PERCEPTUAL DECORRELATOR BASED ON RESONATORS

Jon Fagerström[1], *Nils Meyer-Kahlen*[1], *Sebastian J. Schlecht*[2] *and Vesa Välimäki*[1]

[1]Acoustics Lab, Department of Information and Communications Engineering, Aalto University, Espoo, Finland
[2]Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Germany
fagerstrom.jon@gmail.com, nils.meyer-kahlen@aalto.fi

ABSTRACT

Decorrelation filters transform mono audio into multiple decorrelated copies. This paper introduces a novel decorrelation filter design based on a resonator bank, which produces a sum of over a thousand exponentially decaying sinusoids. A headphone listening test was used to identify the minimum inter-channel time delays that perceptually match ERB-filtered coherent noise to corresponding incoherent noise. The decay rate of each resonator is set based on a group delay profile determined by the listening test results at its corresponding frequency. Furthermore, the delays from the test are used to refine frequency-dependent windowing in coherence estimation, which we argue represents the perceptually most accurate way of assessing interaural coherence. This coherence measure then guides an optimization process that adjusts the initial phases of the sinusoids to minimize the coherence between two instances of the resonator-based decorrelator. The delay results establish the necessary group delay per ERB for effective decorrelation, revealing higher-than-expected values, particularly at higher frequencies. For comparison, the optimization is also performed using two previously proposed group-delay profiles: one based on the period of the ERB band center frequency and another based on the maximum group-delay limit before introducing smearing. The results indicate that the perceptually informed profile achieves equal decorrelation to the latter profile while smearing less at high frequencies. Overall, optimizing the phase response of the proposed decorrelator yields significantly lower coherence compared to using a random phase.

1. INTRODUCTION

Decorrelation refers to decreasing the autocorrelation of a signal or decreasing the cross-correlation of two or more related signals [1, 2]. In the realm of audio, the latter meaning is more common. Therein, decorrelation usually aims to increase the perceived width or spaciousness of an auditory event without changing the timbre of the audio signal, including its perceived spectral and temporal properties. This study investigates the perception of decorrelation in stereo listening over headphones.

In the natural world, decorrelation between the signals arriving at the listener's ears, i.e., a low *interaural cross-correlation* (IACC), can have different causes. One reason for decorrelation is the occurrence of several uncorrelated sound sources in the environment around a listener, like leaves rattling in the wind or waves arriving at the seashore. Another way in which decorrelated sound is created is through reverberation. Moreover, low IACC is found when listening to spatially extended sound sources; if an object with spatial extent vibrates, different points radiate sound with different phase relations. If a listener is close to the object, these paths are summed differently at the two ears.

However, not all natural mechanisms can be virtualized directly when the goal is to create an artificial general-purpose decorrelator using digital signal processing. Reverberation, for example, heavily modifies the timbre of sounds, causing both coloration and temporal smearing. Early decorrelator designs aimed at tackling the mono-to-stereo-upmixing challenge, i.e., to decorrelate a single audio signal into two decorrelated copies via various electrical and electromechanical means, including delay lines, and reverberation chambers [3, 4, 5]. Previous digital methods for virtually creating decorrelation include IIR (infinite impulse response) allpass filters [1, 6, 7, 2], sparse FIR (finite impulse response) filters based on velvet noise [8, 9], short-time Fourier transform (STFT) [10], filter bank-based approaches utilizing frequency-dependent delays or filtering [11, 12, 13], and neural networks [14].

All of these decorrelation methods directly or indirectly depend on introducing a frequency-dependent delay to the signal being decorrelated. The delay could vary throughout the signal, but since this always has the potential for additional, unwanted degradation, we focus on time-invariant decorrelators here. Nevertheless, since introducing frequency-dependent delay naturally leads to temporal smearing, decorrelator design is always a trade-off between the two. Here, we perform a listening experiment to determine upper limits to the required delays to better inform this choice.

This paper proposes a decorrelator based on a bank of resonators, similar to how they have previously been used in the context of modal synthesis of late reverberation [15, 16]. In this work, each resonator produces a group-delay peak, which is inversely proportional to the decay rate of that resonator. Optimization is performed over the initial phases of the resonators to minimize interaural coherence between two decorrelator instances. Additionally, we introduce a perceptually grounded method for estimating coherence using frequency-dependent windowing. The window sizes are derived from a perceptual study that identified the minimal interaural delay per equivalent rectangular bandwidth (ERB) [17] band required to match the perceived width of fully incoherent noise. This approach also establishes a practical upper bound for the group delay of binaural decorrelators, beyond which only perceptual smearing is introduced. The proposed decorrelator is evaluated using three different group delay profiles, demonstrating that the perceptually derived profiles outperform or match prior approaches in terms of effective decorrelation. The system can be implemented using many parallel biquad filters or as an FIR filter, which samples the impulse response (IR) of the decorrelator.

The remainder of this paper is organized as follows. Sec. 2 introduces necessary background on audio decorrelation. Sec. 3

describes the perceptual experiment used to determine appropriate group-delay profiles for perceptually optimized decorrelation as well as frequency-dependent windows for coherence estimation. Sec. 4 proposes the novel resonator-based decorrelator and describes the optimization procedure that minimizes the coherence between two decorrelator instances. Sec. 5 presents objective and qualitative evaluation of the proposed method. Sec. 6 concludes the work.

