

DECORRELATION FOR IMMERSIVE AUDIO APPLICATIONS AND SOUND EFFECTS

Sascha Disch

Fraunhofer IIS
Fraunhofer Institute for Integrated Circuits
Erlangen, Germany
sascha.disch@iis.fraunhofer.de

ABSTRACT

Audio decorrelation is a fundamental building block for immersive audio applications. It has applications in parametric spatial audio coding, audio upmix, audio sound effects and audio rendering for virtual or augmented reality applications. In this paper, we provide insights into the practical design considerations of an audio decorrelator on the example of the decorrelator contained within the upcoming MPEG-I Immersive Audio ISO standard [1]. We describe the desirable properties of such a decorrelator, common approaches for implementation and our particular technology choices for the decorrelator used in MPEG-I for rendering sound sources with homogeneous extent.

1. INTRODUCTION

Historically, many electronic sound effects were designed with the aim to make the sound of an instrument more prominent or “larger” [2]. Usually this came at the price of certain originally unintended changes in sound timbre, modulation or pitch variations. Examples of such effects are (stereophonic) chorus, flanger, phaser and the like or, if it comes to multitrack recording, the studio technique of doubling and layering instrumental parts or voices. However, these artifacts and characteristic sound changes soon were welcomed and accepted by producers and artists and artistically integrated into their art performances.

Opposed to these effects, the aim of audio decorrelation is to as well as possible avoid any perceivable change in the original audio signal and provide decorrelated, but subjectively indistinguishable copies of the original input signal. Such decorrelated signals can then be used e.g. for spatially distributing multichannel sound or modeling spatial source extent. The latter denotes the technique of rendering a sound “larger” than the one emitted from a point source, but rather as if originating from a spatially extended sound source. Ideally, this can be accomplished without any unintended artifacts and timbre changes.

Applications for audio decorrelation are in perceptual coding of audio (parametric stereo/multichannel codecs) [3][4][5], subjective sound enhancement like audio upmix for channel signals and ambisonic signals and also for Virtual Reality (VR) and Augmented Reality (AR) rendering applications [6]. VR and AR rendering applications implement 6-Degrees-of-Freedom (6DoF) audio rendering for an interactive listener position in 3-dimensional space (3DoF) plus allowing for arbitrary listener’s head rotary

movements in all three possible axes (pitch, yaw, roll). One example of such a render is the upcoming MPEG-I Immersive Audio Standard [7] [1][8].

The decorrelation described in this paper has been designed especially for use within said MPEG-I Immersive Audio standard. In the MPEG-I renderer, it is utilized for modeling sound sources with spatial extent [9].

However, the decorrelator is pretty much self-contained and may find uses in other application including spatially enhanced audio effects. A combination of decorrelation with well-known traditional audio effects, eg. echo, delay, panning, tremolo, appears to be interesting to spatially widen and distribute the raw effect signal.

2. BACKGROUND

2.1. Previous Work

In the past, there have been a number of publications on audio signal decorrelation, for instance from Kendall [10], Boueri et al. [11] and Kermit et al. [12]. A publication of Potard et al. [13] addressed the rendering of sound source width through decorrelation and Anemüller et al. [9] have utilized the decorrelator presented in this paper to model spatially extended sound sources.

Purnhagen et al. [14] proposed audio decorrelation for parametric spatial audio coding, e.g. so-called parametric stereo. Here, the decorrelated signal can be seen as the side signal of a Mid/Side (M/S) representation of a stereo signal whereas the original input signal is the mid signal. Two decorrelated output signals are then obtained from the M/S representation by conversion to Left/Right (L/R) stereo. An adaptive M/S mix can be used to adjust the perceived spatial width.

Newer publications discuss the insight that transients should be treated differently from other signal parts in decorrelation as suggested by Disch et al. [15] and Penniman [16]. The reasoning behind this is two-fold: on one hand, unwanted transient dispersion shall be avoided as an audible artefact. On the other hand, convincing multichannel reproduction of spatial width of ambience signals consisting of dense transient events (e.g. applause) is not so much achieved through altered waveforms as produced by common decorrelation techniques, but rather through a widespread spatial redistribution of these individual transient events by fine grain temporal panning [15][17].

