RENDERING OF SOURCE SPREAD FOR ARBITRARY PLAYBACK SETUPS BASED ON SPATIAL COVARIANCE MATCHING

Leo McCormack,[1*] *Archontis Politis,*[2*] *Ville Pulkki*[1]

[1]Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland
[2]Faculty of Information Technology and Communication Sciences, Tampere University, Finland
leo.mccormack@aalto.fi

ABSTRACT

This paper proposes an algorithm for rendering spread sound sources, which are mutually incoherent across their extents, over arbitrary playback formats. The approach involves first generating signals corresponding to the centre of the spread source for the intended playback setup, along with decorrelated variants, followed by defining a diffuse spatial covariance matrix for the confined target spreading area. The mixing matrices required to combine these signals, in a manner whereby the resulting output signals exhibit the target inter-channel relationships for an incoherently spread source, are computed based on an optimised solution which is constrained to preserve signal fidelity. The proposed solution is evaluated in the context of producing extended sound sources for binaural playback. Objective perceptual metrics are computed and shown to be comparable to those derived from an ideal incoherently spread reference. Signal distortion measures are also calculated for speech, musical, and ambience recordings, which indicate higher signal fidelity produced by the proposed constrained spatial covariance matching solution, compared to an unconstrained alternative. These improvements in signal fidelity are further demonstrated by the provided audio examples and open-source audio plug-in.

Index Terms— sound source spreading, spatial audio

1. INTRODUCTION

The ability to create sound sources of arbitrary physical extent is useful for a number of applications; including: immersive content production, cinema audio, and generating sound objects within virtual and augmented reality environments. A trivial solution for this task involves creating several coherent copies of the input signal, and assigning them in directions surrounding the centre of the sound object. This approach has been employed for stabilising the inherent direction-dependent source spread during amplitude panning [1, 2], but has also been used by certain software tools [3, 4] to target the aforementioned applications. Coherent spreading is also how Ambisonics panning fundamentally operates [5, 6]. Incoherent source spreading, on the other hand, is when sound is instead reproduced over the spreading area in a diffuse manner, based on the creation of mutually incoherent copies of the input signal through the application of decorrelators [7, 8]. Note that the intention of this decorrelation is to generate spectrally identical signals, but with incoherent phase responses. Since coherently spread sources have the tendency to collapse into narrower and directionally more ambiguous auditory events [9], and exhibit timbral colourations similar to those associated with lower-order Ambisonics reproduction [10], incoherent source spreading is generally the more favoured approach [11, 12, 13]. However, in practice, as the number of channels increases, solutions for incoherent spreading can become increasingly more challenging; since signal fidelity degradations can occur and be aggregated during the necessary decorrelation operations.

Rather than operating based on coherent or incoherent replicas of the input signal, one popular alternative approach is to instead divide the input signal into frequency bands and then spatialise each band in different directions surrounding the target spreading area [14, 15, 16, 17]. Such an approach has been shown to yield perceptually plausible source spreading in [15, 18], has been employed for spatial upmixing purposes in [19], and applied within the Ambisonics framework using time-varying encoding directions in [20]. The performance of these frequency-dependent spreading approaches does, however, vary depending on the spectral content of the input signal. For example, in the extreme case of a single sine tone input, the spreading function would instead correspond to a directional shift. In [21] the approach was also found to produce stimuli-dependent localisation shifts for musical input, when the spreading was applied over third-octave bands. The frequency resolution of the employed transform and the chosen spreading assignment (random or deterministic) are also cited as important design factors affecting the performance of the approach in [15]. Naturally, such an approach is also not applicable or easily augmented for the task of delivering frequency-dependent source spread.

