Optimization-Based Reproduction of Diffuse Audio Objects

Andreas Franck[1], Filippo Maria Fazi[1], Frank Melchior[2]

July 19, 2015

[1]Institute of Sound and Vibration Research, University of Southampton, Southampton, Hampshire, SO17 1BJ, UK[*]

[2]BBC Research & Development, Dock House, MediaCityUK, Salford, M50 2LH, UK

Abstract

The creation of a diffuse sound event from a single audio signal is an important signal processing task, for instance in spatial audio reproduction or audio coding. Current algorithms based on decorrelation filters or frequency-dependent panning typically cause artifacts due to transients or time-domain aliasing. In this paper, we propose an optimization-based approach to diffusion that creates a set of filters to approximate a desired distribution of frequency-dependent propagation directions to create the perception of a diffuse sound field with a multi-channel audio system. Thus, the diffusion can be optimally adapted to a specific reproduction scenario. In addition, the transient response can be purposefully improved by imposing constraints on the time-domain filter coefficients.

1 Introduction

The creation of diffuse auditory events are an important signal processing task in spatial audio reproduction [1]–[11] as well as in audio coding, e.g., [12], [13]. Typical applications are the transmission of a reverberant sound field in a single audio signal which is turned into a diffuse auditory event by applying an appropriate number of decorrelation filters. For this reason this technique is also termed decorrelation. Another important application is the creation of wide auditory events. This is equivalent to reproduction of a sound source with a perceptual spatial extent.

Early approaches to achieve diffuse sound events date back to [1], [2], when research has been aimed specifically to increase the width of stereo images [3]. Most of the approaches make use of a set of filters to generate decorrelated audio signals. An overview of the use of decorrelation and its perceptual consequences can be found in [4], where the decorrelation filters are formed by specifying unit-magnitude frequency responses with random phase and performing an inverse Discrete Fourier Transform (DFT) to obtain the filter coefficients. However, phase randomization may cause severe artifacts: magnitude errors between DFT bins [5], smearing of transients which might also cause time-domain aliasing [14], [6], [7] as well as coloration of the reproduced signal [3], [15]–[17].

A different approach to diffusion, which also intends to overcome the artifacts of phase randomization, is to vary the incident direction of sound as a function of frequency. This can be either implemented by a frequency-dependent panning (e.g., [8], [9], [18]), or by controlling the frequency-dependent inter-channel time difference (ICTD) [7], [10], [11], [19] using specifically designed allpass filters. The fact that a frequency-dependent variation of the perceived direction of incidence leads to the perception of an extended or diffuse sound event is corroborated by psychoacoustic experiments by Blauert and Lindemann [20]. However, these approaches might still introduce artifacts like transient smearing, time-domain aliasing and coloration, because they do not allow for a direct control of the time-domain filter behavior. Moreover, these techniques are often limited to two channels and difficult to generalize to multichannel scenarios.

2

The methods described so far can be modeled as a set of time-invariant filters applied to the audio signal. In contrast, dynamic decorrelation techniques modify the signals in a time-varying manner, e.g., by time-variant IIR filtering [4], phase randomization of the output signals [14], [21], or time-variant frequency-dependent panning [18]. However, listening experiments suggest that the perceived quality of these approaches is often inferior to that of time-invariant techniques [14], [18], [22]. Audible artifacts due to transient smearing are especially critical for certain signal types, e.g., applause [23]. Several state-of-the-art algorithms therefore perform a separation and a distinct processing of the transient part of a signal [24], [25].

In this paper, we propose a novel approach to the design of time-invariant diffusion filters that is based on convex optimization. Starting from a suitable definition of a diffuse sound event, which relates it to a uniform distribution of the sound propagation direction, we define a deterministic frequency-dependent intensity vector as the target function. The design of the filter coefficients is then stated as a convex optimization problem, where the objective function is to find the best approximation of this target intensity vector. Compared to existing techniques, this approach is able to account for arbitrary unevenly spaced multi-loudspeaker setups. Moreover it allows a degree of control over the transient behavior by imposing constraints on the time-domain response of the diffusion filters. Likewise, the amount of phasiness, an artifact causing image instability and listening discomfort, can be controlled by imposing limits on the reactive component of the intensity. The presented technique is not intended as an alternative to transient-separation approaches, but rather as a potential component of such algorithms which might lower the requirements of the transient separator, decrease the computational complexity, and improve the overall quality.

