Helmut Wittek – "Perception of Spatially Synthesized Sound Fields"

December 03 – v2

Critical Review of Literature

Helmut Wittek

University of Surrey - Institute of Sound Recording, Guildford, Surrey, UK

Institut für Rundfunktechnik, Munich, Germany

Floriansmühlstr.60, 80939 München, Germany

mailto: wittek@irt.de

web: www.hauptmikrofon.de/wittek.html

A. Introduction

To create the "perfect acoustic illusion" – this has been the goal of sound reproduction techniques since their conception. However, due to technical restrictions, this goal has never been realized.
Today, as modern technologies enable audio systems to become ever more complex, it is now conceivable to achieve this illusion (for which the modern term is „Immersion"). These modern technologies allow ideas and techniques formulated in the past to be put into practice.

One of these ideas is the „acoustic curtain", expressed by Steinberg, Snow and Fletcher (Steinberg and Snow, 1934) in the early 30's (see Figure 1). They aimed to transport the acoustic of the recording venue to a reproduction room using microphone and loudspeaker arrays. These scientists quickly noticed that, due to technical constraints, it would not be feasible to put their ideas into practice. As a compromise, they applied three-channel stereophony, accepting that the original aim of recreating the real sound field would no longer be fulfilled. Snow described this precursor of WFS in this way:

"The myriad loudspeakers of the screen, acting as point sources of sound identical with the sound heard by the microphones, would project a true copy of the original sound into the listening area. The observer would *then employ ordinary binaural listening, and his ears would be stimulated by sounds* identical *to those he would have heard coming from the original sound source."* (Snow, 1953)

In the late 80s, the Wave Field Synthesis (WFS) concept was introduced by the Technical University of Delft. One of the Delft authors later mentioned, remembering the work of the aforementioned authors:

„The intuitive acoustic curtain concept is here replaced by a well-founded theory" (Boone, 2001).

The origin of this "well-founded" theory was in 1987 when Berkhout published the book "Applied Seismic Wave theory" (Berkhout, 1987) and "A holographic approach to acoustic control" (Berkhout, 1988) in 1988. Here, he suggested "acoustical holography" (not yet called WFS) to be the ultimate tool for acoustical control systems in theatres. Berkhout introduced the physical basis of WFS by applying algorithms known from seismics to the field of acoustics. The basic work on WFS was continued in "Wave front synthesis: a new direction in electro-acoustics" (Berkhout et al., 1992) and "Acoustic control by wave field synthesis" (Berkhout et al., 1993).

WFS provides a noticeable enhancement on Stereophony in terms of the spatial reproduction of sound scenes. However, before judging the use of this technique, both the advantages and constraints of the system regarding perception should be noted. An overview of the relevant literature is provided here, in order to investigate this further.

Chapter B describes the theoretical background of WFS. Chapter C introduces terminology and techniques required for the assessment of perceptual properties. The existing knowledge about the perception of WFS sound fields is described in chapter D.
References and Figures are given in chapters E and F.

B. Basic Principles and Theoretical Background of the Wave Field Synthesis (WFS) Concept

B.1. Theoretical origin

WFS is based on the Huygens principle that was quantified by Kirchhoff (Kirchhoff-Helmholtz-Integral, see Start (1997 p.13ff) and Figure 2). His theorem states that at any listening point within a source-free volume V, the sound pressure can be calculated if both the sound pressure and the component of the particle velocity are known on the surface S enclosing V. Figure 2 shows the formula and a graphical illustration.

If the surface S degenerates to a plane, separating the listening area from the primary source area the so-called Rayleigh integrals (Start, 1997 p.15ff) can be applied. The next step is the reduction of the plane to a line of secondary sources. Using a mathematical procedure called "Stationary phase approximation" (Bleistein, 1984), the so-called Rayleigh 2½ D (Start, 1997 p.28ff) integrals are derived, leading to the driving signals of a line array of loudspeakers. The so-called "driving function" of the array loudspeakers arises from these integrals. For this reason, it is also called "Rayleigh 2½ D synthesis operator". They are simply the direct mathematical formulation of the "acoustic curtain" with certain directivity characteristics of the microphone/loudspeaker pair plus a position-independent equalization (3dB/Octave boost). Figure 3 illustrates this simple basis of WFS. The synthesis operator can be adapted to the actual directivity characteristics of the array loudspeakers (de Vries, 1995).

This knowledge was derived and described by authors of the TU Delft, for instance Berkhout (1987, 1988), Berkhout and de Vries (1992, 1993), Boone et al. (1995), further in doctoral theses, published as books by Vogel (1993), Start (1997), Verheijen (1998) and Sonke (2000). In addition to their different scientific approaches, these books in particular give an excellent introduction to and overview of the basic theories of WFS.

B.2. Physical Potential of WFS

Figure 4 illustrates the principle characteristics of WFS: For the entire listening area, the acoustic scene remains constant, i.e. the *absolute* setup of the acoustic scene is independent of the listening position. The *relative* acoustic perspective as perceived by the listener changes with movements of the listener. This change involves a realistic change of the sound pressure level when the distance to the virtual source is varied.
This may be called *"motion parallax"*, similar to visual perception. The role of motion parallax for acoustic perception is discussed in chapter D.5 below.