2. DECORRELATION OF AUDIO SIGNALS

The goal of constructing a decorrelator is to achieve strong decorrelation while still maintaining the timbral properties of the signals to which it is applied. Decorrelation is typically characterized either by the cross-correlation between signals, which is

$$r_{\mathrm{x_1x_2}}(l) = \sum_{n=0}^{L-1} x_1(n)x_2(n+l), \tag{1}$$

or, in a frequency dependent manner, often by the magnitude-squared coherence

$$\Phi_{\mathrm{x_1x_2}}(\omega) = \frac{|S_{\mathrm{x_1x_2}}(\omega)|^2}{S_{\mathrm{x_1}}(\omega)S_{\mathrm{x_2}}(\omega)}, \tag{2}$$

where $S_{\mathrm{x_1x_2}}$ is the cross-spectral density (CPSD) between signals x_1 and x_2, which is related to the cross-correlation by DFT:

$$S_{\mathrm{x_1x_2}}(\omega) = \sum_{l=-\infty}^{\infty} r_{\mathrm{x_1x_2}}(l)e^{-\imath\omega l}. \tag{3}$$

Low coherence is achieved essentially by creating frequency-dependent delays between the signals. Accordingly, given two decorrelators $h_1(t)$ and $h_2(t)$, the corresponding inter-channel time difference (ICTD) is defined as [18]

$$\mathrm{ICTD}(\omega) = \tau_2(\omega) - \tau_1(\omega), \tag{4}$$

where τ_1 and τ_2 are the group delays corresponding to the signals

$$\tau_i(\omega) = -\frac{d\angle H_i(\omega)}{d\omega}. \tag{5}$$

Therein, $H_i(\omega)$ denotes the complex frequency response of $h_i(t)$

$$H_i(\omega) = |H_i(\omega)|e^{j\angle H_i(\omega)}, \tag{6}$$

where ω is the frequency in radians, and $|H_i(\omega)|$ and $\angle H_i(\omega)$ are the magnitude and phase response of h_i, respectively.

At the same time as group-delay differences cause decorrelation, the introduced group delay itself causes timbral degradations. Even small group-delay deviations of less than 1 ms can be audible, shown for a single group-delay peak on transient stimuli [19]. For decorrelator design, a much more generous frequency-dependent limit for how much group delay is permissible was proposed by Canfield-Dafilou and Abel [2].

With the goal of a large ICTD on the one hand but low group delay on the other, designing decorrelators becomes a multi-objective optimization problem, where the ideal tradeoff is hard to find. Therefore, it is important to know how much ICTD is required for maximally wide sounds. If the ICTD exceeds this limit, a further increase will not contribute to perceived decorrelation anymore, while still requiring large group delays that cause timbral distortions. To help optimize our decorrelator, we conducted the listening experiment described in the next section.

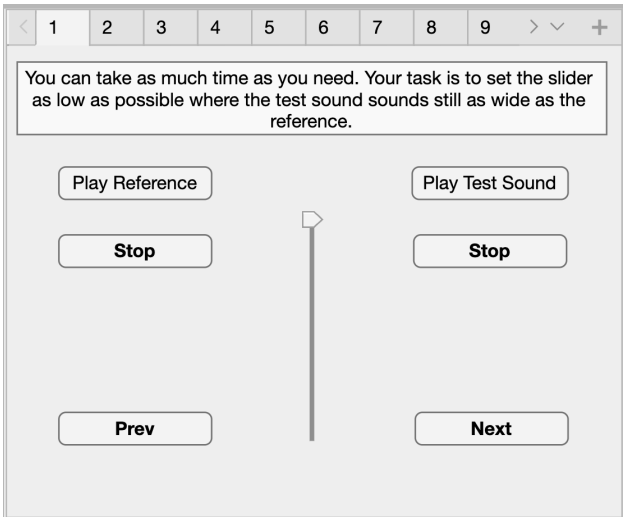


Figure 1: *Graphical user interface of the listening test. Participants were asked to move the slider to the lowest position where the test noise stimulus was perceived as wide as the reference noise.*

3. EFFECT OF TIME DELAY ON BINAURAL COHERENCE

The maximum time delay between the ears that occurs in a natural environment is directly related to the distance between the ears and the speed of sound. An audio event presented in stereo over headphones could be completely lateralized with an ICTD of 630 μs [20]. Depending on the size of the head of a listener, the maximal ICTD can reach 700 μs , but it is clearly below 1 ms. Thus, one might thus assume that ICTDs larger than 1 ms plays no role in spatial perception and is not perceived. This is reflected by the practice of measuring IACC by analyzing the position of its peak with a maximal lag of 1 ms, as it done in room acoustics, see ISO 3382.

In contrast, decorrelation filters are often up to 30 ms long [8], indicating that delays much longer than the maximum interaural time difference might have a role in inducing the sense of width. Bouéri and Kyriakakis [11] proposed to use ICTDs corresponding to the inverse of the frequency, for example. Informal listening confirms that only when increasing the ICTD of noise between the ears well over the maximum interaural time difference range, a similarly wide perception as obtained through fully incoherent binaural noise is perceived.