The remainder of this paper is outlined as follows. Section 2 introduces the chosen definition of diffusion and the acoustic propagation models that defines the relation between the filter coefficients and the reproduced sound field. The filter design problem is stated and translated into a convex optimization problem in Section 3. The performance of the obtained solution is evaluated and compared to existing decorrelation techniques in Section 4, and the findings of this paper are summarized in Section 5.

2 Acoustic Model

In this section, we describe the desired diffuse sound event and the acoustic transfer functions from the diffusion filters to the quantities describing the resulting sound event.

2.1 Modeling the Desired Diffuse Sound Event

There are different definitions of a diffuse sound event, including a) constant energy density in the region of interest; b) a superposition of an infinite number of plane waves with random directions and phases; c) zero time-averaged acoustic intensity , and d) equally probable
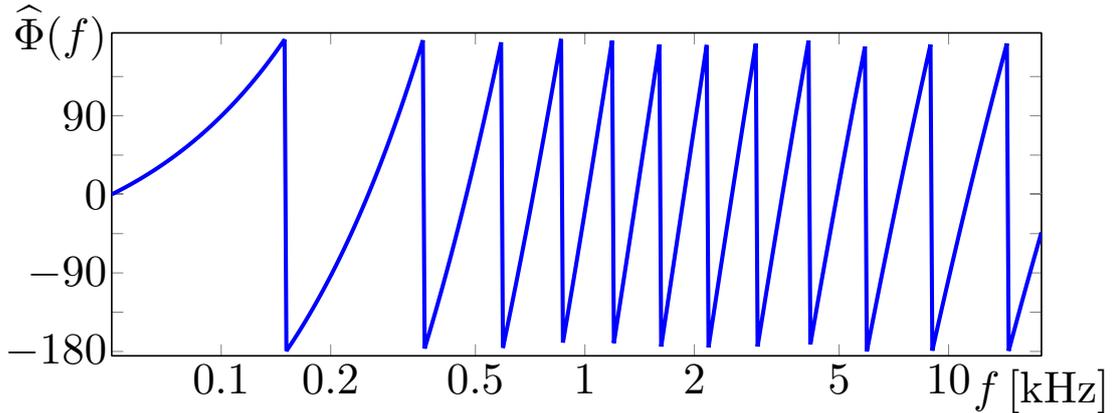
3

Figure 1: Desired frequency-dependent propagation direction.

propagation directions [6], [21], [26], [27]. To synthesize a diffuse sound from a single audio signal using time-invariant diffusion filters, definition d) is the most appropriate. In particular, the zero time-averaged intensity criterion c), which is used as the diffuseness estimator in Spatial Impulse Rendering (SIRR) [28] and Directional Audio Coding (DirAC) [6] is not applicable for optimization-based diffusion filter design, because, for instance, a sound event consisting of coherent plane waves impinging from all directions also fulfills this criterion (cf. [21]).

In this paper, we therefore propose to generate diffuseness by varying the propagation direction as a function of frequency. To this end, we use the acoustic intensity [29], [30] to describe the propagation direction. In the temporal frequency domain, the intensity is a three-dimensional vector

$$\mathbf{I}(f) = \frac{1}{2}P(f)\mathbf{U}(f)^*, \tag{1}$$

where $P(f)$ and $\mathbf{U}(f)$ denote the sound pressure and the particle velocity at frequency f, respectively. The real-valued part of the velocity vector is termed the *active intensity* $\mathbf{I}_a(f)$

$$\mathbf{I}_a(f) = \mathrm{Re}\{\mathbf{I}(f)\}, \tag{2}$$

and represents the propagating part of the sound field with $\mathbf{I}_a(f)$ pointing into the direction of sound propagation.

To create a diffuse sound event, the desired direction of $\mathbf{I}_a(f)$ is varied as a function of frequency. The desired azimuthal direction $\widehat{\Phi}(f)$ used in this paper is depicted in Fig. 1. As motivated in [5], [20], this function should also account for the frequency resolution of the human ear. Therefore the rate of the angular variation is chosen to be proportionally to the critical bandwidth $\mathrm{CB}_c(f)$, e.g., [31]

$$\widehat{\Phi}(f) = 2\pi \left(\int_{f_0}^{f} \frac{r}{\mathrm{CB}_c(f')} \, \mathrm{d} f' \bmod 1 \right), \tag{3}$$

4

where r is a proportionality factor to scale the direction change rate.

