Christoph Hohnerlein

Coloration of virtual sources in Wave Field Synthesis for different loudspeaker spacings

Bachelor Thesis

Thesis supervisor: Prof. Dr.-Ing. Alexander Raake
Thesis supervisor: Prof. Dr.-Ing. Sebastian Möller
Academic advisor: Hagen Wierstorf

Technical University of Berlin
Faculty IV: Electrical Engineering and Computer Science
Department of Telecommunication Systems
Assessment of IP-based Applications

November 4, 2013

Eidesstattliche Erklärung

Hiermit erkläre ich, Christoph Hohnerlein, dass ich die vorliegende Arbeit, betitelt *"Coloration of virtual sources in Wave Field Synthesis for different loudspeaker spacings"*, zu deutsch *"Verfärbung von virtuellen Quellen der Wellenfeldsynthese für verschiedene Lautsprecherabstände"* selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Unterschrift　　　　Berlin, November 4, 2013

Declaration of Authorship

I, Christoph Hohnerlein, declare that this thesis titled, *"Coloration of virtual sources in Wave Field Synthesis for different loudspeaker spacings"* and the work presented in it are my own. I confirm that:

- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signature　　　　Berlin, November 4, 2013

German summary

Wellenfeldsynthese (WFS) versucht, beliebige Druckverteilungen und damit Hörszenen in einem bestimmten Bereich durch eine feste Anordnung von vielen Lautsprechern zu reproduzieren. Durch die physikalische Größe eines einzelnen Lautsprechers ist die Anzahl in einer gewählten Wiedergabestruktur begrenzt. Dadurch kommt es zu unvorhersehbaren Klangverfärbung, vor allem im oberen Spektralbereich.
Die vorliegende Bachelorarbeit untersucht, inwieweit die durch Begrenzung der Lautsprecheranzahl hervorgerufene unsaubere Rekonstruktion und die damit verbundenen Klangverfärbungen wahrnehmbar sind. Da Klangverfärbung eine subjektive, nicht-quantitive Größe ist, wird ein Hörexperiment durchgeführt um folgende vier konkrete Forschungsfragen zu beantworten:

- Wie wirkt sich die Lautsprecheranzahl auf die Klangverfärbung aus?
- Wie viele Lautsprecher werden benötigt, um weniger Klangverfärbung als ein Stereoaufbau zu erzeugen?
- Wie sehr ist die wahrgenommene Klangverfärbung von dem gewählten Stimulus abhängig?
- Welche Abhängigkeit besteht zwischen Klangverfärbung und Zuhörerposition?

Dazu wird eine kreisförmige Lautsprecheranordnung von drei Meter Durchmesser und Lautsprecherabständen von $67.32cm$ bis $26.3mm$ (was einer Lautsprecheranzahl von 14 bis 3584 Stück entspricht) in verschiedenen Hörerpositionen simuliert und die Klangverfärbungen gegen eine Referenz mit Hilfe des MUSHRA (*MUltiple Stimuli with Hidden Reference and Anchor*) Verfahrens ausgewertet.
Die Versuchspersonen können instantan zwischen der gekennzeichneten Referenz und verschiedenen Lautsprecheranordnungen umschalten und sollen die reine Klangverfärbung auf einer kontinuierlichen Skala von “Sehr großer Unterschied” bis “Kein Unterschied” bewerten.
Die Auswertung der gewonnen Daten zeigt, daß die Klangverfärbung wie erwartet mit ansteigender Lautsprecheranzahl abnimmt. Der Verlauf ist annähernd logarithmisch - bei kleinen Arraygrößen wirken sich Änderungen in der Lautsprecheranzahl stärker aus also bei großen Arrays. Auch mit 3584 Lautsprechern ist keine völlige Transparenz gegenüber der Referenz erreicht. Komplexe Stimuli wie Rauschen oder Musik werden insgesamt als verfärbter wahrgenommen als Sprache. Ausserdem gibt es keinen signifikanten Bewertungsunterschied zwischen dem verwendeten Musik- und Rauschstimulus. Bei mittiger Zuhöhrerposition werden 224 Lautsprecher benötigt, um signifikant weniger Klangverfärbung als ein Stereoaufbau aufzuweisen, außerhalb des sog. ‘Sweet Spots’ schneidet ein WFS Aufbau mit nur 14 Lautsprechern signifikant besser ab.

Abstract

Wave Field Synthesis seeks to recreate arbitrary pressure distribution inside a certain arrangement of loudspeakers. Due to physical limitations, the audible scene created inside the array may exhibit considerable audible artifacts. Aside from the non-linear frequency response of the reproduction chain, these limitations are due to the limited amount of loudspeakers that can fit inside the chosen array. This bachelor thesis seeks to determine the correlation between the amount of speakers in a circular Wave Field Synthesis array and the overall perceived coloration of sound by conducting a listening experiment, in which WFS setups of varying sizes and listening positions are compared to a reference.

It is established that increasing the amount of speaker will reduce the perceived coloration in somewhat logarithmic relation, but won't become transparent against a reference even at 3584 speakers. It is found that the reproduction performance of the WFS array surpasses that of a standard two speaker stereo setup when using more than 224 speakers while standing in the sweet spot; outside the centered position even a 14 speaker array is perceived as less colorated. As assumed, complex stimuli such as music or noise demand a bigger array to achieve the same amount of coloration compared to a band-passed signals such as speech. Surprisingly, there is no significant difference between the rating distribution of the music and the noise stimulus. Lastly, it is found that the dependency of perceived coloration is largerly independant of the location inside the array.

It is concluded that a circular array of 224 speakers should be considered as the best compromise between feasibility and reproduction fidelity as it performs not significantly different from an array with 16 times the speakers under most listening conditions and surpasses stereo in perceived coloration.

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of variance
dB	Decibel
GNU	GNU is not Unix
GUI	Graphical User Interface
HRTF	Head Related Transfer Function
IR	Impulse Response
ITU	International Telecommunication Union
LTI	Linear, Time-Invariant
MUSHRA	MUltiple Stimuli with Hidden Reference and Anchor
OPSI	Optimised Phantom Source Imaging
SFS	Sound Field Synthesis
SSR	SoundScape Renderer
WFS	Wave Field Synthesis

1 Introduction

Wave Field Synthesis (WFS) aims to recreate an arbitrary pressure distribution inside a certain arrangement of loudspeaker. It was first proposed by Prof. Berkhout of the TU Delft in 1988 and is based on the idea of breaking up an arbitrary wave front into the superposition of many small elemental wave fronts. This may be recreated by using a horizontal line array, realized either as specialized plane radiator arrays or more commonly as a array of conventional loudspeaker.[1] A more in depth explanation of the technical background is given in section 2.3.

1.1 Status of current research

There are quite a few usable loudspeaker arrays available already: The auditorium *H104* at TU Berlin is fitted with a custom, 832 channel WFS array of 86 meters in length. A total of 2704 loudspeakers have been fitted, with channels spaced at 10 cm.[2] In 2008, the renowned Berlin-based club for electronic music *Tresor* installed a large scale WFS system containing over 800 loudspeakers driven by six audio servers.[3] But still, the concept of Wave Field Synthesis is not yet ready for use in the mass market; the WFS sound system at the Tresor club has since been removed.

Currently, a lot of work is going into finding a suitable compromise between technical feasibility and the auditory demands in competition with the high definition stereo audio, which is widely available. Other configurations such as the "OPSI" (Optimised Phantom Source Imaging) method, suggested by Helmut Wittek, are designed to reduce the number of speakers in a WFS configuration while improving sound reproduction. The OPSI system combines the qualities of WFS and stereophonic reproduction by introducing additional stereophonic phantom sources to playback material above a certain aliasing frequency.[4]

The experiment conducted during this thesis builds on two pieces of open source software, currently being developed at the *Deutsche Telekom Laboratories, Technische Universität Berlin*: The **Sound Field Synthesis Toolbox (SFS toolbox)** was developed by Hagen Wierstorf and Sascha Spors to numerically solve various higher order audio synthesis simulations, including WFS.[5] Secondly, the **SoundScape Renderer** by Matthias Geier, Jens Ahrens and Sascha Spors allows spatial

[1] E.g. Görne (2008), p. 299
[2] E.g. Goertz et al. (2007)
[3] E.g. Neale Lytollis (2008)
[4] E.g. Wittek (2007)
[5] E.g. Wierstorf and Spors (2012)

audio reproduction using a variety of rendering algorithms including WFS and is used for convoluting audio channels with previously generated impulse response in real time.[6] Both of these tools have been very helpful in conducting this experiment.

1.2 Purpose and aim of the conducted experiment

Due to the low amount of complexity and monetary investment, the basic two channel stereo setup has been the go-to audio reproduction system since the 1950's.[7] Forming an equilateral triangle with the two speakers and the listener results in a small sweet spot that allows for the creation of a basic auditory scene using techniques like panning (difference in level) or delay time stereophony.[8]

Even in a high quality stereo setup, large deviations from the ideal flat frequency response are to be expected due to phase cancellation and other effects. These alterations are easily measured but hardly noticed by the listener, especially since a reproduction system is regarded as 'non-colored' after a short amount of time.[9] Adding more speakers, like with the recently popular multi channel surround speaker setups, will add to the problem. Contrary to this, the mathematical description of the WFS theory predicts perfect reproduction properties by using a perfect line array, essentially a infinity amount of infinitesimally small speakers.

This thesis examines the effect of using real loudspeaker with spacings from $67.32cm$ to $26.3mm$, which in the results in 14 to 3584 speakers for a 3 meter array.

Four research questions are being investigated:

- How does the number of speakers used in a WFS array affect the amount of perceived coloration?
- How many speakers are needed to achieve a similar amount of coloration as a stereo setup?
- How does the position of the listener effect the amount of coloration?
- How does the position of the listener effect the amount of coloration?

[6] E.g. Ahrens et al. (2008)
[7] E.g. Snow (1955)
[8] E.g. Snow (1955)
[9] E.g. Fastl and Zwicker (2007)

1.3 Structure of this thesis

In section 1 (page 1), I give a short overview of the history and current situation in the field of wave field synthesis, while section 2 (page 4) contains the theoretical background required to understand the following thesis. Section 3 (page 11) describes the experiment conducted, where first the research questions stated above are refined (section 3.1, page 12) and then the experiment is explained in detail (section 3.2, page 15). The findings are presented in section 4 (page 24), where section 4.1 (page 24) gives an overview of the recorded results and 4.2 (page 28) answers the specific research questions stated in section 3.1. Last, a summary and outlook to further studies is given in section 5 (page 41). Besides additional tables and figures, the appendix (page 45) includes a listing of all digital content on the enclosed CD.

2 Theory

2.1 Sound reproduction

Electroacoustic sound reproduction systems seek to recreate a previous recorded auditory scene to a listener. Reproduction of sound features such as volume, pitch, directivity and quality or coloration respectively (see section 2.2) are crucial but overall fidelity of reproduction needs to be balanced with general feasibility and price.

Sound reproduction systems usually involve one of two types of electroacoustic transducer: Changes in sound pressure reach the listeners ear either through a pair of headphone or through a number of loudspeaker.
Headphone system can be either monaural (a single microphone recording for only one ear), diotic (a single recording for both ears) or binaural (two separate recordings for both ears). Of these three, only the binaural setup can reproduce static directional effects. All headphone recordings eliminate effects of the listening room.[10]

Using conventional speakers, a system can be either monophonic (a single recording for one loudspeaker), stereophonic (two or more separate recordings for the same number of loudspeaker) or multi-speaker (arbitrary amount of recordings for more speakers). All loudspeaker systems with more than one speakers allow the listener to face a virtual source. Compared to headphone systems, the effects of the listening room are going to affect the perceived auditory scene. Furthermore, harmonal distortion may appear through addition or interference/cancellation of frequencies of overlaying audio signals from different speakers.[11]

The reproduction systems investigated in this thesis are wave field synthesis arrays (see section 2.3) - a special type of multi-speaker system, simulated using a binaural headphone setup. For this purpose, it sufficient to be familiar with the use of impulse response in the domain of signal processing. An **impulse response (IR)** simply consists of a system's response to a singular impulse, called Dirac impulse. Ideally, this impulse is infinitely sharp at time $t(0)$, has no temporal spread (values at every other point in time are 0) and the area equals 1, as per equation 1.

$$\delta(t) = \begin{cases} +\infty, & t = 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

[10] E.g. Snow (1955), p 44
[11] E.g. Snow (1955), p 51

An impulse response can easily be transferred to the frequency domain by applying the fourier transform, resulting in the **frequency response** or transfer function of a system.
If a system is **linear and time-invariant (LTI system)**, a recorded impulse response of that system will contain all necessary information to reproduce this system and apply it to any signal. This is done by convoluting the signal with the impulse response of a system, which applies all the system's properties to the signal. This is useful to recreate certain rooms or listening conditions. For example, convoluting a piece of music with the impulse response of a church will make the piece appear as being played back in the church.
A special case of impulse responses are the so-called **head-related transfer functions (HRTF)**, which contain the information of the transmission system from a speaker to the inside of a human ear. These are usually recorded using a mannequin with microphones fitted inside both ear canals and can be used to simulate the relation of a speaker to the human head.

This process is used to simulate all arrays and listening conditions mentioned in this thesis. To generate a simulated array, the impulse response of the array containing the overlaid transfer functions of all the speakers is calculated and convolved with the source signal. The signal is then convolved once more with a head-related transfer function, resulting in signal that contains both the reproduction features of the array and the relation of the listener to that array.

2.2 Sound color

Because of effects like reflections in the listening room or frequency addition and cancellation of a multi-speaker system (see 2.1), the signals reaching the listeners inner ear may exhibit ripples throughout their frequency spectrum.[12] Even though these modulations might be relatively small, they may still lead to a perceivable change in timbre.[13] These changes from the original source can be summarized as the coloration of a sound.

[12] E.g. Rubak and Johansen (2003)
[13] E.g. Fastl and Zwicker (2007), p 142

These features of a sound are to be separated from the loudness (intensity of cochlea stimulation) and the pitch (location of cochlea stimulation) of a sound, as the American Standards Association (*ASA*, now American National Standards Institute, *ANSI*), suggests:

> ***Timbre*** *is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.*
> *Note: Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the wave form, the sound pressure, and the frequency location of the spectrum of the stimulus.*
>
> ***Pitch*** *is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high, such as a musical scale.*
> *Note 1: Pitch depends primarily upon the frequency of the sound stimulus, but it also depends upon the sound pressure and wave form of the stimulus.*
> *Note 2: The pitch of a sound may be described by the frequency of that simple tone, having a specified sound pressure or loudness level, which seems to the average normal ear to produce the same pitch.*
>
> ***Loudness*** *is the intensive attribute of an auditory sensation, in terms of which sounds may be ordered on a scale extending from soft to loud.*
> *Note: Loudness depends primarily upon the sound pressure of the stimulus, but it also depends upon the frequency and wave form of the stimulus.*

Association and of America (1960) - **American standard acoustical terminology**, p 22 & 25

This makes pinpointing the positive definition of the term coloration very difficult, as it tends to be become a pool for all sound qualities excluding pitch and volume. Furthermore, there is no consensus on whether or not directivity of a sound should be included in the term coloration.

Because coloration cannot be compared quantitatively, all measurements in this thesis are recorded subjectively on a continuous scale while changes in sound features of volume, pitch and directivity are avoided.

2.3 Wave Field Synthesis

Dutch mathematician Christian Huygens postulated the idea of an elemental wave, from which any arbitrary wave distribution can be created through superposition.[14] Following this idea, a linear or circular wave front can also originate from a distribution of sources. This is symbolically shown for a circular array of speakers in figure 1, where a concentric wave front originating from a virtual source (black speaker symbol) is reproduced somewhat inadequately using 16 secondary sources (red dots).

