AN OPEN-SOURCE STEREO WIDENING PLUGIN

Orchisama Das

Sonos, Inc
London, United Kingdom
orchisama.das@sonos.com

ABSTRACT

Stereo widening algorithms aim to extend the stereo image width and thereby, increase the perceived spaciousness of a mix. Here, we present the design and implementation of a stereo widening plugin that is computationally efficient. First, a stereo signal is decorrelated by convolving with a velvet noise sequence, or alternately, by passing through a cascade of allpass filters with randomised phase. Both the original and decorrelated signals are passed through perfect reconstruction filterbanks to get a set of lowpassed and highpassed signals. Then, the original and decorrelated filtered signals are combined through a mixer and summed to produce the final stereo output. Two separate parameters control the perceived width of the lower frequencies and higher frequencies respectively. A transient detection block prevents the smearing of percussive signals caused by the decorrelation filters. The stereo widener has been released as an open-source plugin.

1. INTRODUCTION

Although multichannel and binaural audio is becoming more widely adopted in consumer electronics, stereo remains the predominant mode of audio playback. The standard stereo setup consists of loudspeakers at $\pm 30^\circ$ from the listening sweet spot, creating a perceived width of 60°. Stereo width widening aims to increase this perceived image width and makes the content sound more spacious. This is done by decorrelating the two channels, thereby decreasing the interaural cross-correlation (ICC). The simplest stereo widening technique is to add a small delay between the two channels and phase invert one of them[1].

Stereo widening algorithms in the literature propose adding FIR filters to model the direct path (from each speaker to the closest ear), and the cross-talk path, (from each speaker to the further ear), at a new phantom source location [1]. For headphone reproduction [2, 3],the signal is first weakly decorrelated, then the direct and cross-talk path are modeled with head-related transfer functions (HRTFs). The value of the delay between the direct and cross-talk path (interaural time difference - ITD), controls the perceived width of the stereo source. [3] adds one early reflection to the widening circuit to enhance the sense of externalisation.

Commercially available stereo spreading plugins[2] allow an intuitive control over the perceived image width (collapsed to the centre, or very wide). The widening is by alternately distributing the middle frequencies in subbands to the left and right channels. As a rule of thumb, low frequencies are spread less than mid and high frequencies to maintain a strong phantom centre. This is different from our proposed method.

The goal of this paper is to explore the design of a simple stereo width widener that is real-time compatible. While other papers in the literature use interaural cues (ITD) to adjust the stereo width, we rely on decorrelation and mixing. The rationale behind this is that a sense of externalisation can be imparted by manipulating the interchannel coherence [4], which is done by the decorrelation filters. The incoming stereo signal is first decorrelated, then the original stereo signal is mixed with the decorrelated signal. A controllable mixing gain determines the balance of the original stereo to the decorrelated signal, and hence, the perceived stereo width.

The rest of this paper is organised as follows. Some background on the mid/side ratio and decorrelation is discussed in Sec. 2. First, a broadband stereo spreader architecture is explored in Sec. 3. The two stereo channels are decorrelated by convolving with a sparse noise sequence (known as velvet noise [5]), or by passing through a cascade of allpass filters. Then, the original stereo signal and the decorrelated signal are mixed with a parameter controlling their relative weights in the output. A transient handling block prevents transient smearing by passing the input buffer to the output buffer directly when an onset is detected. Next, in Sec. 4, a frequency-dependent stereo spreader architecture is proposed using low-order perfect reconstruction crossover filters to control the perceived width of the stereo image in the lower and higher frequency regions separately. The proposed algorithms are evaluated in Sec. 5 and implemented in the form of an open-source VST3/AU plugin. While the decorrelators used in this paper are borrowed from state-of-the-art methods, the overall architecture, including transient handling, and independent width control in different frequency bands are novel contributions of the paper.

2. BACKGROUND

2.1. Mid-side ratio

The goal of stereo widening is to increase the perceived spatial width of the stereo image. The mid channel (also known as the centre channel), x_M and the side channel (also known as the surround channel), x_S of a stereo signal $x = [x_L, x_R]^T$ are given by,

$$x_M = \frac{1}{\sqrt{2}}(x_L + x_R) \tag{1}$$

$$x_S = \frac{1}{\sqrt{2}}(x_L - x_R). \tag{2}$$

[1] https://www.reddit.com/r/audioengineering/comments/ba338a/heres_a_mixing_trick_stereo_widening_using_phase/

[2] https://support.apple.com/en-gb/guide/logicpro/lgcef240df26/mac

Note that all the signals are time-dependent but the time variable has been omitted in the rest of this paper for brevity. The mid-side ratio, γ, is defined as the ratio of the mid channel to the side channel, and gives the perceived width of the stereo image.

$$\gamma = \frac{x_M}{x_S} = \frac{x_L + x_R}{x_L - x_R}. \tag{3}$$

For signals that are strongly correlated, i.e., have a high interchannel correlation across the two stereo channels, the mid-side ratio is very high, giving a strong centre image and a small perceived width. Likewise, a low mid-side ratio gives a larger perceived width [3].

