Audio Engineering Society

Conference Paper 25

Presented at the AES International Conference on Headphone Technology
2025 August 27–29, Espoo, Finland

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A subjective evaluation methodology for binaural enhancement

Jean-Marc Jot[1], Garrett Treanor[1,2], Shaunak Ranade[1,2], Zhinuo Li[2], Daphna Harel[2], and Agnieszka Roginska[2]

[1]*Virtuel Works LLC*
[2]*New York University*

Correspondence should be addressed to Jean-Marc Jot (jmmjot@gmail.com)

ABSTRACT

Binaural reproduction techniques are essential for personal spatial audio reproduction over headphones, earbuds, head-mounted displays, or loudspeaker arrays. In addition to lateral localization and perceived scene width, the success metrics of binaural reproduction methods include: elevation fidelity, front/back distinction (unambiguous localization of sounds presented in front or back), externalization (the impression that the reproduced sound originates from outside of the listener's head), and timbre reproduction (minimizing artificial room tone or spectral coloration). In this study, we propose a methodology for the subjective evaluation and performance comparison of binaural reproduction enhancement methods. It aims to consistently elicit these subjective attributes in view of the wide variation of audio source types, playback conditions, and signal processing techniques that differentiate binaural reproduction circumstances in consumer applications. We report on some preliminary results of an experimental survey utilizing this methodology, where three binaural rendering methods are compared in a single-source spatialization scenario and a controlled headphone listening environment, without artificial reverberation, listener-dependent personalization, head movement compensation, or accompanying visual cues.

1 Introduction

Binaural reproduction is broadly defined as audio reproduction where the left and right audio signals are transmitted individually to each ear of each listener [1]. A common design objective is to deliver a spatial audio experience as plausible as everyday natural hearing or loudspeaker reproduction, accommodating a variety of experience conditions, including:

- *Variable device types*: e.g., headphones (open or closed), earbuds (obstructing, semi open, fully open), virtual or augmented reality (VR/AR) headsets or glasses, fixed loudspeaker soundbars, headrests, or wearable loudspeaker neckbands
- *Multiple application scenarios*: e.g., music listening (audio only), media watching (audio and video), content production, virtual conferencing, situational awareness, or interactive experiences

- *Audio processing/presentation technology*: e.g., head movement tracking, listener-dependent personalization, presentation of visual cues, or artificial reverberation (including the option of simulating local listening environment acoustics [2]).

Spatial and/or timbral audio reproduction unpredictabilities can limit the confidence of linear immersive audio content creators or interactive experience developers to approach the soundtrack authoring task with sound design choices agnostic to the listener's playback setup – even though that capability is, in theory, an advantage of modern object-based or scene-based immersive audio technology and distribution formats [3, 4]. As reported below, binaural reproduction using headphones or earbuds is often not as convincing as natural listening, or as conventional stereo or multichannel loudspeaker-based reproduction. In frontal loudspeaker soundbar systems that use crosstalk cancellation techniques for immersive audio presentation, the spatial reproduction of rear or lateral sounds may be less robust than for frontal sounds.

The reproduction of binaural audio through headphones presents several perceptual challenges that continue to limit the fidelity of spatial audio experiences. Among the most significant issues is the lack of externalization, where sound sources are perceived as originating from within the listener's head rather than from their intended locations in the external environment. While the left or right localization of a sound is reproduced predictably over headphones, placement in front or back is not consistently perceived unambiguously by listeners. Headphone-based binaural playback may exhibit a characteristic effect of boosted elevation in the frontal region, where sounds positioned directly in front of the listener are perceived as elevated above their specified locations. Spectral coloration introduced by binaural enhancement post-processing can result in unnatural timbral characteristics that deviate from the intended spectral content. These challenges call for the development of subjective evaluation methodologies that can systematically assess the effectiveness of binaural enhancement techniques in addressing these perceptual artifacts.

Given the multiplicity of binaural reproduction situations, a single subjective evaluation survey can only cover a subset of the condition space. In Section 2, a review of prior research motivates the definition of an evaluation approach applicable regardless of the applicative scenario and technological conditions. In Section 3, we propose more specifically an evaluation methodology applicable to headphone audio reproduction enhancement processing methods in a single-source presentation scenario. In Section 4, we present some preliminary results of a perceptual survey utilizing this methodology. In conclusion, we comment on the applicability of the proposed methodology to other binaural reproduction scenarios, including complex sound scenes, head-tracking, interactive experiences, or fixed loudspeaker systems.