The imaginary part of $\mathbf{I}(f)$

$$\mathbf{I}_{\mathrm{r}}(f) = \mathrm{Im}\{\mathbf{I}(f)\} \tag{4}$$

represents the *reactive intensity*, i.e., the locally oscillating, non-propagating part of the sound field [21], [29], [32]. In [33], this imaginary part is related to the level of phasiness [3], a subjective defect of multi-loudspeaker reproduction methods which is associated with listening discomfort and image instability in case of head movements [34]. Thus, using the magnitude of the reactive intensity to quantify the phasiness provides a means to control this artifact.
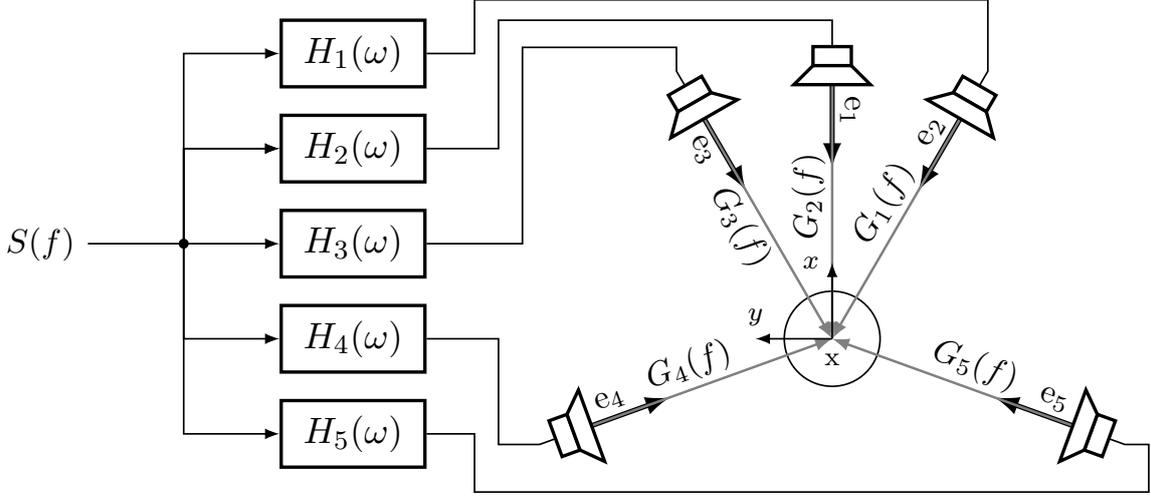
2.2 Acoustic Transfer Function

Figure 2: The transfer functions of the acoustic model.

The acoustic propagation model describes the transfer function from the source signal $S(f)$ to the acoustic intensity at the receiver position $\mathbf{x}$ as depicted in Fig. 2. The source signal $S(f)$ is filtered with a set of discrete-time filters $H_l(\omega)$ to form the driving signals $X_l(f)$, $1 \leq l \leq L$ for the set of L loudspeakers. Here, ω denotes the normalized angular frequency $\omega = 2\pi f/f_s$ with sampling frequency f_s. Within this paper, the filters are modeled as causal general FIR filters whose frequency responses are parameterized by sets of coefficients $b[l, n]$, $0 \leq n < N$

$$H_l(\omega) = \mathrm{e}^{-\mathrm{j}\omega N_0} \sum_{n=0}^{N-1} b[l,n] F_n(\omega) \quad \text{with} \tag{5}$$

$$F_n(\omega) = \mathrm{e}^{-\mathrm{j}\omega(n-N_0)}, \tag{6}$$

where N is the length of the FIR filters and N_0 controls the time shift of the basis functions. In this paper, we use $N_0 = (N-1)/2$, which makes the basis functions conjugate symmetric with

respect to $n = (N-1)/2$.