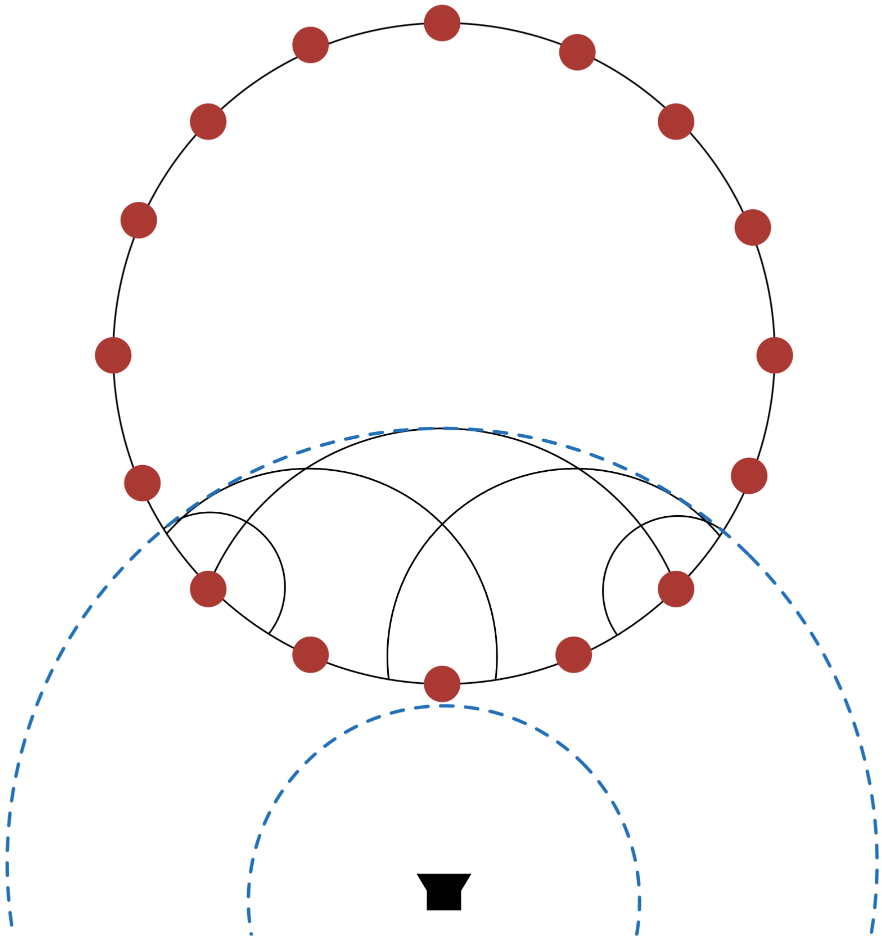


Figure 1 – Schematic of a circular WFS array, reproducing a concentric wave originating from a virtual source outside the array (black speaker) using 16 secondary speakers (red dots)

Please note that the following summary of the theoretical background of Wave Field Synthesis is very short as its full consideration is not within the scope of this thesis. For a more in-depth understanding, please refer to Spors et al. (2008) - The Theory of Wave Field Synthesis Revisited.

[14] E.g. Görne (2008), p 44

Generally speaking, a continuous loudspeaker distribution surrounding the listener can be regarded as an in-homogeneous boundary condition for the wave equation, resulting in the Kirchhoff-Helmholtz integral that describes the solution of the homogeneous wave equation for a bounded region V with respect to in-homogeneous boundary conditions:

$$P(\mathbf{x},\omega) = -\oint_{\partial V} \left(G(\mathbf{x}|\mathbf{x}_0,\omega)\frac{\partial}{\partial \mathbf{n}}P(\mathbf{x}_0,\omega) - P(\mathbf{x}_0,\omega)\frac{\partial}{\partial \mathbf{n}}G(\mathbf{x}|\mathbf{x}_0,\omega) \right) dS_0 \quad (2)$$

Here, $P(\mathbf{x},\omega)$ denotes the pressure field inside a bounded region V enclosed by the boundary $\partial V(\mathbf{x} \in V)$, $G(\mathbf{x}|\mathbf{x}_0,\omega)$ a suitable chosen Green's function, $P(\mathbf{x}_0,\omega)$ the acoustic pressure at the boundary $\partial V(\mathbf{x}_0 \in \partial V)$ and $\mathbf{n}$ the inward pointing normal vector of ∂V . The abbreviation $\frac{\partial}{\partial \mathbf{n}}$ denotes the directional gradient in direction of the normal vector $\mathbf{n}$.
This contour integral can now be simplified into a plane by applying the Sommerfeld radiation condition, resulting in the first Rayleigh integral:

$$P(\mathbf{x},\omega) = -\int\int_{-\infty}^{\infty} D_{3D}(\mathbf{x}_0,\omega)G_{0,3D}(\mathbf{x}|\mathbf{x}_0,\omega)dx_0dz_0 \quad (3)$$

Here, the three dimensional driving function $D_{3D}(\mathbf{x}_0,\omega)$ is determined by the direction gradient of the wave field of a plane wave and the window function as

$$D_{pw,3D}(\mathbf{x}_0,\omega) = -2a_{pw}(\mathbf{x}_0)\frac{\mathbf{n}_{pw}^T\mathbf{n}(\mathbf{x}_0)}{c}j\omega\hat{S}_{pw}(\omega)e^{-j\frac{\omega}{c}\mathbf{n}_{pw}^T\mathbf{x}_0} \quad (4)$$

and the three-dimensional free-field Green's function is given as

$$G_{0,3D}(\mathbf{x}|\mathbf{x}_0,\omega) = \frac{1}{4\pi}\frac{e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_S|} \quad (5)$$

For reasons of feasibility and cost, only a two-dimensional planar array is considered instead of a full three-dimensional sphere. This limits the WFS system to the reproduction of pressure distribution inside the plane of which they span. Therefore, the radiation may be described by the two-dimensional free-field Green's function

$$G_{2D}(\mathbf{x}|\mathbf{x}_0,\omega) = \frac{j}{4}H_0^{(2)}(\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|) \quad (6)$$

where $H_0^{(2)}$ is the zeroth-order Hankel function of second kind.

Furthermore, a problem arises as a result of the mismatch of dimensions when including the properties of real speakers instead of ideal continuous speaker distributions: While the ideal WFS line array radiates only in the two dimensional plane of the two-dimensional free-field Green's function, normal cabinet loudspeakers can be estimated as acoustic point sources emitting omnidirectional in a three dimensional sphere. To account for this mismatch, the two-dimensional Green's function $G_{2D}(\mathbf{x}|\mathbf{x}_0,\omega)$ is approximated as follows:

$$G_{2D}(\mathbf{x}|\mathbf{x}_0,\omega) \approx \sqrt{\frac{2\pi|\mathbf{x}-\mathbf{x}_0|}{j\frac{\omega}{c}}}\underbrace{\frac{1}{4\pi}\frac{e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|}}_{G_{3D}(\mathbf{x}|\mathbf{x}_0,\omega)} \tag{7}$$

This correction results in the following pressure distribution:

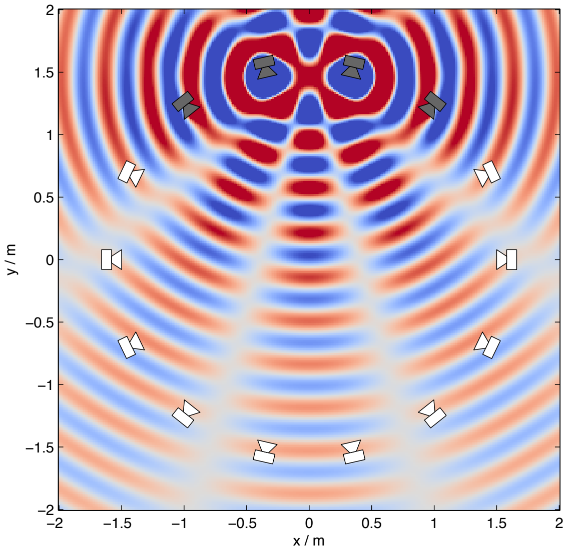
$$P(\mathbf{x},\omega) = -\oint_{\partial V}\sqrt{\frac{2\pi|\mathbf{x}-\mathbf{x}_0}{j\frac{\omega}{c}}}D_{2D}(\mathbf{x}_0,\omega)G_{3D}(\mathbf{x}|\mathbf{x}_0,\omega)dS_0 \tag{8}$$

The correctional term is usually directly applied to the driving function $D_{2D}(\mathbf{x}_0,\omega)$, resulting in the modified driving function:

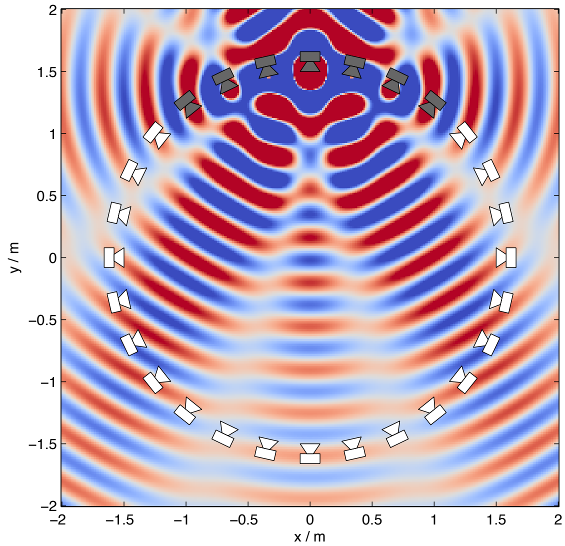
$$D_{2.5D}(\mathbf{x}_0,\omega) = \sqrt{\frac{1}{j\frac{\omega}{c}}}\sqrt{2\pi|\mathbf{x}_{ref}-\mathbf{x}_0|}D_{2D}(\mathbf{x}_0,\omega) \tag{9}$$

$$= -\sqrt{\frac{1}{j\frac{\omega}{c}}}\sqrt{2\pi|\mathbf{x}_{ref}-\mathbf{x}_0|}2a_{pw}(\mathbf{x}_0)\frac{\mathbf{n}_{pw}^T\mathbf{n}(\mathbf{x}_0)}{c}j\omega\hat{S}_{pw}(\omega)e^{-j\frac{\omega}{c}\mathbf{n}_{pw}^T\mathbf{x}_0} \tag{10}$$

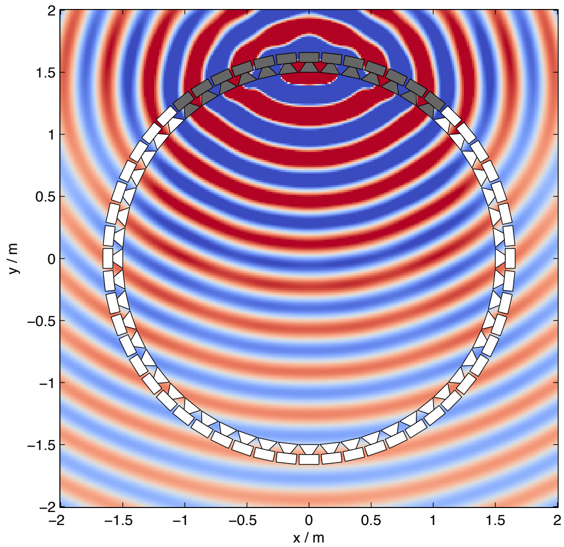
As a more familiar analogy, a WFS array with a discrete number of speakers may be regarded as a sampling device of the original signal. Interestingly, a WFS array exhibits similar properties such as a sampling frequency which marks the upper limit of faithful reproduction.
Figure 3 visualizes this by plotting the frequency responses of various WFS arrays with a seemingly different limiting frequency. Further aliasing effects appear due to the addition of time delayed copies of the same signal, resulting in comb-filter like interference patterns. These can be observed in figure 2, which simulates a three meter circular WFS array with 14, 28, 56 and 112 speakers reproducing a $1kHz$ point source 1 meter outside the array.



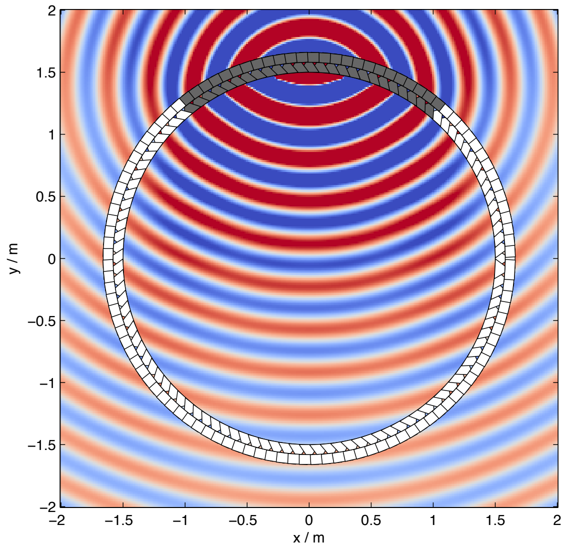
(a) WFS array with 14 speaker



(b) WFS array with 28 speakers



(c) WFS array with 56 speakers



(d) WFS array with 112 speakers

Figure 2 – Virtual WFS arrays of three meter diameter reproducing a point source playing a $1kHz$ sine wave, 1 meter outside the array.
Gray symbols represent active speakers.

3 Experiment: Coloration of Wave Field Synthesis

To illustrate the problem of coloration, the frequency responses of various simulated WFS arrays are depicted in figure 3. It depicts the frequency responses of WFS arrays with a 3 meter diameter and varying number of speakers. It was generated in a centered listening position. A dummy impulse response containing a flat frequency spectrum was used to show the effects of the WFS array only, without any overlayed spectrum of a directional HRTF. This also means that the spatial dimension of the human head have been neglected, since they are small compared to the spatial aliasing of the WFS array. In this exemplary case, both ears practically lay over each other in a central position.

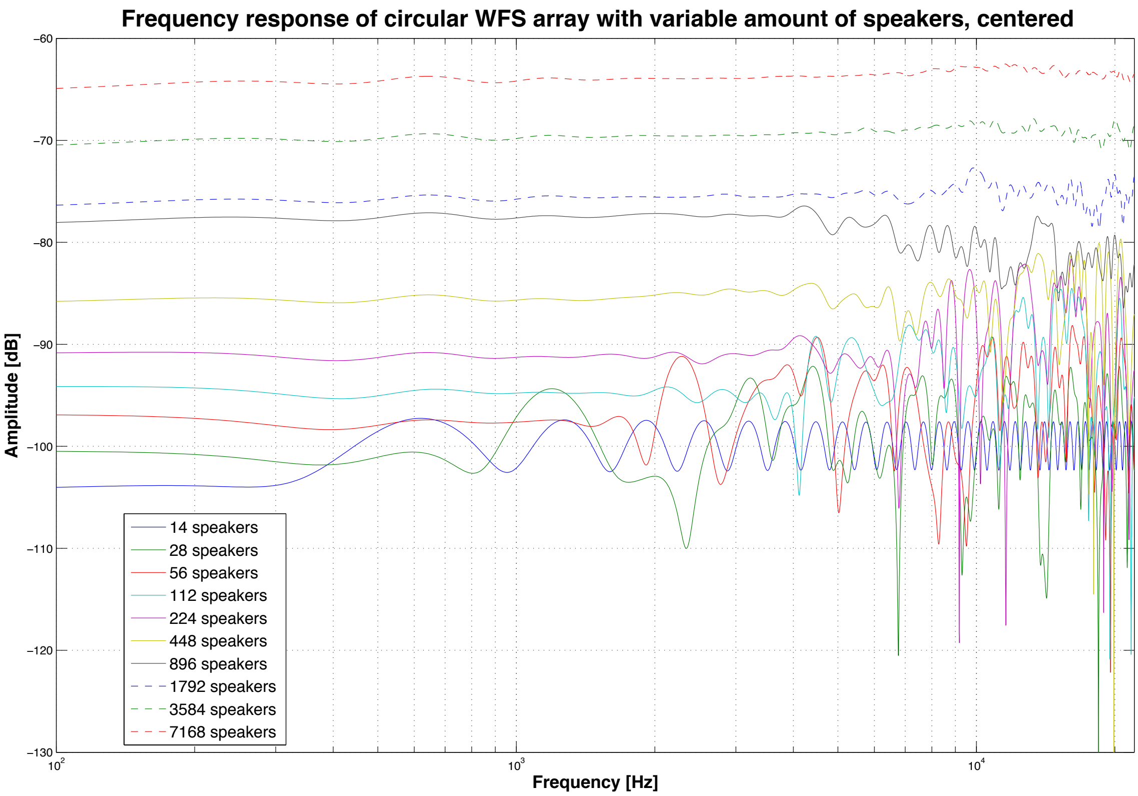


Figure 3 – Frequency responses of a 3 meter circular WFS array with $14 \cdot 2^i, i \in [0..8]$ speakers, centered listening position

It seems clear that increasing the amount of speaker smoothens the overall frequency response of the WFS system. While this does not directly imply a better perceived quality overall, WFS arrays with more speakers should be able to reproduce material closer to the original input. Only arrays with 1792 speakers or higher perform reasonably well up until $5kHz$, which marks the upper limit of the especially sensitive region of the human hearing system (between $2kHz$ and $5kHz$).[15]

[15] E.g. Fastl and Zwicker (2007) Page 204

3.1 Goals of the experiment and research questions

The reason to conduct the following experiment were threefold: Most importantly, the general amount of perceived coloration in correlation to the amount of speakers used in a WFS setup is to be examined. Secondly, the reproduction properties of WFS systems are to be compared to a stereo array. Lastly, the variation of the perceived coloration in different locations inside the WFS array was investigate. The following sections will outline each particular issue and refine the four specific research question stated in section 1.2.

3.1.1 Coloration as a function of speaker number

As stated above, increasing the amount of speakers in a WFS array results in a smoother frequency response of the system. This experiment should now determine how much these changes in frequency response alter the amount of perceived coloration.

All frequency responses used in the experiment have been normalized for two reasons: First and foremost, inexperienced subjects may find it difficult to assess changes in coloration without considering changes in loudness. Effectively removing differences in loudness should improve the precision of the results. Furthermore, louder signals might be at a disadvantage compared to quieter ones, where some frequency components that contain coloration might fall below the hearing threshold and become inaudible[16]. Although normalizing the impulse responses in volume is not the same as adjusting the overall loudness, it is assumed that this approximation is good enough for the purposes of this experiment.

Research question 1 How does the number of speakers used in a WFS array affect the amount of perceived coloration?

3.1.2 Coloration of WFS compared to stereo

Due to price of acquisition and maintenance, availability and convenience, conventional two speaker stereo setups are still the most common use of audio consumption at home.[17], even though the frequency response exhibits large non-linearities.