Considering the above overview, the requirements for an ideal source spreading algorithm may be outlined as follows: to produce incoherent spreading of the input signal, in order to mitigate timbral colourations and localisation ambiguity; employ as few decorrelation operations as possible, or otherwise optimise the algorithm to preserve the signal fidelity of the original signal; involve minimal parameter tuning; and provide support for frequency-dependent spreading. In this paper, a spreading algorithm is formulated based on the covariance domain framework established in [22], which aims to fulfil these requirements. The algorithm may be configured for any output format; such as: head-related transfer functions (HRTFs), microphone array transfer functions, vector-base amplitude panning (VBAP) gains, or spherical harmonic (SH) vectors. The proposed algorithm is evaluated in two parts, in the context of binaural reproduction using HRTFs. The first evaluation involves a comparison of the binaural colouration, interaural level and coherence difference cues, to those provided by an ideal incoherently spread binaural reference. The second part provides insight regarding the signal fidelity of the output signals. An open-source audio-plugin implementation of the proposal is also provided on the companion web-page[1], along with exemplary pre-rendered examples.

[*] Equally contributing authors in this paper.

[1]http://research.spa.aalto.fi/publications/papers/waspaa21-spread/

2. SPREAD SOURCE MODEL

Consider a multi-channel rendering system of Q channels, which, in the general case, can be characterised either by its directional rendering functions (for example, HRTFs for headphone rendering, or VBAP gain vectors for loudspeaker rendering), or by its encoding/capturing functions (for example, SH vectors for Ambisonics encoding, or array responses for surround recording). The directional responses for such systems depend on the direction of arrival $\mathbf{u}$, and potentially also on frequency f, and may be denoted as $\mathbf{h}(f, \mathbf{u}) = [h_1(f, \mathbf{u}), ..., h_Q(f, \mathbf{u})]^{\mathrm{T}}$. Note that such responses may be determined analytically (for example, VBAP gains or SH vectors), modelled numerically (such as solid-sphere HRTFs or geometrical models of the recording setup), or simply measured for a dense grid of K directions around the rendering system. In all cases, the responses may be combined into a directional response matrix $\mathbf{H}(f) = [\mathbf{h}(f, \mathbf{u}_1), ..., \mathbf{h}(f, \mathbf{u}_K)]$.

It is assumed that the setup will render signals $\mathbf{d}(f) = [d_1(f), ..., d_Q(f)]^{\mathrm{T}}$ corresponding to a spread source that is diffuse (i.e. mutually incoherent) over its full extent. The signals corresponding to this spread source may be modelled as

$$\mathbf{d}(f) = \int_{\mathbf{u}\in\mathcal{A}} \mathbf{h}(f, \mathbf{u}) s(f, \mathbf{u}) \mathrm{d}\mathbf{u}, \tag{1}$$

where $s(f, \mathbf{u})$ is the source signal carried by the incident wave from direction $\mathbf{u}$, $\mathrm{d}\mathbf{u} = \cos\theta \mathrm{d}\theta\mathrm{d}\phi$ is the differential surface element on the unit sphere with elevation and azimuth angles (θ, ϕ), and the integration is conducted over all directions within the spreading area $\mathcal{A} \in \mathcal{S}^2$. Since fully incoherent incidence from the source is assumed for each direction, the following assumptions also hold

$$\mathbb{E}[s(f, \mathbf{u}_i) s^*(f, \mathbf{u}_j)] = \begin{cases} 0, & \mathbf{u}_i \neq \mathbf{u}_j \\ P_s(f, \mathbf{u}_i) & \mathbf{u}_i = \mathbf{u}_j, \end{cases} \tag{2}$$

where $P_s(f, \mathbf{u})$ is the incident power for direction $\mathbf{u}$ and $\mathbb{E}[\cdot]$ denotes the expectation operator. The directional power profile may, for example, be derived based on geometrical considerations between the spread source and the receiver. However, in the most common and practical case, where no such information exists, an equal mean power for all directions $P_s(f, \mathbf{u}) = P_s(f)$ can be assumed.

