CHALMERS
UNIVERSITY OF TECHNOLOGY

Decorrelation algorithms with application to artificial reverberation

Master's thesis in Sound and Vibration

GUY FAVILL

Department of Architecture and Civil Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis 2020: ACEX30

Decorrelation algorithms with application to artificial reverberation

GUY FAVILL

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
Gothenburg, Sweden 2020

Decorrelation algorithms with application to artificial reverberation
GUY FAVILL

Supervisor: Jens Ahrens, Division of Applied Acoustics
Examiner: Jens Ahrens, Division of Applied Acoustics

Master's Thesis 2020: ACEX30
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2020

Decorrelation algorithms with application to artificial reverberation
Guy Favill
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

Decorrelation is a process which decreases the cross correlation between signals without destroying other qualities of those signals. In relation to audio, that means that the cross correlation of decorrelated signals is reduced but the audio is perceived as sounding the same. Decorrelation is known to reduce the effect of colouration, phantom sources and precedence effect in a multi-channel audio situation, but can also impart a greater spatial extent, diffuseness or headphone externalisation on sound [1].

This thesis has two main parts. It first investigates and expands upon six different known decorrelation algorithms. The performance of these algorithms is analysed by various metrics, such as the extent to which they can decorrelate, impact upon the frequency content of an input signal and the group delay introduced. Three of the six decorrelation algorithms are selected as the best performing in these regards: IIR allpass cascade decorrelation, Sub-band decorrelation and White Noise decorrelation.

The second part of the thesis is to use these three chosen algorithms to decorrelate the reverberant part of a Room Impulse Resonse (RIR) in order to simulate the diffuse sound field of sound in a room on a multi-loudspeaker setup inside a non-reverberant room. A listening test was performed, where participants were asked to rate sounds with different cross correlation or different decorrelation methods in terms of spaciousness and timbre.

The conclusion drawn as a result of this listening test is that for the RIR with a longer reverberation time, a more decorrelated reverberant part of the audio does not result in a more spacious perception, but for the less reverberant RIR, it does. The Sub-band and White noise decorrelation algorithms produce the most perceptually spacious sound.

The most decorrelated signals tested produce the largest difference in timbre when compared to less decorrelated signals, but it is important to note that all levels of decorrelation have a small to moderate difference in timbre when compared.

Further investigation into this topic would most usefully be focused on the cause of an increase in low frequency content of audio for the least decorrelated audio when compared to less decorrelated. This was noted in informal conversation with most of the subjects that participated in the listening test, but not conclusively evidenced in the results of that listening test. It is suspected that this low frequency "booming" characteristic of the stimuli was percieved as spacious sounding, when it was not necessarily an intended part of any decorrelation algorithm.

Keywords: decorrelation, correlation, room impulse response, listening test

Acknowledgements

Here I would like to express my gratitude to my supervisor Jens Ahrens for his guidance and insight throughout the process of researching, experimenting, implementing, testing and writing that I have gone through during my work on this thesis. I would also like to thank all those who participated in the listening test that forms a part of this thesis, and my classmates in the 2018 year of the Sound and Vibration MSc programme, who were always were happy for me to bounce ideas off of them when we were hanging around in the Acoustics Library.

Guy Favill, Gothenburg, July 2020

Contents

List of Figures

List of Tables

1
Introduction

Correlation between channels of an audio signal introduces a number of effects such as colouration, a phantom source position and a precedence effect. Implementing decorrelation techniques on an audio signal removes these effects and also introduces diffuseness, greater spatial extent and even externalisation, whilst the perceived sound retains the same characteristics. Producing two decorrelated channels from one input signal using decorrelation techniques results in a “stereoised” version of the input. When producing more than two decorrelated channels from one input signal, it is possible to manipulate the perceived spatial extent and diffuse qualities of the percieved sound when playing it through multiple loudspeakers [1].

With modest equipment and a small amount of time it is simple to record a high quality Room Impulse Response (RIR). Alternatively, RIRs of spaces around from the world are readily available on many websites for a small fee or even for free. This represents an opportunity for anyone to simulate a recording in a prestigious venue or space without actually having to go there. Often these RIRs only contain one channel, so an interesting application of decorrelation is to produce reverberant audio for multiple channels from a mono RIR and mono input signal which has spatial qualities, where a simple convolution between that RIR and input signal would produce a simple mono reverberant signal. Combining decorrelation techniques with processing which splits the direct and diffuse parts of an RIR, a convincing and enveloping sound is produced.

This report takes as a starting point some existing decorrelation algorithms and expands on these where possible. The performance of these algorithms is analysed by calculation and comparison of various metrics. A listening test is performed to investigate the perceptual impact of a selected number of these decorrelation methods in a real setting. This setting is a circular array of eight loudspeakers surrounding a listener, where one loudspeaker plays the “direct” sound and the other loudspeakers play the “diffuse” sound - i.e. they simulate the response of the room from which the RIR came from, rather than the room that the loudspeakers and listener are located in.

2

Theory

2.1 Cross correlation

Cross correlation is a measure of the similarity of two signals as a function of the displacement in samples (sometimes called "lag") of one of those signals in relation to the other. Here, two signals x and y have the same length. The cross correlation between x and y, $\Omega_{xy}(m)$, is calculated by:

$$\Omega_{xy}(m) = \sum_{n=1}^{N} x(n)y(n+m) \tag{2.1}$$

Where:

$$\begin{aligned} &m = 1, 2, \ldots, 2N-1 \\ &N = \text{length in samples of the signals} \end{aligned} \tag{2.2}$$

In order to avoid negative indexing, y is prepended and appended with N zeros, and the lag index m runs from $m = 1, 2, \ldots, 2N-1$ rather than from $-N, -N+1, \ldots 0, \ldots N-1, N$. That means that the largest possible negative lag is represented by $m = 1$, the largest positive lag is represented by $m = 2N-1$, and the zero lag position is at $m = N$.

The cross correlation is then normalised by the square root of the product of the autocorrelations of x and y at zero lag ($\Omega_{xx}(N)$ and $\Omega_{xx}(N)$).

$$\Omega_{xy\,\text{normalised}}(m) = \frac{\Omega_{xy}(m)}{\sqrt{\Omega_{xx}(N)\Omega_{yy}(N)}} \tag{2.3}$$

The autocorrelations are largest at $m = N$ (zero lag) because that is when the signal matches itself exactly. The normalised cross correlation coefficient is then:

$$\Omega_{xy\,\text{coefficient}} = \Omega_{xy\,\text{normalised}}(N) \tag{2.4}$$

Normalisation is useful as it means that:

- When the two signals are identical, the cross correlation coefficient is 1 (Figure 2.1).
- When the signals are exactly negatively correlated (they are 180° out of phase) then the cross correlation coefficient is -1 (Figure 2.2).

- When the signals are completely uncorrelated, the cross correlation coefficient is zero (Figure 2.3). Note that even two finite randomly generated noise signals will be slightly correlated in some way. Unless one signal has a value of zero at each sample where the other has a non-zero value, the cross correlation coefficient will be non-zero. The cross correlation coefficient will also be zero if both signals are of infinite length.

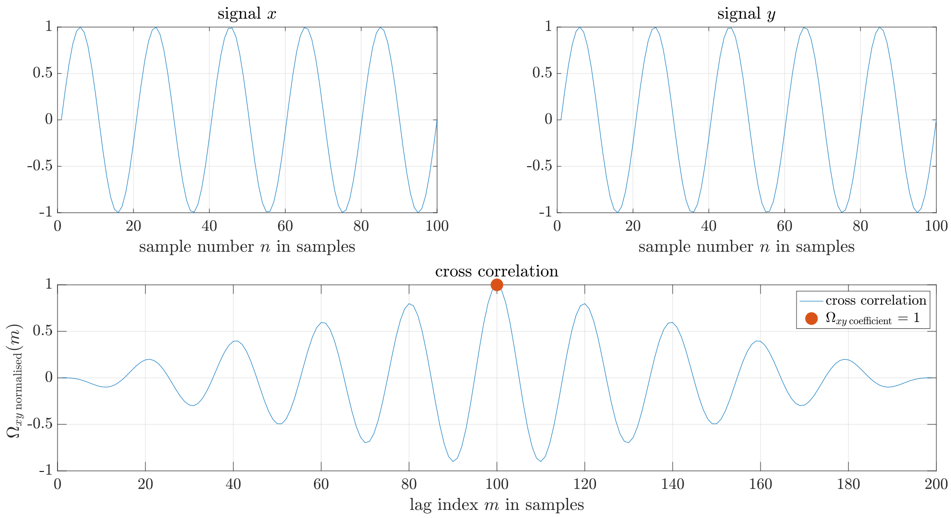

Figure 2.1: Cross correlation of identical signals

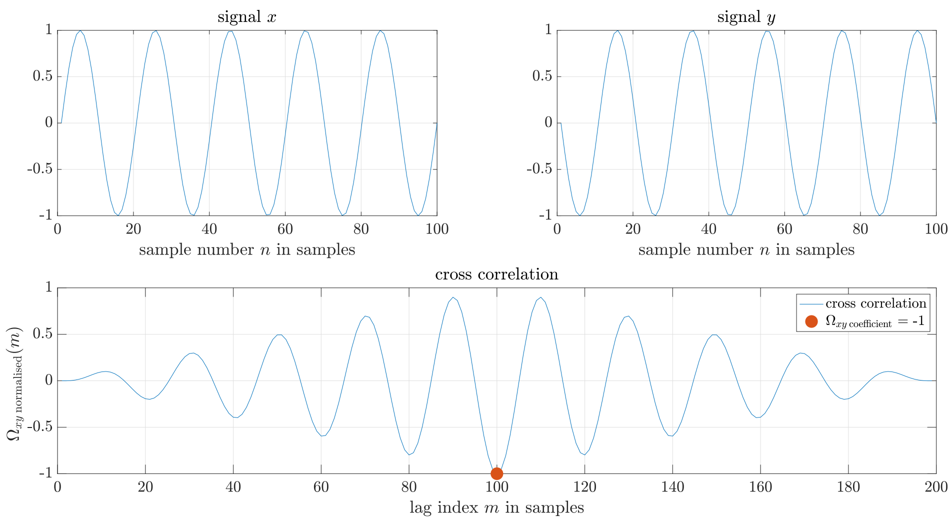

Figure 2.2: Cross correlation of signals 180 ° out of phase

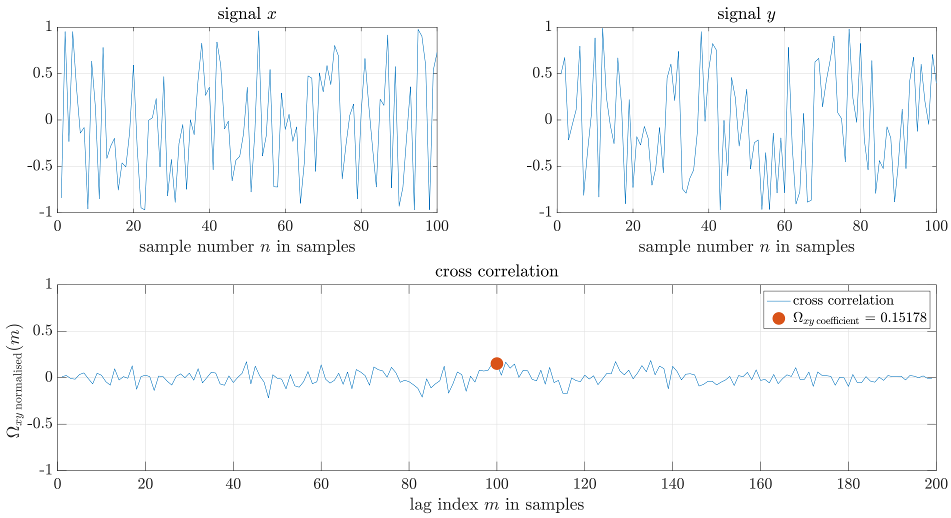

Figure 2.3: Cross correlation of dissimilar signals

A cross correlation coefficient between two output channels of close to zero means a high degree of decorrelation. However it is also necessary to take into account that signals may have different correlation coefficients than expected at other points of the lag index. For example, Figure 2.4 shows a noise signal x and y is a copy of that signal shifted by 20 samples. The cross correlation coefficient is low at almost every value of m as would be expected, but at 20 samples lag there is a much larger correlation. If this signal were played through a stereo loudspeaker and the listener stood in such a way that he was 20 samples closer to the loudpspeaker that played signal x, then they would hear a very correlated signal and a phantom source, which is one of the things that a decorrelation algorithm aims to avoid.

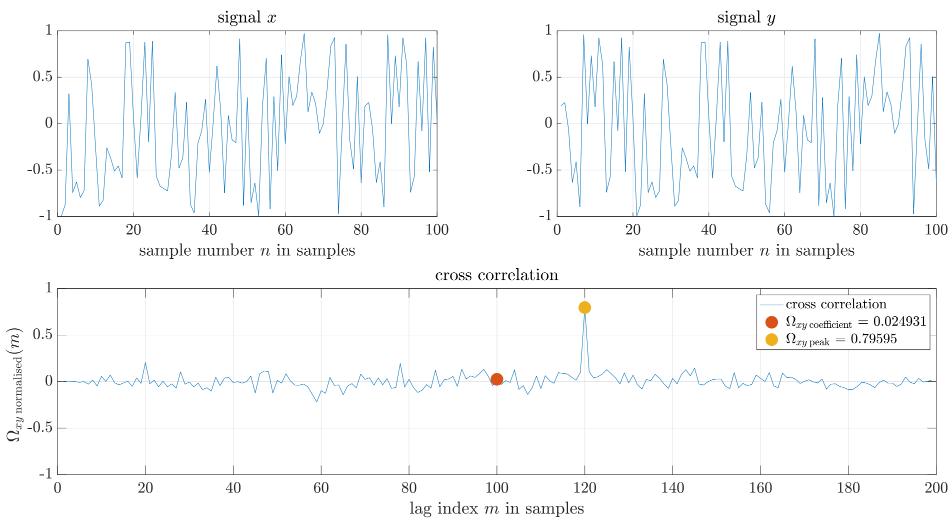

Figure 2.4: Cross correlation of a noise signal and a copy of that signal shifted by 20 samples

2.2 Effects of decorrelation

This section describes the five main effects of decorrelation between audio channels, as explained in *The decorrelation of audio signals and its impact on spatial imagery* by G. S Kendall [1]:

2.2.1 Colouration

When a stereo signal is played through spaced loudspeakers there will be constructive and destructive interference in certain listener positions. When the cross correlation coefficient between the signals played through the loudspeakers approaches zero, this can be reduced and even eliminated.

In a stereo headphone setup, the left and right signals are played directly to the left and right ear, and there is no “cross-talk”. This means that no colouration occurs and so decorrelation can have no effect on the colouration.

2.2.2 Diffuseness

In a concert hall, the sound reaching each ear of a listener will have a low cross correlation coefficient, especially in the later field of a very reverberant hall. The acoustics of the concert hall mean that sound which reaches each ear has taken different paths and there will be small phase changes between the two. Decorrelation is used to replicate this “diffuse” or “spacious” field. The closer the cross correlation coefficient to zero, the more diffuse the sound will be perceived to be. This effect relates to both loudspeaker and stereo headphone setups.

2.2.3 Externalisation

In general, in a stereo headphone setup, sound is perceived as being located inside the head. Decorrelating the audio played through either channel can defeat this and produce perceptual externalisation, and the impression that the audio is coming from around the ears, rather than inside the head. The closer the cross correlation coefficient is to zero, the more pronounced this will be.

In a multiple loudspeaker setup, decorrelation produces the effect that the sound source will be perceived as being located further away than the actual location of the loudspeakers.This can contribute to a “spacious” feeling.

From informal tests carried out using a circular array of eight loudspeakers surrounding the listener, perfectly correlated signals played through all loudspeakers create a strange sensation of the audio being played inside the head and rather than the loudspeakers, which is usually how headphone audio is perceived. Decorrelation of that audio reduces this often unpleasant sensation.

2.2.4 Phantom source

When a correlated signal is played through loudspeakers or headphones, a “Phantom source” will appear between the loudspeakers, which is where the sound appears to be coming from. This can be moved by sufficiently delaying the sound coming from one of the channels by up to 1 ms [2]. The phantom source will be closer to the loudspeaker or headphone channel which plays the signal first. This effect is defeated by decorrelation.

In a stereo headphone setup, a decorrelated sound is localised as a wide source through the head, or from both headphone channels.

In a loudspeaker setup, the decorrelated sound is localised as a wide source between the loudspeakers, or like multiple sources located at the loudspeakers.

The closer the cross correlation coefficient is to zero, the more pronounced this effect is.

2.2.5 Precedence effect

The precedence effect is where a sound is followed by a correlated sound at a different location with a sufficiently short time delay, the listener locates the sound at the location of the first sound. This time delay ranges from approximately 1 ms to 20 ms [2]. Below 1 ms the Phantom source effect is more applicable, and above 20 ms, there is a recognisable echo rather than localisation at the source of the first sound.

This means that a listener standing closer to one speaker in a stereo setup which is playing a correlated signal will perceive the sound to be only coming from that speaker.

Decorrelation will defeat the precedence effect as the sounds are not similar enough to be percieved as one auditory event. The closer the cross correlation coefficient is to zero, the more pronounced this will be. This effect is not observed in headphone setups, as in that case, depending on the delay time, there will be a phantom source instead.

2.3 Transfer function

The H1 transfer function estimator is used to estimate the transfer function between the input signal and one of the output channels, $H(f)$. This is a good way of visualising how each frequency is affected by the decorrelation algorithm and an indicator of what the audible impact will be.

The H1 estimator should be used when any “noise” added (i.e. deviation from the original signal added by the decorrelation function) is not correlated with the input signal [3]. This is the case for all of the decorrelation algorithms used here, so the H1 estimator is used.

A simple way in which to summarise the deviation from the original signal is to compute the mean absolute deviation across all frequencies from a magnitude response of 0 dB re. 1. The length function calculates the length of the vector enclosed in the function - in this case the number of frequency bins in the magnitude response.

$$H_{|\Delta|} = \frac{1}{\text{length}\{H(f)\}} \Sigma |\{20\log_{10}|H(f)|\}| \tag{2.5}$$

The mean transfer function is calculated by:

$$\overline{H} = \frac{1}{\text{length}\{H(f)\}} \Sigma \{20\log_{10}|H(f)|\} \tag{2.6}$$

Which makes the root mean square of the error of the magnitude response from the mean magnitude response:

$$H_{\text{RMSE}} = \sqrt{\frac{1}{\text{length}\{H(f)\}} \Sigma (|H(f)| - \overline{H})^2} \tag{2.7}$$

H_{RMSE} is used as a measure of how "flat" the magnitude response is.

