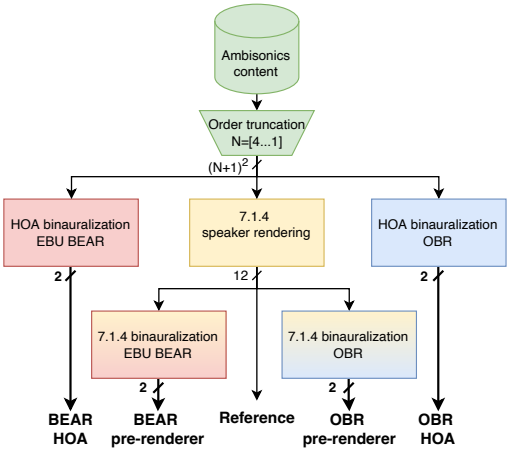

Fig. 2: Generation of rendering conditions

Table 1: Overview of the HOAST content

	HOA Order				
	1	2	3	4	Total
Number of files	3	11	30	12	56
Derived test items	56	53	42	12	163

2.1.3 Objective Loudness Metrics

The following three loudness metrics are computed for all test items and conditions as specified in ITU-R BS.1770 [3] and EBU R128 [15]. Since no dedicated loudness measure for binaural content exists, the binaural test items are treated as conventional two-channel loudspeaker channels, i.e., no directional weighting (i.e. $G = 0$ dB in Fig. 1). For the channel-based 7.1.4 ground-truth, a weighting of $G = 1.5$ dB is applied to the two lateral loudspeakers at $\pm 90°$ as defined in [3].

[1] https://github.com/ebu/ebu_adm_renderer
[2] https://github.com/google/obr
[3] https://github.com/ebu/bear
[4] https://hoast.iem.at

Integrated Loudness is a single overall loudness measure of the audio programme, intended to represent the programme's long-term perceived level. It is computed from K-weighted spectral and G-weighted directional audio energy over the programme duration, with gating applied to exclude very quiet segments from the final average (see Fig. 1). The metric is reported in loudness units relative to full scale (LUFS).

Loudness Range (LRA) is expressed in LU (Loudness Units), quantifies the variability of loudness within a programme, and is based on the statistical distribution of measured loudness as defined in [15]. It is measured in overlapping 3-second time frames.

True Peak (TP) is a technical estimate of the highest signal level that will occur after digital-to-analog reconstruction or sample-rate conversion, including inter-sample peaks that fall between stored samples. A TP above 0 dBFS should be avoided because it may cause clipping in subsequent processes.

2.1.4 Statistical Evaluation

The average differences between the loudness metrics measured with the Binauralizer under Test (L^{SuT}), and the loudness metric estimated from the ground-truth loudspeaker signals (L^{S}) are quantified using the Root Mean Square Error (RMSE, see eq. 1) across all $M = 163$ test items.

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(L_m^S - L_m^{SuT})^2} \tag{1}$$

To examine the worst-case differences, the Maximum Absolute Error (MAE) is given by:

$$\mathrm{MAE} = \max_i(|L_i^S - L_i^{SuT}|) \tag{2}$$

To assess the overall consistency between L^{SuT} and the ground truth L^S, the Pearson correlation coefficient r is computed per eq. 3. Here, $\bar{L}^S$ and $\bar{L}^{SuT}$ are their mean parameter values.

$$r = \frac{\sum_{m=1}^{M}((L_m^S - \bar{L}^S)(L_m^{SuT} - \bar{L}^{SuT}))}{\sqrt{\sum_{m=1}^{M}(L_m^S - \bar{L}^S)^2 \sum_{m=1}^{M}(L_m^{SuT} - \bar{L}^{SuT})^2}} \tag{3}$$

2.2 Results

2.2.1 Integrated Loudness

Figure 3 shows the Integrated Loudness of all four Systems under Test with respect to the reference loudness that was measured from the 7.1.4 loudspeaker feeds. Table 2 shows the RMS error with respect to the reference loudness for individual Ambisonics orders and the RMSE and MAE averaged over all data.

