Audio Engineering Society

Convention Express Paper 56

Presented at the 153rd Convention
2022 October

Binaural Externalization Processing - from Stereo to Object-Based Audio

Jean-Marc Jot[1], Alexey Lukin[2], and Christopher Landschoot[1]

[1] *Virtuel Works LLC, Aptos, CA 95003, United States of America*
[2] *iZotope, Inc., Boston, MA 02111, United States of America*

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

ABSTRACT

In both entertainment and professional applications, conventionally produced stereo or multi-channel audio content is frequently delivered over headphones or earbuds. Use cases involving object-based binaural audio rendering include recently developed immersive multi-channel audio distribution formats, along with the accelerating deployment of virtual or augmented reality applications and head-mounted displays. The appreciation of these listening experiences by end users may be compromised by an unnatural perception of the localization of frontal audio objects: commonly heard near or inside the listener's head even when their specified position is distant. This artifact may persist despite the provision of perceptual cues that have been known to partially mitigate it, including artificial acoustic reflections or reverberation, head-tracking, individualized HRTF processing, or reinforcing visual information. In this paper, we review previously reported methods for binaural audio externalization processing, and generalize a recently proposed approach to address object-based audio rendering.

1 Introduction

Headphone or earbud listening scenarios span from the home or office to mobile and automotive environments, with audio source content formats including two-channel stereo, multi-channel surround, and immersive or object-based music, movie or game soundtracks.

Virtual 3D audio processing techniques informed by spatial hearing models and based on the simulation of Head-Related Transfer Functions (HRTF) were developed with the intent of restoring, during headphone playback, the spatial audio cues experienced in natural or loudspeaker listening [3, 4, 5, 6, 7, 8, 9]. For audio-only content (e.g. music or podcasts), this may reduce listening fatigue and facilitate spatial discrimination. For audio-visual content (e.g. video or teleconference), it can further alleviate cognitive load by improving the spatial coincidence of auditory and visual cues.

Part of this work was carried out at iZotope, Inc. and previously presented in [1]. Additional demonstrations of recent extensions are provided in [2].

An aspiration of the binaural post processing of two-channel or multi-channel recordings is to simulate the auditory experience of attending a live performance, or of listening to a frontal stereo loudspeaker system. However, even with such processing engaged, subjective localization and preference studies have reported artifacts including [10, 11, 12, 13, 14]:

– In-head localization, spurious elevation or front-to-back confusion in the perceived localization of sound events, especially virtual sources whose due localization lies within the listener's field of view;
– Timbre coloration, often attributed in part to the addition of simulated reflections or reverberation processing intended to partially mitigate the above undesired artifacts.

Fig. 1 illustrates a commonly reported subjective experience during the binaural reproduction of a circular motion in the horizontal plane. As cited in [15], Jean Hiraga stated in 1997 *"... the most common case is to feel as though the source moves up as it passes in front..."* (*"... le cas le plus courant est d'avoir l'impression que la source monte en passant devant la tête..."*). This observation is also common when a binaural renderer is applied to a mono waveform, as e.g. in gaming engines or virtual reality audio systems [16].

Factors known to improve the externalization performance of binaural renderers include the simulation of virtual or local room acoustics, the dynamic compensation of the listener's head motion, the individualization of head-related and headphone-related transfer functions, and the provision of congruent visual information [4, 17, 18]. These techniques are not suitable or practical in all application scenarios, or may themselves cause undesirable audio artifacts. In the present paper, we examine an externalization processing approach compatible with these enhancements that may also partially alleviate their need in some use cases, and seeks more particularly to tackle frontal externalization.

In Section 3, we describe a binaural audio externalization processing scheme previously introduced in [1], applicable to the reproduction of conventionally produced stereo recordings. It is based on processing the source signal through a 2-channel quasi-all-pass filter that has the effect of adding a brief reverberation-like decay tail, aiming to restore at the ears of the listener the spatial hearing cues that enable the discrimination of direct vs. indirect sounds in natural listening conditions, while minimizing spectral coloration artifacts.

Fig. 1: Illustration of commonly perceived trajectory in headphone listening to a binaural recording of a sound moving around the listener's head in the horizontal plane (from [15]).

In Section 4, we propose extensions of this scheme and a computationally efficient processing topology for the binaural reproduction of multi-channel or object-based audio content. The resulting methods can be implemented in conjunction with head-tracking, individualized HRTF customization, and artificial or recorded acoustic reverberation. It is applicable to enhancing the decoding and headphone reproduction of linear immersive audio content formats such as Dolby Atmos and MPEG-H [19, 20], or the rendering of interactive audio content over head-worn binaural display devices for virtual or augmented reality applications [21].

