

„Wiedergabe von Ausgedehnten Quellen in der Wellenfeldsynthese“

“Presentation of Virtual Sources with an Increased Auditory Source Width in Wave Field Synthesis”

Bachelor’s Thesis

at the Department of Telecommunication Systems

Assessment of IP-based Applications

Technische Universität Berlin



Author: Konstantin Koslowski (316955)
Thesis advisor: Prof. Dr.-Ing Alexander Raake
Second examiner: Prof. Dr.-Ing Sebastian Möller
Supervisor: Hagen Wierstorf
Submission date: 18.01.2013

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und eigenhändig sowie ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

I hereby declare that I have written this bachelor's thesis completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Berlin, 19. Februar 2013

KONSTANTIN KOSLOWSKI

Zusammenfassung

Diese Bachelorarbeit behandelt die wahrgenommene Breite von Signalen, die auf mehreren virtuellen Quellen mit Wellenfeldsynthese wiedergegeben werden.

Dafür werden drei Methoden implementiert und ausgewertet. Bei zwei dieser Methoden handelt es sich um bekannte Forschungsergebnisse aus dem Bereich der Stereophonie, es wird ein einzelnes, monaurales Eingangssignal verwendet und zu mehreren Ausgangssignalen verarbeitet. Diese werden dann auf einer Stereo-Lautsprecheranordnung wiedergegeben.

Im Verlauf dieser Studie werden beide Methoden an die Wellenfeldsynthese angepasst und implementiert. Dazu kommt eine dritte Methode die aus dem Eingangssignal keine vollständigen Ausgangssignale generiert, sondern dieses in verschiedene Frequenzbereiche zerlegt. Anschliessend wird jeder Frequenzbereich auf einer eigenen virtuellen Quelle wiedergegeben.

Zum Abschluss der Arbeit werden die implementierten Methoden benutzt um Signale für einen Hörversuch zu erzeugen. Dieser wird an einer Gruppe Versuchspersonen durchgeführt und die wahrgenommene Breite und Qualität ausgewertet.

Abstract

This thesis is about increasing the perceived auditory source width of signals presented on multiple virtual sources in a wave field synthesis setup.

For this process three methods have been implemented and evaluated. Two of these methods are known to work on stereo setups, taking a single monaural input signal, processing it into two signals and then presenting them on two loudspeakers. Both methods are now modified to generate a number of signals for presentation in a wave field synthesis setup. Additionally there is a third method that does not generate a number of complete output signals, but splits the input signal into its critical bands which are then individually presented on virtual sources.

Finally, samples generated with these implementations are presented to a group of test subjects to evaluate the auditory source width as well as the perceived quality.

Contents

Eidesstattliche Erklärung	III
1 Introduction	1
2 Theory and Related Work	3
2.1 Kendall: The Decorrelation of Audio Signals	4
2.1.1 Effects of Decorrelation on Multi-Channel Presentation	4
2.1.2 Creating Decorrelated Signals	5
2.2 Bouéri and Kyirakakis: Audio Signal Decorrelation Based on a Critical Band Approach	7
2.3 Critical Bands	8
2.4 Wave Field Synthesis	8
2.5 Conclusion and Deduction of Research Question	9
3 Implementation	11
3.1 Kendall: The Decorrelation of Audio Signals	12
3.2 Bouéri: Audio Signal Decorrelation Based on a Critical Band Approach	14
3.3 Critical Bands	15
4 Experiment	19
4.1 Conduction	19
4.2 Software	23
4.2.1 SoundScape Renderer	23
4.2.2 Matlab/Octave	23
5 Results	27
5.1 Plain Results	27
5.1.1 Auditory Source Width	27
5.1.2 Perceptual Quality	28
5.2 Normalized Results	29
5.2.1 Auditory Source Width	29
5.2.2 Perceptual Quality	30
5.3 Discussion	31
6 Conclusion	33

Bibliography

35

1 Introduction

When listening to a virtual point source it is perceived as very narrow. In sound reproduction with wave field synthesis the presentation of sources with an increased auditory source width is not yet easily possible and there has been little research in this field. Due to this, in most cases point sources are used instead.

The goal of this study is to research the possible methods to increase the auditory source width in wave field synthesis.

In stereophonic reproduction the auditory source width of a virtual source can be increased by utilizing multiple virtual sources rendering the same content but with an inter-aural cross coefficient (IACC) that is close to zero, as discussed by Gary S. Kendall in 1995 [8]. Otherwise effects such as timbral coloration and combing occur and the spatial extent of the presented signals does not increase.

Generation of these signals was subject of many papers, but most focus on stereo setups. One notable example is the paper “The Decorrelation of Audio Signals” [8] by Gary S. Kendall. He was one of the first to research the auditory source width and how to modify it.

Another method was presented in a paper called “Audio Signal Decorrelation Based on a Critical Band Approach” [3] by Maurice Bouéri and Chris Kyirakakis in 2004.

The third and last method investigated in this thesis has not yet been discussed in-depth, it is loosely based on the paper “Phantom Source Widening With Deterministic Frequency Dependent Time Delays” [14] by Franz Zotter et al. released in 2011 and focuses on splitting the signal into different frequency bands.

Taking these methods and results from research regarding stereo setups as a base, this thesis focuses on wave field synthesis. For this the existing methods are adapted and implemented for use in the existing wave field synthesis setup.

As in the previous work, a reference signal is processed into several signals. But then instead of presenting it on a stereo setup, it is presented on a multitude of virtual sources positioned in a circular arrangement at some distance from each other.

After implementing these methods, running some tests and analysis, an experiment will be conducted to evaluate the resulting auditory source width and quality of the generated signals.

In the end, the results from the experiment and suggestions for future research will be discussed.

2 Theory and Related Work

This thesis is founded on a variety of papers about increasing the auditory source width for signals presented in stereo setups. The work from Gary S. Kendall is most notable, being one of the first to research the impact of decorrelated signals on their perceived auditory source width. In his paper *The Decorrelation of Audio Signals*, Kendall states that “[i]n the field of spatial hearing, signal decorrelation is known to have dramatic impact on the perception of sound imagery” [8].

His research mainly focuses on methods to decorrelate a monophonic audio signal into several output signals and its impact on spatial perception.

A different approach was developed by Maurice Bouéri and Chris Kyirakakis. In their paper *Audio Signal Decorrelation Based on a Critical Band Approach* [3] they split the input signal into critical bands and then add a random delay meeting certain requirements before merging them to a single signal again. This is enough to process the given input signal into decorrelated output signals with a decreased IACC.

The term “critical band” was introduced by Harvey Fletcher in the 1940s and describes the bandwidth of the cochlea acting as an auditory filter for hearing in the inner ear. With each subsequent critical band the bandwidth increases.

The last method to focus on is loosely based on the paper *Phantom Source Widening With Deterministic Frequency Dependent Time Delays* [14] by Franz Zotter et al.

Here the input signal is also split into critical bands, like in the work from Bouéri et al. but instead of being provided with a random delay and merged into a single signal, each band is individually presented as a virtual source.

With the achievements in the area of wave field synthesis it is possible to use the previous research and implement it on a bigger scale. Compared to stereo, wave field synthesis offers far more possibilities for rendering sound effects. One notable example of the research on wave field synthesis is the paper *The Theory of Wave Field Synthesis Revisited* [12] by Sascha Spors et al. It discusses the basic foundations on which wave field synthesis is based and presents examples of practical setups.

2.1 Kendall: The Decorrelation of Audio Signals

When presenting signals in a stereo setup the use of decorrelated signals over a single monophonic signal has several positive effects as will be explained in separate sections. Decorrelation can be a by-product of effects such as reverberation or chorusing. Alternatively it can be achieved by recording the same scene twice on separate tracks. In his work Kendall goes into detail on how decorrelated signals can be generated from a monophonic input signal.

2.1.1 Effects of Decorrelation on Multi-Channel Presentation

Presenting decorrelated signals on a multi-channel setup instead of using the same signal for each channel can have the following effects [8].

Elimination of the Perception of Constructive and Destructive Interference

Simultaneous presentation of signals on multiple loudspeakers creates interference patterns. These patterns are either constructive or destructive and can usually be perceived by the listener as “combing” or “coloration”.

In figure 2.1 the first subplot shows the amplitude spectrum of white Gaussian noise. The second displays the average amplitude spectrum of 5 decorrelated samples of white Gaussian noise, constructive and destructive interferences are altering the amplitude spectrum. This is called combing effect.

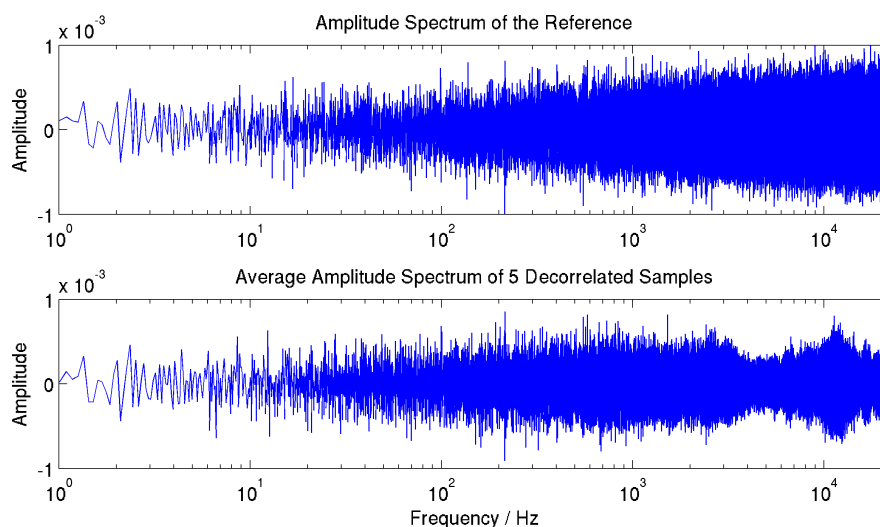


Figure 2.1: Combing effect when presenting multiple decorrelated signals

Creation of Diffuse Sound Fields without Reverberation

In concert hall acoustics the perception of spatial diffuseness correlates with the inter-aural cross-coefficient (IACC), a cross-correlation coefficient between the signals arriving at the left and right ear. An IACC close to 1 means the signals are very similar, values close to 0 imply highly decorrelated signals. When perceiving a reverberating diffuse sound field, the IACC is close to zero and the field is perceived as having a wide spatial extent.