For all four tested systems, there is a very high correlation of loudness estimates between binauralized and loudspeaker-based presentation. Overall the OBR HOA condition achieved the lowest RMSE of 1.92 LU across all test items. The BEAR HOA condition has an RMS error of 3.99 LUFS and has overall a slightly higher loudness estimate, as visible in Fig. 3. For both renderers, the direct Ambisonics binauralization leads to a lower RMS error than when the content is binauralized from the pre-rendered 7.1.4 loudspeaker feeds. The highest MAE of 9.20 LU was found for the OBR pre-rendered condition, whereas the lowest MAE was achieved by the OBR-HOA binauralizer (5.31 dB).

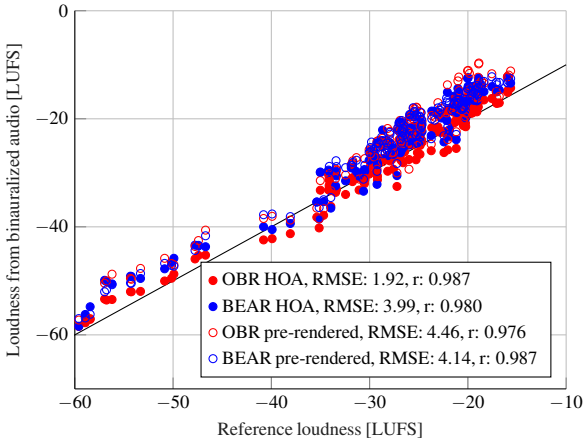

Fig. 3: Loudness [LUFS] pooled across HOA orders

Table 2: Error metrics for Integrated Loudness

	RMSE per HOA Order				Overall	
	1	2	3	4	RMSE	MAE
OBR - HOA	**1.55**	**2.09**	**2.22**	**1.45**	**1.92**	**5.31**
OBR - pre-rendered	5.01	4.37	3.74	4.51	4.46	9.20
BEAR - HOA	3.41	4.37	4.07	4.40	3.99	7.59
BEAR - pre-rendered	4.62	4.04	3.55	4.24	4.14	7.65

2.2.2 Loudness Range

Figure 4 visualizes the Loudness Range results. All rendering conditions show good alignment with the Loudness Range of the reference condition with RMS errors around 1 dB (see Tab. 3). The largest MAE is 5.42 LUs for the BEAR HOA condition.

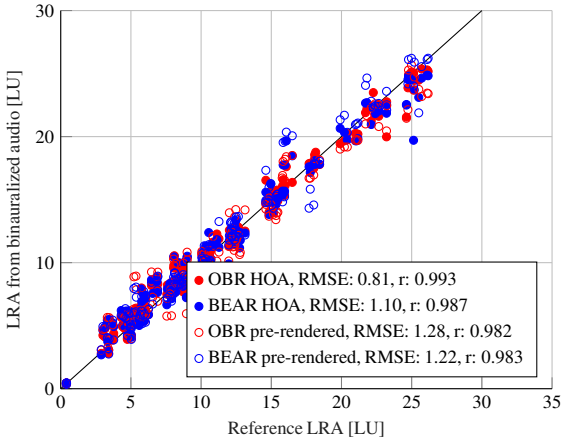

Fig. 4: Loudness Range pooled across HOA orders

Table 3: Error metrics for Loudness Range in LU

	RMSE per HOA Order				Overall	
	1	2	3	4	RMSE	MAE
OBR - HOA	**0.84**	**0.79**	**0.83**	**0.72**	**0.81**	**3.23**
OBR - pre-rendered	1.25	1.23	1.33	1.47	1.28	3.67
BEAR - HOA	1.22	1.23	1.33	1.47	1.10	5.43
BEAR - pre-rendered	1.18	1.18	1.24	1.49	1.22	4.29

2.2.3 True Peak

For all Systems Under Test, the True Peak value is substantially higher (up to 12.5 dB MAE found for OBR pre-rendered) than the True Peak value from the reference 7.1.4 loudspeaker feeds (see Tab. 4 and Fig. 5). This is unexpected because we did not find an equally strong increase in loudness that may explain the increase in TP.
Fig. 5 depicts that the OBR effectively limits the TP value due to its built-in end-of-chain limiter. The BEAR binaurlizer, which does not feature a limiter, produces many more True Peak values above 0 dB.