The theoretical capabilities of WFS to create a quasi-realistic sound field or to recreate an existing sound field are even larger. It is, for instance, possible to simulate a certain directivity characteristic of the virtual source. Furthermore, the location of the secondary (array) loudspeakers is no limitation for the creation of virtual sources. WFS –although not covered by the Kirchhoff-Helmholtz theory- allows the synthesis of virtual sources both in front of and behind the array. In particular, the creation of the so-called focussed sources (sources in front of the array) could make a significant difference to conventional sound reproduction techniques.

From a creative point of view, WFS offers an improvement of flexibility: Both direction- and location-stable sources can be reproduced. The design of the acoustic scene is less limited to the constraints of the reproduction technique in comparison to Stereo. The simulation of a real acoustic scene is more plausible.

These capabilities and their constraints are discussed in the relevant chapters of this review.

B.3. Physical Constraints of WFS

The aim of WFS is the creation of a true copy of a natural sound field. This high aim can - in practice - only be fulfilled with certain restrictions. The main reasons and their consequences are listed in the following box:

Constraint in practice	leads to	Artefact
1. Discretisation of the continuous secondary source distribution to a loudspeaker array	→	Spatial Aliasing (→ B.3.1 below)
2. Finiteness of the array	→	Truncation Effects (→ B.3.2 below)
3. Restriction to a line loudspeaker array in the horizontal plane instead of a planar array (reproduction area: 3D→2D)	→	Amplitude Errors, Localisation restricted to horizontal plane (→ B.3.3 below)

Constraints of WFS in practice and their consequences

These constraints and their consequences are described in the relevant literature. Various authors have suggested methods to deal with these problems or minimize their negative effects.

B.3.1. Spatial Aliasing

"Spatial Aliasing" is an effect that is responsible for both spatial and spectral errors of the synthesized sound field. Spatial Aliasing occurs above a certain maximum frequency which, which is known as the "Spatial Aliasing Frequency ($\mathbf{f_{alias}}$)" (see Figure 5 for an illustration of an aliased wave field). $\mathbf{f_{alias}}$ is determined by the time difference between two successive loudspeaker signals interfering at the listener's position. This time difference depends on the spatial sampling interval, i.e. the loudspeaker/microphone interspacing. Moreover, the maximum wavelength being sampled correctly without Spatial Aliasing occurring depends on the maximum *angle* on the microphone side, as described by Sonke (2000). This interrelationship is illustrated in Figure 6. Accordingly, the maximum wavelength being *received* correctly without Spatial Aliasing occurring depends also on the maximal angle on the receiver side. A correct and consistent declaration of the Spatial Aliasing frequency $\mathbf{f_{alias}}$ is the basis for the comparison of different arrays and the experimental results derived from them. Therefore, by Figure 7, the definition of the relevant $\mathbf{f_{alias}}$ is given at the *receiver position*, meaning that it will describe the actual Spatial Aliasing perceived by the listener. It differs from the definition of the relevant $\mathbf{f_{alias}}$ for the sampled sound field, in which θ^{sec} equals 90° (e.g. Sonke, 2000). That is the reason for differing declarations of $\mathbf{f_{alias}}$ in literature.

Different proposed methods exist to avoid or minimize Spatial Aliasing. As the *perceptual* consequence of Spatial Aliasing is not yet clear, none of these methods is standard practice. The box below gives an overview of some methods:

Spatial Bandwidth Reduction
De Vries et al. (1994) and Start (1997) suggest to minimize the maximum angle on the sampling side (which increases $\mathbf{f_{alias}}$) by applying with frequency increasing directivities to the primary (virtual) sources. Another technique described by these authors works similarly: In order to minimize the maximum angle on the receiver side (which increases $\mathbf{f_{alias}}$ as well) similar directivity behaviour could be applied to the secondary sources (i.e. applying special array loudspeakers). However, these are techniques of simply omitting signal contributions which would cause errors – inevitably leading to a loss of (spatial) information.

Randomisation of high-frequency content
Start (1997) tried to avoid the audible periodicity of the Aliasing artefacts (see Figure 5) and by these means to reduce the quantity as well as the perceptibility of Spatial Aliasing. He realized this by randomising the time-offsets of the high-frequency content. In this way the sound field looses its spatial properties above a certain frequency. This method, however, is complicated and also has perceptual disadvantages, as found out in experiments by Start (→ chapter D.3 below).

"OPSI" – Phantom source reproduction of high-frequency content
This method is called OPSI ("Optimised Phantom Source Imaging of the High Frequency Content in WFS") and it is a proposal of a hybrid WFS + Phantom source reproduction. Wittek (2002) tries to avoid Spatial Aliasing through the omission of the WFS reproduction of the high-frequency content. Instead he proposed to reproduce it through conventional phantom sources which are e.g. created by a few loudspeakers within the array. WFS is applied only below $\mathbf{f_{alias}}$ leading to a perfect reproduction of the wave front. The perceived directions of the WFS and the phantom source part of the virtual source can be matched sufficiently in a large listening area as shown in experiments (Wittek, 2002). There are no negative consequences for the perceived localisation accuracy of the virtual source with respect to normal WFS as found in listening tests (Huber, 2002 for details see chapter D.3 below).

Overview of Methods to minimize Spatial Aliasing

B.3.2. Truncation Effects

In theory, the synthesis of the wave field arises from the summation of an infinite number of loudspeaker signals. In practice, however, the loudspeaker array used will always have a finite length. The finite array can be seen as a window, through which the primary (virtual) source is either visible, or invisible, to the listener. Hence, an area exists which is "illuminated" by the virtual source, together with a corresponding "shadow" area (Sonke, 2000). Applying this analogy, diffraction waves originate from the edges of the finite loudspeaker array. These error contributions appear as after-echoes (and pre-echoes respectively for focussed sources), as can be seen from Figure 8, and – depending on their level and time-offset at the receiver's location – may give rise to colouration.