2.4 Group delay

Decorrelation algorithms work in various ways, but always manipulate the phase of the input signal in order to obtain a low cross correlation coefficient between the output channels.

When the method involves FIR or IIR filtering, a group delay is introduced. The whole signal will be delayed by a certain number of samples, and there can also be a variation in frequency. This is often so small as to be perceptually negligible, but it is important to be aware of it and to compensate for it if necessary.

Otherwise, decorrelation algorithms may work by moving all or parts of the input signal forward or backwards in time by a number of samples. Group delay should not be compensated for here as the decorrelation is created by those delays, so accounting for them will only destroy the effect that has been sought.

Group delay is estimated from the transfer function as follows. fs is the sampling frequency and unwrap is a function which unwraps the vector so that any jump between consecutive angles greater than π radians has multiples of $\pm 2\pi$ added to it until the jump is less than π [4].

$$\tau_g\,(f) = \left(\frac{-\text{unwrap}\{\angle H(f)\}}{2\pi f} \right) \text{fs} \tag{2.8}$$

2.5 Balance

For all decorrelation algorithms, the amplitudes of the output signals will vary in different parts of the signal. It is possible that this can cause a Phantom Source to appear which is close to a channel with a higher amplitude signal, even though both channels are highly decorrelated. A simple way of quantifying this is to calculate the Balance Ratio B of a stereo output signal. If x_{RMS} and y_{RMS} are the RMS values of two output channel signals x and y, then B is defined as the ratio between the larger and the smaller RMS values:

$$B = \frac{\max\{x_{\mathrm{RMS}}, y_{\mathrm{RMS}}\}}{\min\{x_{\mathrm{RMS}}, y_{\mathrm{RMS}}\}} \tag{2.9}$$

A Balance Ratio of close to 1 indicates that one channel has not been overly amplified above the other and that there shall be no Phantom Source effect. Through informal listening testing as part of this thesis, a Balance Ratio of over 1.5 indicates that there may be a slight Phantom Source effect, which increases in extremity with the Balance Ratio.

Poor balance can be corrected with later processing, but it is good to be aware if this should be done when using a certain decorrelation algorithm.

2.6 Decorrelation algorithm success criteria

In general, an ideal decorrelation algorithm will:

- Have a cross correlation coefficient between output channels of as close to the desired cross correlation coefficient as possible. This is not always necessarily zero as partial decorrelation effects may be sought.
- Exhibit minimal impact on the frequency content of the input signal, or in other words, the absolute value of the transfer function between each of the output channels and the input signal shall be as close as possible to 1 at all frequencies.
- Create outputs with group delay between output channels and the input signal of as close to zero as possible over all frequencies.
- Create outputs where no one of the channels has a significantly larger output than the other channels.

To a lesser extent, the following are also important:

- Shall produce output signals with a cross correlation of as close to that which is desired over all possible lag indexes. This would be more important if the output was to be used in a loudspeaker setup where the listener may move around, rather than a headphone setup or stationary listener loudspeaker setup.
- Will not be prohibitively computationally expensive.

In addition to those qualities described above, there is another consideration to be made when selecting a decorrelation algorithm for the purposes of the multi-loudspeaker listening test setup. This consists of eight loudspeakers which play audio which has been processed with decorrelation algorithms and a RIR.

- In order to correctly interface with software which enables continuous playback of an audio signal through decorrelation/RIR filters, those decorrelation/RIR filters must take the form of one impulse response per loudspeaker. The convolution of the signal with these filters takes place inside the playback software. This means that all decorrelation must take place using a single impulse reponse per channel. Decorrelation which varies over time is unsuitable as it cannot be implemented using a single impulse response.

3

Method - decorrelation algorithms

3.1 FIR decorrelation

3.1.1 Background

The FIR decorrelation method is presented first as it is perhaps the simplest to understand and implement. The approach described here is based on that in *The decorrelation of audio signals and its impact on spatial imagery* by G. S Kendall [1].

This method produces a decorrelation impulse response of user determined length by taking the Inverse Fast Fourier Transform (IFFT) of a Frequency Response Function (FRF) specified by a magnitude of 1 and randomly generated phase between $-\pi$ and $+\pi$.

The impulse response can then be treated as a Finite Impulse Response (FIR) filter, and so an audio signal can then simply be convolved with the impulse response or convolved with a RIR in later processing. If multiple impulse responses are created and convolved with the same input, the result will be multiple decorrelated audio signals. The length of the sequence of random phase corresponds to the FIR filter order.

3.1.2 Implementation

1. Firstly, a sequence of phase variables is generated with length equal to the desired FIR filter order:

$$\phi(n) = \text{random number between } -\pi \text{ and } +\pi \tag{3.1}$$

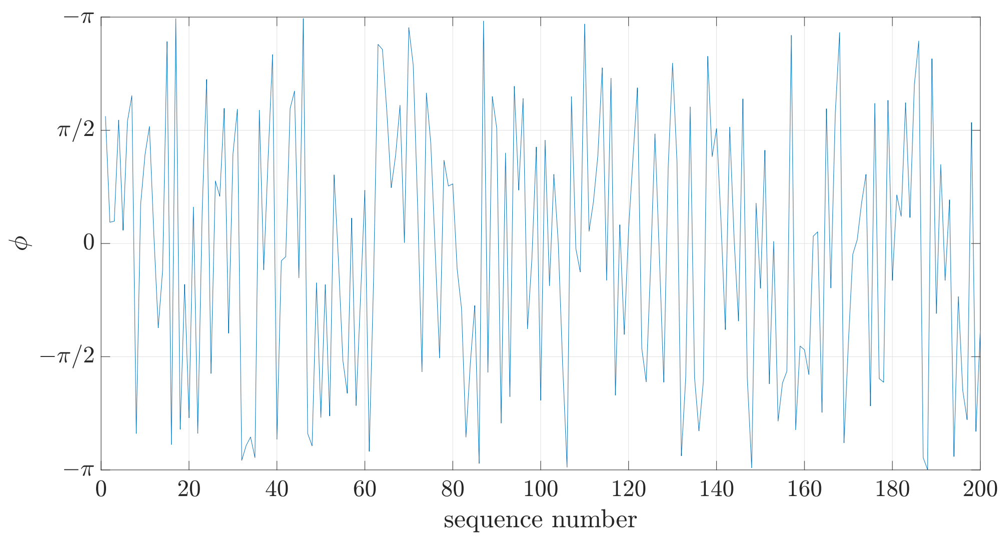

Figure 3.1: Sequence of randomly generated phase

2. A sequence of magnitude variables is generated with length equal to that of the phase sequence:

$$M(n) = 1 \tag{3.2}$$

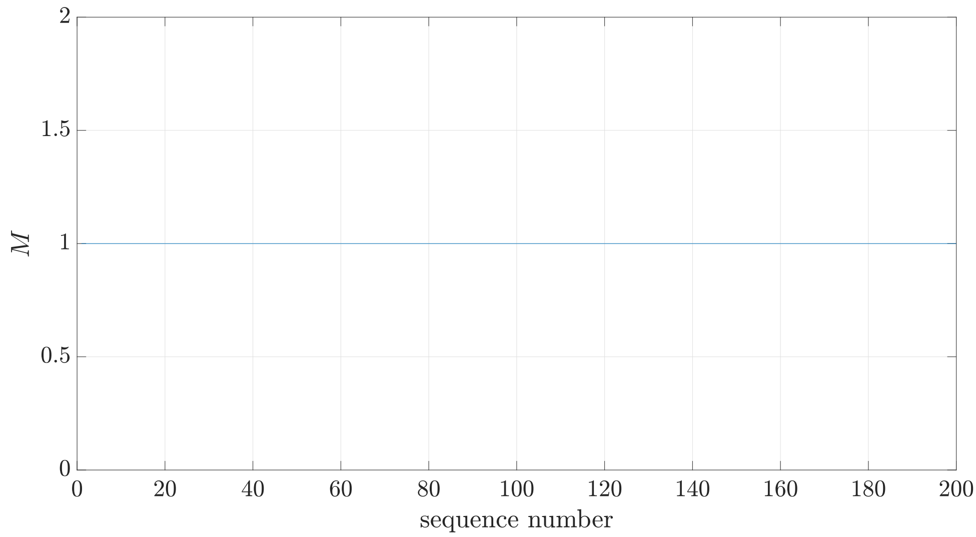

Figure 3.2: Sequence of unit magnitude

3. The magnitude and phase are combined to produce a complex FRF:

$$H(n) = M(n)e^{j\phi(n)} \tag{3.3}$$

Where

$$j = \sqrt{-1} \tag{3.4}$$

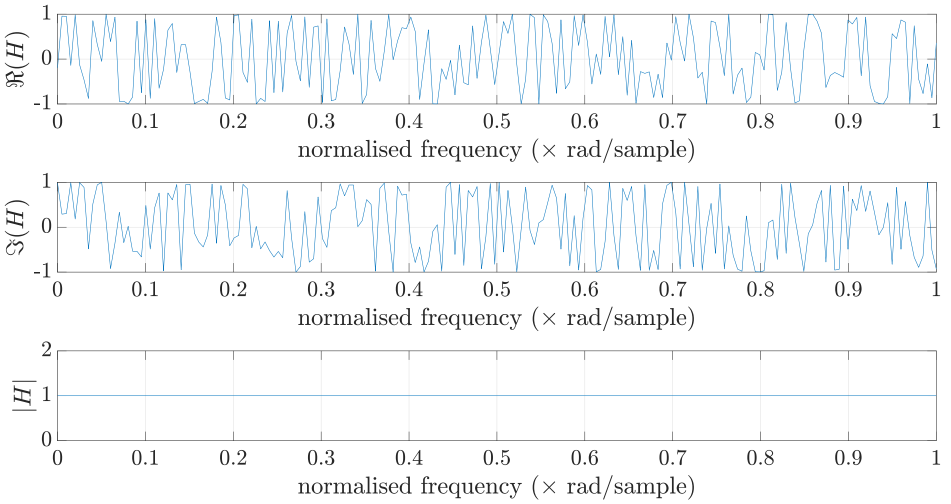

Figure 3.3: Real part, imaginary part and absolute values of the complex FRF

4. The IFFT of the frequency response function is taken to give a decorrelation impulse response:

$$h = \mathfrak{F}^{-1}\{H(n)\} \tag{3.5}$$

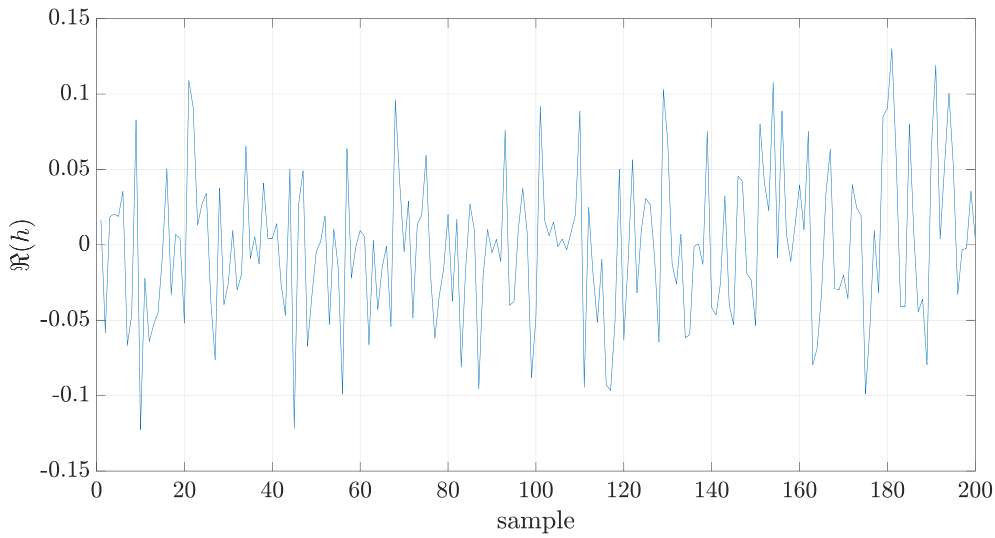

Figure 3.4: Impulse response equivalent to FIR filter coefficients

5. Finally, an input signal x is convolved with the impulse response. New impulse response generated produce outputs (i.e. o) which are decorrelated from each other:

$$o = x * \Re(h) \tag{3.6}$$

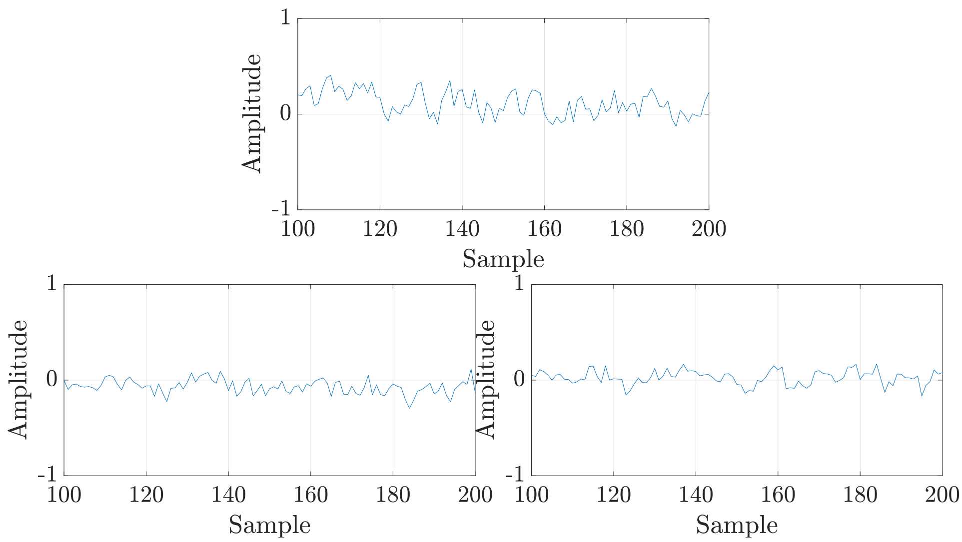

Figure 3.5: Input signal (pink noise, top) and two outputs (bottom) gained from convolution with two different impulse responses

It can be seen in Figure 3.5 that the output signals appear different to the input, but when listened to, they are percieved as similar to the input signal.

3.2 FIR dynamic decorrelation

3.2.1 Background

The FIR dynamic decorrelation method adapted from that presented in *The decorrelation of audio signals and its impact on spatial imagery* by G. S Kendall [1].

This method splits the input signal into many equally sized blocks, then filters each block with a newly generated impulse response generated in the same way as described in Subsection 3.1.2. The blocks are cross faded to avoid sudden obvious jumps in filtering characteristics. This creates an output signal with a feeling of "movement", and decorrelation between each channel - which may or may not be desirable.

3.2.2 Implementation

1. The input signal is split into M blocks of equal length - typically in the range of a few milliseconds up to half a second, depending on the desired effect. An

overlap factor is also selected, which can range from 0 (no overlap) to half of the block length. The overlap is the length over which a block is cross-faded into the next block.

2. M impulse responses per channel are generated in the same way as described in Subsection 3.1.2.
3. Each input block is convolved with the impulse responses for each channel to create M output blocks per channel.
4. Each output block is cross-faded with the next output block by the relevant part of a Hann window (rising or falling) over a length equal to the given overlap. This ensures that there is no sudden and audible "jump" between the output of two different filters. The use of a Hann window shape ensures that the power of the signal is kept as constant as possible and there will be no "ducking" effect of a decrease in volume at any point in the cross-fade. The blocks are then pieced together to form the final output channel, and this is repeated for each channel to produce decorrelated output channels. See Figure 3.6.

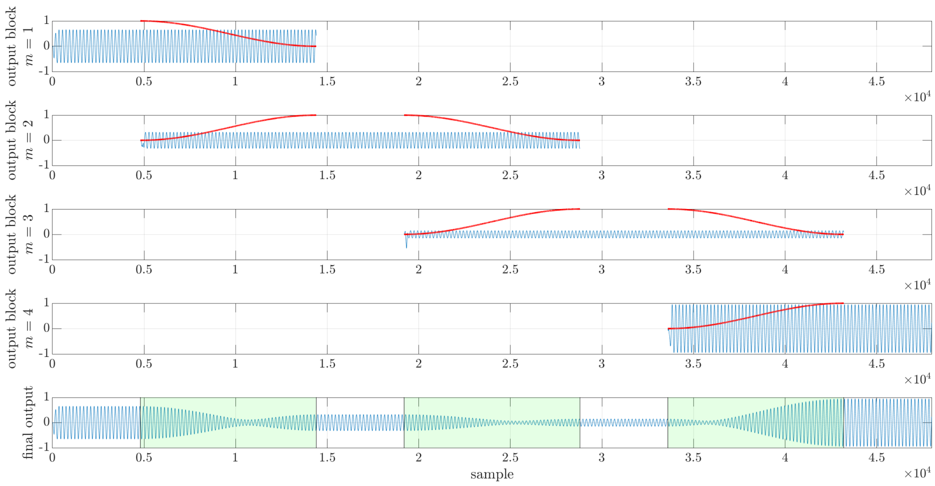

Figure 3.6: Visualisation of how each output block is cross-faded to create a final output channel. A 250 Hz sin wave was used as an input signal. Each of the output blocks which have been passed through a different decorrelation filter are cross-faded by a factor indicated by the red lines. The green areas in the final output are those which have been cross faded, and have a length equal to the selected overlap.

3.3 Sub-band decorrelation

3.3.1 Background

The Sub-band decorrelation method presented here is based on that described in *Audio Signal Decorrelation Based on a Critical Band Approach* by M. Bouéri and C. Kyirakakis [6].

Various band-pass filters with upper and lower frequency limits described by the critical bandwidth of the human auditory system are created. Each of these band-pass filters is then delayed by a random number of samples which is limited to be inversely proportional to the centre frequency of the critical band. The band pass filters are then summed back together to create a single decorrelating impulse response. The process is repeated for the desired number of channels, creating multiple impulse responses which can be convolved with an input signal to produce decorrelated copies of that input signal. It is noted that the band pass filters do introduced their own delay, which is compensated for in later processing.