This effect can be achieved by presenting signals with a low cross-correlation coefficient to the listener over a loudspeaker setup. If there is just a single signal available, Kendall states that “a single channel of reverberation can be made spatially diffuse by decorrelation” [8].

Reduction of “Image Shift” of Diffuse Sound Fields & Elimination of the Precedence Effect

Listening to multiple sources presenting the same signal but with a time delay of less than 1.0 ms creates a single sound image between the sources. According to Barron, who in 1971 discovered this “Image Shift Effect” [1], the perceived location of this sound image is close to the signal that arrived first.

The “Precedence Effect” is characterized by the perceived sound location being the original location, ignoring reflections. This is dependent on the actual delay between original and reflected sounds, it decreases after about 10 ms and completely vanishes after 50 ms [15].

2.1.2 Creating Decorrelated Signals

In order to gain the advantages of decorrelated signals without recording multiple tracks it is possible to transform a given input signal into several decorrelated output signals that still sounds perceptually equal to the input signal.

The correlation between two signals y_1 and y_2 is called the inter-aural cross coefficient and calculated by a cross-correlation function displayed in equation 2.1.

$$IACC(\Delta t) = \lim_{t \rightarrow \infty} \int_{-T}^{+T} y_1(t) \cdot y_2(t + \Delta t) dt \quad (2.1)$$

The results from this equation are within a range of $[-1, 1]$ where 1 equals to both signals being identical, -1 means that they are identical but phase shifted by 180 degrees, while results close to 0 stand for heavily decorrelated signals.

The method used by Kendall simulates a filter with a finite impulse response (FIR) by convolving the input signal $x(nT)$ with a number of filter coefficients $h_i(nT)$, shown in figure 2.2. The IACC of the filter coefficients $h_i(nT)$ is maintained through the convolution and as a result the IACC of the output signals $y_i(nT)$ is of similar dimensions, as displayed in equation 2.2.

$$IACC(h_i(nT)) \approx IACC(y_i(nT)) \quad (2.2)$$

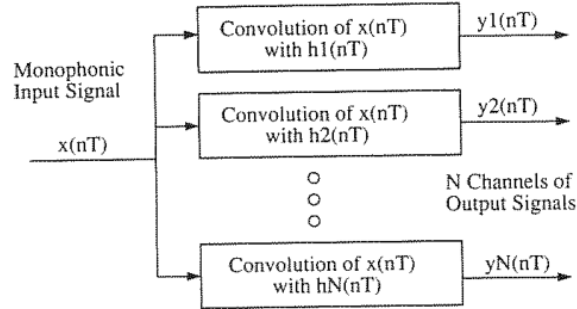


Figure 2.2: Decorrelation with monophonic input and multiple output channels [8]

To create a set of FIR filter coefficients $h_i(nT)$, magnitude and phase must be specified in the frequency-domain and transformed back into the time-domain with an inverse fast Fourier transformation (IFFT). With the magnitude set to a constant value of 1 the resulting FIR filter will be an all-pass. To create a set of these filter coefficients with near-zero correlation coefficients the phase must be randomized with specified frequency and wrapped around to be inside the interval $[-\pi, \pi]$. The higher the frequency, the lower the correlation coefficients. In fact the whole range for correlation coefficients from $[-1, 1]$ is only dependent on the phase. This allows preparation of filter signals with the desired correlation coefficients and the phase only has very little effect on the timbre of the generated output signals but a “dramatic impact on spatial perception” [8, p. 73].

However, there are some restrictions that apply to the FIR filter coefficients to avoid perceptual changes. The filter’s impulse response may not exceed a temporal duration of 20 ms without causing diffusion in the time domain and smearing the input signals transient properties. Also the length of the filter coefficients needs to be variable, depending on the type of input. For example speech requires fewer coefficients to keep its timbral coloration.

But regardless of those circumstances the randomized phase may always lead to varying results, and as a consequence the generated signals should always be evaluated by ear.

2.2 Bouéri and Kyirakakis: Audio Signal Decorrelation Based on a Critical Band Approach

Bouéri and Kyirakakis’ paper was released in 2004 [3] and starts with explaining that by splitting a signal into critical bands and applying a random time shift to each band, the signal will become decorrelated. As mentioned before there are certain advantages of using decorrelated signals over uncorrelated ones when presenting multiple signals on a loudspeaker setup. It is important not to alter the perceptual parts of the signal and yet change them in a way that affects their correlation.

The critical bands are obtained by filtering the signal with an equivalent rectangular bands (ERB) filterbank with a high enough order to feature sharp edges. They represent the frequency bandwidth of the cochlea in the human ear.

In order to achieve a decorrelation the random time offsets must exceed a certain value but may not exceed 20 ms because otherwise they will also alter the perceptual information contained in the signal. When choosing values between -20 and 20 ms “the precedence effect is destroyed” [3, p. 2]. Because of the higher frequencies in the upper critical bands they are also more sensitive to time shifts and should be handled with more care than the lower bands. Figure 2.3 shows the allowed maximum time shift in ms for each critical band. Another aspect that should be taken into account is that

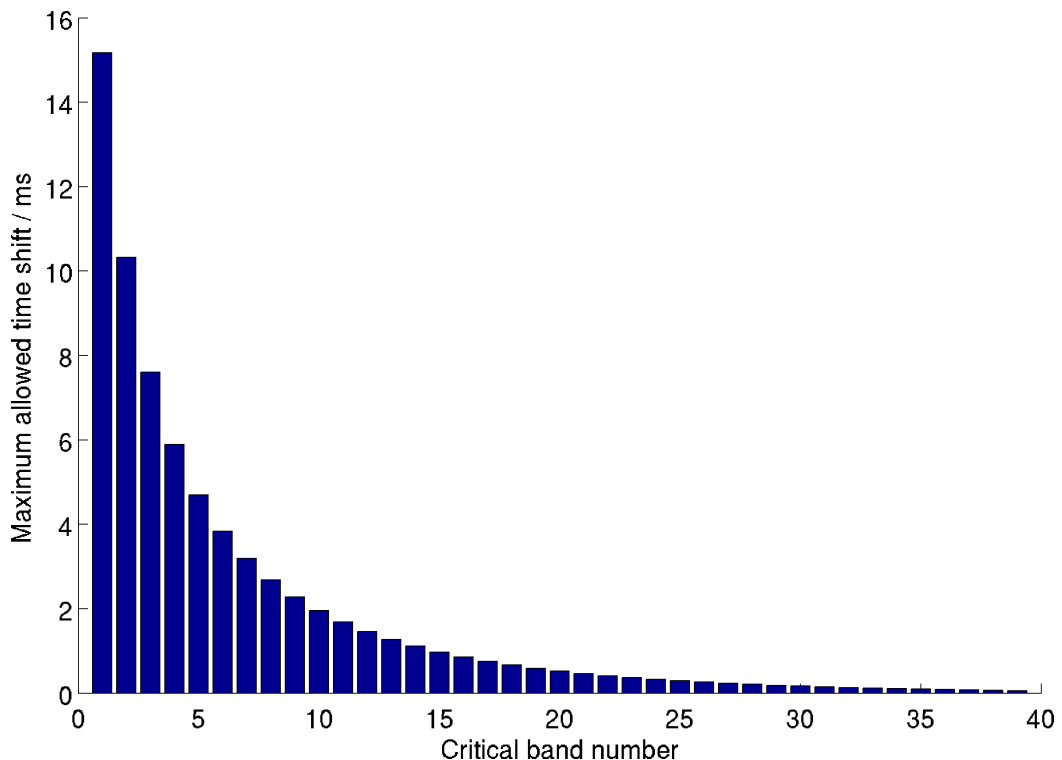


Figure 2.3: Maximum time shifts for each critical band / ms

when reassembling the modified critical bands into a single signal the time shifts of adjacent frequencies may cause a destructive addition. Therefore any possible time shift

t_{shift} has to be a multiple of the time corresponding to the frequency f_{lim} between the adjacent bands as in equation 2.3.

$$t_{\text{shift}} = n \cdot \frac{1}{f_{\text{lim}}}, \quad n \in \mathbb{N} \quad (2.3)$$

With these conditions met the decorrelation algorithm is supposed to generate signals with perceptually unaltered content but increased auditory source width when presented to a listener.

2.3 Critical Bands

After investigating the previously described methods by Kendall and Bouéri et al. and the work by Franz Zotter et al. another possibility comes up. The generated output signals need to have an IACC near zero to each other and when presented by a single virtual source there should be no noticeable difference from the original signal. But instead of generating multiple signals each of which does comply with these conditions, it is sufficient that the presented scene does.