One way of increasing the perceived width is to lower the interchannel ratio by boosting the surround channel. This can be done by decorrelating the two stereo channels. The decorrelated signals, x'_L, x'_R have a new mid-side ratio,

$$\gamma' = \frac{x'_L + x'_R}{x'_L - x'_R}, \tag{4}$$

$$\gamma' < \gamma. \tag{5}$$

2.2. Decorrelation

There are many proposed techniques for decorrelation in the literature and studies on how it affects spatial imagery [6, 7]. Here, we explore two methods that are efficient and can be run in real-time.

The first is convolution with a sparse sequence of plus and minus ones known as "velvet noise". Convolution with a velvet noise (VN) sequence can be implemented very efficiently with a tapped delay line [8]. Alary et. al [9] suggest adding an exponential decay to the sequence to prevent transient smearing, and distributing the impulse locations logarithmically so that energy is concentrated at the start of the sequence.

The second option for the decorrelator is a cascade of biquad allpass filters with randomised phases as proposed in [10]. The pole radii and angle are chosen from a uniform distribution. The angles are further warped in the ERB scale to maintain a constant pole density across the critical bands of human hearing. The pole radius is bounded by the maximum group delay in each frequency band. As suggested in [10], a cascade of $200-250$ biquads need to be used, thereby making this algorithm more inefficient than velvet noise convolution. Velvet noise sequences used for decorrelation have lengths of $15 - 30$ ms with a pulse density of 50 pulses/ms [9, 11], which is equivalent to a delay line with $750 - 1500$ taps.

On the flip side, the allpass filters maintain the original signal's magnitude response. Although velvet noise has been shown to have a flat power spectrum [12], audible comb filtering is heard with this method for short sequences. An optimised velvet noise sequence for decorrelation was proposed by Schlecht et al. in [11]. The non-convex optimisation cost function minimises the deviation of the magnitude response VN sequence from a flat spectrum in one-third octave bands. The impulse locations and gains are optimised within constraints. The resulting sequence, the optimised velvet noise (OVN), has less colouration than a standard VN sequence. However, the cost function is not explicitly designed to reduce the correlation among pairs of velvet noise sequences; therefore a Monte Carlo search is carried out by optimising many velvet noise pairs initialised with different seeds and choosing one that minimises the weighted sum of the mean absolute interchannel coherence and spectral flatness.

Such an optimisation routine can be carried out on initialisation of the plugin, however it is unlikely to converge within a reasonable amount of time. A better option is to do the optimisation offline and load the saved filter coefficients on startup but this restricts the decorrelation filters to be static.

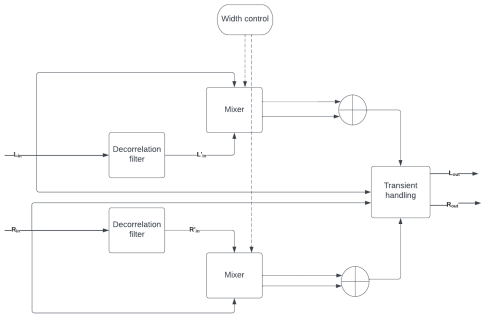
2.3. Mixer

A mixer distributes the incoming signal into the outgoing signals using some kind of mapping. Here, we use sine-law mixing to distribute the incoming signals [13]. For an incoming signal, $\boldsymbol{x} = [x_1, x_2]^T$, the outgoing signal $\boldsymbol{y} = [y_1, y_2]^T$ is given by,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \sin(\beta) & 0 \\ 0 & \cos(\beta) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{6}$$

where $\beta \in [0, \pi/2]$ is the mixing angle. The mixing angle controls the distribution of the two input signals, x_1, x_2 into the output signals.

3. BROADBAND STEREO WIDENER



(a) Broadband stereo widener block diagram

Figure 1: Block diagrams of the broadband stereo widener.

First, we will discuss a broadband stereo widener that applies the same width parameter to all frequencies. The block diagram for the frequency-independent stereo widener is shown in Fig. 1. The input stereo signal is passed through the decorrelation filters, and then the original stereo and decorrelated signal are cross-faded in the output. A width parameter, which controls the mixing angle β, is used to control the perceived stereo width by weighting the original stereo signal and the decorrelated stereo signal.

For input stereo signals, x_L, x_R and decorrelated signals x'_L, x'_R, the output of this circuit is simply a weighted sum of the original stereo and decorrelated signals:

$$\begin{aligned} x_{\{L_{\text{out}}, R_{\text{out}}\}} &= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \sin(\beta) & 0 \\ 0 & \cos(\beta) \end{bmatrix} \begin{bmatrix} x'_{\{L,R\}} \\ x_{\{L,R\}} \end{bmatrix} \\ x_{L_{\text{out}}} &= \sin(\beta) x'_L + \cos(\beta) x_L \\ x_{R_{\text{out}}} &= \sin(\beta) x'_R + \cos(\beta) x_R \end{aligned} \tag{7}$$

The mid-side ratio is given by,

$$\gamma_{\text{out}} = \frac{\sin(\beta)(x'_L + x'_R) + \cos(\beta)(x_L + x_R)}{\sin(\beta)(x'_L - x'_R) + \cos(\beta)(x_L - x_R)}. \tag{8}$$