3.3.2 Implementation

3.3.2.1 Critical bandwidth

Critical bandwidth is best understood as bandwidth of human auditory filters. It is the frequency bandwidth within which a tone will interfere with the perception with another another tone at the centre frequency of the critical band by auditory masking [7]. The Equivalent Rectangular Bandwidth (ERB) is used to estimate this bandwidth for a given centre frequency f, as defined in [8]:

$$ERB = 24.7\left(\frac{4.37f_{centre}}{1000} + 1\right) \tag{3.7}$$

The ERB corresponding to a centre frequency can then be used to calculate the relevant upper and lower frequency limits of that bandwidth:

$$\begin{aligned} f_{lower} &= f_{centre} + \frac{1}{2}ERB \\ f_{upper} &= f_{centre} - \frac{1}{2}ERB \end{aligned} \tag{3.8}$$

A bank of band-pass filters with Equivalent Rectangular Bandwidth are created as follows:

1. A first centre frequency and last centre frequency are chosen, here 100 Hz and 20 kHz.
2. For the first filter, the upper limit is calculated and then the lower limit is set at 0 Hz. Frequencies so low are unlikely to be required, but this is done for completion.
3. For the second filter, the lower limit is taken as the upper limit of the previous filter, and the upper limit is calculated. This repeats until the newest created filter overlaps with the chosen last centre frequency.
4. At this last filter, the upper limit is set to fs/2.
5. All of the frequency limits are used to generate band-pass filters using an FIR design method which ensures the steepest possible cutoff curves. Here, Matlab's *fir1* is used. Steep cutoff curves enable better reconstruction when the bands are summed back together, but a balance has to be struck between this and the delay introduced by a high filter order.

Figure 3.7 shows the filter bank created in this manner. The filters have been created in such a way that all frequencies (from 0 Hz to fs/2 Hz) of an input signal would be preserved in separate channels when filtered with all of the filters in the bank. It is possible then to sum all of those channels back together and receive a signal which only differs from the input by insignificant rounding errors, ensuring that the perceptual impact of this decorrelation algorithm is kept to a minimum. An FIR filter order of 1000 was used to ensure this. The same order is used for each band in order to ensure that the delay introduced by each filter will be approximately the same and therefore easier to compensate for.

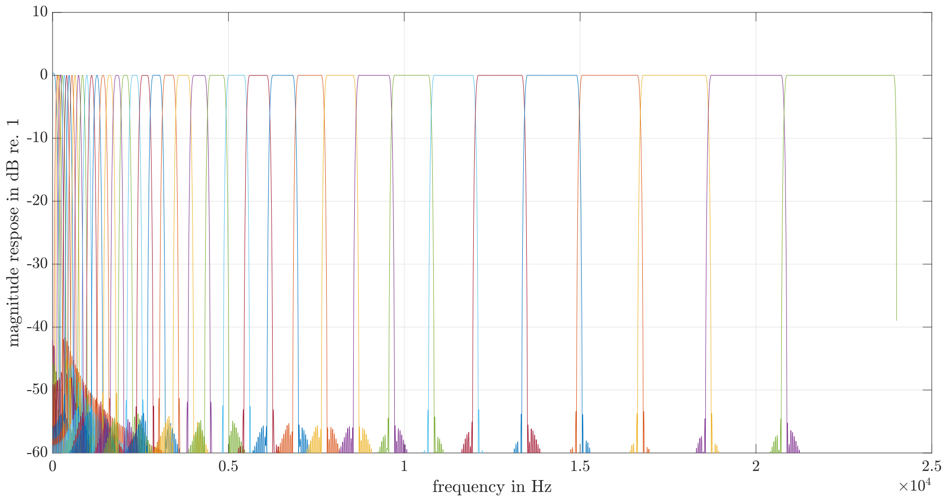

Figure 3.7: ERB (Equivalent Rectangular Bandwidth) filter bank. Filter order: 1000, first centre frequency: 100 Hz, last centre frequency: 20000 Hz

3.3.2.2 Decorrelation

The method by which decorrelation is achieved in this implementation is to delay the samples in each frequency band by different random amounts, which are different for each created channel. If all frequency bands were delayed by the same amount, then a delay or the precedence effect would be percieved by the listener, but processing each frequency differently ensures that this is defeated and decorrelates the audio.

1. An absolute maxiumum delay, τ_{max} is selected, which depends on the effect which is desired. A higher τ_{max} will create strange spatial effects whereas a lower value will create minimal decorrelation. A value of 20 ms is used here, which resulted in a sufficiently low cross correlation coefficient.

2. Limits are selected for each frequency band on what the largest delay will be. These should decrease from τ_{max} for the first frequency band as large shifts at high frequencies can cause easily perceptable audio artefacts. The amount of decrease depends on what is desired as the result. In [6], an exponential decrease to almost 0 ms is used, but here, good results were found using a linear decrease down to 0 ms for the highest frequency band. Experimentation with large maximum delays and limiting each band differently can be rewarded with

unusual effects.

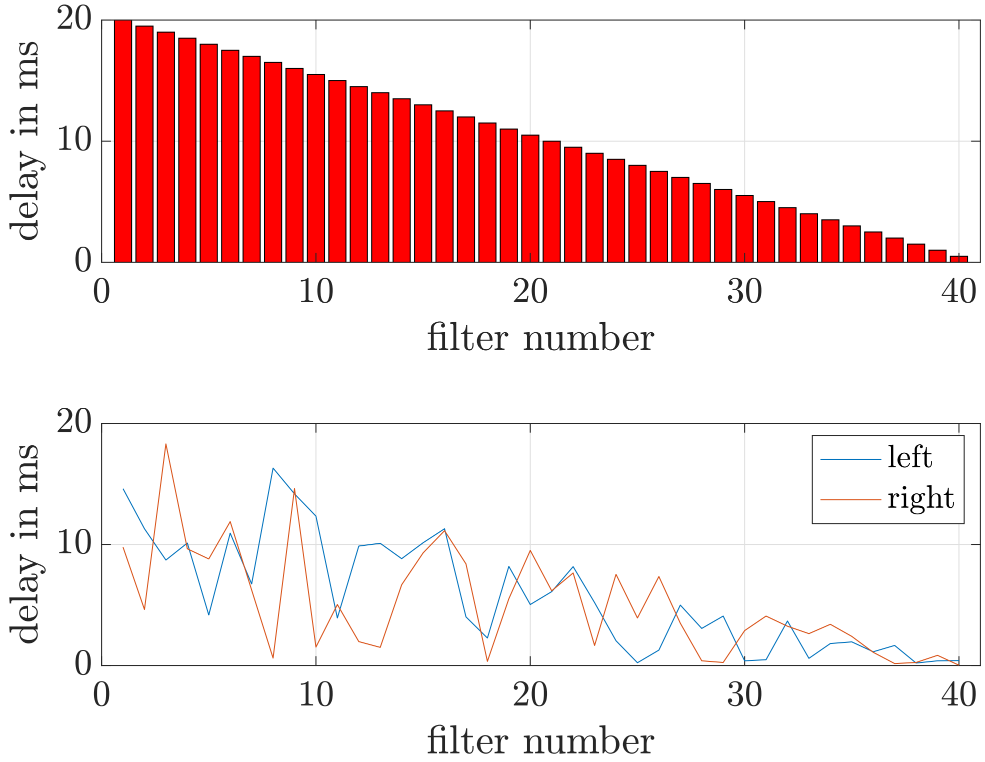

Figure 3.8: Limits on allowable delay (top), and an example of randomly generated delays for each channel (bottom)

3. The delays randomly generated for each frequency band with the limits described above, rounded to the nearest sample then turned into impulse responses - for example a delay of 5 samples would be implemented by an impulse response of 5 zeros and then a 1 in the 6th sample position.
4. The delay impulse responses are then convolved with their relevant band-pass impulse response from the band-pass filter bank. The newly created impulse responses are then simply summed to form a final decorrelation impulse response. The process is repeated for the amount of channels or impulse responses that is desired. When convolved with an input signal, various decorrelated output signals will be the result.

3.4 IIR allpass cascade decorrelation

3.4.1 Background

This approach is based on that described in *Signal decorrelation using perceptually informed allpass filters* by E. Kermit-Canfield and J. Abel [9].

a and b filter coefficients for an IIR biquad allpass filter are generated. A Matlab function is used to turn these IIR filter coefficients into an impulse response. Several impulse responses are created in this way and convolved with each other, creating a long decorrelating impulse response consisting of approximations of many iterations of cascading IIR allpass filters. Multiple decorrelated output channels can be created by convolving an input signal with a number of decorrelating impulse responses created in this way.

3.4.2 Implementation

1. The following is a biquad allpass filter with poles inside the unit circle and zeros with the same magnitude and frequencies, in other words the zeros are reflected across the unit circle. The z-transform is given by:

$$H(z) = \frac{z^{-2} - 2\mathbb{R}(\kappa)z^{-1} + |\kappa|^2}{1 - 2\mathbb{R}(\kappa)z^{-1} + |\kappa|^2 z^{-2}} \tag{3.9}$$

   Where:

$$\kappa = \psi e^{2\pi\omega j} \tag{3.10}$$

2. The following coefficients for the biquad allpass filter are then calculated as follows.

$$\begin{aligned} a_0 = 1,\ a_1 = -2\mathbb{R}(\kappa),\ a_2 = |\kappa|^2 \\ b_0 = |\kappa|^2,\ b_1 = -2\mathbb{R}(\kappa),\ b_2 = 1 \end{aligned} \tag{3.11}$$

3. There are some restrictions set out in [9] on ψ and ω, but there seems to be no disadvantage in defining them as follows, as per informal testing undertaken as part of this thesis:

$$\begin{aligned} \psi &= \text{random number between 0 and 1} \\ \omega &= \text{random number between } -\pi \text{ and } +\pi \end{aligned} \tag{3.12}$$

4. These filter coefficients are then turned into an impulse response using the Matlab function *impz*. Because the coefficients correspond to an *Infinite* Impulse Response, a true representation of them by an impulse response would be infinitely long, so a cutoff point is chosen in order to approximate those coefficients. In this implementation, the impulse response is limited to $s_{\text{limit}} = 300$ samples. In testing over several impulse responses generated in this way, the impulse response had decayed by at least 60 dB by 300 samples, so it was judged that any samples after would not have a significant perceptual impact on the sound.

5. This impulse response is saved, and then a new one is created with new values of ψ and ω. This is convolved with the previous impulse response, creating a new one which is saved ready to be convolved with the next created impulse response, and so on. This process is repeated for a desired number of iterations, with new values of ψ and ω.

6. The resulting final decorrelating impulse response is then limited to the following number of samples, where i is the chosen number of iterations, in order to reduce convolution time.

$$s_{\text{cutoff}} = s_{\text{limit}} \frac{i}{100} \tag{3.13}$$

   This can be done with minimal impact to the effectiveness of the impulse response as by the time of the cutoff the values are very close to 0.

7. Because each filter used has an allpass characteristic, thousands of iterations can be used with minimal impact on an input signal's frequency content, only it's phase characteristics which means many decorrelated copies of an input signal can be created with impulse responses with newly generated ψ and ω for each iteration. The higher the number of iterations, the closer the cross correlation coefficient will be to zero, but too high and the resulting impulse response will be too long and the characteristics of an input signal will not be preserved. In this thesis, 1500 iterations were used and found to be a reasonable compromise.

3.5 Velvet noise decorrelation

3.5.1 Background

The velvet noise decorrelation method is adapted from *Velvet-noise decorrelator* by B. Alary, A. Politis and V. Välimäki [10].

Impulse responses of randomly and logarithmically spaced, exponentially decaying velvet noise are generated. Each impulse in the impulse response also has a random sign. The properties of the impulse response accord to a user specified density of impulses and decay of impulses. A number of impulse responses corresponding to the desired number of output channels are generated.

The impulse responses are then convolved with an audio input signal and the result is multiple output signals which are decorrelated.

Velvet noise is defined as sparse random noise with values of of -1 or 1, with the remaining values as 0. It has been used in reverberation modelling and has been shown to be smoother sounding than white noise for this purpose [11]. Velvet noise is also prefered over white noise as there are convolution operations where the coefficient is zero can be skipped, making the operation computationally efficient. In the listening test implementation, this advantage is not exploited as the convolution takes place in an external piece of software.

3.5.2 Implementation

1. M is defined as the total number of non-zero impulses which shall be present in the impulse response. It is calculated as follows, where ζ is the chosen impulse density in impulses per second, and L is the desired length of the impulse response (or in other words, the filter order). fs is the sampling frequency.

$$M = \frac{L\zeta}{\mathrm{fs}} \tag{3.14}$$

2. ϵ is the decay constant, where L_{dB} is the desired decay over the whole impulse response:

$$\epsilon = \frac{-\ln 10^{\frac{-L_{\mathrm{dB}}}{20}}}{M} \tag{3.15}$$

3. The amplitude and sign of each non-zero impulse, s are given as follows:

$$s(m) = (2\,\mathrm{round}\{r_1\} - 1)\,\mathrm{e}^{-\epsilon m} \tag{3.16}$$

   Where m and r_1 are defined:

$$\begin{aligned} m &= 1, 2, 3, \ldots, M \\ r_1 &= \text{random number between 0 and 1} \end{aligned} \tag{3.17}$$

4. Now it is necessary to create a vector of indexes, k, which determine the position of each impulse s_e in the full impulse response. The impulses should be logarithmically spaced - i.e. there is a higher density of them at the beginning of the impulse response than the end. k_{test} is the vector from which these indexes shall be picked:

$$\begin{aligned} k_{test} = &\text{ vector of logarithmically spaced integers} \\ &\text{between 1 and } L, \text{ with length } M^3 \end{aligned} \tag{3.18}$$

   $k(m)$ values are then chosen:

$$k(m) = \text{randomly chosen non} - \text{duplicate values from } k_{test} \tag{3.19}$$

   And then the vector $k(m)$ is sorted in ascending order (so that the impulses will appear in an exponentially decaying curve).

5. The impulse response h is produced from the impulse values and locations. All remaining values of h which do not have indexes specified in k are left as zero.

$$h(k(m)) = s(m) \tag{3.20}$$

6. The result is a logarithmically spaced and exponentially decaying velvet noise impulse response (see Figure 3.9). Multiple impulse responses are then generated for each output channel required, then convolved with the input signal. The impulse response decays exponentially to prevent smearing of transients

when using long impulse response lengths, and the logarithmic spacing also contributes to the same effect by concentrating the majority of the impulses in the early part of the impulse response.

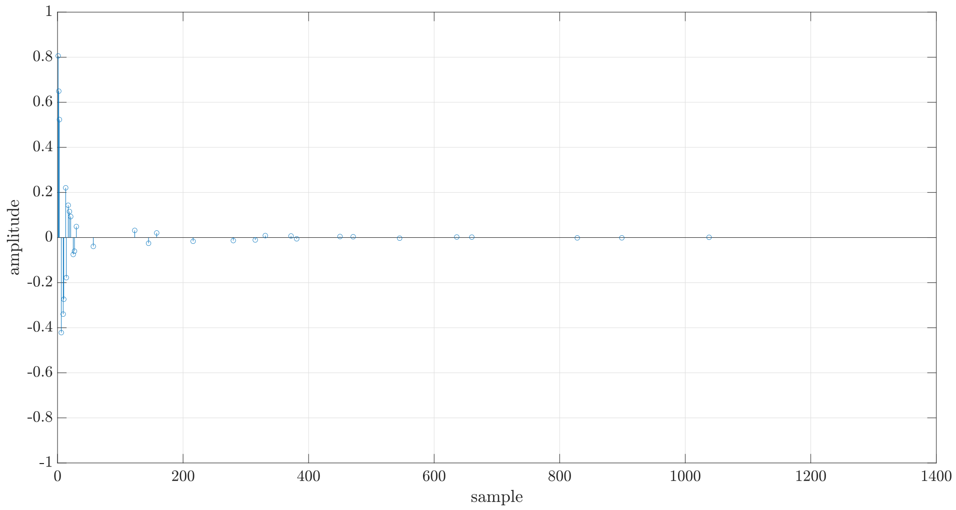

Figure 3.9: Logarithmically spaced, exponentially decaying, velvet noise impulse response. Impulse decay: 60 dB, impulse density: 1000

3.6 White noise decorrelation

3.6.1 Background

The white noise decorrelation method is adapted from *Velvet-noise decorrelator* by B. Alary, A. Politis and V. Välimäki [10].

A number of impulse responses equal to the desired number of output channels are created. These consist of exponentially decaying white noise.

The impulse responses are then convolved with an audio input signal and the result is multiple output signals which are decorrelated.

3.6.2 Implementation

1. ϵ is defined as the decay constant, where L_{dB} is the desired decay over the whole impulse response and M is the desired filter length:

$$\epsilon = \frac{-\ln 10^{\frac{-L_{\mathrm{dB}}}{20}}}{M} \tag{3.21}$$

2. An envelope e is created:

$$e = \mathrm{e}^{-\epsilon v} \tag{3.22}$$

   v is a vector of length M where:

$$v = 1, 2, 3, \ldots, M \tag{3.23}$$

3. h is then the impulse response of length M, where n is a vector of white noise of length M, with values randomly selected between -1 and 1.

$$h = en \tag{3.24}$$

4. Multiple h are created, one for each desired output channel, with new randomly generated white noise vectors. Each h is then convolved with the input signal to create multiple decorrelated output channels.

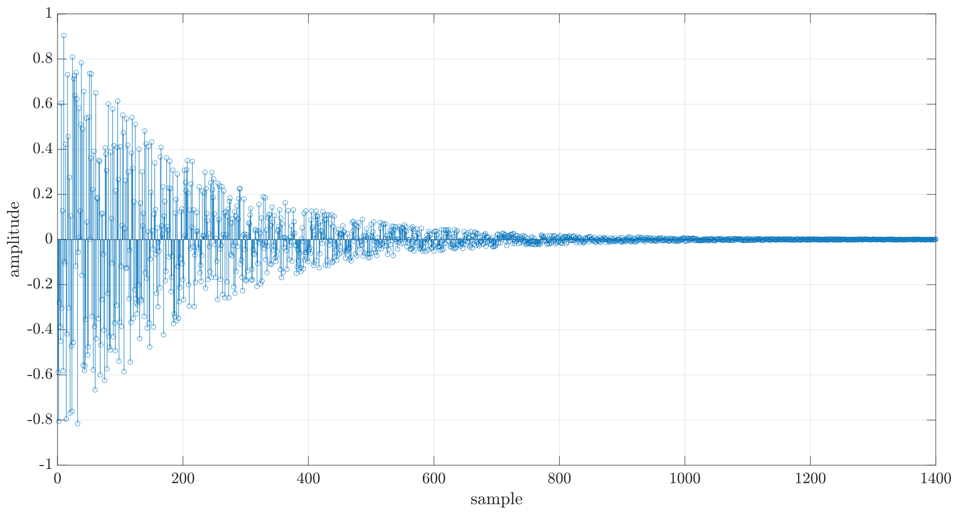

Figure 3.10: Exponentially decaying white noise impulse response. Impulse decay: 60 dB

4

Method - further processing

4.1 Correlation optimisation

Correlation optimisation is included in all decorrelation methods which can be implemented using an impulse response - FIR decorrelation, Sub-band decorrelation, IIR allpass cascade decorrelation, Velvet noise decorrelation and White noise decorrelation. The goal of the optimisation is to ensure that when the decorrelation impulse responses are finally convolved with the input signal to create multiple decorrelated output channels, then the cross correlation coefficient between each of those output channels is as close as possible to the desired value. Optimisation is achieved by creating many decorrelation impulse responses (several hundred is a good place to start, from informal experimentation as part of the work of this thesis) and then testing them against this criteria. The implementation for the different setups of stereo or multi-loudspeaker setup are as follows.