The idea is to split the signal into critical bands similar to the method of Bouéri et al. and then present the critical bands as virtual sources. When perceiving all of them simultaneously they should add up to something similar to the input signal but as each features a different frequency domain their IACC is very close to zero.

When increasing the distance between the virtual sources presenting the critical bands, according to the previous research on stereo setups, auditory source width should increase notably.

2.4 Wave Field Synthesis

Wave field synthesis (WFS) is a technique for reproducing spatial sound fields, initially presented by A. J. Berkhout in 1988 [2].

It uses a high number of individually driven loudspeakers surrounding the listening area. Each of these speakers emits a wave front so that the combination of these wave-fronts creates the wave field of the virtual source. In contrast to well known techniques like stereo or surround sound, the location of the created virtual sources is not dependent on the listener's position.

The paper “The Theory of Wave Field Synthesis Revisited” [12] released in 2008 by Sascha Spors et al. provides a detailed explanation on the subject.

2.5 Conclusion and Deduction of Research Question

After describing two previously researched methods for decorrelating audio signals on stereo setups and one possible method for improving the auditory source width the methods now need to be implemented and evaluated on the wave field synthesis setup. First, each method has be realised in a programming environment. This will be done in Matlab [9] and Octave [4] respectively as they already do feature a lot of signal processing algorithms, are widely used and very well documented. Additionally there is a variety of extensions available for the programs.

Each method will be implemented optimized and perceptually tested in the existing wave field synthesis setup.

Once that is done, an experiment will be conducted to verify the expectations.

3 Implementation

In the following the tests conducted for this thesis will be described. The goal is to process an input signal into multiple decorrelated output signals presented as virtual sources in a wave field synthesis setup, thus increasing the auditory source width. As explained in the previous section all algorithms used were originally designed for stereo setups, because of this modifications are necessary to apply them on a wave field synthesis setup. These modifications are detailed in this section. Each method takes a reference signal as input and transforms it into multiple output signals to be presented on the wave field synthesis setup. These output signals are decorrelated with an IACC that is aimed to be close to zero.

When processing the signals to decorrelate them, they are supposed to be perceptually equal to the given input signal. This may not always succeed and is going to be evaluated in the experiment in addition to the auditory source width.

Important applications used are Matlab [9] and the open source implementation Octave [4] which achieves compatibility with Matlab code in most ways. All algorithms used for this thesis are implemented in Matlab code. This section will end with a detailed explanation of the experiment carried out in the course of this thesis during which the test subjects evaluated a series of sound samples created using the modified algorithms. The sound samples are presented in a room equipped with a setup of loudspeakers for wave field synthesis.

This setup consists of an array of 56 loudspeakers in a circular arrangement with 3m diameter. The speakers are controlled by the software SoundScape Renderer [5] and the scenes are prepared with PureData [7]. To switch between the defined scenes a simplified graphical user interface written in Python is used.

In the previous section the methods from Gary S. Kendall [8], Maurice Bouéri and Chris Kyriakakis [3] for decorrelating input signals into multiple output signals and a method splitting the signal into critical bands have been introduced. The first two methods were developed for stereo setups, the third was not yet discussed in-depth.

In the following section the implementation of these methods for a wave field synthesis setup will be developed out step by step.

3.1 Kendall: The Decorrelation of Audio Signals

First Kendall's method from 1995 is implemented for wave field synthesis. The implementation is done in Matlab and tries to stay close to the original method presented in the paper [8]. The goal of this method is to generate a specified number of decorrelated signals from a reference input signal. In comparison to stereo setups the necessary number of signals is a lot higher for presentation on wave field synthesis setups. This means that each output signal must have a low IACC when compared to every other output signal in order to avoid the occurrence of perceptual artefacts.

To achieve this the input signal is taken and convoluted with a coefficient array of the same length. These coefficients consist of a constant amplitude and randomized phase. The phase starts at zero degrees, then increases in randomized steps with a specified frequency. Before convolving the input signal with this coefficient array, it needs to be transformed from the frequency into the time domain with an inverse Fourier transformation.

The first step generates a randomized phase array (see listing 3.1). Each value is generated as a sum of the previous and a step calculated with the desired phase frequency. As the actual step is multiplied with a random value $[0, 1]$, it has an average value of about $\text{step}_{\text{avg}} = \frac{1}{2} \cdot \text{step}$ and needs to be doubled. When exceeding the range $[-\pi, \pi]$, the value is wrapped around. To remove sudden changes and irregularities a cubic spline interpolation of medium smoothness is applied to the resulting phase.

Listing 3.1: rndphase.m

```

1 function phase = rndphase(len, fp, fs, start)
2
3 % first element
4 randm(1) = start;
5 last = start;
6
7 % calculate step
8 step = fp*2*pi/fs;
9
10 % randomization slows down phase change, increase step
11 % this may of course result in a higher phase frequency
12 step = 2*step;
13
14 % increase previous value
15 for k=2:1:len
16
17     current_step = rand*2*step;
18     current = last+current_step;
19
20     % wrap when exceeding range [-pi, pi]
21     if current > pi
22         old_current=current;
23         current = rem(current, pi)-pi;

```

```

24 elseif current < -pi
25     old_current=current;
26     current = rem(current, pi)+pi;
27 end
28
29 last = current;
30 rndm(k)=current;
31 end
32
33 % interpolate the signal with cubic splines
34 x = 1:len;
35 y = rndm(1,:);
36 % smoothness of interpolation [0;1]
37 p = 0.5;
38 cs=csaps(x,y,p,x);
39 % cubic spline interpolation tends to exceed range [-pi; pi]
40 % ignore those values and use the randomized ones
41 ymax = 3.1;
42 index_rnd = find(abs(cs) > ymax );
43 index_cs = find(abs(cs) <= ymax );
44 phase(index_rnd) = rndm(index_rnd);
45 phase(index_cs) = cs(index_cs);

```

The generated phase is then used to compute the coefficients applied to the convolution in listing 3.2. For this it is mirrored and adapted to conform to an array that was Fourier transformed. The complex spectrum is set as $\text{compspec} = e^{i \cdot \text{phase}}$ and is transformed to the time-domain with an inverse fast Fourier transformation and only its real part is used.

Listing 3.2: rndcoeff.m

```

1 function coeff = rndcoeff(len, cols, rows, fp, fs)
2
3 for ii=1:1:cols
4     for jj=1:1:rows
5         start = 0;
6
7         % rndphase length, keep in mind that the result is mirrored
8         lenphase = len/2+1;
9         phase = rndphase(lenphase, fp, fs, start);
10        % mirror to conform to a fourier transformed array
11        phase = [ phase -phase(end-1:-1:2) ];
12        compspec = exp(i*phase);
13
14        % apply inverse fast fourier transformation
15        coeff(ii,jj,:) = real(iff(compspec));
16    end
17 end

```

Finally the generated coefficients with appropriate length and phase frequency are convolved with the original signal to generate output signals in listing 3.3. In this process the correlation coefficient between the randomized coefficients is kept at a very similar value for the resulting output signals.

Listing 3.3: decorrelate-kendall.m

```

1 function outsig = decorrelate(insig, num, phase_frequency, fs)
2
3 % input can consist of multiple signals
4 len = size(insig,1);
5 rows = size(insig,2);
6
7 % generate $num sets of random coefficients
8 coeff = rndcoeff(len, num, rows, phase_frequency, fs);
9
10 % generate output signals
11 for ii=1:num
12     for jj=1:size(insig,2)
13         % convolute signal
14         conv_signal = conv(insig(:,jj), squeeze(coeff(ii,jj,:)));
15         % cut off unwanted part
16         outsig(:,ii,jj) = conv_signal(end-len+1:end);
17     end
18 end

```

3.2 Bouéri: Audio Signal Decorrelation Based on a Critical Band Approach

Secondly the method used for generating signals by Maurice Bouéri and Chris Kyriakakis is implemented to generate signals for presentation on the wave field synthesis setup. Released in 2004 [3] their method creates another possible approach to generate decorrelated output signals besides the one developed by Kendall [8]. Like the previous method it was originally designed for stereo setups. With an increasing number of decorrelated signals their IACC does not constantly stay at the desired value close to zero.

The used code is displayed in listing 3.4, it starts with defining the limit frequencies between the adjacent critical bands. These frequencies are used to calculate limit samples as a base for the possible time shifts applied to each critical band.

Now a gammatone filter bank, as explained in “Frequency analysis and synthesis using a Gammatone filterbank” [6] by V. Hohmann in 2002, is applied to the input signal. The filter bank is part of the LTFAT toolbox [11] and features separation of a given signal into critical bands. While not using ideal filters and therefore generating overlapping bands, it provides necessary algorithms to reassemble the bands without causing destructive or constructive additions. Each critical band undergoes testing to generate a randomized time shift that is within the allowed range of $[-20, 20]$ ms and also is a multiple of the time corresponding to the frequency limit between the adjacent bands f_{lim} . After completing this procedure the generated time shifts are applied and all bands are reassembled to form the output signal using the gammatone filter bank again. Finally, the volume level is restored to match the level of the input signal.