Based on the above, the spatial covariance matrix (SCM) of the rendered spread source signals is given as

$$\begin{aligned} \mathbb{E}[\mathbf{d}(f)\mathbf{d}^{\mathrm{H}}(f)] &= P_s(f) \int_{\mathbf{u}\in\mathcal{A}} \mathbf{h}(f, \mathbf{u})\mathbf{h}^{\mathrm{H}}(f, \mathbf{u}) \mathrm{d}\mathbf{u}, \\ &= P_s(f)\mathbf{D}(f), \end{aligned} \tag{3}$$

where $\mathbf{D}(f)$ refers to a spread diffuse coherence matrix (SDCM) for the renderer. This SDCM can be computed analytically for certain array configurations, or approximated numerically from models or measurements [23]

$$\begin{aligned} \mathbf{D}(f) &= \int_{\mathbf{u}\in\mathcal{A}} \mathbf{h}(f, \mathbf{u})\mathbf{h}^{\mathrm{H}}(f, \mathbf{u}) \mathrm{d}\mathbf{u}, \\ &\approx \sum_{k|\mathbf{u}_k\in\mathcal{A}} w_k \mathbf{h}(f, \mathbf{u}_k)\mathbf{h}^{\mathrm{H}}(f, \mathbf{u}_k), \\ &\approx \mathbf{H}(f)\mathbf{W}\mathbf{H}^{\mathrm{H}}(f), \end{aligned} \tag{4}$$

where w_k are integration weights if the modelling/measurement points are not uniformly distributed over the sphere (otherwise $w_k = 1/K$), and $\mathbf{W}$ is a diagonal matrix constructed from them. Note that the formulation of the SDCM in (4) can readily accommodate more general directional power distributions, as in (2), using $\mathbf{D}(f) = \mathbf{H}(f)\mathbf{W}\mathbf{P}\mathbf{H}^{\mathrm{H}}(f)$; where $\mathbf{P}$ is a $K \times K$ diagonal matrix of power distribution values for the grid points, which are normalised so that $\mathrm{tr}[\mathbf{P}] = 1$.

3. SOURCE SPREADING

3.1. Coherent spreading

This straightforward approach serves as the baseline for this study, since it is often employed in software solutions [3, 4]; owing to its simplicity and computational efficiency. Coherent spreading is based on the principle that the input source signal may be replicated and reproduced over many directions surrounding the intended spreading area. By employing the appropriate directional responses, the intention is that this combination of the respective interaural cues should generate a perception of source width. This principle has also been used to stabilise the apparent shifts in source spreading depending on direction; for example, for spatialisation via amplitude panning [1, 2] or Ambisonics [6]. It can be formulated as

$$\mathbf{d}_{\mathrm{coh}}(f) = s(f) \sum_{k|\mathbf{u}_k\in\mathcal{A}} w_k \mathbf{h}(f, \mathbf{u}_k) = s(f)\mathbf{h}_{\mathrm{coh}}. \tag{5}$$

Note that coherent spreading does not involve any signal decorrelation, but instead simply averages the same signal spatialised over multiple directions. It preserves high signal fidelity, since artefacts commonly associated with decorrelation are avoided. On the other hand, coherently spread sources can have the tendency of collapsing into narrower auditory events, due to the summing localisation phenomenon [9], and is otherwise considered to be perceptually less convincing than incoherent spreading [11, 12, 13].

3.2. Direct incoherent spreading

The simplest approach to achieving an incoherent source spread, which closely follows the assumed model, is to decorrelate and spatialise multiple copies of the input signal for all available directional responses within the spreading area. Decorrelation approaches suitable for this task include short FIR filters of white noise sequences, sparse noise sequences [24], or time-frequency domain operations; such as: phase randomisation [7], sub-band delays [25], or networks of all-pass filters [26, 11]. If all K points are employed for spreading, then $K - 1$ decorrelated copies of the original source signal $s(n)$ are obtained $\mathbf{s}_{\mathrm{dec}}(f) = [s(f), s_{\mathrm{dec},1}(f), ..., s_{\mathrm{dec},K-1}(f)]^{\mathrm{T}}$, with an approximate SCM of $\mathbb{E}[\mathbf{s}_{\mathrm{dec}}(f)\mathbf{s}^{\mathrm{H}}_{\mathrm{dec}}(f)] \approx P_s(f)\mathbf{I}$. This direct incoherent spreading approach is conducted based on the spatialisation of all $\mathbf{s}_{\mathrm{dec}}$ signals as

$$\mathbf{d}_{\mathrm{inc}}(f) = \mathbf{H}(f)\mathbf{W}^{1/2}\mathbf{s}_{\mathrm{dec}}(f). \tag{6}$$

However, while this direct solution is theoretically correct and generates the appropriate inter-channel properties of $\mathbf{D}$, the excessive decorrelation required to create large extended sources can impair the signal quality in practice. Therefore, available incoherent spreading solutions typically employ and distribute far fewer points in the spreading area to mitigate such decorrelation artefacts [8].