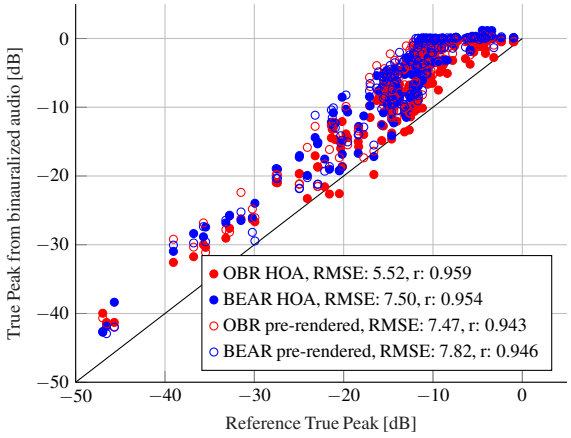

Fig. 5: True Peak (TP) pooled across HOA orders

Table 4: Error metrics for True Peak in dB

	RMSE per HOA Order				Overall	
	1	2	3	4	RMSE	MAE
OBR - HOA	**6.23**	**5.70**	**4.39**	**4.59**	**5.52**	**9.47**
OBR - pre-rendered	8.77	7.27	5.72	7.12	7.47	12.55
BEAR - HOA	7.84	7.77	6.62	7.48	7.50	11.64
BEAR - pre-rendered	8.91	7.73	6.33	7.40	7.82	12.03

3 Experiment B: Effect of HRTF preprocessing

The second experiment investigates the impact of the deployed HRTFs and different HRTF preprocessing methods on the Integrated Loudness.

3.1 Methodology

Three sets of HRTFs from different sources were chosen and converted into the spherical harmonic (SH) domain and all test items used in Experiment 1 were directly binauralized in the SH domain.

The HRTFs are processed with three different HRTF preprocessing methods as described below, leading to 3 preprocessing methods x 3 HRTFs = 9 conditions.

3.1.1 HRTF Sets

Two sets of KEMAR dummy head HRTFs, originated from different HRTF providers (named KEMAR A and KEMAR B) and a set of KU100 HRTFs, are used.

The HRTFs were selected to explore loudness differences between the same type of head model originating

from different datasets (i.e. KEMAR A vs. KEMAR B) and differences across heads (i.e. KEMAR A/B vs. KU100). The KEMAR sets consisted of 828 and 836 measurement directions, respectively, whereas the KU100 was measured at 8802 directions. All HRTFs have a sampling rate of 48 kHz.

3.1.2 HRTF preprocessing

1. Least-squares (LS) conversion (Baseline): the LS method is the simplest method to convert HRTFs into the SH domain. As shown in eq. 4, the conversion is achieved by the Spherical Harmonic Transform of $\mathbf{h}_M$ up to a given order N, by deploying the pseudoinverse $(\cdot)^\dagger$ of $\mathbf{Y}_M$ where M is a dense set of directions around the listener for which the set of HRTFs $\mathbf{h}_M$ is defined.

$$\min_w ||\mathbf{Y}_M\mathbf{w}-\mathbf{h}_M||_2^2, \text{ with } \mathbf{w}_{LS}=\mathbf{Y}_M^\dagger\mathbf{h}_M=\mathcal{SHT}_N^M(\mathbf{h}_M) \tag{4}$$

2. MagLS conversion: exploits that humans are insensitive to interaural time differences (ITDs) at high frequencies, therefore only fitting the HRTFs' magnitude values at higher frequencies to spherical harmonics. This results in reduced interaural level difference (ILDs) errors as proposed in [16].

3. Diffuse-field equalization prior MagLS: The Diffuse-field equalization preprocessing removes the direction-independent components within a dense set of M HRTFs, measured for the source directions θ,φ per eq. 5, resulting in the so-called Directional Transfer Functions (DTF) which are then converted into the SH domain via the MagLS approach.

$$\mathrm{DTF}(\theta,\varphi,f)=\frac{H(\theta,\varphi,f)}{\sqrt{\frac{1}{M}\sum_{i=1}^{M}|H(\theta_i,\varphi_i,f)|^2}} \tag{5}$$

