Signal decorrelation for diffuse sound fields

Audio engineering project

Matthias Blochberger

Supervisor: Dr. Franz Zotter

Graz, March 9, 2019

institut für elektronische musik und akustik

Abstract

The goal of signal decorrelation is to minimize the correlation coefficients of existing signals or to generate an arbitrary number of decorrelated signals from one source signal. This should however not introduce any disturbing amount of perceivable reverberation or other perceivable change or destruction in the resulting signals. The target method is applied for example in adaptive systems for multichannel parametric audio processing as these often rely on decorrelated signals to work efficiently or at all. A typical goal in parametric 3D audio processing deals with the creation of diffuse sound fields for resolution enhancements of first-order or low-order Ambisonic recordings. This work confirms that filters based on randomized group delays yield good results at relatively small computational effort. Using the relation between group delay and phase the corresponding all-pass filters can be generated. These all-pass structures are used to generate an encoded higher-order Ambisonics signal with diffuseness by applying a new strategy to distribute the signals. The new algorithm is implemented in a filter design GUI and as a VST plugin and evaluated in different variations as well as compared to known algorithms.

Zusammenfassung

Das Ziel von Methoden zur Signaldekorrelation ist, die Korrelationskoeffizienten mehrerer Signale zu minimieren beziehungsweise aus einem Signal eine beliebige Anzahl an unkorrelierten Signalen zu erzeugen. Dies soll allerdings die Signale nicht soweit beeinträchtigen, dass hörbarer Nachhall oder ein Verschmieren der Signalstruktur eintritt. Die Methoden finden Anwendung bei adaptiven parametrischen Mehrkanalsystemen, da die Funktionsfähigkeit bzw. Effizienz dieser meist dekorrelierte Signale voraussetzt. Ein typischer Anwendungsfall in parametrischer 3D-Audiosignalverarbeitung ist die Diffusfelderzeugung zur Erhöhung der Auflösung von ambisonischen Aufnahmen erster oder allgemein niedriger Ordnung. Auf zufällig erzeugten Gruppenlaufzeiten basierte Filter liefern gute Ergebnisse bei relativ geringem Rechenaufwand. Über Gruppenlaufzeit und Phase können Allpassfilter generiert werden. Die erzeugten Allpass-Gruppenlaufzeitentwürfe werden in Kombination mit einer neuen Signalverteilungsstrategie verwendet, um ein diffuses ambisonisches Signal hoher Ordnung zu erzeugen. Der erarbeitete Algorithmus ist in einer Benutzeroberfläche zum Filterdesign und einem VST-Plugin implementiert und wurde in verschiedenen Varianten evaluiert sowie mit bekannten Algorithmen verglichen.

Contents

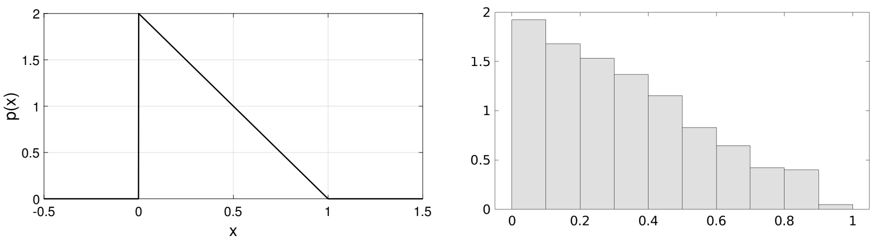
1 Introduction

The main focus of this project is to implement and test a system to generate an arbitrary amount of decorrelated signals from one single channel signal source without changing the perceived sound quality of the signals. The main motivation for this is the application in phantom source widening for sound sources in 3D audio.
Early works about the topic are found in [WFO92], [Ken95] and they deal with the manipulation of signal spectrum in terms of amplitude and phase to create multiple copies of a signal. An overview and evaluation of decorrelation techniques for apparent sound source width can be found in [PB04]. Existing methods using deterministic and short FIR-filters for decorrelation used in phantom source widening are described in [ZFMS11], [ZFKC14], [ZF17]. Other approaches use randomized group delays [Pes11], [CDA18] or randomized phase values [SS18].
This work focuses on refining the proposed system by Elliot K. Canfield-Dafilou and Jonathan S. Abel [CDA18] which uses randomly generated group delay within reasonable constraints and hereby promises filters that avoid producing artifacts that impede sound quality.
The distribution of sound in a 3D audio playback system to achieve a desired auditory impression width can be approached by distributing frequency components of a signal as done in [ZFKC14] via a varying Ambisonics encoding or in [FFM15] by using optimization. Another method is the placement of virtual decorrelated virtual sources as shown in [ZF17] and [SS18]. In this work it is done in similar fashion by placing them evenly distributed on a spherical cap of interest.

2 Method

2.1 Random generation of group delay

A number of $\frac{N}{2}$ samples are drawn from a probability density function with a triangular shape, so that values closer to zero are produced more often. N is the desired length of the impulse response.

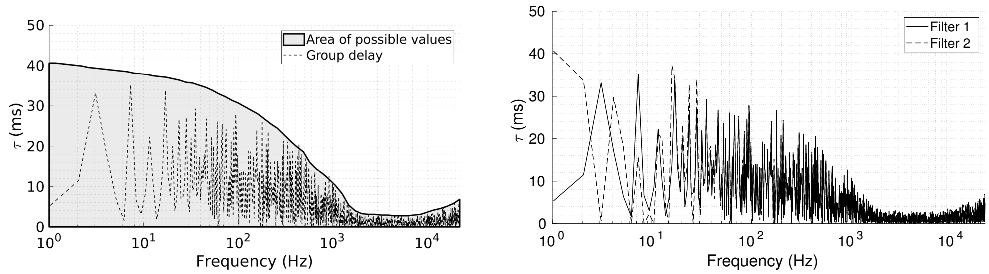


(a) PDF for sampling (b) Histogram in pdf form of sampled points

Figure 1 – Sampling of random variable

According to the distribution the values of these samples range from 0 to 1 and have to be scaled by the values of the maximum group delay specified at each frequency. These are defined by predetermined curves saved as *.mat* files or drawn as input by the means of a graphical user interface (sec. 3.2).
The density of the random group delay sequence over the audible frequency range is adjustable by various warping functions. The consistently best sounding results were achieved with the *ERB* (fig. 3) warping which ensures sufficient variation in the low frequency range.
However, phase differences of π and multiple of it can lead to annoying or even disturbing artifacts in the lower frequency range when listening. In this case the solution is to limit the differences in the group delays to values which result in phase differences contained in the interval $[-\frac{\pi}{4} \quad \frac{\pi}{4}]$.