The relation between the signal of loudspeaker l to its contribution to the sound pressure at the listener position, denoted $P_l(f)$, is modeled by the transfer function $G(f)$, which can be either measured or simulated. In this paper, we use a free-field plane wave propagation model for a receiver position $\mathbf{x}$ in the far field given by

$$G_l(f) = \mathrm{e}^{-2\pi\mathrm{j}\frac{f}{c}\mathbf{x}\cdot\mathbf{e}_l}, \tag{7}$$

where c denotes the speed of sound. The unit vector $\mathbf{e}_l$ represents the direction of sound propagation for loudspeaker l. Using this plane wave assumption, the particle velocity vector $\mathbf{U}_l(f)$ due to loudspeaker l is related to the sound pressure $P_l(f)$ by [30]

$$\mathbf{U}_l(f) = \frac{1}{Z_0}P_l(f)\mathbf{e}_l, \tag{8}$$

where Z_0 denotes the characteristic acoustic impedance of the medium.

The total sound pressure and intensity at the listener position $\mathbf{x}$ are formed by summing the loudspeaker contributions

$$P(f) = \sum_{l=1}^{L} P_l(f), \quad \mathbf{U}(f) = \sum_{l=1}^{L} \mathbf{U}_l(f), \tag{9}$$

and the resulting intensity vector is formed according to (1). Combining (5) and (9), the sound pressure $P(f)$ at $\mathbf{x}$ can be expressed as a function of the time-domain filter coefficients $b[l,n]$, that is,

$$P(f) = \sum_{l=1}^{L}\sum_{n=0}^{N-1} H_{l,n}^{P}(f)b[l,n] \quad \text{with} \tag{10}$$

$$H_{l,n}^{p}(f) = G_l(f)F_n(2\pi f/f_s). \tag{11}$$

Likewise, the velocity is formed by combining (5), (8) and (9)

$$\mathbf{U}(f) = \sum_{l=1}^{L}\sum_{n=0}^{N-1} \mathbf{H}_{l,n}^{\mathbf{U}}(f)b[l,n] \quad \text{with} \tag{12}$$

$$\mathbf{H}_{l,n}^{\mathbf{U}}(f) = \frac{1}{Z_0}\mathbf{e}_l G_l(f)F_n(2\pi f/f_s). \tag{13}$$

3 Optimization Method

Based on the acoustic model above, the design of the filter coefficients $b[l,n]$ to yield the best approximation of a desired frequency-dependent propagation direction is transformed into a

convex optimization problem. To this end, the coefficients $b[l, n]$ are arranged into a column vector

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_L \end{bmatrix}^T \tag{14}$$

with $\mathbf{b}_l = \begin{bmatrix} b[l, 0]\, b[l, 1] \, \cdots \, b[l, N-1] \end{bmatrix}$. The frequency interval of interest is discretized to a vector of K frequencies $W = \begin{bmatrix} f_1 \, f_2 \, \cdots \, f_K \end{bmatrix}$. In this way, discretized representations of $P(f)$ and $\mathbf{U}(f)$ are obtained as

$$\widetilde{\mathbf{P}} = \begin{bmatrix} P(f_1)\, P(f_2) \, \cdots \, P(f_K) \end{bmatrix}^T \qquad\qquad = \widetilde{\mathbf{H}}^P \mathbf{b} \tag{15}$$

$$\widetilde{\mathbf{U}} = \begin{bmatrix} \mathbf{U}(f_1)^T \, \mathbf{U}(f_2)^T \, \cdots \, \mathbf{U}(f_K)^T \end{bmatrix}^T \qquad\qquad = \widetilde{\mathbf{H}}^{\mathbf{U}} \mathbf{b} . \tag{16}$$

The elements of the transfer matrices $\widetilde{\mathbf{H}}^P$ and $\widetilde{\mathbf{H}}^{\mathbf{U}}$ are determined by (11) and (13), respectively. Note that $\mathbf{U}(f)$ (12) and $\mathbf{H}^{\mathbf{U}}_{l,n}(f)$ (13) are three-element column vectors. Thus $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{H}}^{\mathbf{U}}$ are formed by concatenating these elements vertically.