3.3. Incoherent spreading through SCM matching

Ideally, no more than Q decorrelators should be employed for Q channels in the playback system. To demonstrate how that may be achieved, the spreading operation is posed as an optimal mixing problem; i.e., determining the mixing matrix $\mathbf{M}$ to apply to Q uncorrelated signals, in order to generate output signals that exhibit the appropriate inter-channel relationships defined by the SDCM

$$\mathbf{d}_{\mathrm{scm}}(f) = \mathbf{M}(f)\mathbf{s}_{\mathrm{dec}}(f), \quad \text{with} \tag{7}$$

$$\mathbf{M}(f)\mathbf{M}^{\mathrm{H}}(f) = \mathbf{D}(f). \tag{8}$$

In the specific case of binaural rendering (employing HRTFs as directional responses), the above mixing would directly generate the binaural cues corresponding to the assumed incoherent spread source model. The general solution to this mixing problem is given by any appropriate decomposition of the Hermitian matrix $\mathbf{D} = \mathbf{M}\mathbf{M}^{\mathrm{H}}$. Among other options, a straightforward solution may be derived based on the eigenvalue decomposition (EVD) of $\mathbf{D} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{H}}$, resulting in the mixing solution fulfilling (8) as

$$\mathbf{M}(f) = \mathbf{E}(f)\mathbf{\Lambda}^{1/2}(f). \tag{9}$$

4. PROPOSED INCOHERENT SPREADING THROUGH CONSTRAINED SCM MATCHING

The SCM matching solution of Sec. 3.3 generates signals with the correct target inter-channel relations, but not necessarily with any across-frequency consistency. The resulting multichannel spreading filters in $\mathbf{M}(f)$, combined with the decorrelated signals, may result in temporal artefacts and thus reduce the signal fidelity and the perceived sound quality of the spreader. On the other hand, it is apparent that there is an infinite number of solutions fulfilling (8), since $\mathbf{M}\mathbf{Q}\mathbf{Q}^{\mathrm{H}}\mathbf{M}^{\mathrm{H}} = \mathbf{D}$ holds true for any arbitrary unitary matrix $\mathbf{Q}$. Therefore, additional degrees of freedom are available, which may be employed to minimise such artefacts and improve the signal fidelity. To that end, the optimal upmixing framework described in [22], which has been used previously by the authors in [27, 28, 29], is employed. Incoherently spreading a mono signal is therefore realised as an upmixing optimisation task.

The proposed method involves first generating prototype multichannel signals, which only partially exhibit the required inter-channel relationships, but nevertheless have high signal fidelity. Such prototypes are usually derived based on a linear time-invariant mixing process, without signal decorrelation. A suitable candidate for the present scenario is $\mathbf{d}_{\mathrm{cen}}(f) = \mathbf{h}_{\mathrm{cen}}(f)s(f)$, where $\mathbf{h}_{\mathrm{cen}}(f)$ corresponds to the central spreading direction. In the second stage, these prototype signal are enhanced by applying a mixing matrix $\mathbf{M}$ such that the resulting signals match the target SDCM. Since the target SDCM is generally not fully reached by linearly mixing the prototype spread signals, some decorrelated signal energy is also introduced via a secondary mixing matrix $\mathbf{M}_{\mathrm{dec}}$ to fulfil the remaining target inter-channel relationships. Contrary to the solution of Sec. 3.3, which mixes fully decorrelated signals, the proposed solution here only introduces the minimum amount of decorrelation energy that is needed; therefore, decorrelation artefacts are minimised in the output. Additionally, the mixing matrix is fully optimised to both achieve the target SDCM and minimise signal distortion.