2 Review of Prior Research and Proposed Approach

2.1 Review of Relevant Literature

Previous investigative works that evaluated binaural reproduction have both implicitly and explicitly expressed the need for a methodological framework that is considerate of, and applicable to, a range of testing conditions, including varying playback device types, application scenarios, and signal processing techniques [5, 6]. A desirable framework would be robust enough to reliably assess a set of perceptual attributes across those conditions. However, existing methods for evaluating binaural enhancement are largely limited in their scope, either in terms of testing conditions or evaluation criteria [5, 7]. Many works may provide valuable guidance without explicitly proposing general evaluation methodologies [6, 8]. As such, while rigorous research may draw on best practice principles to devise the design of an individual published subjective experiment, there are still large methodological variations between studies which make cross-comparisons difficult.

Overwhelmingly, the most prevalent challenges that binaural enhancement techniques aim to address are externalization, elevation, front/back distinction, and spectral coloration [9]. When compared with individualized head-related transfer functions (HRTFs), generic or non-individualized HRTFs are consistently rated lower in each of these four subjective attributes [10, 11, 12], but are often inaccessible or difficult for the layperson to acquire [13, 14]. Given this, it has become a focal point of binaural technology research to investigate and improve the subjective quality of generic or non-individualized binaural rendering [15].

Vertical localization upward bias is among the most commonly reported artifacts of headphone reproduction, especially when evaluating dummy-head recordings or non-individualized HRTFs [12, 16], and may become more prevalent when the specified position of the sound source is in the frontal region of the listener's head [17]. These vertical errors may be caused by inaccurate spectral content [18, 19, 20, 21, 22], incorrectly assumed symmetry of the applied HRTF models [23], or by the listener's impression of the space [24, 25].

Coinciding with efforts to mitigate vertical localization bias, there are equally intensive efforts to improve externalization. As detailed in [26], established approaches often rely on the simulation of virtual loudspeakers (a.k.a. binaural virtualization) [27, 28, 29, 30]. Previous studies have isolated a number of factors that may impact externalization [31], including HRTF selection [32, 33], binaural room impulse response (BRIR) length [34, 35], and real-time compensation of head movements [15, 36].

A significant challenge associated with binaural virtualization processing is coloration or timbral alteration [37, 38]. HRTF-based filters inherently introduce spectral alterations, which may be more or less noticeable depending on the similarity between the applied HRTF and the listener's own HRTF [10]. Motivated by situations where HRTF individualization is inaccessible or impractical, methods to decrease perceived coloration of non-individualized HRTFs have yielded promising results [30]. However, many commercially available binaural enhancement approaches have relied on artificial reverberation. These may introduce further coloration, often to the subjective preferential detriment of the listener [6, 37, 39, 40, 41], which has generated interest in timbrally transparent binaural enhancement [26, 42, 43].

The final frequently evaluated and addressed difficulty related to binaural virtualization is front-vs-back confusion. One of the most common methods to mitigate these confusions is through dynamic binaural synthesis compensating head-rotation in real time (a.k.a. head-tracking) [34, 44, 45]. However, dynamic virtualization processing is not always a possibility, and efforts to reduce front-back confusions under static binaural conditions have traditionally focused on HRTF individualization [10, 11], addition of artificial acoustic reverberation or reflections [25], or directionally-dependent spectral modifications to either the applied HRTF [46], or the program content [47, 48].

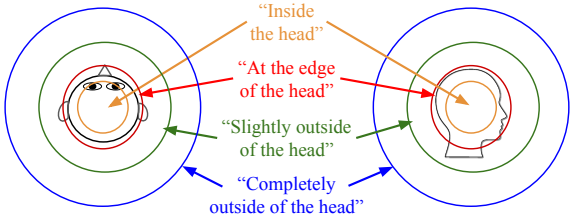


Fig. 1: Illustration for "degree of externalization"

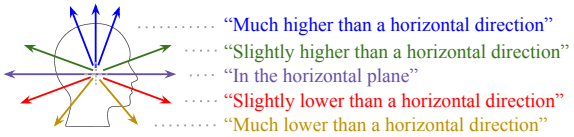


Fig. 2: Illustration for "degree of elevation"

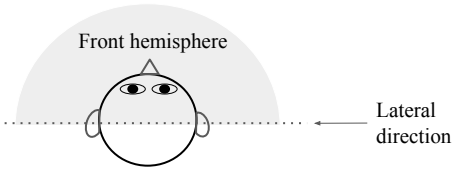


Fig. 3: Illustration for "front hemisphere"

2.2 Proposed General Approach

In view of the aforementioned challenges of binaural reproduction, we propose a minimal set of subjective attributes and their associated subjective rating scales, defined to be applicable to any psychoacoustic evaluation and comparison of binaural reproduction enhancement methods:

- **Degree of externalization** relative to the edge of the head, rated on a 4-step scale – see Figure 1
- **Degree of elevation** relative to a horizontal plane containing the listener's ears (a.k.a. the equatorial plane), rated on a 5-step scale – see Figure 2
- **Degree of front/back distinction** relative to a lateral direction of sound arrival (a.k.a. the frontal plane) – see Figure 3
- **Presence of room tone**, defined here as the sound of acoustic reverberation or reflections in a room
- **Spectral coloration**, defined here as the alteration of a sound's spectral content.