A headphone listening experiment was devised to find the smallest delays per ERB at which the perceived width of the delayed noise matches the width of incoherent noise. Figure 1 shows the user interface of the listening test. The reference sound was a gammatone-filtered completely incoherent binaural noise, i.e., each stereo channel had a different random noise instance. The test sound was fully coherent noise, and the slider controlled the ICTD between the two headphone channels from 0 ms at the bottom to 60 ms at the top of the scale. During the test, the participants were asked to set the slider to the lowest point at which the test noise sounded as wide as the reference noise. The test consisted of 16 trials, each of which had a different ERB-spaced gammatone filter center frequency between 100 Hz–10 kHz. Fur-

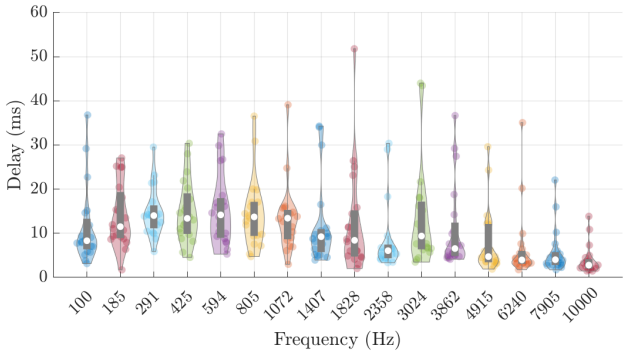


Figure 2: *Listening test results with ERB-filtered noises plotted using the* `violinplot` *function [21]. The boxplot (grey, box) is included in the center of the violin. The white dot indicates the median, and the bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The violin outline shows the kernel density estimation.*

Table 1: *ERB center frequencies for the gammatone filters used and the median and 75th percentile ICTD results of the listening test.*

Center Frequency	Median	75th Percentile
100 Hz	8.3 ms	13.0 ms
185 Hz	11.5 ms	19.3 ms
291 Hz	14.0 ms	16.3 ms
425 Hz	13.4 ms	19.0 ms
594 Hz	14.2 ms	17.9 ms
805 Hz	13.7 ms	17.0 ms
1072 Hz	13.4 ms	15.2 ms
1407 Hz	9.2 ms	10.9 ms
1828 Hz	8.4 ms	15.1 ms
2358 Hz	6.1 ms	7.3 ms
3024 Hz	9.4 ms	17.1 ms
3862 Hz	6.5 ms	12.3 ms
4915 Hz	4.7 ms	12.0 ms
6240 Hz	4.0 ms	5.9 ms
7905 Hz	4.0 ms	5.7 ms
10000 Hz	2.8 ms	4.1 ms

thermore, each of the ten participants completed the test twice to give a total of 320 data points. The order of the center frequencies and the headphone channel to which the delay was added was randomized for each participant and both repetitions.

The results of the listening test are shown in the violin plots of Fig. 2. The median (white dot) and the 75th percentile of each tested ERB are further summarized in Table 1. The results show the general trend that lower frequency bands need more delay than higher bands to induce the perception of width. However, the required amount of delay does not follow the previously proposed $1/f$ curve and instead shows a much higher relative amount of delay beyond the very low frequencies.

These results help the design of decorrelators; increasing the ICTDs far beyond these limits will only contribute to smearing, but not to perceived width. Incorporating the obtained curve into the decorrelator design is achieved by modifying how the coherence is estimated. Practical estimation of the coherence between two signals (2) requires windowing and is obtained as the time average over the windows:

$$\hat{\Phi}_{\text{uni}}(\omega) = \frac{\sum_{t=0}^{T-1} |H_{\text{L}}(t,\omega) H_{\text{R}}^*(t,\omega)|^2}{\sum_{t=0}^{T-1} |H_{\text{L}}(t,\omega)|^2 \sum_{t=0}^{T-1} |H_{\text{R}}(t,\omega)|^2}, \tag{7}$$

where t is the time-frame index, T is the number of STFT frames, $H_{\text{L}}(t,\omega)$ and $H_{\text{R}}(t,\omega)$ are the STFT coefficients of the left and the right channel signals, respectively, and $.^*$ denotes complex conjugation. In this work, we used a Hann window with 50% overlap. Typically, a uniform frequency-independent window size is used for the coherence estimation. We now incorporate a frequency-dependent window by first filtering the signals with ERB-spaced gamma tone filters and computing the STFT of each band-passed signal with different window sizes corresponding to the 75th percentile delays found in our listening test. This perceptual frequency-dependent coherence is formulated as

$$\hat{\Phi}_{\text{per}}(\omega) = \sum_{b=0}^{B-1} \frac{\sum_{t_b=0}^{T_b-1} |H_{b_{\text{L}}}(t_b,\omega) H_{b_{\text{R}}}^*(t_b,\omega)|^2}{\sum_{t_b=0}^{T_b-1} |H_{b_{\text{L}}}(t_b,\omega)|^2 \sum_{t_b=0}^{T_b-1} |H_{b_{\text{R}}}(t_b,\omega)|^2}, \tag{8}$$

where $H_{b_{\text{L}}}$ and $H_{b_{\text{R}}}$ are the STFT coefficients of the bth ERB filtered signals of the left and right channels, respectively, t_b is the time-frame index of the bth ERB, and T_b is the number of time-frames for the STFT of the bth ERB signals. In this work, we used $B = 16$ ERB bands for the estimation.