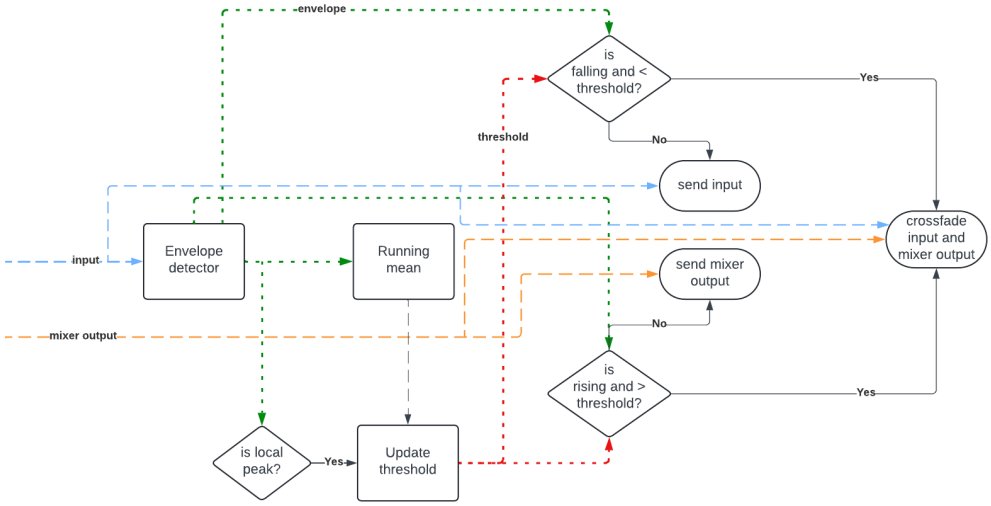


Figure 2: Transient handling logic. Coloured dashed lines represent different signals. Blue is the input signal, orange is the output signal from the widener, green is the input signal's envelope, and red is the adaptive threshold for onset/offset detection.

When $\beta = 0$, $\gamma_{\text{out}} = \gamma$ and the original stereo width is maintained. When $\beta = \pi/2$, $\gamma_{\text{out}} = \gamma'$, the maximum width is achieved. $\beta = \pi/4$ gives equal weighting to the original stereo and decorrelated stereo signals. Thus, $\beta \in [0, \pi/2]$ is a controllable parameter that gives an intuitive control over the perceived stereo width by balancing the weight of the original stereo input with the decorrelated signal.

3.1. Interchannel cross correlation

The interchannel cross correlation of the input stereo signal is given by,

$$\text{ICC}_{\text{in}} = \mathbb{E}(x_L x_R) \tag{9}$$

The interchannel cross correlation of the output signal is given by,

$$\begin{aligned}\text{ICC}_{\text{out}} &= \mathbb{E}(x_{L_{\text{out}}} x_{R_{\text{out}}}) \\ &= \mathbb{E}\left((\sin(\beta)x'_L + \cos(\beta)x_L)(\sin(\beta)x'_R + \cos(\beta)x_R)\right) \\ &= \sin^2(\beta)\mathbb{E}(x'_L x'_R) + \cos^2(\beta)\mathbb{E}(x_L x_R) \\ &+ \sin(\beta)\cos(\beta)\left(\mathbb{E}(x_L x'_R) + \mathbb{E}(x'_L x_R)\right) \\ &\approx \cos^2(\beta)\mathbb{E}(x_L x_R) \leq \text{ICC}_{\text{in}}\end{aligned} \tag{10}$$

Ideally, the cross-correlation coefficient of the decorrelated signals among themselves and with the original signal is negligible, i.e., $\mathbb{E}(x'_L x'_R) = \mathbb{E}(x'_L x_R) = \mathbb{E}(x_L, x'_R) \approx 0$. The weighting of $\cos^2(\beta)$ ensures that the ICC of the output signal is lesser than the ICC of the input signal.

3.2. Transient handling

Both allpass and velvet noise decorrelators will smear transients in time, thereby degrading performance for percussive signals. Inspired by decorrelators in the literature that have a transient handling block [14], we implement a similar logic in our plugin. A transient detection block is placed at the end of the signal chain. If an onset is detected, the input and widener output buffers are cross-faded. While a transient is active, only the input buffer is passed to the output. During an offset, the widener output and input buffer are again cross-faded. Between an offset and the next onset, only the widener output is passed to the output. A hold counter ensures that a false offset is not detected, and an inhibit counter ensures that a minimum number of frames exist between onsets [14].

To detect an onset, the input signal envelope and a running mean of the envelope is tracked with a leaky integrator. The leaky integrator has a fast attack time, τ_a s and a slow release time, τ_r s. The signal envelope, $e(n)$ is calculated from the input signal, $x(n)$, as

$$\begin{aligned}&\text{if } x(n) > e(n): \\ &\qquad e(n) = e(n-1) + (1 - e^{\frac{-1}{\tau_a f_s}})\,(|x(n)| - e(n-1)) \\ &\text{else:} \\ &\qquad e(n) = e(n-1) + (1 - e^{\frac{-1}{\tau_r f_s}})\,(|x(n)| - e(n-1))\end{aligned} \tag{11}$$