A reduction of these truncation effects, for the mentioned reasons also referred to as "diffraction effects", can be achieved by applying a so-called tapering window to the array signals. This means that a decreasing weight is given to the loudspeakers near the edges of the array. In this way the amount of diffraction effects is substantially reduced, however, at the cost of a reduction of the listening area. For details see Boone et al. (1995) and Sonke (2000).

De Vries et al. (1994) depict an alternative solution to deal with diffraction effects: After approximating of the diffraction contributions on a fixed reference position, these *"can be interpreted as scaled point sources with a specific directivity pattern radiating"* (de Vries et al., 1994) from the edges of the array. Hence, the compensation of these error signals is possible, albeit leading to an erasure only for the reference position. One important disadvantage is the accompanying introduction of even stronger colouration outside the listening area.

B.3.3. 3D → 2D

Theory does not restrict WFS to the horizontal plane. Komiyama et al. (1991) and Ono et al. (1997, see Figure 9) actually built their "loudspeaker wall" in two array dimensions. In practice, too few convincing arguments exist for a WFS array to be installed in two dimensions. However, this reduction of the array dimension to a line and the synthesis dimensions to the horizontal plane does have two main consequences.

First, only virtual sources within the horizontal plane can be synthesized. Conventional WFS is capable of creating the correct directional localisation only for sources within the horizontal plane. This can be explained by the "array as a window" analogy of subchapter B.3.2 above. An easy solution to overcome this restriction is to apply different techniques (e.g. single elevated stereophonic loudspeakers similar to existing stereo formats like 10.2 or transaural signals) to satisfy the less sensitive localisation capabilities of the auditory system in the median plane (for localisation in the median plane see e.g. Blauert, 1997, p.41 and p.44).

For listeners outside the horizontal plane, the wave field of a WFS linear loudspeaker array will be distorted in terms of localisation. This is presumably less disturbing than the directional distortions in the reverse case of Wave Field Synthesis, which is Wave Field Analysis (→ chapter B.4 below). Now, contributions from other directions will misleadingly be considered as coming from the horizontal plane, which will cause errors.

Second, one must be aware of the fact that in WFS (conventional 2D-case) no real spherical waves, but waves with some cylindrical components are created. This can be understood when the reproduction of a plane wave in WFS is observed: In the display of the horizontal section (Figure 10a) the plane wave seems to be perfect, whereas in the vertical section (Figure 10b) the cylindrical waveform is represented as a circular waveform emitted from the array. Arising from this, the main difference for the plane wave in the horizontal plane is the increased level roll-off (3dB/doubling of distance) in comparison with the ideal plane wave (no roll-off).
These "Amplitude errors" or "Spatial Decay errors" are described and quantified by Sonke et al. (1998) and Sonke (2000). Figure 11 illustrates the deviation of real and desired sound fields. Sonke depicts methods to handle and reduce these errors, for example by applying secondary line sources instead of point sources (loudspeakers) for remote source positions.

Boone et al. (1999) depict solutions for the special case of the improvement of the spatial amplitude decay for virtual Surround Sound reproduction. Sound reinforcement systems in concert halls are considered by Start (1997). He studied the effect of conflicting primary source (e.g. an actor on the stage) and notional source (created by a WFS array) positions (Start, 1997 p.117ff).

B.4. Wave Field Analysis and Auralisation

The reverse case of Wave Field Synthesis is Wave Field Analysis (WFA). It is closely related to Wave Field Synthesis because it shares the same principles and algorithms. Berkhout et al. (1997) and de Vries et al. (1996) give an introduction into special techniques for the analysis of sound fields (e.g. the "Radon transformation": transformation from time/*offset*-domain to time/*direction*-domain, i.e. a directional analysis of a sound field). Linear microphone arrays were proposed to "scan" the sound field.

Hulsebos et al. (2002,1 and 2002,2) optimised this technique by using circular microphone arrays, reducing the disadvantages of truncation effects. The so-called "plane-wave decomposition" is proposed as an (microphone and loudspeaker) array-design-independent format to describe a wave field. This is a powerful tool not only for the analysis of a wave field, with the help of which one is able to e.g. determine level, spectrum and direction of certain reflections, but also as a tool to "auralise" a sound field.

The technique of "Auralisation" aims to reproduce the acoustics of a room at a remote location. The general process of Auralisation is illustrated in Figure 12. The first step is the acquisition of a set of impulse responses through a Wave Field Measurement in the desired room. As microphone and loudspeaker positions in general do not agree with each other, extrapolation steps (described in Berkhout, 1988) or transformations like "plane-wave decomposition" are necessary. Subsequently, the sound field can be re-synthesized using the extrapolated filters and applying the relevant WFS operators. The temporal and spatial properties of the synthesized sound field are very similar to the original sound field with regard to the horizontal plane. Hulsebos (2001 and 2002,2) also investigated ways of reducing the necessary data amounts for these applications. He proposes spatial data reduction for the diffuse field (e.g. the reduction of the spatial resolution for the reverb tail) and a parameterisation of the obtained impulse responses.

B.5. Further and Associated Topics

De-Reverberation, Room-Compensation

The use of WFS as a tool to manipulate the reproduction room acoustics is a subject of current research. "De-Reverberation", or the compensation for discrete, annoying reflections still requires techniques to be developed. In principle, WFS could be the ultimate tool for these applications, as it offers the possibility to synthesize disturbing contributions very accurately and thus is predestined to cancel them. Corteel and Nicol (2003) and Spors et al. (2003) discuss the possibilities of WFS in principle for this task.