[16] E.g. Fastl and Zwicker (2007), Page 204
[17] E.g. Snow (1955)

This may be seen in figure 4, which depicts the frequency response of a two speaker stereo setup at a distance of 2,5 meters when sitting in the sweet spot. The *ideal* frequency response is measured right at the center, with no influence of the room or cross-talk between the ears. The *real* frequency response shows an approximation of the real spectrum, as it is the overlay of four transfer functions: The signal of both speakers reach both ears, which are $10cm$ to each side of the center. Effects of the room have been excluded.
While the ideal response shows no signs of alteration whatsoever, the spatial distance of both ears leads to a severe comb-filter like distortion of the flat spectrum.

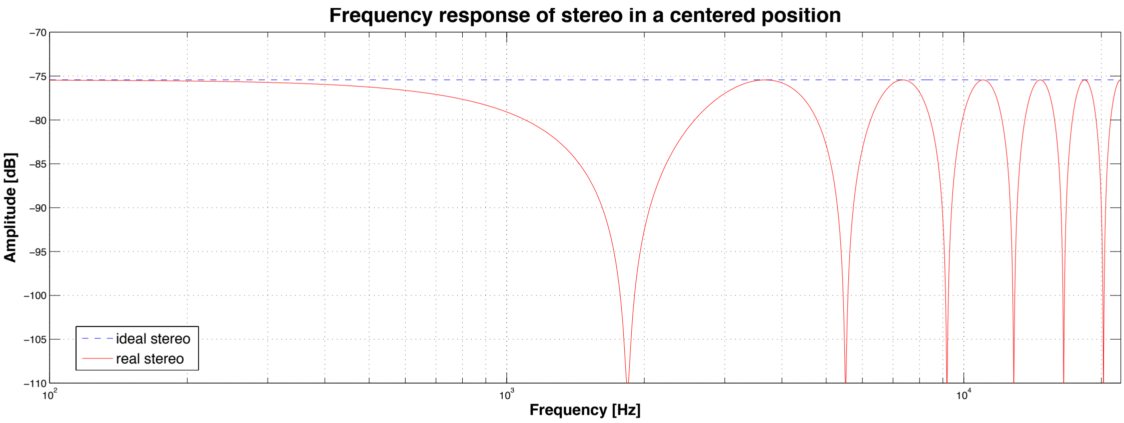


Figure 4 – Frequency spectrum of two speaker stereo setup at $2.5m$ distance in centered listening position. The real stereo response is created using the superposition of two ears located $0.1m$ to each side of the listening position.

Figure 5 depicts a similar setting with the listening position moved 1 meter to the left. This further reinforces the problem. While even the ideal spectrum shows strong com-filter overlay, the approximated real frequency response exhibits strong non-linear alterations.

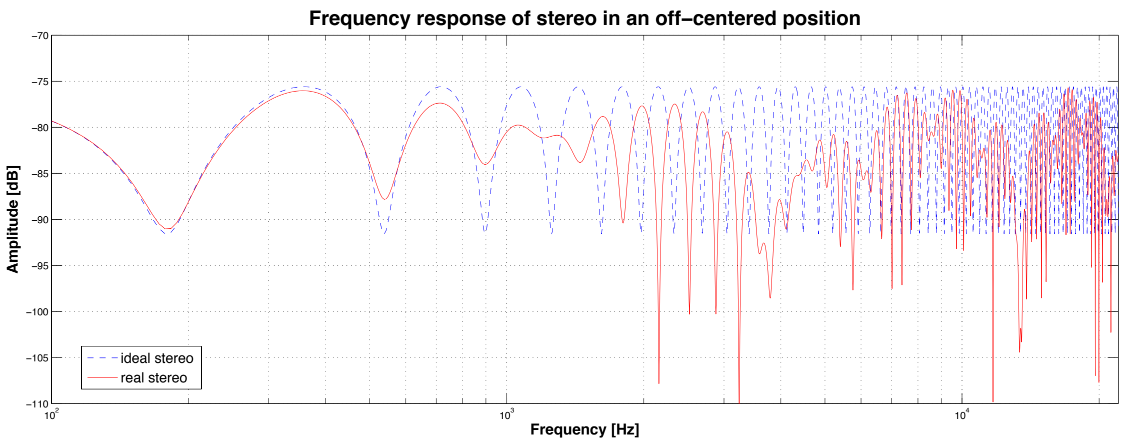


Figure 5 – Frequency spectrum of two speaker stereo setup at $2.5m$ distance, 1 meter left of the center. The real stereo response is created using the superposition of two ears located $0.1m$ to each side of the listening position.

Although more coloration does not automatically equate worse quality, surpassing the amount of coloration of a traditional stereo setup should be a reasonable goal of Wave Field Synthesis.

Research question 2 How many speakers are needed to achieve a similar amount of coloration as a stereo setup?

3.1.3 Effects of stimulus choice on perceived coloration

Different applications have demand different fidelity of reproduction - band-passed signals like speech might not be affected by spatial effects as much due to the little information carried in higher frequency bands. For example, this can be seen in figure 7, where the frequency spectra of all stimuli used in this experiment are depicted. On the other hand, complex piece of music or pure noise signals might require much more speakers to be faithfully reproduced.

Research question 3 How much depends the perceived amount of coloration on the stimulus?

3.1.4 Coloration as a function of listener location

While the ideal WFS setup is completely independent of listening position, lowering the amount of speakers will introduce comb-filter like spectral aliasing due to the modified driver function, as described in section 2.3. This might change the perceived coloration depending on the location of the listener inside the array. As an example, figure 6 depicts the frequency responses of ten circular WFS arrays with a diameter of 3 meter in listening position 1 meter to the left of the center and a varying number of speakers. Although direct comparison to figure 3, which shows the same arrays in a central listening position, is not straightforward, major changes can be observed, especially in the higher frequency regions.

Research question 4 How does the position of the listener effect the amount of coloration?

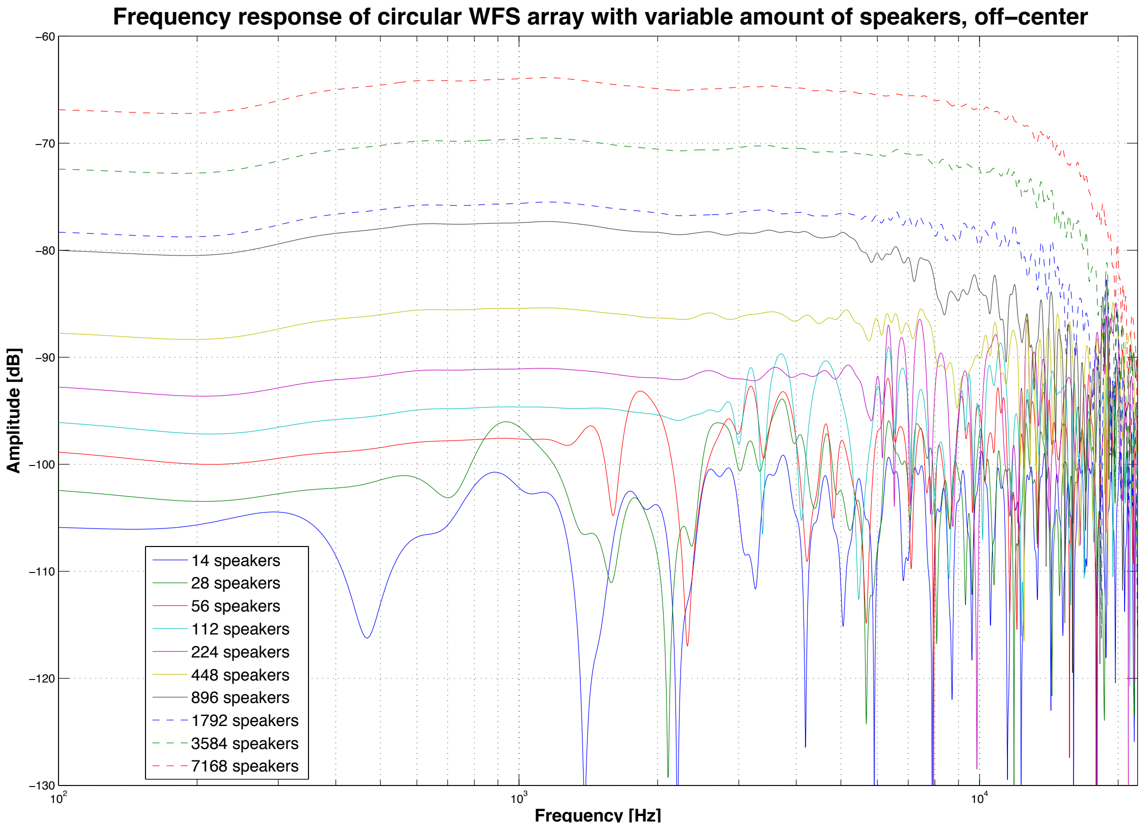


Figure 6 – Frequency responses of a 3 meter circular WFS array with $14 \cdot 2^i, i \in [0..8]$ speakers, listening position 1 meter off-center

3.2 Experiment setup

To answer the research questions stated in sections 1.2 and 3.1, the following experiment was designed.

3.2.1 Methodology

Test subjects blindly rate WFS arrays with a varying amount of speakers against a point source reference on a scale from "no difference" to "very different". The stimuli used include looped audio signals of a speech sample, two bars of piece of electronic music and a impulse train of pink noise.

For practical reasons, all listening conditions are simulated on headphones using impulse responses created beforehand using the SFS toolbox. This way, practically unlimited amounts of speaker can be simulated and seamlessly switched. Furthermore, it was decided to use static HRTFs (Head Related Transfer Functions) to reduce to complexity of the presented listening scene. This means virtual sources do not appear fixed but instead move with the participants head. As the subjects were sitting in a somewhat stationary position, this was not deemed to be an important issue.

Test method To ensure that small differences in sound coloration can be accounted for, a MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) test setup was implemented. It is recommended by the *International Telecommunication Union - Radiocommunication Sector (ITU-R)* in *BS.1534: Method for the subjective assessment of intermediate quality levels of coding systems* because it fulfills the requirement for an absolute scale, comparison with a reference signal and small confidence intervals with a reasonable number of subjects at the same time and is therefore suitable for evaluation of intermediate audio quality and gives accurate and reliable results.[18]

To follow these recommendations, the graphical user interface (GUI) depicted in Figure 11 was created using python. Every page (Figure 11 shows the training page) contains one particular stimulus (speech, music or noise). Clicking any of the play buttons to the left immediately starts the playback of a particular setup. Once a signal other than the references has been selected, the corresponding slider to the left becomes active and may be placed anywhere on the scale between "Very different" and "No difference". Selecting a different condition changes the playback immediately to ensure maximal acoustic memory.

Once all 12 signals have been rated, the subject may continue to the next page containing a new stimulus. After all three stimuli have been rated, the subject may take a short break while the examiner exchanges the set of impulse responses.

Conditions Three different listening conditions were simulated, all of which incorporate the same basic circular array layout described further below. The first condition is run twice and averaged per test subject to reduce the effect of outliers.

- **Subject centered** Position of the virtual listener is centered inside the circular WFS array, amount of speakers is varied. See figure 9a through 9d.
- **Subject off-center** Position of the virtual listener is moved one meter to the left of the center, amount of speakers is varied. See figure 9e.
- **Subject moving** Position of the virtual listener is moved throughout the array, amount of speakers is fixed. Head orientation was fixed towards the source to minimize spatial distraction. See figure 9f.

The order of these conditions was randomized per subject to avoid effects of fatigue and adaptation.

[18] E.g. ITU Radiocommunication Sector (ITU-R) (2001)

Stimuli The same three stimuli have been used in in the same order in all three conditions. The frequency spectra of all stimuli may be compared using Figure 7.

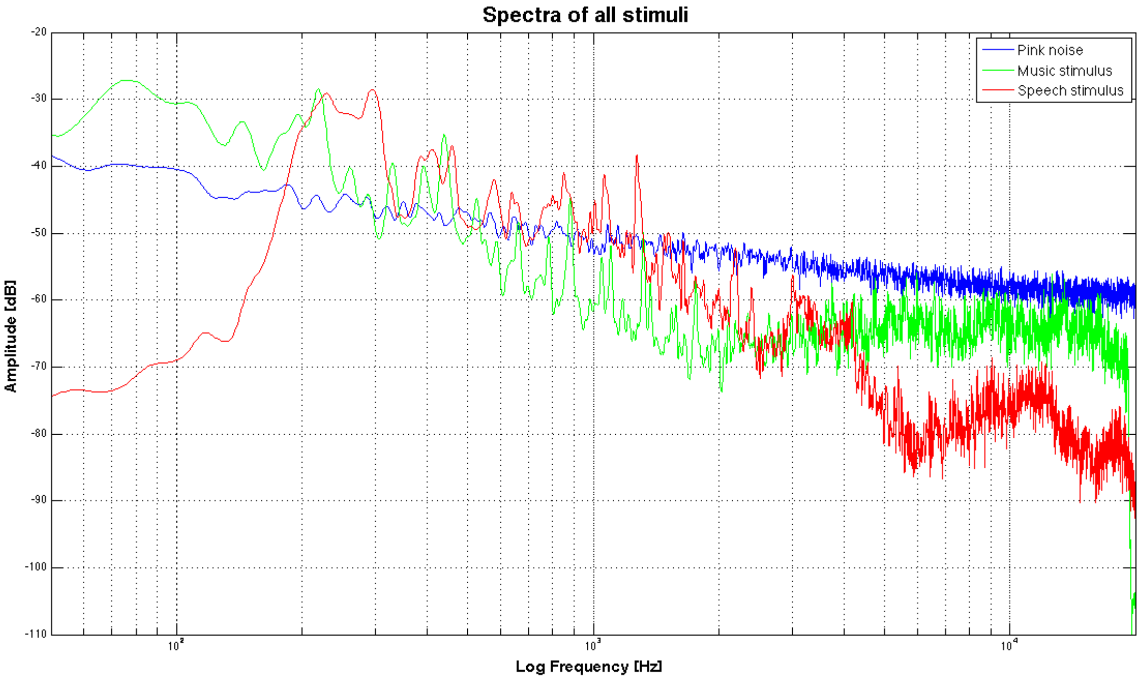


Figure 7 – Frequency spectra of a all stimuli used in this experiment

- **Speech** An eight second looped speech sample of a female voice uttering the German sentences *"Mit einem Male kam die Sonne wieder. Hat das denn gar kein Ende? Die Frau ist klug und weise."*
- **Music** A twelve second loop of the electronic song *"Luv deluxe"* by *"Cinnamon Chasers"*. This piece was chosen for being instrumental and non-intrusive. It contains features such as delicate cymbals and subtle white noise that may help revealing small alterations in the high frequency spectrum.
- **Noise** A pulsed pink noise ($^1/_f$) train composed of 800ms noise bursts with 50ms fade in and fade out time and 500ms of silence. This combination has been shown to be most sensitive to changes in sound coloration.[19] The stimulus was generated entirely using Matlab, figure 8 depicts the timing of the noise train.

