

Audio Signal Decorrelation for Phantom Source Widening

Seminararbeit aus Algorithmen in Akustik und Computermusik 2, SE

Hannes PESSENTHEINER

Betreuung: Dr Franz ZOTTER, DI Matthias FRANK

Graz, April 15, 2011



institut für elektronische musik und akustik



Abstract

This work has a focus on audio signal decorrelation for phantom source widening. It describes different implementations of a dual-channel audio signal decorrelator, based on a single allpass filter, and its characteristics. These characteristics are shown in spectral-, cepstral- and time-domain to distinguish the relation between these domains by changing spectral and cepstral coefficients of the transformed group delay in a certain manner, e.g., the computation of an oscillation with one or more periods spread over the whole spectrum, or changing the cepstral coefficients randomly. Furthermore, the signal pathway of the overall system and its mathematical description are discussed in detail.

Contents

1	Introduction and Motivation	4
2	Decorrelation and Its Effects	4
3	New Techniques to Generate Decorrelated Signals	5
3.1	State of the Art	5
3.2	Design in the Imaginary Cepstrum	5
3.3	The Implementations	7
3.3.1	The Stochastic Manipulation of the Group Delay	7
3.3.2	The Limited Stochastic Manipulation of the Group Delay	12
3.3.3	The Deterministic Manipulation of the Group Delay	14
4	Conclusion	16
5	References	17

1 Introduction and Motivation

The output of mono audio signals with two loudspeakers, e.g., used at the computer workstation, generates a phantom source, which is located in front of the listener, if both loudspeakers are placed symmetrically around the screen. To modify the sound experience, a dual-channel audio signal decorrelator is implemented to widen the image of the phantom source. This may lead to the effect of perceiving a stretched image and being surrounded of the sound. In some configurations, decorrelation may also introduce spatiality, width and depth occurring in natural environments.

The effect of source widening can also be used in multi-channel audio systems, e.g., Ambisonics, to avoid comb filters or phasing effects when many loudspeakers play the same signal. To handle this problem, a multi-channel audio signal decorrelator, which is an extension of the model described in this work, can be used to expand the image of the sound source generated by an active loudspeaker, so above-mentioned artefacts (interference effects) are attenuated.

2 Decorrelation and Its Effects

As mentioned previously, audio signal decorrelation leads to a widening of a mono sound source, which can be described as a stretched sound image, e.g., all instruments stay at the same position, but they increase their size horizontally.

According to [Ken95] decorrelation features five effects on the perception of spatial imagery:

1. timbral coloration, because of some unavoidable constructive and destructive interferences,
2. spatial diffuseness in the reproduced sound field,
3. externalization in headphone reproduction,
4. no image shift of the sound source,
5. no precedence effect, if group delays are not too big and uniform;

A prerequisite to generate the effect of decorrelation is the introduction of two or more channels. In this work, a dual-channel system with two loudspeakers is assumed.

A single audio source signal is transformed into two signals, where both signals exhibit different phase, i.e. group delay, spectrums to produce decorrelation. Despite both signals may have different waveforms due to phase changes, both signals sound equally if they are played separately. According to [Abe06] and [Bou04], pure phase changes are allpass filters, which clearly avoid changes in the amplitude spectrum to maintain the original spectral characteristics of the sound.

3 New Techniques to Generate Decorrelated Signals

3.1 State of the Art

A simple and well known technique to produce decorrelated replicas of a given audio signal is the convolution of an audio signal with a variety of unique allpass impulse responses in parallel as shown in figure 1 and mentioned in [Ken95]. These impulse responses are based on random phase or random group delay calculations performed in the frequency and/or cepstral domain.

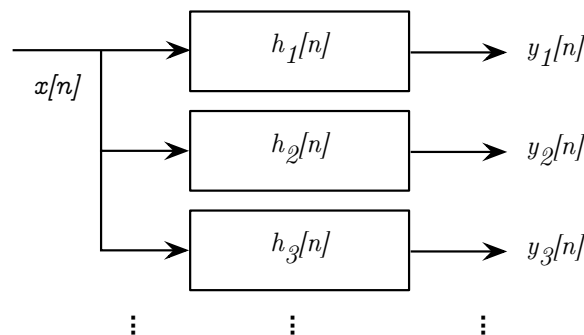


Figure 1: An audio signal is convolved with the impulse response of a variety of allpass filters in parallel. Each convolution leads to a signal with a manipulated group delay and phase spectrum.

As it turned out during this work, uniformly distributed random phase or group delay might be unrewarding, and its outcome is hard to control. E.g., with a random group delay, the same maximum group delay may occur quite often at many frequencies, yielding poor impulse behaviour, it will split phantom source images and will lead to timbral changes. Thus, an ordinary random process might not offer sufficient control for satisfying results. Moreover, a finite impulse response might suffer from cyclic wrap-around artifacts.