4.1.1 Stereo loudspeakers or stereo headphones

1. Create t number of decorrelation impulse responses. t should be greater than the number of channels to be created.
2. Convolve the first impulse response with the input signal to create a test first output channel. This first impulse response is the first channel's winning decorrelation impulse response.
3. For the second output channel, convolve all of the impulse responses with the input to create t test outputs. Calculate the cross correlation coefficient between each of the test outputs and the first test output. The test output which produces the closest cross correlation coefficient to that which is desired is the winner, and the impulse response which produced it is the winning decorrelation impulse response for the second channel.

4.1.2 Multiple loudspeakers

1. Create t number of decorrelation impulse responses. t should be greater than the number of channels to be created.
2. The first impulse response is simply 1 in the first sample and then zeros for the remainder of the impulse response length. This means that the first output

channel will be identical to the input. The corresponding loudspeaker will be the “source” loudspeaker, and in the final setup the audio source will appear to come from here, whereas the reverberant field of the created virtual room will appear to come from the remaining loudspeakers.

3. For the remaining channels, the cross correlation coefficient is calculated between each test output and the outputs of the previous channels winning decorrelation impulse responses. The winner is selected, which has the impulse response which produces an output which has a cumulative cross correlation coefficient closest to the desired one - i.e. the sum of the absolute difference between the new channel’s cross correlation coefficients and the desired cross correlation coefficient is minimised.

Implementing this procedure has the implication that the computational cost of running increases exponentially with the number of required output channels - as an exponentially increasing amount of cross correlation coefficients must be calculated.

This means that without using an impractical number of t, a significant number of channels will always mean that there is a fairly wide variation of cross correlation coefficients between each channel.

A good measure in order to quantify the correlation of all of the channels is the absolute mean cross correlation coefficient, which gives a single figure with which to compare the desired cross correlation coefficient to. This is simply the mean of the cross correlation coefficients between each channel.

In further discussion, results and analysis in this thesis, it should be noted that the correlation between output channels is referred to by its *target* correlation - for example correlation 0, correlation 0.25 or correlation 0.75. These numbers are not necessarily the exact correlation between each channel of the output, as this can vary considerably, but is a simple metric to compare differently correlated sets of signals.

4.2 Room Impulse Response processing

Artificial reverberation is added to a signal by convolving it with a Room Impulse Response (RIR). The RIR is split into two parts - the direct part and the diffuse part, at the diffuse point, which is the sample in the RIR where these overlap. The diffuse point is picked manually to be just after the first parts of the impulse response which obviously correspond to the first reflections of the room. The direct part of the RIR is set to be from the first sample of the original RIR to the diffuse point. For the diffuse part of the RIR, all samples before the diffuse point are set to zero and then the rest are kept.

Figure 4.1 shows an example - the diffuse point is picked to be just after the first two reflections, then it is clear that the remaining reverberation trail corresponds to the diffuse part of the RIR.

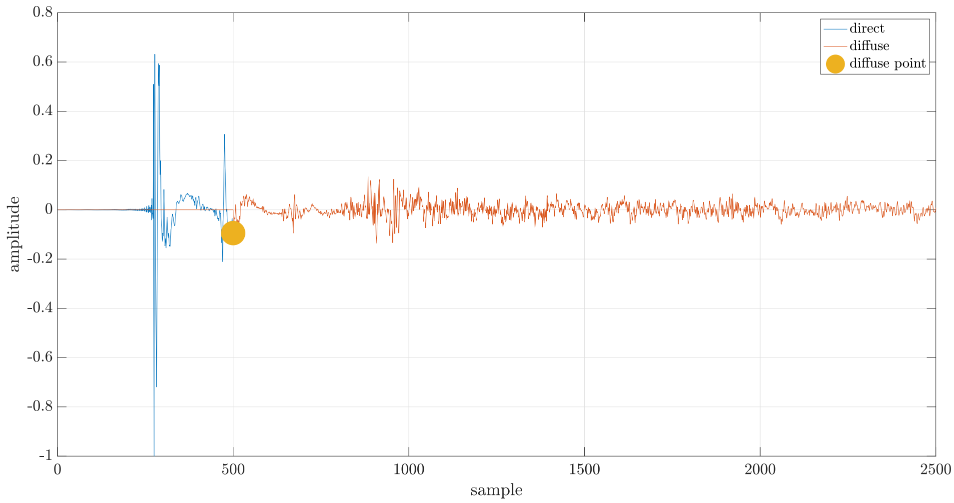

Figure 4.1: The first 2500 samples of a RIR which has been split into direct and diffuse parts at the diffuse point. Omnidirectional impulse reponse of Lady Chapel, St. Albans Cathedral [5]

The diffuse part of the RIR is also clipped when the main part of the decay has finished - in order not to introduce unwanted noise into the output signal, and to reduce the the total convolution time taken.

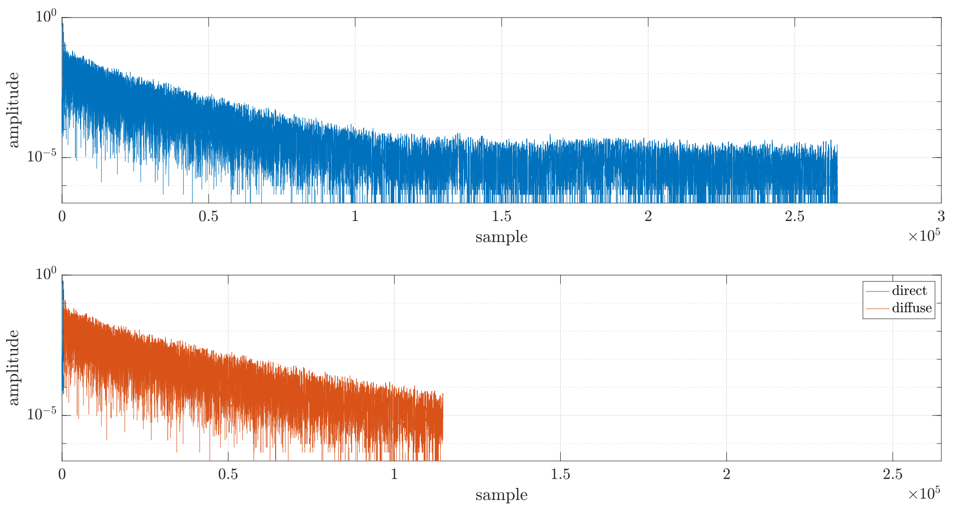

Figure 4.2: Wider view of a RIR which has been split into direct and diffuse parts at the diffuse point (top). The diffuse part of the RIR has been cut after the main part of it's decay (bottom). Omnidirectional impulse reponse of Lady Chapel, St. Albans Cathedral [5]

4.2.1 Stereo loudspeakers or stereo headphones

In the case that a regular stereo output signal is required, the input signal is convolved with the direct part of the RIR. The input is also then put through any of the created decorrelation methods first (with optimisation as per Sub-section 4.1.1), then convolved with the diffuse part of the RIR. The results of these two operations are then simply added together. The consequence is that only the diffuse part of the RIR has an output which is decorrelated, where the direct sound is not, which is more consistent with what a real room sounds like - a strong direct sound and first reflections which will be correlated and then a more wide sounding diffuse field.

4.2.2 Multiple loudspeakers

In the case of a multiple loudspeaker setup, a number of decorrelation filter impulse responses equal to the number of loudspeakers are required (as per Sub-section 4.1.2).

1. The first impulse response, which corresponds to the "source" loudspeaker, is convolved with the direct part of the RIR. It is increased by a factor, then added to the diffuse part of the RIR. The 8 "diffuse" sounds which will play from the loudspeakers sum incoherently if they are completely decorrelated which means an increase in power of 3 dB for each doubling of the number of sources. To go from 1 source to 8 source means doubling 3 times, a factor of 9 dB, so 9 dB is the amount that is added to the direct part of the RIR before it is added to the diffuse part of the RIR, to ensure that the original qualities of the whole RIR are kept consistent. This is imperfect when the "diffuse" sounds are not completely decorrelated - see Chapter 8 for discussion.
2. The remaining impulse responses corresponding to the non-source loudspeakers are simply convolved with the diffuse part of the RIR.
3. Any group delay which is introduced by the decorrelation algorithm that is being used must be compensated - for example if the group delay introduced is 250 samples, then the diffuse part of the RIR for the diffuse channels must be advanced by 250 samples. This ensures that the timing of the RIR is preserved whilst being able to retain the whole part of the decorrelation impulse response.
4. The diffuse loudspeaker RIR/decorrelation impulse responses are then normalised in such a way that their root-mean-square amplitude is the same as the root-mean-square of the diffuse part of the first (source) loudspeaker's impulse response.

The resulting decorrelation/RIR combined impulse responses are then ready to be convolved with the input signal in order to be played through loudspeakers, creating a sound source in a simulated room, or loaded into some software to enable playback.

5

Method - listening test

5.1 Background

A listening test was performed in order to ascertain the relationship between correlation between sound sources, different decorrelation algorithms and two metrics: “spaciousness” and “timbre”.

Listeners are played two sets of stimuli (A and B) from eight loudspeakers - one of these loudspeakers is the “source” loudspeaker, and the rest play sound which simulates the reverberant field of a room. A and B either consist of audio which has been decorrelated by different amounts but use the same decorrelation algorithm, or decorrelated by the same amount but use different algorithms. Two different rooms were simulated - a church and a drum room, but each test only contained one room. Some subjects undertook both listening tests (both rooms), but the order in which they took them was randomised. The remaining participants only took one test (one room).

The listener is asked to rate on a sliding scale which stimuli they perceived to be more “spacious” and whether they perceived any difference in “timbre” between the two stimuli (A and B). Written definitions for both of these metrics are given to the listener.

5.2 Choices

A number of decisions were made regarding the form that the listening test should take, which are justified here.

5.2.1 Decorrelation algorithms

Due to constraints on time, it was decided that the listening test only need take around 15 minutes. It was then decided to only include three of the six decorrelation algorithms that have been investigated in this thesis.

These were:

1. IIR allpass cascade decorrelation
2. Sub-band decorrelation

3. White noise decorrelation

These algorithms were chosen because they are the three which have the least and flattest impact on the frequency content of an input signal (see Section 6). IIR allpass cascade decorrelation and Sub-band decorrelation do introduce a significant amount of group delay, but this was compensated for in the Room Impulse Response convolution processing.

As a powerful computer was available to undertake convolution playback, the length of the impulse response required to produce decorrelation was not taken into account.

5.2.2 Cross correlation coefficients

The target cross correlation coefficients chosen for listening tests were 0.0, 0.25 and 0.75. The *actual* absolute mean cross correlation coefficients between loudspeakers 2, 3, 4, 5, 6, 7 and 8 are summarised in Tables 5.1 and 5.2. These were calculated by calculating the cross correlation coefficient between each of the outputs (in turn, calculated by convolving each decorrelation/RIR impulse response with the input signal), then taking the absolute value of those cross correlation coefficients, then taking the mean. Loudspeaker 1 was excluded as it mostly contained the direct sound and so is always decorrelated from loudspeakers 2-8, so is not relevant to the calculation.

It can be seen that particularly for the target cross correlation coefficient of 0, there are significant differences between the target and actual values. This, however was the best that could be achieved whilst keeping the calculation time reasonable, as it is difficult to ensure that the cross correlation between so many channels is low. It was judged that each of the different cross correlation coefficients did represent audibly different levels of correlation in informal listening. The cross correlation coefficients between each channel can be viewed in Appendix A.2. In this appendix it can be seen that there are often some large outlying numbers which skew the absolute mean cross correlation coefficient to be somewhat different to the target cross correlation coefficient, where otherwise the values are close to the target cross correlation coefficient.

	Target cross correlation coefficient		
Algorithm	**0**	**0.25**	**0.75**
IIR allpass cascade	0.20	0.26	0.70
Sub-band	0.17	0.30	0.76
White noise	0.11	0.18	0.77

Table 5.1: Target, and actual absolute mean cross correlation coefficients for the Drum Room stimuli

	Target cross correlation coefficient		
Algorithm	**0**	**0.25**	**0.75**
IIR allpass cascade	0.24	0.27	0.69
Sub-band	0.14	0.24	0.70
White noise	0.12	0.29	0.78

Table 5.2: Target, and actual absolute mean cross correlation coefficients for the Church stimuli

5.2.3 Metrics

One of the main effects of decorrelation is an impartation of diffuseness, or spaciousness onto sound, and that is one of the main reasons that decorrelation would be performed in relation to audio, so it was decided that this must be included as a question in listening test. Subjects are asked to rate on a sliding scale which of two stimuli are more spacious - from A significantly more spacious than B, equally spacious, B significantly more spacious than A to everything in between.

Change in timbre is one of the largest disadvantages to using decorrelation. An algorithm which produces effective decorrelation effects but completely changes how the input signal sounds is not desirable in most situations. Subjects are asked to rate on a sliding scale how different the timbre of two stimuli are, from no different to a large difference.

During literature study, it was not found that any formal experiments had been performed testing different correlations or different decorrelation methods for spaciousness perception with a multiple loudspeaker setup, which also encourages this selection of metrics.

5.2.4 Stimuli

A 10 second, 44.1 kHz sampling frequency, mono drum loop was used as a stimulus in the listening testing. This had little or no reverberation on it to begin with and looped for continuous playback. This was chosen as it has a wide span of frequency contents and transient qualities, which would mean that any audible phase artefacts produced by decorrelation would be more apparent. A more continuous sound may hide such artefacts and so the drum audio is a more “difficult” test for the decorrelation algorithm

The first Room Impulse Response was chosen as that from Lady Chapel, St. Albans Cathedral [5]. This was chosen as it has a long reverberation time (2 - 3 seconds) and so would showcase the decorrelation taking place in an extreme and obvious way. There is also a clear division between the direct and diffuse parts of the Room Impulse Response.

The second Room Impulse Response which was chosen was a “Nice Drum Room” [12]. It has a shorter reverberation time of under 1 second and would provide a less extreme type of reverberation whilst still having some audible reverberation trails.

5.2.5 Hardware

Informal initial trials were undertaken using a headphone setup with eight virtual loudspeakers (created by convolution with Head Related Impulse Responses), but these were found to create an uninspiring sound where it was difficult to tell the difference between different decorrelation methods and degrees of decorrelation. A hardware setup of eight real loudspeakers fared better informally. Eight loudspeakers was judged to be a sufficient quantity to avoid any localisation cues meaning that a listener would perceive sound coming from a specific loudspeaker.

5.3 Implementation

5.3.1 Subjects

No data on the subjects which participated in the listening test was recorded excepting their responses to the listening test. They were all students of age between 20 and 35 with at least some previous experience of listening tests.

For the Church RIR, eight listening tests were performed. For the Drum Room RIR, seven listening tests were performed, giving a total of 15 listening tests.

5.3.2 Equipment

- 1x Apple iMac computer
- 8x Neumann KH80 DSP active loudspeakers
- 8x Loudspeaker stands (tables, books etc.)
- 1x Antelope Orion 32 AD/DA converter
- 1x D-SUB 25 to XLR cable
- XLR cables
- 1x Chair with large arm rest for using computer mouse
- Sweets given to subjects as a thank you for participating

5.3.3 Graphical User Interface and software

A Graphical User Interface (GUI) was designed and run in Matlab. It's purpose is to allow the user to seamlessly switch between the A and B stimuli at their leisure and choose which is the more spacious and how significant the difference in timbre is. It also prompts the subject to make decisions about the timbre and spaciousness of the stimuli by not allowing them to continue to the next stimulus before listening to A and B and moving both sliders.

Various other software is also used to ensure continuous playback of the audio and switch stimuli:

- Jack [13]
- Pure Data [14]
- Sound Scape Renderer [15]

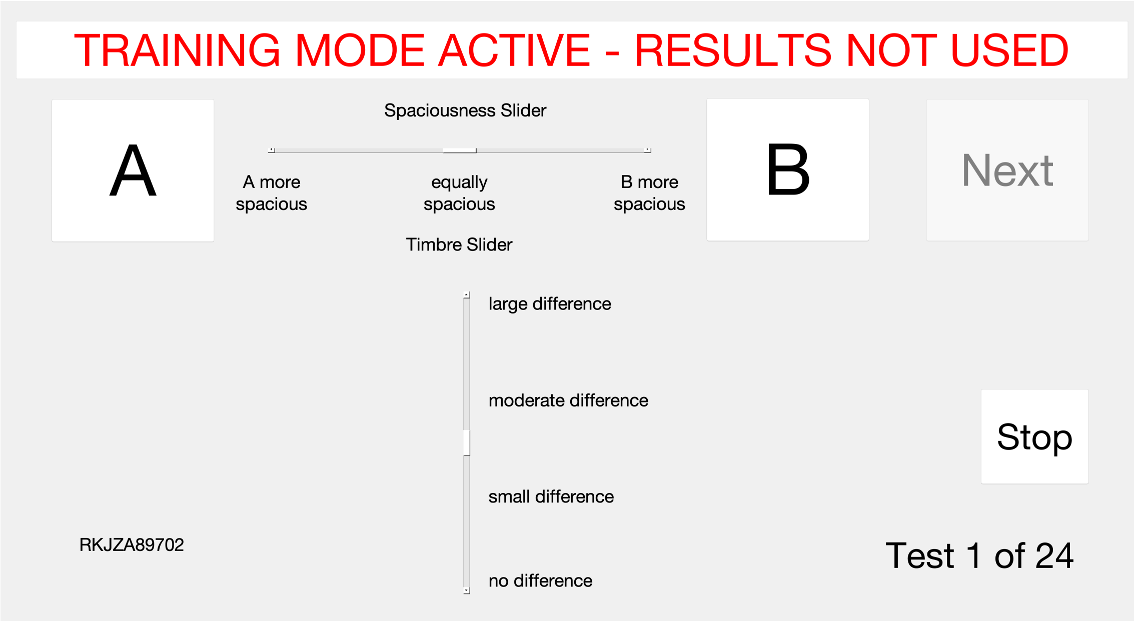

Figure 5.1: Screenshot of the GUI

5.3.4 Setup

The listening test was conducted in a (approximately 4 m by 8 m rectangular) room with many soft absorbing surfaces to minimise unwanted reverberation which could confuse the decorrelated reverberation deliberately imparted on the stimuli.

The loudspeakers were adjusted to the following settings and placed in a circle 1.5m from the listener position, at approximately ear height:

- **Auto standby:** on
- **Acoustical control:** free standing
- **Output level:** 94 dB SPL
- **Input gain:** -15 dB

The computer running the GUI was placed on the floor in order to create minimal interference with the sound played by the loudspeakers. The computer mouse was placed on the arm rest of the chair that the listener sat in so that the GUI could be used whilst seated listening to the sound.