[19] E.g. Wittek (2007) Page 142

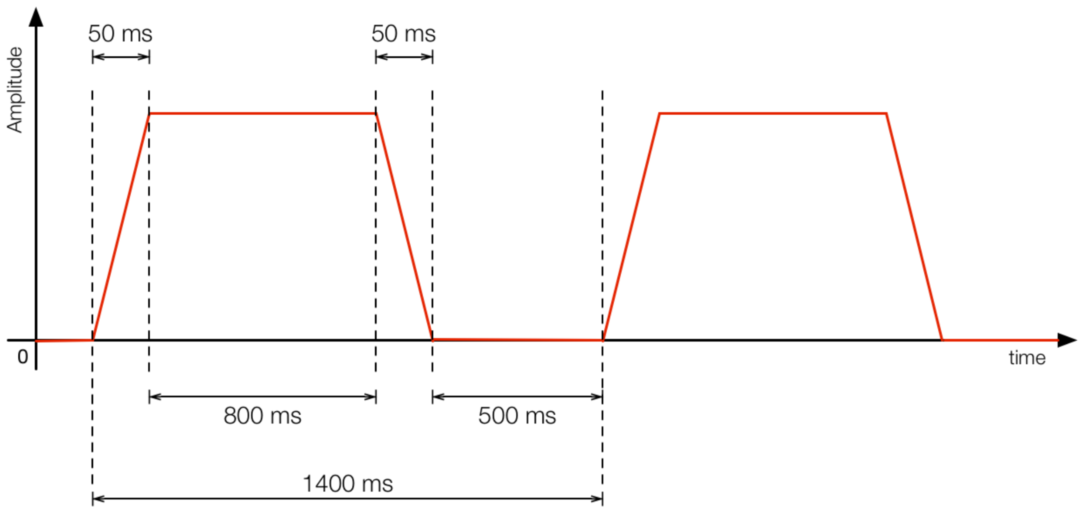


Figure 8 – Timing of the noise train used in the experiment

Simulated arrays All simulated WFS arrays are circular with a 3 meter diameter and evenly distributed speakers. The reference consists of a point source, positioned 2.5 meters from the center of the array. The amount of speakers were generated using the formula $14 \cdot 2^i, i \in [0..8]$. This was chosen to leave the possibility of reproducing the results of the three smallest arrays (14, 28 and 56 speakers) in the room Pinta at Telekom Laboratories which is equipped with a 3 meter, 56 speaker WFS setup. This resulted in the following 9 array sizes:

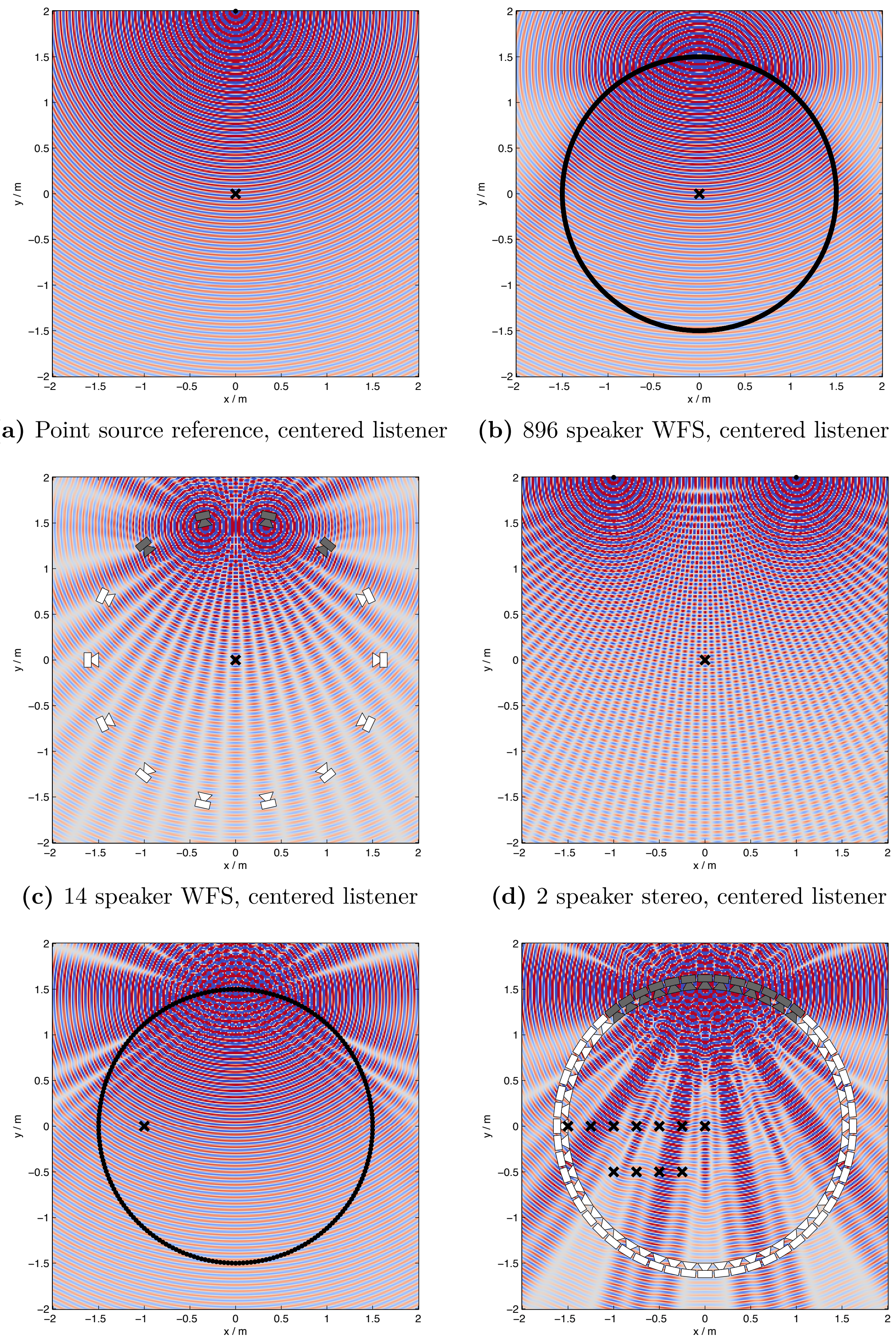
- WFS array with 14 Speakers
- WFS array with 28 Speakers
- WFS array with 56 Speakers
- WFS array with 112 Speakers
- WFS array with 224 Speakers
- WFS array with 448 Speakers
- WFS array with 896 Speakers
- WFS array with 1792 Speakers
- WFS array with 3584 Speakers

Furthermore, a stereo array was added to help answering research question 2. It consists of 2 speakers at ±30° from the listener's position at a distance of $tan(\alpha) \cdot 2.5 = tan(^{\pi}/_{6}) \cdot 2.5 = 1.4434$, see figure 9d.
To comfy to the recommendations of the MUSHRA experiment setup (see section above), two additional impulse responses have been added for a total of 12 signals:

- A hidden reference, an exact replica of the labeled reference point source.
  This marks the upper end of the scale, as a signal should always be rated as “not colored” compared to itself. It also helps detecting potential subject-related issues, such as insufficient hearing performance or random answering.
- An anchor, composed of the reference signal after applying a second order Butterworth high-pass filter with a cutoff frequency of $5kHz$.
  The anchor marks the lower end of the scale, as the drastic change in sound quality should exceed the amount of coloration induced by the arrays under test. Similar to the hidden reference, deviating results for the anchor may also hint at potential subject-related issues, although the lower end of the scale is much more subjective and the results are expected to be more scattered.

A visualization of some of the array configurations and listening conditions may be seen in figure 9, where a point source playing back a $5kHz$ sine wave is simulated. Figure 9a to 9d show examples of the first condition, where 9a depicts the reference point source, figure 9c and 9b show WFS arrays and Figure 9d shows a simulated stereo setup, all playing back a single point source, 2.5 meters in front of the listener.

Figure 9e is an exemplary setup for the second condition, where the virtual listening position is moved one meter to the left. Figure 9f visualizes the third condition, where the speaker number is set to 56 and the position of the listener is moved throughout the array. The size of 56 was chosen to allows a replication of the experiment in room Pinta, see above.



(a) Point source reference, centered listener

(b) 896 speaker WFS, centered listener

(c) 14 speaker WFS, centered listener

(d) 2 speaker stereo, centered listener

(e) 224 speaker WFS, off-centered listener

(f) 54 speaker WFS, moving listener

Figure 9 – Simulated speaker arrays of three meter diameter reproducing a point source playing a $5kHz$ sine wave, 2.5 meters from the listening position. Gray symbols represent active speakers, 'x' marks the listener position. Black dots have been used for arrays larger than 100 speakers.

Further annotations It should be noted that for reasons of efficiency, the SFS toolbox implements WFS using a delay-line with different amplitudes. The whole signal is convoluted once with a so called pre-equalization filter, which raises the amplitude starting from a certain frequency in a logarithmic fashion of $+3db$ per decade. Figure 10 indicates that when the spectral energy of the spatial aliasing, occurring from a certain array-depending frequency upwards, is added to the signal, the filters added amplitude is resulting in a $3dB$ per octave raise in overall amplitude. The pre-equalization filter therefore has to be limited to the frequency spectrum before the aliasing frequency.

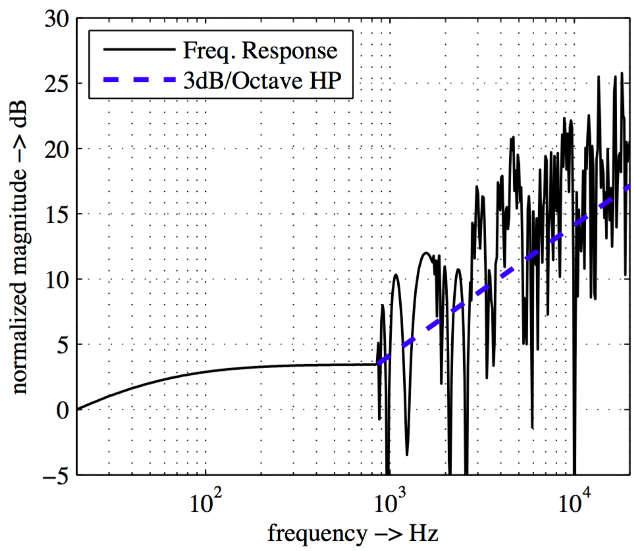


Figure 10 – Frequency response of a traditional WFS driving function using an pre-equalization filter without upper limit[20]

To automate the process, two Matlab scripts were written to find both the appropriate starting point and the the upper limit of the pre-equalization filter. The first script calculates the frequency response of the array without any pre-equalization and finds the first point in the frequency spectrum where the amplitude leaves a certain amplitude range. This is then set as the starting point of the filter.
The second script simulates the same array and applies a $3dB$ per octave filter, starting at the previously calculated frequency. It then predicts the end of the pre-equalization filter using the same technique.

The filter itself in the time domain is given in equation 11.[21]

$$h(t) = \mathcal{F}^{-1}\left\{\sqrt{\frac{i\omega}{c}}\right\} \tag{11}$$

[20] E.g. Spors and Ahrens (2010)
[21] E.g. Wierstorf et al. (2013)

3.2.2 Equipment used

Hardware The computer used to conduct the experiment runs *Debian GNU/Linux 6.0* on 16 AMD Opteron 6128 processors and 32 GB RAM. The subjects are seated inside an auditory booth, all interaction is done using a KVM switch connected to the audio server via Ethernet. The subjects are wearing AKG K601 headphones, amplified using a Millennium HP3 headphone amplifier at a fixed volume setting. The amplifier is fed directly from the stereo output of the audio server.
A prerecorded set of compensational filters for the AKG K601 headphones have been used as a compensation for the coloration introduced by the headphones itself. While this may alter the results slightly, it is assumed that the effects of the compensation filter are negligible compared to the coloration of the WFS arrays.

Software On the audio server, several programs are running simultaneously to conduct the experiment. The *SoundScape Renderer (SSR, version 0.4.0-pre3)* is convoluting 12 channels with a set of 12 impulse responses in real time. These sets have been created beforehand using *Matlab 2013a (8.1.0.604)* and the *SFS toolbox (version 0.2.2)*.
Stimulus playback is implemented using a PureData (*Pd version 0.42-6*) patch. It provides to ability to play, pause and select uncompressed audio files in *.wav* format. In order to accomplish the seamless switching between two stimuli, this patch also switches the channel on which the audio signal is fed to the SSR Renderer. All internal routing is done using the *Jack Audio Connection Kit (Qt GUI Interface Version 0.3.6)*. Lastly, the graphical interface which communicates all user interaction with the underlying PD patch is provided using python and GLIDE. A screenshot of the interface can be seen in figure 11.

3.2.3 Subjects

The 16 test persons used in this experiment were chosen by the examiner and consisted of trained and untrained subjects of both genders. They were allowed to have a short break after each set of conditions (see 3.2.1). After the experiment, one individual reported a slight hearing impairment on one ear which resulted in skewed answers during the third condition where the majority of the audio signal is reproduced on one ear. These results have been omitted.
Furthermore, one individual reported a misunderstanding of the scale during his first run. These answers have been restored with a arithmetical operation to extend the range of his/her ratings. No prior aural examination was performed.

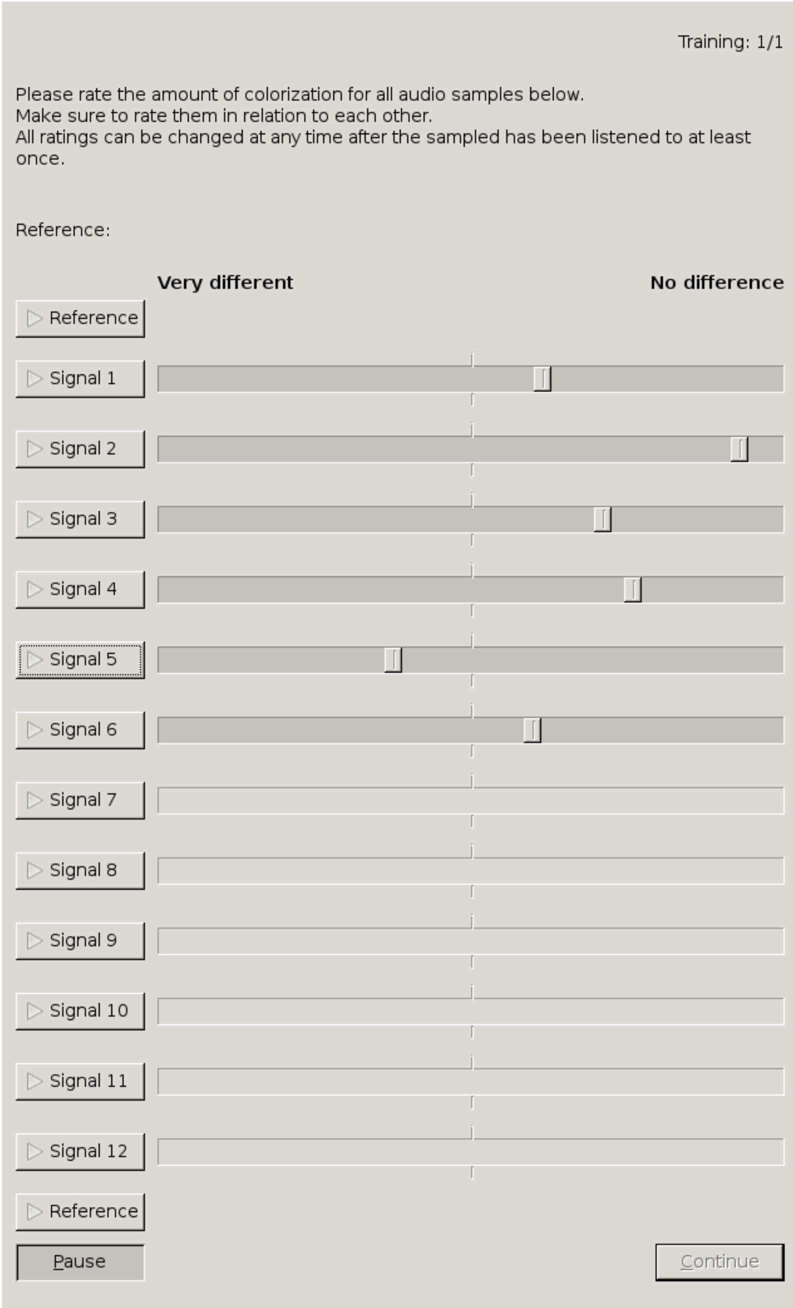


Figure 11 – Screenshot of the graphical interface used in the experiment

3.3 Execution

All subjects were seated in an auditory booth with no visual and audible distractions. They were introduced to the experiment using a verbal explanation and a training page, where answers would be omitted and the subjects were able to get accustomed to the GUI and stimuli.

4 Analysis of results

In this chapter the results found in the hearing experiment are presented and analyzed. First, the general distribution of scores divided by listening condition and stimulus are exploratively analyzed in section 4.1. These preliminary insights are then used in section 4.2 to answer the research questions formulated in section 3.1.

Please note that the term “better” does not directly apply to the measured ratings, as coloration does not directly infer quality of sound. Nevertheless, it may be used on a scale of faithfulness of reproduction in a fashion of “less colored”, while “worse” describes a more colored sound respectively. Furthermore, timbral fidelity has been shown to be strongly correlated with basic audio quality (BAQ).[22]

4.1 Exploratory evaluation of experiment conditions

When gathering very subjective data such as psychoacoustic comparisons from human subjects, the distribution of results is expected to be scattered. Effects such as fatigue, distraction and even imagination can randomly skew the results of even the same person. Therefore, the median has been chosen to rank and interpret all collected data instead of the mean. It is more robust against outliers which are inevitable and represents the true average more accurately in a skewed distribution. This means that the 95% confidence intervals are just approximation, as they are given by the nearest rank to the calculated upper and lower limit. The lower 95% confidence limit is given by the $\frac{n}{2} - \frac{1.96\sqrt{n}}{2}$ ranked value, the upper 95% confidence limit by the $1 + \frac{n}{2} + \frac{1.96\sqrt{n}}{2}$ ranked value. This may result in confidence intervals that are not centered around the median.

[22] E.g. Rumsey et al. (2005)

4.1.1 Experiment condition: Centered listener

Figure 12 shows the median of all scores with a centered listening position over the tested array, separated by stimulus. This main condition has been run twice, the scores represent the median of the average of both runs. The bars indicate the 95% confidence interval as discussed above. The anchor is included to the very left, the hidden reference on the very right. All other scores are ordered in ascending order according to the number of loudspeaker, from the two speaker stereo to the 3584 speaker WFS setup.