3.2 Design in the Imaginary Cepstrum

The imaginary Cepstrum provides the computational means to design all upcoming all-pass techniques in this work. Its schematic is plotted below, and it is complemented by the spectral derivation theorem of the Fourier transform for group-delay design:

Figure 2 is divided into two parts. The shaded part may be interesting, if an impulse response in time-domain would be generated, which exhibits allpass characteristics in case of signal decorrelation. Indeed, it is very difficult to design an impulse response in

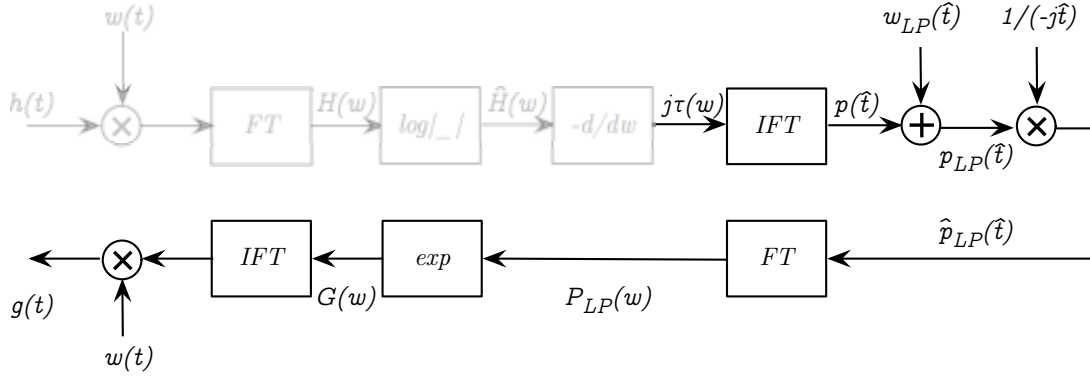


Figure 2: FIR Allpass design in the imaginary Cepstrum is divided into two parts. The shaded part is only depicted because of completeness of the involved transform relations.

time-domain, which exhibits all necessary properties. Thus, this part can be ignored and is only depicted because of completeness.

Considerations start in frequency-domain. First, let's have a focus on the frequency response of an allpass filter:

$$H(\omega) = 1e^{j\phi(\omega)}. \quad (1)$$

Eq. (1) describes an allpass filter, where all frequencies have a constant magnitude of $|H(\omega)| = 1$. Only the phase consists of non-constant values. The frequency response is manipulated by using the natural logarithm. This yields the following equation:

$$\hat{H}(\omega) = \ln(1) + j\phi(\omega) = j\phi(\omega). \quad (2)$$

After this, the manipulated frequency response is derivated with respect to the radial frequency ω and multiplied with -1 , which leads to the group delay

$$j\tau(\omega) = -\frac{dj\phi(\omega)}{d\omega}. \quad (3)$$

Afterwards, the group delay is inverse Fourier transformed, which results in a cepstral representation of the group delay. This is the starting point of all further computations

$$p(\hat{n}) = \frac{1}{2\pi}c(\hat{n}) \int_{-\pi}^{+\pi} \frac{d(-j\phi(\omega))}{d\omega} e^{j\omega\hat{n}} d\omega = \frac{1}{2\pi} \frac{1}{(-j\hat{n})} \int_{-\pi}^{+\pi} \frac{d(-j\phi(\omega))}{d\omega} e^{j\omega\hat{n}} d\omega. \quad (4)$$

As one can see, an additional expression appears in eq. (4):

$$c(\hat{n}) = \frac{1}{(-j\hat{n})}; \quad (5)$$

This expression leads to a cepstral representation of the phase spectrum without using an integrator. A convenient theorem is used to obtain the phase representation:

$$H(\omega) = \int_{-\infty}^{+\infty} h(t)e^{-j\omega t} dt, \quad (6)$$

$$\frac{dH(\omega)}{d\omega} = \int_{-\infty}^{+\infty} h(t)(-jt)e^{-j\omega t} dt. \quad (7)$$

Thus, the group delay can be calculated in the following way:

$$j\tau(\omega) = -\frac{dj\phi(\omega)}{d\omega} = \sum_{-\infty}^{+\infty} p(\hat{n})(-j\hat{n})e^{-j\omega\hat{n}} = DTFT\{p(\hat{n})(-j\hat{n})\}. \quad (8)$$

Two essential steps can be considered to control the resulting impulse response length:

- truncation of the cepstrum in \hat{n} reduces the variation of the group delay over frequency, and thus the impulse response length,
- limiting/scaling the magnitude of the group delay influences the impulse response length.

After calculating $p(\hat{n})$ and multiplying it with the additional expression mentioned above, it has to be Fourier transformed as shown in figure 1. The resulting spectrum has to be used in the exponent of an Euler-Function to invert the logarithm considered before the cepstral domain. One last inverse Fourier transformation of the resulting spectrum has to be calculated to obtain the desired impulse response.