2.2. Decorrelator Design Criteria

Audio decorrelators for practical real-time applications ideally have to fulfill the following requirements:

- Reasonable computational complexity

Copyright: © 2023 Sascha Disch. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, adaptation, and reproduction in any medium, provided the original author and source are credited.

- Low latency
- Stream processing capability (single pass processing)

A general audio signal may conceptually be composed of a mixture of many different signal classes, like steady tones, transients, speech, talking or singing voice or noise. All these signal classes contained in the signal need to be processed by the decorrelator with little artifacts. So, a well-designed decorrelator should have a number of perceptual properties:

- Shall preserve sound timbre
- Shall not sound reverberated
- Shall avoid modulations of amplitude or pitch
- Shall allow precise reproduction of transients
- Shall decorrelate mixtures of spatially distributed transient events (applause, rain drops, fire crackling)

Audio decorrelation processing with reasonable latency and computational complexity is most successfully attempted by application of filters [18]. Allpass filters have the useful property to heavily influence the phase of a signal through their phase response while having a flat magnitude response. Although the signal waveform at a given point in time is dissimilar to the original signal, yet its timbre is still the same. The most simple allpass filter is a pure delay corresponding to a linear phase response. However, when input signal and filter output are combined, strong regularly spaced comb-filtering results and timbre colorization occurs.

Therefore, allpass filters that sufficiently alter the signal's phase in a more uneven way are preferably used, e.g. a combination of different allpass filters in series or multiple allpass filters in a nested structure, where the delay element of one allpass filter is replaced by a second allpass filter. These filters usually have a long, ringing impulse response that is audible as a reverberant tail of the processed sound.

For that reason, key to successful audio decorrelation using allpass filters is the preservation or the reconstruction of the signal's temporal and spectral envelope. Any reverberant tail caused by the long impulse response of the filters can be suppressed by application of envelope shaping and thereby will become inaudible.

The phase response of suitable allpass filters is also likely to impair transients in the signal mix through phase dispersion, e.g. for sharp signal onsets like plugged strings or piano tones. A dedicated handling of such transients within the decorrelator processing can further improve perceptual quality.

3. DECORRELATOR CONCEPT

3.1. Overview

The MPEG-I decorrelator is designed to match the criteria described above with respect to perceptual quality but also with respect to latency, complexity and stream processing capability.

It is based on overlapping DFTs for analysis of the incoming signal and synthesis of the decorrelated signal. A combination of a pre-delay and an allpass filter chain is applied in the (subsampled) DFT subband domain followed by a time/Frequency (t/F) envelope shaper. The envelope shaper applies an individual temporal envelope shaping in the DFT's spectral subbands. A short DFT block size of 256 samples at 48kHz sampling rate keeps latency low at ~2.7ms and also enables direct t/F envelope shaping downstream in DFT domain with sufficient temporal resolution.

A dedicated transient handling excludes transients from being subjected to phase dispersion. Excluding onset transients from decorrelation of most point sources (apart from e.g. mono recordings of multi-transient mixtures) appears to not impair the overall listening impression of spatial extent, which is the main task of the MPEG-I decorrelator. As explained in subsection 2.1, for said ambient mixtures of spatially distributed transient events, the individual transient events would have to be spatially re-distributed by other techniques e.g. by rapid amplitude panning [15].

Consequently, transient signal blocks are copied unaltered into the decorrelator output signal. When adding/subtracting original signal and processed signal in the output assembly stage, a scaling factor accounts for the addition of these coherent signal parts and ensures equal level with the incoherently added/subtracted decorrelated signals.

The transient handling entirely aims at an artefact-free reproduction of transient onsets. In consequence, relevant time constants of the transient handling are designed to be shorter than pre-delay/filterdelay such that any traces of delayed and dispersed transient onsets are excluded from the output signal.

Figure 1 depicts the block diagram of the decorrelator. To create two decorrelated output signals y_{Out1} and y_{Out2} from one input signal x , the MPEG-I decorrelator performs the steps that are described in the following.