2 Existing post-processing methods

Binaural processors for stereo reproduction over headphones commonly rely on the simulation of "virtual loudspeakers" [6, 7, 8, 9] often augmented with some of the following enhancements:

– The synthesis of artificial reverberation or reflections to simulate a natural listening experience [22, 23]
– Direct-diffuse decomposition to render reverberation or ambience components already present in the source material as diffuse sound components [22]
– Up-mixing techniques to mitigate the incorrect matching of natural HRTF cues for sources panned across two or more virtual loudspeakers [24, 25]
– Decorrelation techniques: mitigating localization and timbre preservation issues for center-panned sound components [26]
– Augmented Reality artificial reverberation: estimating and matching reverberation properties of the local environment [27, 28].

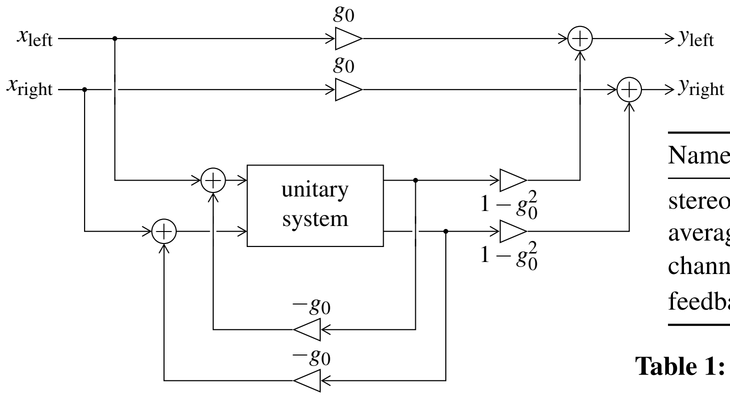

Fig. 2: Basic 2-in, 2-out allpass filter.

Name	Value	Unit
stereo cross-feed angle, θ	$\pi/4$	radians
average delay, $(m_1+m_0)/2$	2.943	ms
channel delay diff., $\frac{m_1-m_0}{m_1+m_0}$	28.74	%
feedback gain, g_0	0.7214	

Table 1: All-pass filter parameter settings used in [1].

3 A new approach to the externalization processing of stereo recordings

In [1], the authors describe a post-processing strategy that seeks to offer some of the advantages of previously reported methods while minimizing signal modification, prioritizing two main objectives:

(a) The preservation of the timbre of the original recording, through an algorithm design method founded on a 2-in, 2-out all-pass filter topology;

(b) The externalization of sound components, with particular attention to frontal and center-panned elements, through the selection of algorithm design variables and the insertion of minimal additional processing.

The design intent departs from the simulation of virtual loudspeakers in a virtual room. It concentrates on delivering binaural cues experienced consistently in everyday natural listening, in the form of spatial relations between direct and diffuse sound-field components. By adopting a minimal processing approach, we seek to avoid the drawbacks of some previously studied methods, including computational complexity and audible artifacts for some source material.

3.1 Basic timbre-preserving 2×2 system

We start with a general 2-channel all-pass filter prototype (Fig. 2) implementing the multichannel generalization of Schroeder's all-pass filter by Gerzon [29, 30]. If the block "unitary system" is energy-preserving, then the overall system will be as well: the power spectrum of the output signal pair (y_{left}, y_{right}) will be equal to the power spectrum of the input (x_{left}, x_{right}).

The stability condition is $|g_0| < 1$. For realizability, the 2-in, 2-out unitary system must be causal, with at least one-sample delay in both the left and right channels. In the topology proposed in [1], shown in Fig. 3, the internal unitary system is realized as the cascade of a pair of nonzero delay lines (z^{-m_0}, z^{-m_1}) with a 2×2 rotation matrix $\mathbf{R}(\theta)$, as defined in Eqn. (1) and summarized in Table 1.

$$\mathbf{R}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad , \quad \theta \in [0, \pi/4] \qquad (1)$$

where the extremal values represent no mixing and maximum mixing between the channels:

$$\mathbf{R}(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (2)$$

$$\mathbf{R}(\pi/4) = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \qquad (3)$$

3.2 Comparison to reverberation processing

The topologies of Fig. 2 and Fig. 3 may be compared to artificial reverberation processing, as proposed previously for the purpose of enhancing binaural audio externalization (e. g. [22, 23]). In this section, we highlight analogies and differences.

The block labeled "diffuse tail" in Fig. 3 is comparable to a 2-in, 2-out artificial reverberator realized by a 2-channel Unitary-Feedback Delay Network (UFDN) [31, 32], with design exceptions ensuring that the overall system (adding the reverberated and original signals) must be spectrally neutral - thus preventing potential timbral artifacts arising from simulated room reflections or reverberation, such as coloration or comb-filtering.