3.3 The Implementations

Three ideas of decorrelating audio signals are realized:

1. the Stochastic Manipulation of the Group Delay,
2. the Limited Stochastic Manipulation of the Group Delay,
3. the Deterministic Manipulation of the Group Delay,

All implementations are based on manipulations in the frequency and/or the cepstral domain and generate two allpass impulse responses.

3.3.1 The Stochastic Manipulation of the Group Delay

This realization of a dual-channel audio signal decorrelator is based on a stochastic manipulation of the group delay, which is transformed in cepstral domain to simplify calculations and reduce computation time.

First, one-third octave bands from $20Hz$ to $16255Hz$ are considered to comply with the spectral resolution of hearing. Random values without a proper weighting will lead to a bad impulse behaviour, e.g., a single impulse may sound like a frequency sweep. To avoid this, a weighting function with a decreasing exponential behaviour at higher frequencies is calculated first:

```

1 %GD is the abbreviation of group delay
2 GDmax = zeros(length(bands),1);
3 %GDmax is the maximum group delay in seconds
4 GDmax(1) = 0.015;
5 for i=1:length(bands)-1
6     %Exponential function used for weighting
7     GDmax(i+1) = GDmax(1)*exp(-i/8);
8 end

```

As mentioned before, these computations yield a weighting function GDmax used to decrease the maximum group delay at higher frequencies. The random values are generated as follows:

```

1 %Calculation without steady-component
2 RVmean = zeros(length(GDmax),1);
3 %Generate new amplitude distribution of group delay
4 GDrand = zeros(max(bands)+1,length(bands));
5 for i=1:length(bands)
6     if i==1
7         RV = RVmean(i)+GDmax(i)*2*(rand(bands(i),1)-0.5);
8         GDrand(1:bands(i),i) = RV;
9     else
10        BW = length(bands(i-1)+1:bands(i));
11        RV = RVmean(i)+GDmax(i)*2*(rand(BW,1)-0.5);
12        GDrand(bands(i-1)+1:bands(i),i) = RV;
13    end
14 end

```

As one can see, random values are generated and weighted with the weighting function GDmax. Thus, each random value within a band will exhibit the same weighting value. After further calculations according to figure 2, the results look as follows:

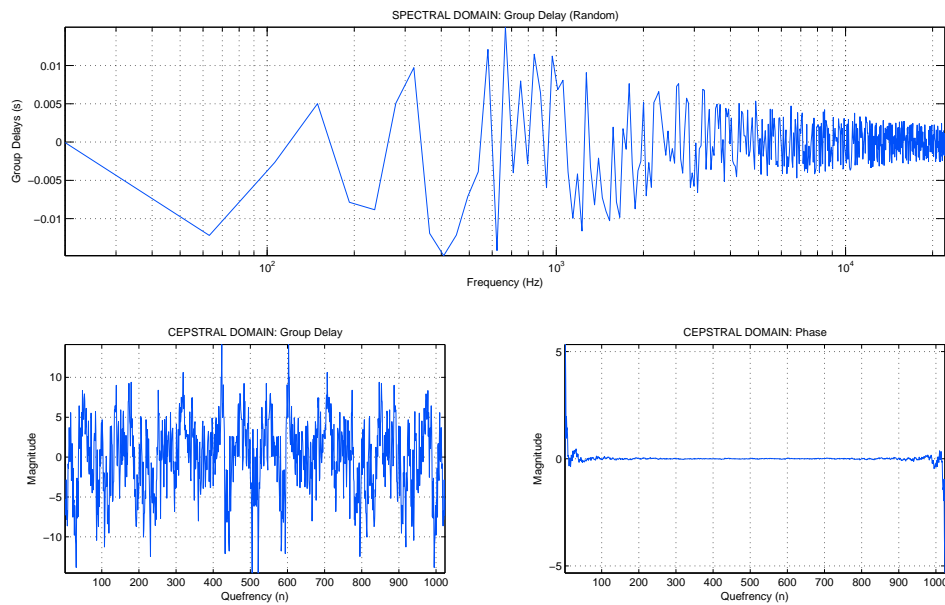


Figure 3: Random group delay and random phase of the **left** channel. Above: Group delay in spectral domain. Left: Group delay in cepstral domain. Right: Phase in cepstral domain.