Listing 3.4: decorrelate-boueri.m

```

1 function sig = decorrelate_boueri(sig,fs)

```

```

2
3 % limit frequencies between filters (compare Fig. 3)
4 limit_freqs = [...
5     89.5   125.8   166.2   211.2   ...
6     261.4   317.3   379.5   448.9   ...
7     526.1   612.2   708.0   814.8   ...
8     933.7   1066.3  1213.9  1378.2   ...
9     1561.5  1765.5  1993.0  2246.1   ...
10    2528.1  2842.4  3192.6  3582.7   ...
11    4016.9  4500.8  5039.8  5640.5   ...
12    6309.4  7054.6  7884.6  8809.3   ...
13    9839.5 10987.6 12265.5 13689.1   ...
14    15275.5 17042.5 ...
15 ];
16 limit_samples = floor(1./limit_freqs*fs);
17
18 % apply filterbank
19 [filtered_sig,cf] = filterbank(sig,fs,'analysis');
20 level_sig = db(rms(sig));
21 % tuning factor, 1 == 1/center_frequency
22 % 1: normal, 0.5: improved decorrelation
23 fact = 0.9;
24 % getting maximum possible time shift (compare Fig. 4)
25 max_shifts = floor(1./(fact.*cf).*fs);
26
27 % get the random starting shift
28 random_number = rand(1)*2-1; % -1..1
29 sample_shift = floor(random_number*1/(fact*cf(1))*fs);
30 filtered_sig_shifted(1,:) = shift_sig(filtered_sig(1,:),sample_shift);
31 % shift the other frequency channels, but considering the constraint that
32 % adjacent shifts could only be a multiple of the time associated to
33 % 1/limit_freq, where limit_freq is the frequency limit between the two filters
34 k = -100:100; % integer multiples
35 for ii = 2:size(filtered_sig,1)
36     % calculate possible shifts as multiple of the limit frequency between
37     % adjacent bands
38     possible_shifts = sample_shift+k.*limit_samples(ii-1);
39     % use only those shifts which are below the maximum shifting threshold
40     possible_shifts = possible_shifts(abs(possible_shifts)<max_shifts(ii));
41     %length(possible_shifts)
42     % choose randomly one of the possible shifts
43     sample_shift = possible_shifts(round(rand*(length(possible_shifts)-1))+1);
44     % shift the signal in this frequency band
45     filtered_sig_shifted(ii,:) = shift_sig(filtered_sig(ii,:),sample_shift);
46 end
47
48 % synthesize signal
49 sig = filterbank(filtered_sig_shifted,fs,'synthesis');
50 sig = gaindb(sig,level_sig-db(rms(sig)));

```

3.3 Critical Bands

The third implemented method differs from the methods discussed by Gary S. Kendall [8] and Maurice Bouéri et al. [3]. It does not generate signals that are perceptually indistinctive compared to the original signal, but splits the signal into critical bands

and generates an individual output signal from each band.

Those output signals are highly decorrelated, as shown in table 4.2, although this was not achieved by applying any kind of transformation, because each of them resides in a different frequency domain of the signal bandwidth. Splitting the signal into critical bands and presenting each of them as an independent virtual source can have an impact on the perceived source location. The virtual sources are likely to be arranged in an array with some distance between them while each critical band covers its own frequency domain within the signal bandwidth. For samples like white Gaussian noise, that are equally spread across the entire bandwidth, this results in an equal sound intensity from all virtual sources, nevertheless for all samples with a different spectrum the sound intensity will deviate.

The chosen solution to this is to align the bands according to their energy level and evenly arrange them around the desired central virtual source.

The algorithm for this is displayed in listing 3.5. It starts with setting the required configuration for the gammatone filter bank analyzer and synthesizer from the Auditory Modelling Toolbox [10] (AMT). These objects are initialized and used to filter the input signal into critical bands and to adjust their time delays and volume levels. The 39 generated critical bands are then added into groups of 2 to 3 bands since there are only 15 virtual sources available.

As a last step in processing the input signal the 15 output signals need to be arranged according to their energy level, otherwise when mapping them to the virtual sources their perceived location may shift from the center.

This is done using the function displayed in listing 3.6, the signals are simply summed up and then sorted in descending order.

Listing 3.5: decorrelate-bands.m

```

1 % Generate a wav file containing decorrelated channels using the bands
2 % decorrelation function
3
4 %% === Configuration =====
5 fs = 44100;
6 %% === Input Files =====
7 sig = wavread('./samples/raw/wgn.wav');
8 outfile = './samples/experiment/wgn_bands.wav';
9
10 % === Get the gammatone filterbank analyzer and synthesizer =====
11 % Filterbank settings
12 flow = 70;           % / Hz
13 fhigh = 20000;       % / Hz
14 fbase = 1000;        % / Hz
15 desired_delay = 0.004; % / s
16 filters_per_erb = 1;
17 filter_order = 4;
18 % do not narrow or widen filters
19 bandwidth_factor = 1;
20
21 analyzer = gfb_analyzer_new( ...
22     fs, flow, fbase, fhigh, ...
23     filters_per_erb, filter_order, ...

```



```

24     bandwidth_factor ...
25 );
26 synthesizer = gfb_synthesizer_new(...
27     analyzer, ...
28     desired_delay ...
29 );
30
31 %% ==== Split the signal in different bands =====
32 % Here nothing will be decorrelated, but the different bands will be stored
33 % as single files, in order to place at different positions within the SSR.
34 % Before storing them as wav files they will be adjusted accordingly.
35
36 % apply filterbank
37 filtered_sig = gfb_analyzer_process(analyzer, sig);
38 % adjust time delays
39 filtered_sig = gfb_delay_process(synthesizer.delay, filtered_sig);
40 % adjust levels
41 outsig = bsxfun(@times,filtered_sig,synthesizer.mixer.gains');
42 outsig=filtered_sig;
43
44 % add bands to the available 15 channels
45 clear outsig_added;
46 disp('make_decorrelated_sigs_bands) adding up to 15 channels')
47 delta=[1 1 1 2 2 2 2 2 2 2 2 2 2 1 1 1];
48 cur=1;
49 for ii=1:15
50     next=cur+delta(ii);
51     outsig_added(ii,:)=sum(outsig(cur:next,:),1);
52     cur=next+1;
53 end
54 outsig=outsig_added;
55
56 % generate proper format
57 outsig=outsig';
58
59 % sort_bands
60 outsig=make_sorted_bands(outsig);
61
62 % save_files
63 outsig = outsig./(max(abs(outsig(:)))+0.01);
64 wavwrite(outsig,fs,16,outfile);

```

Listing 3.6: sort-bands.m

```

1 function outsig = make_sorted_bands(insig)
2 num=size(insig,2);
3
4 for ii=1:num
5     sigsum(ii)=sum(abs(insig(:,ii))));
6 end
7
8 for ii=1:num
9     index(ii)=find(sigsum==max(sigsum));
10    sigsum(index(ii))=0;
11    outsig_sorted(:,ii)=insig(:,index(ii));
12 end
13
14 outsig=outsig_sorted;

```


4 Experiment

To conduct the experiment a number of scenes were prepared to evaluate the three implemented methods. A scene refers to a selected number of signals, generated from a sample, that are presented on virtual sources.

Now instead of presenting a limited amount of sources in a stereo setup, the wave field synthesis setup allows a much higher number. Unfortunately, as the signals need to be convolved in real-time to their desired location and this was only possible with up to 15 virtual sources during this research. This drastic increase of signals presented can have effects on both auditory source width as well as the perceived quality so both aspects need to be evaluated.

4.1 Conduction

A total of 18 sets was prepared and all use the same wave field synthesis setup. The virtual sources are covering a 28° angle at a radius of 2.5 m while the wave field synthesis array consists of 56 speakers with a 1.5 m radius. Detailed in figure 4.1 is the position of the loudspeakers assembling the wave field synthesis array, the position of the virtual sources and the listener. During the experiment the test subjects were seated in the central position of the array and faced the center of the virtual sources.

From those 18 sets, there are 6 for each of the three methods. Three sets to evaluate the auditory source width and the remaining three for quality. Each set uses one of four input signals that was decorrelated and set up in 6 different scenes, comparable to the guidelines of Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) the input signal is presented as a single point source in the first scene.

The other scenes, as shown in figure 4.2 for the methods after Kendall and Bouéri et al. and in figure 4.3 for the third method, present an increasing number of decorrelated signals on the virtual sources. Starting with three sources close to each other in scene 2, then increasing the distance between the sources in scene 3. Scene 4 features 5 sources, scene 5 features 7 and scene 6 presents signals on all available 15 sources. The active virtual sources are always spread equally to both sides of the central source.