The desired sound pressure $\widehat{\mathbf{P}}$ can be chosen depending on the application, e.g., $|\widehat{P}(f)| = 1$ for a flat magnitude response. The target particle velocity vector $\widehat{\mathbf{U}}$ is formed by discretizing $\widehat{\mathbf{\Phi}}(f)$ (3), scaling it to a desired velocity magnitude $|\widehat{\mathbf{U}}(f)|$, and arranging it into a column vector with the same layout as $\widetilde{\mathbf{U}}$. For the examples in this paper, we use $|\widehat{\mathbf{U}}| = 1/Z_0\widehat{\mathbf{P}}$. As both $P(f)$ and $\mathbf{U}(f)$ are complex-valued in general, there are many possible choices for specifying $\widehat{P}(f)$. In this paper, we restrict the sound pressure to be real-valued, corresponding to a zero-phase frequency response. In this way, the phase of $\mathbf{U}(f)$ and of the complex intensity $\mathbf{I}(f)$ are identical. Consequently, the particle velocity components contributing to the active and reactive intensity are determined by $\mathrm{Re}\{\widetilde{\mathbf{U}}\} = \mathrm{Re}\{\widetilde{\mathbf{H}}^{\mathbf{U}}\}\mathbf{b}$ and $\mathrm{Im}\{\widetilde{\mathbf{U}}\} = \mathrm{Im}\{\widetilde{\mathbf{H}}^{\mathbf{U}}\}\mathbf{b}$, respectively.

Using these discretized representations, the optimization problem can be stated in matrix form as

$$\underset{\mathbf{b}}{\mathrm{argmin}} \left\| \mathrm{Re}\{\widetilde{\mathbf{H}}^{\mathbf{U}}\}\mathbf{b} - \widehat{\mathbf{U}} \right\|_p \quad \text{subject to } \widetilde{\mathbf{H}}^P \mathbf{b} = \widehat{\mathbf{P}} , \tag{17}$$

for a suitable norm L_p, e.g., $p = 2$ for a least-squares or $p = \infty$ for a Chebyshev (minimax) norm. In addition, the reactive part of the sound field can be controlled by adding the inequality constraint

$$\ldots \text{ subject to } \left\| \mathrm{Im}\{\widetilde{\mathbf{H}}^{\mathbf{U}}\}\mathbf{b} \right\|_p \leq \varepsilon_r , \tag{18}$$

to the optimization problem. The constant ε_r is chosen, for instance, as a percentage of the propagating energy [3].

To improve filter stability and robustness, the optimization problem is augmented with a

regularization term

$$\underset{\mathbf{b}}{\operatorname{argmin}} \left\| \operatorname{Re}\{\widetilde{\mathbf{H}}^{\mathbf{U}}\}\mathbf{b} - \widehat{\mathbf{U}} \right\|_p + \|\mathbf{\Gamma b}\|_p \quad \text{subject to } [\,] , \tag{19}$$

where $\mathbf{\Gamma}$ is a typically a diagonal matrix. Two weightings are used in this paper: a constant diagonal value β which uniformly penalizes all filter coefficients, and a progressive weighting that imposes a larger penalty on filter coefficients $b[l, n]$ with larger n

$$\beta(l, n) = \left(\frac{n}{M - 1} \right)^k , \quad k \geq 0 . \tag{20}$$

This progressive weighting is applied to make the filter impulse responses more compact, thus improving the transient response.

The optimization specifications (17)–(19) have been formulated and solved using CVX, a software for modeling and solving convex problems [35], [36].

4 Evaluation

In this section the proposed diffusion filter design method is evaluated and compared to existing decorrelation techniques using simulations. A five-channel loudspeaker configuration according to ITU recommendation ITU-R BS.775-2 as depicted in Fig. 2 is chosen for the simulations. The proposed design method is used with a discrete logarithmically spaced 2048-element frequency grid $f \in [200\,\text{Hz} \dots 18\,\text{kHz}]$ to generate FIR filters of length $N = 256$. A non-uniform regularization according to (20) with $k = 1$ is applied to control the transient response. The frequency-dependent propagation direction was chosen according to (3) with scaling factor $r = 0.5$ (see Fig. 1).

For comparison, we selected an allpass FIR decorrelation filter design with randomized phase [4] and a deterministic frequency-dependent amplitude panning according to [9], using the same filter length $N = 256$ as for the proposed technique. Likewise, the frequency-dependent panning directions are chosen identically to the proposed method.

