Audio Engineering Society

Convention Express Paper 382

Presented at the AES 159th Convention

2025 October 23-25, New York, NY, USA

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express 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.

Spectral and Spatial Discrepancies Between Stereo and Binaural Spatial Masters in Headphone Playback: A Perceptual and Technical Analysis

Richard King[1,2] and Marc Thériault[3]

[1] *Graduate Program in Sound Recording, McGill University, Montréal, Canada*

[2] *Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT), Montréal, Canada*

[3] *Le Lab Mastering, Montréal, Canada*

Correspondence should be addressed to Richard King (richard.king@mcgill.ca) or Marc Thériault (marctheriault@lelabmastering.com)

ABSTRACT

This paper investigates the key perceptual and measurable differences between traditional stereo masters and binaural spatial audio masters during headphone playback. With headphone use as the dominant format for music listening, the translation of immersive audio formats into binaural renderings introduces measurable shifts in frequency response, stereophonic imaging behaviour, and loudness characteristics. Through comparative testing of stereo versus binaural content across calibrated headphone systems, the tonal tilt induced by HRTF-based rendering is quantified, along with the deviation in phantom center and stereo width behaviour, and variations in integrated loudness and peak-to-loudness ratios.

1 Introduction

Edison Research (2025) reports that 59% of U.S. persons aged 12 and over own Bluetooth-enabled wireless headphones, while only 17–22% listened to audio via smart speakers in the past week/month [1]. Additionally, 87% of Americans say they use headphones for music listening. Headphones are on track to become the most used playback technology for consuming music, while the use of loudspeakers in a component or hi-fi system is becoming an increasingly rare practice. Many publications have previously investigated comparisons between stereo and binaural headphone listening, and many authors have proposed solutions for improved binaural immersive listening in general, for example, [2,3,4].

2 Context

The application of cross-feed in headphone listening has been investigated ever since the 1960s, in an effort to simulate the experience of listening to loudspeakers [5]. The desired outcome is an externalization of the soundstage, as is the case when listening with loudspeakers. Overall, the image is presented outside the head, in front of the listener, at the perceived listening angle of approximately +/- 30 degrees, as is common in a stereo loudspeaker playback configuration. A more complete investigation into the effects of cross-feed and phantom center signals can be found in King et al, 2023 [6].

The difference in image width is quite possibly the most common observation of listeners who are new to immersive audio playback, as they find that stereo reproduction over headphones sounds much wider than a binaural immersive mix of the same musical program.

The fact that the image is inside the listeners' head is not always perceived as a negative attribute, since the wide presentation is reported by some listeners to be a positive benefit over the more natural image width of a binaural presentation [7]. However, the principal advantage of binaural signal processing specifically in immersive audio production, lies in its capacity to reconstruct a perceptually accurate frontal soundstage while simultaneously enhancing spatial resolution and listener envelopment.

Cross-feed is the result of signals coming from each loudspeaker being received by both the left and right ear of the listener, at slightly different timings and intensities, as demonstrated in Figure 1. The longer signal paths are heard as delayed, and the levels are affected by head shadowing. As signals pass to the ear on the opposite side of the source, they are attenuated by the listener's head. Figure 2 shows transfer function of the left channel (normalized), compared to the cross-feed on the right channel (as heard at the right ear of the listener). A smoothing filter is applied so that only one instance of phase wrap is shown on the graph.

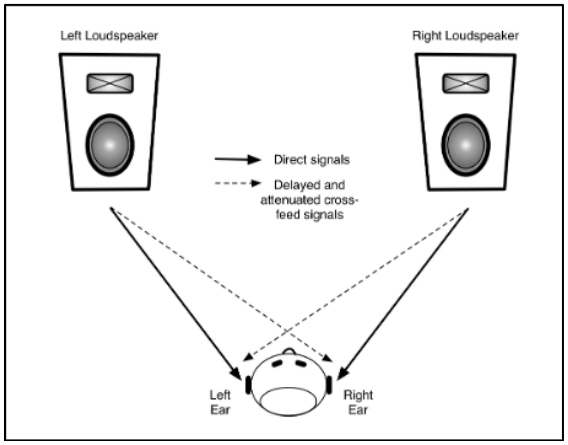


Figure 1. Cross-feed in loudspeaker listening (from [3]).