For the methods after Kendall and Bouéri et al. each active virtual source presents one of the decorrelated signals which is perceptually indistinctive to the original signal. But the signals generated by the third method just contain a single critical band and

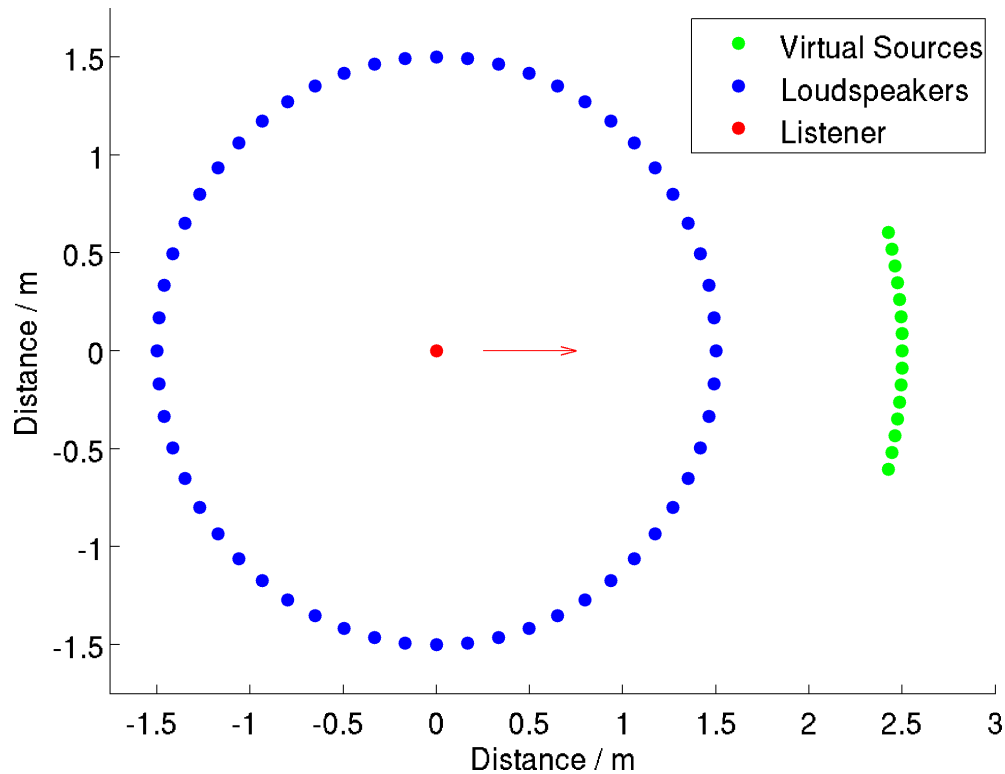


Figure 4.1: Source, loudspeaker and listener locations in the experiment

therefore is it necessary to always present all of them simultaneously. This is done as shown in table 4.1. The virtual sources are numbered with 0 for the central source, descending for the left side until -7 for the leftmost and ascending for the right side until 7 for the rightmost virtual source.

With each scene more signals are combined on each virtual source until in scene 6 all 15 signals are presented on an unique virtual source. The active virtual sources chosen for the scenes in the third method very similar to those in the other two methods.

During the experiment the order of the scenes is randomized and different for all test subjects.

For the experiment the following samples are used:

1. Classical Orchestra
2. Speech
3. Castanets
4. White Gaussian Noise

These samples have a length of 10 s, beginning with a fade-in and ending with a fade-out. They describe most standard situations. The classical orchestra covers most frequency domains while speech is limited to roughly 80 Hz to 12 kHz. Castanets feature very short attacks and white Gaussian noise (WGN) equally covers all frequency domains.

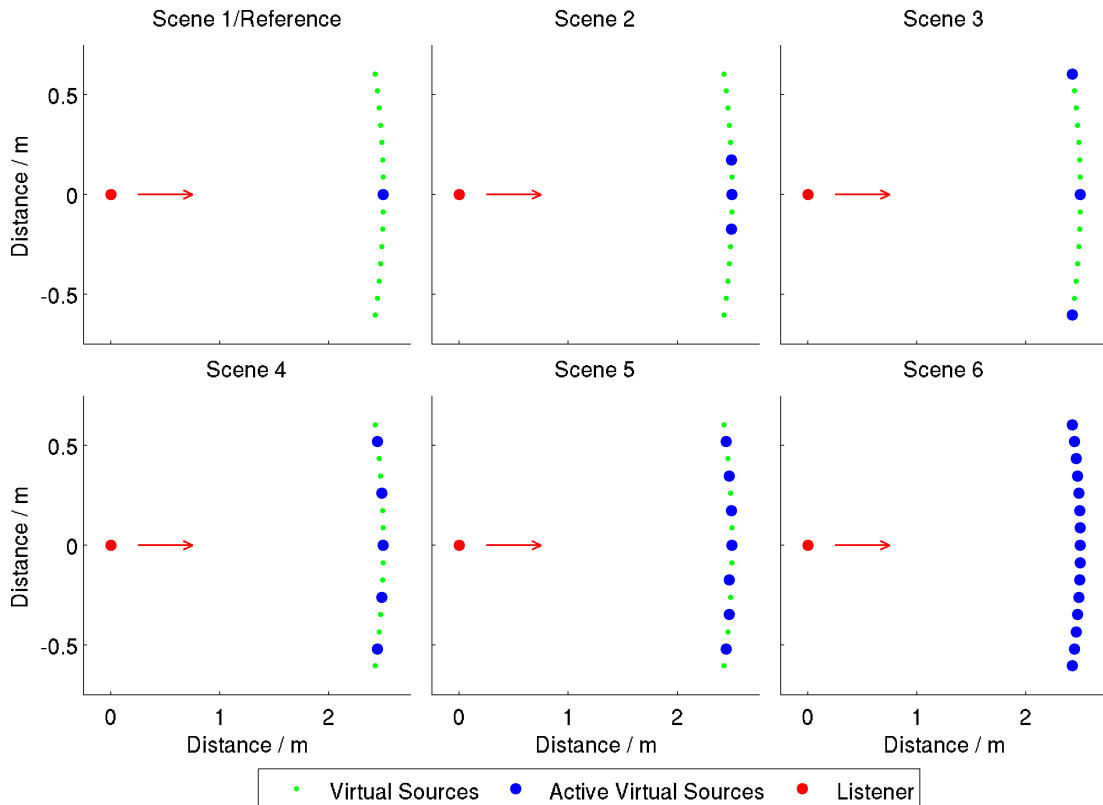


Figure 4.2: Virtual sources used by the first and second method

Out of these the samples classical orchestra, speech and WGN are used to evaluate the auditory source width. They do come with a predefined spatial extent however. A classical orchestra is expected to have a wide spatial extent and speech to be narrow. The spectrum of white Gaussian noise has the same level over all frequencies and is therefore an ideal sample for evaluation.

For analyzing the perceived quality, the samples classical orchestra, speech and castanets are selected. The sound of a classical orchestra and speech is well known to most listeners, castanets feature fast attacks and when the signal is processed it may smear easily. So those samples should provide a good basis for evaluation of the quality after being decorrelated with the implemented methods.

This decorrelation results in the average inter-aural cross coefficients between the generated signals that are displayed in table 4.2, each of the samples was processed by the three implemented methods, the method after Kendall is abbreviated as M1, the method after Bouéri et al. as M2 and M3 stands for critical bands.

The first channel corresponds to the reference for the first two methods, as the third method needs all channels presented simultaneously to reassemble a complete signal, the first channel is not used here. As shown in table 4.2, the IACC for the methods after Kendall is and Bouéri is never close to zero. For the third method the IACC is almost zero. The environment of the experiment is the room “Pinta” in the Quality & Usability Department. It features an array of 56 loudspeakers arranged in a circle with a diameter of 3 m. These speakers can be lowered to be on par with the listener

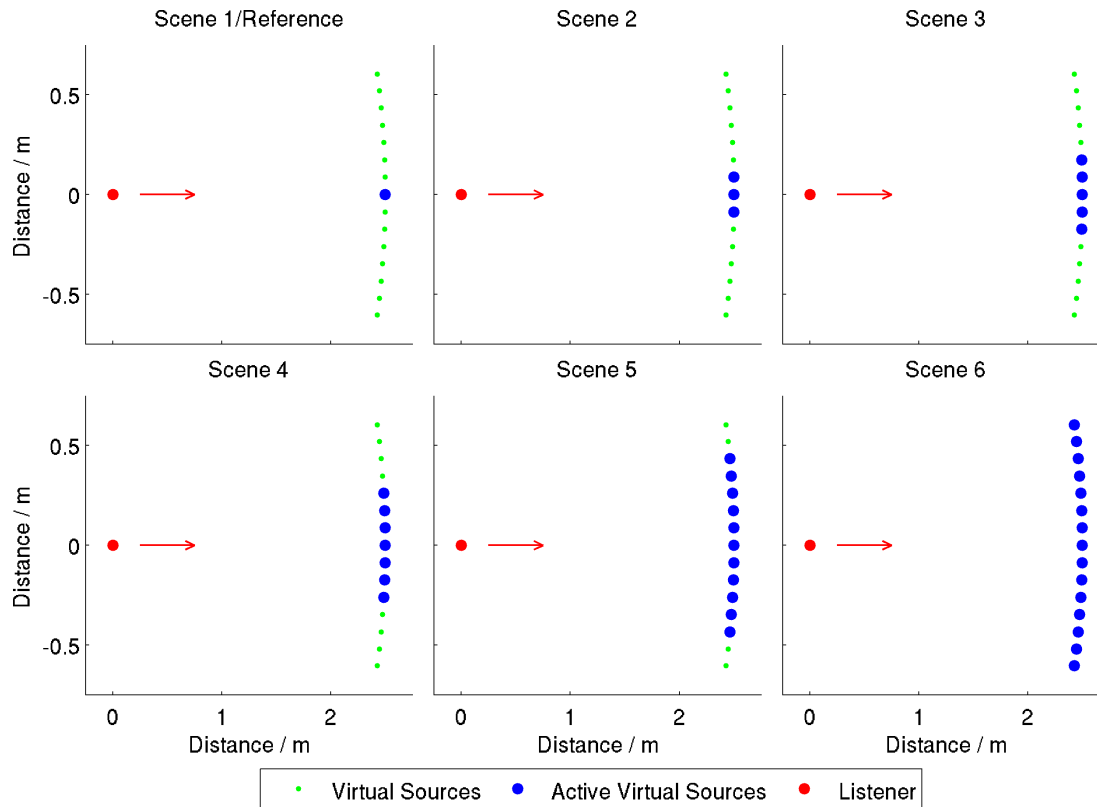


Figure 4.3: Virtual sources used by the “Critical Bands” method

and provide the best possible listening experience. They are driven by the SoundScape Renderer, controlled by PureData.