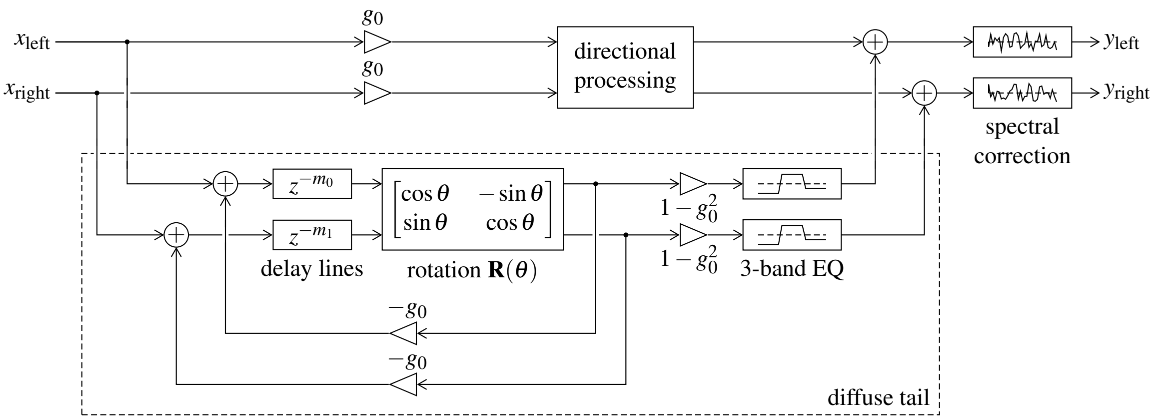

Fig. 3: Signal flow diagram of the externalization processing system originally proposed in [1].

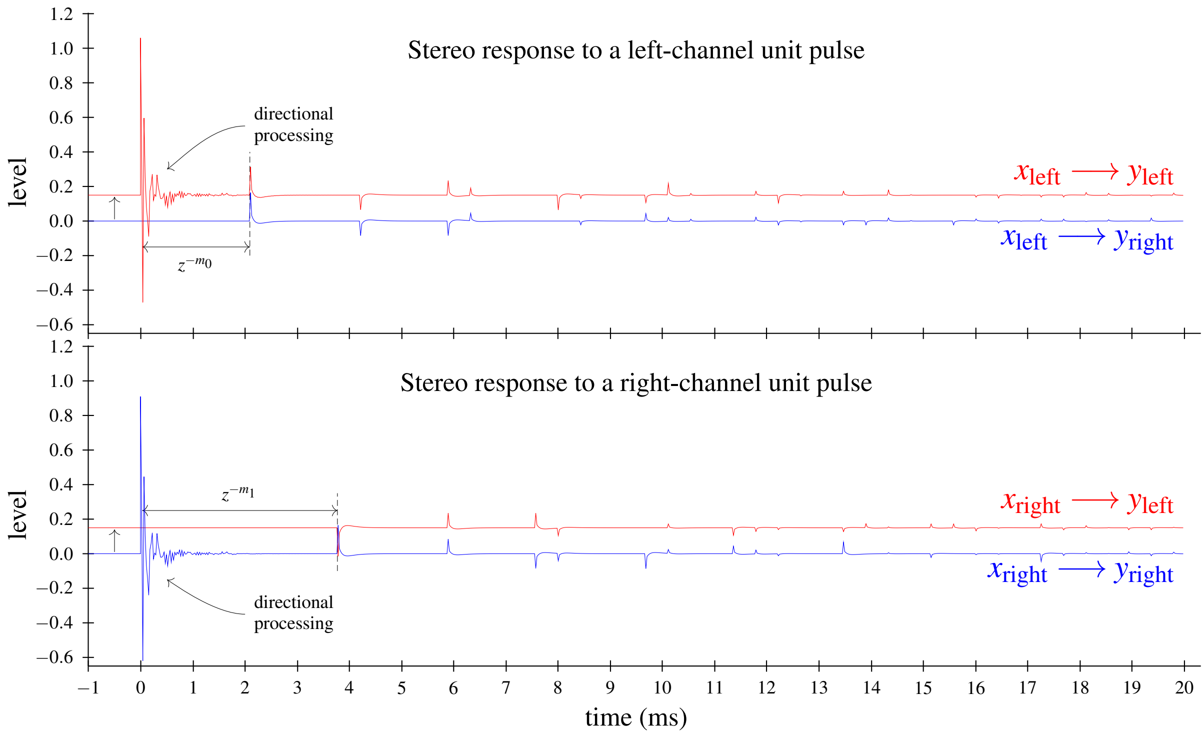

Fig. 4: Time-domain response of the 2-in, 2-out system of Fig.3 with basic parameters set per Table 1 and directional processing per Fig. 7 [1].