Figure 5.2: Photograph of the listening test setup

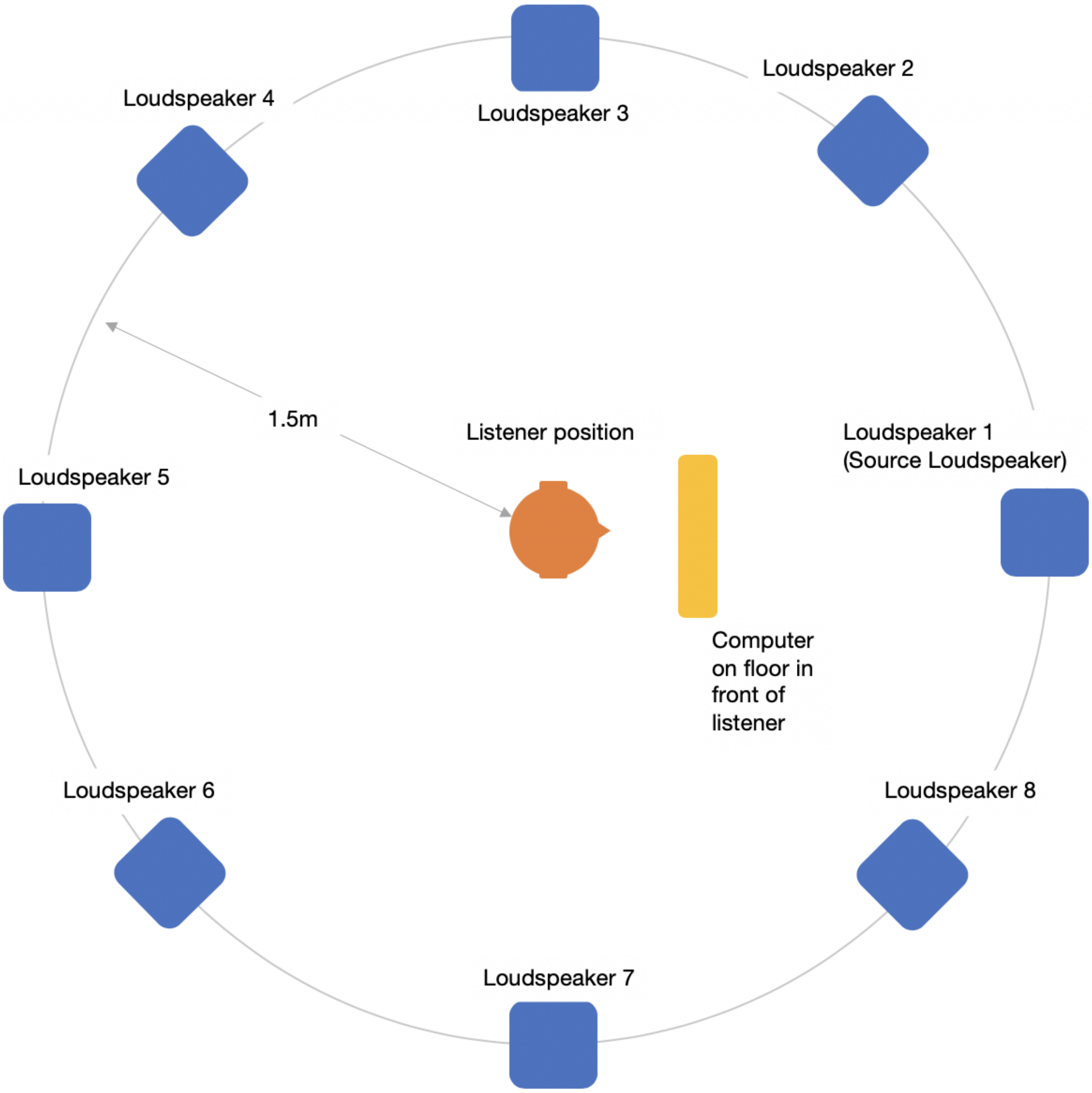

Figure 5.3: Diagram of the listening test setup

SPACIOUSNESS

Evaluate how spacious the audio played by button A and button B is.

- Moving the slider all the way to the left means that you think A is significantly more spacious than B.
- Moving the slider all the way to the right means that you think B is significantly more spacious than A.
- Moving the slider to the centre means that you think both are equally spacious.

The virtual room in which you are placed in by stimuli A or B can be considered more spacious than the other if:

- You perceive the sound as coming from further away.
- You perceive the virtual room as sounding as if it is larger.
- The sound you hear is more diffuse.
- The sound you hear is more enveloping.

TIMBRE/SOUND COLOUR

Evaluate how timbrally different the audio played by button A and button B is.

- Moving the slider all the way to the bottom means that you think that there is no timbral difference between A and B.
- Moving the slider all the way to the top means that you think that there is a large timbral difference between A and B.

The greater the extent that you perceive the following should correspond to a higher rating on the timbre scale:

- A definite difference in frequency content between A and B - for example, an increase or decrease in low frequency content.
- A difference in frequency content where you are unable to say in what way.
- Artefacts in the sound - for example, clicking, popping or buzzing noises.

PLEASE WRITE DOWN YOUR TEST ID BELOW

Figure 5.4: Test instructions given to subjects

5.3.5 Procedure

1. The subject is asked to sit in a chair in the listening position and is given a brief description of the thesis and what will happen during the test. They are then asked to read the test instructions (see Figure 5.4).
2. If the subject confirms that they understand the definitions of “spaciousness” and “timbre” and understand what will occur in the test, the GUI is launched on the computer by their feet and the test begins.
3. The subject will first enter a “training mode” of six comparisons in order to ensure that they understand what is being asked of them and give some context to what they have read in the test instructions. The training comparisons are

always the same for each subject, and simulated room, in the same order, as follows:

Stimulus A	**Stimulus B**
IIR 0.0	IIR 0.75
Sub-band 0.75	Sub-band 0.0
White Noise 0.0	White Noise 0.75
IIR 0.75	IIR 0.0
Sub-band 0.0	Sub-band 0.75
White Noise 0.75	White Noise 0.0

Table 5.3: Listening test training stimuli, where the number represents the approximate cross correlation coefficient between each channel, and the words are the method of decorrelation

4. The subject then enters "test" mode, where the spaciousness and timbre slider positions are recorded for each comparison. The comparisons as follows have their order randomised and their status as stimulus A or B randomised. Each decorrelation method is compared against the other for the same cross correlation coefficient. Each cross correlation coefficient is compared against the other for each method. This totals 18 comparisons of spaciousness and timbre.

Stimulus A	**Stimulus B**
IIR 0.0	IIR 0.25
IIR 0.0	IIR 0.75
IIR 0.25	IIR 0.75
SubBand 0.0	SubBand 0.25
SubBand 0.0	SubBand 0.75
SubBand 0.25	SubBand 0.75
WhiteNoise 0.0	WhiteNoise 0.25
WhiteNoise 0.0	WhiteNoise 0.75
WhiteNoise 0.25	WhiteNoise 0.75
IIR 0.0	SubBand 0
IIR 0.0	WhiteNoise 0
SubBand 0.0	WhiteNoise 0
IIR 0.25	SubBand 0.25
IIR 0.25	WhiteNoise 0.25
SubBand 0.25	WhiteNoise 0.25
IIR 0.75	SubBand 0.75
IIR 0.75	WhiteNoise 0.75
SubBand 0.75	WhiteNoise 0.75

Table 5.4: Listening test stimuli, where the number represents the approximate cross correlation coefficient between each channel, and the words are the method of decorrelation

5. After the last comparison the subject is asked to stop the audio using the button on the GUI and leave the room, then given some sweets as a thank you for participating.

6

Results and analysis - decorrelation algorithms

The numerical results which represent the performance of each of the six investigated decorrelation algorithms is presented here. These results, in combination with informal listening, informed choices made regarding the formal listening tests which were undertaken for this thesis.

A 5 second long pink noise signal with a sampling frequency of 44.1 kHz was chosen as an input signal to investigate the properties of each decorrelation algorithm. This was done in order to ensure that as many frequencies as possible were processed by the algorithm and because pink noise signal is closer than white noise signal to signals which are likely to be used in a realistic situation - i.e. speech or musical audio.

A target cross correlation coefficient of 0 was aimed for when generating these decorrelation algorithms. This is because it is assumed that this will produce the most extreme transfer functions between input signal and a decorrelated copy, as the decorrelated copy will be least similar to the input signal.

6.1 Impulse responses

Here (Figures 6.1 and 6.2) it is clear that the shortest impulse response used is the FIR decorrelation algorithm. It can also be seen that the FIR algorithm consists of a burst of what is essentially white noise, and the Velvet Noise and White Noise algorithms exponentially decay. The Sub-band and IIR allpass cascade algorithms have a “ramp” up to a certain peak, then an exponential decay, which is important as it means that energy will be gradually added to any signal that it is convolved with. In informal listening, this was not found to be preceptually problematic but it is important to be aware of the properties of the impulse response.

Impulse responses used by the Dynamic algorithm are not shown as that algorithm consists of multiple cross-faded FIR algorithm impulse responses.

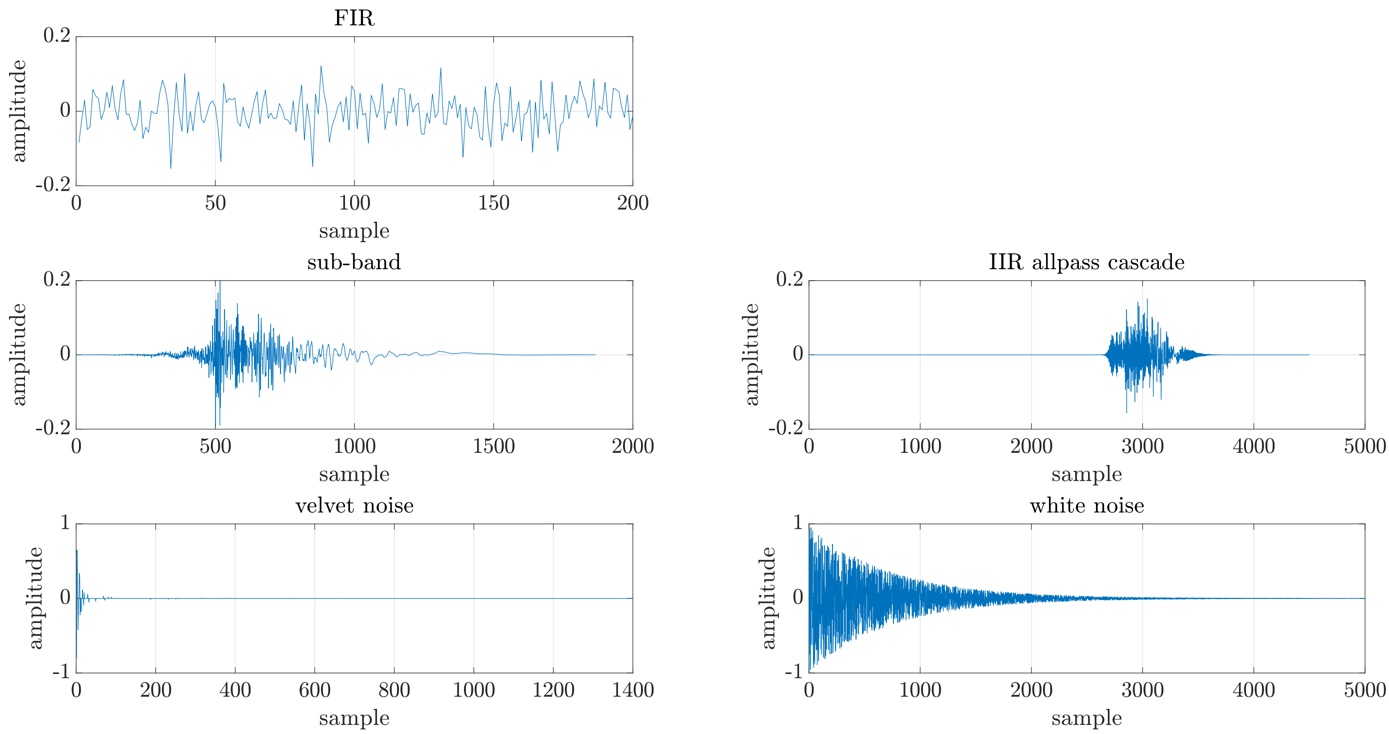

Figure 6.1: Impulse responses designed to produce the first of two decorrelated channels with an input signal of a 5 second long pink noise signal

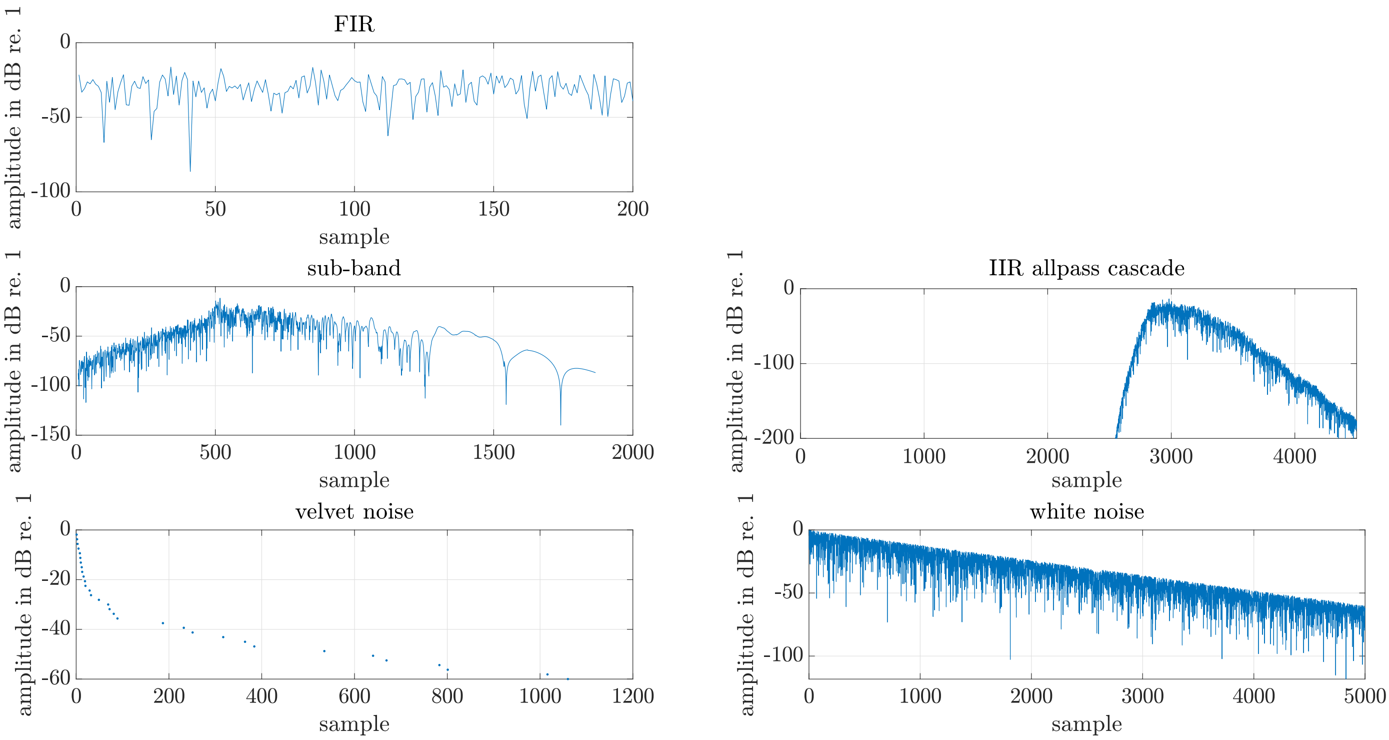

Figure 6.2: Impulse responses designed to produce the first of two decorrelated channels with an input signal of a 5 second long pink noise signal, in dB re. 1

6.2 Transfer function

The IIR cascade and Sub-band decorrelation algorithms have the smallest deviations from a transfer function of 0 dB re. 1 out of all six methods tested (Figure 6.3), and also the flattest in the frequency domain (Figure 6.5).

The Sub-band algorithm has a slight boost and a more erratic transfer function at low frequencies which is due to the close proximity of adjacent critical bands which make up the band pass filters at low frequencies. This causes constructive or destructive intereference between audio in multiple adjacent critical bands. This effect can be also seen to a lesser extent at the meeting points between the frequencies at higher critical bands, but it is less extreme as only two bands interfere, where at low frequencies it can be more than two.

The FIR, Dynamic and White Noise algorithms exhibit various cuts in frequency bands randomly placed in the frequency spectrum (these are different for every randomly generated decorrelation filter). These are most densely placed for the Dynamic algorithm and least for the FIR algorithm. As the phase of the signal is changed by decorrelation, cancellation of certain parts of the signal is inevitable, which is the cause of these cuts in the frequency domain.

The Velvet Noise algorithm exhibits large cuts at some frequencies which are determined by the random nature of placement of impulses within it's impulse response.