3.2 Results

The top part of Table 5 summarizes the RMSE and MAE values for the nine HRTF conditions w.r.t the 7.1.4 ground-truth reference. For the LS Baseline condition, the Integrated Loudness varies substantially across the three tested HRTF sets, with RMSE values ranging from 2.61 to 10.93 LU. The corresponding MAE values relative to the 7.1.4 reference are similarly high, ranging between 8.83 and 15.93 LU, and indicates notable differences in perceived loudness. Both KEMAR HRTFs show larger RMSE and MAE values overall, suggesting greater loudness deviations compared to the KU100. The MagLS method slightly degrades both error metrics. In contrast, applying diffuse-field equalization remarkably improves performance for the KEMAR HRTFs, thus reducing the error metrics to approximately 3 LU, while slightly increasing the KU100's RMSE by 0.65 LU. In all conditions, the MAE decreases substantially and remains below 7 LU after equalization. Although not shown here, also the error metrics for the True Peak values decrease after diffuse-field equalization from more than 16 dB MAE to about 10.5 dB and 5.0 dB with respect to the 7.1.4 reference and OBR-HOA, respectively.

Table 5: RMSE and MAE of Integrated Loudness in comparison to 7.1.4 reference and OBR-HOA

Method	Kemar B		Kemar A		KU100	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Comparison to 7.1.4 Reference Loudness						
LS Baseline	10.93	15.93	7.89	14.33	**2.61**	8.83
+ MagLS	11.84	18.36	8.75	16.67	3.35	9.87
+ DTF	**2.87**	**6.16**	**3.31**	**6.68**	3.26	**6.51**
Comparison to OBR-HOA Loudness						
LS Baseline	11.59	22.76	8.41	10.73	1.93	6.76
+ MagLS	12.41	23.02	9.15	13.30	2.89	7.95
+ DTF	**1.35**	**5.40**	**1.72**	**4.55**	**1.64**	**4.10**

Of particular interest is the performance of the different HRTF sets relative to the OBR-HOA binauralizer, which yielded the best results in Experiment A. The corresponding RMSE and MAE metrics are reported in the lower part of Tab. 5. Applying diffuse-field equalization substantially decreases both RMSE and MAE across all three HRTF sets, with RMSE values falling below 1 LU and an MAE of 2.15 LU. These results indicate that diffuse-field equalization effectively helps align perceived loudness across binauralizers and HRTF sets. The comparison of the loudness estimates of the three different HRTFs with each other is presented in Tab. 6 and in Fig. 6. Especially Fig. 6 clearly shows that diffuse-field equalization helps align the loudness estimates across sets of HRTFs. RMSEs between different sets decrease from the baseline conditions with up to 10 LUs to less than 1 LU. Also, the MAEs decrease from up to 15.51 LU to as little as 0.83 LU. Additionally, Tab. 7 shows the differences of True Peak values across HRTF sets and how the diffuse-field equalization helps align the True Peak values to about 1 dB RMSE.

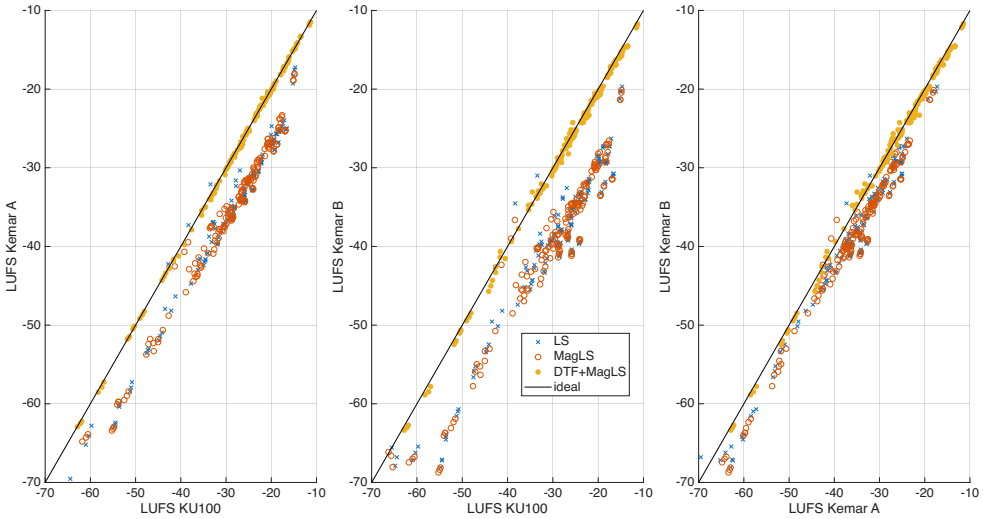

Fig. 6: Differences of Integrated Loudness between pairs of HRTF sets as a function of preprocessing methods.