Limiting the group delay differences: Take a copy of the group delay curve of the first channel and add random values sampled from a uniform distribution on the interval $\left[-\frac{1}{4f} \quad \frac{1}{4f}\right]$ for the other channels. After limiting the values to a minimum of 0 and a maximum of the maximal group delay defined by hand, these group delay curves are then smoothed and cross-faded with the previously generated delay curves. The cross fade in this case happens linearly from $600Hz$ to $1400Hz$. The limited differences at low frequencies can be observed in fig. 2b.
Refer to section 3.1 for details on the MATLAB implementation.



(a) Group delay curve inside maximum delay constraints (b) Group delay curves of two channels

Figure 2 – Generated group delay

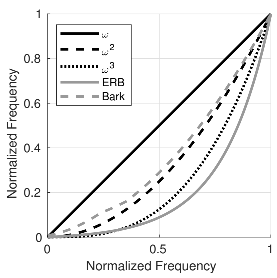


Figure 3 – Available warping types to specify the density of the group delay variation over frequency

2.2 All-pass filter from group delay

The relation between group delay $\tau(\omega)$ and the phase $\phi(\omega)$ is known as

$$\tau(\omega) = -\frac{\mathrm{d}\phi(\omega)}{\mathrm{d}\omega}. \tag{1}$$

Assuming any group delay curve over frequency, the construction of an all-pass filter is straightforward. Integrating over the negative group delay $-\tau(\omega)$ yields the phase $\phi(\omega)$

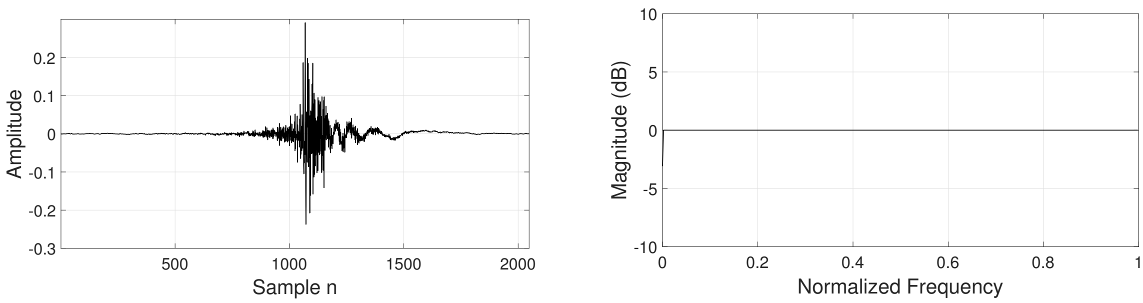
$$\phi(\omega) = -\int_0^\omega \tau(\omega)\mathrm{d}\omega \tag{2}$$

and the impulse response in the time domain is

$$h(t) = \mathcal{F}^{-1}\left[e^{j\phi(\omega)}\right]. \tag{3}$$

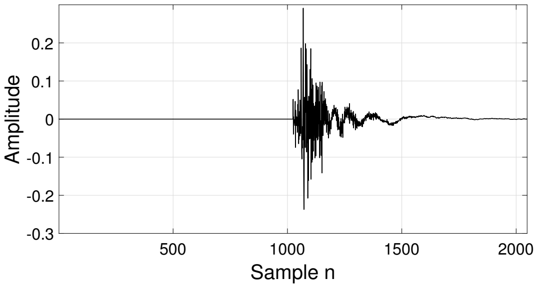
2.2.1 Temporal smearing and shaping of fast onsets

Depending on the chosen maximum group delay curve the impulse response of the resulting filters smear the temporal structure of the signals. Figures 4a, 4b demonstrate one of these filters where the resulting pre ringing is visible. These slow onsets are perceived as a degradation of the sound quality by destroying the transients of the source signal.
In [ZF17] it is demonstrated that the pre ringing of a decorrelating impulse response can be truncated with the effect of losing the pseudo all-pass structure, while improving the transient behavior as much to still improve the overall perceived sound quality. In contrast to deterministic methods that permit truncation of the anti-causal part, a suitable time instant has to be found. At the moment there is no clear method to choose this time instant for fast onsets hence it is rather done by trial and error. In general as more pre ringing is removed, the frequency response further diverges from a perfect all-pass filter. This effect is a noise-like diversion stemming from the initial random generation of the filter and removing parts of it as figures 5a to 5d show.

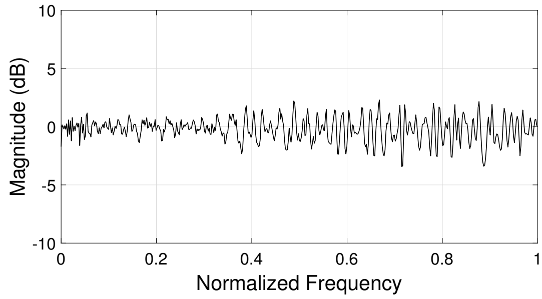


(a) $h[n]$ (b) $|H|_{dB}$

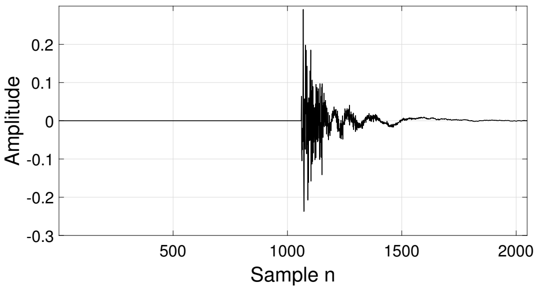
Figure 4 – Example for generated filter impulse response.



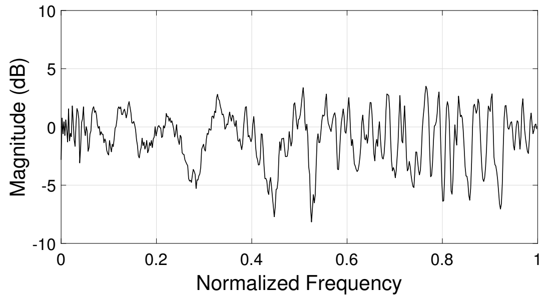
(a) $h[n]$ with $n_r = 1023$



(b) $|H|_{dB}$ with $n_r = 1023$



(c) $h[n]$ with $n_r = 1059$



(d) $|H|_{dB}$ with $n_r = 1059$

Figure 5 – Examples for generated filter impulse responses. n_r denotes the number of samples starting from 0 which have been truncated to value 0 to remove the pre ringing.