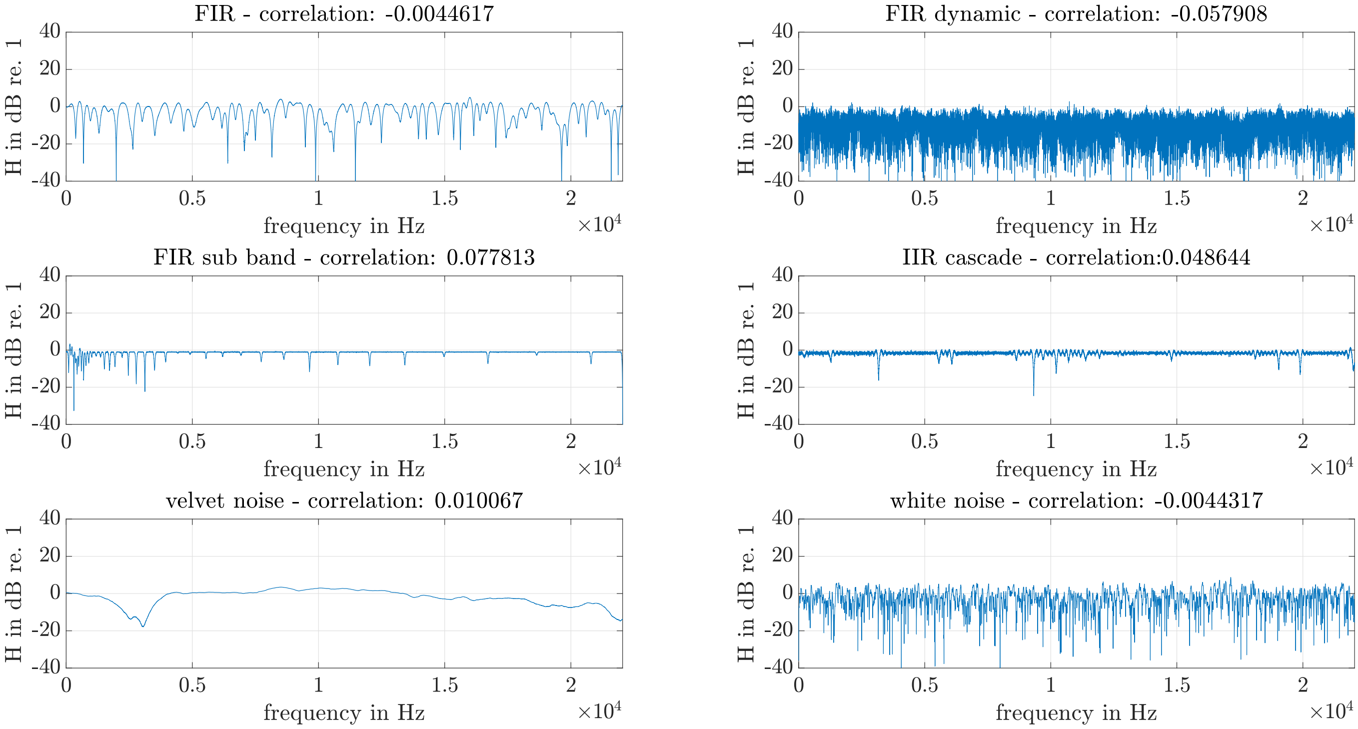

Figure 6.3: Transfer function (H1 estimator), between a 5 second long pink noise input signal and a decorrelated copy of that signal. 'correlation: ' refers to the cross correlation coefficient between two decorrelated copies of the input signal (i.e. left and right channels of a stereo output)

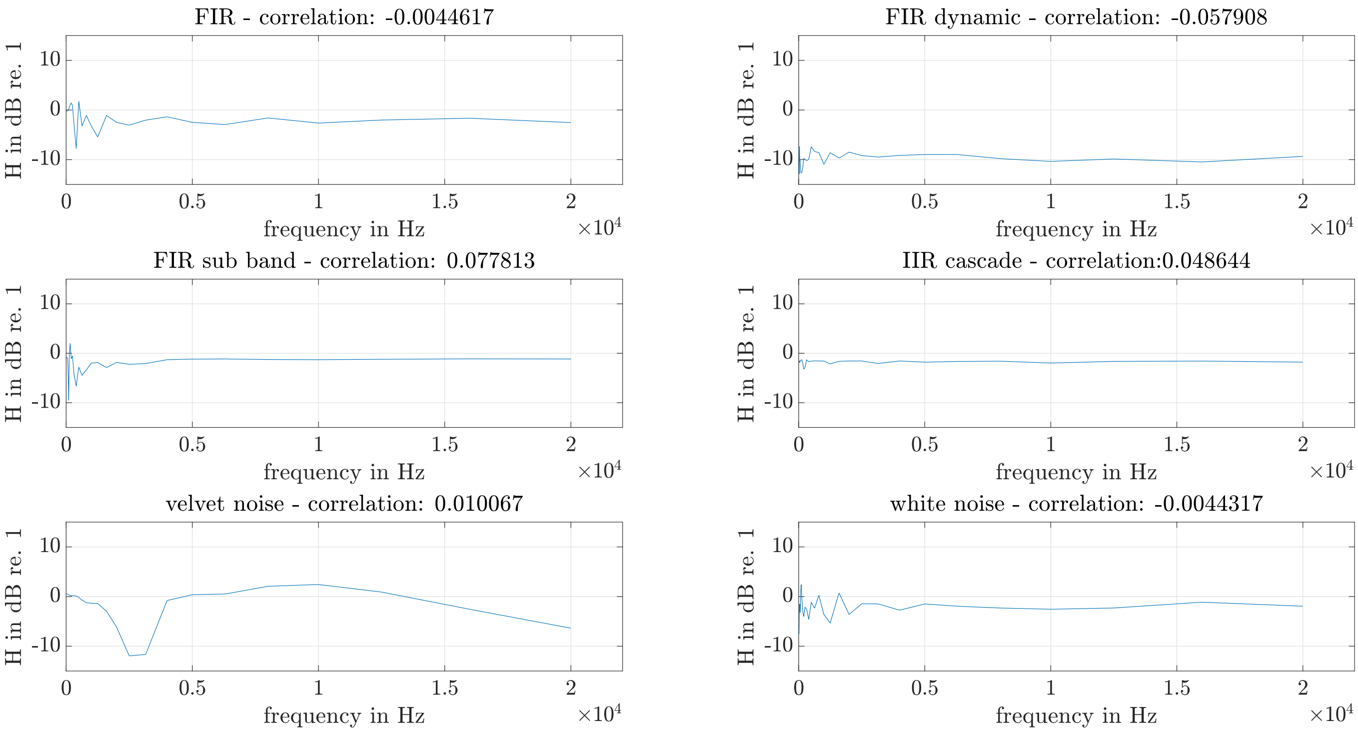

Figure 6.4: Mean transfer function (H1 estimator) within each third octave band, between a 5 second long pink noise input signal and a decorrelated copy of that signal

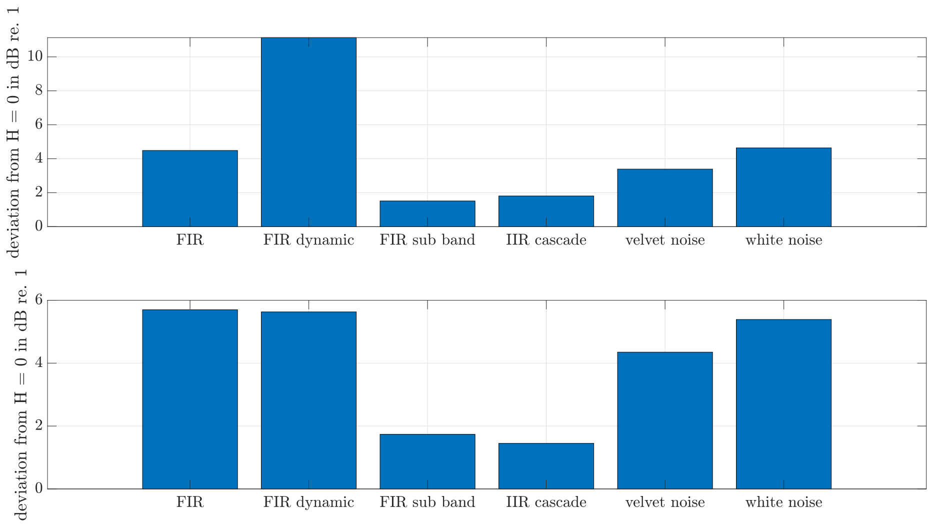

Figure 6.5: Mean average (top) and rms deviation (bottom) from transfer function (H1 estimator) of 0 dB re. 1, between a 5 second long pink noise input signal and a decorrelated copy of that signal

6.3 Cross correlation coefficient

Using correlation optimisation as described in Section 4.1 on all methods except the dynamic decorrelation (for which it is not possible), it is possible to obtain a very low cross correlation coefficient between output channels when the target cross correlation coefficient is 0. This figure generally lies between -0.10 and +0.10 but is different for each new set of randomly generated decorrelation filter. Figure 6.3 shows typical figures from randomly generated decorrelation.

6.4 Group delay

Group delay introduced by decorrelation is important to be aware of so it is possible to compensate for it correctly in later processing. The IIR cascade algorithm has a large group delay due to the large number of allpass filter iterations required to produce decorrelated signals. For example, 1500 iterations were used to produce the signal in Figure 6.6.

FIR, Dynamic, Velvet Noise and White Noise algorithms produce a more even group delay across the whole frequency spectrum because they use decorrelation impulse responses which either decay exponentially in amplitude in time (Velvet Noise and White Noise) or have a random amplitude in time which does not decay (FIR and Dynamic). The Impulse Responses used by the Sub-band and IIR cascade algorithms both increase from zero for some time to a peak, then decay in amplitude, which means that there is a larger variation in group delay across the frequency spectrum.

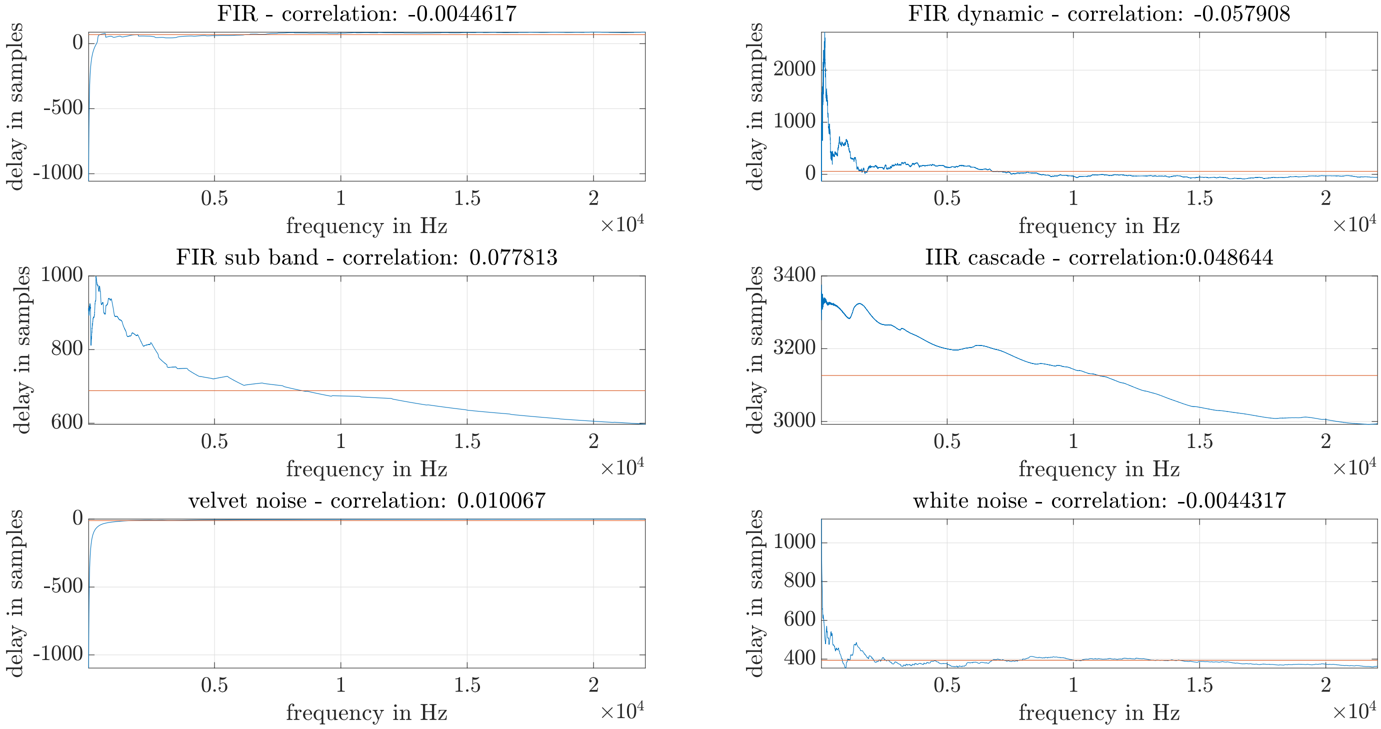

Figure 6.6: Group delay between a 5 second long pink noise input signal and a decorrelated copy of that signal

6.5 Balance

Balance, as described in Section 2.5, is the ratio between the root mean square of the amplitudes of two decorrelated copies of an input signal.

Balance will be different for each randomly generated algorithm, and also it must be noted that it can be corrected for in later processing if that algorithm is to be used for a real purpose.

The Sub-band, IIR cascade and White Noise algorithms have the best (closest to 1) balance. The Velvet Noise algorithm has balance problems because of the way that impulses are randomly distrubuted during its creation.

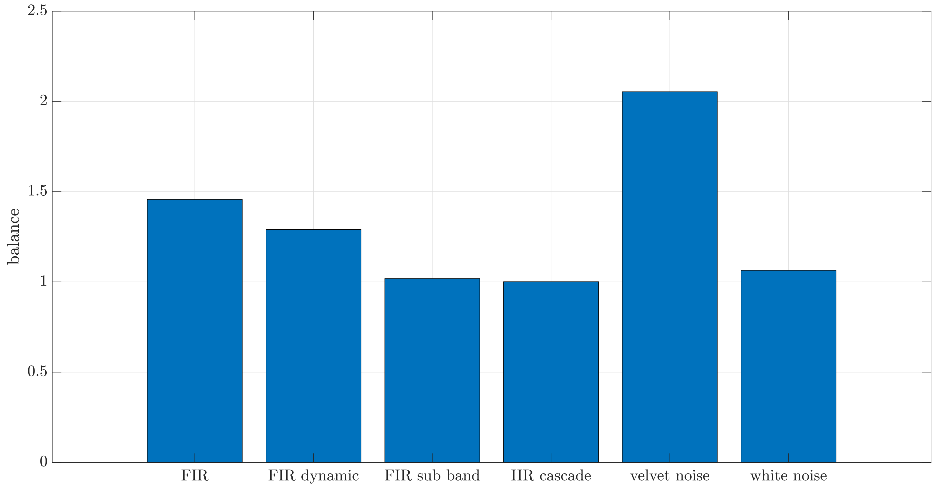

Figure 6.7: Balance between two decorrelated copies of a 5 second pink noise input signal as per Equation 2.9

6.6 Informal listening

In this section, a brief description and notes on how decorrelation *sounds* is given, for various algorithms, cross correlation coefficients and input signals. The aim is to show what is possible with decorrelation and what is not. To give the clearest impression and achieve this aim, an input signal is only decorrelated to produce a stereo output signal, and no processing involving room impulse responses is undertaken. The decorrelated "stereoised" audio is compared to the mono input signal on stereo headphones. In general, it is found that decorrelation in this setting works best and sounds most natural in a musical setting.

6.6.1 Pink noise signal

6.6.1.1 FIR decorrelation

Target cross correlation: 0.75 Slightly larger spatial extent compared to input signal. Audibly different frequency content - perhaps in mid and high frequencies.

Target cross correlation: 0.25 Larger spatial extent compared to input signal. Audibly different to input in mid and high frequencies. Audibly different to 0.75 but hard to say in which way.

Target cross correlation: 0 Largest spatial extent - sounds like two different pink noise signals played from each ear. Phantom source in middle of head almost completely defeated. Sounds more diffuse than input and there appears to be filtering of frequencies in middle and high frequencies, as with 0.75 and 0.25.

6.6.1.2 Dynamic decorrelation

Target cross correlation: 0.75 Signal completely destroyed and turned into a whooshing noise which moves from ear to ear. Very obvious filtering.

Target cross correlation: 0.25 Signal completely destroyed and turned into a whooshing noise which moves from ear to ear. Very obvious filtering.

Target cross correlation: 0 Signal completely destroyed and turned into a whooshing noise which moves from ear to ear. Very obvious filtering. Audibly greater spatial extent than 0.75 and 0.25.

6.6.1.3 Sub-band decorrelation

Target cross correlation: 0.75 Larger spatial extent compared to input. There is an apparent doubling effect as if two pink noise signals are being played at the same time. The timbre is different between this and input signal but difficult to say in what way.

Target cross correlation: 0.25 Same as 0.75 but there are noticably less high frequencies and the spatial extent is slightly larger.

Target cross correlation: 0 Low frequencies still localised in the head but mid and high frequencies panned to be localised on both ears. Sounds like two different pink noise signals played in each ear.

6.6.1.4 IIR allpass cascade decorrelation

Target cross correlation: 0.75 Clicking noise or artefact at beginning of signal. Slightly wider spatial extent. Less loud than input.

Target cross correlation: 0.25 Clicking noise or artefact at beginning of signal. Wider spatial extent. Less loud than input.

Target cross correlation: 0 Clicking noise or artefact at beginning of signal. Sound totally located on ears, phantom source seems to be completely defeated.

6.6.1.5 Velvet noise decorrelation

Target cross correlation: 0.75 No impact on spatial extent. Sounds like a high pass filter.

Target cross correlation: 0.25 No impact on spatial extent. Sounds like a low pass filter.

Target cross correlation: 0 One ear high pass, one ear low pass. Greater spatial extent.

6.6.1.6 White noise decorrelation

Target cross correlation: 0.75 Different in timbre but hard to say how. Slightly greater spatial extent.

Target cross correlation: 0.25 Different in timbre but hard to say how. Greater spatial extent.

Target cross correlation: 0 Different in timbre but hard to say how. Much greater spatial extent - phantom source effect defeated.

6.6.2 1000 Hz sine signal

6.6.2.1 FIR decorrelation

Target cross correlation: 0.75 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0.25 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0 Very slight increase in spatial extent. Very lightly less high frequency content.

6.6.2.2 Dynamic decorrelation

Target cross correlation: 0.75 LFO filter effect means signal is effectively destroyed. Localisation moves from side to side in the head.

Target cross correlation: 0.25 LFO filter effect means signal is effectively destroyed. Localisation moves from side to side in the head.

Target cross correlation: 0 LFO filter effect means signal is effectively destroyed. Localisation moves from side to side in the head.

6.6.2.3 Sub-band decorrelation

Target cross correlation: 0.75 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0.25 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0 Very slight increase in spatial extent. Very lightly less high frequency content.

6.6.2.4 IIR allpass cascade decorrelation

Target cross correlation: 0.75 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0.25 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0 Very slight increase in spatial extent. Very lightly less high frequency content.

6.6.2.5 Velvet noise decorrelation

Target cross correlation: 0.75 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0.25 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0 Very slight increase in spatial extent. Very lightly less high frequency content.

6.6.2.6 White noise decorrelation

Target cross correlation: 0.75 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0.25 Very slight increase in spatial extent. Very lightly less high frequency content.

Target cross correlation: 0 Very slight increase in spatial extent. Very lightly less high frequency content.

6.6.3 Orchestral music (L. van Beethoven (1770-1827) Symphony no. 7, I movement, bars 1-53 [16])

6.6.3.1 FIR decorrelation

Target cross correlation: 0.75 Increased spatial extent. No difference in timbre.

Target cross correlation: 0.25 Even greater spatial extent. Slight boost to low frequencies.

Target cross correlation: 0 Very wide and natural sounding spatial extent. Slight boost to low frequencies.

6.6.3.2 Dynamic decorrelation

Target cross correlation: 0.75 Sound source moves around the centre of the head. Able to tell that the filters are changing over time so there is significant difference in timbre, but not unpleasant.