The experiment is conducted with 25 test subjects, mostly students from the TU Berlin all from the age 20 to 30. They are presented with the 18 sets and supposed to rate the auditory source width and quality.

To enable the test subjects to easily switch between the scenes and evaluate them, there is a simplified graphical interface written in Python. It communicates with PureData via TCP that way it can control which of the predefined scenes is presented to the test subject and switch seamlessly between them when the test subjects wants to listen to the next one. The gui is visible in figure 4.4 showing the evaluation of the auditory source width, the evaluation of quality being similar to this. With this interface the test subjects are able to evaluate the selected attributes of each scene, the auditory source width of the scene and the perceived quality. By clicking the numbers the test subject can choose what scene to listen to in any order, repeat them as often as they wish and set the slider to a level they think is appropriate. Once they listened to all scenes and set all sliders they can continue to the next set.

The results are then saved as comma-separated values for later evaluation.

Channel	Critical Bands	Frequency / Hz	Scene/Virtual Source					
			1	2	3	4	5	6
1	1,2	0 - 125	0	0	0	0	0	0
2	3,4	125 - 211	0	-1	-1	-1	-1	-1
3	5,6	212 - 317	0	1	1	1	1	1
4	7,8,9	317 - 526	0	0	-2	-2	-2	-2
5	10,11,12	526 - 814	0	-1	2	2	2	2
6	13,14,15	814 - 1.21k	0	1	0	-3	-3	-3
7	16,17,18	1.21k - 1.77k	0	0	-1	3	3	3
8	19,20,21	1.77k - 2.52k	0	-1	1	0	-4	-4
9	22,23,24	2.52k - 3.58k	0	1	-2	-1	4	4
10	25,26,27	3.58k - 5.03k	0	0	2	1	-5	-5
11	28,29,30	5.03k - 7.05k	0	-1	0	-2	5	5
12	31,32,33	7.05k - 9.84k	0	1	-1	2	-1	-6
13	34,35	9.84k - 12.23k	0	0	1	-3	1	6
14	36,37	12.23k - 15.28k	0	-1	-2	3	-2	-7
15	38,39	15.28k - 20k	0	1	2	0	2	7

Table 4.1: Mapping of Critical Bands to Virtual Sources

4.2 Software

4.2.1 SoundScape Renderer

For the generation of virtual sources within Wave Field Synthesis the software SoundScape Renderer (SSR) [5] is used. It features real-time rendering of spatial audio for several arbitrary rendering methods.

While direct generation of virtual sources with given input signals for the experiment conducted for this thesis it is necessary to change the input signals within the runtime to present several scenes without any notable delay or transition. This is only possible by using the SSR as generic renderer. In this mode the SSR acts as a multiple-input-multiple-output convolution engine and requires an impulse response for each virtual source, generated with the Sound Field Synthesis Toolbox (SFS) [13]. With these impulse responses the locations of the virtual sources is determined and cannot be changed during run-time. The SSR utilizes these impulse responses and convolve them with the input signals in real-time to present the output signals.

In the experiment PureData [7] is used to control the input signals for SSR so that a seamless change between different scenes is possible.

4.2.2 Matlab/Octave

Matlab [9] is an interactive environment for numerical computing. Its name is derived from Matrix Laboratory and it is commercially developed by MathWorks. Featuring a wide variety of mathematical functions that are implemented in C and Fortran and it is very versatile and highly optimized.

#	Castanets			Orchestra			Speech			WGN		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
1	0.841	0.778	-	0.912	0.593	-	0.898	0.666	-	0.657	0.759	-
2	0.779	0.731	0.031	0.899	0.618	0.111	0.862	0.611	0.071	0.576	0.767	0.015
3	0.835	0.715	0.025	0.820	0.635	0.088	0.856	0.588	0.039	0.501	0.766	0.020
4	0.782	0.749	0.053	0.890	0.500	0.065	0.887	0.503	0.068	0.540	0.728	0.029
5	0.805	0.668	0.020	0.907	0.622	0.111	0.888	0.561	0.023	0.529	0.538	0.027
6	0.770	0.721	0.049	0.907	0.593	0.052	0.890	0.545	0.026	0.535	0.730	0.016
7	0.798	0.745	0.034	0.908	0.541	0.046	0.893	0.646	0.023	0.605	0.671	0.016
8	0.838	0.739	0.048	0.859	0.549	0.048	0.886	0.614	0.017	0.607	0.768	0.017
9	0.768	0.740	0.014	0.909	0.595	0.050	0.851	0.669	0.017	0.484	0.659	0.016
10	0.826	0.726	0.017	0.902	0.469	0.039	0.889	0.569	0.015	0.588	0.739	0.016
11	0.761	0.720	0.045	0.880	0.557	0.034	0.862	0.677	0.021	0.577	0.728	0.017
12	0.845	0.711	0.029	0.910	0.617	0.027	0.862	0.530	0.017	0.585	0.588	0.020
13	0.823	0.768	0.024	0.910	0.557	0.035	0.880	0.658	0.030	0.569	0.762	0.031
14	0.830	0.706	0.036	0.900	0.447	0.053	0.851	0.587	0.036	0.605	0.733	0.043
15	0.804	0.698	0.042	0.892	0.587	0.239	0.883	0.568	0.037	0.591	0.764	0.049
16	0.828	0.730	0.025	0.905	0.597	0.124	0.891	0.683	0.027	0.505	0.715	0.026

Table 4.2: Average IACC for samples decorrelated with the implemented decorrelation methods. (M1: Kendall, M2: Bouéri, M3: Critical Bands)

An alternative to Matlab is GNU Octave [4], mostly compatible to Matlab but open source and freely available under the GNU General Public License (GPL). For this thesis Matlab code was used to implement the decorrelation methods and to visualize the results.

To provide additional functionality to Matlab and Octave three toolboxes were used.

- Sound Field Synthesis Toolbox [13]
- Linear Time/Frequency Analysis Toolbox [11]
- Auditory Modelling Toolbox [10]

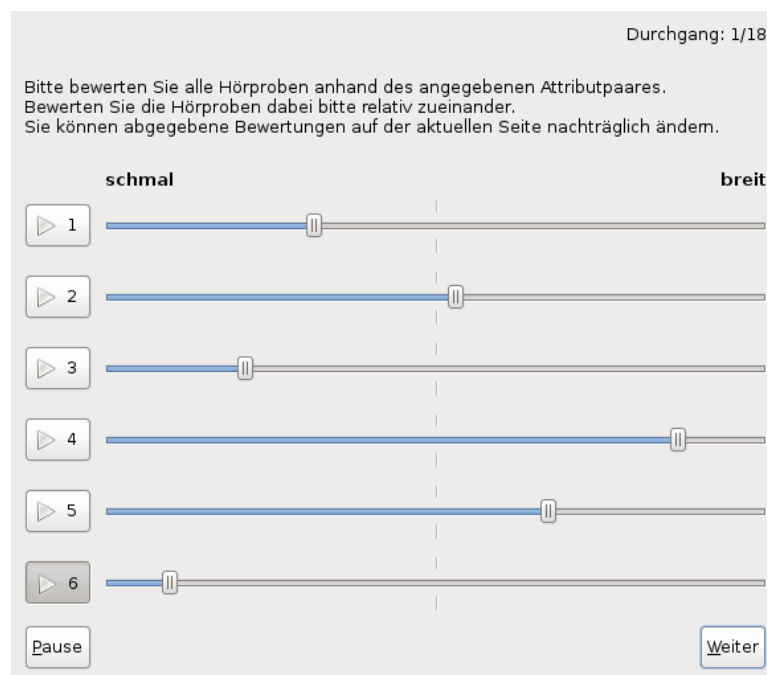


Figure 4.4: Experiment GUI: evaluate auditory source width [narrow, wide]

5 Results

The experiment was conducted with 25 test subjects who evaluated the auditory source width and the sound quality of the three methods implemented. For each method there were three tests for the perceived auditory source width and three for quality.

This chapter contains four plots displaying the results of the experiment. In the first section the results for each scene will be displayed as they are, while in the second section the results for each sample are normalized with the rating of the reference sample.

Each of these plots consists of three subplots, one for each method. Every subplot shows the average rating for each sample and the corresponding 95% confidence interval for each scene. To distinguish between the selected samples, their x coordinates have been modified with a slight offset, this way they appear next to each other in the figure.

The scale ranges from 0 for very narrow/low quality to 1 for a very wide/high quality presentation. Scene 1 is the reference, a single virtual source presenting the unprocessed input signal, the other scenes each use an increasing number of virtual sources.

5.1 Plain Results

This section presents the results of the experiment with two figures containing the ratings regarding auditory source width and signal quality for the three implemented methods.

5.1.1 Auditory Source Width

In figure 5.1 the ratings for auditory source width are shown for all three methods, starting with the first subplot for the method after Gary S. Kendall. The perceived auditory source width rises from scene 1 presenting the reference to the third scene. From there it begins to fall again and the last scene presenting 15 signals on all virtual sources is rated lower than the reference. The rating for all three samples is mostly similar, white Gaussian noise (WGN) is generally perceived with the largest source width.