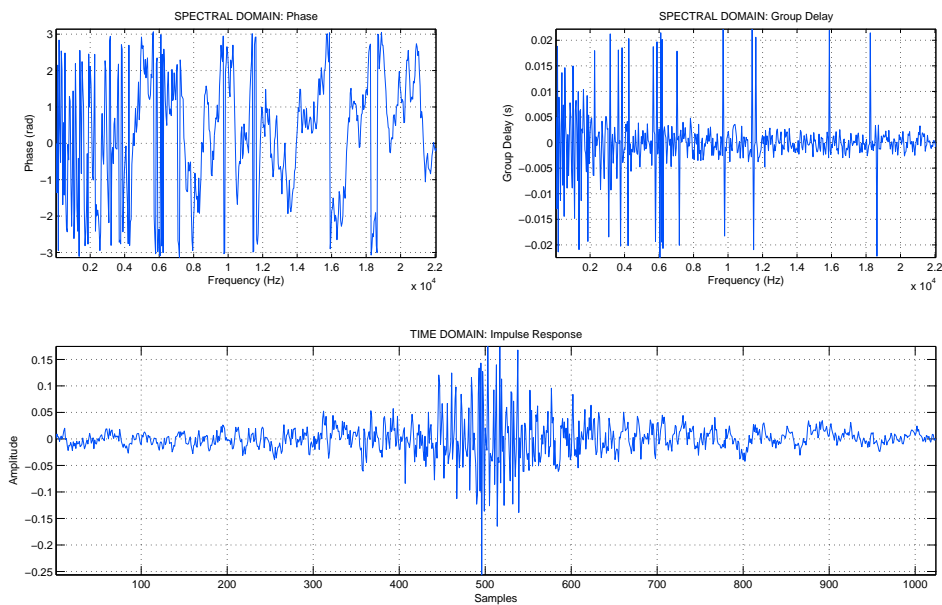


Figure 4: Random group delay, phase and impulse response of the **left** channel. Left: Phase in spectral domain. Right: Group delay in spectral domain. Below: Resulting impulse response.

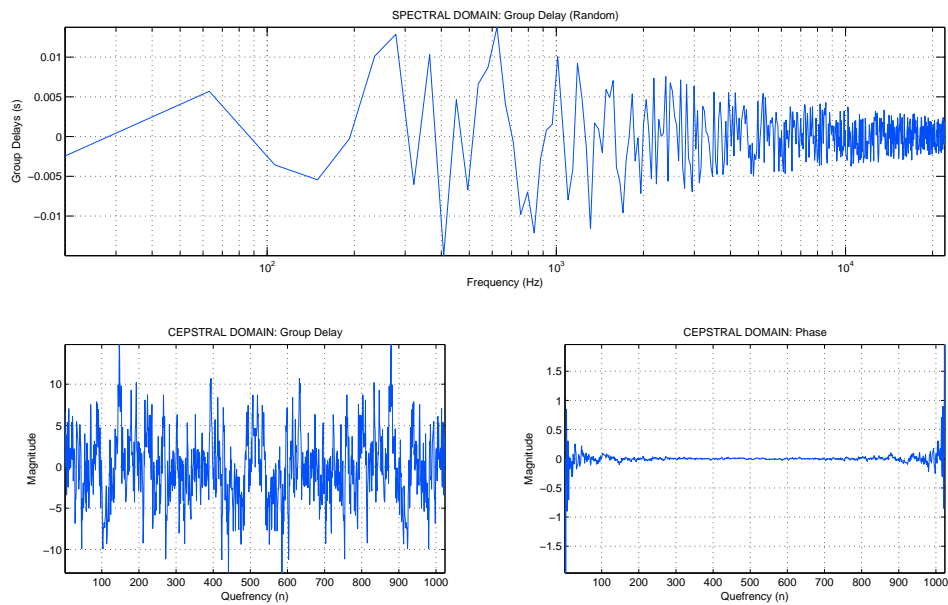


Figure 5: Random group delay and random phase of the **right** channel. Above: Group delay in spectral domain. Left: Group delay in cepstral domain. Right: Phase in cepstral domain.

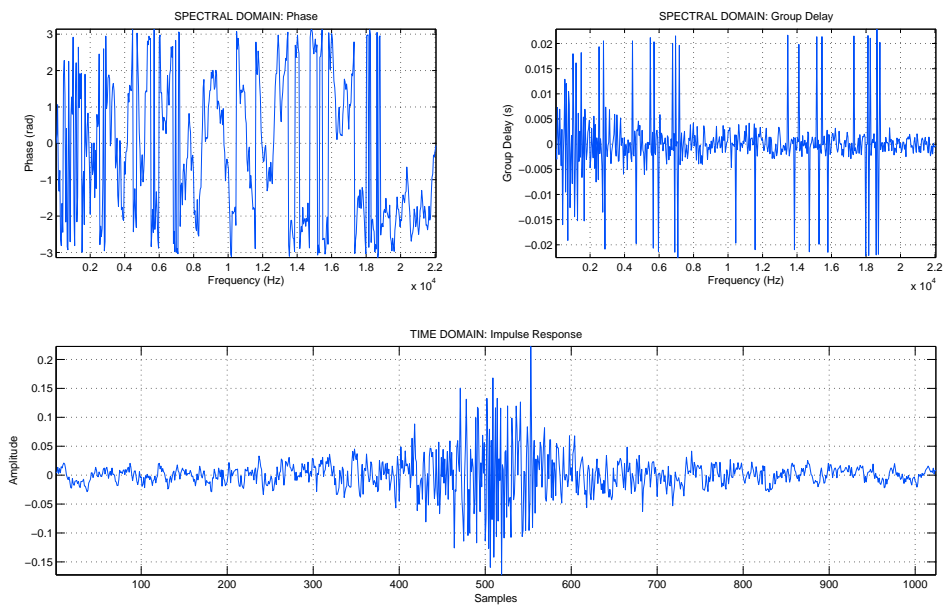


Figure 6: Random group delay, phase and impulse response of the **right** channel. Left: Phase in spectral domain. Right: Group delay in spectral domain. Below: Resulting impulse response.