Figure 3 shows the effect of incorporating such windowing on an example. Fig. 3 (a) shows the coherence of noise that has an ICTD corresponding to the 75th percentile of the listening test results (see Table 1). The same is used for the windowing, so that using the perceptual window, a coherence near 0 is found by definition. Here, the same would be obtained for a short window of 5 ms. A large window, however, would overestimate the perceived coherence. Fig. 3 (b), less ICTD is used (the median from Table 1). While the small window still yields rather low values, perceptual windowing suggests that the amount may not be sufficient at low frequencies, whereas the large window provides larger values again. We will incorporate the perceptual windowing in the optimization of the proposed resonator-based decorrelator described in the following section. The signals of Fig. 3, as well as a fully incoherent reference, are available online. [1]

4. PROPOSED DECORRELATOR DESIGN

This section presents the proposed decorrelator design based on a bank of resonators. An optimization scheme to minimize the perceptual coherence derived above between two channel decorrelators is also discussed.

4.1. Resonator-Based Decorrelator

The proposed design is based on exponentially decaying sinusoids, which are linearly distributed on the ERB scale across frequencies throughout the audio range. A sum of K decaying sinusoids is

[1] `http://research.spa.aalto.fi/publications/papers/dafx25-reso-deco/`

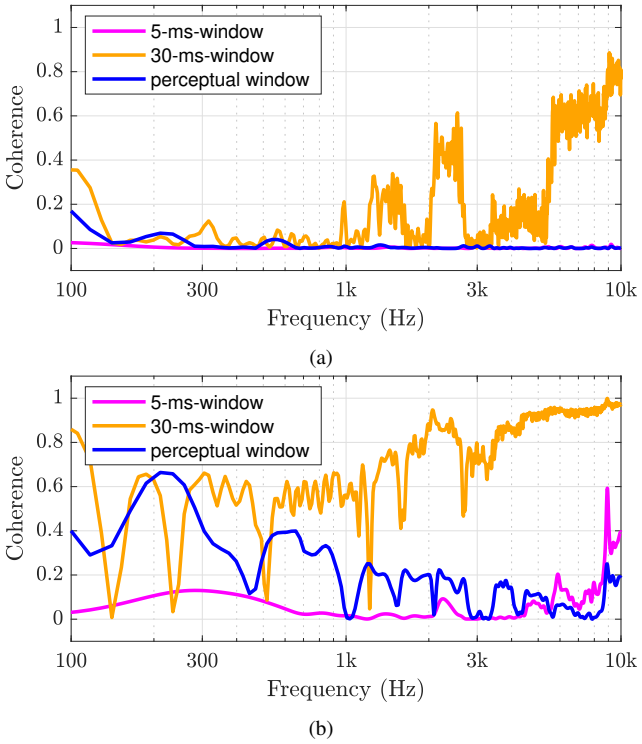


Figure 3: *Coherence estimates of delayed ERB noise configured with the (a) 75th percentile and (b) median ICTDs obtained from the listening test data of Fig. 2. The coherence estimates use one of the two constant window sizes (5 ms and 30 ms) or the proposed perceptual frequency-dependent windowing based on the 75th percentile delays.*

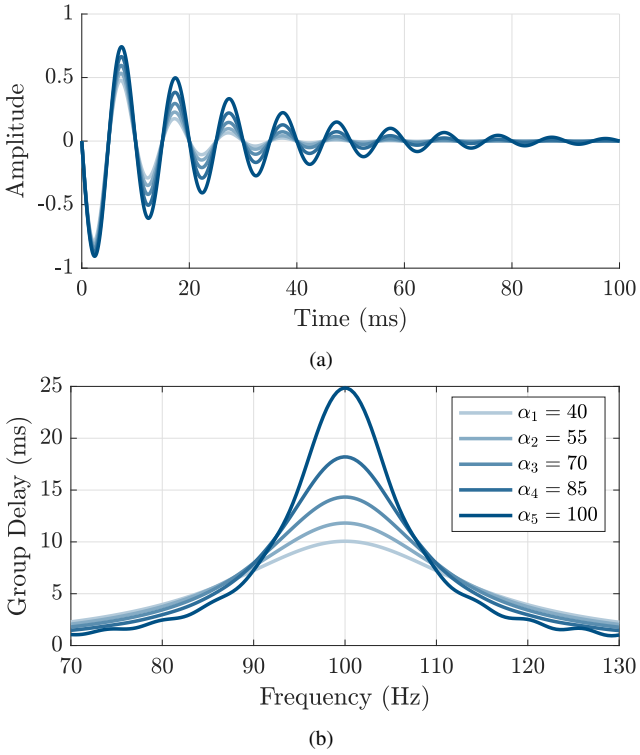


Figure 4: *(a) Impulse responses and (b) group delays of single 100-Hz resonators with different decay constants α_k.*

obtained as the IR of a bank of second-order resonators [15]:

$$h(n) = \sum_{k=0}^{K-1} A_k R_k^n \cos(\omega_k n / f_s + \phi_k), \tag{9}$$

where A_k is the initial amplitude, $R_k = e^{-\alpha_k / f_s}$ is the pole radius defining the exponential decay term, and ω_k and ϕ_k are the frequency and phase offset in radians of the kth partial, respectively. The decay rate of each resonator controls the group-delay peak centered at the resonator frequency with

$$\tau(\omega_k) = \frac{1}{\alpha_k}. \tag{10}$$

Figures 4a and 4b show the impulse responses and group delays, respectively, of a 100-Hz resonator with different values of the decay constant α_k. The resonators with slower decays can be seen as having a higher group-delay peak, which is in line with Eq. (10).