After the above-mentioned modifications MATLABs `grpdelay` is used to verify the general functionality of the algorithm as figure 6 shows s similar trajectory as the curves in figure 2. Refer to section 3.3 for details on the implementation of the algorithm.

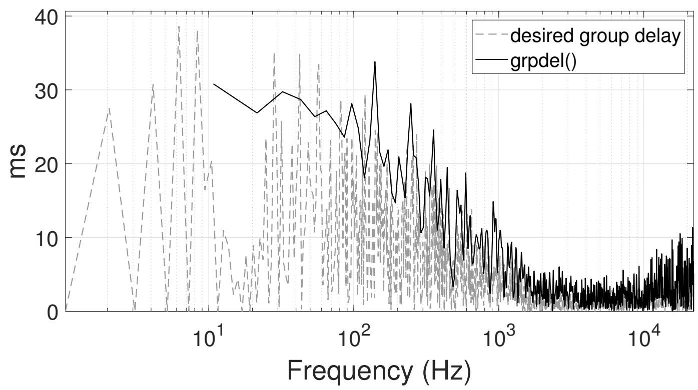


Figure 6 – Output of the `grpdelay` function compared to the desired random group delay values

3 Details on the implementation

The function to generate finite impulse responses (FIR) was implemented as a MATLAB program with user interface (fig. 7). This filter design tool allows to easily change the parameters such as the number of decorrelated impulse responses, FIR length, sampling rate, etc. The *smoothing* parameter is for smoothing the random group delay curve after random generation. As smoothing provided no initial perceivable effect it was not further explored or evaluated in the scope of this project; all data and examples in this work are done without ($K = 0$).

3.1 Random generation

To randomly generate the samples for the group delay curve, MATLABs `random` command is used. It samples from a triangular probability density function where figure 1 shows details.

```
pd = makedist('Triangular','a',0,'b',0,'c',1);
range_x = 0:0.001:1;
```

```
raw_dels = random(pd,N,L);
```

For each frequency bin, an independent sample drawn from the random process defines the group delay. Figure 3 shows warping curves for a non-equidistant mapping of the random samples to the frequency scale. The non-equidistant mapping has to be interpolated to an equidistant one for the inverse Fourier transformation (sec. 3.3). As mentioned above, the warping type *ERB* consistently yielded the best results.

3.2 GUI and definition of maximum group delay

As mentioned the random generation is constrained by zero and a curve of maximal group delays over frequency. The definition of this happens per MATLAB GUI. The user can draw a curve by dragging the curve. Figure 7 shows the interface with all options.

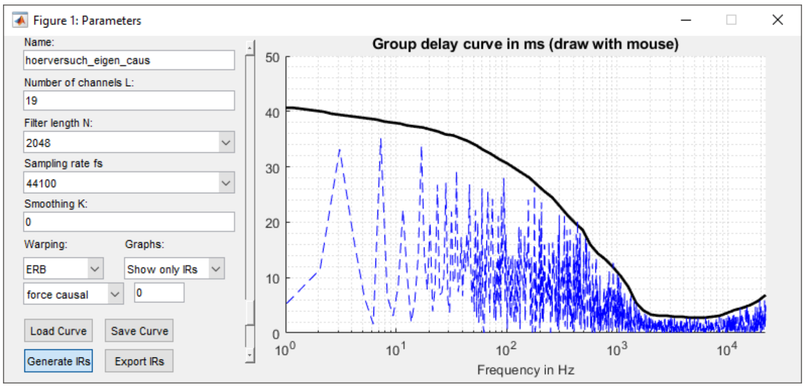


Figure 7 – Matlab tool for impulse response generation

GUI parameters:

Name: The description of the set of generated impulse responses for exporting.

Number of channels $\mathbf{L}$: The number of impulse responses that are generated for this set.

Filter length $\mathbf{N}$: The number of samples/tabs of the impulse response, which is used for the *IFFT* (sec. 3.3).

Sampling rate $\mathbf{f_s}$: Sampling rate which is used for the *IFFT*.

Smoothing (window) $\mathbf{K}$: length of the *Hann* window used for smoothing the randomly generated delay curve. Length is $2K + 1$.

Warping: Selection of frequency warping method. [*linear*, ω^2, ω^3, *ERB*, *bark*]

Onset Shaping: Selection to set impulse responses to zero at samples 0 to $\frac{N}{2} + m$ where m is the input number.

Load/Save Curve: Save the current curve and load a saved curve.

Generate IRs: Generate the impulse responses using the parameters.

Export IRs: Export *.conf* and *.wav* file for the *mcfx_convolver* VST plugin [Kro14].

3.3 Phase to impulse response

Section 2.2 outlines the mathematical method which yields an all-pass filter. The group delay at this point is given by a set of discrete points on a warped frequency axis. These are resampled to an equidistantly sampled frequency scale. The integral in continuous domain in equation (2) is approximated by using the *trapezoidal rule*. The MATLAB function `cumtrapz` allows to get the approximated integral for all frequency values.

```
dels_linspace = interp1(warp*fs*pi,final_dels.',linspace
    (0,fs*pi,N*10) ,'linear');
phase_linspace = -cumtrapz(linspace(0,fs*pi,N*10),
    dels_linspace);
```

The phase is mirrored at the origin to create a skew-symmetric spectrum. This ensures a real-valued impulse response after the inverse fast Fourier transformation.

```
phase_negative = flipud(phase_linspace);
phase = [ -phase_11(1:end-1,:); phase_linspace];
phase = interp1(linspace(-fs/2,fs/2,N*20-1),phase ,
    linspace(-fs/2,fs/2,N) ,'linear');
```

```
spec = exp(1j*phase);
H = fftshift(ifft(ifftshift(spec,1),[],1,'symmetric'),1);
```

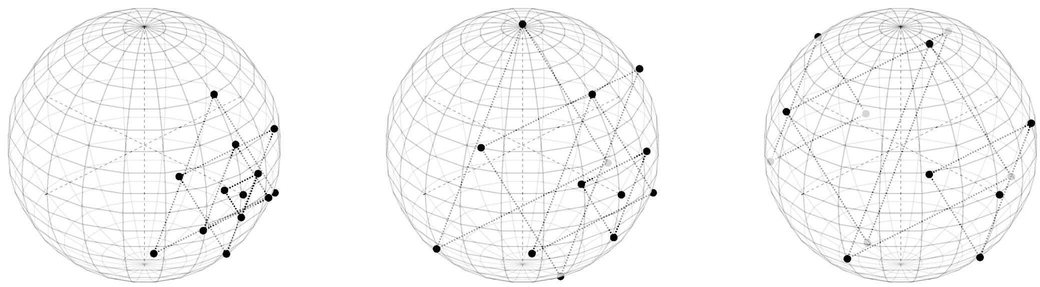
The matrix $\mathbf{H} \in \mathbb{R}^{NxL}$ (in code `H`) consists of the generated impulse responses as columns, each of which is normalized to

$$\sqrt{\sum_{i=0}^{N-1} h[i]^2} = 1 \tag{4}$$

where h is the impulse response of length N.