Multichannel Equalization, DML loudspeakers

WFS offers control of the sound field over a wide listening area. A method to equalize over a two-dimensional area is to apply a multichannel adaptive algorithm, which makes use of microphone array measurements within the listening area, trying to adaptively control the loudspeaker driving functions in order to achieve the desired result. Corteel et al. (2002) developed and applied an algorithm to WFS with flat-panel loudspeakers. These flat-panel loudspeakers, the so-called "DML" (Distributed Mode Loudspeakers) or "MAP" (Multi Actuator Panels) require these equalization algorithms anyway because of their complex behaviour regarding both spectrum and directivity.

The use of DML panels for WFS has been discussed by Boone and de Bruijn (2000) and Corteel et al. (2002) and is being further investigated by various other institutes at the moment. Both authors point out the use of the Multi-Exciter DML panels for WFS, because of their practicability (compared with many single loudspeaker chassis) and the encouraging experimental results.

Recording Techniques for WFS

The use of WFS as a sound reproduction tool for Sound Engineers has been under investigation for several years. Generally, it is possible to simply reproduce virtual Stereo[1] loudspeakers and thus be compatible with existing formats, albeit without significant sound quality enhancement. Boone et al. (1995) and de Bruijn et al. (1998) have already depicted some possibilities to combine point sources reproducing the direct sound of a source with plane waves reproducing the acoustics. Sonke (1997) investigates reverb reproduction though plane waves.

Theile et al. (2002, 2003) introduce a technique called "VPS" (Virtual Panning Spots), enabling the use of stereophonic methods for WFS. By applying this method, a flexible and scalable sound quality enhancement in comparison to the Stereo formats is possible. Reisinger, M. (2002) and Reisinger G. (2003) explored this method for use in practical recording situations and also investigated a flexible recoding technique for atmo and location sound.

---

[1] By the term "Stereo (-phony)" a sound reproduction technique that creates phantom sources through few loudspeakers is meant. This terminology is used further on as well.

Ambisonics

As shown in Daniel et al. (2003), WFS and Ambisonics (for more about Ambisonics see Daniel et al, 2003) are two similar types of sound field reconstruction. Though they are based on different representations of the sound field - the Kirchhoff-Helmholtz Integral for WFS and the Spherical Harmonic Expansion for Ambisonics – their aim is congruent and their properties are alike. Daniel et al. analysed the existing artefacts of both principles and – for a circular setup of array loudspeakers – came to the conclusion that HOA (Higher Order Ambisonics), more exact near-field-corrected HOA and WFS "*meet similar limitations*".
However, an important drawback of HOA is the need for a circular setup.

C. Methods for the measurement of perceptual properties

A distinct terminology is required for an investigation into perceptual properties and a validation of existing studies. Previously existing attributes will be organized and redefined.

Some techniques of measuring certain perceptual attributes are discussed, in order to assess their validity and compare them with each other. This enables an objective summary of existing studies as presented in chapter D.

C.1. Classification of perceptual attributes

This study concentrates on the properties of a WFS system regarding the spatial reproduction. Its particular interest is in the perceptual consequences of different physical aspects. Therefore, an underlying reservoir of perceptual attributes is needed which fits the purpose of this study. It is necessary to organize these attributes into certain categories. These categories facilitate a global overview of the diverse properties, and enable a structured proceeding.

It is a difficult task to find an adequate structure of attributes as well as adequate and unambiguous attributes themselves in literature. The proposals found do not offer a proposal for a structure of attributes which is suitable (e.g. too much stress on timbral aspects as in loudspeaker evaluation) or which is complete enough for this investigation. Existing sets are the sound quality scales of Gabrielsson et al. (1985) or the terminology of spatial sound attributes of Rumsey (2002).

Therefore, a categorisation (see Table 1) is applied that allows the separation of different perceptual dimensions. The categorisation is led by the need for a description of the spatial properties of a sound reproduction system. It is designed by applying the hypothesis that the three dimensions depicted in Table 1 (Localisation, Source content, Environment) are independent concerning their *perceptual* classification. This means the perception of a sound source is based on the following independent steps (listed in arbitrary order):

- Perception of the location and dimension of the source itself
- Perception of the source content, the signal itself
- Perception of the environment

The separation of "Source"- and "Environmental"-related attributes can be found in Rumsey (2002) as well. Rumsey applies an acoustic scene-based approach that subdivides into individual-source-related, source-ensemble-related and environmental properties. The distinct separation between individual source and source ensemble is discarded in this categorisation to simplify matters.

The additional separation of the source content was derived from the conviction that the auditory system separately analyses the source content (the "Gestalt") after detecting the source location. This was proposed by Theile (1980). The environmental attributes could be separated in the same sense into dimension-related and timbral attributes. This is discarded here for simplicity.

Localisation	Signal Content, "Gestalt"	Environment, "Room"
Geometry: Direction Distance * Width ***Quality:*** *Focus* *Locatedness* *Stability, Robustness* *Externalisation,* etc	Loudness Sound Colour Familiarness, Plausibility etc	Depth Room dimensions Envelopment Presence Naturalness of the room Room timbre etc

Table 1: Three Classes of attributes of sound source perception

The attribute "Distance" is a source attribute that cannot exist without a room. In spite of that, it is considered no attribute of the room itself.