Left Channel: Figure 3 shows the resulting group delay of the left channel, which exhibits the decreasing exponential behaviour. The higher the frequency, the lower the group delay. In the lower left part of the figure, the symmetric group delay in cepstral domain is depicted. In the lower right part, the phase with its anti-symmetric behaviour in cepstral domain is figured. Figure 4 depicts the resulting phase and group delay in spectral domain and the resulting impulse response, which is spread over 1024 samples.

Right Channel: Figures 5 to 6 describe the right channel. Both figures show the same behaviours as mentioned in the previous figures.

Figure 7 depicts the group delay differences and the auto-correlation between both channels. Obviously, both impulse responses are highly decorrelated as one can see in the lower part of the figure. Using the resulting impulse responses will not always lead to satisfying results, because of its random behaviour over the whole frequency spectrum. Each realization in combination with another one sounds a little different, especially the image width of the phantom source won't be always the same. Therefore, it is difficult to generate impulse responses, which lead to a certain spatial behaviour, i.e. very often the phantom source seems to be perceived as two separated sources at the position of both loudspeakers. Sometimes, uniformly different group delays at low frequencies shift the phantom source to one side.

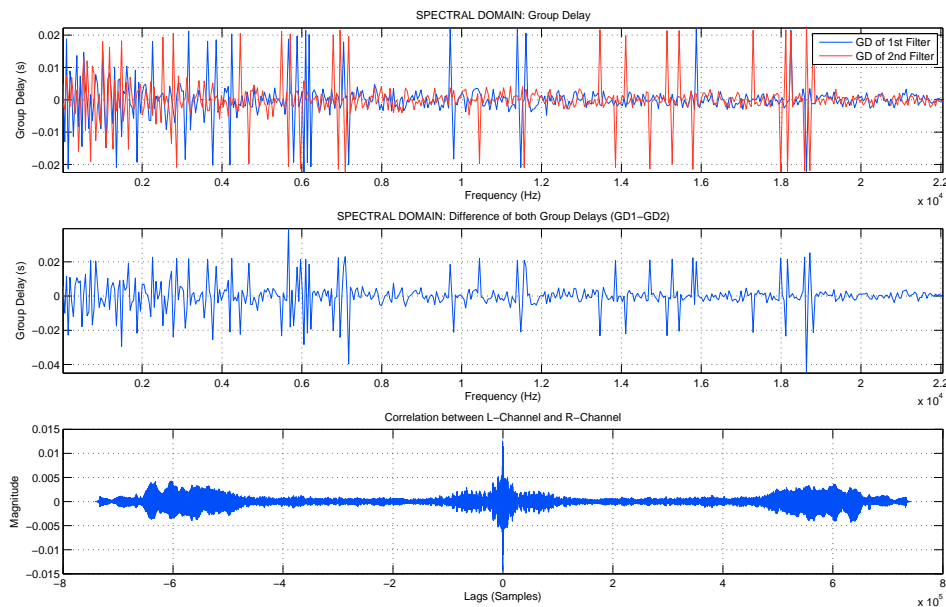


Figure 7: Comparison of both channels (L+R). Above: Group delays of both channels in spectral domain. Middle: Difference between the group delays of both channels in spectral domain. Below: Correlation of the impulse responses between the left and the right channel.

3.3.2 The Limited Stochastic Manipulation of the Group Delay

This realization of a dual-channel audio signal decorrelator is based on a limited stochastic manipulation of the group delay, which is transformed in cepstral domain to simplify calculations and reduce computation time.

In contrast to the previous implementation, this one only exhibits random behaviour for lower frequencies. It yields a group delay with less ripple and stronger low-frequency oscillations as depicted in figure 8 and 9. Therefore, the resulting impulse response is much shorter. It consists of about 100 samples with magnitudes higher than $-50dBFS$, if the impulse response is figured logarithmically. Samples with smaller magnitudes are truncated. The differences between each generated impulse response is much smaller than in the previous implementation. Unfortunately, it is still difficult to find impulse responses with satisfying behaviour, i.e. sometimes the phantom source seems to be perceived as two separated sources at the position of both loudspeakers. However, the use of this implementation yields bad acoustic results and is only discussed, because of demonstrative reasons.