Novel contributions - to our best knowledge - of our new decorrelator design are the application of low order Schroeder allpass filters in the subsampled DFT domain on complex spectral subband signals, the transient handling functionality and the computationally efficient co-use of the DFT for subband filtering and subband envelope estimation and shaping.

3.2. Input Buffering

The decorrelator has an internal processing cycle, managed by an input ring buffer, of 256 sample frames with a 128-sample overlap between subsequent processing frames. A sine window according to Equation 1 is applied where $m = 0 \dots M - 1$ is the time index within a frame n and $M = 256$.

$$x_{Win}(m) = x_{In}(m) \cdot \sin((m + 0.5) \cdot \pi/M) \quad (1)$$

Next, a 256-point Discrete Fourier Transform (DFT) is performed on the windowed input frame $x_{Win}(m)$ to obtain $K = 129$ complex valued frequency bins $X_{DS}(n)$ for each frame n . In the following processing, the temporal sequences of individual DFT coefficients are treated as a set of spectral subband signals with n being the time index subsampled by a factor of 128 with respect to m .

3.3. Pre-Delay and Allpass Filters

The $k = 0 \dots K - 1$ complex DFT coefficients of each DFT subband (index k is omitted where possible for better readability) are passed through a delay, as illustrated in Equation 2, where D_d is the pre-delay. This pre-delay is set to 4 frames (~10.7ms).

$$X_{Del}(n) = X_{DS}(n - D_d) \quad (2)$$

Next, the delayed signal $X_{Del}(n)$ is allpass filtered according to Equation 3

$$H(z) = \frac{\gamma + z^{-D}}{1 + \gamma \cdot z^{-D}} \quad (3)$$

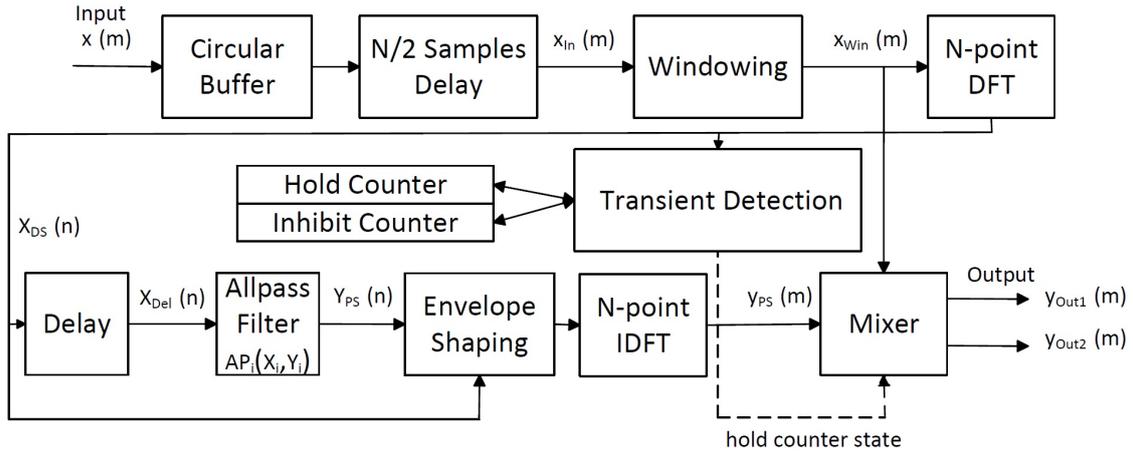


Figure 1: Decorrelator signal flow overview. Based on overlapping DFTs for analysis of the incoming audio signal, a combination of pre-delay and allpass filter chain is applied in the DFT subband domain followed by a time/Frequency (t/F) envelope shaper and the output assembly stage controlled by a transient handling. The transient detection information (dashed line) influences the internal mixing process of the output audio signals.