There are a lot of terms belonging to the first class, the *localisation* of sound sources. Often they are individually defined for a certain investigation or sometimes even remain diffuse in their meaning. The consequence is a lack of consistence between the different definitions (or applied meanings) and a significant difficulty in comparing different results. Some terms, as they are used in literature, can have different meanings. This discussion makes clear that there is a certain need for a distinct investigation into the terminology of localisation.
The relevant terms found in literature are listed in Table 2 below together with definitions by this author unless otherwise noted. Attributes written in *Italics* are not used as localisation attributes in this paper further on.

Localisation	General mapping law between the location of an auditory event and a certain attribute of the sound source. Definition according to Blauert (1997)
Direction	The direction in which the source is perceived
Distance	Perceived range between listener and reproduced source Definition according to Rumsey (2002)
Depth	*Sense of perspective in the reproduced scene as a whole* *Definition according to Rumsey's "scene depth" (2002),* *- belongs to the environmental properties -*
Stability	The degree to which the perceived location of a source changes with time. Sub-attribute of Direction and Distance

Robustness	The degree to which the perceived location of a source changes with movement of the listener. Sub-attribute of Direction and Distance
Accuracy	The degree to which the desired and the actually perceived source agree with each other. This "agreement", unless defined differently involves all attributes of the source. Often the term accuracy is used for only the "directional accuracy" which means the agreement concerning the sound source direction. The relevant numerical measure for this is the "Directional Error" of a source/system.
Resolution	The achievable precision of the synthesized sound field in terms of direction and/or distance. Sub-attribute of Direction and/or Distance
Individual Source Width ISW, Apparent Source Width ASW	Perceived width of the source Definitions according to Rumsey (2002) and Griesinger (2001)
(Image) Focus	The degree to which the energy of the perceived source is focussed in one point.
Diffuseness	*Inverse of Image Focus*
Blur	*Inverse of Image Focus*
Locatedness	The degree to which an auditory event can be said to be clearly in a particular location. Definition according to Blauert (1997)
Certainty of Source localisation	*Similar to "Locatedness", used by Lund (2000)*
Localisation quality, Localisation performance	*These attributes describe a mix of attributes. They are about the* overall *performance of a certain localisation. They should be defined individually, because they can have ambiguous meanings ("quality" of accuracy, sound colour, focus or an "average" quality?).*
Externalisation	The degree to which the auditory event is out of the head
Spaciousness	*- belongs to the environmental properties -*

Table 2: Collection of potential localisation attributes as found in literature, with definitions by this author unless otherwise noted

C.2. Experimental and statistical methods for the assessment of localisation attributes

C.2.1. Dispersion measurement equals quality measurement?

The experiments of Vogel (1993), Start (1997) and Verheijen (1998) gained in exploring the localisation properties of WFS virtual sources. The Mean Run standard deviation $<\bar{s}>$ of the perceived auditory event directions serves as a measure for the "overall localisation quality" of the systems[2]. This procedure may be regarded as valid if it is done with respect to a reference, a single loudspeaker, having small source width, sharp focus and good locatedness by definition. One or more of these three attributes are expected to change when the standard deviation increases, and a change in one or more of these attributes can be interpreted as a decrease of the overall quality of the localisation. However, there are two important objections to this method:

1. It cannot be judged which of those attributes changed when a certain standard deviation is measured. It is believed (see e.g. Corey et al., 2002 and Rumsey, 2002) that there can in fact be a difference between the perception of source width, focus and locatedness, and this is believed to be true in particular in the case of WFS by this author.
2. It may be possible to judge from a change in the standard deviation on a change in the overall "localisation quality", but the reverse is not proven: An existing change of the "localisation quality" does not necessarily lead to a change of the measured standard deviation. This was observed in the investigations of this author (can be found in Huber, 2002 and Wittek, 2001). Vogel, Start and Verheijen, however, deduced the localisation quality from the measured standard deviations only and therefore came to different results.

The same problem occurs for measurements of the Minimal Audible Angle (MAA).
By mathematical analysis, two other figures can be extracted (see Hartmann, 1983) from the data derived from a measurement of the perceived auditory event directions in the case of the existence of a predefined reference direction (e.g. the desired direction of a virtual source or the actual position of a single loudspeaker): The RMS error *D* is the RMS average of the deviation of the perceived direction from the reference direction. It is quite similar to the standard deviation *s* except for the reference from which the deviation is measured. The RMS error *D* references to a predefined direction in contrast to the standard deviation *s* which references to the mean value of all perceived directions. Hartmann takes the Mean Run RMS error $<\bar{D}>$ for the most suitable parameter to describe the "localisation performance". Start takes over this definition in his analyses. The mentioned problem of the standard deviation applies to the RMS error even more so: The reason for an increasing RMS error may be found in a changed focus, width, locatedness or direction of the perceived source. Thus, this parameter can describe only the "overall localisation performance" of a system that has to be accurate both in direction and shape and quality of the (virtual) source. Imagine a system which can synthesize the desired directions only with a bias of +/-5°, but is capable of presenting the sources with very

---

[2] The standard deviation s here is defined as the deviation of all assessments of one person and one stimulus. By averaging the standard deviations from all test items the Run standard deviation $\bar{s}$ is calculated. Averaging all test subjects' Run standard deviations $\bar{s}$ results in the Mean Run standard deviation $<\bar{s}>$

good imaging characteristics, thus leading to a standard deviation of less than 1°. The RMS error of this measurement would be bigger than 5° in spite of its sharp image.