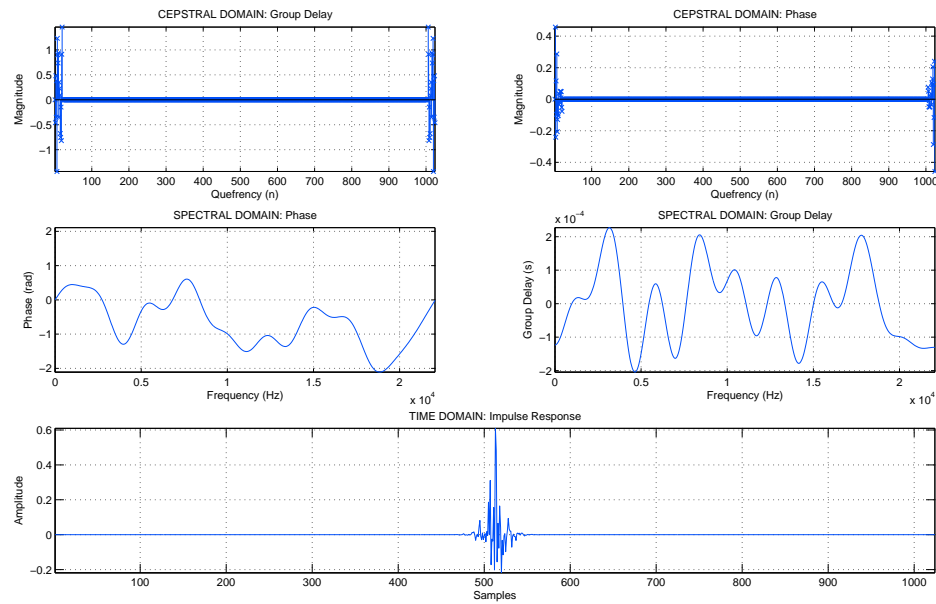


Figure 8: Impulse response properties of the **left** channel. Top-Left: Group Delay in cepstral domain. Top-Right: Phase in cepstral domain. Middle-Left: Phase in spectral domain. Middle-Right: Group-Delay in spectral domain. Below: Impulse response.

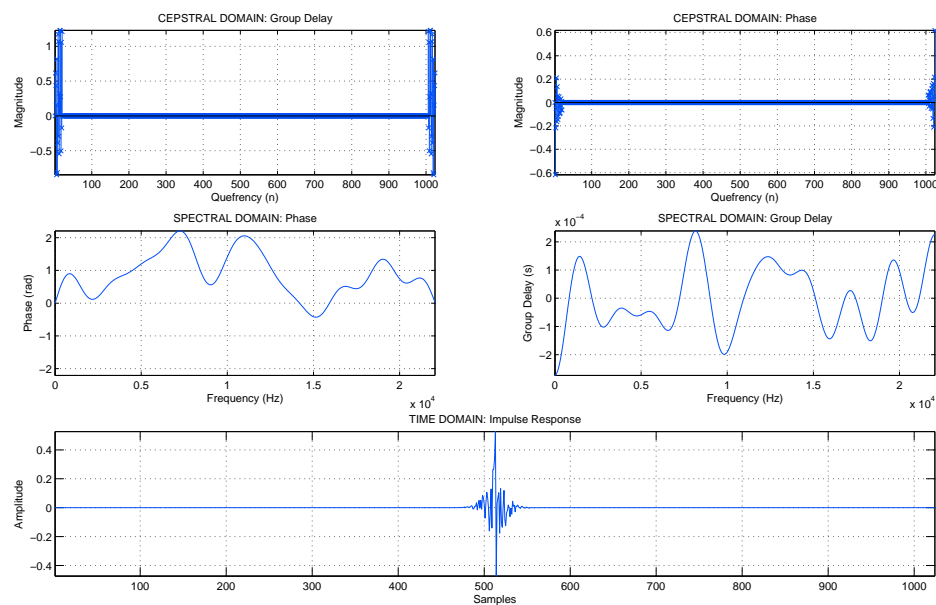


Figure 9: Impulse response properties of the **right** channel. Top-Left: Group Delay in cepstral domain. Top-Right: Phase in cepstral domain. Middle-Left: Phase in spectral domain. Middle-Right: Group-Delay in spectral domain. Below: Impulse response.