3 Methodology

We present a comprehensive subjective evaluation methodology for evaluating binaural reproduction enhancement methods across multiple perceptual dimensions, considering headphone-based audio reproduction in a single-source presentation scenario. The evaluation framework aims to provide a consistent and reliable assessment of binaural enhancement techniques across varying audio source types and processing conditions.

The methodology consists of three complementary subjective tests designed to evaluate the key attributes of binaural reproduction identified in Section 2.2: externalization, elevation perception, front/back distinction, room tone, and spectral coloration. Three subjective test applications were developed using MATLAB App Designer to explore different modes of subjective assessment:

1. **Absolute Judgment Test**: Individual assessment of audio stimuli on perceptual scales for each sample under test.
2. **Relative Comparison Test**: Pairwise comparative evaluation of samples under test.
3. **Spectral Coloration Test**: Focused assessment of timbral differences of each sample under test relative to a reference sample.

In the perceptual survey presented below, the samples under test differ by binaural processing method used. Alternatively, with the same proposed methodology, they could differ by source content, application scenario, or other presentation factors (such as HRTF individualization, artificial reverberation, or head-tracking – see Section 1 for examples of possible variations).

3.1 Source Stimuli

Two distinct audio stimuli were selected to represent different acoustic characteristics:

- **Female speech**: Representing typical voice content recorded in anechoic conditions
- **Rain stick**: Selected for its stationary noise-like characteristics, providing an example of broadband spectral content.

All stimuli were sampled at 96 kHz with 24-bit resolution. For each trial, a random 5-second segment was extracted from the source audio file to minimize adaptation effects and ensure stimulus variety.

3.2 Spatial Reproduction Conditions

Four processing conditions were compared:

1. **Unprocessed mono**: Baseline monophonic signal
2. **Binaural processing**: Standard binaural rendering using HRTF convolution [1]
3. **Binaural with enhancement #1**: Enhanced binaural processing, variant #1 (see Section 3.3)
4. **Binaural with enhancement #2**: Enhanced binaural processing, variant #2.

Spatial rendering was performed using a diffuse-field compensated HRTF dataset of a Neumann KU-100 dummy head, collected during the BiLi measurement campaign [49]. Stimuli were generated for 11 equatorial locations selected from the following set of azimuth angles: 0° (front), 15°, 30°, 45°, 60°, 90°, 120°, 135°, 150°, 165°, and 180°. In order to reduce experiment duration while maintaining spatial coverage, left/right hemisphere symmetry in subjective judgment was assumed. Each participant was presented with stimuli at 0° and 180° positions, with the remaining 9 equatorial locations randomly assigned for each subject between the left and right hemispheres. Content was played back in a controlled listening environment (quiet listening booth, RT_{60} of 0.2 s) over closed-back headphones (Audio-Technica ATH-M40x) at a comfortable listening level.

3.3 Binaural Processing Methods Compared

Besides standard binaural rendering, two enhancement processing methods were compared. Both apply the diffuse tail processing principle introduced in [26] and demonstrated in [50]. Figure 4 illustrates variant #2, where g_0 denotes an attenuation factor ensuring amplitude normalization. In variant #1, the diffuse tail processing module is fed by a mono downmix of the output signals of the binaural rendering module.

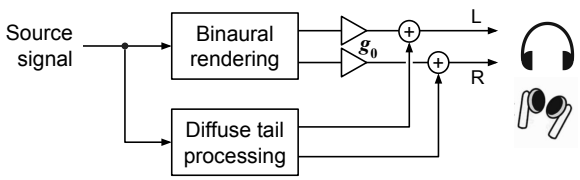


Fig. 4: Signal-flow diagram of the binaural enhancement processing method evaluated here