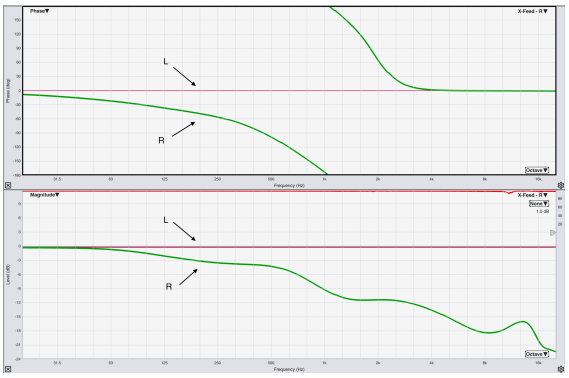


Figure 2. Cross-feed frequency and phase response.

Figure 3 presents a Mid/Side (M/S) encoded version of the measurement in Figure 2. M/S processing is explained fully in Section 3.3. It is interesting to note that the Mid component shows the same curve we see when analysing phantom center, with a bass boost and a dip in the mid-range around 1kHz, as in [6].

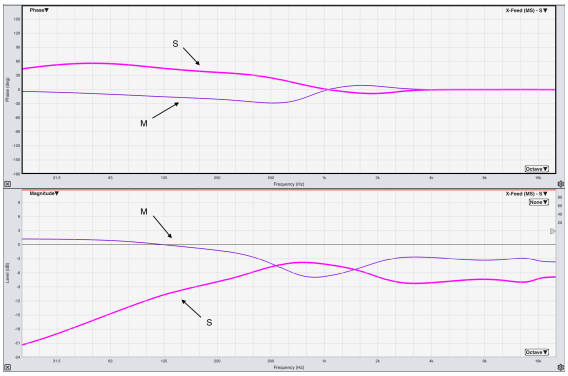


Figure 3. M/S encoding of cross-feed frequency and phase response.

Cross-feed is generally applied in binaural rendering using a profile based on an "average" Head Related Transfer Function, or HRTF. While this process results in the externalization of the auditory scene, it alters the frequency response of the audio content, as filtered through a virtual or measured head, torso, pinnae and ear canal system. Perceived loudness will also be different, as low frequency information in the center of the image will be boosted. Moreover, the listener is effectively experiencing two HRTF filters – one that is applied in the binaural renderer, and the natural HRTF of the listener. Many binaural profiles also introduce some form of room simulation to

increase the externalization effect – this additional "spatialization" often sounds unpleasant and further modifies the original (stereo) master, in terms of timbral and phase artifacts.

Headphone manufacturers have put a great deal of effort into matching the timbre of a stereo loudspeaker listening experience by adding low frequency content that is normally enhanced by the effect of cross-feed in loudspeaker listening, at least in popular music mixes where most low frequency resides in the phantom center image. Additionally, certain mid-frequencies may be attenuated to simulate the cancellation in phantom center elements. When standard headphones are used for binaural listening, there is a risk of a double frequency filter effect – one applied by the headphone manufacture, and one introduced by the cross-feed in the binaural filter. Therefore, most commercial binaural profiles introduce a filter that reduces the low frequency range to offset the boost applied by the headphone designers.

In stereo headphone listening, widely spread bass content is experienced due to the 100% separation of the left and right channels, i.e. zero cross-feed is present. AB or spaced microphone recording techniques, originally developed for loudspeaker listening, exhibit low frequency decorrelation that provides increased envelopment. However, when listening over headphones in stereo, the low frequency width is exaggerated. This is where the cross-feed in binaural processing can help to create a more realistic presentation. Highly decorrelated low frequency content, as is common in classical orchestral recordings and modern electronic music may exhibit some "narrowing" when the cross-feed process is applied. This is actually a correction, as the result more closely resembles loudspeaker listening.

3 Metering

Beyond frequency response and loudness measurements, metering tools provide quantitative insight into the spatial and phase behaviour of stereo and binaural signals. Three categories of meters are particularly relevant for this study:

3.1 Correlation Meter

The inter-channel correlation coefficient is defined as:

$$r = \frac{\mathrm{E}[(L(t) - \mu_L)(R(t) - \mu_R)]}{\sigma_L \sigma_R}, \quad -1 \le r \le 1 \qquad (1)$$

In stereo headphone playback, low frequencies often appear decorrelated (r ≈ 0) because there is no acoustic cross-feed, which exaggerates bass width and destabilizes the phantom center.

In binaural renderings however, HRTF-based cross-feed increases low-frequency correlation (r→ +1, stabilizing bass and anchoring the phantom image. At mid and high frequencies, controlled decorrelation (r < 1) is deliberately introduced to increase spaciousness and envelopment.