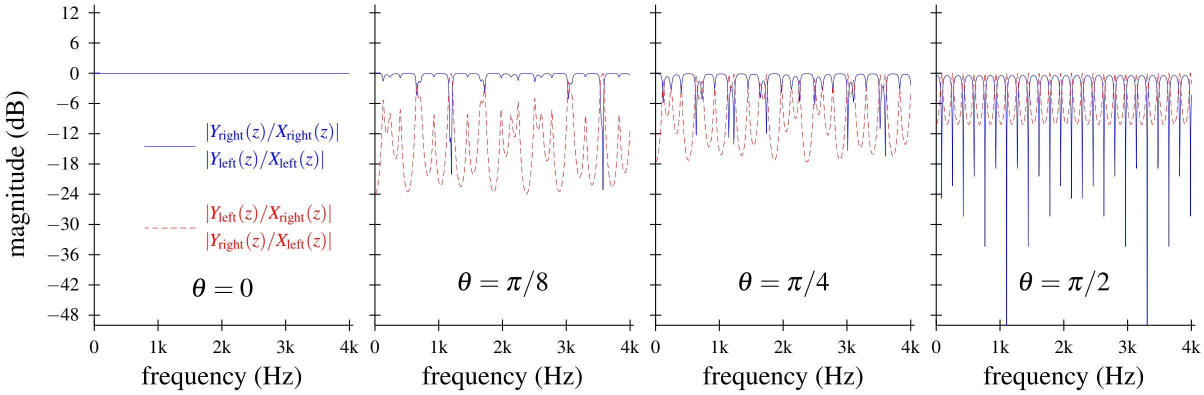

Fig. 5: Magnitude frequency response of the 2-in, 2-out system of Fig. 3 with directional processing, EQ and spectral correction blocks disabled, for several settings of the feedback cross-feed angle $\theta \in \{0, \pi/8, \pi/4, \pi/2\}$ (for compactness, only the frequency range 0–4 kHz is shown) [1].

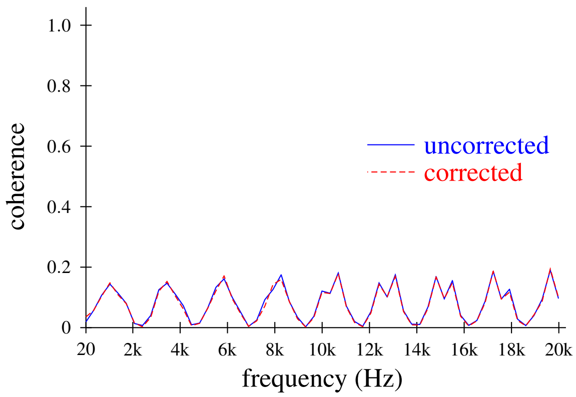

Fig. 6: Inter-channel coherence in the tail response when a unit impulse is input to both channels, with parameters set according to Table 1 – computed by the Welsh method using a length-128 Hann window length and 50% overlap [1].

In the system of Fig.3, the onset delay time of the "diffuse tail" (reverberation) and the diffuse-to-direct ratio (wet/dry mix) are imposed by the values of m_0, m_1 and g_0, which also jointly determine the decay rate in the time-domain response (Fig. 4). In [1], they were chosen as shown in Table 1, under the constraint of maintaining the effective duration of the overall impulse response within approximately 50 ms, in order to prevent audible spectral or temporal artifacts in the presence of percussive input signals.

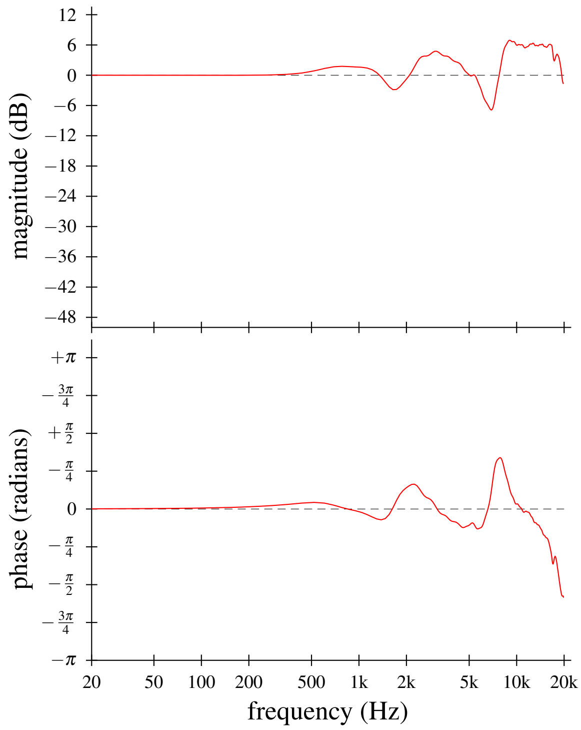

Fig. 7: Frequency response of the directional processing block employed in [1]: minimum-phase, diffuse-field compensated HRTF at 0° azimuth and 0° elevation, derived from [33] (subject 'KU 100'), with normalization to 0 dB gain a low frequencies.