The optimisation process can be summarised as [22]

$$\mathbf{d}_{\mathrm{opt}}(t,f) = \mathbf{M}(t,f)\mathbf{d}_{\mathrm{cen}}(t,f) + \mathbf{M}_{\mathrm{dec}}(t,f)\mathcal{D}[\mathbf{d}_{\mathrm{cen}}(t,f)], \tag{10}$$

where $\mathcal{D}[\cdot]$ denotes a decorrelation operation on the enclosed signals. Omitting time and frequency indices, and denoting e.g. $\tilde{\mathbf{A}}$ as a diagonal matrix containing the diagonal entries of matrix $\mathbf{A}$, the following quantities are defined: $\mathbf{C}_{\mathrm{inc}} = P_s\mathbf{D}$ is the target spread source SCM; $\mathbf{C}_{\mathrm{cen}} = \mathbb{E}[\mathbf{d}_{\mathrm{cen}}\mathbf{d}_{\mathrm{cen}}^{\mathrm{H}}] = P_s\mathbf{h}_{\mathrm{cen}}\mathbf{h}_{\mathrm{cen}}^{\mathrm{H}} = P_s\mathbf{H}_{\mathrm{cen}}$ is the SCM of the prototype signals; and $\mathbf{G} = \tilde{\mathbf{C}}_{\mathrm{inc}}\tilde{\mathbf{C}}_{\mathrm{cen}}^{-1}$ is a matrix that matches the channel energies of the prototype signals with the target energies. The optimisation problem is then expressed as

$$\underset{\mathbf{M},\mathbf{M}_{\mathrm{dec}}}{\arg\min}\ \mathbb{E}[||\mathbf{d}_{\mathrm{opt}} - \mathbf{G}\mathbf{d}_{\mathrm{cen}}||^2], \quad \text{subject to} \tag{11}$$

$$\mathbf{M}\mathbf{C}_{\mathrm{cen}}\mathbf{M}^{\mathrm{H}} + \mathbf{M}_{\mathrm{dec}}\tilde{\mathbf{C}}_{\mathrm{cen}}\mathbf{M}_{\mathrm{dec}}^{\mathrm{H}} = \mathbf{C}_{\mathrm{inc}}. \tag{12}$$

The solution to this problem is given by

$$\mathbf{M} = \mathbf{K}_{\mathrm{inc}}\mathbf{V}\mathbf{U}^{\mathrm{H}}\mathbf{K}_{\mathrm{cen}}^{-1}, \tag{13}$$

where the decompositions $\mathbf{D} = \mathbf{K}_{\mathrm{inc}}\mathbf{K}_{\mathrm{inc}}^{\mathrm{H}}$ and $\mathbf{H}_{\mathrm{cen}} = \mathbf{K}_{\mathrm{cen}}\mathbf{K}_{\mathrm{cen}}^{\mathrm{H}}$ are defined similarly as in Sec 3.3. The $\mathbf{U}, \mathbf{V}$ are obtained from the singular value decomposition $\mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{H}} = \mathbf{K}_{\mathrm{cen}}^{\mathrm{H}}\mathbf{G}\mathbf{K}_{\mathrm{inc}}$, while $\mathbf{G}$ reduces to $\mathbf{G} = \tilde{\mathbf{D}}\tilde{\mathbf{H}}_{\mathrm{cen}}^{-1}$. Note that $\mathbf{M}$ is computed first, while $\mathbf{M}_{\mathrm{dec}}$ is obtained only if decorrelation is required to reach the remaining target SCM $\mathbf{C}_{\mathrm{inc}} - \mathbf{M}\mathbf{C}_{\mathrm{cen}}\mathbf{M}^{\mathrm{H}}$, which may not be always the case; e.g. when the spreading area is comparatively narrow.

5. EVALUATION

The implementation of the proposed algorithm employed the alias-free short-time Fourier transform (afSTFT) filterbank described in [30], configured with a hop size of 128 samples and with the additional hybrid filtering of the lower-bands to obtain 133 frequency bands in total. Decorrelators based on cascaded lattice all-pass filters, as described in [26] and implemented in the open-source Spatial_Audio_Framework[2], were employed for the decorrelation.