4 Application: Phantom source widening

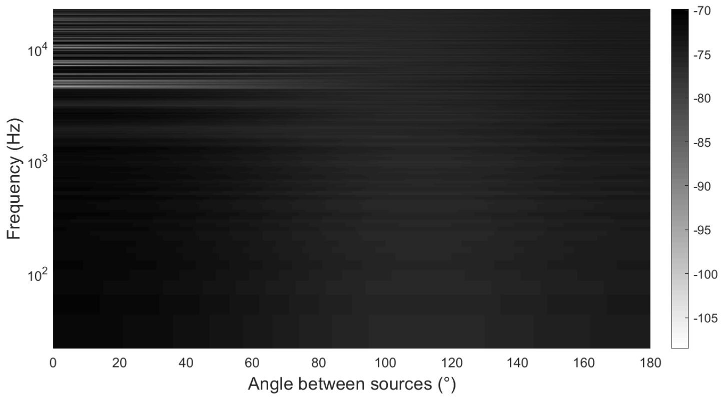
An application of this type of decorrelation is to create a method of phantom source widening. In stereo applications one can simply create two decorrelated signals using this method and change the perceived source width by scaling the maximal group delays which leads to more or less pronounced decorrelation. A 3D application in Ambisonics can achieve better results than the 2-channel stereo approach. The proposed Ambisonic method takes advantage of an arbitrary number of decorrelation filters, signals and virtual sources for a more diffuse appearance. Using a higher number of decorrelated signals and distributing them in meaningful constellations leads to a rich diffuse sound field.
In [SS18] the width of a phantom source is controlled by using a triangular pyramid with its apex aligned with the source direction and the aperture by scaling the size of the triangular base around the source. The Ambisonics order is reduced using gain correction on the Ambisonics signals so higher orders are faded out with bigger aperture. This approach is adapted by using a set of pyramids with an equilateral triangle as a base instead of one single tetrahedron to provide a even distribution of sound across the spherical cap of interest (fig. 8).



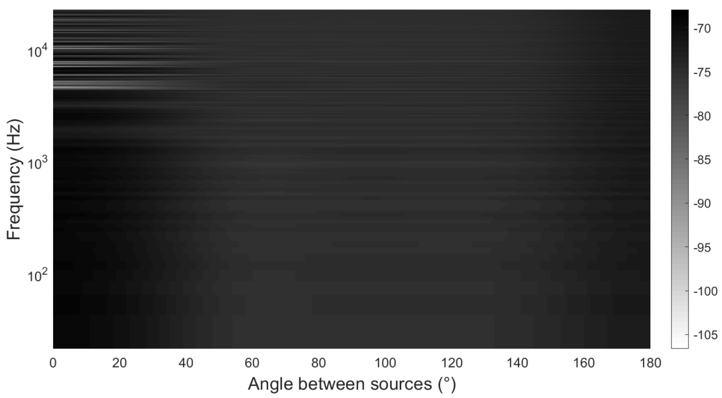
(a) 90° aperture (b) 180° aperture (c) 270° aperture

Figure 8 – Example of distribution of decorrelated signals on the sphere

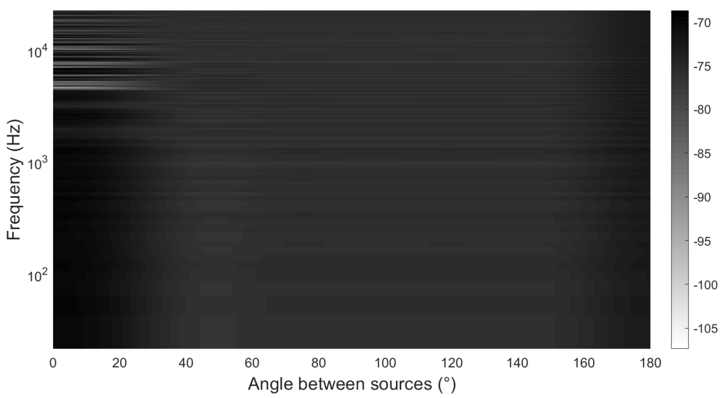
A prototype for a plugin providing phantom source widening for a single source was developed based on the *MultiEncoder* VST plugin from the IEMPlugin suite [Rud19]. The plugin has a single channel input and the number of outputs needed for the chosen Ambisonics order. Part of the functionality was readily provided by Daniel Rudrichs implementations for the *IEMPlugin suite*.
For the extension elaborated for the application of this work, the filters are loaded into the plugin via a *.wav* file with L channels, each containing one generated impulse response. The filtering of the single-channel input signal with every filter happens via fast convolution in the frequency domain. The resulting signals are distributed on the sphere as mentioned above.



(a) Ambisonics order 2



(b) Ambisonics order 5



(c) Ambisonics order 7

Figure 9 – **Overlapping virtual sources**: $10\log_{10}\sum_L |S_l(f)|^2$ where $S_l(f)$ are the spectra of the loudspeaker signals after decoding with L as number of loudspeakers. This shows the destructive interference in some high frequencies and constructive interference in low frequencies when two virtual sources are moved close to each other. This simulation was done on an array with 64 loudspeakers and using two filters (ch.-nr. 1 & 2) from $del_{2,onset}$.

The gains are calculated using

$$g_l(\theta) = \overbrace{\left[\frac{1+\cos\left(\theta_{l,end}\right)}{2}\right]^D}^{\text{Directional focus}} \cdot \underbrace{\left[\frac{1}{1+\exp\left(-\zeta_l\left(\theta_l-\theta_{l,trans}\right)\right)}\right]}_{\text{Gain reduction for overlapping sources}} \tag{5}$$

with $D \in [0, 10]$ as the directional focus parameter and $\theta_{l,end}$ as the end position elevation of the used virtual source constellation. The parameter ζ_l is a configuration parameter defining the rate at which the virtual sources are faded in and out and $\theta_{l,trans}$ is the transition point at which the fading happens. These parameters should be chosen according to the Ambisonics order to attenuate the interfering sources (fig. 9).