The frequency-dependent propagation directions $\Phi(f)$, i.e., the azimuth of $\mathbf{U}_a(f)$, generated by the three methods are shown in Fig. 3(a). The random-phase allpass design achieves rapidly changing propagation directions for all but the lowest frequencies. The relatively constant direction up to about $150\,\text{Hz}$ common to all designs can be attributed to the finite resolution of the filter. Thus, it can be resolved by increasing the filter length at the expense of a higher computational complexity. For medium frequencies up to $2\,\text{kHz}$, the propagation directions generated by amplitude panning are mainly from front to back (close to $\pm 180°$). This can be explained by the non-uniform spacing of the chosen loudspeaker setup, which lacks speakers in the far rear. In contrast, the proposed optimization-based filter design creates a controlled, relatively rapid variation of the propagation direction starting at about $200\,\text{Hz}$, covering all
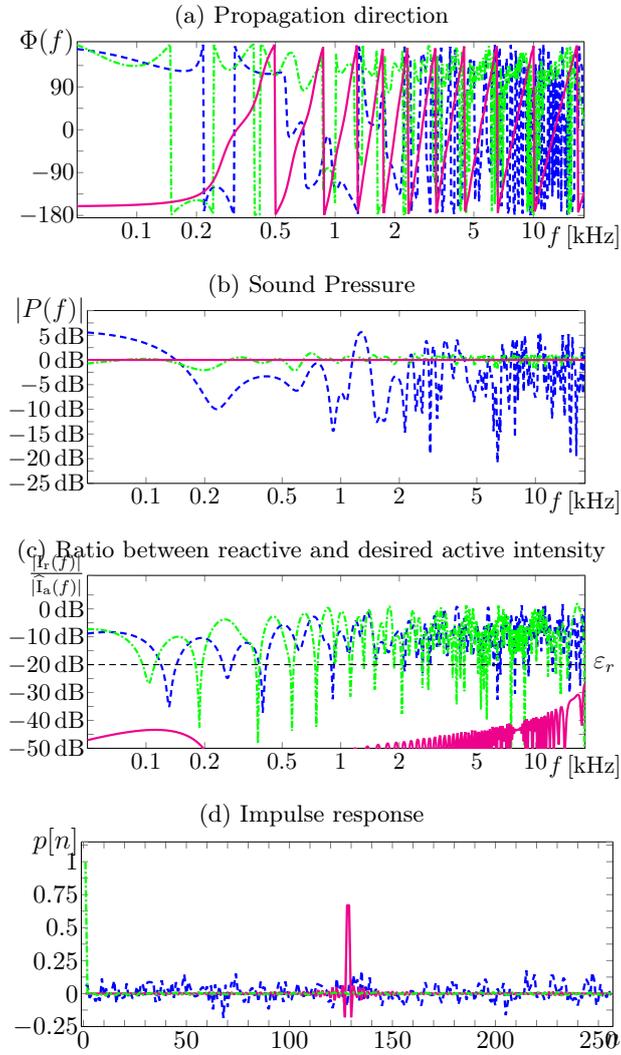
8

Figure 3: Sound propagation direction, sound pressure, level of reactive intensity, and impulse response at the central listener position. - - - randomized-phase allpass, -·-·- frequency-dependent panning, ——— optimization-based design.

9

directions relatively evenly despite the non-uniform loudspeaker distribution.

The sound pressure magnitude at the listener position $|P(f)|$ is depicted in Fig. 3(b). For this example assuming ideal plane wave sources, the desired sound pressure response is flat $|\widehat{P}(f)| = 1$. For comparison, the filter sets for the three designs are normalized to equal sound energy at position $\mathbf{x}$. The random-phase allpass shows relatively large variations over frequency, which can be explained by the frequency-dependent random constructive and destructive interference patterns of the loudspeaker contributions. The boost at very low frequencies is due to the constructive interference at the lowest DFT frequency and the limited frequency resolution of this example filter. Again, this can be alleviated by increasing the DFT length (see, e.g., [14]). Diffusion based on amplitude panning exhibits only minor, narrowband magnitude response deviations, which are likely to be perceptually irrelevant in most cases. In contrast, for the proposed design method, the sound pressure at the listener position is incorporated as an optimization constraint. Consequently, it achieves a constant sound pressure magnitude within a very small tolerance over the entire frequency range of interest. Conversely, this implies that this design method can be used to impose a desired magnitude response for diffuse sound events.