3.2 Panorama Meter

A panorama index can be expressed as the normalized energy ratio:

$$p(t) = \frac{E_L(t) - E_R(t)}{E_L(t) + E_R(t)}, \quad -1 \le p \le +1 \qquad (2)$$

In stereo mixdowns (e.g., ITU BS.775 to 2.0 or 2/0 [8]), panorama is constrained to the L/R axis via amplitude panning.

In binaural renderings, perceived lateralization depends not only on level balance but also on ITD and ILD cues introduced by HRTFs. As a result, *p(t)* alone does not fully capture perceived positioning in binaural playback.

3.3 Mid/Side Spectrum Analysis

Stereo signals can be transformed into Mid (*M*) and Side (*S*) components:

$$M(t) = \frac{L(t) + R(t)}{2}, \quad S(t) = \frac{L(t) - R(t)}{2} \qquad (3)$$

In stereo headphone playback, the Side spectrum often shows strong low-frequency content due to the complete channel separation, producing exaggerated bass width.

In binaural renderings, HRTF-based cross-feed reduces low-frequency Side energy, concentrating bass into the Mid channel and anchoring it at the

phantom center. At higher frequencies, the Side spectrum remains richer, providing stereo width and immersive spaciousness.

4 Technical Interpretation

Together, these metering approaches reveal the core trade-offs:

Stereo 2.0 mixdowns (BS.775 to 2.0 or 2/0):
Lower bass correlation (r closer to 0), strong low-frequency Side energy and wide but unstable phantom image.

Binaural renderings:
Improved bass correlation ($r \rightarrow +1$), Mid-dominant low-frequency spectrum, anchoring the phantom center, and controlled high-frequency Side energy, increasing spaciousness.

These differences highlight why binaural mixes, despite sometimes sounding narrower to listeners accustomed to stereo headphones, often reproduce a listening experience closer to loudspeaker reference conditions, commonly presented as a +/-30º listening angle in the horizontal plane.

5 Loudness Considerations

An important factor when comparing stereo and binaural masters is program loudness. Conventional stereo masters in popular music are typically delivered around -8 LUFS integrated, reflecting the loudness optimization practices of contemporary pop and commercial music production. By contrast, binaural renderers and immersive audio workflows generally require much lower output levels to preserve headroom, often targeting -18 LUFS integrated.

This difference in baseline level is critical when conducting perceptual comparisons. If the stereo master at -8 LUFS is compared directly to a binaural render at -18 LUFS, the listener will perceive the stereo version as louder, and by extension more impactful, even if the spatial quality of the binaural version is superior. For fair evaluation, both versions must be loudness-matched, either through normalization to a common LUFS target or by applying calibrated gain adjustments.

From a production standpoint, the lower nominal loudness of binaural delivery offers an additional benefit: it allows for greater dynamic range. With an average level 10 LU lower than typical stereo mastering practice, transient detail and musical dynamics can be preserved without risk of clipping or excessive limiting. This dynamic headroom is especially advantageous in immersive contexts, where spatial cues and reverberant energy depend on “micro dynamics” for naturalness and envelopment.

In summary, loudness alignment is not merely a technical detail but a prerequisite for valid perceptual testing of stereo versus binaural reproduction. Moreover, the lower-level ceiling of binaural workflows provides an opportunity to rethink mastering aesthetics, balancing loudness with fidelity and spatial impact.

6 Measurements

The following figures show comparisons of stereo and binaural content for overall frequency response, image width by frequency, loudness, directional response and phase relationships. The next two figures are taken from meter displays in the Nuendo audio workstation (Steinberg Media Technologies, [9]).

Figure 4 (below) shows a Mid/Side (M/S) spectrum analysis, multi-band panorama and multi-band correlation of de-correlated stereo pink noise. Here the Mid and Side components are almost equal throughout the spectrum, and there is a great deal of low frequency width evident in the panorama meter. The correlation meter shows very little focus in low frequency content.

Figure 5 reports the same measurements, but of binaurally encoded de-correlated pink noise. By contrast to Figure 4, Figure 5 exhibits a much lower Side component as compared to the Mid signal, and the panorama meter shows much less width in the low frequency range. Much stronger focus in low frequencies is exhibited in the correlation meter.