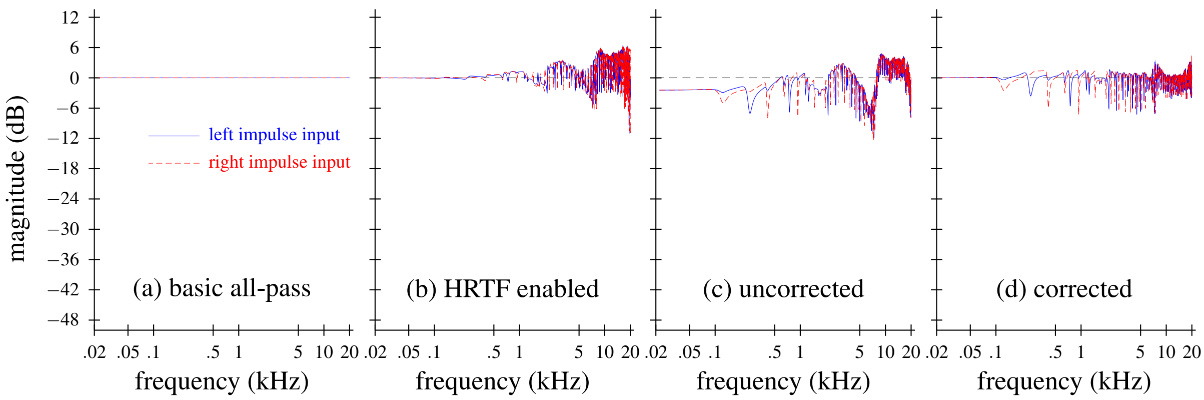

Fig. 8: Magnitude frequency response of the overall externalization processing system of Fig. 3, derived by power summation of the left and right outputs for a left-only or right-only input [1] – (a) overall spectral conservation by the basic 2×2 system; (b) enabling the directional processing function (Fig. 7); (c) enabling the 3-band EQ function; (d) enabling the overall spectral corrector.

The setting of angle θ determines the amount of cross-feed from left input to right output (and vice-versa) in the diffuse tail processing block. If $\theta = 0$, it is equivalent to a parallel pair of independent single-channel Schroeder all-pass filters [29], with no cross-feed between them. When $\theta = \pi/4$, the cross feed is maximum. Fig. 6 shows that diffuse tail processing outputs a signal having low inter-channel correlation, as needed for the binaural simulation of a diffuse sound field [34].

Fig. 5 shows that, when $\theta > 0$, the individual matrix transfer function entries of the overall system are no longer all-pass (even though the overall 2×2 system is, by construction): the red dotted line which displays the overall magnitude frequency response from left input to right output (and vice-versa) exhibits spectral dips which are made up for in the transfer function to the opposite ear.

3.3 Directional and spectral corrections

In [1], the "directional processing" block (see Fig. 3) applies an identical filter to the left and right direct-path channels. It has a minimum-phase transfer function derived from a diffuse-field compensated HRTF measurement provided in the SADIE II database by York University [33], taken on the Neumann KU 100 dummy head for the frontal direction of sound incidence (0° azimuth and 0° elevation).

In this way, the overall system of Fig. 3 provides a frontal localization emphasis via its direct-path output, presenting a natural spectral difference against the diffuse tail. Additionally, a gain correction is applied to ensure that the directional processing filter is neutral in the low-frequency range (see Fig. 7). Therefore, the all-pass behavior of the overall 2×2 system is exactly preserved at low frequencies. That is confirmed in Fig. 8(b), which plots the power frequency spectrum of the overall 2-channel output signal when an impulse is fed to the left or right input. Fig. 8(b) also shows that, at frequencies above approximately 300 Hz, the overall 2×2 system is no longer strictly all-pass, as a result of inserting the directional processing filter.

An output spectral correction filter, shown in Fig. 3, may be engaged in order to compensate for the effects of directional processing and 3-band EQ (or "tail EQ") on overall tonal preservation. The 3-band EQ filter is a second-order dual-shelving filter [35] applied to the diffuse tail output. In [1], it was tuned by ear as a cursory modification to reduce the overall left-to-right and right-to-left cross-feed at high frequencies, and to prevent excessive decorrelation in the low frequency range. As seen in Fig. 8(c), this modification had the undesirable effect of causing an audible decrease in overall loudness at low frequencies, which made the output spectral correction all the more necessary to preserve tonal balance (as shown in Fig. 8(d)).