Target cross correlation: 0.25 Sound source moves around the centre of the head in a more extreme way. Able to tell that the filters are changing over time so there is significant difference in timbre, but not unpleasant.

Target cross correlation: 0 Sound source moves from ear to ear in a random pattern. Able to tell that the filters are changing over time so there is significant difference in timbre, but not unpleasant.

6.6.3.3 Sub-band decorrelation

Target cross correlation: 0.75 Increased spatial extent. No difference in timbre.

Target cross correlation: 0.25 Even greater spatial extent. No difference in timbre.

Target cross correlation: 0 Sound located on both ears, but it still sounds as if each ear hears exactly the same as the input signal. Sounds very good and natural.

6.6.3.4 IIR allpass cascade decorrelation

Target cross correlation: 0.75 Increased spatial extent. No difference in timbre.

Target cross correlation: 0.25 Even greater spatial extent. No difference in timbre.

Target cross correlation: 0 Sound located on both ears, but it still sounds as if each ear hears exactly the same as the input signal. Sounds very good and natural.

6.6.3.5 Velvet noise decorrelation

Target cross correlation: 0.75 High pass filter effect and the sound source is closer to the right ear, but no increased spatial extent.

Target cross correlation: 0.25 High frequency boost and increased noise at high frequencies. Increase in spatial extent.

Target cross correlation: 0 High frequency boost and increased noise at high frequencies. Larger increase in spatial extent.

6.6.3.6 White noise decorrelation

Target cross correlation: 0.75 Natural sounding increase in spatial extent, timbrally indistinguishable from input signal.

Target cross correlation: 0.25 Natural sounding and greater increase in spatial extent, timbrally indistinguishable from input signal.

Target cross correlation: 0 Sound located on both ears, but it still sounds as if each ear hears exactly the same as the input signal. Sounds very good and natural.

6.6.4 Speech (Male Talker 1 [17])

6.6.4.1 FIR decorrelation

Target cross correlation: 0.75 Slightly increased spatial extent and timbrally indistinguishable from input signal, but there is a slight reverberation added, similar to a very small room.

Target cross correlation: 0.25 Increased spatial extent and timbrally indistinguishable from input signal, but there is a slight reverberation added, similar to a very small room.

Target cross correlation: 0 Much greater spatial extent and timbrally indistinguishable from input signal, but there is a slight reverberation added, similar to a very small room.

6.6.4.2 Dynamic decorrelation

Target cross correlation: 0.75 Moving spatial extent, small difference in timbre. Movement is distracting. Small amount of reverberation added.

Target cross correlation: 0.25 Moving spatial extent, small difference in timbre. Movement is distracting. Small amount of reverberation added.

Target cross correlation: 0 Moving spatial extent, small difference in timbre. Movement is distracting. Small amount of reverberation added.

6.6.4.3 Sub-band decorrelation

Target cross correlation: 0.75 Slightly increased spatial extent. Unpleasant high frequency artefacts and distracting reverberation.

Target cross correlation: 0.25 Increased spatial extent. Unpleasant high frequency artefacts and distracting reverberation.

Target cross correlation: 0 Much greater spatial extent. Unpleasant high frequency artefacts and distracting reverberation.

6.6.4.4 IIR allpass cascade decorrelation

Target cross correlation: 0.75 Slightly increased spatial extent. Unpleasant high frequency artefacts and distracting reverberation.

Target cross correlation: 0.25 Increased spatial extent. Unpleasant high frequency artefacts and distracting reverberation.

Target cross correlation: 0 Much greater spatial extent. Unpleasant high frequency artefacts and distracting reverberation.

6.6.4.5 Velvet noise decorrelation

Target cross correlation: 0.75 Very similar to input but sound source has moved slightly to the left.

Target cross correlation: 0.25 Very similar to input but sound source has moved far to the right.

Target cross correlation: 0 Very similar to input but sound source has moved far to the right.

6.6.4.6 White noise decorrelation

Target cross correlation: 0.75 Reverberation added (not unpleasant). Slight increase in spatial extent.

Target cross correlation: 0.25 Reverberation added (not unpleasant). Increase in spatial extent.

Target cross correlation: 0 Reverberation added (not unpleasant). Much greater spatial extent.

7

Results and analysis - listening test

Results of the listening test are presented as box plots and as a percentage of the total possible selection, and separated by the room which the listening test simulated (Church or Drum Room).

In box plots, the central red line is the median, the blue box has upper and lower lines which are the first and third quartiles, the black horizontal lines are the minimum and maximum, and any red crosses are outliers.

In the "percentage of total possible ..." plots, if every participant had chosen the highest possible value on the slider for a certain correlation or algorithm every time it appeared, then the plot would read 100%. It is the sum of the percentage of the maximum from the listening test sliders from every subject.

7.1 Spaciousness

7.1.1 Correlation

For the Church RIR, it is clear from Figure 7.1 that a lower correlation does not mean a more perceptually spacious sound, and that there is a wide variety of opinion. For the 0 vs. 0.75 comparison, the difference between the first and third quartiles is the largest of any comparison for spaciousness, and the minimum and maximum span from a correlation of 0 as the maximum available spaciousness to a correlation of 0.75 as the maximum available spaciousness. All subjects were given the same definition of spaciousness, so this suggests that there is a large difference in opinion on whether the decorrelated "sound" contributes to the feeling of spaciousness for the Church RIR which has a longer reverberation time.

For the Drum Room RIR, Figure 7.1 shows a more direct link between decorrelation and spaciousness. The lower cross correlation coefficient of each comparison was rated as the most spacious. The variation in opinion is smaller than that for the Church RIR.

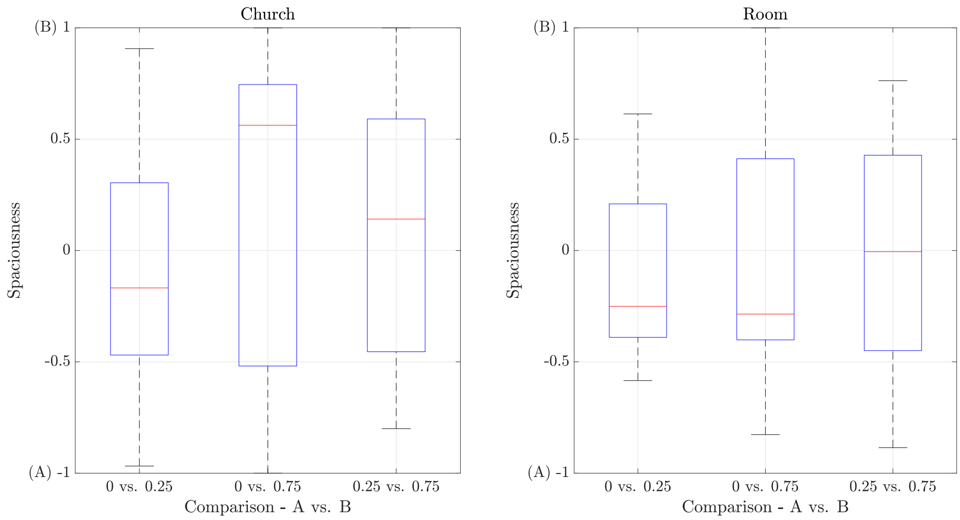

Figure 7.1: Box plot of spaciousness for each correlation comparison, over all algorithms and all subjects

Figure 7.2 shows that of the times it was preferred, the correlation 0.75 stimulus actually had the greatest percentage of the total available spaciousness for the Church RIR, wheras for the Drum Room RIR, the correlation 0 stimulus had the greatest percentage of the total available spaciousness. A conclusion that can be drawn is that there is some fundamental difference in the interaction between decorrelation algorithm and Room Impulse Response with respect to RIRs with different reverberation times.

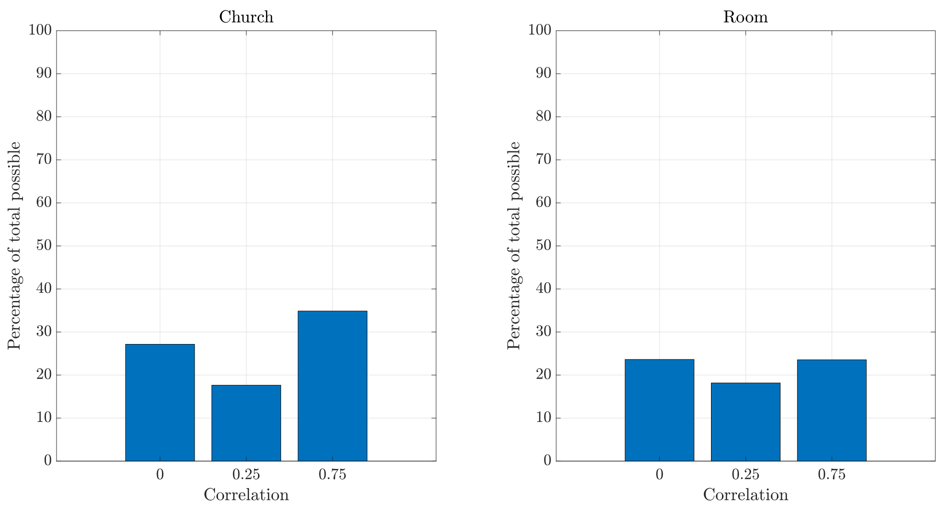

Figure 7.2: Percentage of total possible spaciousness for each correlation comparison, over all algorithms and all subjects

7.1.2 Algorithm

Figure 7.3 shows that the preferred algorithm in terms of spaciousness was White Noise for both RIRs, but again with a wide range of opinion.

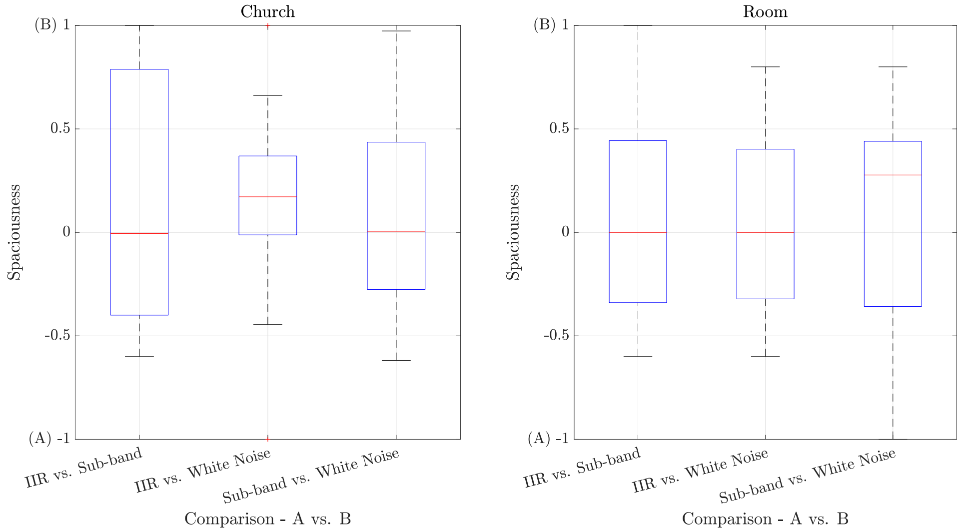

Figure 7.3: Box plot of spaciousness for each algorithm comparison, over all correlations and all subjects

Figure 7.4 shows that the IIR algorithm was the least preferred in total, with comparable scores for the Sub-Band and White Noise algorithms.

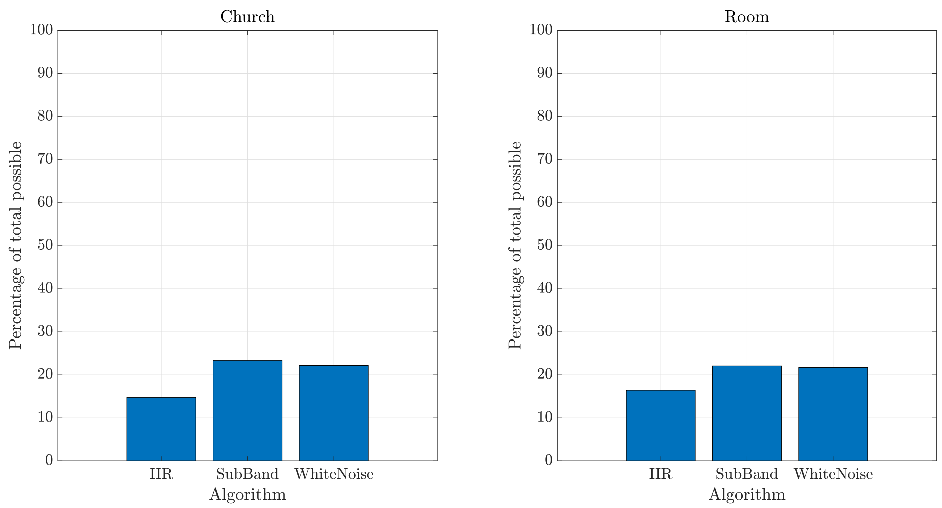

Figure 7.4: Percentage of total possible spaciousness for each algorithm comparison, over all correlations and all subjects

7.2 Timbre

7.2.1 Correlation

Figures 7.5 shows that there is an even larger difference in opinion regarding timbre rating than spaciousness, but for both RIRs, the comparison between 0 and 0.25 or 0 and 0.75 correlation stimuli produced the largest difference in timbre, where comparisons between 0.25 and 0.75 produced a significantly smaller difference in percieved timbre.

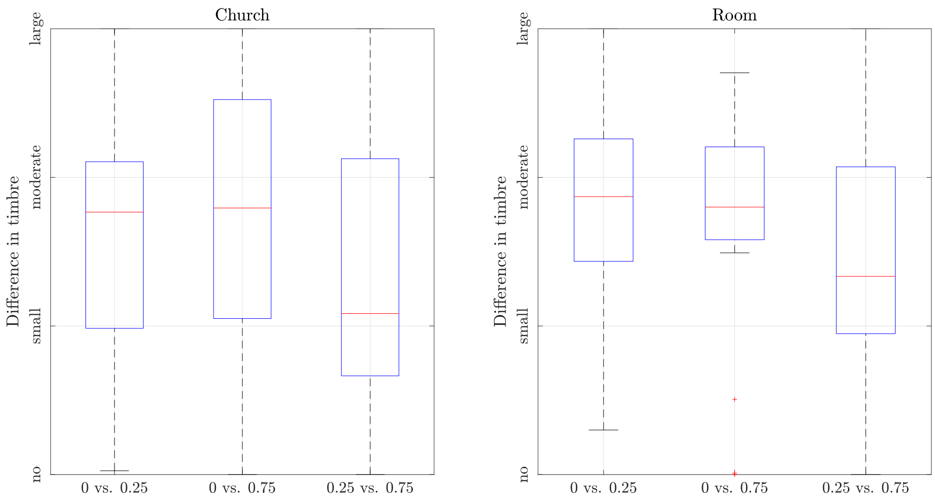

Figure 7.5: Box plot of difference in timbre for each correlation comparison, over all algorithms and all subjects

Figure 7.6 shows agreement that the correlation 0 stimuli produced the largest difference in percieved timbre when compared to either of the other two correlation stimuli.

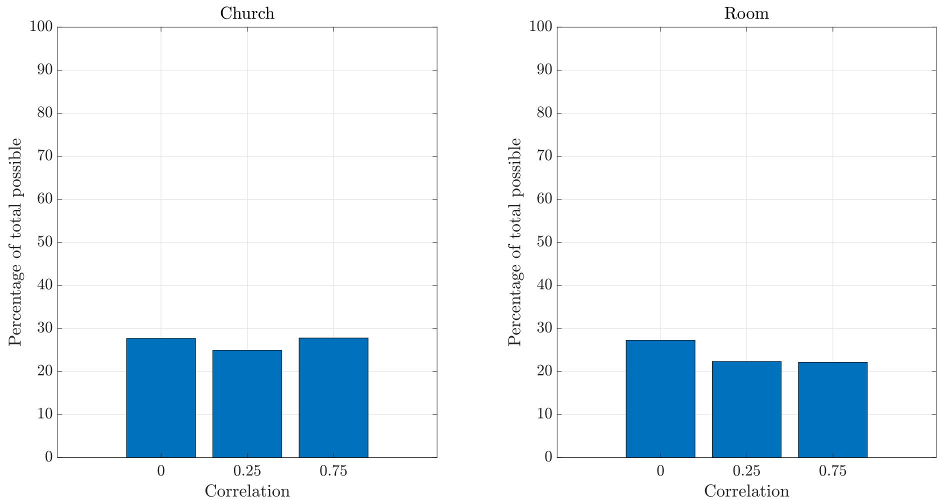

Figure 7.6: Percentage of total possible difference in timbre for each correlation comparison, over all algorithms and all subjects

7.2.2 Algorithm

Figures 7.7 and 7.8 show that there is no significant link between any decorrelation algorithm and how different in timbre stimuli sounded, only that they all produced differences in timbre in a small to moderate way.