In the second subplot showing the rating for the second implemented method after Maurice Bouéri and Chris Kyirakakis there is also a slight rise in ratings from the reference to scene 2, except for WGN. The third scene shows comparable results to

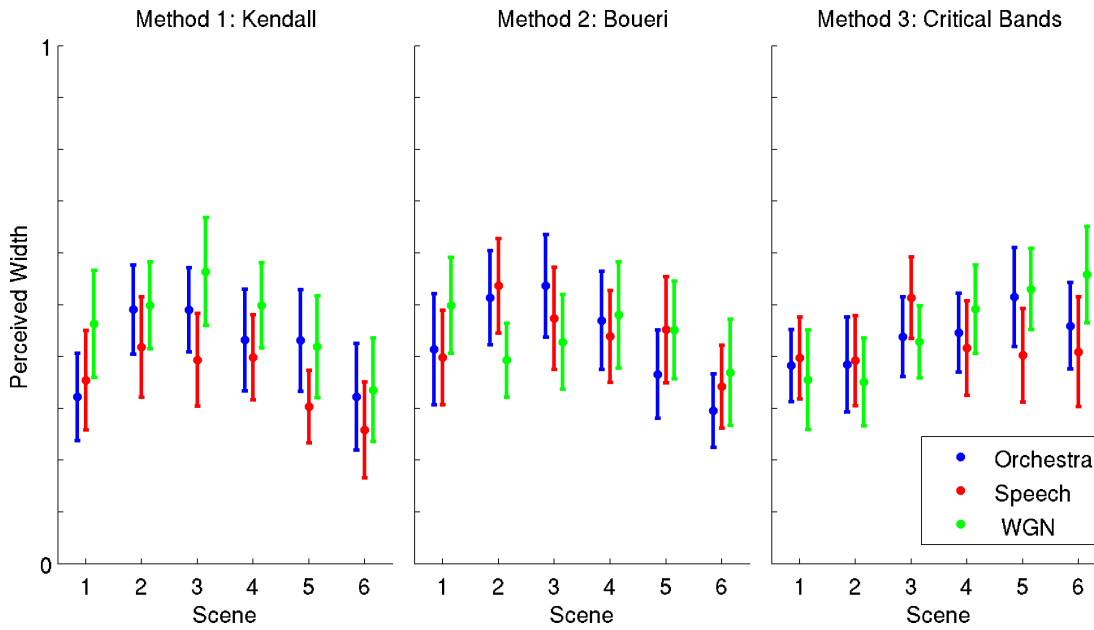


Figure 5.1: Evaluation for the perceived auditory source width

the second one and from there on the ratings fall again. Scene 6 is rated with lower auditory source width than the reference.

In contrast to the first and second method, the ratings in the third method for the classical orchestra rise to a maximum in scene 5, scene 6 is still rated higher than the reference. Speech is about the same for the reference and scene 2, reaches the highest rating at scene 3 and then falls back to a rating just barely above the reference. WGN rises constantly throughout all 6 scenes.

5.1.2 Perceptual Quality

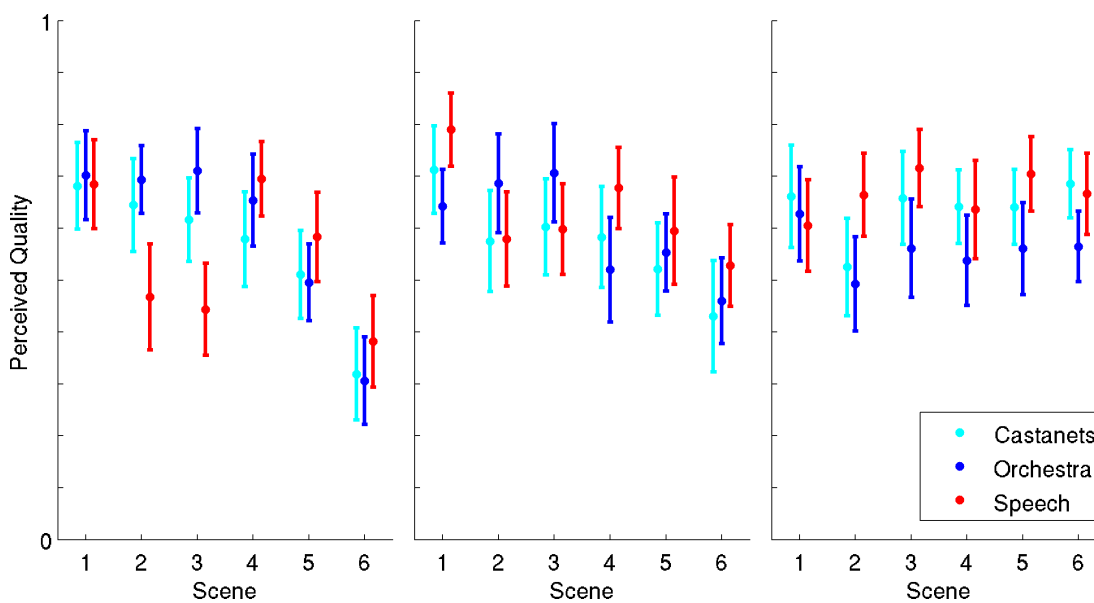


Figure 5.2: Evaluation for the perceived quality

The rating of the perceptual quality of each sample is displayed in figure 5.2. For the first method the three reference signals receive a similar rating, then in the second and third scenes the speech sample is rated far lower than castanets and classical orchestra which only drop a little below the reference. In the scene 4 the fall of castanets and orchestra continues while speech reaches its maximum. In scenes 5 and 6 the rating of all signals continuously decreases, with scene 6 presenting the minimum rating for all samples.

Moving on to the next plot, which shows the quality rating for the second implemented method, the pattern is similar to the first plot. The reference signal for speech is rated higher than the other two. In the second scene castanets and speech fall down a bit while orchestra rises. These ratings do not change significantly for the third scene. Scene 4 increases the rating for speech, castanets fall a little while orchestra falls a lot. This continues to scene 5 and in the last scene all samples receive their lowest rating.

The last method shows very close ratings for castanets, except in scene 2 where the rating is rather low. Orchestra also has the lowest rating in scene 2, and then rises up, but stays below the reference rating. Speech is constantly rated above the reference but contains slight variations.

5.2 Normalized Results

In the previously shown figures the reference samples did not receive an equal rating. In this section the ratings for each sample is calculated based on the rating of the reference sample. This improves the significance of the ratings, as each signal generated is based on the reference. It also reduces the range of the 95% confidence interval by a notable amount.

5.2.1 Auditory Source Width

The evaluation of auditory source width, normalized to the reference is shown in figure 5.3. In the first subplot the ratings for all three signals rise in the second scene, only change slightly in scene 3 and then begin to fall, reaching their minimum in the last scene.

The middle plot shows the results for the second method. Orchestra and speech rise for scenes 2 and 3, while noise is rated lower than the reference. From scene 4 onward they receive a similar rating and fall down to reach their minimum in scene 6.

Shown in the last subplot is the rating for the third method. With this method the rating for noise rises continuously, orchestra has its maximum in scene 5 and speech in scene 2, but all samples never fall below the reference rating.