The running mean of the signal envelope, $m(n)$, is tracked. An an adaptive threshold, $t(n)$, is calculated dynamically from the running mean, and is updated with a forgetting factor, α. The forgetting factor switches between two states depending on whether a local maxima in the signal envelope is detected. If a local maxima is detected, the higher forgetting factor ensures that more emphasis is placed on the current running mean value in calculating the current threshold. Otherwise, the last threshold value is weighted more.

$$\begin{aligned}m(n) &= \frac{1}{n}\,e(n) + \left(1 - \frac{1}{n}\right)\,m(n-1) \\ t(n) &= \alpha\,m(n) + (1-\alpha)\,t(n-1) \\ \alpha &= \begin{cases}0.99, & \text{if } e(n-2) < e(n-1) \text{ and } e(n-1) > e(n) \\ 0.01, & \text{otherwise}\end{cases}\end{aligned} \tag{12}$$

An onset is detected if the signal envelope is rising and crosses the threshold. An offset is detected if an onset was detected previously, and the signal envelope is falling and is below the threshold. The logic of the transient handling algorithm is shown in Fig. 2.

4. FREQUENCY-DEPENDENT STEREO WIDENING

As mentioned in Sec. 1, commercial stereo widening plugins have different width controls in different frequency bands. Lower frequencies should sum coherently for a strong phantom centre, and the mid-and high-frequencies can be spread more to increase the sense of perceived width. To control the width of the high frequencies and low frequencies separately, we propose a frequency-dependent widening architecture, shown in Fig. 3 and described below.

The output from the decorrelators and the original stereo signal is passed through Linkwitz-Riley crossover filters [15] to get the lowpassed signals and highpassed signals respectively. Let the lowpassed and highpassed versions of the original stereo signal be $x_{L,R_{\text{low,high}}}$ and those of the decorrelated signals be $x'_{L,R_{\text{low,high}}}$. The amplitude-preserving filterbank ensures that,

$$\begin{aligned} x_{L,R} &= x_{L,R_{\text{low}}} + x_{L,R_{\text{high}}} \\ x'_{L,R} &= x'_{L,R_{\text{low}}} + x'_{L,R_{\text{high}}} \end{aligned} \tag{13}$$

Alternately, a power-preserving filterbank may also be used, such as a high-order Butterworth crossover. There is an option in the plugin to switch between amplitude preserving and power preserving filterbanks.

Now, two controllable parameters, β_{low} and β_{high} determine the low and high frequency mixing angles respectively. The output signal is given by,

$$\begin{aligned} x_{L,R_{\text{out}}} =& \sin(\beta_{\text{low}}) x'_{L,R_{\text{low}}} + \sin(\beta_{\text{high}}) x'_{L,R_{\text{high}}} \\ &+ \cos(\beta_{\text{low}}) x_{L,R_{\text{low}}} + \cos(\beta_{\text{high}}) x_{L,R_{\text{high}}} \end{aligned} \tag{15}$$

Similarly, the mid-side ratio is given by (14). If $\beta_{\text{low}} = \beta_{\text{high}}$, then the frequency dependent widener gives the same output as the frequency-independent widener. Generally, $\beta_{\text{low}} < \beta_{\text{high}}$ will ensure that the low frequencies are spread less than the high frequencies. This is to mimic the fact that longer wavelengths sum coherently whereas shorter wavelengths are scattered. The crossover frequency is another controllable parameter that determines the cutoff between the low and high frequencies. Similar to the broadband architecture, a transient handling block is added at the end of the signal chain.

5. EVALUATION

5.1. Decorrelation evaluation

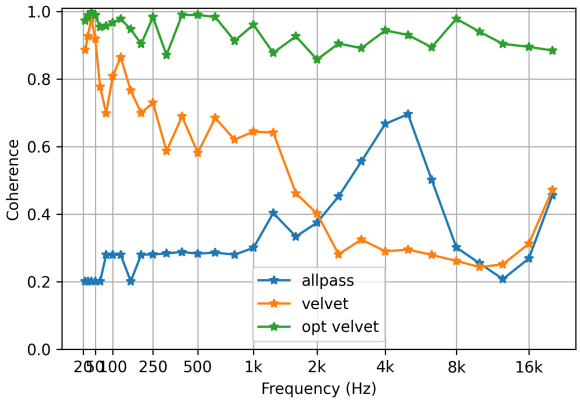


Figure 4: Decorrelation filters' median interchannel coherence over 100 different initialisations, in third octave bands.

To evaluate the different methods of decorrelation discussed in the paper, we pass an impulse through the decorrelating filters and measure the interchannel coherence for the entire sequence, as well as the magnitude spectrum of the filters. Two impulses of length 15 ms starting at time zero are passed through the allpass filters. The allpass filters have a total of 200 biquads and a maximum group delay of 30 ms. The VN and OVN filters are of the same length, 15 ms, have a density of 1 sample/ms and a decay of -60 dB. The resulting IRs obtained are the decorrelating filters for the input stereo signal. The experiment is repeated 100 times with different initialisation seeds.