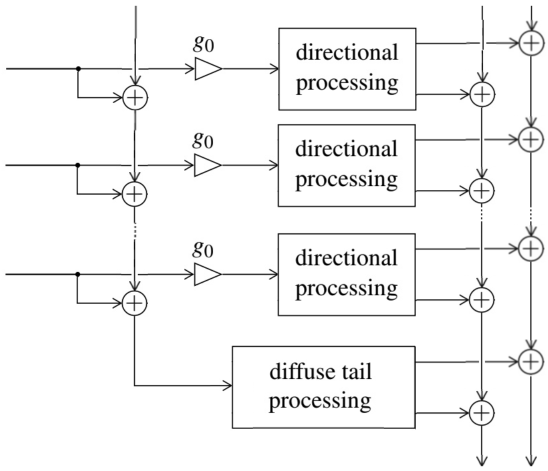

Fig. 9: Binaural externalization processing of multiple 3D audio objects or channels with shared diffuse tail processing module.

4 Generalization: externalized binaural rendering of multiple 3D audio objects or channels

In [1] and Section 3.3, the directional processing block of Fig. 3 applies the same HRTF filter to all sound elements regardless of their individual panning positions in the input mix. However, more generally, it may incorporate any binaural rendering method, including those reviewed in Section 2 – for instance: loudspeaker virtualization processing combined with up-mixing techniques such that the direct-path positional emphasis cues convey more faithfully the respective localization of the different elements in the mix (given by the left-to-right "panpot" setting of each source track in a stereo production), as proposed in [22, 24].

Fig. 9 represents a scenario where the input audio content is a collection of 3D audio objects individually panned to different positions. Each object is provided as a mono waveform signal assigned a localization (azimuth and elevation). Collectively, the set of audio objects may constitute an immersive multichannel signal wherein each audio input channel is assigned a fixed position on a virtual sphere centered on the listener, referenced to the "front center" direction.

4.1 HRTF normalization and individualization

Each directional processing module in Fig. 9 outputs a dry binaural signal, by simulating the pair of HRTF filters for the direction assigned to its corresponding input object or channel. As in Section 3.3 and Fig. 7, these HRTF filters are diffuse-field compensated minimum-phase filters, normalized to 0 dB amplitude gain at low frequencies. The discussions and results reported in Section 3, when applied to a center-panned ("dual-mono") input, apply identically here in the case of an audio object panned to the front-center direction.

By employing diffuse-field compensated HRTF filters, we ensure that setting one of the directional processing modules to simulate a different position in 3D space does not require modifying the spectral equalization of the diffuse tail, whose computation can therefore be shared among all objects (as shown in Fig. 9). For the same reason, diffuse tail processing is not affected by HRTF individualization (customization of the directional processing to account for HRTF data representative of a different listener or head morphology). By normalizing all HRTF filters to 0 dB gain at low frequencies, we simulate far-field sound incidence and ensure that the effect of the directional processing module at low frequencies is only an inter-channel time difference – the Interaural Time Delay (ITD), which varies according to the specified 3D audio object panning position and vanishes if that position is set to front center (or in the median plane).

4.2 Diffuse tail processing

In the design discussed in Section 3.2, the diffuse tail processing block is comparable to a 2-channel Unitary-Feedback Delay Network (UFDN) [31], where the gain g_0 and the delay lengths m_0 and m_1 are tuned to produce a very brief diffuse impulse response while achieving a chosen direct-to-diffuse power ratio.

More generally, the design requirement for the diffuse tail processing module in the externalizers of Fig. 3 or Fig. 9 is to subject any input signal through a 2-channel quasi-all-pass filter that preserves the temporal locality of transients and adds a binaural response tail replicating the interaural coherence properties of natural diffuse fields [34], while the direct path is subject to binaural processing that conveys the primary localization cues for each object. In doing so, we aim to restore at the ears of the listener the spatial hearing cues that enable the discrimination of direct vs. indirect sounds in natural listening conditions.

4.3 Evolutionary interpretation

We hypothesize that the binaural externalization processing method proposed here exploits an evolutionary developmental ability of humans, acquired from prior listening experiences involving the spatial audio perception of concurrent free-field and diffuse-field audio components emanating from the same sound source.

Indeed, our hypothesis is that objective triggers to the percept of externalization that are robust across listening conditions rely on the discrimination of direct vs. diffuse sound field components, enabled by interaural coherence and spectral differences that are intrinsic to the listener's bilateral HRTF properties but do not depend on listening room acoustics. This inspires a binaural externalization processing strategy that prevents timbral coloration of the source material by de-emphasizing room-dependent timbre coloration cues, without substantially undermining the perception of externalization in comparison to methods employing artificial room reverberation or reflections.

5 Demonstrations

In [1], the authors include audio examples illustrating the effect of the binaural externalization processing scheme described in Section 3, for the reproduction of conventional 2-channel stereo recordings (including both full mix and soloed center-panned vocals). In addition to the unprocessed original, the following examples are provided for comparison:

- directional processing without diffuse tail;
- directional processing with externalization processing, with or without overall spectral correction.

When listening to these examples, we suggest directing one's attention to the following aspects of the listening experience (this may be facilitated by closing one's eyes or listening in dark conditions, in order to avoid visual distractions). Do the overall width and timbre seem different from one example to another? Does the localization of the voice seem to change from one example to another? Does it appear to be localized inside the head, or rather in front of the listener?