The second available measure is the signed error *E,* which is a measure for the average deviation of the perceived direction to the predefined direction, in which the sign of the deviation is taken into account. This is a measure for the "directional accuracy" of a system.

Summarizing, the bias of the perceived directions in comparison to a reference direction is indicated by these two additional statistical measures. It is known that the measurements of the system bias require a larger amount of participants to be able to separate the individual bias of the test participants from the system bias.

C.2.2. Measurement methods for Localisation Focus and Locatedness

Various authors describe measurements of the localisation focus.
Martin et al. (1999) presented stimuli pairs requiring the test subjects to indicate the more focussed of the two stimuli. In his definition, the focus of a (phantom) source is dependant on the *expected* image size (in this case the human voice). Martin states: "*When a phantom image is larger or wider than the anticipated size of the actual sound source [...] the image is perceived as being unfocussed."* This definition makes clear that the focus of a source does have a clear relationship to its width, but not in a direct sense. That means that large sources can exist which are not perceived as being unfocussed and vice versa. Martin's results showed quite clear distinctions between the five different systems under investigation in terms of the assessed focus of the sources. They also performed IACC (Interaural Cross Correlation) measurements of the same stimuli using a dummy head, which did *not* reveal these distinctions. Wittek (2001), in his investigations, presented stimuli (phantom sources) in comparison to a reference, this being a single loudspeaker. The subjects were asked to assess the difference in the focus using a five-grade scale. He came to quite clear results that could *not* be concluded simply from the presumed differences in the spread of the collected directional data. The subjective differences showed a clear trend whereas the measured standard deviations showed no significant differences. (For further details, see Wittek, 2001). His definition of focus can be regarded as similar to Martin's (Wittek similarly used a human voice as the stimulus).

Huber (2002) measured the locatedness[3] of virtual sources of WFS using a five-grade scale without reference and anchor. The test subjects were required to assess how easy a source could be localized. Once again, it turned out that just by measuring the spread of the directional data the clear deviations between the systems would not have been captured (for results of this investigation see chapter D.3 below). Lund (2000) introduced a "Consistency scale" consisting of five grades and being described by three attributes at once: "Certainty of angle", "Robustness" and "Diffusion". According to his scale the best validation would get a source which is localized with "no doubt", very robust and has no diffusion. Corey et al. (2002) invokes Lund's scale and measured the "Certainty of the source location" on a five-grade scale. He additionally measured the incidence direction of the stimulus (phantom source) and the time in which the response was given. By this procedure, these different parameters could be compared to each other. It was found that there was indeed a negative correlation between response time and the certainty of

[3] *in the sense of the attribute definition of Table 2. Huber named it "Lokalisationsqualität" = "localisation quality".*

source location. However, regarding the spread and the bias of the directional data that he calls "accuracy", he states: *"In comparing the localisation accuracy with certainty, it was found that there was not a significant correlation between the variables. From this we can conclude that confidence in source location does not always translate into accurate or consistent localisation ability"*.

Start (1997) investigated the "spaciousness" of virtual sources of WFS with an experiment in which he asked the test subjects to assess the relative *width* of the presented stimuli in comparison to reference. It is mentioned here because he compared his results with objective measurements of the IACC and found some agreements. In contrast to the findings of Martin (1999) above, his subjective results where yielded from headphone listening of dummy head recordings in the anechoic chamber. That points to a possible drawback in his experiment if it is assumed that in this way natural listening conditions cannot be simulated.

D. Perceptual Properties of WFS

Wave Field Synthesis is a significant step forward from stereophonic sound reproduction. It offers a noticeable enhancement of a sound field's spatial properties. Nevertheless, there seems to be a broad unclarity about what the perceptual benefits of WFS actually are. As a consequence, these benefits of WFS may be underestimated, or more likely overestimated. Berkhout (1988) asserted: "*As holographically reconstructed sound fields cannot be distinguished from true sound fields, it is argued that holographic sound systems are the ultimate in sound control.*" This possibly optimistic assertion refers to the enhanced possibilities of WFS to simulate the true acoustics of a room. A similar statement is given in Brix et al. (2001) where the capabilities of WFS are described as follows: "*WFS permits the generation of sound fields, which fill nearly the whole reproduction room with correct localisation and spatial impression*".

Of course, these (typical) statements are not descriptions of distinct attributes of spatial sound that might be characteristic for WFS. Rather, they are complex observations of its performance in comparison to other techniques. The lack of distinct and scientifically approved descriptions of the perceptual properties of WFS causes misunderstandings.

The need to detect and describe the potential of WFS including both its advantages and shortcomings is apparent. Moreover, it should be described clearly by means of suitable physical and psycho-acoustical attributes. The description of WFS on the physical side is at an advanced stage. Investigations into the perceptual properties of WFS, however, have thus far been performed less often and thoroughly.

In subchapter D.1 the basic principles of sound reproduction are depicted. With that a classification of WFS and its properties might be enabled in the future. Subchapters D.2, D.3, D.4 and D.5 discuss certain attributes and their existence in WFS literature.

D.1. Fundamental Psycho-acoustical Principles of Sound Reproduction

A useful organization of the different existing principles of perception with regard to the reproduction of sound is given in Theile et al. (2002, 2003). It is stated that the reproduction of a sound event can be done using three fundamentally different techniques. Every sound reproduction technique can be traced back to one of these or a combination of these. These three techniques are:

1. Loudspeaker Stereophony
2. Reconstruction of the ear signals (Binaural/Transaural audio)
3. Synthesis of the sound field (Ambisonics/WFS)

Theile hypothesizes that these principals fundamentally differ from each other.
The difference between principles 1 and 3 in particular is not evident. Therefore the principles are described here and underlined with the relevant literature.