The chained series of i allpass filters AP_i is further illustrated in Equation 4, where filter coefficients are $\gamma = 0.7$ and D_{ap_i} represents the amount of delay of each allpass filter, which is set to 1, 2, 3, 5 frames for $i = 1, 2, 3, 4$.

$$Y_i(n) = \gamma \cdot X_i(n) + X_i(n - D_{ap_i}) - \gamma \cdot Y_i(n - D_{ap_i}) \quad (4)$$

In the following steps, $X_{DS}(n)$ is denoted the direct signal (DS), while $Y_{PS}(n)$, which went through the delay and the series of allpass filters, is called the processed signal (PS).

As a side note - if needed in future applications, a set of mutually independent decorrelators can be obtained by variation of pre-delay/allpass filter parameters and structures as sketched in subsection 2.2.

3.4. Transient Handling

In the decorrelator's output, transients are excluded from the pre-delay and allpass processing. This is managed by the transient handling. A transient in the current frame is detected by calculating whether the energy of the current frame, summed over $k = 4 \dots K$ frequency bins, is stronger than the previous frame by a threshold $T = 2.8$. The accuracy of transient detection is not overly critical; a false positive transient detection result, for example, will just exclude the falsely classified signal part from decorrelation causing no further artefacts. For this reason, the implemented transient detection algorithm was kept rather simple and computationally efficient.

Two counters control the transient processing: a so-called hold counter and an inhibition counter. The hold counter controls the duration of the detected transient event and the inhibit counter enforces a minimum temporal separation between detected transient events. Both are initially set to their inactive state 0. When a transient is detected $t_n = 1$ and the inhibition counter is inactive, a hold counter is started for the next 8 frames ($\sim 21.3\text{ms}$) to control a muting of the PS in the output mix. Also, the inhibition counter starts counting to prevent a hold counter start in the next 56 frames ($\sim 149.3\text{ms}$). In addition, if another transient is detected during this

inhibition time, this will re-start the inhibition counter, and the inhibition time will be increased from 56 to 64 frames. With these time constants, pulse trains with a pulse frequency greater than approximately 6Hz will not be subjected to the dedicated transient processing. The threshold of 6Hz was chosen since at that frequency pulses are still clearly perceived as separate events in fast succession (where preservation of transient perceptual quality is essential) rather than a continuous tonal sound (where good decorrelation is desirable). When active counters reach their maximum count, they are reset to their inactive state 0.

The durations reflected in the counters for transient handling have been chosen to be larger than pre-delay and allpass filter delay to suppress any traces of dispersed transients in the PS. Additionally, a hysteresis is implemented for avoiding on/off toggling that may happen e.g. when the input signal contains low frequency pulse trains with slight temporal variations of pulse distance around the inhibition time.

The energy in current frame is calculated using the DS. The energy of the current frame is smoothed by a first order IIR lowpass filter with "forgetting factor" $\delta = 0.4$. Equation 5 illustrates how energy of the current frame, $E_{Tr}(n)$, is calculated from the energy of the previous frame, $E_{Tr}(n-1)$, and the DS, $X_{DS,k}(n)$. Equation 6 defines the calculation of the transient detection state parameter t_n .

$$E_{Tr}(n) = \delta \cdot \left(\sum_{k=4}^K X_{DS,k}(n)^2 \right) + (1 - \delta) \cdot E_{Tr}(n-1) \quad (5)$$

$$t_n = \begin{cases} 1 & \text{if } E_{Tr}(n) > T \cdot E_{Tr}(n-1) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

3.5. Envelope Shaping

The following equations explain the t/F envelope shaping. For each time frame n , in the output $\hat{Y}_{PS}(n)$, each frequency bin of the PS is amplified or attenuated if it is weaker or stronger by a factor of $\beta = 1.5$ as compared to the DS. Equation 7 and Equation

5 illustrate the estimation of the energy of the current PS, $E_{p,k}(n)$, and of the current DS, $E_{d,k}(n)$, smoothed by a first order IIR low-pass filter with “forgetting factor” $\alpha = 0.4$. Equation 9 represents the t/F shaping process depending on the energy relations. Finally, the PS is scaled with a fixed factor $f = 1.1$ that was empirically determined to adjust the mean signal energy to match the DS energy.