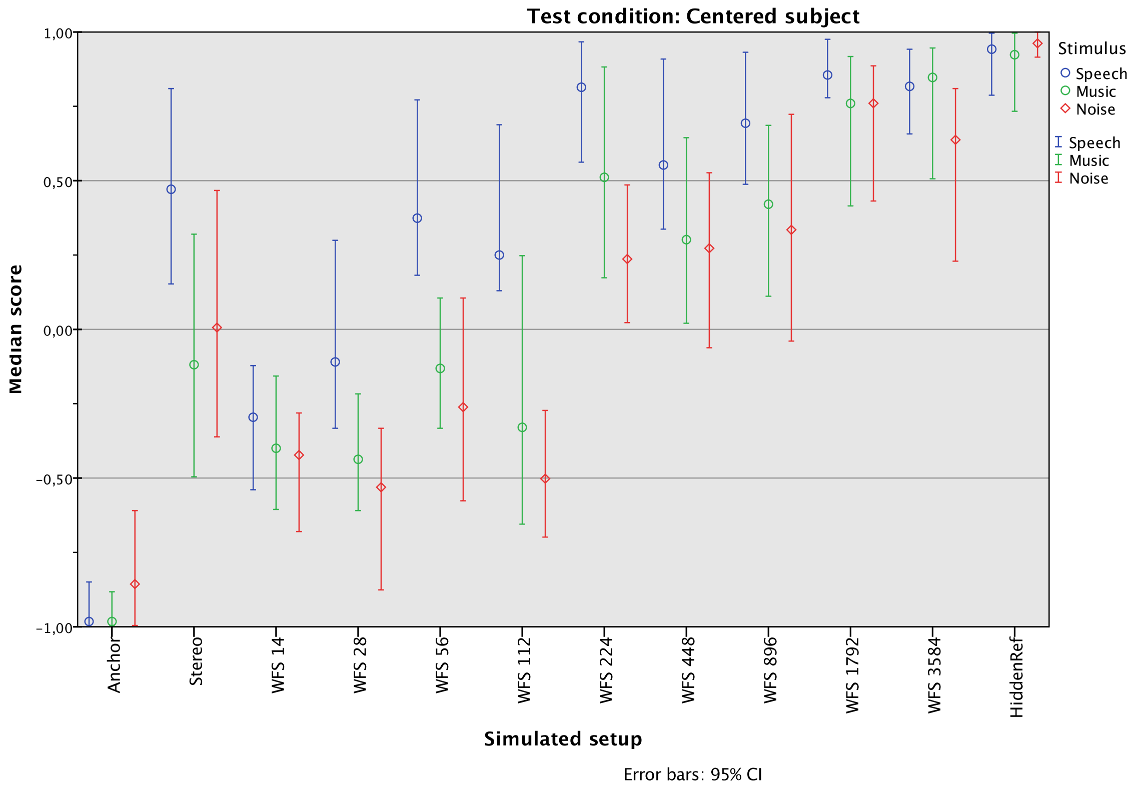


Figure 12 – Median of results with condition „listener centered“ split by stimulus over the reproduction array. A score of +1 indicates no coloration, −1 indicates a strong coloration compared to a reference.

A few things may be observed: The anchor is regarded as very strongly colored, while the hidden reference is regarded as not colored; all other scores rank in between. This means the anchor was chosen appropriately as it marks the lower end of the scale but is not too different to diminish the subjective scale. As expected, overall coloration is decreasing with higher number of speakers in WFS arrays but does not appear to become transparent against the reference, even at 3584 speakers.
The speech stimulus is generally rated as the least colored. For all simulated setups but the reference, anchor and the stereo setup, the noise stimulus is rated as the most colored. The music stimulus is usually rated slightly above the noise stimulus.

4.1.2 Experiment condition: Off-center listener

Similarly to figure 12, figure 13 again shows the distribution of the median of all scores, this time with the listener in an off-center position 1 meter to the side. As this condition was only run once instead of twice, less data points have been recorded and the 95% error bars are larger compared to above.
The general distribution seem to be similar to the centered condition above, with the exception of the stereo setup, which is rated much lower. This seems to indicate the bad reproduction properties of a stereo setup outside the sweet spot, which will be further discussed in section 4.2.2. Again, the anchor was regarded as very colored while the hidden reference was rated transparent. The same general trends can be noted here as well - overall coloration goes down with increased loudspeakers. Again, the speech stimulus was regarded as least colored in all cases, while the noise stimulus results were generally rated most colored.

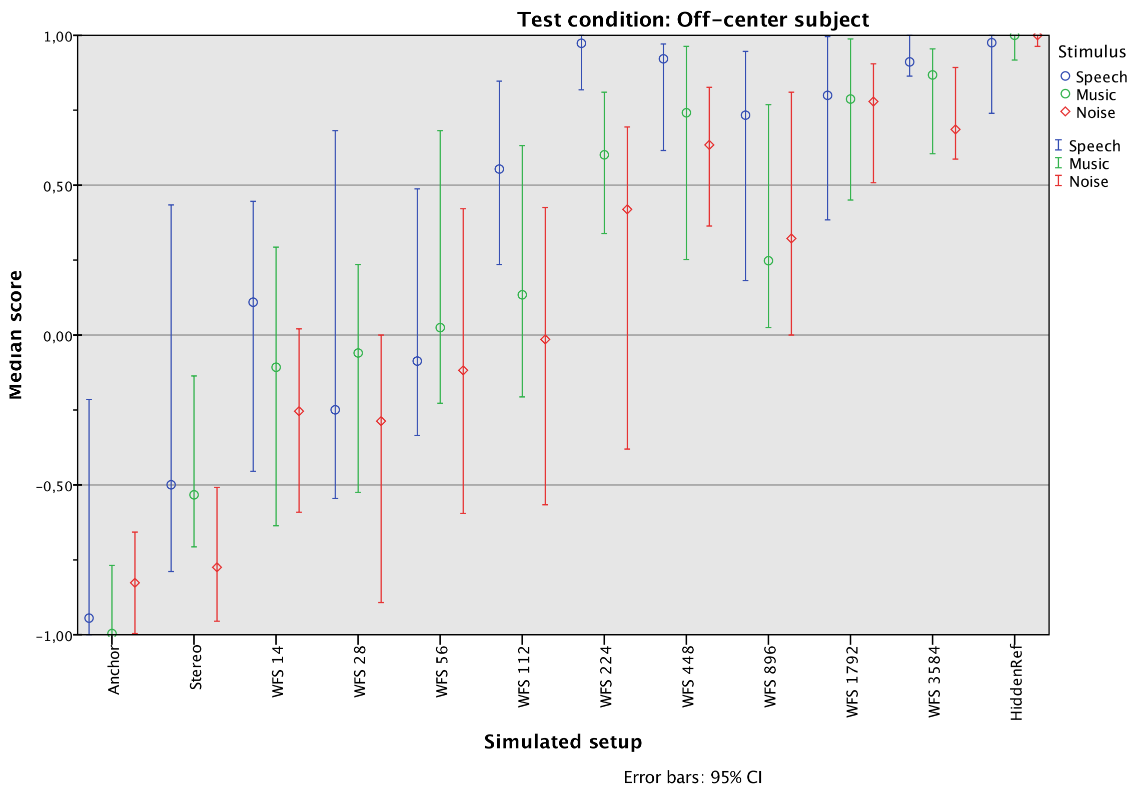


Figure 13 – Median of results with condition „listener off-center“ split by stimulus over the reproduction array. A score of +1 indicates no coloration, −1 indicates a strong coloration compared to a reference.
Bars indicate the 95% confidence interval.

4.1.3 Experiment condition: Moving Listener

Figure 14 show the distribution of scores when moving through the virtual WFS array. Position 1 through position 6 mark *0.25m* steps to the left of the center. Position 7 to position 10 follow these spots 0.25m behind the center. For a visual representation, please refer to figure 9f.
The large error bars suggest that ranking the amount of perceived coloration reliably was more difficult compared to the conditions of section 4.1.1 and 4.1.2, even though the virtual head direction was always turned towards the source and the level of sound was normalized. This could be due to the lack of experience of the subjects with assessing more subtle differences in coloration. A anchor was deemed unnecessary as no absolute ranking was required.

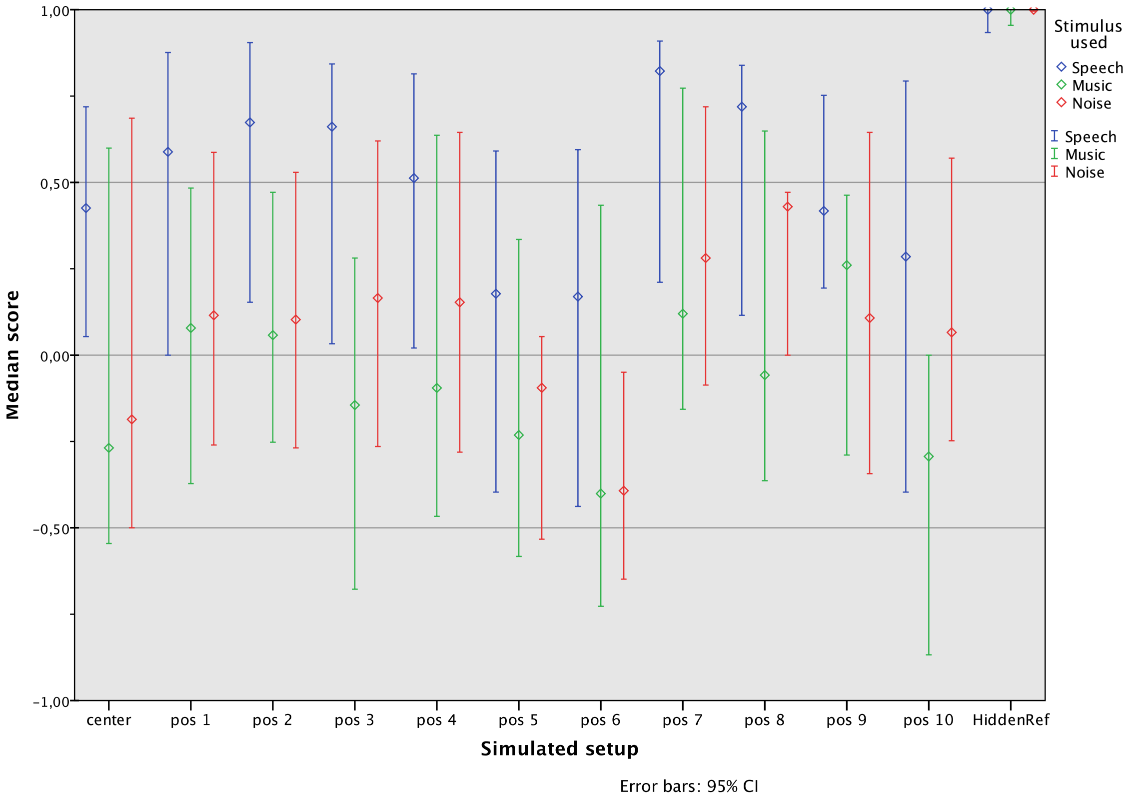


Figure 14 – Median of results with condition “moving listener”.

A slight increase in coloration can be observed when approaching the secondary sources, as can be seen from the slow decline in score from the centered position to position 6. Returning to the center and moving 25cm back, position 7 through 10 exhibit a similar pattern. The hidden reference was recognized as such and rated as not colored.

4.2 Evaluation of research questions

After the exploratory analysis of the results in section 4.1, the research questions stated in 3.1 are evaluated. Because the distribution of the scores cannot be assumed to be normally distributed, only two non-parametric test for independent samples are used: The Independent-Samples Mann-Whitney U Test for the comparison of exactly two distributions and the Independent-Samples Kruskals-Wallis Test for comparison of three or more distributions.[23] Furthermore, a short regressional analysis is conducted to find the general shape of the distribution of coloration over loudspeaker number.

4.2.1 Question 1: Coloration as a function of speaker number

Combining the findings of 4.1.1 and 4.1.2 yields interesting results. It seems that some specific array sizes are less likely to result in perceived coloration than other array with a higher amount of speakers, which was not predicted by the frequency responses, as per figure 3.

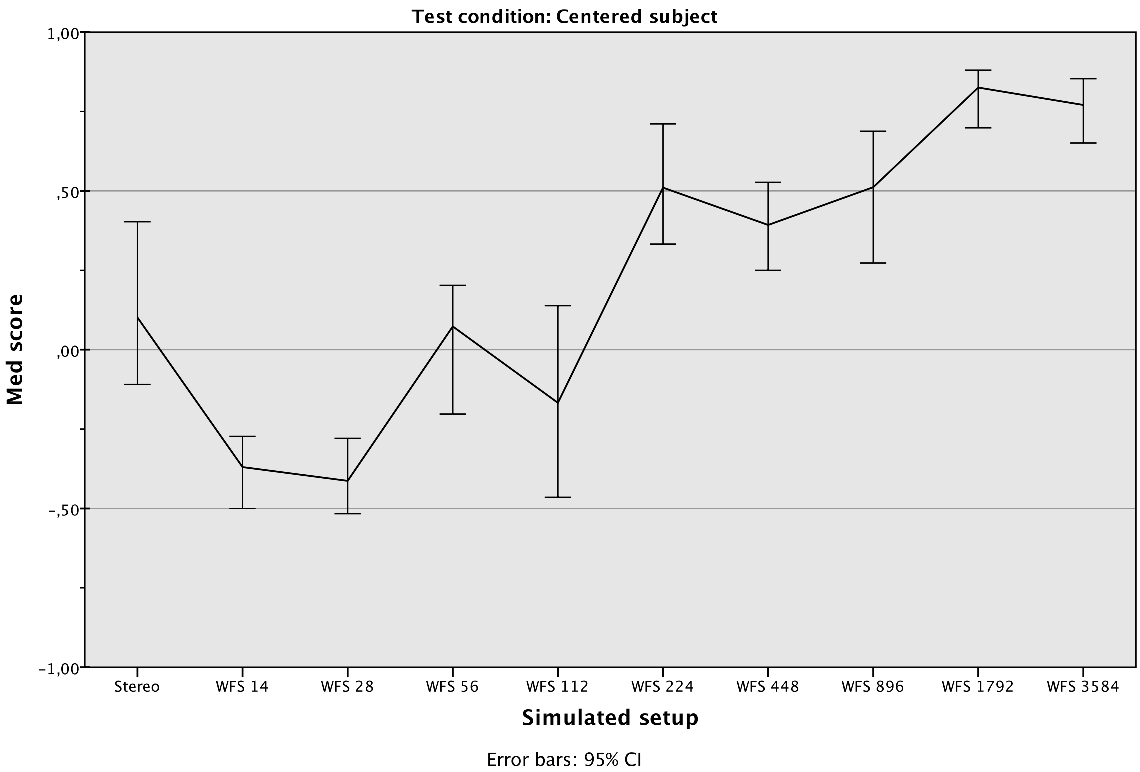


Figure 15 – Median of results across all setups in a centered listening position in ascending order, combining all stimuli. Error bars show 95% confidence interval

[23] E.g. Field (2009)

This becomes evident looking at figure 15, which depicts the combined median in ascending order of all three stimuli (speech, music and noise). It indicates two things: The correlation generally approaches transparency but there seems to be a few WFS configurations outperforming the array with twice the speakers. Especially the WFS array with 224 speakers was ranked surprisingly well.
Please note that although all rated stimuli are probable audio signals, the consolidation is a only for reasons of clarity, as the distribution across the stimuli is significantly different for some setups (see section 4.2.3, table 4). The same graph separated by stimulus can be found in the appendix (figure 23).

Sorting the ratings by rank (median) shows the overall trend of coloration, as depicted in figure 16.

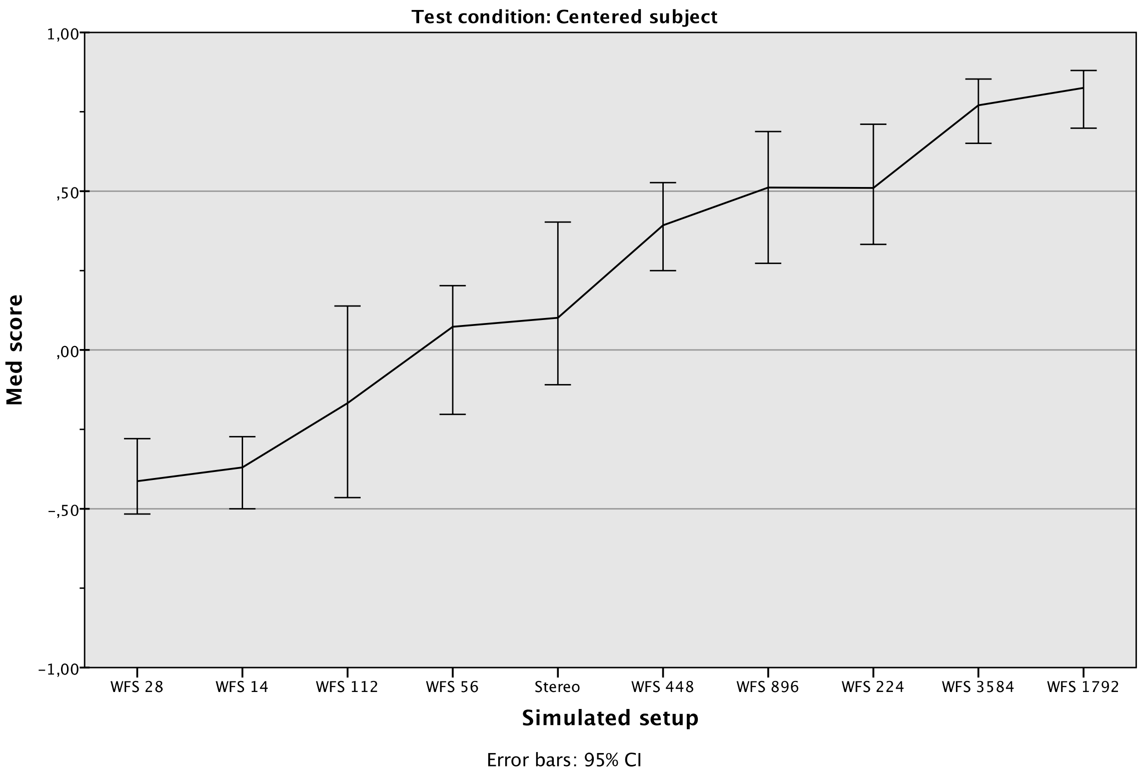


Figure 16 – Median of results across all setups in a centered listening position, combining all stimuli. Error bars show 95% confidence interval

The Independent-Samples Mann-Whitney U Test can be used to test whether result distribution of consecutive setups are significantly different from each other or not and therefore may be grouped together. A significance value below 0.05 means the Null Hypthesis *The distribution of results across categories of setup* has been rejected and the tested setup configurations have been rated significantly different.