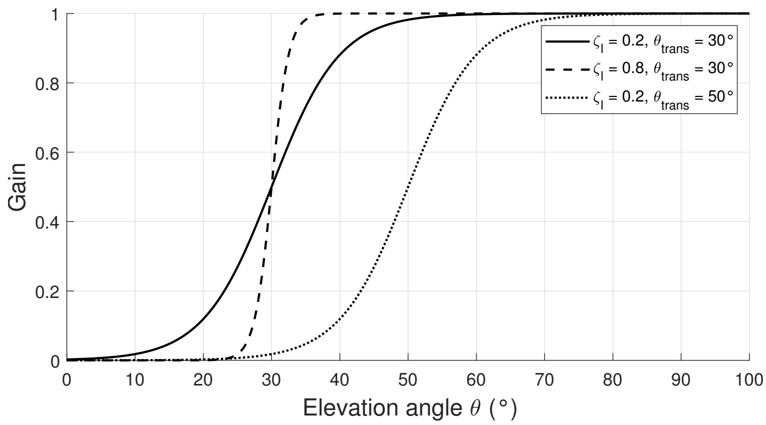


Figure 10 – Examples of *sigmoid* functions for the fade out of overlapping virtual sources.

The directional focus control seen in figure 11 additionally controls the gain scaling so the width and direction can be accentuated even with a wide distribution of signals. This provides more control over the width of the diffuse sound field controlling the directional fade-out behavior of the virtual source distribution. Gain normalization for constant loudness is necessary after calculation.

Note: The constellations are defined in *.json* configuration files by elevation and roll. Therefore all calculations can be done purely on the elevation of the virtual sources before individual roll and the following transformation with the *Master*-azimuth,-elevation and -roll which can be seen in fig. 11.

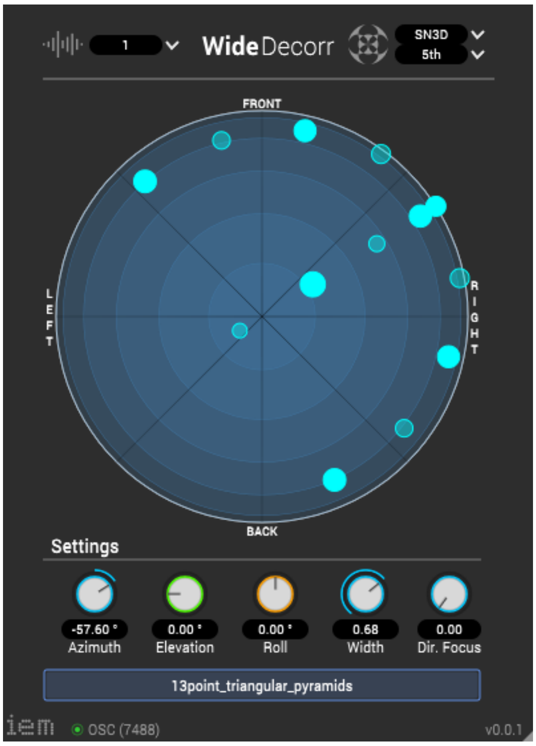


Figure 11 – VST for phantom source widening application. The plugin allows to load configurations of signal constellations saved in a *.json* format.

At the time of authoring this work, the VST plugin was not ready for release, as it should still benefit from thinkable smaller improvements.

5 Evaluation

For the evaluation of the algorithm three group delay curves have been chosen as pictured in fig. 12:

del_1 group delay curve that sounded promising while designing the algorithm.

del_2 The maximum group delay as proposed in [CDA18].

del_3 The values are pushed above the proposed maximum group delay to explore potential decrease in sound quality.

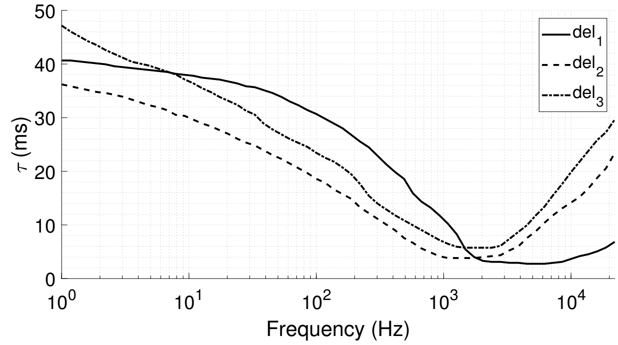


Figure 12 – The chosen delay curves

5.1 Evaluating the filter decorrelation

The cross correlation function (XCF)

$$\Phi_{xy}[l] = \sum_{m=0}^{N-1} x[m]y[l+m] \tag{6}$$

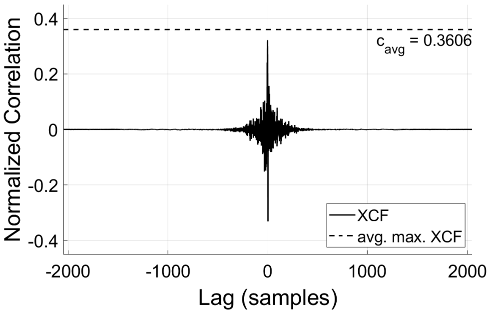
of two generated filters is computed and normalized by

$$\frac{1}{\sqrt{\Phi_{xx}[0]\Phi_{yy}[0]}}.$$

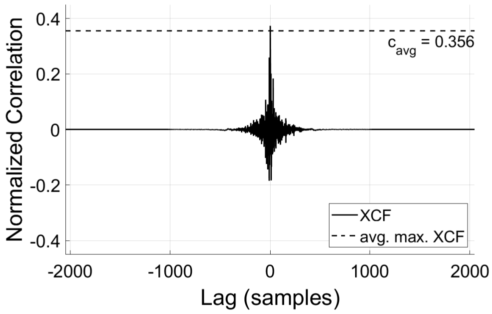
To get a general measure for decorrelation the average correlation coefficient is calculated over all combinations of generated filters as

$$c_{avg} = \frac{2}{(L-1)L} \sum_{k\in\{1,\cdots,L\}} \sum_{l\in\{k+1,\cdots,L\}} \max\left(|\Phi_{kl}|\right) \tag{7}$$

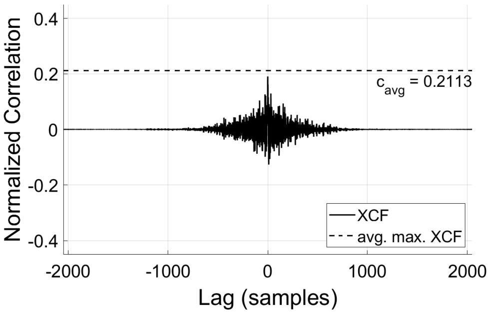
with $L = 19$ as the number of generated filters for each group delay curve. The number of unique channel combinations is $\frac{L(L+1)}{2}$. Excluding the auto correlation functions when $k = l$ leads to $\frac{(L-1)L}{2}$ as the normalization factor in equation (7).