Table 6: Integrated Loudness differences between pairs of HRTFs with different preprocessing methods

	Kemar A vs. Kemar B		Kemar B vs. KU100		KU100 vs. Kemar B	
Method	RMSE	MAE	RMSE	MAE	RMSE	MAE
LS Baseline	3.44	6.81	6.79	8.81	10.05	15.51
+ MagLS	3.58	6.79	6.63	8.71	10.05	15.27
+ DTF	**0.62**	**2.15**	**0.28**	**0.83**	**0.62**	**1.67**

Table 7: True Peak differences between pairs of HRTFs with different preprocessing methods

	Kemar A vs. Kemar B		Kemar B vs. KU100		KU100 vs. Kemar B	
Method	RMSE	MAE	RMSE	MAE	RMSE	MAE
LS Baseline	2.87	6.92	6.41	9.43	8.86	15.31
+ MagLS	2.79	6.05	6.18	9.15	8.69	14.49
+ DTF	**0.83**	**3.58**	**0.90**	**3.10**	**1.03**	**4.09**

4 Discussion

The two presented experiments provide insight into how binauralization process affects the objective loudness. This can have implications on regulatory compliance and quality control. Experiment A shows the presence of loudness differences between Ambisonics binauralizers of about 2.5 LU RMSE, but with up to 4.51 LU RMSE and 9.2 LU MAE compared to the ground truth loudspeaker feeds. For comparison, the CALM Act allows deviation from the specified target loudness generally up to 2 dB which is the user's comfort zone [1]. Figure 3 suggests that most of the loudness is overestimated compared to the reference loudness, resulting in a quieter level than intended. The direct binauralization method seems to result in slightly more aligned loudness values compared to binauralizing the pre-rendered loudspeaker feeds. The loudness range seems well preserved in all tested binauralization condition, as desired. On the other hand, the True Peak increased significantly up to 12.5 dB. This may be related to the peaky shape of the HRIRs. End-of-chain limiter may be required to prevent clipping in D/A conversion or other subsequent processing blocks. It is unclear whether the quality of the binaural rendering can be affected by increased True Peak values.

Experiment B further explores how state-of-the-art HRTF processing methods may affect loudness. The results strongly suggest that diffuse-field equalization does not only slightly improve localization accuracy in Ambisonics [17] or reduce (spectral) differences between HRTF corpora [18, 19], but also has the positive effect of: a) eliminating loudness differences and True Peak values of binauralized content created with different HRTFs; and b) better aligns objective loudness between loudspeaker feeds and binauralized signals.

Consequently, this study suggests that diffuse-field equalization is an essential process to enable consistent

loudness of binauralized audio, especially beneficial for binauralizers that can import 3rd-party HRTFs.

4.1 Limitations and future work

Although the rendering configuration used to generate the reference condition in Experiment A was a conscious choice, using a different reference configuration may arguably lead to somewhat different objective values. An option to circumvent this uncertainty would be to signal not only the programme loudness as metadata in the audio bitstream, but also auxiliary information about the renderer configuration used for production. However, the relative differences between the investigated rendering methods would remain similar.

Future work could extend this analysis by conducting formal subjective loudness comparison tests where subjects have to switch between listening to the reference signal via loudspeakers and binauralized audio over headphones. Furthermore this study was limited to HRTFs and did not include Binaural Room Impulse Responses (BRIRs) which may be utilized instead of HRTFs to enhance externalization. Applying diffuse-field equalization to BRIRs may not be desired since it can alter the room-specific characteristics.

5 Summary and Conclusion

This study examined how binauralization affects objective loudness estimation for Ambisonics content under ITU-R BS.1770. The results show that Integrated Loudness is affected moderately by the choice of the binauralizer, whereas True Peak values can differ substantially from the 7.1.4 reference. The experiments also showed that HRTF choice and preprocessing have a noticeable impact on loudness estimation. In particular, HRTF diffuse-field equalization consistently reduced True Peak and loudness differences across HRTF sets and improved agreement with the reference, indicating that it is an effective preprocessing step for binauralizers, especially those that support personalization via 3rd-party HRTF import. Overall, these findings indicate that loudness compliance for binauralized audio cannot be assumed from loudspeaker-based rendering. More reliable binaural loudness estimation will likely require both binauralizer-aware evaluation and careful HRTF preprocessing.