The following groups are not are not significantly different from each other:

- WFS 14 and WFS 28, two sided significance = 0,738
- WFS 56 and WFS 112, two sided significance = 0.831
- WFS 224, WFS 448 and WFS 896, two sided significance = 0.091
- WFS 1792 and WFS 3584, two sided significance = 0.450

All other combinations are significantly different with a two sided significance well below the cutoff of $p = 0.05$.

Further examination of the WFS 224 array Of the groups above, the WFS 224 array again stands out as being not significantly different to an array with 4 times the speakers. To further investigate the performance of the WFS 224 array, it is compared to the WFS 896 and the WFS 3584 array.

Table 1 – Significance of differences in rating distribution, compared to the WFS 224 array

	Centered Position			Off-Centered Position		
	Speech	Music	Noise	Speech	Music	Noise
WFS 896	**0.287**	**0.270**	**0.491**	0.019	**0.094**	**0.696**
WFS 3584	**0.780**	**0.128**	0.001	**0.897**	0.004	**0.067**

Results of the Independent-Samples Mann-Whitney U Test
Null hypothesis: *The distribution of score is the same compared to the WFS 224 array.*
Significance level is 0.05, all retained hypotheses are marked bold.

Interestingly, there is little significant difference between the distribution of the result of the WFS 224 array and the much larger WFS 896 and WFS 3584 array. Splitting up the tests allows for a finer analysis. There are three significant cases where the null hypothesis was rejected: In the category off-center speech, the WFS 224 was rated significantly **less colored** as the WFS 884 array. Therefore, the WFS 224 array should be preferable in all cases.
While the median suggests a better overall performance of WFS 3584 array, it only significantly differs from the WFS 224 array in the categories centered noise and off-centered music. The improvements are therefore only justifiable when using very complex signals. It is not clear why the result distribution of the category off-center noise across WFS 224 and WFS 3584 are not significantly different.

Regressional analysis Although the prediction of an overly accurate mathematical or graphical description is hard due to the immense scattering of the result distribution, the general shape of the increase of reproduction fidelity over number of loudspeaker may be examined.

For this reason, all data points of the centered position are plotted over their real amount of loudspeaker in figure 17. Because of the large span of used speaker numbers, the x-axis was logarithmized. Therefore, a logarithmic curve appears as a linear slope while a linear regression appears as an exponential function.

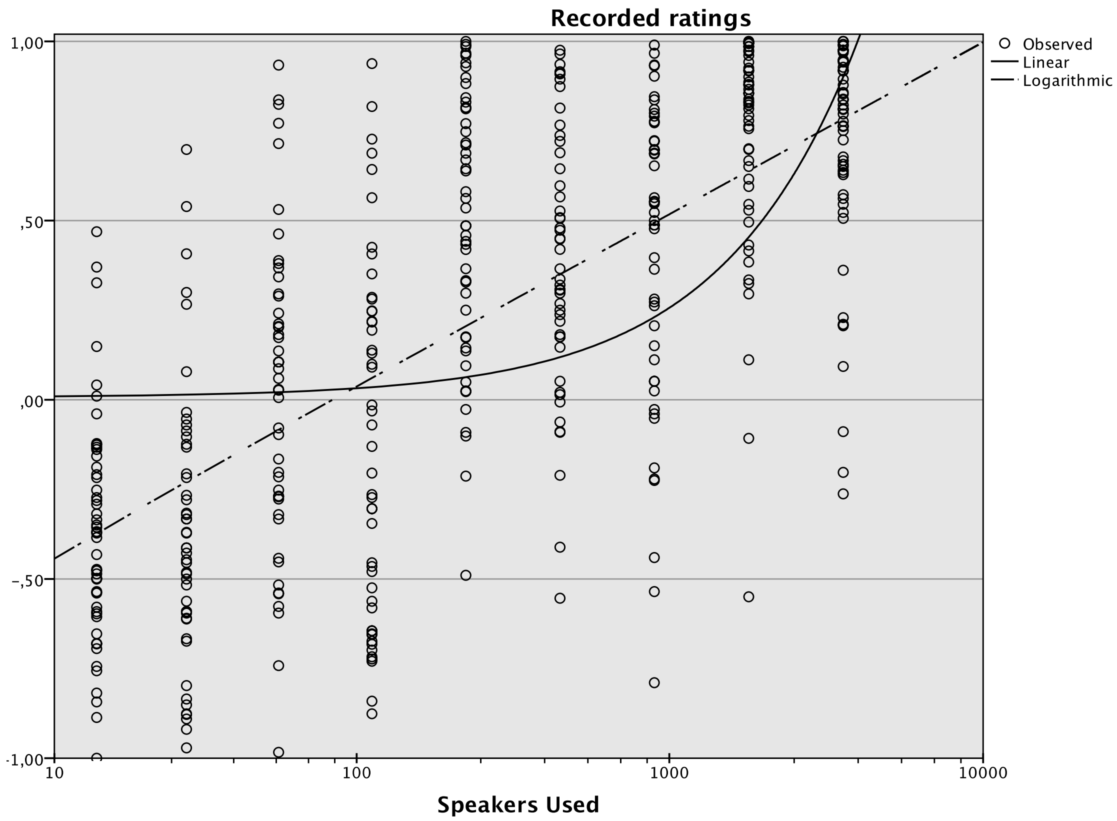


Figure 17 – Linear and logarithmic regression through the recorded score of perceived coloration over the logarithmic speaker number.

Both types of regressions were fitted to the available data. A comparison of the linear and the logarithmic regression shown in figure 17 using ANOVA (analysis of variance) parameters can be seen in table 2. The R^2 value, which represents the explained variation divided by the total variation, of the logarithmic model is over 76% larger compared to the linear case, indicating a much better fit of the logarithmic regression. In the logarithmic regression, 45.6% of the variance of the data can be explained by the model. Furthermore, the residual sum of squares of the logarithmic regression is more than 26% smaller compared to the linear case.

Table 2 – Comparing linear and logarithmic regression

	Model Summary			ANOVA			
	R	R^2	Std. Error		Sum of Squares	Mean Square	F
Lin	0.507	0.258	0.477	Regression	34.119	34.119	149.787
				Residual	97.947	0.228	-
Log	0.675	0.456	0.409	Regression	60.200	60.200	360.194
				Residual	71.867	0.167	-

4.2.2 Question 2: Coloration of WFS compared to stereo

Comparing the results for stereo in a centered listening position (section 4.1.1, figure 12) with the off-center position (section 4.1.2, figure 13), it seems obvious that the performance of stereo dramatically depends on the positioning of the listener. Figure 18 illustrates this further.

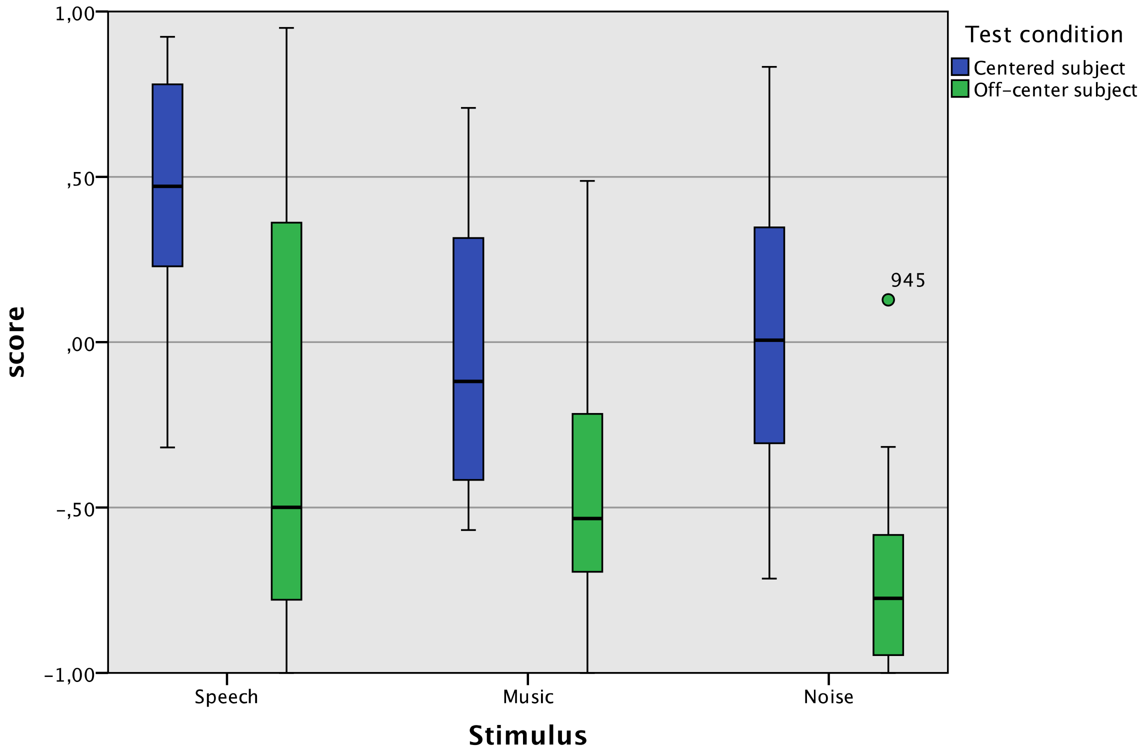


Figure 18 – Box-and-whiskers plot of the ratings for the stereo setup, separated by stimulus and listener location. The median is marked inside the boxes, which indicate the 95% confidence interval as per section 4.1, whiskers indicate 98% confidence interval.

This is confirmed by the Independent-Samples Mann-Whitney U test of table 3, which shows that the results of the off-center stereo array are significantly worse.

Table 3 – Comparing rating distribution of stereo setup over listening position

	Speech	Music	Noise
Asymptotic significance	0.003	0.014	0.00

Null Hypothesis: The distribution of results is the same across categories of listening position
Significance level is 0.05, therefore **all hypotheses are rejected**.

While in the case of a centered listener, stereo was on par with an WFS 224 array in regards to coloration (see figure 16), moving the listening position one meter to the side leaves the stereo setup in last place by a big margin (see figure 21). It can be concluded that at least 224 speakers are required to safely perform better than a regular stereo system, as the distribution of ratings is significantly better across listening conditions (two sided significance = 0.000 with the Null Hypothesis being the distribution of rating of cases stereo and WFS 224 are across the category listening position).

4.2.3 Question 3: Effects of stimulus choice on perceived coloration

Looking at figure 7, which shows the frequency spectra of all stimuli, it seems obvious that different stimuli demand different WFS sizes to be reproduced faithfully. This is confirmed graphically by figures 12 through 14. Table 4 shows the the results of running the Independent-Samples Kruskal-Wallis Test over all setups, separated by conditions. The Null Hypothesis is that the distribution of score is the same across all stimuli.

Table 4 – Significance of differences in rating distribution based on stimulus

	Listener location	
	center	off-center
Stereo	0.005	0.046
WFS 14	**0.507**	**0.274**
WFS 28	0.002	**0.260**
WFS 56	0.000	**0.320**
WFS 112	0.000	0.014
WFS 224	0.000	0.000
WFS 448	0.038	0.021
WFS 896	**0.075**	**0.117**
WFS 1792	**0.128**	**0.970**
WFS 3584	**0.177**	0.009
Anchor	**0.099**	**0.295**
Hidden Ref	**0.301**	**0.345**

Null Hypothesis: Distribution of score is the same across all stimuli
Significance level: 0.05, retained hypotheses are bold.

It seems that the difference in stimulus is not significant in very large and very small arrays. This means that combing the results of all stimuli, as done in figure 15 and 16, can only be seen as an approximation for arrays between 28 and 448 speakers.

Figure 12 further suggests that the difference in rating between the music and the noise stimulus was very little. This suspicion is confirmed by running the Independent-Samples Mann-Whitney U Test for these two stimuli, see table 5.

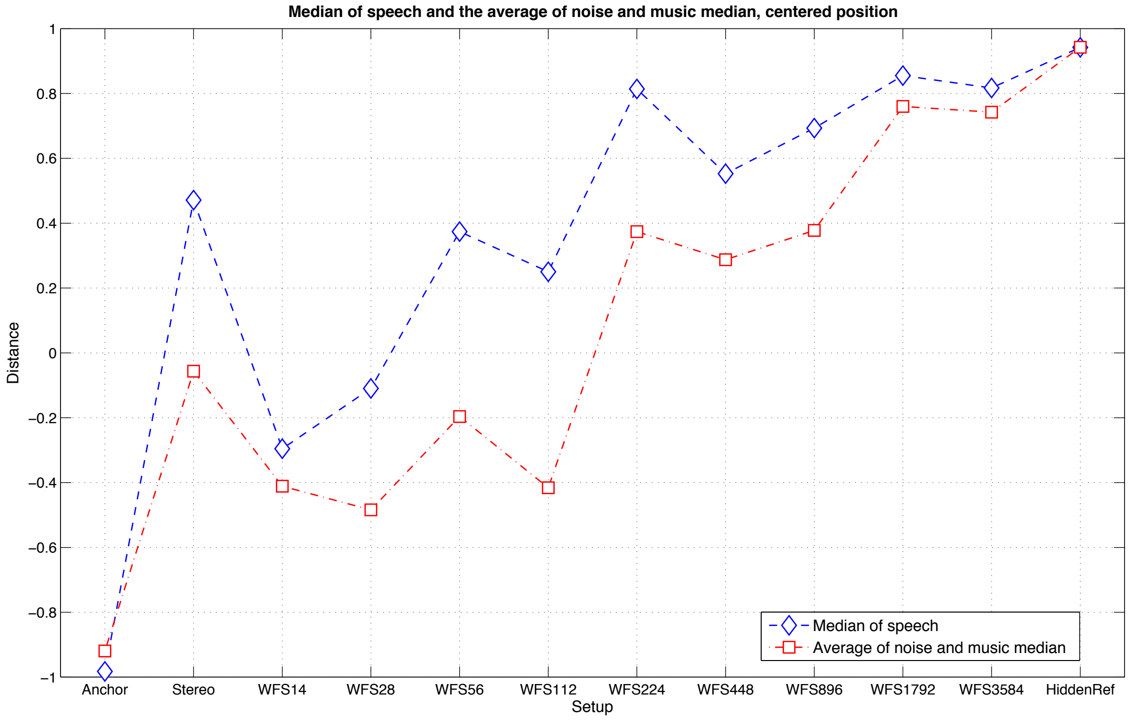
Table 5 – Significance of rating distribution of music and noise

	Listener location	
	center	off-center
Stereo	**0.696**	0.019
WFS 14	**0.780**	**0.590**
WFS 28	**0.287**	**0.149**
WFS 56	**0.323**	**0.184**
WFS 112	**0.239**	**0.224**
WFS 224	0.019	**0.254**
WFS 448	**0.564**	**0.184**
WFS 896	**0.642**	**0.590**
WFS 1792	**0.956**	**0.867**
WFS 3584	**0.184**	**0.184**
Anchor	**0.061**	**0.102**
Hidden Ref	**0.138**	**0.491**

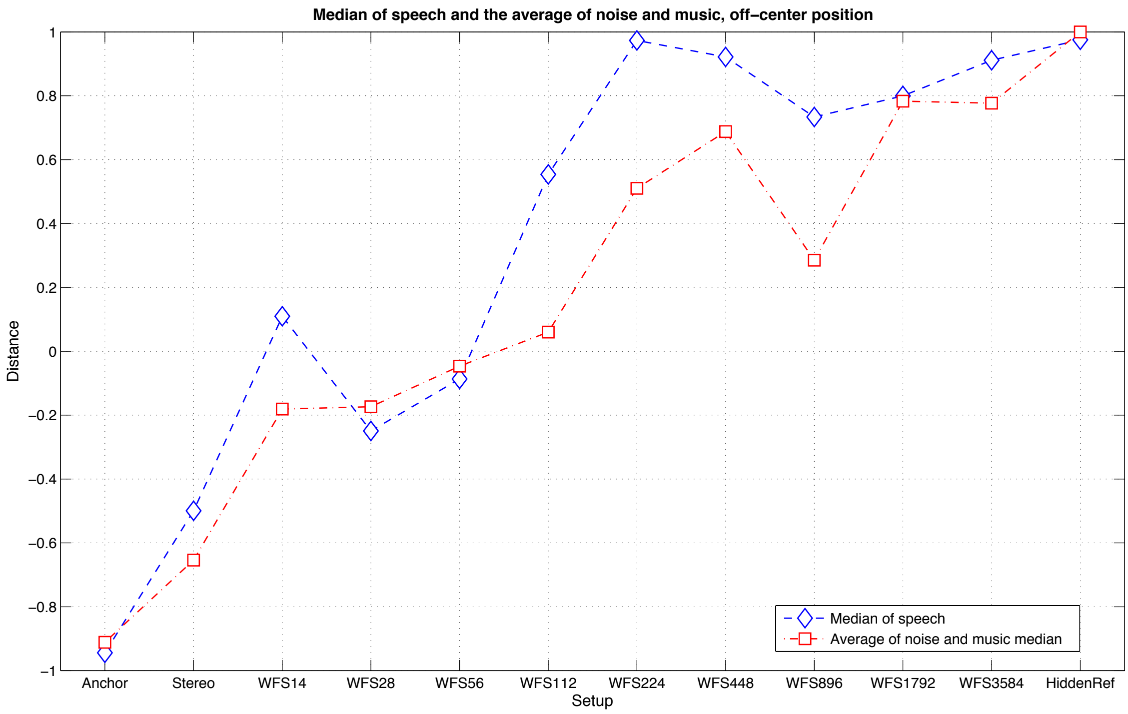
Null Hypothesis: Distribution of score is the same across stimulus music and noise
Significance level: 0.05, retained hypotheses are bold.