The interchannel coherence (IC) between two signals, $\boldsymbol{x}_L, \boldsymbol{x}_R$, per frequency band, f is defined as:

$$\rho_{\boldsymbol{x}_L,\boldsymbol{x}_R}(f) = \frac{\boldsymbol{x}_L(f)^T \boldsymbol{x}_R(f)}{\sqrt{\boldsymbol{x}_L(f)^T \boldsymbol{x}_L(f)\, \boldsymbol{x}_R(f)^T \boldsymbol{x}_R(f)}}. \tag{16}$$

An IC value of 1.0 means the signals are fully correlated, whereas that of 0 indicates they are fully decorrelated.

The median IC (over all repetitions) for the three filters, calculated in third octave bands, are shown in Fig. 4. The allpass filters perform especially well in the lower frequencies upto 1 kHz. However, the ICs of the allpass filters peak in the region between $2-8$ kHz. The VN filters perform better in this frequency range, although its low frequency performance is poor. The OVN filters perform worst of the lot, with a consistently high median IC. The cost function of the OVN does not actually minimise the correlation between any two sequences, but optimises them individually to get a flatter response. Therefore, they perform worse than the VN sequences in terms of the median IC.

The magnitude response plots for the IR pair, which gives the minimum mean absolute coherence for all three methods, are shown in Fig 5. The best OVN filter pairs minimises a weighted sum of the mean absolute coherence and spectral flatness to achieve a trade-off [11]. As expected, the allpass filters have the flattest response, followed by the optimised velvet noise and the velvet noise filters. The optimisation does a good job of flattening the magnitude response of the original velvet noise sequence.

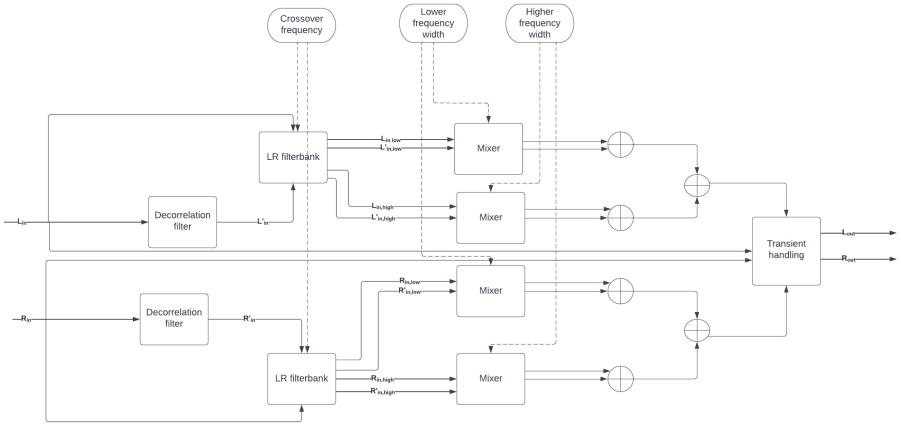


Figure 3: Frequency-dependent stereo widener block diagram.

$$\begin{aligned}\gamma_{\text{out, amp}} &= \frac{\sin(\beta_{\text{low}})(x'_{L_{\text{low}}} + x'_{R_{\text{low}}}) + \sin(\beta_{\text{high}})(x'_{L_{\text{high}}} + x'_{R_{\text{high}}}) + \cos(\beta_{\text{low}})(x_{L_{\text{low}}} + x_{R_{\text{low}}}) + \cos(\beta_{\text{high}})(x_{L_{\text{high}}} + x_{R_{\text{high}}})}{\sin(\beta_{\text{low}})(x'_{L_{\text{low}}} - x'_{R_{\text{low}}}) + \sin(\beta_{\text{high}})(x'_{L_{\text{high}}} - x'_{R_{\text{high}}}) + \cos(\beta_{\text{low}})(x_{L_{\text{low}}} - x_{R_{\text{low}}}) + \cos(\beta_{\text{high}})(x_{L_{\text{high}}} - x_{R_{\text{high}}})} \\ &= \frac{\sin(\beta)(x'_L + x'_R) + \cos(\beta)(x_L + x_R)}{\sin(\beta)(x'_L - x'_R) + \cos(\beta)(x_L - x_R)} \iff \beta_{\text{low}} = \beta_{\text{high}} = \beta.\end{aligned} \tag{14}$$