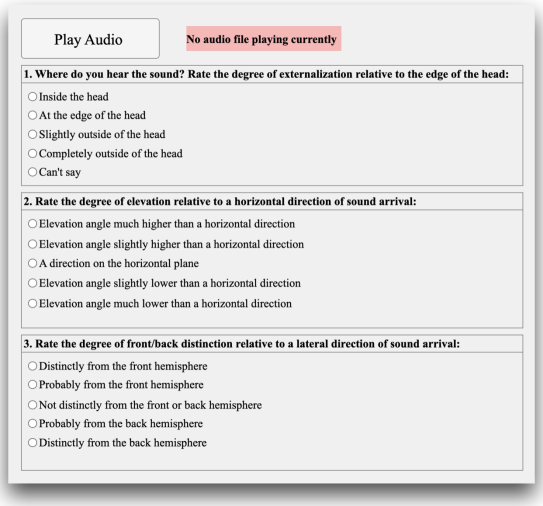


Fig. 5: Screenshot of Absolute Judgment test

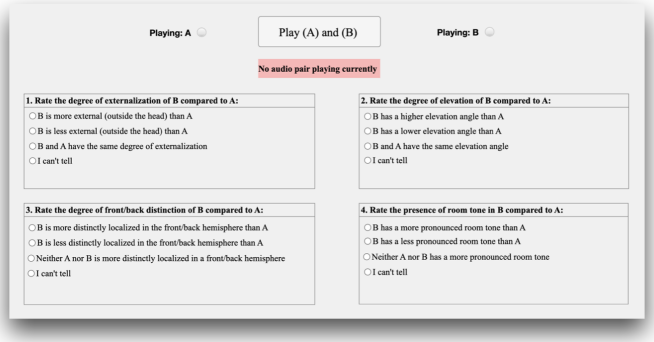


Fig. 6: Screenshot of Relative Comparison test

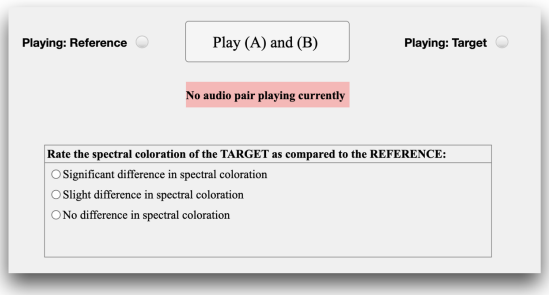


Fig. 7: Screenshot of Spectral Coloration test

3.4 Absolute Judgment Test Protocol

The Absolute Judgment test required participants to evaluate individual audio stimuli on three perceptual dimensions:

1. **Externalization**: "Where do you hear the sound? Rate the degree of externalization relative to the edge of the head"
2. **Elevation**: "Rate the degree of elevation relative to a horizontal direction of sound arrival"
3. **Front/Back Distinction**: "Rate the degree of front/back distinction relative to a lateral direction of sound arrival."

Each participant was randomly assigned the female speech or the rain stick stimulus. To ensure reliability, each condition was presented twice with the stimuli duplicated and order randomized, yielding 68 trials: 11 locations, 3 processing methods, and an additional mono trial, with each trial repeated twice. A total of 13 subjects participated in the Absolute Judgment test.

3.5 Relative Comparison Test Protocol

The Relative Comparison protocol used pairwise comparisons in which participants evaluated stimulus B relative to stimulus A, along four perceptual dimensions:

1. **Externalization**: "Rate the degree of externalization of B compared to A"
2. **Elevation**: "Rate the degree of elevation of B compared to A"
3. **Front/Back Distinction**: "Rate the degree of front/back distinction of B compared to A"
4. **Room Tone**: "Rate the presence of room tone in B compared to A"

Six pairwise comparisons were implemented: 1) Mono vs Binaural; 2) Mono vs Enhancement #1; 3) Mono vs Enhancement #2; 4) Binaural vs Enhancement #1; 5) Binaural vs Enhancement #2; 6) Enhancement #1 vs Enhancement #2. In order to minimize order effects, each pair was presented in original and reversed order (thus effectively providing duplicate presentations for reliability assessment), yielding 134 trials: 12 pairs for each of the 11 locations (6 original + 6 reversed) and 2 mono vs mono trials. A total of 13 subjects participated in the Relative Comparison test.

3.6 Spectral Coloration Test Protocol

The Spectral Coloration test focused specifically on timbral differences using a pairwise comparison methodology. Participants rated the difference in spectral coloration between a target sample and a reference sample, with the unprocessed monophonic signal serving as the reference sample for all comparisons. Three comparison pairs were evaluated: 1) Mono vs Binaural; 2) Mono vs Enhancement #1; 3) Mono vs Enhancement #2. Similarly to the Relative Comparison test, each pair was presented randomly, in both original and reversed order to control for order effects, yielding 68 trials: 6 pairs for each of the 11 locations (3 original + 3 in reverse order), and 2 mono vs mono. A total of 26 subjects participated in the Spectral Coloration test.

3.7 Experimental Controls

The trials within each test were presented in a randomized order to minimize systematic biases and learning effects. Each experimental condition was presented twice across all tests to assess response consistency and improve measurement reliability. In the Absolute Judgment protocol, this was achieved through stimulus duplication. In the Relative and Spectral Coloration tests, reversed pair presentations served as duplicates. Participants were randomly assigned to one of two stimulus conditions (female speech or rain stick), in order to assess the generalizability of findings across different audio content types. The experimental session duration varied from 1.5 to 2 hours per participant, depending on the specific test assignment and individual response patterns.