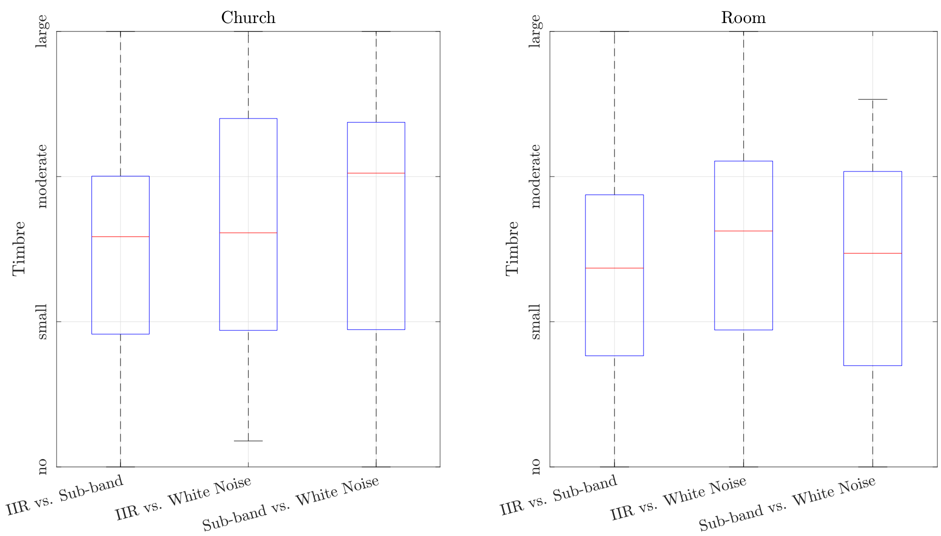

Figure 7.7: Box plot of difference in timbre for each algorithm comparison, over all correlations and all subjects

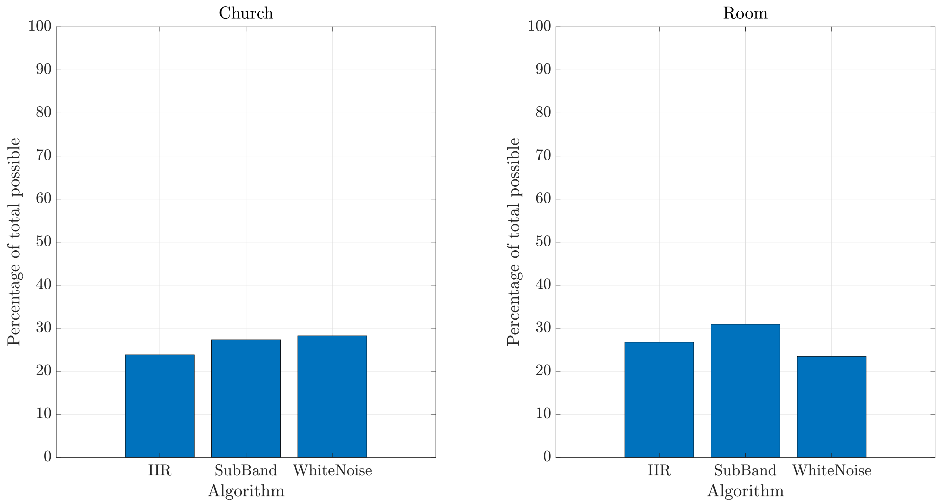

Figure 7.8: Percentage of total possible difference in timbre for each algorithm comparison, over all correlations and all subjects

8

Discussion

Some subjects were already versed in what decorrelation sounds like - often this can be unmistakable feeling of negative pressure around the ears and the perception that the sound is around the head rather than inside it. A common comment was that the stimuli which were clearly more correlated had an emphasised low frequency content - especially on the kick drum and snare drum of the drum loop. It must be noted however, that this observation is not supported by the results of the timbral perception of the stimuli. This low frequency emphasis, whether consciously perceived or not, can contribute to a perception of a longer reverberation time or a more spacious sound, which explains why there is a lot of disparity between results regarding spaciousness for the larger RIR (the Church). In conversation with subjects, this did not seem to be as noticeable for the smaller RIR (Drum Room).

The first loudspeaker always played exactly the same signal for each RIR, so this percieved boost in low frequencies for more correlated diffuse loudspeaker sounds can be attributed to constructive interference at low frequencies between the sound that the diffuse loudspeakers are playing at the listener position. If this were the case, decorrelated sound would not constructively interfere in the same way and so not create the same increase of low frequency content. Another explanation is that there is some pscyhoacoustic effect which leads to an increase in perception of low frequencies when the sound is perceived as being located more inside or closer to the head than outside of it. This explanation has support in that during informal headphone testing, the low frequency emphasis was still present.

This effect is unfortunate as it is desirable to study the absolute effect of decorrelation whilst removing timbral effects completely. A perfect decorrelation algorithm would produce decorrelation between channels without introducing any effects in the frequency domain. A suggestion for further work would be to investigate whether constructive interference is the cause of this effect by using a microphone to record audio at the listener position and comparing that audio to the played signals. If there is some tangible or perceptual increase in low frequencies, then this could be compensated for in future decorrelation algorithms used in this way.

In addition to this, it is possible that stimuli with lower cross correlation will have a "source" loudspeaker which sounds louder in comparison to the "diffuse" loudspeakers than stimuli with lower cross correlation. When processing the signals, it is assumed that the "diffuse" sounds sum up incoherently, so 9 dB is added to the direct part of the RIR for the first loudspeaker (see 4.2.2). When the "diffuse" sounds are not completely decorrelated, they will not meet the assumption of incoherent summation and so increase in power by some factor which is less than 9 dB, meaning that the “source” loudspeaker will appear louder than the “diffuse” loudspeakers. This may be accountable for variation in listening test results.

A perceptual effect which was commonly reported by some subjects was the impression that the reverberation for some stimuli was coming from one particular direction (e.g. left or right) rather than from all directions, which is what was intended. This was reported to make it difficult to decide on which stimuli was more spacious, as the perceived room would move around between A and B stimuli. The effect is attributed to one or more of the “diffuse” loudspeakers unintentionally playing louder sound than loudspeakers on the other side of the loudspeaker circle. All decorrelating/room impulse responses were normalised to have the same rms amplitude, but that does not necessarily mean that when they have been convolved with the input signal that the output will have exactly the same rms amplitude, meaning there is a chance that some loudspeakers are louder than others. To eliminate this effect in the future, a further rms normalisation after convolution could be implemented, but this would not have been possible using the current setup of the convolution taking place inside external software.

In informal conversation with subjects it was also reported that some stimuli sounded “better” or “more realistic” than others. This (as with spaciousness and timbre) is a subjective assertion, but further study could assess whether there is any link between stimuli that sound “better” or “more realistic” and a certain decorrelation algorithm or correlation.

A decrease in high frequencies can correspond to a perception of a source sounding further away (as high frequencies are attenuated at long distances), it can be said that “spaciousness” and “timbre” are interlinked and are therefore poor metrics to assess at the same time. This may account for wide disagreement in responses observed in the results of the listening test. The sound which is perceived may either be interpreted as being more spacious *or* being different in timbre, not both or neither. “Spaciousness” and “timbre” are linked in many other ways which cannot easily be described and so it may be desirable to investigate them separately or find different qualities to assess.

Finding another listening test format for which it would be possible to determine whether subject responses were inconsistent would also be another avenue for further study. This could be done by repeating the same comparisons in randomised order to see if the choices made by the subject are the same. Subjects who made too many inconsistent judgments would have their whole set of responses discarded. This scheme may reduce disparity between overall responses and produce a more definite conclusion.

Another suggestion for further study would be to use a greater number of participants. This would lend a greater weight of statistical significance to listening test results.

A wider number of cross correlation coefficients could also be used - for example, something closer to perfectly correlated diffuse loudspeaker sounds, such as a cross correlation coefficient of 0.9, could be used to further investigate the effect of decorrelation.

9
Conclusion

Based on the results of numerical and perceptual analysis, the following can be concluded:

- The decorrelation algorithm which will give the best transfer function between input and output (IIR allpass cascade and Sub-band) are those which have impulse responses which increase from zero to a maximum, then decay exponentially again. For this reason they are also the decorrelation algorithms which have the most uneven impact on group delay over the frequency spectrum, so it is important to take this into account when choosing a decorrelation algorithm for a certain purpose, depending on that purpose.
- A lower cross correlation coefficient between loudspeakers playing the diffuse part of the RIR will mean a more spacious perception for a RIR with a short reverberation time, but a less spacious perception for a RIR with a longer reverberation time.
- Of the three algorithms tested (IIR allpass cascade, Sub-band and White Noise), the Sub-band and White Noise algorithms produce the most perceptually spacious sound.
- The lowest cross correlation coefficient between loudspeakers (0) produced a small to moderate difference in timbre when compared to the higher cross correlation stimuli (0.25 and 0.75).
- The stimulus with lowest cross correlation coefficient between diffuse loudspeakers (0) produced a median moderate difference in timbre when compared to higher cross correlation stimuli (0.25 and 0.75), but when comparing those stimuli, the median difference in timbre was small. This suggests that large differences in timbre are introduced when decorrelating signals to an extreme extent, which can be expected as the signals sent to each loudspeaker are more disimilar to each other.
- Different decorrelation algorithms were perceived as having small to moderate differences in timbre between them, but not in a way that appears as coherent in the listening test results. This could be expected by the result of the numerical analysis as the Sub-band and IIR cascade algorithms had transfer functions between input and output of close to 1 at all frequencies (not all), but it is less expected for the White Noise algorithm which has a more erratic transfer function.

- Investigation into what causes a reported (but not observed by the results of the listening test) increase at low frequencies for more highly correlated diffuse loudspeaker signals would be desired in order to further study the perceptual effects of decorrelation.
- The cause of the perception of reverberation coming from a particular direction rather than from all directions could be ascertained.
- Further investigations could also be made in the subject by using more listening test subjects, greater variety of correlations, different criteria such as "realism" or different listening test schemes.

Bibliography

[1] G. S Kendall “The decorrelation of audio signals and its impact on spatial imagery”, 1996, Computer Music Journal, December 1996

[2] R. Litovsky, H. S. Colburn, W. A. Yost, S. J. Guzman, “The precedence effect”, 1999, The Journal of the Acoustical Society of America 106, November 1999

[3] The MathWorks, Inc “Transfer function estimate” 2020, [Online] Available: https://www.mathworks.com/help/signal/ref/tfestimate.html

[4] The MathWorks, Inc “Shift phase angles” 2020, [Online] Available: https://se.mathworks.com/help/matlab/ref/unwrap.html

[5] M. Gorzel, G. Kearney, A. Foteinou, S. Hoare, S. Shelley “Lady Chapel, St Albans Cathedral” 2010, [Online] Available: https://openairlib.net/?page_id=595

[6] M. Bouéri and C. Kyirakakis “Audio signal decorrelation based on a Critical Band Approach”, 2004, Convention Paper 6291, Proceedings of the 117th Audio Engineering Society Convention, San Fransisco, CA, USA, 28th - 31st October, 2004

[7] H. Fastl, E. Zwicker ”Pscyhoacoustics: Facts and Models” 2007

[8] B. R. Glasberg and B. C. J. Moore ”Derivation of auditory filter shapes from notched-noise data” 1990 Hearing Research, Vol. 47, Issues 1-2, p. 103-138

[9] E. Kermit-Canfield and J. Abel “Signal decorrelation using perceptually informed allpass filters”, 2016, Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16), Brno, Czech Republic, 5th–9th September, 2016

[10] B. Alary, A. Politis and V. Välimäki ”Velvet-noise decorrelator” 2017 Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17), Edinburgh, UK, 5th – 9th September, 2017

[11] M. Karjalainen and H. Järveläinen ”Reverberation modelling using velvet noise” 2007 Proceedings of the Audio Engineering Society 30th International Conference, Saariselkä, Finland, 15th - 17th March, 2007

[12] Voxengo ”Free Reverb Impulse Responses” 2020, [Online] Available: https://www.voxengo.com/impulses/

[13] Jack Audio Connection Kit “Jack Audio Connection Kit” 2020, [Online] Avail-

able: https://jackaudio.org

[14] Pure Data “Pure Data” 2020, [Online] Available: https://puredata.info

[15] spatialaudio.net “SoundScape Renderer” 2020, [Online] Available: http://spatialaudio.net/ssr/

[16] T. Lokki, J. Pätynen and V. Pulkki “Anechoic recordings of symphonic music”, 2008, [Online] Available: https://users.aalto.fi/~ktlokki/Sinfrec/sinfrec.html

[17] Starkey ”Starkey Research: Open Access Stimuli: The "Rainbow Passage" Narrative” 2020, [Online] Available: https://starkeypro.com/research/research-resources/open-access-stimuli

A

Appendix - decorrelation algorithm settings

A.1 Settings used to produce numerical results

output channels	2
filter length	200 samples
number of test impulse responses (t)	10

Table A.1: FIR decorrelation settings for numerical results

output channels	2
filter length	200 samples
block length	500 ms
overlap	250 ms

Table A.2: Dynamic decorrelation settings for numerical results

output channels	2
ERB filter length	1000
maximum delay	20 ms
number of test impulse responses (t)	10

Table A.3: Sub-band decorrelation settings for numerical results

output channels	2
iterations	1500
number of test impulse responses (t)	10

Table A.4: IIR allpass cascade decorrelation settings for numerical results

output channels	2
filter length	1400
impulse density	1000
impulse decay	60 dB
number of test impulse responses (t)	10

Table A.5: Velvet noise decorrelation settings for numerical results

output channels	2
filter length	5000
impulse decay	60 dB
number of test impulse responses (t)	10

Table A.6: White noise decorrelation settings for numerical results

A.2 Settings used to produce listening test stimuli

output channels	8 (1 source and 7 diffuse)
iterations	1500
number of test impulse responses (t)	500

Table A.7: IIR allpass cascade decorrelation settings for listening tests

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.13						
4		0.38	0.07					
5		-0.14	-0.27	-0.14				
6		-0.24	0.12	-0.02	-0.13			
7		0.12	0.09	0.54	-0.03	-0.06		
8		0.29	0.32	0.50	-0.36	0.10	0.20	

Absolute Mean:	0.20

Table A.8: IIR allpass cascade decorrelation, target correlation 0, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.27						
4		0.30	0.49					
5		0.20	-0.08	-0.01				
6		0.40	0.41	0.28	-0.03			
7		0.54	0.10	0.09	0.43	0.25		
8		0.11	0.58	0.26	-0.15	0.28	-0.11	

Absolute Mean:	0.26

Table A.9: IIR allpass cascade decorrelation, target correlation 0.25, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.83						
4		0.48	0.46					
5		0.83	1.00	0.46				
6		0.48	0.46	1.00	0.46			
7		1.00	0.83	0.48	0.83	0.48		
8		0.83	1.00	0.46	1.00	0.46	0.83	

Absolute Mean:	0.70

Table A.10: IIR allpass cascade decorrelation, target correlation 0.75, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.23						
4		0.39	0.10					
5		-0.20	-0.39	-0.18				
6		-0.21	0.24	-0.04	0.00			
7		0.14	0.13	0.49	-0.06	-0.13		
8		0.35	0.33	0.59	-0.51	0.07	0.24	

Absolute Mean:	0.24

Table A.11: IIR allpass cascade decorrelation, target correlation 0, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.11						
4		0.23	0.61					
5		0.19	-0.23	-0.15				
6		0.44	0.29	0.27	-0.14			
7		0.65	-0.09	-0.06	0.36	0.29		
8		-0.05	0.65	0.35	-0.21	0.21	-0.19	

Absolute Mean:	0.27

Table A.12: IIR allpass cascade decorrelation, target correlation 0.25, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.88						
4		0.46	0.41					
5		0.88	1.00	0.41				
6		0.46	0.41	1.00	0.41			
7		1.00	0.88	0.46	0.88	0.46		
8		0.88	1.00	0.41	1.00	0.41	0.88	

Absolute Mean:	0.69

Table A.13: IIR allpass cascade decorrelation, target correlation 0.75, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

output channels	8 (1 source and 7 diffuse)
ERB filter length	1000
maximum delay	20 ms
number of test impulse responses (t)	500

Table A.14: Sub-band decorrelation settings for listening tests

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.18						
4		0.20	0.03					
5		-0.05	0.03	-0.01				
6		0.30	-0.15	0.16	-0.17			
7		-0.11	0.08	-0.70	0.20	-0.16		
8		0.62	0.06	0.06	-0.11	0.10	-0.06	

Absolute Mean:	0.17

Table A.15: Sub-band decorrelation, target correlation 0, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.31						
4		0.40	0.28					
5		0.17	0.13	0.27				
6		0.42	0.21	0.07	0.16			
7		0.45	0.14	0.62	0.30	0.39		
8		0.25	0.77	0.37	0.07	0.27	0.24	

Absolute Mean:	0.30

Table A.16: Sub-band decorrelation, target correlation 0.25, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.82						
4		0.74	0.75					
5		0.75	0.79	0.73				
6		0.70	0.77	0.84	0.73			
7		0.64	0.76	0.75	0.65	0.71		
8		0.82	1.00	0.75	0.79	0.77	0.76	

Absolute Mean:	0.76

Table A.17: Sub-band decorrelation, target correlation 0.75, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.04						
4		-0.06	0.09					
5		-0.37	0.04	0.07				
6		-0.04	-0.12	-0.07	-0.09			
7		0.14	0.06	-0.54	0.10	-0.02		
8		-0.53	0.29	-0.02	0.10	-0.12	0.05	

Absolute Mean:	0.14

Table A.18: Sub-band cascade decorrelation, target correlation 0, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.17						
4		0.19	0.13					
5		0.17	0.13	0.17				
6		0.35	0.04	0.32	0.21			
7		0.62	0.50	0.19	0.10	0.36		
8		0.28	0.41	0.28	0.34	-0.08	0.01	

Absolute Mean:	0.24

Table A.19: Sub-band cascade decorrelation, target correlation 0.25, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.74						
4		0.78	0.74					
5		0.74	0.70	0.70				
6		0.66	0.65	0.69	0.80			
7		0.60	0.70	0.59	0.76	0.68		
8		0.63	0.65	0.65	0.73	0.74	0.72	

Absolute Mean:	0.70

Table A.20: Sub-band cascade decorrelation, target correlation 0.75, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

output channels	8 (1 source and 7 diffuse)
filter length	5000
impulse decay	60 dB
number of test impulse responses (t)	500

Table A.21: White noise decorrelation settings for listening tests

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.00						
4		0.15	0.09					
5		0.07	-0.12	-0.05				
6		-0.11	0.08	-0.18	0.01			
7		0.11	0.22	-0.17	-0.15	0.05		
8		-0.28	0.09	-0.09	-0.09	0.14	-0.05	

Absolute Mean:	0.11

Table A.22: White noise decorrelation, target correlation 0, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.26						
4		-0.02	0.21					
5		0.19	0.12	0.16				
6		0.24	0.27	0.04	0.03			
7		-0.18	0.11	0.35	0.36	0.03		
8		-0.04	0.26	0.32	0.08	0.34	0.15	

Absolute Mean:	0.18

Table A.23: White noise decorrelation, target correlation 0.25, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1								
2								
3		0.69						
4		1.00	0.69					
5		0.69	1.00	0.69				
6		0.69	0.66	0.69	0.66			
7		1.00	0.69	1.00	0.69	0.69		
8		0.69	1.00	0.69	1.00	0.66	0.69	

Absolute Mean:	0.77

Table A.24: White noise decorrelation, target correlation 0.75, Drum Room - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1								
2								
3		0.09						
4		0.04	0.14					
5		-0.15	-0.09	-0.04				
6		-0.02	0.18	-0.28	0.03			
7		-0.09	0.24	-0.06	-0.09	-0.06		
8		-0.25	0.05	-0.02	0.15	0.23	-0.20	

Absolute Mean:	0.12

Table A.25: White noise decorrelation, target correlation 0, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1								
2								
3		0.23						
4		0.24	0.30					
5		0.27	0.16	0.24				
6		0.31	0.28	0.29	0.24			
7		0.31	0.21	0.25	0.32	0.40		
8		0.51	0.12	0.38	0.17	0.44	0.33	

Absolute Mean:	0.29

Table A.26: White noise decorrelation, target correlation 0.25, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8

Channel	1	2	3	4	5	6	7	8
1								
2								
3		0.70						
4		0.70	0.65					
5		0.78	0.73	0.79				
6		1.00	0.70	0.70	0.78			
7		0.83	0.73	0.74	0.80	0.83		
8		1.00	0.70	0.70	0.78	1.00	0.83	

Absolute Mean:	0.78

Table A.27: White noise decorrelation, target correlation 0.75, Church - cross correlation coefficients between the output signals played by loudspeakers 2-8