5.2. Stereo widener evaluation

5.2.1. *Broadband widener evaluation*

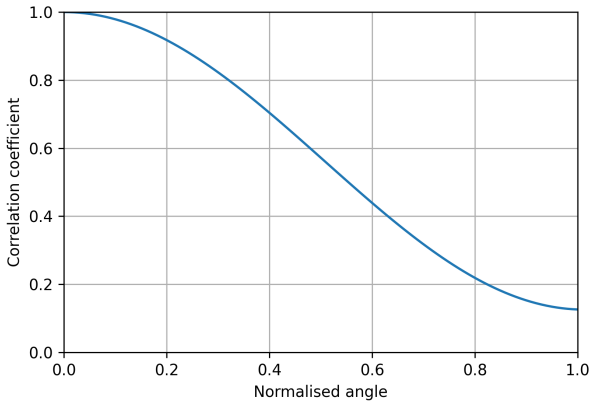


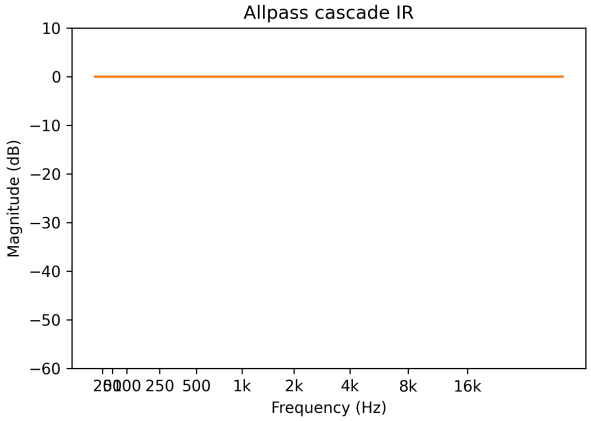
Figure 6: Correlation coefficient vs $\beta/\frac{\pi}{2}$ for the broadband stereo widener

To evaluate the proposed stereo widener architecture, we plot the ICC as a function of β for the broadband widener (Fig. 6). The input signal is a 2 s sinusoidal chirp in the left and right channels (duplicated mono, fully correlated). The allpass decorrelator is used as it was shown to be the most timbrally neutral. As expected, the ICC decreases with increasing β, it is maximum at $\beta = 0$ and minimum at $\beta = \pi/2$. The shape of the ICC as a function of β matches (10), but zero ICC at $\beta = \pi/2$ is not achieved because the decorrelating filters are not ideal. Nevertheless, the broadband stereo widening architecture proposed in the paper achieves its desired effect.

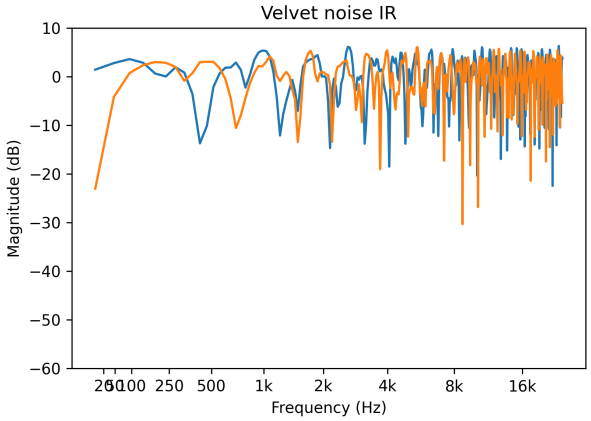
To check the transient response, we passed a stereo drum and bass track through the widener and with and without transient handling. We used the allpass decorrelators with a maximum width of $\beta = \pi/2$. Spectrograms of a 2 s excerpt of the output signal that zooms in on the transients are shown in Fig. 7. Starting from the left, the first spectrogram shows the inout signal, the middle spectrogram shows the widener output without transient handling and the last shows the widener output with transient handling. The two strongest transients in Fig. 7b have non-uniform group delay across frequencies, where the lower frequencies are delayed more (encircled in figure). This is audible as a chirping effect. The transient handling block mitigates this effect significantly, as shown in Fig. 7c.

5.2.2. *Frequency-dependent widener evaluation*

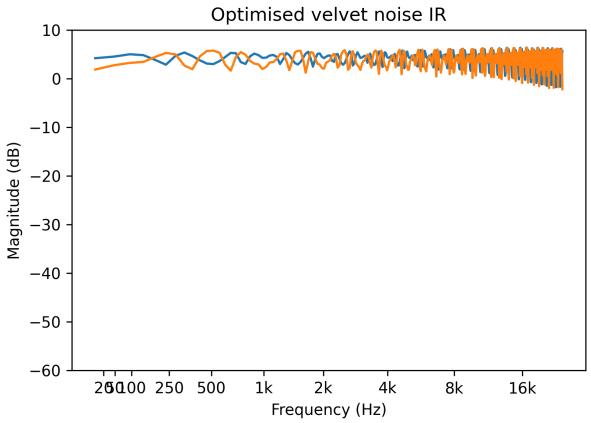
To test the frequency-dependent stereo widener, we pass a sinusoidal chirp signal through the proposed architecture in Fig. 3 that has a Linkwitz Riley filter with a crossover at 250 Hz. The allpass decorrelation filters are initialised with random seeds and the experiment is repeated 100 times. We plot the median of the interchannel coherence in third octave bands for two cases – $\beta_{\text{low}} = 0, \beta_{\text{high}} = \pi/2$ and $\beta_{\text{low}} = \pi/2, \beta_{\text{high}} = 0$ in Fig. 8. We can clearly see the crossover at 250 Hz. The expected outcome is to see the blue lines dip at the crossover frequency and stabilise at that value, and the orange line peak at the crossover frequency and then stabilise. The latter is achieved but the former is not. The blue line is not monotonically decreasing but the overall shape shows a reduction in ICC except near the Nyquist frequency.



(a) Allpass



(b) Velvet noise



(c) Optimised Velvet noise

Figure 5: Magnitude response of the different decorrelators that give the minimum average IC.