Here, these two stimuli seem to not be significantly different in all but two arrays, stereo off-center and WFS 224 centered. This means that the median of these two stimuli can be averaged to further compare it to the speech stimulus.
In a follow-up to the experiment, most subjects noted that the repetition of the music signal was much more pleasant compared to the pink noise train. Because of the results of table 5, it might be reasonably to replace the unpleasant noise sample with this specific music sample, especially if longer test runs are necessary.

Figure 19 shows both the median of speech and the mean of the median of noise and music to compare the band-passed speech signal to the more complex noise and music signal.



(a) Median of speech compared to the average of the noise and music median in centered position.



(b) Median of speech compared to the average of the noise and music median in an off-center position

Figure 19 – Comparison of median depending on stimulus and split by listening position

To further visualize the gap, figure 20 depicts the numerical distance of both values over each setup and split by listening condition. As expected, both the anchor and the hidden reference are rated very similar and therefore show very little difference of median for both listening positions. Looking at the blue graph of the centered position, a large increase in distance between median of speech and noise/music can be seen. This is explained by the relatively low demands for speech reproduction. Looking at figure 19a, speech approaches transparency at a faster rate than the other stimuli. This gap peaks at the 112 WFS array, after which increasing the number of speaker leads to a big change in coloration of music and noise, closing the gap.

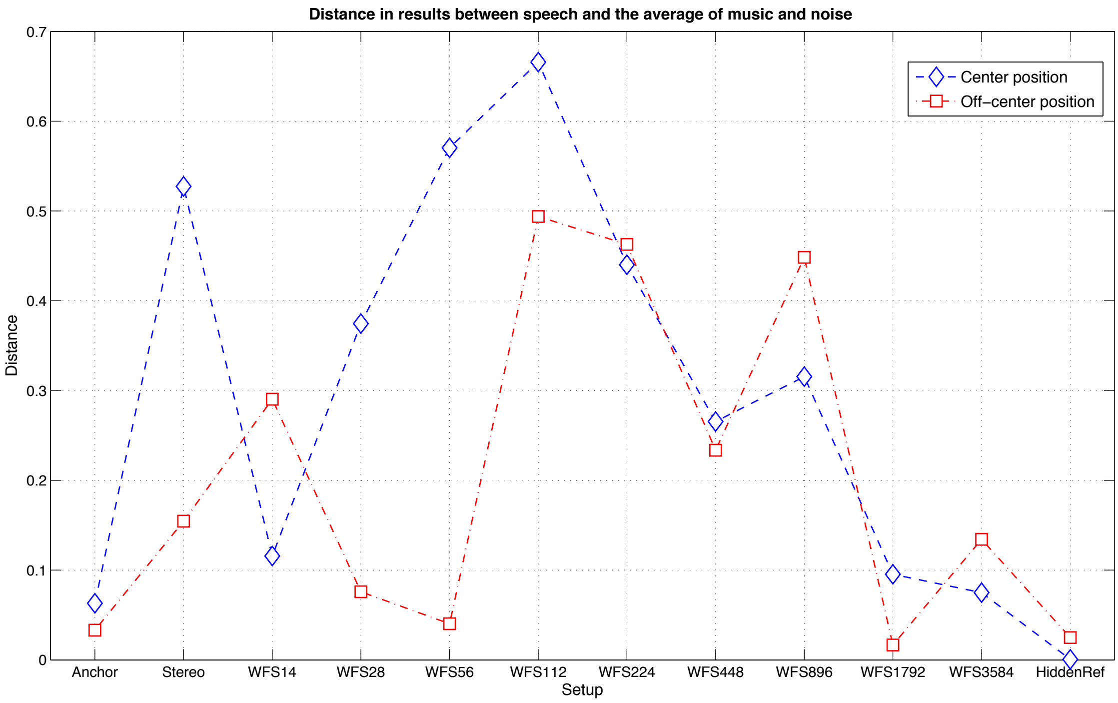


Figure 20 – Distance of speech from the average of music and noise score for a centered and off-center listening position

Comparing these finding to the off-center location, two things may be noted: Below 112 speakers, big differences to the centered listening position were recorded. These may be due to the fact that outside of the centered sweet spot, spatial aliasing drowns out differences between stimuli, shrinking the gap in ratings. This doesn't explain the higher off-center sensitivity for the WFS14 array.

Upwards of 112 speakers, the graphs behave very similarly, as the ratings are approaching the upper limit of the scale and the arrays are less prone to spatial aliasing.

4.2.4 Question 4: Coloration as a function of listener location

Off-center location Calculating the median of the result distribution over all setups combining all stimuli in an off-centered listening location and comparing it to the graph of figure 15 gives surprising results, as can be seen in figure 21.

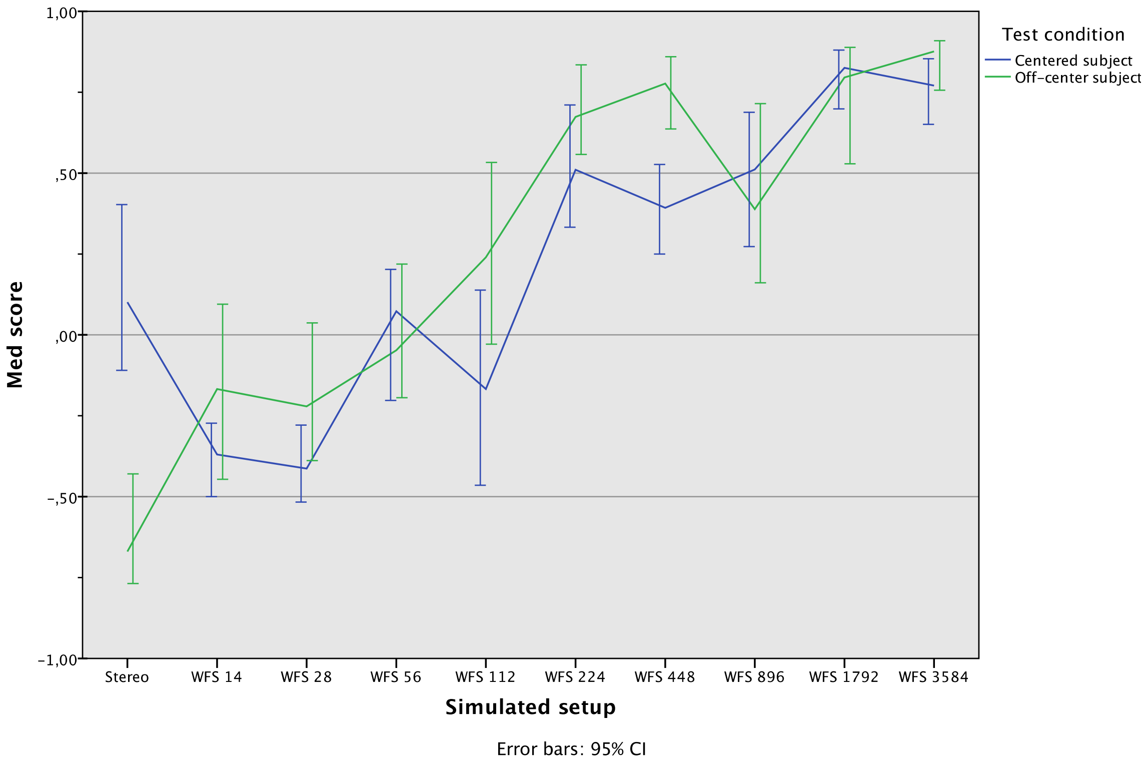


Figure 21 – Median of results in centered (blue) and off-centered (green) listening position across all 9 WFS setups combining all stimuli.
Error bars show 95% confidence interval

It appears that with the exception of the stereo setup, the distribution of ratings generally follow the same shape. Interestingly, the off-center position was rated similarly or better compared to the centered case. This is confirmed by table 6, where each setup is compared to itself across the category listening position. Bold values mark if the null hypothesis "The Distribution of scores is the same across categories of listening position" is retained. If this is not the case, an arrow indicates whether the off-center position was rated higher (upwards arrow, significantly lower coloration) or lower (downwards arrow, significantly higher coloration).

Table 6 – Significance of differences in rating distribution across the category listening location

	Stimulus		
	Speech	Music	Noise
Stereo	0.003 ↘	0.014 ↘	0.000 ↘
WFS 14	**0.094**	**0.210**	**0.149**
WFS 28	**0.956**	0.039 ↗	**0.171**
WFS 56	0.008 ↘	**0.110**	**0.381**
WFS 112	**0.160**	**0.051**	0.039 ↗
WFS 224	0.024 ↗	**0.985**	**0.564**
WFS 448	0.035 ↗	0.032 ↗	0.029 ↗
WFS 896	**0.809**	**0.642**	**0.590**
WFS 1792	**0.361**	**0.752**	**0.809**
WFS 3584	0.014 ↗	**0.616**	**0.539**

Results of the Independent-Samples Mann-Whitney U Test
Null hypothesis: *The distribution of score is the same compared to the WFS 224 array.*
Significance level is 0.05, all retained hypotheses are marked bold.
Arrows mark the performance of the off-center compared to the centered position if distribution of ratings is significantly different (null hypothesis rejected).

While the performance of the stereo setup is significantly lower across all stimuli (also see section 3.1.2), WFS generally stayed the same or even improved. Most noticeable was this trend in the WFS 448 array, where the off-center location was regarded as significantly less colored across all stimuli. This thesis offers no satisfactory explanation for this trend, further studies are warranted to determine the exact cause.

Moving listener Figure 22 shows the distribution of ratings when moving through a circular WFS setup of 3 meter diameter when the speaker number is fixed at 56. The center of the coordinate system $(0:0)$ is set at the center of the array. Interestingly, approaching the secondary sources (increasing x coordinate) only increases perceived coloration at around 1.25. Before, the WFS array shows no significant differences.
Moving $0.5m$ behind the center, a slight increase of the median can be obsevered, which drowns out when moving towards the secondary sources. This increase was not significant in any case, as may be seen by the strongly coinciding confidence intervalls and table 7, which lists the significance values of running an Independent-Samples

Mann-Whitney U Test to compare the result distribution at each x coordinate across the category y_coordinate, separated by stimulus.

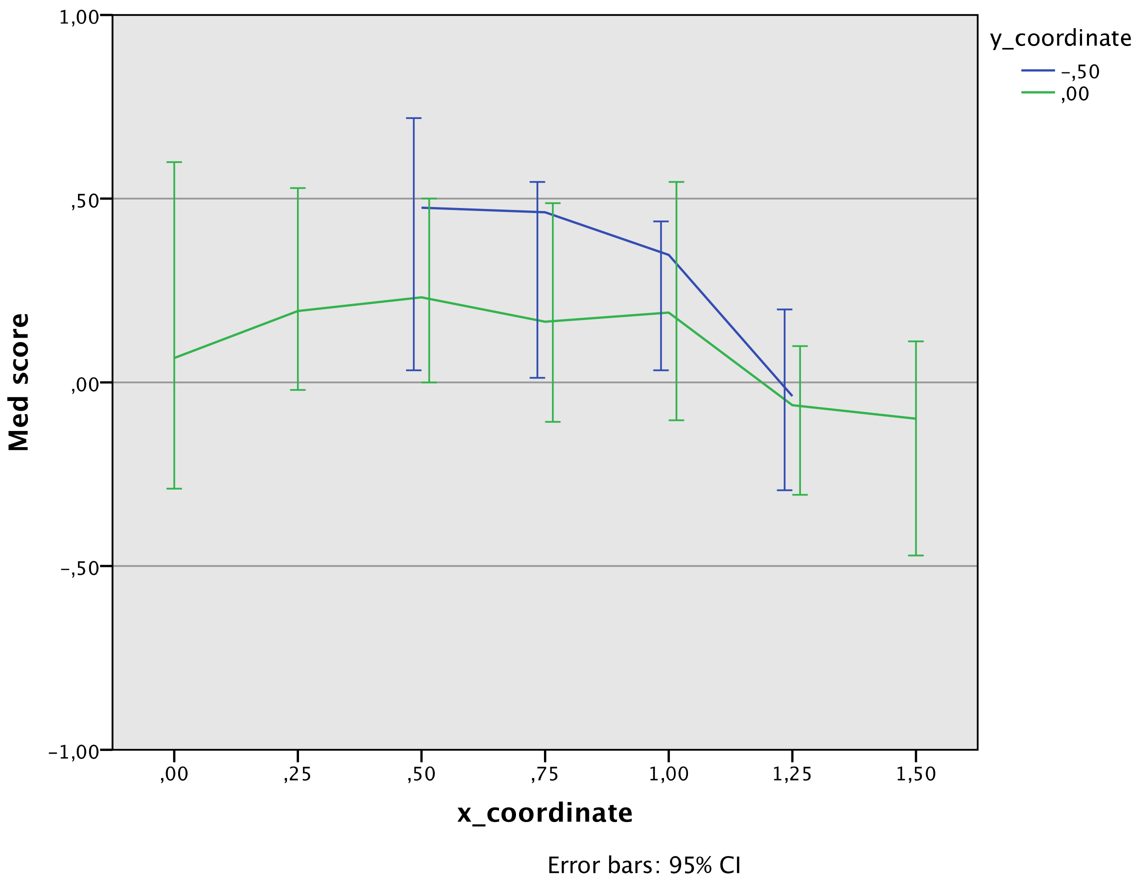


Figure 22 – Median of results when moving through a 56 speaker WFS array. The green line depicts the ratings when moving from the center to the left in steps of $0.25m$, the blue line follows the same steps $0.5m$ behind. Error bars show 95% confidence interval

Table 7 – Significance of differences in rating distribution across the category y-coordinate

		Stimulus	
x coordinate	Speech	Music	Noise
0.50	**0.838**	**0.412**	**0.436**
0.75	**0.775**	**0.345**	**0.461**
1.00	**0.902**	**0.935**	**1.000**
1.25	**0.775**	**0.367**	**0.089**

Results of the Independent-Samples Mann-Whitney U Test
Null hypothesis: *The distribution of score is the same across the categories y coordinate.*
Significance level is 0.05, all hypotheses are retained!

5 Summary

While outlining every detail of the conducted experiment is not practical, the following pages should do a good job of summarizing the findings.

5.1 Conclusions

Coloration strongly depends on the amount of speakers used in a three meter circular WFS array. The overall reproduction capabilities seem to correlate in a somewhat logarithmic manner to the number of speakers. They increase rather quickly until about 224 speakers, from where doubling the amount doesn't improve the systems performance significantly in most cases.

Up until 224 Speakers, a simple stereo setup was regarded as less colored when sitting inside the sweet spot. Moving the listening position one meter to the left strong decreases the performance of the stereo array, where it shows significantly more coloration than even a WFS 14 array. To conclude, a circular WFS 224 array performs significantly better in all listening positions.

The stimulus has a visible impact on the recorded scores, as speech signal was rated significantly higher than both a piece of electronic music and a pink noise impulse train in all but very small and very large arrays. This effect is less renowned in an off-center position. In a centered position, this effect is especially pronounced in medium sized arrays from 28 to 896 speakers and the stereo setup. Lower amounts of speaker produce more spatial aliasing which in turn decrease the sensitivity of the listener to smaller changes between stimuli; more speaker result in better reproduction capabilities, closing the gap between stimuli. The first effect is especially pronounced in an off-center position, where up until 112 speakers, stimuli are rated very similarly.
Surprisingly, the music and noise stimulus don't show significant differences in all but two cases (centered WFS 224 and off-center stereo). This indicates that using a piece of music alone might be suitable to test for coloration, as most subjects regarded the noise stimulus as unpleasant.

Looking at the listeners location it was shown that the performances of the WFS arrays stay stable or even improve significantly, which can not be explained at this point. Again, perceived coloration decreases roughly in a logarithmic manner until 224 speakers from where it only slightly declines. Especially the WFS 448 array improves in an off-center location across all stimuli.