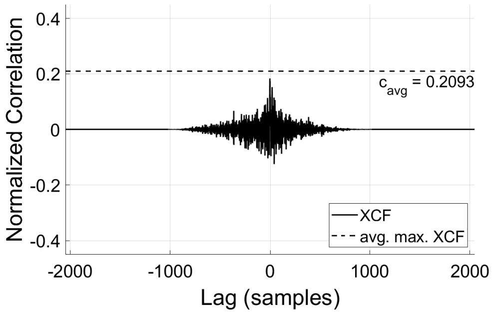
(a) del_1



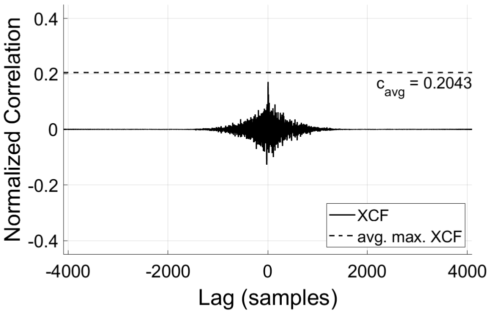
(b) $del_{1,onset}$



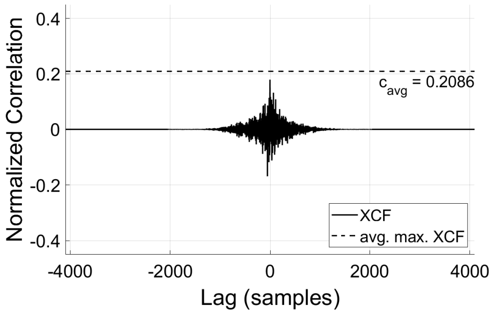
(c) del_2



(d) $del_{2,onset}$



(e) del_3



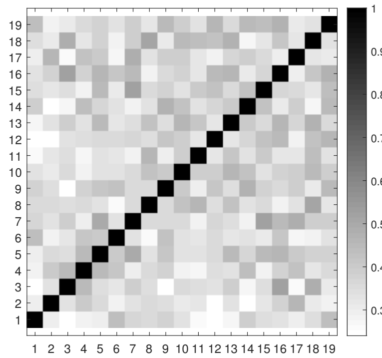
(f) $del_{3,onset}$

Figure 13 – Calculated correlation coefficients for group delay Φ_{kl} with $k = 5$ and $l = 10$ as example and the average maximum correlation coefficient c_{avg} over all channel combinations.

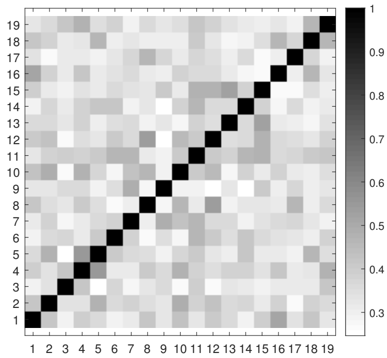
The cross correlation coefficient matrices

$$\mathbf{C} \in \mathbb{R}^{LxL} \text{ with } C_{k,l} = \max\left(|\Phi_{kl}|\right) \tag{8}$$

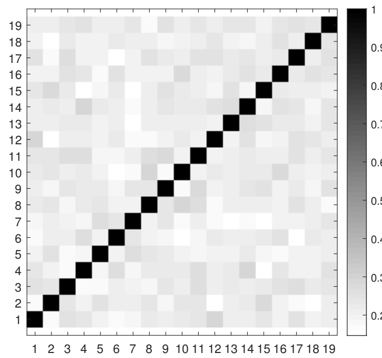
where $k, l = 1...L$ and $L = 19$ as the number generated filters are shown in figure 14.



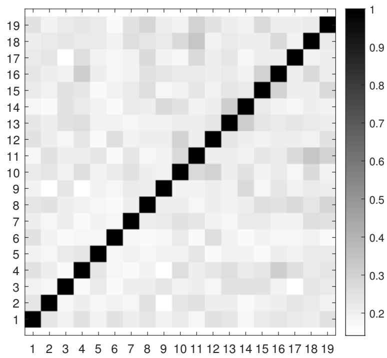
(a) Cross correlation del_1



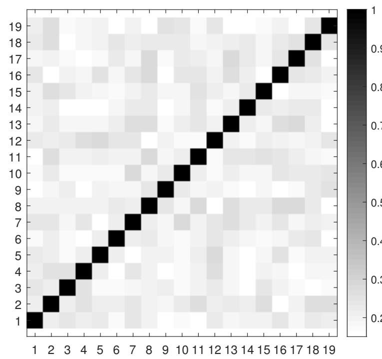
(b) Cross correlation $del_{1,onset}$



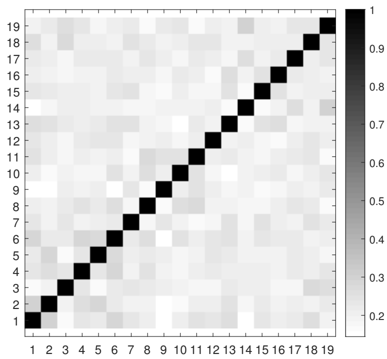
(c) Cross correlation del_2



(d) Cross correlation $del_{2,onset}$



(e) Cross correlation del_3



(f) Cross correlation $del_{3,onset}$

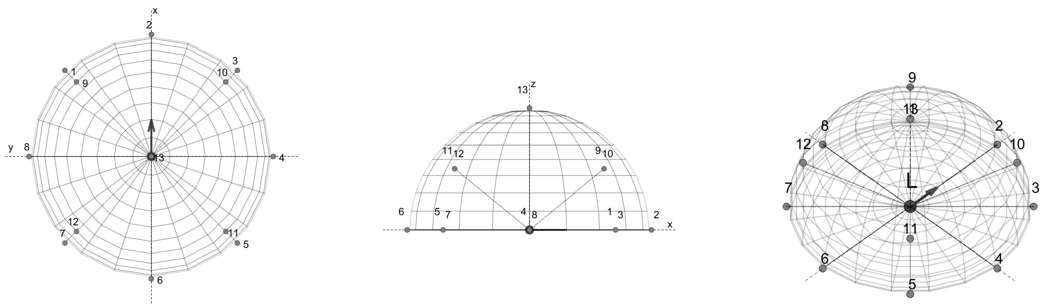
Figure 14 – The cross correlation matrices for group delay curves.