5.3. Plugin

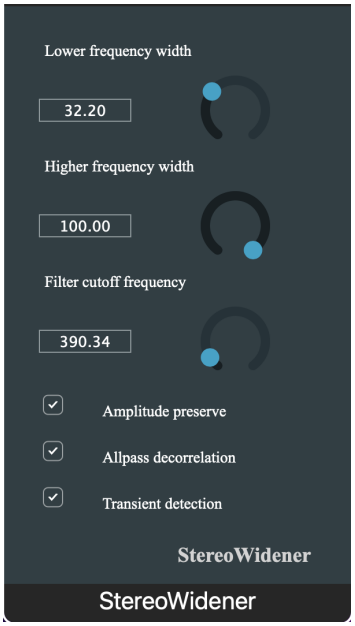
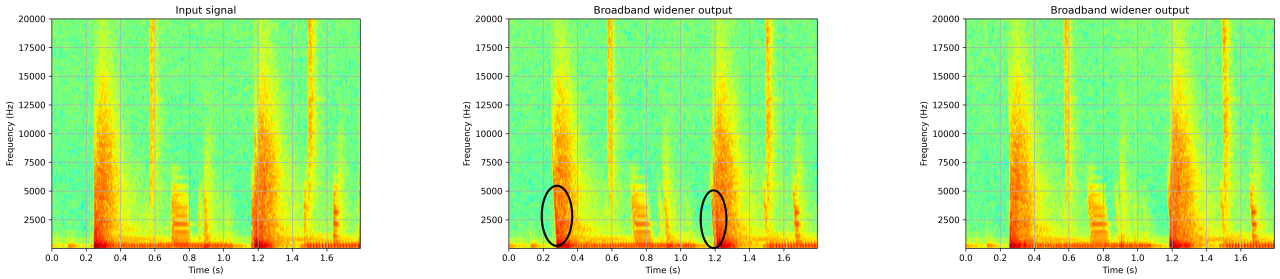


Figure 9: Stereo widening plugin interface.

The plugin was implemented using the JUCE framework[3]. Its interface is shown in Fig. 9. The plugin is open source and is available on Github[4]. The low frequency and high frequency width knobs control β_{low} and β_{high} respectively. Their values range from $0-100$ which is mapped linearly to the range $0-\frac{\pi}{2}$ rad. The filter cutoff frequency knob controls the crossover frequency of the filterbank, and ranges from $100-4000$ Hz. All controllable parameters are smoothed with a one-pole filter. The check-boxes allow switching between 1) amplitude and energy preserving filterbanks, 2) allpass filters (implemented as a cascade of biquads) and velvet noise filters (implemented with tapped delay lines), 3) turn transient handling on and off. After some listening, we decided to proceed with the optimised velvet noise filter pair that gave the minimum absolute coherence whilst maintaining a flat magnitude response among the 100 sequences in our previous experiment, as it reduced comb filtering artifacts compared to the non-optimised velvet filters.

For comparison, we also implemented an offline version of the ITD based stereo widener [2]. For this, we used the spherical head HRTF model proposed by Brown et al. [16]. We created a set of HRIRs corresponding to source angles in the range of $[-90°, +90°]$, with the positive angles denoting a source on the left and negative angles denoting a source on the right. For a standard stereo setup, the direct and cross-talk paths are simulated in the output for a source angle of $\pm 30°$. In our implementation, the source angle is a tunable parameter that varies from $[0, 90°]$, where $0°$ gives the original stereo track, and $90°$ gives a maximally wide stereo track (the source angle is negated for the right speaker). Depending on the value of the source angle, the corresponding direct and cross-talk HRIRs are looked up from the HRIR set, and convolved with the input signal.

[3] https://juce.com/
[4] https://github.com/orchidas/StereoWidener



(a) Input signal

(b) Broadband widener output without transient handling

(c) Broadband widener output with transient handling

Figure 7: Spectrograms of an excerpt of a drum and bass track with strong transients. The first plot shows the spectrogram of the input signal, the next plot shows the spectrogram of the broadband widener without transient handling and the last plot shows the spectrogram of the broadband widener with transient handling.

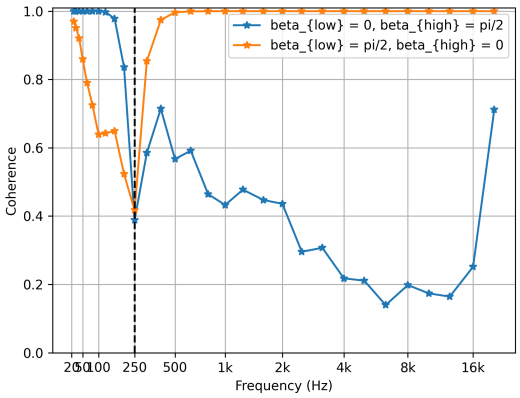


Figure 8: Interchannel coherence for the frequency-dependent stereo widener for different pairs of $\beta_{\text{low}}, \beta_{\text{high}}$. The black dashed line represents the cutoff frequency of the filterbank.

Sound examples are available at `https://ccrma.stanford.edu/~orchi/StereoWidener/audio_examples.html`. The allpass filters are better at preserving the overall colouration, but at a width of 100 %, the dispersive effects are audible, due to the non-uniform group delay of the filters. These timbral degradations are somewhat mitigated by activating the transient handling. The optimised velvet noise filters produce a much stronger widening effect, whilst introducing some timbral changes. There is an overall reduction in brightness with these filters, however, no major degradation is introduced, and the transient response is much better. In general, when a transient is detected, there is audible gain mismatch between the input and the widener's output. This will be fixed in a future release of the plugin.