$$E_{p,k}(n) = \alpha \cdot Y_{PS,k}(n)^2 + (1 - \alpha) \cdot E_{p,k}(n-1) \quad (7)$$

$$E_{d,k}(n) = \alpha \cdot X_{DS,k}(n)^2 + (1 - \alpha) \cdot E_{d,k}(n-1) \quad (8)$$

$$\hat{Y}_{PS}(n) = \begin{cases} f \cdot Y_{PS}(n) \cdot \sqrt{\frac{\beta \cdot E_{d,k}(n)}{E_{p,k}(n)}} & \text{if } E_{p,k}(n) > \beta \cdot E_{d,k}(n) \\ f \cdot Y_{PS}(n) \cdot \sqrt{\frac{E_{d,k}(n)}{\beta \cdot E_{p,k}(n)}} & \text{if } E_{d,k}(n) > \beta \cdot E_{p,k}(n) \\ f \cdot Y_{PS}(n) & \text{otherwise} \end{cases} \quad (9)$$

A 256-point Inverse Discrete Fourier Transform (IDFT) is performed on $\hat{Y}_{PS}(n)$ to convert the processed frequency bins back to 256 time domain samples. To prepare for overlap-add in the output signal assembly, the signal is windowed with the same sine window as used for analysis.

3.6. Output Signal Assembly

The decorrelated output is generated as illustrated in Equation 10. In regular decorrelator operation, when the hold counter is inactive, the two decorrelated output frames are obtained from the M/S representation by conversion to L/R stereo as seen from the sum and difference computation of $x_{Win}(m)$ and $y_{PS}(m)$, respectively. If a transient was detected and the hold counter is activated, the two output frames will carry an identical signal consisting of the windowed input signal weighted by a scaling factor. The scaling factor $\frac{1}{\sqrt{2}}$ adjusts the energy of the “transient part” $2 \cdot x_{Win}(m)$ to match the energy of the “non-transient part” that is obtained by combining two incoherent signals $x_{Win}(m)$ and $y_{PS}(m)$. The windowing (computationally efficient inherited from the overlapping DFT frames) ensures a smooth cross fade when switching between the regular and the transient operation modes.

$$\begin{cases} \left. \begin{aligned} y_{Out,1}(m) &= x_{Win}(m) + y_{PS}(m) \\ y_{Out,2}(m) &= x_{Win}(m) - y_{PS}(m) \end{aligned} \right\} & \text{if hold cnt inactive} \\ \left. \begin{aligned} y_{Out,1}(m) &= \frac{2}{\sqrt{2}} \cdot x_{Win}(m) \\ y_{Out,2}(m) &= \frac{2}{\sqrt{2}} \cdot x_{Win}(m) \end{aligned} \right\} & \text{if hold cnt active} \end{cases} \quad (10)$$

Finally, the windowed frames are re-combined in an overlap-add procedure with 128-sample overlap between subsequent processing frames.

4. COMPLEXITY

The decorrelator was designed for high perceptual quality at low-est possible computational complexity. Allpass filtering is implemented using very simple subband IIR filters. The application of

pre-delay and allpass filtering is done in (subsampling) DFT domain. The analysis DFT of the input signal - the direct signal X_{DS} - can thereby be efficiently used for both filtering and as input to t/F-envelope shaping. The DFT windowing is co-used for transient handling controlled output signal cross fade as can be seen in Equation 10.

If a set of many mutually decorrelated output signals is desired, the entire set may advantageously share said analysis DFT and parts of the envelope shaping calculations, specifically the energy estimation according to Equation 8 and the transient handling processing according to Equations 5 and 6, since these are solely dependent on the common direct input signal $X_{DS}(n)$.

The exact complexity is determined by DSP hardware and on the DSP implementation, so here a pen-and-paper estimate is given. The following Table 1 provides an overview on the number of arithmetic operations of the different parts of the decorrelator. Operations (OP) are multiplications (MUL) or additions (ADD); Divisions (DIV) and square roots (SQRT) are assumed to equal 25 OP each (which is a very conservative assumption that might be too high on modern DSP hardware).