As outlined in Sec. 2.1, the reactive intensity $\mathbf{I}_r(f)$ represents the locally oscillating, non-propagating part of the sound field, which is commonly related to the sensation of phasiness. The ratio between the magnitude $|\mathbf{I}_r(f)|$ and the desired active velocity magnitude $|\widehat{\mathbf{I}}_a(f)|$ is shown in Fig. 3(c) for the three considered methods. It can be seen that the allpass FIR decorrelators generate a relatively high level of reactive intensity for most of the frequency range. This can be explained by the superposition of phase-incoherent signals. The technique based on amplitude panning creates similar levels of reactive intensity, which can be attributed to phase variations introduced by the inverse DFT to generate FIR filters from amplitude panning weights. In this example, the proposed design method yields very low levels of reactive intensity, because the obtained filters have a nearly linear phase. For design specifications which do not have this property, it is possible to control the amount of reactive energy by adding an inequality constraint to the optimization problem. The limit $\varepsilon_r = -20\,\text{dB}$ used in the example is indicated in Fig. 3(c).

To evaluate the transient behavior, the impulse responses of the sound pressure at the central listening position $p[n]$ are depicted in Fig. 3(d). The impulse response of the randomized-phase allpass design shows that the signal amplitude is distributed over long time interval, causing a time smearing of transients that is typical for this class of decorrelators. In contrast, the amplitude panning and the proposed optimization-based design show relatively localized impulses, corresponding to a good preservation of transients. For the proposed design, the impulse is located at about $N/2$ due to the choice of N_0 (see (5)). It has been observed that varying this parameter enables a trade-off between the filter latency and the level of reactive intensity generated by the proposed diffusion filters.

10

5 Conclusion

This paper proposes an optimization-based approach to synthesize diffuse sound events from single audio signals over multi-loudspeaker setups with arbitrary geometry. To this end, we optimally design a set of FIR filters to vary the sound field propagation direction over frequency in order to achieve the perception of a diffuse sound event.

This first investigation of this approach shows several advantages over existing techniques. Firstly, it enables a more even distribution of propagation directions over frequency, especially for non-uniform loudspeaker arrangements. Secondly, it enables a better control over the time-domain properties of the diffusion filters in order to improve the quality of transient signals. Thirdly, controlling the amount of reactive intensity provides a means to limit phasiness, a subjective artifact that degrades the quality of multi-loudspeaker reproduction. Incorporating these aspects into the filter design optimization process enables application-specific trade-offs between these criteria.

Future research will focus on exploring the filter design options, suitable specifications for the frequency-dependent propagation distribution, and subjective evaluations of the technique.

6 References

[1] M. R. Schroeder, "An artificial stereophonic effect obtained from a single audio signal," *J. Audio Eng. Soc.*, vol. 6, no. 2, pp. 74–79, Apr. 1958.

[2] R. Orban, "A rational technique for synthesizing pseudo-stereo from monophonic sources," *J. Audio Eng. Soc.*, vol. 18, no. 2, pp. 157–164, Apr. 1970.

[3] M. A. Gerzon, "Signal processing for simulating realistic stereo images," in *Proc. 93rd AES Convention*, San Francisco, CA, USA, Oct. 1992.

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

[5] M. Boueri and C. Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *Proc. 117th AES Convention*, San Francisco, CA, USA, Oct. 2004.

[6] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, Jun. 2007.

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

[8] V. Pulkki, M.-V. Laitinen, and C. Erkut, "Efficient spatial sound synthesis for virtual worlds," in *Proc. AES 35th Int. Conf.*, London, UK, Feb. 2009.

[9] M.-V. Laitinen, T. Pihlajamäki, C. Erkut, and V. Pulkki, "Parametric time-frequency representation of spatial sound in virtual worlds," *ACM Transactions on Applied Perception*, vol. 9, no. 2, 8:1–8:20, Jun. 2012.

[10] J.-W. Choi, "Source-width extension technique for sound field reproduction systems," in *Proc. AES 52nd Int. Conf.*, Guildford, UK, Sep. 2013.

[11] F. Zotter and M. Frank, "Efficient phantom source widening," *Archives of Acoustics*, vol. 38, no. 1, pp. 27–37, 2013.

[12] J. Engdegård, H. Purnhagen, J. Rödén, and L. Liljeryd, "Synthetic ambience in parametric stereo coding," in *Proc. 116th AES Convention*, Berlin, Germany, May 2004.