The ITD based widener also successfully widens the perceived width, and allows for more accurate control of the stereo angle. However, our proposed method has certain advantages, as it operates solely in the time-domain (no overlap-add needed for HRIR filtering), is very lightweight, and the playback can be both out-loud and on headphones. For ITD based wideners, out-loud playback is an issue as accurately simulated interaural cues will be distorted by the cross-talk path from the physical speakers, thus giving less control over the width parameter. The ITD based stereo widener architecture can also be implemented in the time domain, using dynamically interpolating delay lines [17] to vary the ITD (and hence, source width). However, this is more expensive than the static tapped delay line implementation of the velvet filters.

6. CONCLUSION

In this paper, we have proposed a stereo widening architecture that is based on decorrelation and mixing (as opposed to interaural cue modification), and have provided its real-time implementation as an open-source plugin. The plugin can be used to widen a single source, or in the mastering chain to widen a whole scene.

The incoming stereo (or dual mono signal) is first decorrelated, and the original and decorrelated signal are mixed based on a controllable "perceived width" parameter. A transient handling block prevents the smearing of transients that are audible when the width parameter is high. We have compared different decorrelators, and given the user flexibility to choose among them when using the plugin. Two architectures have been proposed – a broadband stereo widener, which has only one controllable width parameter, and a frequency-dependent stereo widener, that has two controllable width parameters in the low and high frequency bands respectively, with a controllable cutoff frequency. Our experiments have shown that the broadband is able to lower the interchannel correlation and frequency-dependent widener is able to lower the interchannel coherence of the output stereo signal when the input signal is dual-mono. Sound examples have been provided for different input tracks, and compared with an ITD-based widener. Further blind listening tests should be conducted to evaluate the efficacy of the proposed architectures compared to other stereo wideners proposed in the literature. Gain matching during onsets and offsets for transient handling should also be implemented.

7. REFERENCES

[1] Ronald M Aarts, "Phantom sources applied to stereo-base widening," *Journal of the Audio Engineering Society*, vol. 48, no. 3, pp. 181–189, 2000.

[2] Ole Kirkeby, "A balanced stereo widening network for headphones," in *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.

[3] SMA Basha, Abhinav Gupta, and Anshul Sharma, "Stereo widening system using binaural cues for headphones," in *Proc. of International Conference on Signal Processing and Communication Systems, Australia*, 2007, pp. 378–382.

[4] Fritz Menzer and Christof Faller, "Stereo-to-binaural conversion using interaural coherence matching," in *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

[5] Vesa Välimäki, Heidi-Maria Lehtonen, and Marko Takanen, "A perceptual study on velvet noise and its variants at different pulse densities," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1481–1488, 2013.

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

[7] Guillaume Potard and Ian Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3d audio displays," in *Proceedings of the International Conference on Digital Audio Effects (DAFx 2004)*, 2004.

[8] Vesa Välimäki, Bo Holm-Rasmussen, Benoit Alary, and Heidi-Maria Lehtonen, "Late reverberation synthesis using filtered velvet noise," *Applied Sciences*, vol. 7, no. 5, pp. 483, 2017.

[9] Benoit Alary, Archontis Politis, Vesa Välimäki, et al., "Velvet-noise decorrelator," in *Proceedings of the International Conference on Digital Audio Effects (DAFx 2017)*, 2017, pp. 405–411.

[10] Elliot Kermit-Canfield and Jonathan Abel, "Signal decorrelation using perceptually informed allpass filters," in *Proceedings of the International Conference on Digital Audio Effects, (DAFx 2016)*, 2016, pp. 225–31.

[11] Sebastian J Schlecht, Benoit Alary, Vesa Välimäki, Emanuël AP Habets, et al., "Optimized velvet-noise decorrelator," in *Proceedings of the International Conference on Digital Audio Effects (DAFx-17) (DAFx 2018)*, 2018, pp. 87–94.

[12] Nils Meyer-Kahlen, Sebastian J Schlecht, and Vesa Välimäki, "Colours of velvet noise," *Electronics Letters*, vol. 58, no. 12, pp. 495–497, 2022.

[13] David Griesinger, "Stereo and surround panning in practice," in *Audio Engineering Society Convention 112*. Audio Engineering Society, 2002.

[14] Sascha Disch, "Decorrelation for immersive audio applications and sound effects," in *Proceedings of the International Conference on Digital Audio Effects (DAFx 2023)*, 2023.

[15] Siegfried H Linkwitz, "Active crossover networks for noncoincident drivers," *Journal of the Audio Engineering Society*, vol. 24, no. 1, pp. 2–8, 1976.

[16] C Phillip Brown and Richard O Duda, "A structural model for binaural sound synthesis," *IEEE transactions on speech and audio processing*, vol. 6, no. 5, pp. 476–488, 1998.

[17] Andreas Franck, "Efficient algorithms and structures for fractional delay filtering based on lagrange interpolation," *Journal of the Audio Engineering Society*, vol. 56, no. 12, pp. 1036–1056, 2009.