The evaluation of the proposed spreading algorithm was based on HRTFs and split into two parts. The first evaluation involved the use of white noise stimuli as the input, and compared the output of the spreading algorithm to incoherent noise signals distributed over the same spreading area which served as the reference. Objective perceptual metrics, namely: the binaural colouration, inter-aural level and coherence differences (ILD & IC), were then computed for the reference (Ref), the coherent spreading baseline of Sec. 3.1 (BL), the unconstrained approach described in Sec. 3.3 (EVD), and the proposed optimal-mixing approach described in Sec. 4 (OM). Since, in this case, the target is binaural, the auto- and inter-channel contributions between the left and right (l, r) channels are given as

$$\mathbf{C}_{\mathrm{d}}(f) = \begin{pmatrix} c_{d_{ll}}(f) & c_{d_{lr}}(f) \\ c_{d_{rl}}(f) & c_{d_{rr}}(f) \end{pmatrix} = \mathbb{E}[\mathbf{d}(f)\,\mathbf{d}^{\mathrm{H}}(f)], \tag{14}$$

from which the perceptual metrics can be derived as follows:

$$\mathrm{Colouration}(f) = 10\log_{10}[c_{d_{ll}}(f) + c_{d_{rr}}(f)], \tag{15}$$

$$\mathrm{ILD}(f) = 10\log_{10}[c_{d_{ll}}(f)/c_{d_{rr}}(f)], \tag{16}$$

$$\mathrm{IC}(f) = \frac{\mathrm{real}[c_{d_{lr}}(f)]}{\sqrt{c_{d_{ll}}(f)c_{d_{rr}}(f)}}. \tag{17}$$

An example of these metrics plotted over frequency for a white noise input signal is depicted in Fig. 1. Note that all perceptual metrics were averaged over one second.

[2] https://github.com/leomccormack/Spatial_Audio_Framework

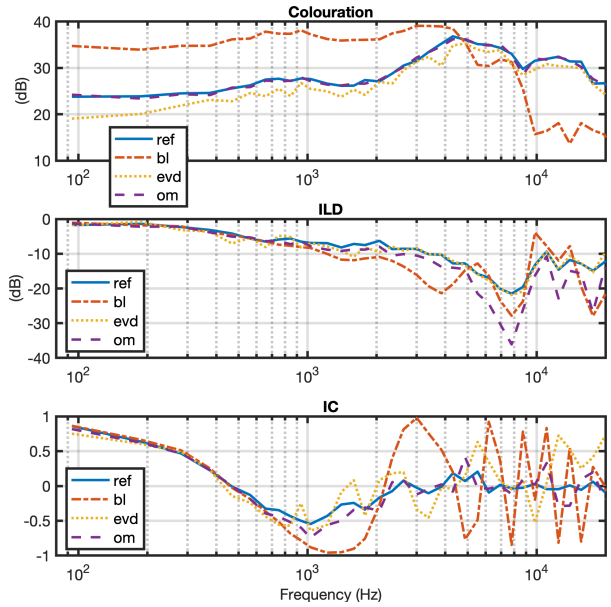

Figure 1: An example of binaural perceptual parameters plotted over frequency for a source at 45 degree azimuth, 0 elevation, and with 60 degrees of spread.

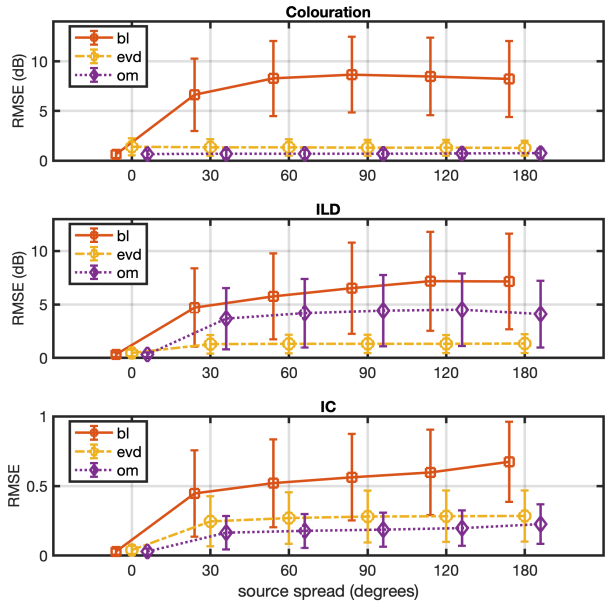

Figure 2: The RMSE values and standard deviations for the perceptual parameters, which were averaged over all 836 directions and ERB frequencies for different degrees of source spreading.