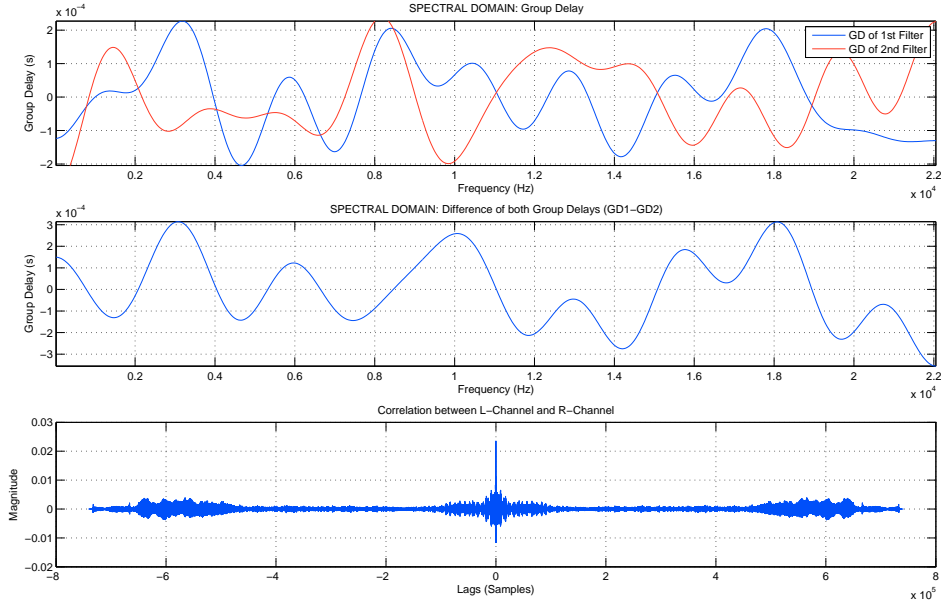


Figure 10: Comparison of both channels (L+R). Above: Group delays of both channels in spectral domain. Middle: Difference between the group delays of both channels in spectral domain. Below: Correlation of the impulse responses between the left and the right channel.

3.3.3 The Deterministic Manipulation of the Group Delay

In comparison to both previous implementations, the following one does not exhibit any random processes. Only a single component in the cepstral group delay is set to a value higher than zero. The magnitude used in this work is exactly 1.5 and the affected quefrequency bin is calculated as follows:

$$\text{round}\left(\frac{f_s}{10000\text{Hz}}\right) = \text{round}\left(\frac{44100\text{Hz}}{10000\text{Hz}}\right) = 4 \quad (9)$$

Thus, there are only two periods of a low-frequency oscillation spread over the whole group delay and phase spectrum. In order to obtain decorrelation, the cepstral component of the second channel group delay has to be inverse in sign, i.e. -1.5 . Both figures (figure 11 and 12) show the same properties. The only difference is the phase-shift of the oscillations. Both resulting impulse responses are very short and similar in shape. Again, truncation occurs for samples with magnitude beneath -50dBFS to reduce the length of the impulse response. The phantom source is widened satisfactorily and perceived as a single source between both loudspeakers. Unfortunately, it is not easy to generate more than two decorrelated impulse responses with this simple scheme.

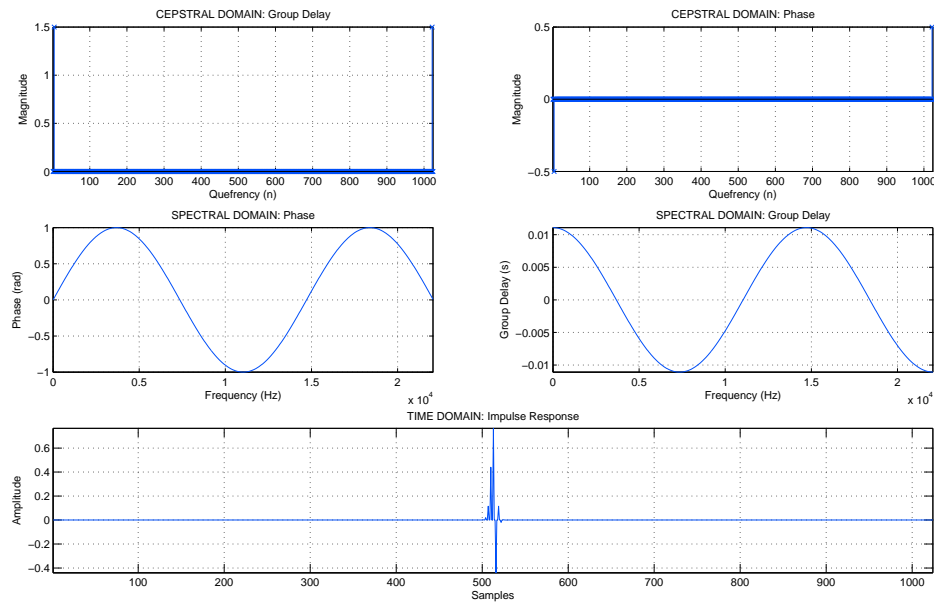


Figure 11: Impulse response properties of the **left** channel. Top-Left: Group Delay in cepstral domain. Top-Right: Phase in cepstral domain. Middle-Left: Phase in spectral domain. Middle-Right: Group-Delay in spectral domain. Below: Impulse response.