4.2. Equalization

The final pre-processing step in our design is to correct the magnitude response. Since the decorrelator IR is obtained as the sum of closely and non-uniformly spaced (linearly spaced on the ERB scale) sinusoids with variable initial phases, the magnitude response tends towards lowpass behavior with notches and peaks resulting from constructive and destructive interferences between neighboring resonators. To compensate for this, we apply linear prediction (LP) to design an FIR equalizing filter to flatten the magnitude response [22, 23].

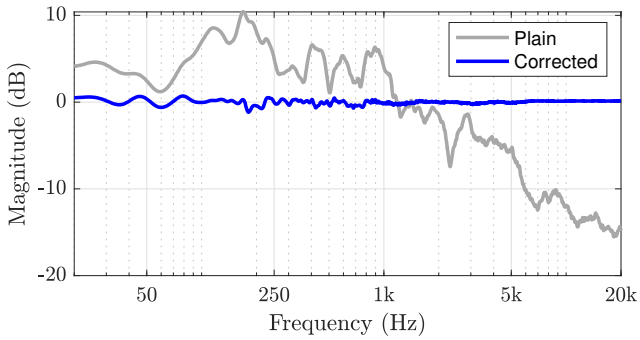


Figure 5: *An example third-octave smoothed magnitude response of the resonator-based FIR decorrelator before (gray) and after applying the LP-based equalizing filter of order N=960 (blue).*

Figure 5 shows the third-octave smoothed magnitude spectrum of a single proposed decorrelator instance, with $K = 1600$ resonators, before (gray) and after applying the LP filter (blue). The order of the LP filter is $N = 960$, which was chosen based on informal listening. The corrected magnitude response is flat within ± 1.2 dB. A higher-orders LP filter could completely flatten the response at even lower frequencies. However, there is a trade-off in that using a higher LP order renders the equalizing filter response audible as metallic ringing.

4.3. Practical Implementation

The proposed resonator-based decorrelator can be implemented either in IIR or FIR form. The IIR version can be implemented in biquad form using

$$y(n) = \sum_{k=0}^{K-1} \Big[b_{0_k} x(n) + b_{1_k} x(n-1) + b_{2_k} x(n-2) + a_{1_k} y(n-1) + a_{2_k} y(n-2) \Big], \tag{11}$$

where

$$\begin{aligned} b_{0_k} &= A_k, \\ b_{1_k} &= -2A_k \cos(\phi_k), \\ b_{2_k} &= A_k, \\ a_{1_k} &= -2R_k \cos\left(\frac{\omega_k}{f_s}\right), \\ a_{2_k} &= R_k^2. \end{aligned} \tag{12}$$

Alternatively, the IR of the decorrelator given by Eq. (9) can be truncated and sampled as coefficients of an FIR filter. We set the FIR filter order N relative to the slowest decaying partial in the resonator bank with

$$N = \max(T_{60}(\omega_k)), \tag{13}$$

where $T_{60}(\omega_k)$ denotes the time required for the amplitude of the kth partial to decay by 60 dB. This time is related to the group delay peak of the resonator via

$$T_{60}(\omega_k) = 3\ln(10)\tau(\omega_k) \approx 6.91\tau(\omega_k). \tag{14}$$

Note that the decay time T_{60} is considerably larger than the peak group-delay value.

4.4. Optimization

To obtain an effective decorrelator using the proposed resonator-based technique, we minimize the perceptual coherence, Eq. (8), between two instances. The optimization was implemented using the genetic algorithm (GA) in the MATLAB optimization toolbox. GA is an optimization method inspired by natural selection to find optimal or near-optimal solutions in complex search spaces [24]. The optimizable variable is the initial phase ϕ_k of each resonator and we set the upper and lower bound for the optimization to be π and $-\pi$, respectively. Before optimization, the phase of each resonator was initialized to uniform random values between the bounds and the initial amplitudes A_k between -1 and 1.

As the loss function, we used the mean squared error of the proposed perceptual coherence function, Eq. (8), in dB on the equalized decorrelator IR:

$$J = 10\log_{10}\left(\sqrt{\frac{1}{M}\sum_{m=0}^{M-1}\hat{\Phi}_{per}^2(\omega_m)}\right). \tag{15}$$

To speed up the IR recomputation for the optimization, we implemented each resonator in the all-pole form derived from the full biquad form of Eq.(11) with the second-order difference equation

$$h(n) = \sum_{k=0}^{K-1} a_{1_k} h_k(n-1) - a_{2_k} h_k(n-2), \tag{16}$$

To set the correct initial phase and amplitude, each resonator's difference equation is initialized with

$$h(0) = \sum_{k=0}^{K-1} A_k \cos(\phi_k), \tag{17}$$

$$h(1) = \sum_{k=0}^{K-1} A_k R_k \cos(\omega_k / fs + \phi_k). \tag{18}$$

Note that this all-pole form is only usable in computing the IR of the system and not in implementing the filtering of arbitrary signals.

5. EVALUATION

In this section, we discuss the evaluation of our proposed resonator-based decorrelator. In addition to the objective results presented here, sound examples are provided online in Footnote 1 for the reader to assess the perceptual qualities of the method.