In [2], we provide examples illustrating more specifically the processing of object-based source content per Section 4, including the case of an individual audio object assigned a moving position on a horizontal trajectory around the listener, as illustrated in Fig. 1.

6 Perspectives and applications

In this paper, we reported some preliminary results of a study that attempts to help address a persistent challenge commonly encountered in the design and application of binaural technologies: the difficulty of obtaining a natural-sounding externalization of headphone audio images and of reproducing the due localization of frontal sounds, without introducing detrimental timbral coloration or other audio artifacts.

Opportunities for further exploration of the algorithm's design abound – including, e. g.:

- Exposing a sound-designer friendly high-level control interface for the fine tuning of internal low-level parameters such as m_0, m_1 and g_0 defined in Section 3. (Note that, as visible in Figures 3 and 9, the application of the attenuation factor g_0 to direct-path components is necessary for loudness preservation when externalization processing is applied by enabling the "diffuse tail processing" function described here.)
- The use of Velvet Noise decorrelators [36] or alternative IIR network designs such as nested all-pass filters and feedback delay networks [37, 38] or time-varying all-pass networks [32].
- Diffuse tail processing algorithm designs that preserve mono compatibility or employ alternative approaches to control low-frequency interaural coherence in order to match the natural properties of diffuse sound fields (see Fig. 10 and [34]).

Possible future extensions and applications of the present study include:

- Psychophysical investigations of the perception of concurrent free-field and diffuse-field components of a sound event (see also Section 4.3).
- Producing 2-channel stereo recordings that feature externalization enhancements beneficial for headphone playback, but remain compatible with loudspeaker playback.
- Producing binaural recordings that leverage familiar 2-channel stereo production techniques (such as stereo flanger or chorus effects), or apply externalization processing to selected tracks or stems in a stereo mix (e.g. a vocal track).
- Virtual meeting and augmented or virtual reality experiences [21], combining the proposed externalization processing scheme with head tracking and HRTF individualization.

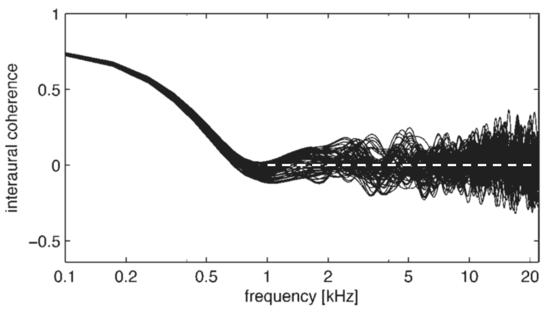

Fig. 10: Measured frequency-dependent interaural coherence in diffuse reverberation [34] and substitution above 1kHz.

7 Acknowledgments

The authors would like to thank Kurt Werner for his contributions in initial algorithm design discussions and for producing the diagrams and plots included in [1] and reproduced in the present paper. We are grateful to Matt Carlson and Evan Allen for procuring and selecting example source material, and for producing the audio demonstration files presented in [1] (accessible in [2]). We salute the wizardry of Roth Michaels in encapsulating the initial prototype implementation into a Max/MSP module, thus facilitating subsequent evaluation.

References

[1] Jot, J.-M., Lukin, A., Werner, K. J., and Allen, E., "Binaural Audio Externalization Processing," in *151st conv. Audio Eng. Soc.*, 2021.

[2] Landschoot, C. and Jot, J.-M., "Binaural Externalization Processing Demonstration," 2022, <https://bit.ly/3RtKJFl>.

[3] Blauert, J., *Spatial Hearing*, MIT Press, Cambridge, MA, 1977.

[4] Best, V. et al., "Sound Externalization: A Review of Recent Research," *Trends in Hearing*, 24, pp. 1–14, 2020.

[5] Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., W., and M., E., "On the externalization of auditory images," *Presence: Teleoperators & Virtual Environments*, 1(2), p. 251–257, 1992.

[6] Rubak, P., "Headphone Signal Processing System for Out-of-the-head Localization," in *90th conv. Audio Eng. Soc.*, 1991.

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

[8] Kirkeby, O., "A Balanced Stereo Widening Network for Headphones," in *Proc. 22nd Conf. Audio Eng. Soc.*, 2002.

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

[10] Mason, R. and Rumsey, F., "An Assessment of the Spatial Performance of Virtual Home Theatre Algorithms by Subjective and Objective Methods," in *108th conv. Audio Eng. Soc.*, 2001.

[11] Olive, S., "Evaluation of Five Commercial Stereo Enhancement 3D Audio Software Plug-ins," in *110th conv. Audio Eng. Soc.*, 2001.

[12] Lorho, G. et al., "Round Robin Subjective Evaluation of Stereo Enhancement Systems for Headphones," in *Proc. 22nd Conf. Audio Eng. Soc.*, 2002.