Table 3 summarizes the below-mentioned comparisons between the different sound reproduction principles. These hypothesized statements are partially taken from literature, the sources are described in the following sub-chapters.

Parameter	**Stereophony**	**Binaural**	**Sound Field Synthesis (WFS)**
Reproduction means	Few loudspeakers	Headphones *Transaural: few loudspeakers*	Loudspeaker array
Type of reproduced source	"Phantom source"	"Virtual source"	"Virtual source"
Reproduction principle	Psycho-acoustic merge of few correlated sources	Reproduction of the ear signals	Physical synthesis of many single sources
Degree to which it is possible to move within the listening area	Possible to some degree only within a small area: Phantom source image shifts towards nearer loudspeaker, phantom source levels change	Headphones: question irrelevant *Transaural: Very low degree. Binaural illusion collapses, localisation and sound colour artefacts get audible.*	High degree: Virtual source image does not change, virtual source level changes fairly realistic
Number of stable source positions in terms of direction	A limited number of source positions is reproduced stable (= only the loudspeaker positions)	Headphones: question irrelevant *Transaural: No stable source position*	Infinite number of stable source positions (including focussed sources in front of the array)
Spatial Reproduction: Direction (Localisation)	For the "sweet spot" it is possible to create phantom sources in all directions between the loudspeakers	Perfect spatial reproduction only with some constraints (see below)	It is possible to create virtual sources in all directions in front of and behind the array
Spatial Reproduction: Distance	In particular for Surround Sound: Through the reproduction of room acoustics it is possible to create phantom sources in a distance which is bigger than the loudspeaker distance	Perfect spatial reproduction only with some constraints (see below)	A distance-dependent wave front curvature is possible, for its perceptual validity see chapter D.5 below

Table 3: General Comparison between sound reproduction principles, as hypothesized and described below

D.1.1. Principle 1: Loudspeaker stereophony

Snow (1953) pointed out, regarding the basic difference between the n-channel acoustic curtain and 3-channel stereophony: *"This arrangement* (3-channel stereophony, see Figure 1) *does indeed give good auditory perspective, but what has not been generally appreciated is that conditions are now so different from the impractical <infinite screen> setup that a different hearing mechanism is used by the brain."*

In spite of that, it is generally agreed that Stereophony works due to of a physical superposition of the two loudspeakers building a phantom source. This is called "summing localisation" and has been investigated for a long time, e.g. by de Boer (1940) and Wendt (1963), for more sources see Blauert (1997). Nowadays this principle is investigated by Pulkki (1999, 2001) who searched for a similarity between the binaural signals of real and phantom sources. He found agreements but he also left questions open regarding the sound colour of phantom sources, see Ono and Pulkki (2002).

A different approach is made by Theile (1980) in his "Association model". He states that the physical superposition of the loudspeaker signals does not create localisation, but rather that each binaural signal pair emerging from a loudspeaker causes a "localisation stimulus" separately. The auditory system is able to *associate* this signal pair to a stored signal pattern that determines a distinct sound incidence direction. In the case of Stereo hearing there are two localisation stimuli that merge together to a phantom source in a psycho-acoustic process after their signal content was detected to be congruent. The congruence of the signal content ("Gestalt") is checked after an inverse filtering of the binaural signals: This filter process is made possible through the distinct assignment of the sound incidence direction, because then the HRTF are known and their influence can be removed.
There are some conclusions from Theile's theory which are important for this study:

- The perception of a phantom source and the perception on a natural source rely on *different* processes → therefore in sound reproduction theory there should be a clear separation between "phantom" source on one side and "virtual" source on the other side. (see Principle 3)
- The binaural signals created through the loudspeaker signal superposition in stereo listening only partially give information about the properties of the phantom source.

There are a number of open questions regarding the nature of the phantom source. Although many of its properties have been described in several publications so far, the general type of perception for Stereo listening is not yet clear enough. Theile's theory gives an explanation of phenomena related with the phantom source, but it is rather general and cannot measure or foresee its properties in a quantitative way.

D.1.2. Principle 2: Reconstruction of the ear signals

This type of sound reproduction aims to deliver *binaural* signals, i.e. signals that include the influence of the pinna to the listener. The advantage is an easy coupling of recording and reproduction side through the use of headphones.
These signals are obtained by dummy head recordings. The dummy head replaces the influence of the listener's own head and pinna, which causes problems because of disparities.

Certain techniques enable the replacement of dummy head and headphones respectively. The transfer functions of the dummy head (HRTF) can be taken from a database and then convolved with dry signals to simulate a binaural recording as described by Horbach et al. (1999). That may happen in real-time through a PC or a Digital Signal Processor (DSP).

The technique of Transauralisation enables the reproduction through loudspeakers. The double pinna influence is erased through inverse filtering, see Kyriakakis (2001).
Modern techniques of binaural reproduction like the BRS method, see Horbach et al. (1999), enable the simulation of head rotations through the use of head-trackers. These measure the rotation of the listener's head and so determine the corresponding HRTF with which the input signals are convolved.