5.1. Group-Delay Profiles

The main design parameter of the proposed decorrelator is the group-delay profile, which is controlled by the decay constants of each resonator, see Eq. (10). In this work we compare three different choices for the group-delay profiles: one based on the 75th percentile listening test results of Sec. 3, a second based on the smearing limits defined by Canfield-Dafilou and Abel [2] and a third relative to the longest wavelength within each ERB band, i.e. $1/f$ profile [11]. In the following, we refer to them as the 'proposed', 'smearing limit', and '$1/f$' group-delay profiles, respectively.

Figure 6 shows the three different group-delay profiles tested. For the synthesis, we needed to interpolate all of the curves as well as extrapolate our proposed curve to span the whole audio range 20 Hz–20 kHz. and to match the number of resonators. We chose to use $K = 1600$ resonators based on preliminary testing for each configuration to produce a noise-like IR for the decorrelator. For the interpolation, we used the spline method, and for the extrapolation, we simply used the nearest neighbor method to extend the low and high bands. The resulting group-delay profiles used for synthesis are shown Fig. 6.

5.2. Optimized Coherence

Figure 7 shows the loss of the GA optimization for each of the three group-delay profiles. The population size used was 20, and the optimization ran until reaching the function tolerance, which was set to 1×10^{-6}. The $1/f$ did not optimize almost at all with running only for 272 generation and with a final loss $J = -0.2$ dB. The proposed and smearing limit group-delay profiles performed better while running for 20,728 and 16,218 generations with a final loss $J = -23.4$ dB and $J = -24.1$ dB, respectively.

The coherence estimates with initial random parameters (blue, dashed) and optimized parameters (red, solid) are shown in Fig. 8 and confirm the informtion obtained from the loss values; the proposed and smearing limit group-delay profiles are shown to optimize to close to zero coherence across frequencies, whereas the $1/f$ profile fails to provide low coherence with the exception of single sharp notches and a low-frequency roll-off.

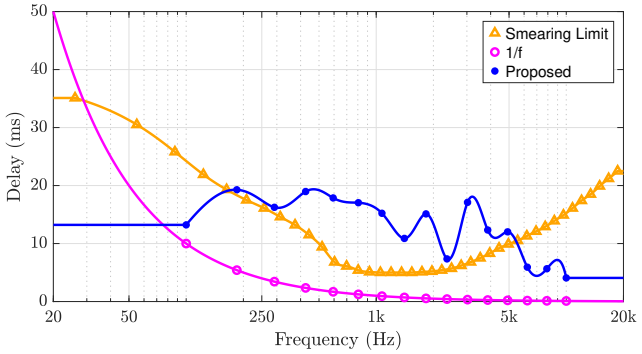


Figure 6: *Maximum group-delay limits based on the smearing limit [2] (orange, triangle), 75th percentile delays from Fig. 2 (blue, dot), and the delay limits proportional to the ERB (pink, circle) as proposed in [11]. The markers show the underlying data points and the curves are the profiles after interpolation/extrapolation to fit the number of resonators used in the synthesis,*

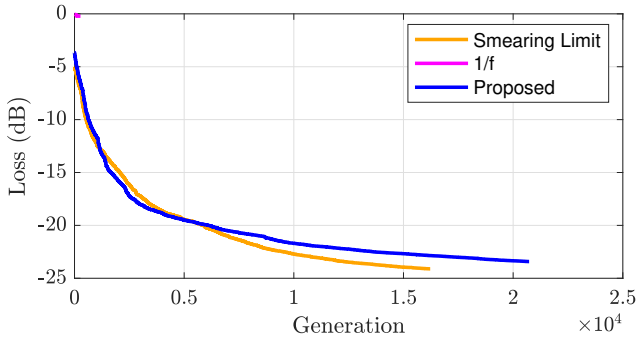


Figure 7: *Coherence loss with the tested group-delay profiles after running the GA with population size* 20*, converged after reaching function tolerance of* 1×10^{-6}.

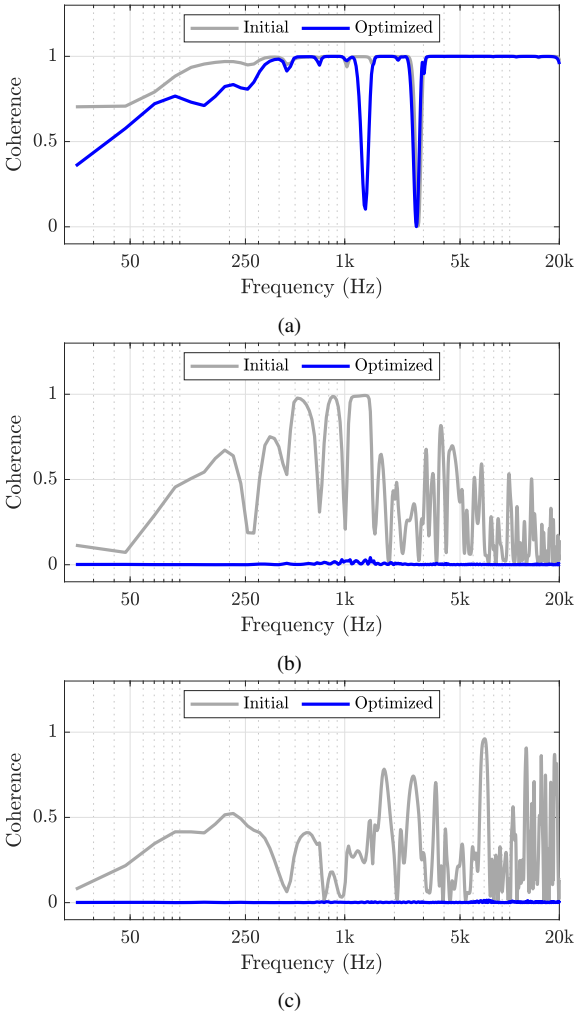


Figure 8: *Initial (gray), and optimized (blue) coherence of a resonator-based decorrelator using (a) the* $1/f$ *[11], (b) the smearing limit [2], and (c) the proposed perceptual group-delay profiles.*