Figure 13 and 14 give a good comparison of the correlation properties of the filters generated from the selected group delay curves. The delay curves with greater values in the high frequency range yield lower correlation in theory, which is expected. The onset-shaping by removing the pre ringing does not seem to have significant impact on the correlation values.

5.2 Listening evaluation

5.2.1 Setup

The listening evaluation took place in the *Lehrstudio* (Room Nr.: IG1003010) in one of the IEM facilities. The room includes a sound system made up of 13 loudspeakers on the upper hemisphere and one sub woofer and the necessary equipment to drive the system. Figure 15 shows the loudspeaker layout and the listening position.



(a) View from above (b) Side view (c) Isometric view

Figure 15 – Loudspeaker directions and listener position. The listener is facing loudspeaker 2 as the arrow indicates. Refer to table 1.

Loudspeaker Nr.	Azimuth [°]	Elevation [°]
1	45	0
2	0	0
3	-45	0
4	-90	0
5	-135	0
6	-180	0
7	135	0
8	90	0
9	45	30
10	-45	30
11	-135	30
12	135	30
13	0	90

Table 1 – Azimuth and elevation of the loudspeakers on the hemisphere. Azimuth is the angle counter-clockwise on the x-y-plane. Elevation is measured from the x-y-plane towards the z-axis. The listener position **L** is the centre of the hemisphere.

5.2.2 Method

Using the method of phantom source widening described in section 4, a small scale evaluation was done of different settings with additional comparison to an established method.
The evaluation method uses a multi-stimulus design (MUSHRA-like). A set of 8 sound samples and a reference sound samples are presented on the user interface. The person is requested to order the sound samples regarding the posed question/hint using the sliders. The sliders represent a position on the relative rating scale. In addition to the visible reference a hidden reference is included in the 8 samples.
Table 3 lists the 8 settings. The initial order of the samples on the user interface is randomized and hidden from the listener.
Table 2 shows the questions/hints presented to the listener. Two audio sources were chosen: One music sample (20 seconds) and one speech sample (10 seconds). Each question/hint appeared twice for each audio source which results in 8 posed questions.
Movement/rotation of the head was allowed as it facilitates better width perception.

Question/Hint	Scale
Gesamteindruck der Klangqualitaet im Vergleich verschiedener Aufweitungsverfahren. Transientes Verhalten, etc.	*schlecht–gut*
Translation: Overall Sound quality comparing phantom source widening methods. Transients, etc.	bad–good
Relative Breite der Klangbeispiele.	*schmal–breit*
Translation: Relative width of sound samples.	narrow–wide

Table 2 – Overview of evaluation questions and data

ID	Encoding	Decoding
ref	Reference: Single mono source encoded with *AMBIX-ENCODER* [Kro14].	Rendered to 13 loudspeaker channels using the *AllRADecoder* [Rud19].
del_1	Phantom source widening plugin (sec. 4). Group delay curve: del_1 (fig. 12). Width 180°. Directional focus value 0.0.	
$del_{1,onset}$	Phantom source widening plugin (sec. 4). Group delay curve: del_1 (fig. 12). Width 180°. Directional focus value 0.0. Onset shaping $n_r = 1023$.	
del_2	Phantom source widening plugin (sec. 4). Group delay curve: del_2 (fig. 12). Width 180°. Dampening value 1.0.	
$del_{2,onset}$	Phantom source widening plugin (sec. 4). Group delay curve: del_2 (fig. 12). Width 180°. Directional focus value 0.0. Onset shaping $n_r = 1023$.	
del_3	Phantom source widening plugin (sec. 4). Group delay curve: del_3 (fig. 12). Width 180°. Directional focus value 0.0.	
$del_{3,onset}$	Phantom source widening plugin (sec. 4). Group delay curve: del_3 (fig. 12). Width 180°. Directional focus value 0.0. Onset shaping $n_r = 1023$.	
$ambi$	3-axis *ambix_widening_o5* [Kro14] + mono source at center at $-9dB$; Parameters: Table 4; Method: [ZFKC14, Sec. 4.3]	

Table 3 – The compared algorithms and settings. All sources have the main direction along the x-axis facing the listener and are encoded/decoded to/from 5th order Ambisonics.

Axis	Mod Depth	Mod T
x	90.72 deg	1.875 ms
y	90.72 deg	1.665 ms
z	90.72 deg	2.294 ms

Table 4 – Parameters of the *ambix_widening_o5* VST plugin [Kro14]. In addition a single mono source encoded at the source position at $-9dB$

5.2.3 Evaluation results

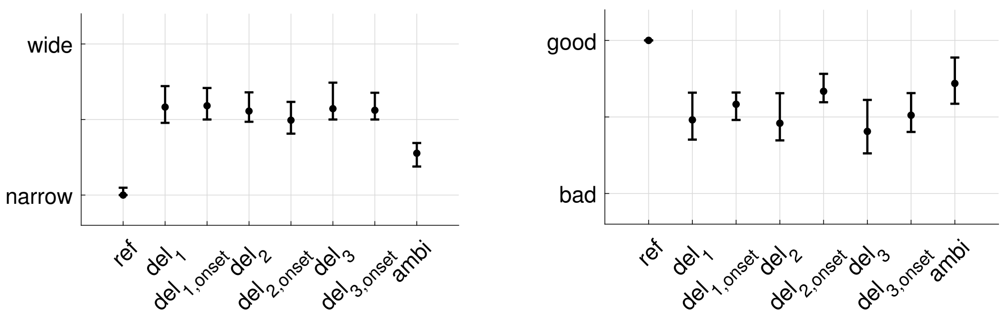
7 listeners took part in the evaluation with an average individual evaluation time of 20 minutes.

Width

The evaluation results in fig. 16a, 17a, 17c show no significant differences in width perception among the group delay based approach at different settings. The pairwise analysis of statistical significance was done with the *ANOVA* as well as *Kruskal-Wallis* method. Setting del_2 tends to be narrower. The results for the comparison of relative width were not as desired for the setting $ambi$ that appeared narrower than indented.

Sound Quality

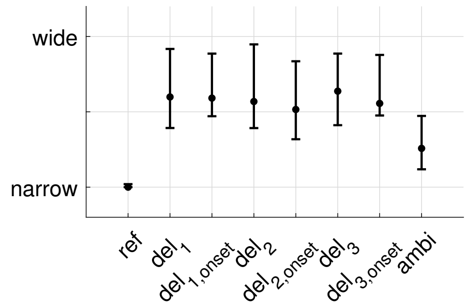
In perceived sound quality the tendency towards the fast-onset settings $del_{1,onset}$, $del_{2,onset}$, $del_{3,onset}$ compared to their slow-onset counterparts del_1, del_2, del_3 is visible and most pronounced at the pair $del_{2,onset}$, del_2. As expected the setting with group delay values beyond the proposed maximum values tends to be perceived as the worst of the group. Statistical significant differences (*ANOVA*, *Kruskal-Wallis*) exist at the pairing $del_{2,onset}$, del_3. Unexpectedly del_1, which in development was thought to be a better set of values, is not perceived as such and judging from figure 16b, it is save to say that $del_{2,onset}$ can be named as the configuration with the best perceived quality among the group delay based filters, on the same level as $ambi$.