[13] C. Faller, "Parametric multichannel audio coding: Synthesis of coherence cues," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 1, pp. 299–310, Jan. 2006.

[14] V. Pulkki and J. Merimaa, "Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests," *J. Audio Eng. Soc.*, vol. 54, no. 1/2, pp. 3–20, 2006.

[15] M. Brüggen, "Sound coloration due to reflections and its auditory and instrumental compensation," PhD thesis, Ruhr-Universität, Bochum, Germany, 2001.

[16] A. Raake, *Speech Quality of VoIP — Assessment and Prediction.* Chichester, West Sussex, UK: John Wiley & Sons Ltd., 2006.

[17] A. M Salomons, "Coloration and binaural decoloration of sound due to reflections," Dissertation, Delft University, Delft, The Netherlands, 1995.

[18] T. Pihlajamäki, O. Santala, and V. Pulkki, "Synthesis of spatially extended virtual sources with time-frequency decomposition of mono signals," *J. Audio Eng. Soc.*, vol. 62, no. 7/8, pp. 467–484, Jul. 2014.

[19] F. Zotter, M. Frank, M. Kronlachner, and J.-W. Choi, "Efficient phantom source widening and diffuseness in ambisonics," in *Proc. EAA Joint Symp. Auralization and Ambisonics*, Berlin, Germany, Mar. 2014.

[20] J. Blauert and W. Lindemann, "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *J. Acoust. Soc. Am.*, vol. 79, no. 3, pp. 806–813, 1986.

[21] J. Merimaa, "Analysis, synthesis, and perception of spatial sound — binaural localization modeling and multichannel loudspeaker reproduction," PhD thesis, Helsinki University of Technology, Espoo, Finland, Apr. 2006.

[22] G. Potard and I. Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays," in *Proc. 7th Int. Conf. Digital Audio Effects (DAFx'04)*, Naples, Italy, Oct. 2004, pp. 280–284.

[23] M.-V. Laitinen, F. Küch, S. Disch, and V. Pulkki, "Reproducing applause-type signals with directional audio coding," *J. Audio Eng. Soc*, vol. 59, no. 1/2, pp. 29–43, Jan. 2011.

[24] R. Penniman, "A general-purpose decorrelator algorithm with transient fidelity," in *Proc. 137th AES Convention*, Los Angeles, CA, USA, Oct. 2014.

[25] A. Kuntz, S. Disch, T. Bäckström, and J. Robilliard, "The transient steering decorrelator tool in the upcoming MPEG unified speech and audio coding standard," in *Proc. 131st AES Convention*, New York, NY, USA, Oct. 2011.

[26] F. Jacobsen, "The diffuse sound field," Acoustics Laboratory, Technical University of Denmark, Lyngby, Denmark, Technical Report 27, 1979.

[27] H. Nélisse and J. Nicolas, "Characterization of a diffuse field in a reverberant room," *J. Acoust. Soc. Am.*, vol. 101, no. 6, pp. 3517–3524, Jun. 1997.

[28] J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: Analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, Dec. 2005.

[29] F. Fahy, *Sound Intensity*, 2nd edition. London, UK: Spon Press, 1995.

[30] J. Merimaa, "Energetic sound field analysis of stereo and multichannel loudspeaker reproduction," in *Proc. 123th AES Convention*, New York, NY, USA, Oct. 2007.

[31] E. Zwicker and E. Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," *J. Acoust. Soc. Am.*, vol. 68, no. 5, pp. 1523–1525, Nov. 1980.

[32] G. Schiffrer and D. Stanzial, "Energetic properties of acoustic fields," *J. Acoust. Soc. Am.*, vol. 96, no. 6, pp. 3645–3653, 1994.

[33] M. A. Gerzon, "General metatheory of auditory localisation," in *Proc. 92nd AES Convention*, Vienna, Austria, Mar. 1992.

[34] C. Guastavino, V. Larcher, G. Catusseau, and P. Boussard, "Spatial audio quality evaluation: Comparing transaural, ambisonics and stereo," in *Proc. 13th Int. Conf. Auditory Display*, Montreal, Canada, Jun. 2007, pp. 53–59.

[35] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., Springer, 2008, pp. 95–110.

[36] ——, *Cvx: Matlab software for disciplined convex programming, version 2.1*, http://cvxr.com/cvx, Mar. 2014.