The processing takes place in 50% overlapped windowed 256 sample blocks. According to [19], a 256-point real-data FFT/IFFT consumes 3022 operations. All further operations in DFT domain are to be calculated on 129 complex subband signals, hence the factors of 129 and 2, respectively. To arrive at operations per sample, all numbers are normalized on the 128 sample time domain stride of the overlapped DFT.

The computational complexity of the decorrelator is a moderate 7.4 MOPS (Million Operations per Second).

5. QUALITY

Some properties of the new decorrelator design presented in this paper have been compared with the ones of the MPEG Surround (MPS) decorrelator [5] as a baseline. Figures 2 and 3 display plots of mean correlation of the two output signals, Figures 4 and 5 display plots of mean magnitude differences of the two output signals as a function of frequency for the MPS decorrelator and the new decorrelator design. The measurements were obtained using pink noise as an input signal. From the plots it is apparent that the remaining correlation of the output channels is always below 0.045 and the magnitude differences are always less than 0.3 dB in the relevant audio frequency range. So, the new decorrelator design is on par with the performance of the MPS decorrelator while avoiding the much higher delay and computational costs caused by the Quadrature Mirror Filterbanks (QMF) of the MPS decorrelator. Thorough expert listening confirmed a high overall perceptual quality and a transparent transient reproduction of the new design. Nevertheless, a formal perceptual evaluation will be done in the future.

6. APPLICATION IN MPEG-I

In MPEG-I Immersive Audio, sources with homogeneous spatial extent are modelled by a number of decorrelated point sources that are spatially distributed in the VR scene geometry. The necessary decorrelated point source signals are derived from one or more “original” input signals that are the direct signals input to the decorrelator. One example of such a source with spatial extent is the park fountain in the MPEG-I VR test scene depicted in Figure 6. In the MPEG-I Immersive Audio Reference Software, the

Table 1: Decorrelator complexity estimation. Operations (OP) equal multiplications (MUL) or additions (ADD); Divisions (DIV) and square roots (SQRT) are assumed to count 25 OP each.

	Operations (OP) per 128 sample frame	Sum of OP/frame
Analysis window (Equation 1)	256 · MUL	256
256-point real-data FFT/IFFT	2 · FFT (see [19])	2 · 3022 = 6044
4 Schroeder allpass filters on complex data (Equation 4)	4 · 129 · 2 · 2 · MUL; 4 · 129 · 2 · 2 · ADD	32 · 129 = 4128
Energy envelope estimation (Equation 78)	2 · 3 · 129 · MUL; 2 · 2 · 129 · ADD;	(6 + 4) · 129 = 1290
Transient detection (Equation 5)	3 · 126 · MUL; 2 · 126 · ADD	(3 + 2) · 129 = 645
t/F Envelope shaper (Equation 9)	2 · 129 · MUL; 1 · 129 · DIV; 1 · 129 · SQRT	(2 + 25 + 25) · 129 = 6708
Synthesis window	256 · MUL	256
Overlap+add	256 · ADD	256
Output assembly (Equation 10)	2 · 128 · ADD	256
Total		19839
	Total OP/sample	OP/sec@ $f_s = 48kHz$
Decorrelator	155	approx. 7.4 MOPS

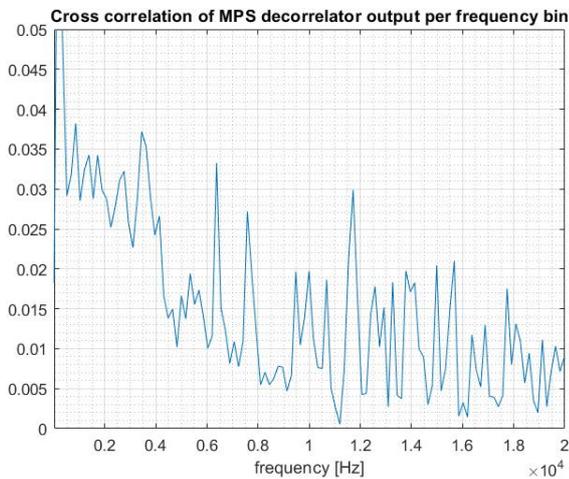


Figure 2: Baseline MPS Decorrelator: mean correlation between decorrelator output channels as a function of frequency.