5.3. Spectrogram Analysis

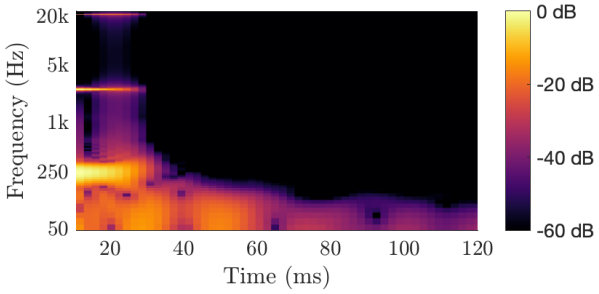
In this section, the spectrogram analysis of the three tested decorrelator configurations is discussed. The spectrograms of the optimized decorrelators for the $1/f$, smearing limit, and proposed group delay profile configurations are shown in Figs. 9a, 9b, and 9c, respectively. A Hann window of $L = 1024$ samples long with 50% overlap was used to compute the spectrograms. The FFT size was 4096 samples, with a sample rate $f_s = 48$ kHz.

Overall, the spectrograms in Fig. 9 reveal the expected shapes when compared with the used group-delay profiles of Fig. 6. The smearing limit profile results in the decorrelator IR (cf. Fig. 9b) that is shorter at the mid frequencies and extends longer at low and high frequencies. Note that the IR length is again much larger than the group delays, which is in line with Eq. (14). The IR related to $1/f$ in Fig. 9a is so short at the higher bands that the STFT fails to capture it accurately. Finally, the proposed group-delay profile in Fig. 9c presents a pattern that is essentially the inverse of the smearing limit plot shown in Fig. 6: the IR is short at low and high frequencies, while the mid frequencies exhibit a longer decay.

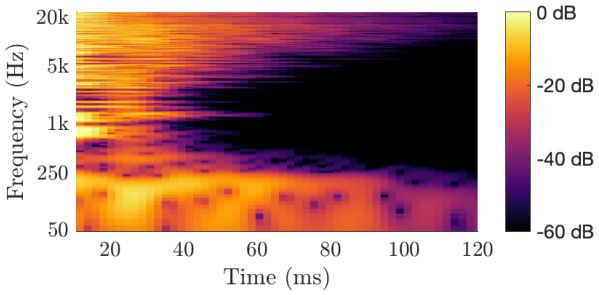
5.4. Computational Cost

Computational costs of the two alternative implementation strategies are compared in this section. The number of multiplications for the IIR and FIR implementations of the proposed method are presented in Table 2. The numbers are computed for the configuration using the proposed group-delay profile, with $K = 1600$ resonators and with the FIR equalizing filter of order $N = 960$.

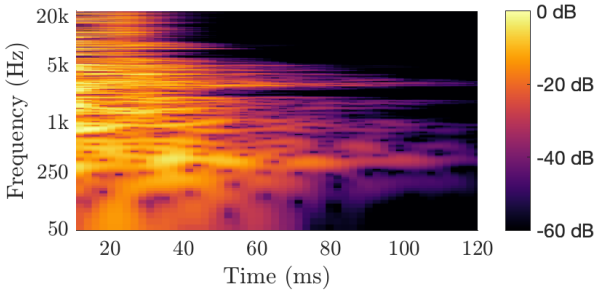
For the IIR filter implementation, each biquad requires five multiplications per sample and the equalizer consumes $N = 960$ multiplications per sample, resulting in a total of $5 \times 1600 + 960 = 8960$ multiplications. In comparison, the FIR configuration is slightly more efficient in this case, requiring $0.134 \times 48{,}000 = 6432$ multiplications per sample. Note that with the FIR implementation, the equalizer incurs no additional cost, as it can be applied offline to the IR. However, increasing the maximum group delay would require a longer IR for the FIR implementation, cf. Eq. (14), which would increase its computational cost. In



(a)



(b)



(c)

Figure 9: *Spectrograms of the left-channel optimized resonator-based decorrelators using (a) the* $1/f$ *[11], (b) the smearing limit [2], and (c) the proposed perceptual group-delay profiles, cf. Fig. 6.*

summary, the computational cost of the IIR implementation scales with the number of resonators and is independent of the maximum group delay, whereas the cost of the FIR implementation scales with the maximum group delay and is independent of the number of resonators.

6. CONCLUSIONS

A resonator-based teechnique is proposed in this paper as an optimizable framework for designing decorrelation filters for stereo audio. The proposed decorrelator's IR is synthesized using a bank of resonators, where the decay rate of each resonator controls the peak group delay introduced at its corresponding frequency. Optimization is applied to the initial phase of each resonator in the bank, while minimizing the coherence between two decorrelator instances. The initial phases of each resonator control the constructive and destructive interference between the adjacent resonators, resulting in reduced coherence. The method also requires an equalizer to whiten the magnitude response of the resonator bank. The new decorrelation filter can be implemented as an IIR filter consisting of many parallel second-order IIR resonators together with an equalizing filter, or as a single high-order FIR filter, which samples the truncated impulse response of the decorrelator.