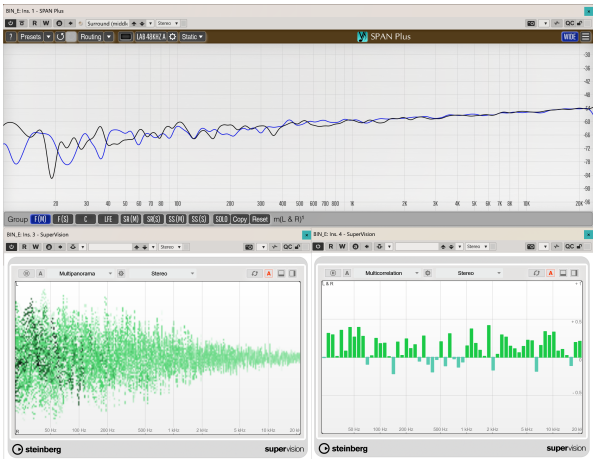


Figure 4. M/S spectrum analysis (top), multi-band panorama (bottom left) and multi-band correlation (bottom right) of de-correlated pink noise in stereo.

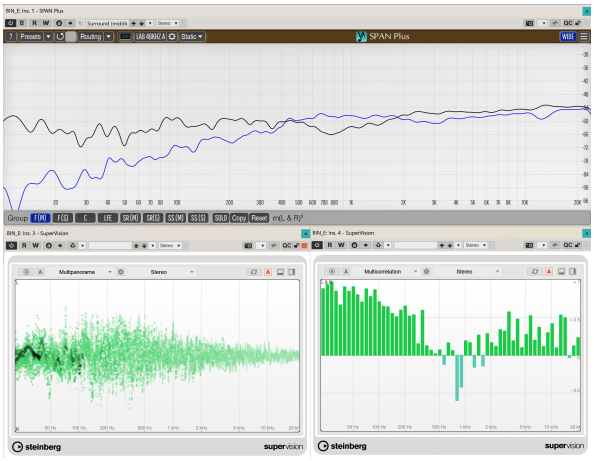


Figure 5. M/S spectrum analysis (top), multi-band panorama (bottom left) and multi-band correlation (bottom right) of de-correlated pink noise in a binaural render.

7 Conclusions

Headphone reproduction of stereo content is often exaggerated in the low frequencies, and the center image resides inside the head. Introducing cross-feed into a stereo signal provides good externalization during headphone reproduction (out of head experience) and a more realistic listening experience, closer to that of loudspeaker reproduction. Understanding the differences between stereo and binaural headphone listening is crucial when it comes to comparing the two formats side by side.

8 Acknowledgements

Thanks to the Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT), and Social Sciences and Humanities Research Council (SSHRC) for supporting this publication.

9 References

[1] Edison Research. 2025. The Infinite Dial 2025. Edison Research. Accessed September 8th, 2025. https://www.forbes.com/sites/bradadgate/2025/03/26/study-repots-listening-to-digital-audio-platforms-at-an-all-time-high.

[2] Y. Ueno, M. Mizumachi, & T. Horiuchi, "Perceptual evaluation of binaural rendering and stereo width control in headphone reproduction", Proc. 148th Convention of the Audio Eng. Soc., Online, June 4th (2020).

[3] D. Griesinger, "Accurate reproduction of binaural recordings through individual headphone equalization and time domain crosstalk cancellation", Proc. 147th Convention of the Audio Eng. Soc., New York, October (2019).

[4] E. Manor, W. Martens, A. Marui and D. Cabrera, "Nearfield Crosstalk Increases Listener Preferences for Headphone-Reproduced Stereophonic Imagery", J Audio Eng. Soc., Vol 63 No. 5, pp. 324 – 335, May (2015).

[5] B. Bauer, "Stereophonic Earphones and Binaural Loudspeakers", J Audio Eng. Soc., Vol 8 No. 2, pp. 148 – 151, April (1961).

[6] R. King, M. Theriault and G. Massenburg, "A Practical Approach to the Use of Center Channel in Immersive Music Production", Proc. 155th Convention of the Audio Eng. Soc., New York, October 26th (2023).

[7] Y. Ueno, M. Mizumachi, & T. Horiuchi, "Perceptual optimization of stereo width shrinkage method in headphone reproduction", Acoustical Science and Technology (2022), 43(2), 170–178.

[8] ITU-R BS.775-2, "Multichannel stereophonic sound system with and without accompanying picture," Rec., International Telecommunications Union, Geneva, Switzerland, 2010. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-2-200607-S!!PDF-E.pdf

[9] Steinberg Media Technologies GmbH, Hamburg, Germany, 2025. https://www.steinberg.net/nuendo/