Moving the listener throughout the array shows that the coloration of WFS doesn't become significantly worse when approaching the secondary speakers until about $0.25m$ from the edge of the surrounding speaker array.

5.2 Future outlook

Following this thesis, a lot of experiments may be designed to further investigate the coloration in WFS. One obvious suggestion is the constructing a similar experiment that implements non-static, dynamic HRTFs and a head tracker or even a real array; although feasibility of the latter might be restricted for now. Furthermore, the amount of speaker for absolute transparency may be investigated, setting a definitive upper limit for speaker number in WFS arrays.

The findings of this thesis can also be generalized by testing other array sizes and forms, as there probably is a correlation between the number of speaker and the spatial size of the array. Especially the surprising performance of the WFS 224 array might demand further research, especially in different configurations. Although established, the connection between perceived coloration and perceived quality may be further explored in the context of WFS.

5.3 Final thoughts

Despite not providing full transparency, a circular WFS array with 224 speaker seems to be a very promising setup size. Although intricate to build, it is still viable to design and build such an array. It seems to rival much larger systems on reproduction capabilities both centered and throughout the array, making it the near perfect compromise of quality and feasibility. Of course, it is still to be determined how such a system would perform if physically build but the results of the simulated array used in this thesis were very promising.

References

ITU Radiocommunication Sector (ITU-R). Bs.1534 : Method for the subjective assessment of intermediate quality levels of coding systems, January 2001. URL `http://www.itu.int/rec/R-REC-BS.1534-1-200301-I/en`.

Jens Ahrens, Matthias Geier, and Sascha Spors. The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. In *Audio Engineering Society Convention 124*, 5 2008. URL `http://www.aes.org/e-lib/browse.cfm?elib=14460`.

American Standards Association and Acoustical Society of America. *American standard acoustical terminology: (including mechanical shock and vibration)*. ASA Standard. American Standards Association, 1960. URL `http://books.google.de/books?id=t9oIAQAAIAAJ`.

H. Fastl and E. Zwicker. *Psychoacoustics: Facts and Models*. Springer series in information sciences. Springer, 2007. ISBN 9783540688884. URL `http://books.google.de/books?id=eGcfn9ddRhcC`.

A. Field. *Discovering Statistics Using SPSS*. ISM (London, England). SAGE Publications, 2009. ISBN 9781847879073. URL `http://books.google.de/books?id=a6FLF1YOqtsC`.

Anselm Goertz, Michael Makarski, Christoph Moldrzyk, and Stefan Weinzierl. Entwicklung eines achtkanaligen lautsprechermoduls für die wellenfeldsynthese. 2007.

Thomas Görne. *Tontechnik*. Hanser Fachbuchverlag, 2008. ISBN 9783446415911. URL `http://amazon.com/o/ASIN/3446415912/`.

Neale Lytollis. The idea of being somewhere else: Iosono's wave field synthesis, October 2008. URL `http://www.eyho-blog.com/2008/10/15/iosono-wave-field-synthesis/`.

Per Rubak and Lars G. Johansen. Coloration in natural and artificial room impulse responses. In *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, May 2003. URL `http://www.aes.org/e-lib/browse.cfm?elib=12332`.

Francis Rumsey, Slawomir Zielinski, and Rafael Kassier. On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America*, 118(2):968–976, 2005. doi: http://dx.doi.org/10.1121/1.1945368. URL `http://scitation.aip.org/content/asa/journal/jasa/118/2/10.1121/1.1945368`.

W. Snow. Basic principles of stereophonic sound. *Audio, IRE Transactions on*, AU-3(2):42–53, 1955. ISSN 0096-1981. doi: 10.1109/TAU.1955.1165407.

Sascha Spors and Jens Ahrens. Analysis and improvement of pre-equalization in 2.5-dimensional wave field synthesis. In *Audio Engineering Society Convention 128*, May 2010. URL `http://www.aes.org/e-lib/browse.cfm?elib=15418`.

Sascha Spors, Rudolf Rabenstein, and Jens Ahrens. The theory of wave field synthesis revisited. In *124th AES Convention*, pages 17–20, 2008.

H. Wierstorf, A. Raake, M. Geier, and S. Spors. Perception of Focused Sources in Wave Field Synthesis. *Journal of the Audio Engineering Society*, 61(1/2):5–16, January 2013.

Hagen Wierstorf and Sascha Spors. Sound Field Synthesis Toolbox. In *132nd Convention of the Audio Engineering Society*, 04 2012. URL `http://www.aes.org/e-lib/browse.cfm?elib=16605`.

Helmut Wittek. *Perceptual differences between wavefield synthesis and stereophony*. PhD thesis, University of Surrey, 2007.

Appendix

CD content

A CD containing all relevant figures and MATLAB scripts is attached to each physical copy of this thesis. The content is as follows:

- Full digital copy of the presented thesis in the pdf format
- All figures depicted in the presented thesis as vector graphics in the pdf format
- Matlab scripts

  BRS_preparation.m Generates multiple 720 Channel .wav files containing the IR for a particular array size on the first two channel, variable parameter is amount of speakers

  BRS_moving_listener.m Generates multiple 720 Channel .wav files containing the IR for a particular array size on the first two channel, variable parameter is listener position

  freqresp_wfs.m Draws the impulse response of WFS arrays

  freqresp_stereo.m Draws the impulse response of stereo arrays

  findhprelow.m Method to find the starting frequency of the correctional WFS filter

  findhprehigh.m Method to find the upper frequency of the correctional WFS filter

  generate_noise.m Method to generate white or pink noise with adjustable length and fade-in/out times

  generate_noise_train.m Method to generate customized noise train

- A copy of the sfs-toolbox used to generate all relevant impulse responses

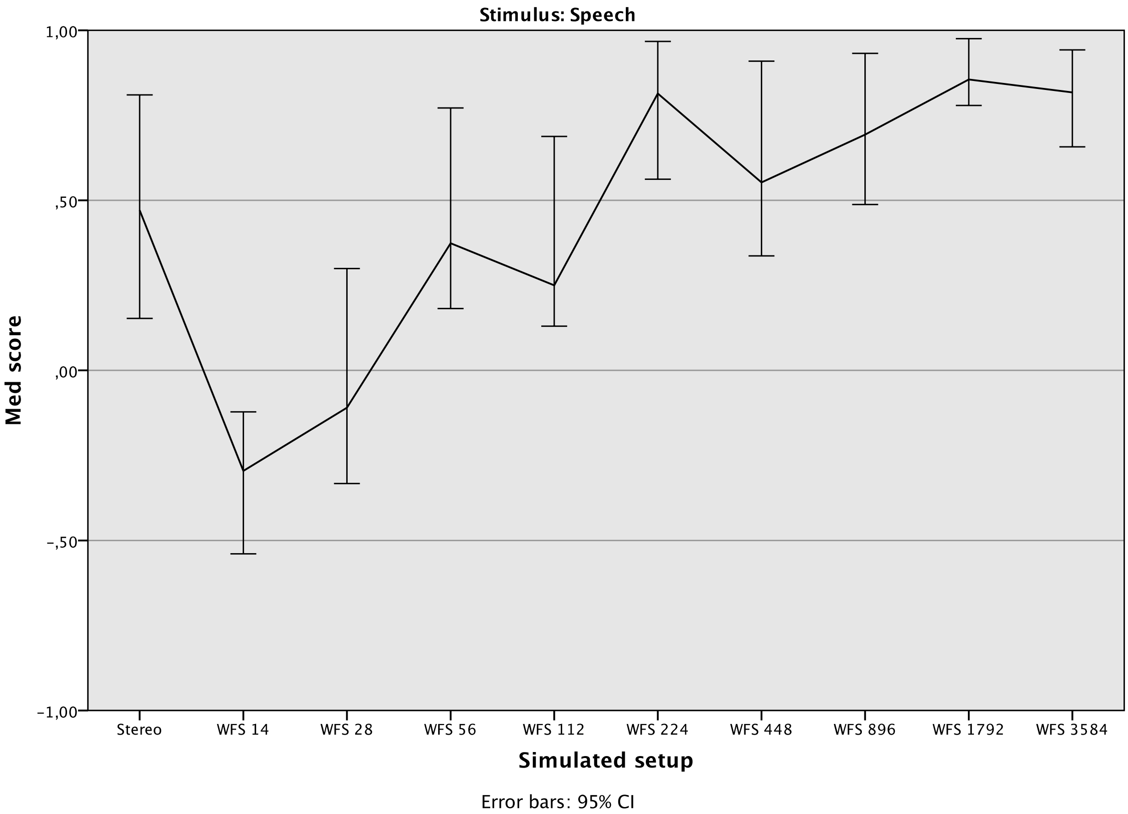
  sfs-master-0.2.2

- A copy of all three stimuli used in this experiment as uncompressed .wav files

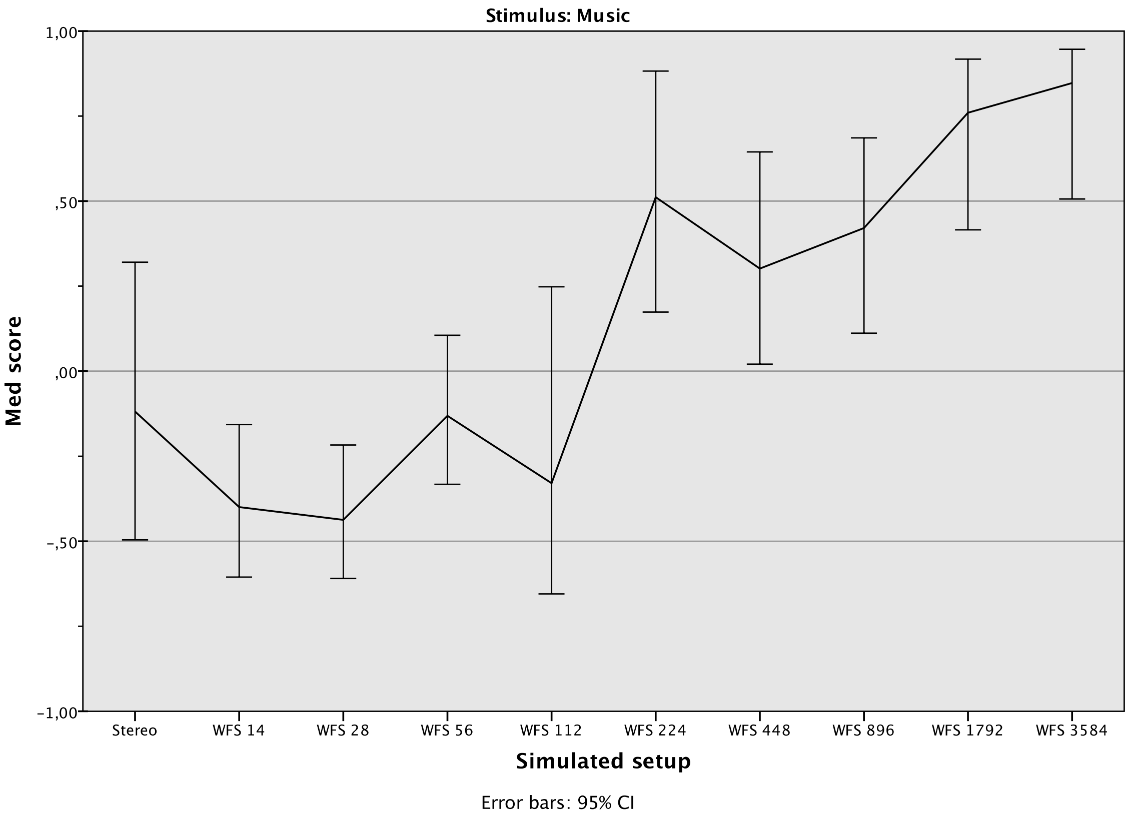
  stimuli/speech.wav
  stimuli/music1.wav
  stimuli/pnoise_pulse.wav

- A copy of the headphone compensation filter and HRTF set

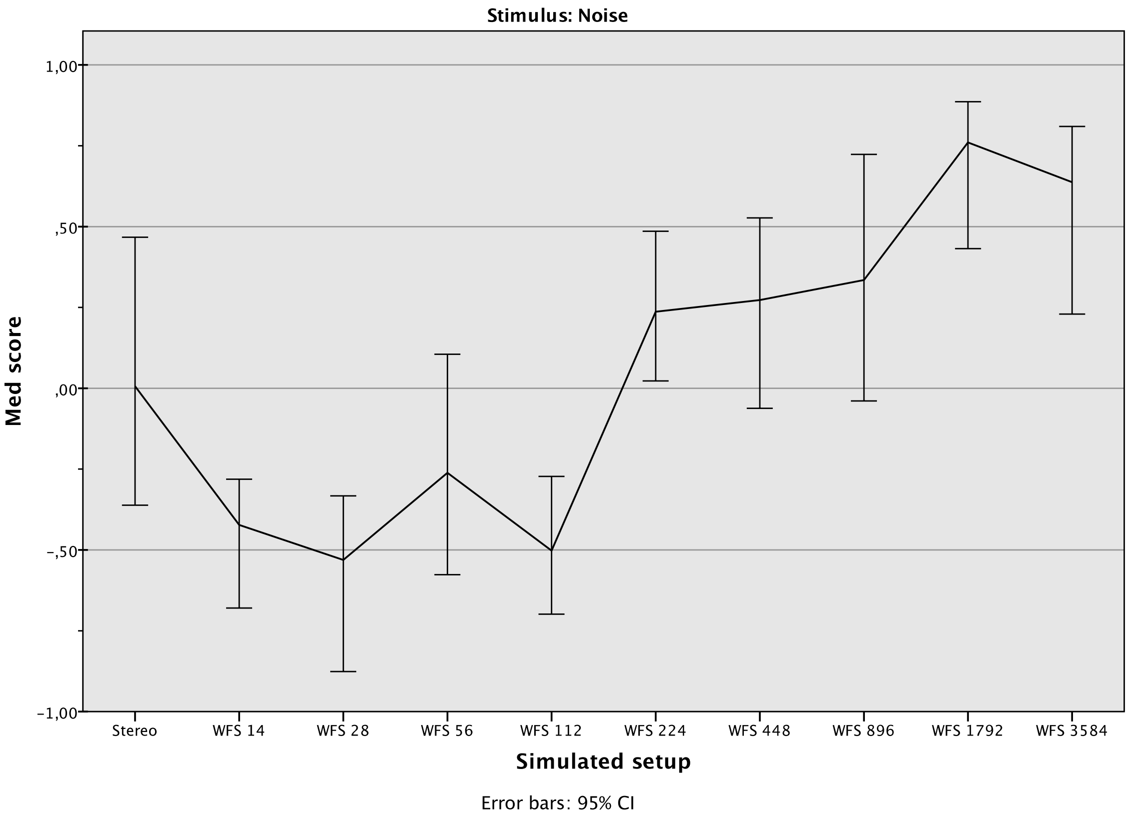
  compensation/QU_KEMAR_AKGK601_hcomp.wav
  HRTF/QU_KEMAR_anechoic_3m.mat



(a) Median for speech stimulus in centered position



(b) Median of music stimulus in centered position



(c) Median of noise stimulus in centered position

Figure 23 – Median of results across all simulated setups split by stimulus in a centered listening position)
Error bars show 95% confidence interval

Table 8 – Significance of distribution difference between two runs in centered listening position

	Stimulus					
	Speech		Music		Noise	
Setup	MWU	K-S	MWU	K-S	MWU	K-S
WFS 14	0.809	0.699	0.869	0.941	0.468	0.699
WFS 28	0.224	0.021	0.196	0.211	1.000	0.941
WFS 56	0.239	0.415	0.515	0.415	1.000	0.941
WFS 112	0.423	0.094	0.642	0.699	0.642	0.941
WFS 224	0.956	0.941	0.086	0.211	0.110	0.415
WFS 448	0.539	0.699	0.468	0.699	0.423	0.415
WFS 896	0.809	0.094	0.696	0.699	0.468	0.415
WFS 1792	0.724	0.699	0.224	0.699	0.564	0.699
WFS 3584	0.239	0.699	0.867	1.000	0.056	0.094
Stereo	0.073	0.415	0.491	0.415	0.491	0.699

MWU: 2-sided asymptotic significance of the Mann-Whitney U test for independent samples
K-S: 2-sided asymptotic significance for the Kolmogorov-Smirnov test for independent samples
Null Hypothesis: The distribution of the variable result is the same across both runs.
The **significance level** is 0.05, therefore **all hypotheses are retained**.

Table 8 shows the results of running both the Mann-Whitney U (MWU) and the Kolmogorov-Smirnov (K-S) test for independent samples. The null hypothesis for all tests is that the distribution of results is the same across both runs. All significance values are above the cutoff value of 0.05, the Null Hypothesis is therefore retained. It can be safely assumed that the hearing experiment is easily reproduced.
While the Mann-Whitney U test is sensitive to differences in rank, the Kolmogorov-Smirnov test also accounts for differences in shape of the distribution. Furthermore, it performs better with small population samples ($N < 25$). For these reasons, both tests have been used to achieve the highest accuracy.[24]

[24] E.g.: Field (2009)