Table 2: *Computational costs of the IIR resonator and FIR filter implementations of the proposed method. The resonator method contains* $K = 1600$ *resonators and an LP-based equalizer of order* $N = 960$*. The FIR filter is truncated to the length of* $134\,ms$ *at the sample rate* $f_s = 48\,kHz$*.*

Implementation	Multiplications
Resonators and equalizer	8960
FIR filter	6432

In addition to the novel decorrelator design, we provide a more general framework for estimating the coherence based on frequency-dependent windowing, where the window sizes are derived from a perceptual study that aimed to find the minimal ICTD of gamma-tone-filtered noise, per ERB, that induce the same perceived width as the corresponding fully incoherent filtered noise. These minimal delays also reveal the maximum group delay that a binaural decorrelator should produce, as any additional delay would only cause unwanted temporal smearing in the output signal.

The proposed decorrelator was configured with three different group-delay profiles to compare the proposed profile to the previous profile based on the relation $1/f$ and to another previous perceptual profile defined as the informally tested maximum delays found not to cause temporal smearing. The results indicate that the $1/f$ has insufficiently small delays everywhere except at very low frequencies to induce any decorrelation. On the other hand, the proposed perceptual profile and the previous smearing limit-based profile perform objectively equally well.

However, the extended low and high-frequency group-delay range of the smearing limit profile does not seem to provide any benefit over the proposed lower group delays in those frequency regions. Conversely, it could be said that the larger mid-frequency group-delay range we found based on the listening test might not bring considerable improvement compared to the smaller smearing limits.

The findings of this paper are useful in the design of optimized decorrelation filters, applicable in the widening of mono sounds. Future research could consider the joint optimization of multiple decorrelators for multi-channel upmixing scenarios and include a formal listening test to assess the perceptual quality of the proposed method against state-of-the-art alternatives.

7. REFERENCES

[1] 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.

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

[3] 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.

[4] B. B. Bauer, "Some techniques toward better stereophonic perspective," *J. Audio Eng. Soc.*, vol. 17, no. 4, pp. 410–415, Aug. 1969.

[5] 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.

[6] M. A. Gerzon, "Signal processing for simulating realistic stereo images," in *Proc. Audio Eng. Soc. Conv. 93*, San Francisco, CA, USA, Oct. 1992, paper 3423.

[7] C. Uhle and P. Gampp, "Mono-to-stereo upmixing," in *Proc. Audio Eng. Soc. Conv. 140*, Paris, France, May 2016.

[8] B. Alary, A. Politis, and V. Välimäki, "Velvet-noise decorrelator," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Edinburgh, UK, Sep. 2017, pp. 405–411.

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

[10] S. Disch, "Decorrelation for immersive audio applications and sound effects," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Copenhagen, Denmark, Sep. 2023.

[11] M. Bouéri and C. Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *Proc. 117th Audio Eng. Soc. Conv.*, San Francisco, CA, Oct. 2004.

[12] G. Potard and I. S. Burnett, "Control and measurement of apparent sound source width and its applications to sonification and virtual auditory displays," in *Proc. Int. Conf. Auditory Display (ICAD)*, Sydney, Australia, Jul. 2004.

[13] R. Penniman, "A general-purpose decorrelator algorithm with transient fidelity," in *Proc. AES Conv. 137*, Los Angeles, CA, USA, Oct. 2014, paper 9170.

[14] C Anemüller, O. Thiergart, and E. A. P. Habets, "Neural audio decorrelation using generative adversarial networks," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2023, pp. 1–5.

[15] M. Karjalainen and H. Järveläinen, "More about this reverberation science: Perceptually good late reverberation," in *Proc. 111th AES Conv.*, New York, USA, Sep. 2001.

[16] J. S. Abel, S. Coffin, and K. Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *Proc. 137th AES Conv.*, Los Angeles, USA, Oct. 2014, paper 9208.

[17] B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.*, vol. 152, pp. 942–953, Sep. 2022.

[18] F. Zotter, M. Frank, G. Marentakis, and A. Sontacchi, "Phantom source widening with deterministic frequency dependent time delays," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Paris, France, Sep. 2011, pp. 307–312.

[19] J. Liski, A. Mäkivirta, and V. Välimäki, "Audibility of group-delay equalization," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 29, pp. 2189–2201, Jun. 2021.

[20] J. Blauert, *Spatial Hearing - Revised Edition: The Psychoacoustics of Human Sound Source Localisation*, MIT Press, Cambridge, MA, 1997.

[21] B. Bechtold, "Violin Plots for Matlab, Github Project," Available at https://github.com/bastibe/Violinplot-Matlab, accessed Mar 27, 2024.

[22] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, vol. 12, Springer-Verlag, 1976.

[23] J. O. Smith, *Spectral Audio Signal Processing*, accessed July 2, 2025, Available at https://ccrma.stanford.edu/~jos/sasp/. Online book, 2011 edition. See Section "Linear Prediction Spectral Envelope".

[24] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Boston, MA, USA, 1989.