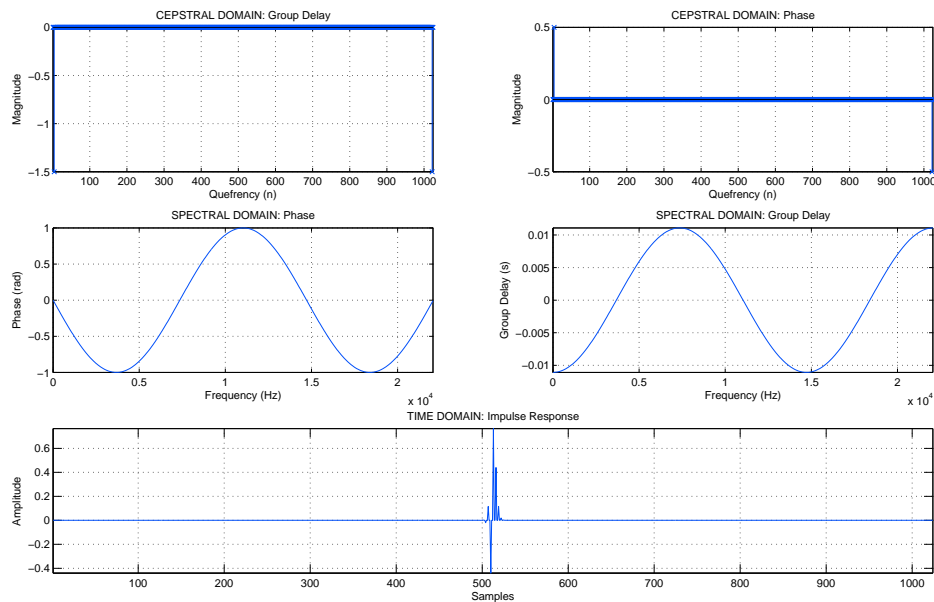


Figure 12: Impulse response properties of the **right** channel. Top-Left: Group Delay in cepstral domain. Top-Right: Phase in cepstral domain. Middle-Left: Phase in spectral domain. Middle-Right: Group-Delay in spectral domain. Below: Impulse response.

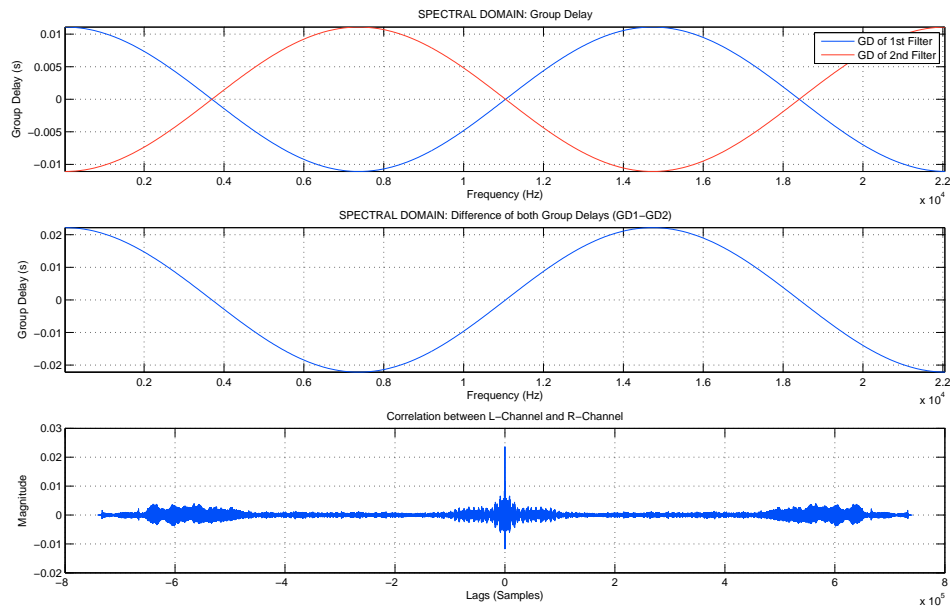


Figure 13: Comparison of both channels (L+R). Above: Group delays of both channels in spectral domain. Middle: Difference between the group delays of both channels in spectral domain. Below: Correlation of the impulse responses between the left and the right channel.

4 Conclusion

Implementations based on random processes can be used to generate an infinitely large number of different impulse responses for audio signal decorrelation. However, not every one will exhibit satisfying properties, i.e. decorrelation is so high that the phantom source seems to be perceived as two separated sources at the position of both loudspeakers. In contrast to the random processes, the third implementation mentioned in this work, i.e. the deterministic manipulation of the group delay, works perfectly for a system with two loudspeakers. The image of the phantom source is widened satisfactorily and perceived as a single source between both loudspeakers. A nice future challenge will be to generate impulse responses for, e.g., a multi-channel system, using a deterministic implementation.

5 References

- [Abe06] J. Abel, "Robust Design of very High-Order Allpass Dispersion Filters," in *Proc. of the 9th Int. Convergence on Digital Audio Effects (DAFx-06)*, Montreal, Canada, September 2006.
- [Bou04] M. Bouéri, "Audio Signal Decorrelation Based on a Critical Band Approach," *AES Convention Paper 6291*, 2004.
- [Ken95] G. Kendall, "The Decorrelation of Audio Signals and its Impact on Spatial Imagery," *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, 1995.