Simuli	BL	EVD	OM
Speech (male)	0.2199	1.6233	0.3555
Speech (female)	0.5136	1.2979	0.5731
Drums	0.4139	2.1727	0.4266
Strings	0.1661	1.3856	0.1735
Seagulls	0.3507	1.5155	0.3653
Waves	0.3646	1.5636	0.3639
Average	**0.3381**	**1.5931**	**0.3763**

Table 1: The SDE values for a source at 90 degree azimuth, 0 elevation, with 60 degrees of spread, for different stimuli.

These metrics were then computed for all 836 directions in the employed HRTF dataset, and for target spreading angles: [0 30 60 90 120 150 180] degrees. The metrics for the three spreading modes were then compared to those provided by the reference, based on the root-mean-square-error (RMSE) averaged over all 836 directions and over the perceptually-motivated equivalent rectangular bandwidths (ERB) frequency axis. These error values and their standard deviations are depicted in Fig. 2. It can be observed that the colouration and IC errors for the proposed and unconstrained EVD solutions are significantly closer to the reference compared to the baseline method. The ILD errors for the proposed method are then closer to the baseline than to the reference and the unconstrained EVD solution. However, while this first part of the evaluation provides insight into the binaural cues and colouration of the proposed method, it does not demonstrate how its additional constraints allow it to better retain the original signal fidelity compared to the EVD solution. Therefore, the second evaluation involved computing a signal distortion error (SDE) metric based on the output time-domain signals as

$$\text{SDE} = \sqrt{\frac{\sum_n ||\mathbf{d}(n) - \mathbf{d}_{\text{ref}}(n)||^2}{\sum_n ||\mathbf{d}_{\text{ref}}(n)||^2}}, \tag{18}$$

where $\mathbf{d}_{\text{ref}}$ is a signal corresponding to the central spreading direction. This metric was computed for the three spreading modes, using a number of different input stimuli. The results are given in Table. 1. It is shown that while the EVD approach can deliver similar or lower RMSE values for the perceptual metrics compared to the proposed OM approach, it does so with the penalty of increased signal distortion. Audio files which also demonstrate this distortion can be found on the companion web-page, or revealed by using the audio plug-in when set to the EVD spreading mode.

6. CONCLUSION

This paper has proposed an algorithm for rendering incoherently spread sound sources over arbitrary playback setups. It employs an optimised covariance domain solution to synthesise output signals exhibiting inter-channel relationships defined by a target spatial covariance matrix. In this case, the target is a diffuse covariance matrix for the specified confined spreading area. The solution is then constrained to mix decorrelated signal energy into the output, only to the degree necessary to fulfil the remaining target inter-channel relationships after a purely linear combination of the input signals has first been conducted. Objective evaluations, in the context of synthesising binaural signals corresponding to spread sound sources, demonstrate that the proposed method provides low binaural colouration and interaural coherence errors when compared to: a perfectly incoherent reference case, a coherently spread baseline and an unconstrained incoherently spread alternative. The proposed constrained method fairs less favourably with regard to binaural interaural level difference errors, compared to the unconstrained approach. However, it may be argued that this is nonetheless a permissible compromise, as the proposed constraints yield higher signal fidelity; as verified based upon an objective distortion metric, and informal listening of the provided sound examples.

7. REFERENCES

[1] V. Pulkki, "Uniform spreading of amplitude panned virtual sources," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452)*. IEEE, 1999, pp. 187–190.

[2] A. Franck, F. M. Fazi, and E. Hamdan, "An optimization approach to control sound source spread with multichannel amplitude panning," in *24th International Congress on Sound and Vibration*, 2017.

[3] V. Pulkki, "Generic panning tools for MAX/MSP," in *ICMC*, 2000.

[4] L. McCormack and A. Politis, "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, 2019.