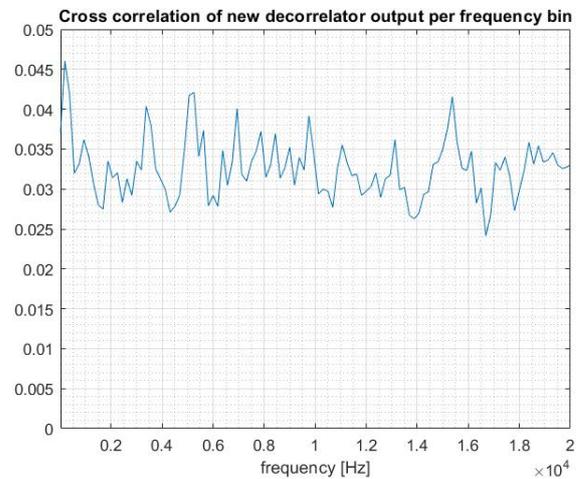


Figure 3: New decorrelator design: mean correlation between decorrelator output channels as a function of frequency.

presented decorrelator is available as an implementation in C++ [8].

7. CONCLUSIONS

In this paper, we describe the design considerations behind the decorrelator that is part of the upcoming MPEG-I Immersive audio standard and that is used for rendering sound sources with homogeneous extent. We present details of the actual implementation and provide computational complexity numbers showing a moderate total complexity of 7.4 MOPS for two decorrelated signals at 48kHz sampling rate. We also provide measurements on correlation and spectral magnitude differences of our new decorrelator design compared to the MPEG Surround (MPS) decorrelator as a baseline, showing that regarding these measurements the new design is on par with the baseline while minimizing processing delay and computational costs. We suggest that the new decorrelator design at hand might also be useful for integration in other spatial audio applications and audio effects.

8. REFERENCES

- [1] Jürgen Herre and Sascha Disch, “MPEG-I Immersive Audio - reference model for the new virtual / augmented reality audio standard,” *JAES Special Issue on Audio for Virtual and Augmented Reality*, 2023.
- [2] Harald Bode, “History of electronic sound modification,” *Journal of the Audio Engineering Society (JAES)*, vol. 32/10, pp. 730–739, 1984.
- [3] Heiko Purnhagen, Jonas Engdegard, Jonas Roden, and Lars Liljeryd, “Synthetic ambience in parametric stereo coding,” in *Audio Engineering Society Convention 116*, May 2004.
- [4] Christof Faller and Frank Baumgarte, “Binaural cue coding-part ii:schemes and applications,” *Speech and Audio Processing, IEEE Transactions on*, vol. 11, pp. 520 – 531, 12 2003.
- [5] J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén, W. Oomen,

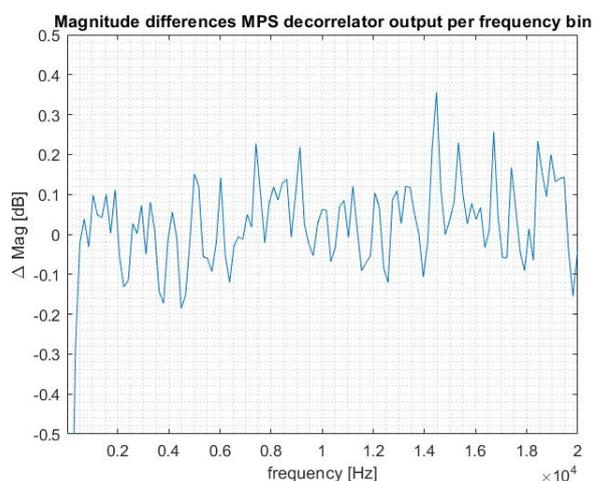


Figure 4: Baseline MPS Decorrelator: mean magnitude differences between decorrelator output channels as a function of frequency.