4 Preliminary Results

We collected data from 26 participants who were randomly assigned to one of two experimental conditions: 13 participants completed the Absolute Judgment and Spectral Coloration study, while the remaining 13 participants completed the Relative Comparison and Spectral Coloration study. This paper presents the statistical analysis and results for the Absolute Judgment and Spectral Coloration study only. Detailed results, including comprehensive analysis of the Relative Comparison study, will be reported in a subsequent publication. To evaluate differences in perceived externalization, elevation, and spectral coloration across processing methods, we fit a linear mixed effects model that included fixed effects predictors of processing method and stimulus, and random intercepts for each participant to account for the clustered nature of the data [51].

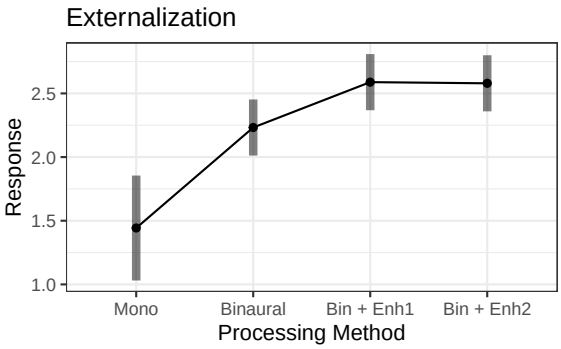


Fig. 8: Predicted externalization ratings, aggregated across tested positions in the equatorial plane

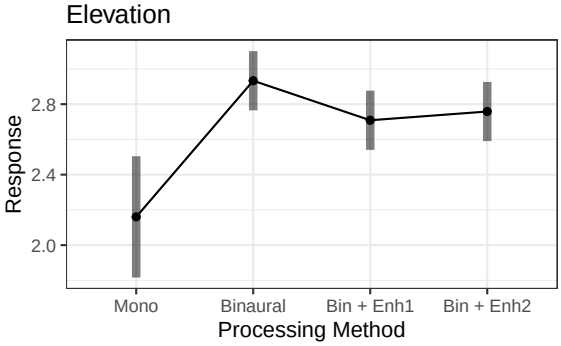


Fig. 9: Predicted elevation ratings, aggregated across tested positions in the equatorial plane

Although the linear mixed effects model does not account for the ordinal nature of the perceptual ratings, we validated our results through a sensitivity analysis by fitting a Cumulative Link Mixed Model from which the results did not differ from the main analyses (results not shown) [52]. Across all outcome measures, we did not observe a main effect of stimulus (rainstick and speech; $p > 0.05$ for all models).

4.1 Externalization

In Figure 8, we observe a sharp increase in externalization ratings when moving from Mono to Binaural rendering ($\Delta_{Bin\ vs\ Mono} = 0.79$ (95% CI 0.40, 1.12), $p < 0.001$), and a modest increase with Enhancement #1 ($\Delta_{Enh1\ vs\ Bin} = 0.35$ (95% CI 0.20, 0.52), $p < 0.001$). Ratings for Enhancement #2 were significantly higher than Binaural alone ($\Delta_{Enh2\ vs\ Bin} = 0.35$ (95% CI 0.19, 0.51) $p < 0.001$). There was no evidence of a difference between Enhancement #1 and #2 ($p = 0.91$).

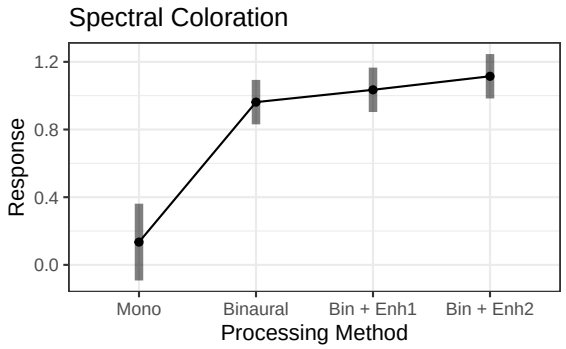


Fig. 10: Predicted coloration ratings, aggregated across tested positions in the equatorial plane

4.2 Elevation

In Figure 9, we observe a sharp increase in elevation ratings when moving from the Mono condition to the Binaural rendering method ($\Delta_{Bin\ vs\ Mono} = 0.77$ (95% CI 0.44, 1.11), p < 0.001), followed by a modest decrease with Enhancement #1 ($\Delta_{Enh1\ vs\ Bin} = -0.22$ (95% CI 0.36, -0.09), p = 0.001). Although there was no evidence of a difference between Enhancements #2 and #1 (p = 0.48), the ratings for Enhancement #2 were lower than Binaural alone, but with less of a difference than compared to Enhancement #1 ($\Delta_{Enh2\ vs\ Bin} = -0.17$ (95% CI -0.31, -0.04) p = 0.01). In future research, we'll investigate the dependence of elevation ratings on azimuth angle in the equatorial plane, and assess whether Enhancements #1 or #2 can mitigate the "frontal boost" effect reported in Sections 1 and 2 (illustrated, for instance, in Figure 1 of [26]).