[5] T. Carpentier, "Ambisonic spatial blur," in *Audio Engineering Society Convention 142*. Audio Engineering Society, 2017.

[6] N. Epain, C. Jin, and F. Zotter, "Ambisonic decoding with constant angular spread," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 928–936, 2014.

[7] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, 1995.

[8] G. Potard and I. Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays," in *Proc. Int. Conf. on Digital Audio Effects (DAFx'04)*, 2004.

[9] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[10] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2711–2721, 2013.

[11] E. Kermit-Canfield and J. Abel, "Signal decorrelation using perceptually informed allpass filters," in *Proceedings of the 19th International Conference on Digital Audio Effects*, 2016, pp. 225–31.

[12] F. Zotter and M. Frank, "Phantom source widening by filtered sound objects," in *Audio Engineering Society Convention 142*. Audio Engineering Society, 2017.

[13] C. Gribben and H. Lee, "The perception of band-limited decorrelation between vertically oriented loudspeakers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 876–888, 2020.

[14] T. Hirvonen and V. Pulkki, "Center and spatial extent of auditory events as caused by multiple sound sources in frequency-dependent directions," *Acta acustica united with acustica*, vol. 92, no. 2, pp. 320–330, 2006.

[15] T. Pihlajamäki, O. Santala, and V. Pulkki, "Synthesis of spatially extended virtual source with time-frequency decomposition of mono signals," *Journal of the Audio Engineering Society*, vol. 62, no. 7/8, pp. 467–484, 2014.

[16] A. Franck, F. M. Fazi, and F. Melchior, "Optimization-based reproduction of diffuse audio objects," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2015, pp. 1–5.

[17] M. Hakala, "Synthesis of spatially extended sources in virtual reality audio," Master's thesis, School of Electrical Engineering, Aalto University, 2019.

[18] H. Su, A. Marui, and T. Kamekawa, "The auditory source widening effect in binaural synthesis with spatial distribution of frequency bands," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 691–704, 2019.

[19] H. Lee, "Perceptual band allocation (PBA) for the rendering of vertical image spread with a vertical 2D loudspeaker array," *Journal of the Audio Engineering Society*, vol. 64, no. 12, pp. 1003–1013, 2016.

[20] F. Zotter, M. Frank, M. Kronlachner, and J.-W. Choi, "Efficient phantom source widening and diffuseness in ambisonics," in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, vol. 3, 2014, p. 5.

[21] H. Su, A. Marui, and T. Kamekawa, "Virtual source width in binaural synthesis with frequency-dependent directions," in *Audio Engineering Society Convention 142*. Audio Engineering Society, 2017.

[22] J. Vilkamo, T. Bäckström, and A. Kuntz, "Optimized covariance domain framework for time–frequency processing of spatial audio," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 403–411, 2013.

[23] A. Politis, "Diffuse-field coherence of sensors with arbitrary directional responses," *arXiv preprint arXiv:1608.07713*, 2016.

[24] S. J. Schlecht, B. Alary, V. Välimäki, E. A. Habets, *et al.*, "Optimized velvet-noise decorrelator," in *Proc. Int. Conf. Digital Audio Effects (DAFx-18), Aveiro, Portugal*, 2018, pp. 87–94.

[25] M. Bouéri and C. Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *Audio Engineering Society Convention 117*. Audio Engineering Society, 2004.

[26] J. Herre, H. Purnhagen, J. Breebaart, C. Faller, S. Disch, K. Kjörling, E. Schuijers, J. Hilpert, and F. Myburg, "The reference model architecture for MPEG spatial audio coding," in *Audio Engineering Society Convention 118*, 2005.

[27] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852–866, 2015.

[28] A. Politis, L. McCormack, and V. Pulkki, "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2017, pp. 379–383.

[29] L. McCormack and S. Delikaris-Manias, "Parametric first-order ambisonic decoding for headphones utilising the cross-pattern coherence algorithm," in *EAA Spatial Audio Signal Processing Symposium*, 2019, pp. 173–178.

[30] J. Vilkamo and T. Bäckström, "Time-frequency processing: Methods and tools," *Parametric Time-Frequency Domain Spatial Audio*, p. 3, 2017.