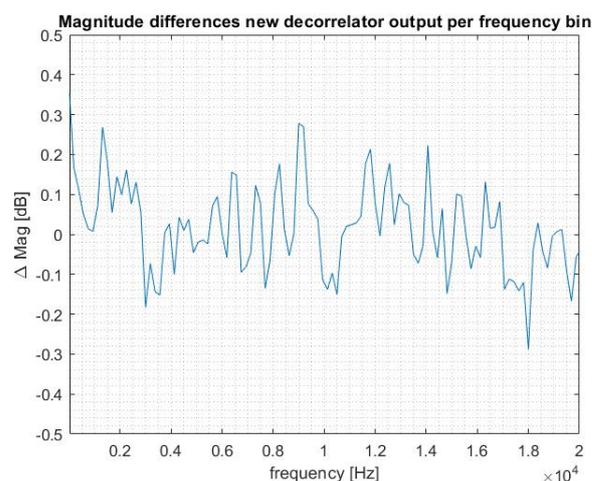


Figure 5: New decorrelator design: mean magnitude differences between decorrelator output channels as a function of frequency.

K. Linzmeier, and K. S. Chong, “MPEG Surround - the ISO/MPEG standard for efficient and compatible multichannel audio coding,” *Journal of the AES*, vol. 56, pp. 932–955, nov 2008.

- [6] Jean-Marc Jot, “Efficient models for reverberation and distance rendering in computer music and virtual audio reality,” in *International Conference on Mathematics and Computing*, 1997.
- [7] Schuyler R. Quackenbush and Jürgen Herre, “MPEG standards for compressed representation of immersive audio,” *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1578–1589, 2021.
- [8] “<https://www.mpeg.org/standards/mpeg-i/4/>,” 2023.
- [9] Carlotta Anemüller, Alexander Adami, and Jürgen Jürgen Herre, “Efficient binaural rendering of spatially extended sound sources,” *JAES Special Issue on Audio for Virtual and Augmented Reality*, 2023.
- [10] Gary Kendall, “The decorrelation of audio signals and its impact on spatial imagery,” *Computer Music Journal*, vol. 19, 12 1996.
- [11] Maurice Boueri and Chris Kyriakakis, “Audio signal decorrelation based on a critical band approach,” in *Audio Engineering Society Convention 117*, Oct 2004.
- [12] Elliot Kermit-Canfield and Jonathan Abel, “Signal decorrelation using perceptually informed allpass filters,” *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, 2016.
- [13] Guillaume Potard and Ian Burnett, “Decorrelation techniques for the rendering of apparent sound source width in 3d audio displays,” in *Proc. of the 7th Int. Conference on Digital Audio Effects (DAFx’04)*, Naples, Italy, Oct 2004.
- [14] Heiko Purnhagen, “Low complexity parametric stereo coding in mpeg-4,” *Proc. of the 7th Int. Conference on Digital Audio Effects (DAFx’04)*, Naples, Italy, 2014.



Figure 6: MPEG-I test scene “Park” with the fountain’s sound modeled as extended sound source.

- [15] Sascha Disch and Achim Kuntz, *A Dedicated Decorrelator for Parametric Spatial Coding of Applause-Like Audio Signals*, pp. 355–363, Springer-Verlag, January 2012.
- [16] R. Penniman, “A general-purpose decorrelation algorithm with transient fidelity,” *137th Audio Engineering Society Convention 2014*, pp. 99–107, 01 2014.
- [17] Florin Ghido, Sascha Disch, Jürgen Herre, Franz Reutelhuber, and Alexander Adami, “Coding of fine granular audio signals using high resolution envelope processing (HREP),” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 701–705.
- [18] Manfred Schroeder, “Natural sounding artificial reverberation,” *Journal of the Audio Engineering Society*, vol. 10, no. 3, pp. 219–223, 1962.
- [19] S. G. Johnson and M. Frigo, “A modified split-radix FFT with fewer arithmetic operations,” *IEEE Transactions on Signal Processing*, 2006.
- [20] Carlotta Anemüller, Oliver Thiergart, and Emanuel A. P. Habets, “A Data-Driven Approach to Audio Decorrelation,” *IEEE Signal Processing Letters*, vol. 29, pp. 2477–2481, 2022, Conference Name: IEEE Signal Processing Letters.