4.3 Coloration

As shown in Figure 10, there is a reported increase in perceived spectral coloration from Mono to Binaural conditions ($\Delta_{Bin\ vs\ Mono} = 0.83$ (95% CI 0.62, 1.03) p < 0.001), with Enhancement #1 and Enhancement #2 conditions both providing further increases to perceived coloration ($\Delta_{Enh1\ vs\ Mono} = 0.90$ (95% CI 0.69, 1.11) p < 0.001; $\Delta_{Enh2\ vs\ Mono} = 0.98$ (95% CI 0.77, 1.19) p < 0.001). Although the differences between Enhancement #1 and Enhancement #2, and between Enhancement #1 and Binaural were not statistically significant (p = 0.06; p = 0.09, respectively), these results still indicate that each successive processing method introduces additional coloration compared to the Mono condition, with the highest ratings observed for the Enhancement #2 condition.

5 Discussion and Conclusions

This study presents a comprehensive subjective evaluation methodology designed to systematically assess binaural enhancement techniques across a set of key perceptual attributes that characterize successful headphone-based spatial audio reproduction. The proposed framework addresses a significant gap in the field, by providing a standardized approach for evaluating binaural enhancement methods that can be applied consistently across different research contexts and technological conditions.

The methodology's three-component structure (absolute judgment, relative comparison, and spectral coloration assessment) represents an advancement over existing evaluation approaches that are often limited in scope or focus on single perceptual dimensions. By incorporating multiple complementary evaluation methods, the framework provides a more comprehensive assessment of binaural enhancement performance, while maintaining practical feasibility for research and development applications. A key strength of the proposed methodology is its focus on the fundamental perceptual challenges that have consistently emerged in binaural reproduction research: externalization, elevation perception, front/back distinction, and spectral coloration. This targeted approach ensures that the evaluation framework directly addresses critical aspects of binaural reproduction quality, while ensuring that subjective survey results obtained from independent experiments can be compared and aggregated.

The preliminary results demonstrate the effectiveness of the methodology in detecting meaningful differences between processing methods across all evaluated dimensions. The framework successfully captured expected patterns in externalization and elevation perception while revealing nuanced differences in spectral coloration that might not have been apparent through single-dimension evaluations. These findings validate the methodology's sensitivity and suggest its potential for identifying perceptually relevant differences between enhancement conditions. The scalability of the methodology to more complex scenarios remains an important consideration for future development. While the current framework provides a solid foundation for evaluating binaural enhancement techniques, its extension to include head-tracking, multi-source scenarios, and interactive applications, to name a few, will require adaptation to evaluate its effectiveness.

References

[1] Hammershøi, D. and Møller, H., "Methods for binaural recording and reproduction," *Acustica United with Acta Acustica*, 88(3), pp. 303–311, 2002, ISSN 1610-1928.

[2] Jot, J.-M. and Lee, K. S., "Augmented Reality Headphone Environment Rendering," in *Audio Eng. Soc. Int. Conf. on Audio for Virtual and Augmented Reality*, 2016.

[3] Roginska, A., *Binaural Audio Through Headphones*, Routledge, New York, NY, USA, 2018, ISBN 1138900001, 978-1138900004, doi:10.4324/9781315707525.

[4] Jot, J.-M., Carpentier, T., and Warusfel, O., "Acoustic Objects: bridging immersive audio creation and distribution systems," in *Audio Eng. Soc. 158th Conv.*, 2025.

[5] Mason, R. and Rumsey, F., "An assessment of spatial performance of virtual home theatre algorithms by subjective and objective methods," in *Audio Eng. Soc. 108th Conv.*, 5137, 2000.

[6] Lorho, G., "Evaluation of spatial enhancement systems for stereo headphone reproduction by preference and attribute rating," in *Audio Eng. Soc. 118th Conv.*, 6514, 2005.

[7] Schlesinger, A., Ramirez, J.-P., and Boone, M. M., "Evaluation of a speech-based and binaural Speech Transmission Index," in *Audio Eng. Soc. 40th Intl. Conf.*, 4-6, pp. 1–10, 2010.

[8] Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J., Brinkman, F., and Weinzierl, S., "A Spatial Audio Quality Inventory (SAQI)," *Acta Acustica*, 100(5), pp. 984–994, 2014.

[9] Reardon, G., Calle, J. S., Genovese, A., Zalles, G., Olko, M., Jerez, C., Flanagan, P., and Roginska, A., "Evaluation of Binaural Renderers: A Methodology," in *Audio Eng. Soc. 143rd Conv.*, 359, 2017.

[10] Jenny, C. and Reuter, C., "Usability of Individualized Head-Related Transfer Functions in Virtual Reality: Empirical Study With Perceptual Attributes in Sagittal Plane Sound Localization," *JMIR Serious Games*, 8, 2020.