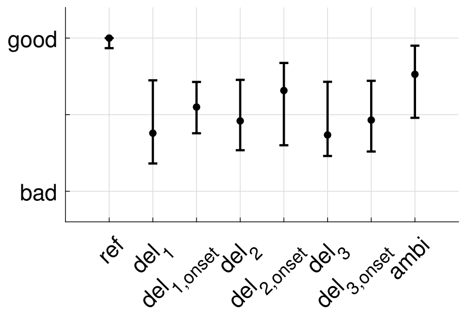
(a) Perceived width taking into account speech and music

(b) Perceived quality taking into account speech and music

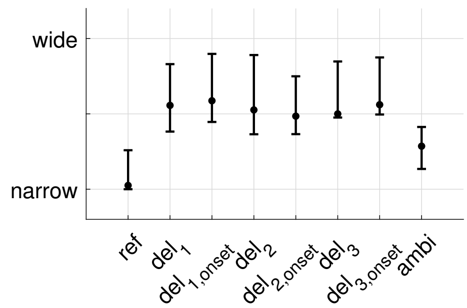
Figure 16 – Results of listening evaluation. Mean and confidence intervals calculated with `CI2p.m` from the IEM OpenData Archive [IEM19]



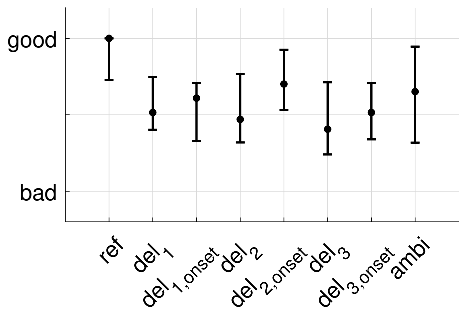
(a) Perceived width taking into account only music



(b) Perceived quality taking into account only music



(c) Perceived width taking into account only speech



(d) Perceived quality taking into account only speech

Figure 17 – Results of listening evaluation. Mean and confidence intervals calculated with `CI2p.m` from the IEM OpenData Archive [IEM19]

Discussion

Verbal feedback from the subjects leads to the conclusion that the perception of differences is significantly easier when confronted with music samples as opposed to speech samples. Interpreting the results leads to the assumption that the in [CDA18] proposed maximum values are correct. The decorrelation is more effective in theory (sec. 5.1) as well as practice (sec. 5.2.3) when including greater group delay values in the higher frequency range.
Judging from these results it is necessary to think about a more thorough listening evaluation. Especially the comparison with the *ambix* approach has to be redone. Using the results of this evaluation, one can limit the parameters for the group delay based approach as $del_{2,onset}$ proved more successful and should be used in further tests.

The following points are proposed as a focus in a follow-up evaluation:

Source Material: carefully selected audio source material (recordings from IEM OpenData Archive [IEM19])

Comparison of apertures: behavior of listening experience at different aperture settings.

***AMBIX-WIDENING*:** correct parameterization for comparison at *different* aperture settings.

Stability: evaluation of the stability of the listening experience (Sweet-spot size)

Artifacts: evaluation of types of artifacts at different apertures and settings.

6 Conclusion and future work

The result of this project is a system to create an arbitrary number of impulse responses to create decorrelated signals with given random group delay curves. These signals can be used to achieve high quality phantom source widening, which was shown and tested for a proposed multi-decorrelated-source constellation encoded in 3D Ambisonics.
Using group delay based all-pass filters with or without pre ringing gives good decorrelation that permits impulse response generation at low computational effort. The filter responses are relatively long when compared to deterministic decorrelation methods, which can be seen as a drawback especially in application with a large number of filters, however not necessarily so when fast convolution methods are available.
Further listening tests should certainly be a useful extension of this project work in particular for a comprehensive evaluation in a larger listening space (see sec. 5.2.3). Topics interesting for further investigation are the stability and size of the sweet area where the diffuse sound field is not collapsing into single loudspeakers as well as a more thorough test on different forms of maximum group delay curves.
The VST plugin developed in this work is still under tests and will be added to the IEMPlugin suite as soon as the results and performance have been checked to be all-satisfactory.

References

[CDA18] E. K. Canfield-Dafilou and J. S. Abel, "A group delay-based method for signal decorrelation," in *Convention Paper, 144th Conv. Audio Eng. Soc.*, Milan, Italy, 2018.

[FFM15] A. Franck, F. M. Fazi, and F. Melchior, "Optimization-based reproduction of diffuse audio objects," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 2015.

[IEM19] *IEM OpenData Archive*. IEM, 2019, https://opendata.iem.at/.

[Ken95] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, vol. 19, no. 24, 1995.

[Kro14] M. Kronlachner, "Plug-in suite for mastering the production and playback in surround sound and ambisonics," in *Gold award at the AES Students Design Competition 136th AES Convention Berlin*, Berlin, Germany, 2014.

[PB04] G. Potard and I. 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*, Naples, Italy, 2004.

[Pes11] H. Pessentheiner, "Audio signal decorrelation for phantom source widening," Seminar Project, Institut für Elektronische Musik und Akustik, Kunstuni Graz, Technical University Graz, Graz, A, 2011.

[Rud19] D. Rudrich, *IEM Plugin Suite*. IEM, 2019, https://plugins.iem.at/.

[SS18] T. Schmele and U. Sayin, "Controlling the apparent source size in ambisonics using decorrelation filters," in *Convention Paper, 142nd Conv. Audio Eng. Soc.*, Tokyo, Japan, 2018.

[WFO92] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation based on decorrelation," Technical Report, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, 1992.

[ZF17] F. Zotter and M. Frank, "Phantom source widening by filtered sound objects," in *Convention Paper, 142nd Conv. Audio Eng. Soc.*, Berlin, Germany, 2017.

[ZFKC14] 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*, Berlin, Germany, 2014.

[ZFMS11] F. Zotter, M. Frank, G. Marentakis, and A. Sontacchi, "Phantom source widening with deterministic frequency dependent time delays," in *Proc. of the 14 th Int. Conference on Digital Audio Effects (DAFx-11)*, Paris, France, 2011.