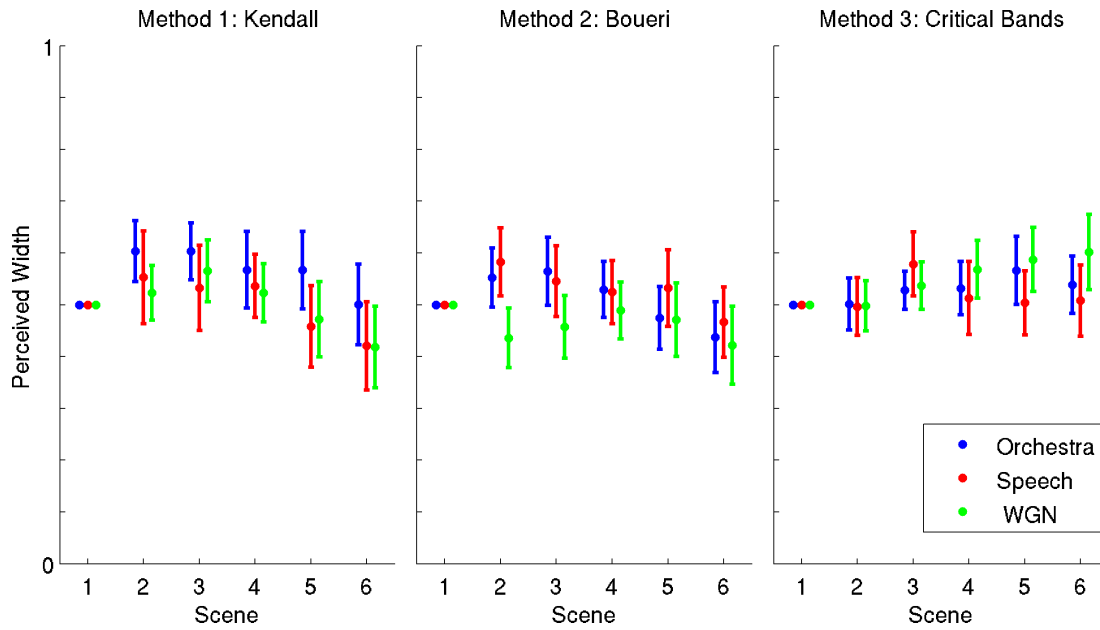


Figure 5.3: Normalized evaluation for the perceived auditory source width

5.2.2 Perceptual Quality

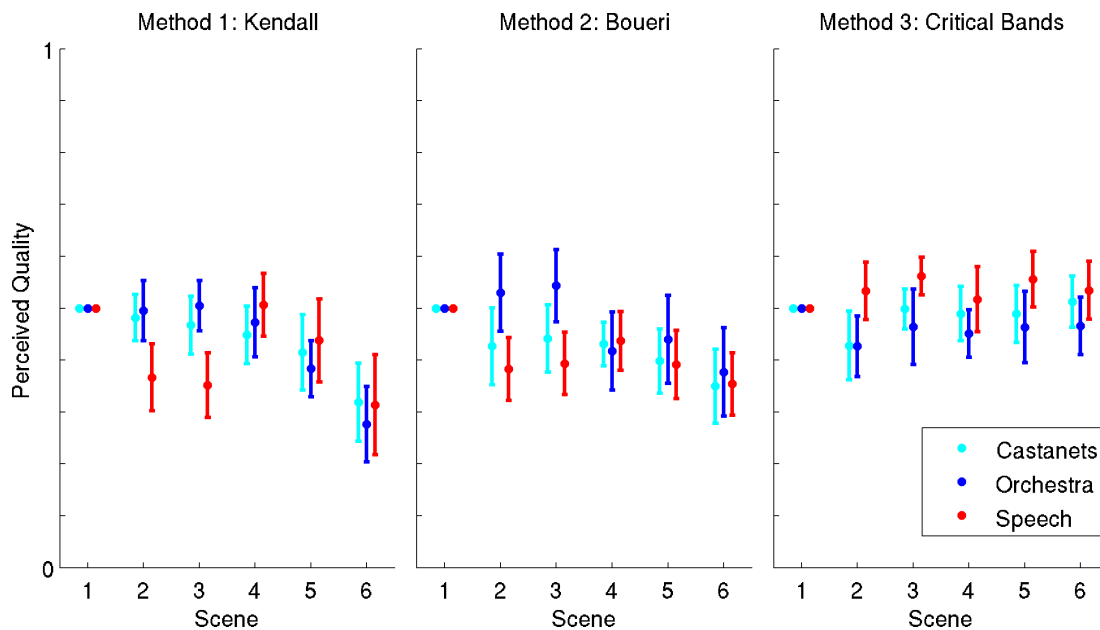


Figure 5.4: Normalized evaluation for the perceived quality

In figure 5.4 the normalized results for the evaluation of signal quality are displayed. The first subplot of this figure shows that for the first method speech was rated considerably lower than castanets and orchestra in scene 2 and 3. Those two are only slightly behind the reference. In Scene 4 speech recovers while castanets and orchestra continue to fall. For scene 5 and 6 this fall continues for all samples, reaching their minimum in the last scene.

For the second method castanets and speech receive low ratings in scene 2 and 3 while

orchestra receives a higher rating than the reference. From scene 4 onward the rating of all samples constantly decreases.

The last method shows low quality for castanets and orchestra in scene 2. The rating recovers and for the following scenes the perceived quality of castanets is almost as high as the reference. Orchestra is continuously rated slightly lower and speech higher than reference.

5.3 Discussion

In figure 5.3 the normalized results for the three implemented methods are shown. When comparing the results for the first two methods there is a clear tendency, for the first scenes the auditory source width does increase. With each scene the number of active virtual sources is increased, once this number reaches a certain point, the auditory source width decreases again.

This is most likely caused by the IACC between the generated signals, as shown in table 4.2, it is in most cases far from zero.

The first two methods were designed for stereo setups, the chance for creating a number of signals with a low IACC to each other is sinking rapidly on an increasing number of signals. In the implementation of those methods there are certain parameters that affect the resulting IACC, like the frequency for the phase change in the method after Gary S. Kendall, but changing of these parameters also causes a change in the quality of the resulting signals. So it is not easily possible to reduce the IACC without also reducing the quality and for evaluation the parameters were adjusted in a way that did not notably reduce the quality of a single signal presented. And as shown in figure 5.4, the quality of the generated signals is already depending on the number of signals presented. For the first two methods, the average quality falls slowly in scenes 2 to 4 and then begins to decrease rapidly, for the third method it does not change significantly.

Additionally the auditory source width does not increase infinitely with each virtual source added, but only rises to a certain peak level before falling again. For the conducted experiment this peak level for the auditory source width in methods 1 and 2 is around 5 to 7 virtual sources.

The third method is not affected by this regression. This method is splitting the signal into its critical bands and between those, the IACC is always close to zero because they only share a very limited frequency spectrum between the bands. As visible in the third subplot of figure 5.3, the average auditory source width rises constantly throughout scene 1 presenting the signals from a single virtual source, to scene 6 presenting each of the 15 signals as a unique virtual source.

6 Conclusion

The goal of this thesis is to increase the auditory source width of signals presented in wave field synthesis. To achieve this, the input signal is processed into multiple signals to be presented on virtual sources in a wave field synthesis setup.

For this procedure three methods were implemented, the first was published by Gary S. Kendall in 1995[8], the second by Maurice Bouéri and Chris Kyirakakis in 2004[3]. These methods were designed for stereo setups and have now been modified to work for wave field synthesis. The third method was loosely based on a paper by Franz Zotter et al. from 2011[14] and instead of generating output signals that are perceptually equal to the input, but still differ in their cross correlation coefficients, the input signal is split into critical bands and each band is presented individually.

With these methods several samples were processed and used in an experiment to verify the expectations and compare the three methods.

The last section presented the results from the experiment without further investigation, this section will go into detail and explain the results, where possible. For this, the normalized results will be used as a base.

In conclusion it can be said that it is possible to present multiple signals on virtual sources that feature an increased auditory source width in a wave field synthesis setup. However the methods for doing this are not yet entirely satisfactory, especially methods 1 and 2 were not able to produce results with an increased auditory source width and consistent quality. The third method proved to work as expected but there is certainly room for improvement.

For example future research could take into account that certain samples, like speech, will not easily be perceived as wide since they naturally originate from a very narrow source. This is also reflected in the results as figure 5.3 shows.

Another possible aspect in the third method that could be evaluated is the arrangement of the critical bands along the virtual sources.

Some signals contain actual information only in a small number of critical bands and with the method used in this study they are presented only on virtual sources that are located close to the center. With an arrangement that aims to increase or even maximize the distance between the active virtual sources, the auditory source width could be notably increased.

Bibliography

- [1] BARRON, M. : The Subjective Effects of First Reflections in Concert Halls - The Need for Lateral Reflections. In: *Journal of Sound and Vibration* 15 (1971), S. 475–494 2.1.1
- [2] BERKHOUT, A. J.: A Holographic Approach to Acoustic Control. In: *Journal of the Audio Engineering Society* 36 (1988), December, S. 977–995 2.4
- [3] BOUÉRI, M. ; KYIRAKAKIS, C. : Audio Signal Decorrelation Based on a Critical Band Approach. In: *Audio Engineering Society* 117 (2004), S. 1–6 1, 2, 2.2, 3, 3.2, 3.3, 6
- [4] EATON, J. W. ; MANY OTHERS: *GNU Octave*. <http://www.gnu.org/software/octave/>. Version: December 2012 2.5, 3, 4.2.2
- [5] GEIER, M. ; SPORS, S. : Conducting Psychoacoustic Experiments with the SoundScape Renderer. (2010) 3, 4.2.1
- [6] HOHMANN, V. : Frequency analysis and synthesis using a Gammatone Filterbank. In: *ACTA Acustica United With Acustica* 88 (2002), S. 433–442 3.2
- [7] IEM INSTITUTE OF ELECTRONIC MUSIC AND ACOUSTICS: *Pure Data*. <http://puredata.info>. Version: December 2012 3, 4.2.1
- [8] KENDALL, G. S.: The Decorrelation of Audio Signals. In: *Computer Music Journal* 19:4 (1995), S. 71–87 1, 2, 2.1.1, 2.1.1, 2.2, 2.1.2, 3, 3.1, 3.2, 3.3, 6
- [9] MATHWORKS: *Matlab - The Language of Technical Computing*. <http://www.mathworks.de/products/matlab/>. Version: December 2012 2.5, 3, 4.2.2
- [10] SØNDERGAARD, P. L. ; CULLING, J. F. ; DAU, T. ; GOFF, N. L. ; JEPSEN, M. L. ; MAJDAK, P. ; WIERSTORF, H. : Towards a binaural modelling toolbox. In: *Proceedings of the Forum Acusticum 2011*, 2011 3.3, 4.2.2
- [11] SØNDERGAARD, P. L. ; TORRÉSANI, B. ; BALAZS, P. : The Linear Time Frequency Analysis Toolbox. In: *International Journal of Wavelets, Multiresolution Analysis and Information Processing* 10 (2012), Nr. 4. <http://dx.doi.org/10.1142/S0219691312500324>. – DOI 10.1142/S0219691312500324 3.2, 4.2.2
- [12] SPORS, S. ; RABENSTEIN, R. ; AHRENS, J. : The Theory of Wave Field Synthesis Revisited. In: *Audio Engineering Society* 124 (2008), S. 1–19 2, 2.4

-
- [13] SPORS, S. ; WIERSTORF, H. : Sound Field Synthesis Toolbox. In: *Audio Engineering Society, Convention e-Brief* 132 (2012) 4.2.1, 4.2.2
- [14] ZOTTER, F. ; FRANK, M. ; MARENTAKIS, G. ; SONTACCHI, A. : Phantom Source Widening With Deterministic Frequency Dependent Time Delays. In: *Int Conference on Digital Audio Effects* (2011) 1, 2, 6
- [15] ZUREK, P. M.: The Precedence Effect and Its Possible Role in the Avoidance of Interaural Ambiguities. In: *Journal of the Acoustical Society of America* 67 (1980), S. 952–964 2.1.1