[11] Oberem, J., Richter, J.-G., Setzer, D., Seibold, J., Koch, I., and Fels, J., "Experiments on localization accuracy with non-individual and individual HRTFs comparing static and dynamic reproduction methods," *bioRxiv preprint*, pp. 1–11, 2020.

[12] Frankenbach, C., Herzog, C., Seitz, J., and Fels, J., "Evaluating Localization Accuracy of Dynamic Binaural Synthesis: Differences Between Individual, Estimated, and Generic HRTFs," *Journal of the Audio Engineering Society*, 73(1/2), pp. 5–23, 2025.

[13] Katz, B. F. G. and Parseihian, G., "Perceptually based head-related transfer function database optimization," *The Journal of the Acoustical Society of America*, 131(2), pp. EL99–EL105, 2012, ISSN 0001-4966, doi:10.1121/1.3672641.

[14] Mendonça, C., Campos, G., Dias, P., Vieira, J., Ferreira, J. P., and Santos, J., "On the Improvement of Localization Accuracy with Non-individualized HRTF-Based Sounds," *Journal of the Audio Engineering Society*, 60, pp. 821–830, 2012.

[15] Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B. F., and de Boishéraud, C., "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," *Journal of the Acoustical Society of America*, 141(3), pp. 2011–2023, 2017.

[16] Pietrzak, A. P., "Azimuth and elevation errors in binaural reproduction of ambisonics sound," *Vibrations in Physical Systems*, 35(2), pp. 1–8, 2024, ISSN 0860-6897, doi:10.21008/j.0860-6897.2024.2.03.

[17] Lagnel, B., "Voix Intérieure Binaurale Natif et DAW," https://www.lesonbinaural.fr/edit/pdf/voix_interieure_binaural_natif_daw.pdf, 2022.

[18] Bloom, J. P., "Creating Source Elevation Illusions by Spectral Manipulation," *Journal of the Audio Engineering Society*, 25, pp. 560–565, 1977.

[19] Avendano, C., Algazi, V. R., and Duda, R. O., "A head-and-torso model for low-frequency binaural elevation effects," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99*

(Cat. No.99TH8452), pp. 179–182, 1999, doi: 10.1109/ASPAA.1999.810879.

[20] Iida, K., "Estimation of sound source elevation by extracting the vertical localization cues from binaural signals," in *Proc. Mtgs. Acoust.*, volume 4, pp. 1–11, 2008, ISSN 1939-800X, doi: 10.1121/1.2980020.

[21] Macpherson, E. A. and Sabin, A. T., "Vertical-plane sound localization with distorted spectral cues," *Hearing Research*, 306, pp. 76–92, 2013, ISSN 0378-5955, doi:https://doi.org/10.1016/j.heares.2013.09.007.

[22] Yao, D., Li, J., Xia, R., and Yang, F., "A Parametric Method for Elevation Control," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 181–185, 2018, doi: 10.1109/IWAENC.2018.8521349.

[23] Genovese, A., Juras, J., Miller, C., and Roginska, A., "The Effect of Elevation on ITD Symmetry," in *Audio Eng. Soc. Conf. on Headphone Technology*, 2016.

[24] Roßkopf, S., Kroczek, L. O., Stärz, F., Blau, M., de Par, S. V., and Mühlberger, A., "The impact of binaural auralizations on sound source localization and social presence in audiovisual virtual reality: converging evidence from placement and eye-tracking paradigms," *Acta Acust.*, 8, pp. 1–21, 2024, doi:10.1051/aacus/2024064.

[25] Carlini, A., Bordeau, C., and Ambard, M., "Auditory localization: a comprehensive practical review," *Frontiers in Psychology*, 15, 2024, ISSN 1664-1078, doi:10.3389/fpsyg.2024.1408073.

[26] Jot, J.-M., Lukin, A., and Landschoot, C., "Binaural Externalization Processing - from Stereo to Object-Based Audio," in *Audio Eng. Soc. 153rd Conv.*, 2022.

[27] Rubak, P., "Headphone Signal Processing System for Out-of-the-Head Localization," in *Audio Eng. Soc. 90th Conv.*, 3063, 1991.

[28] Jot, J.-M., Larcher, V., and Warusfel, O., "Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony," in *Audio Eng. Soc. 98th Conv.*, 3980, 1995.

[29] Kirkeby, O., "A Balanced Stereo Widening Network for Headphones," in *Audio Eng. Soc. 22nd Intl. Conf. on Virtual, Synthetic, and Entertainment Audio*, 000249, 2002.

[30] Merimaa, J., "Modification of HRTF Filters to Reduce Timbral Effects in Binaural Synthesis," in *Audio Eng. Soc. 127th Conv.*, 7912, 2009.