[13] Lorho, G. and Zacharov, N., "Subjective Evaluation of Virtual Home Theatre Sound Systems for Loudspeakers and Headphones," in *116th conv. Audio Eng. Soc.*, 2004.

[14] Lorho, G., "Evaluation of Spatial Enhancement Systems for Stereo Headphone Reproduction by Preference and Attribute Rating," in *118th conv. Audio Eng. Soc.*, 2005.

[15] Lagnel, B., "Internalisation d'une Voix," *Le Son Binaural*, 2022, (course notes) <https://www.lesonbinaural.fr>.

[16] Reardon, G., Zalles, G., Genovese, A., Flanagan, P., and Roginska, A., "Evaluation of Binaural Renderers: Externalization, Front/Back and Up/Down Confusions," in *144th conv. Audio Eng. Soc.*, 2018.

[17] Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K., "A Summary on Acoustic Room Divergence and its Effect on Externalization of Auditory Events," in *Proc. IEEE 8th Int. Conf. Quality of Multimedia Experience (QoMEX)*, 2016.

[18] Griesinger, D., "Accurate Timbre and Frontal Localization Without Head Tracking Through Individual Eardrum Equalization of Headphones," in *121st conv. Audio Eng. Soc.*, 2016.

[19] Robinson, C. Q., Mehta, S., and Tsingos, N., "Scalable Format and Tools to Extend the Possibilities of Cinema Audio," *SMPTE Motion Imaging Journal*, 2012.

[20] Herre, J., Hilpert, J., Kuntz, A., and Plogsties, J., "MPEG-H Audio - The New Standard for Universal Spatial / 3D Audio Coding," in *137th conv. Audio Eng. Soc.*, 2014.

[21] Jot, J.-M. et al., "Rendering Spatial Sound for Interoperable Experiences in the Audio Metaverse," in *Proc. Int. Conf. Immersive and 3D Audio (I3DA)*, 2021, <https://arxiv.org/pdf/2109.12471.pdf>.

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

[23] Davidson, G. et al., "Design and Subjective Evaluation of a Perceptually-Optimized Headphone Virtualizer," in *140th Conv. Audio Eng. Soc.*, 2016.

[24] Goodwin, M. M. and Jot, J.-M., "Binaural 3-D Audio Rendering Based on Spatial Audio Scene Coding," in *123rd Conv. Audio Eng. Soc.*, 2007.

[25] Breebaart, J. and Schuijers, E., "Phantom Materialization: a Novel Method to Enhance Stereo Audio Reproduction on Headphones," *IEEE Trans. Audio, Speech and Lang. Proc.*, 2008.

[26] Jot, J.-M. and Walsh, M., "Center-Channel Processing In Virtual 3-D Audio Reproduction over Headphones or Loudspeakers," in *128th conv. Audio Eng. Soc.*, 2010.

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

[28] Murgai, P., Rau, M., and Jot, J.-M., "Blind Estimation of the Reverberation Fingerprint of Unknown Acoustic Environments," in *143rd conv. Audio Eng. Soc.*, 2017.

[29] Schroeder, M. R. and Logan, B. F., " 'Colorless' Artificial Reverberation," *IRE Trans. Audio*, 9(6), pp. 209–214, 1961.

[30] Gerzon, M. A., "Unitary (Energy-Preserving) Multichannel Networks with Feedback," *Electronics Letters*, 12(11), pp. 278–279, 1976.

[31] Dahl, L. and Jot, J.-M., "A Reverberator Based on Absorbent All-Pass Filters," in *Proc. Int. Conf. Digital Audio Effects*, 2000.

[32] Werner, K. J., Germain, F. G., and Goldsmith, C. S., "Energy-Preserving Time-Varying Schroeder Allpass Filters and Multichannel Extensions," *J. Audio Eng. Soc.*, 69(7/8), pp. 465–485, 2021.

[33] Armstrong, C., Thresh, L., Murphy, D., and Kearney, G., "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Applied Sciences*, 2018.

[34] Menzer, F. and Faller, C., "Investigations on an Early-Reflection-Free Model for BRIRs," *J. Audio Eng. Soc*, 2010.

[35] Jot, J.-M., "Proportional Parametric Equalizers — Application to Digital Reverberation and Environmental Audio Processing," in *139th conv. Audio Eng. Soc.*, 2015.

[36] Alary, B., Politis, A., and Välimäki, V., "Velvet-Noise Decorrelator," in *Proc. 20th Int. Conf. Digital Audio Effects*, 2017.

[37] Jot, J.-M., "Efficient Models for Reverberation and Distance Rendering in Computer Music and Virtual Audio Reality," in *Proc. Int. Computer Music Conf.*, 1997.

[38] Gardner, W. G., "Reverberation Algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, Springer, 2002.