References

[1] ATSC, "A/85 - Techniques for establishing and maintaining audio loudness for digital television," 2013.

[2] EBU, "Tech R128v5 - Loudness Normalisation And Permitted Maximum Level Of Audio Signals," 2023.

[3] ITU-R, "BS.1770-5 Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," 2023.

[4] Silzle, A., Sporer, T., Liebetrau, J., and Ode, S., "Progress in Standardization of 3D-Audio Loudness in ITU-R BS. 1770," in *Proc. of the Intl. Conference on Spatial Audio (ICSA)*, 2015.

[5] Norcross, S., Nanda, S., and Cohen, Z., "ITU-R BS.1770 Based Loudness for Immersive Audio," in *Proc. of the 140th AES Convention*, 2016.

[6] Iwasaki, T., Kubo, H., and Oode, S., "Loudness of Next-Generation Audio contents depending on rendering conditions," in *Proc. of the 154th AES Convention*, 2023.

[7] Peters, N. and Epain, N., "Loudness Perception of Scene-Based Audio across Loudspeaker Configurations and HOA Orders," in *Proc. of the 154th AES Convention*, 2023.

[8] Ben-Hur, Z., Alon, D. L., Rafaely, B., and Mehra, R., "Loudness Stability of Binaural Sound with Spherical Harmonic Representation of Sparse Head-Related Transfer Functions," *EURASIP Journal on Audio, Speech, and Music Processing*, 2019, doi:10.1186/s13636-019-0148-x.

[9] Agrawal, S., Bech, S., De Moor, K., and Forchhammer, S., "Influence of changes in audio spatialization on immersion in audiovisual experiences," *Journal of the Audio Engineering Society*, 70(10), pp. 810–823, 2022.

[10] Rudzki, T., Kearney, G., and Skoglund, J., "On the Design of the Binaural Rendering Library for Eclipsa Audio Immersive Audio Container," in *Proc. of the 158th AES Convention*, 2025.

[11] EBU, "Tech 3369 - Binaural EBU ADM Renderer (BEAR) for Object-Based Sound over Headphones," 2023.

[12] Shivappa, S., Morrell, M., Sen, D., Peters, N., and Salehin, S., "Efficient, compelling, and immersive vr audio experience using scene based audio/higher order ambisonics," in *AES International Conference on Audio for Virtual and Augmented Reality*, 2016.

[13] Deppisch, T., Meyer-Kahlen, N., Hofer, B., Latka, T., and Zernicki, T., "HOAST: A higher-order ambisonics streaming platform," in *Proc. of the 148th AES Convention*, 2020, https://hoast.iem.at.

[14] Narbutt, M., O'Leary, S., Allen, A., Skoglund, J., and Hines, A., "Streaming VR for immersion: Quality aspects of compressed spatial audio," in *23rd International Conference on Virtual System & Multimedia (VSMM)*, IEEE, 2017.

[15] EBU, "Tech 3342 - Loudness Range: A Measure to Supplement Loudness Normalisation in Accordance with EBU R128," 2023.

[16] Schörkhuber, C., Zaunschirm, M., and Höldrich, R., "Binaural rendering of ambisonic signals via magnitude least squares," in *Proc. of DAGA 44*, 2018.

[17] McKenzie, T., Murphy, D. T., and Kearney, G., "Diffuse-field equalisation of binaural ambisonic rendering," *Applied Sciences*, 8(10), 2018.

[18] Wen, Y., Zhang, Y., and Duan, Z., "Mitigating Cross-Database Differences for Learning Unified HRTF Representation," in *2023 Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, IEEE, 2023.

[19] Bahu, H., Jot, J.-M., Carpentier, T., Noisternig, M., Mihocic, M., Majdak, P., and Warusfel, O., "Towards improved consistency between databases of head-related transfer functions," *Journal of the AES*, 74(1/2), 2026.