[31] Best, V., Baumgartner, R., Lavandier, M., Majdak, P., and Kopčo, N., "Sound Externalization: A Review of Recent Research," *Trends in Hearing*, 24, pp. 1–14, 2020.

[32] Hartmann, W. and Wittenberg, A., "On the externalization of sound images," *The Journal of the Acoustical Society of America*, 99, pp. 3678–88, 1996, doi:10.1121/1.414965.

[33] Kim, S.-M. and Choi, W., "On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach," *The Journal of the Acoustical Society of America*, 117(6), pp. 3657–3665, 2005, ISSN 0001-4966, doi: 10.1121/1.1921548.

[34] Begault, D. R., Wenzel, E. M., and Anderson, M. R., "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, 49(10), pp. 904–916, 2001.

[35] Catic, J., Santurette, S., and Dau, T., "The role of reverberation-related binaural cues in the externalization of speech," *Journal of the Acoustical Society of America*, 138(2), p. 1154–1167, 2015.

[36] Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A., "The Contribution of Head Movement to the Externalization and Internalization of Sounds," *PLoS ONE*, 8(12), pp. 1–12, 2013, doi:10.1371/journal.pone.0083068.

[37] Jot, J.-M. and Avendano, C., "Spatial Enhancement of Audio Recordings," in *Audio Eng. Soc. 23rd Intl. Conf.*, 15, 2003.

[38] Davidson, G., Darcy, D., Fielder, L., Shuang, Z., Graff, R., Breebaart, J., and Crum, P., "Design and Subjective Evaluation of a Perceptually-Optimized Headphone Virtualizer," in *Audio Eng. Soc. 140th Conv.*, 9588, 2016.

[39] Olive, S., "Evaluation of Five Commerical Stereo Enhancement 3D Audio Software Plug-ins," in *Audio Eng. Soc. 110th Conv.*, 5386, 2001.

[40] Lorho, G., Isherwood, D., Zacharov, N., and Huopaniemi, J., "Round robin subjective evaluation of stereo enhancement systems for headphones," in *Audio Eng. Soc. 22nd Intl. Conf. on Virtual, Synthetic, and Entertainment Audio*, 2002.

[41] Lorho, G. and Zacharov, N., "Subjective Evaluation of Virtual Home Theatre Sound Systems for Loudspeakers and Headphones," in *Audio Eng. Soc. 116th Conv.*, pp. 1–17, 2004.

[42] Breebaart, J. and Schuijers, E., "Phantom Materialization: A Novel Method to Enhance Stereo Audio Reproduction on Headphones," *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8), pp. 1503–1511, 2008, doi:10.1109/TASL.2008.2002983.

[43] Goodwin, M. and Jot, J.-M., "Binaural 3-D audio rendering based on spatial audio scene coding," in *Audio Eng. Soc. 123rd Conv.*, volume 3, pp. 1282–1293, 2007.

[44] Perrett, S. and Noble, W., "The contribution of head motion cues to localization of low-pass noise," *Perception Psychophysics*, 59, pp. 1018–1026, 1997, doi:10.3758/BF03205517.

[45] Wightman, F. L. and Kistler, D. J., "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, 105(5), pp. 2841–2853, 1999, ISSN 0001-4966, doi:10.1121/1.426899.

[46] Frank, M. and Zotter, F., "Simple Reduction of Front-Back Confusion in Static Binaural Rendering," in *Fortschritte der Akustik - DAGA*, pp. 329–332, 2018.

[47] Brimijoin, W. O. and Akeroyd, M., "The Effects of Hearing Impairment, Age, and Hearing Aids on the Use of Self-Motion for Determining Front/Back Location," *Journal of the American Academy of Audiology*, 27, pp. 588–600, 2016, doi:10.3766/jaaa.15101.

[48] Zhang, P. X. and Hartmann, W. M., "On the ability of human listeners to distinguish between front and back," *Hearing Research*, 260(1), pp. 30–46, 2010, ISSN 0378-5955, doi:https://doi.org/10.1016/j.heares.2009.11.001.

[49] Carpentier, T., Bahu, H., Noisternig, M., and Warusfel, O., "Measurement of a Head-Related Transfer Function Database with High Spatial Resolution," in *7th EAA Forum Acusticum*, Kraków, Poland, 2014.

[50] Landschoot, C. and Jot, J.-M., "Colorless Binaural Externalization Processing Demonstration," https://www.chrislandschoot.com/binaural-externalization, 2025.

[51] Gałecki, A. and Burzykowski, T., *Linear Mixed-Effects Model*, pp. 245–273, Springer New York, New York, NY, 2013, ISBN 978-1-4614-3900-4, doi:10.1007/978-1-4614-3900-4_13.

[52] Christensen, R. H. B. and Brockhoff, P. B., "Analysis of sensory ratings data with cumulative link models," *Journal de la société française de statistique*, 154(3), pp. 58–79, 2013.