D.1.3. Principle 3: Synthesis of the sound field

In contrast to the aforementioned methods, the complete sound field is synthesized in the reproduction room. That means the physical properties of real and virtual sound field agree on the main lines. This is achieved by using loudspeaker arrays which are driven in a way that they rebuild the desired sound field through superposition. This basic principle of WFS is described in detail in chapter B above.

In terms of psycho-acoustics, it is particularly important whether certain physical agreements are of importance or not and whether physical disagreements are disturbing or not. In the following chapters this question is addressed.

D.1.4. Transition between Stereo and WFS ?

Subject to the condition that stereophonic perception (D.1.1) and the perception of synthesized sound fields (D.1.3) are based on different principles, as stated, at a certain point there could be a transition between them. In other words it would depend on the array design whether the loudspeaker signals are perceived as single localisation stimuli or as a whole after the physical interference.

This hypothesis can be illustrated by an experiment of Vogel (1993, p.130ff). He conducted the very first experiments into WFS. The WFS linear (all loudspeakers in one line) array setup used for this investigation consisted of 12 loudspeakers, located 45 cm (!) from each other. This system has quite a poor performance with regard to Spatial Aliasing, which starts at $\mathbf{f}_{\mathbf{alias}}$= 380 Hz.

A very important question now arises: Is this WFS at all?

To answer this question, much more knowledge about the general way of perceiving superimposed sound fields would be necessary. Vogel commented in his experimental results, shown in Figure 14: *"[…] it can be concluded that the wave fields […] contain the desired directional information. The spatial aliasing in the simulated wave fields does not disturb this information."* The conclusion is correct, but the reasoning behind it could possibly be doubted. Vogel assumed that the correct synthesis at frequencies below $\mathbf{f}_{\mathbf{alias}}$ (in this case 380 Hz) is responsible for the correct localisation. However, the precedence effect is a cue he did not mention.

The precedence effect supports the localisation of non-focussed WFS signals because the signal from the array loudspeaker in the direction of the virtual source is always the first (and also often the strongest). This can be understood when Figure 1 is studied: The total path length of a signal from source via microphone/loudspeaker to the receiver is shortest for the microphone/loudspeaker pair that is next to the line between source and receiver. For the receiver this loudspeaker is nearest to the virtual source's direction. This simple phenomenon creates a localization cue for *all* frequencies.

This cue is hypothesized by this author to be crucial at least for the situation of Vogel's experiment described above. One could mention that a bigger directional error occurs if not the synthesized wave front – which is perfectly corresponding to the virtual source's direction – but the nearest array loudspeaker is localized. First, this is only true if not the source is actually perceived between the two array loudspeakers with the shortest path lengths similar to stereophonic localisation. Second, in the described experiment by Vogel, the loudspeaker distance was 45 cm. This system would have, if only the precedence effect would affect source localisation, a mean directional error of a quarter of the loudspeaker distance, in this case 11.25 cm. Vogel's results do not comprise a smaller mean directional error (Figure 14). Therefore, it cannot be concluded from his results that the system under investigation has a better accuracy.

The result of these considerations is: Maybe the precedence effect or stereophonic principles are more valid for the perception WFS than generally supposed. The result of Vogel's test applying *focussed* sources would probably be quite different, because in this case the first wave front would *not* arrive from the virtual source's direction leading to conflicting cues.

At a certain point there could be a transition between stereophonic perception and the perception of synthesized sound fields. In other words it would depend on the array design whether the loudspeaker signals are perceived as single localisation stimuli or as a whole after the physical interference.

D.2. Directional Accuracy

The directional accuracy of a sound reproduction system is good if it is capable of reproducing a source in a direction without too much system-caused bias.

WFS is in theory capable of creating accurate wave fronts and as a consequence accurate virtual source directions below $\mathbf{f_{alias}}$. However, WFS is not capable of synthesizing the wave fronts correctly above $\mathbf{f_{alias}}$ and this leads to an incorrect directional representation of these sound contributions. As this incorrect directional representation doesn't in theory have a constant shift across all frequencies, a certain direction bias caused by the Synthesis process probably does not exist. Instead, the various sound incident angles for different frequencies above $\mathbf{f_{alias}}$ will probably cause a decrease of locatedness and/or a blur of the virtual source (→ next chapter D.3).

Other reasons, however, may cause a certain bias, which are an influence of the reproduction room or the design of the test signal (a single sine wave certainly is biased to a distinct direction). There exists a third possibility of the presence of a certain bias of WFS: The special case when WFS does not work, i.e. there is no effective synthesis.

Then, according to the considerations of chapter D.1.4 above (about the validity of the precedence effect in WFS), the resolution of WFS could be determined by the angular resolution of the array loudspeakers themselves, in other words, the grid of the array loudspeakers is congruent with the grid of actually synthesisable source directions.

The directional accuracy as it is defined above can be measured by the signed error *E* of the data (see chapter C.2 above). By this measure, the bias of the mean perceived direction to the desired direction is indicated. Start (1997) uses the term "accuracy", defining the RMS Error *D* as an indication of the "*overall accuracy of localisation*". He comprises both bias and focus/locatedness into the term accuracy. In this way, he measures if there is any difference between the different systems, regardless whether it is bias, focus or locatedness. From his experiment's results, which found no significant difference between a broadband and a low-pass stimulus regarding the Mean Run RMS Error $<\overline{D}>$, he concluded: "*Apparently, the effect of spatial aliasing above 1.5 kHz does not degrade localisation performance for the broadband noise stimulus*." This statement has to be checked for validity because a closer look at the signed errors $<\overline{E}>$ reveals discrepancies. In two of the three experiments, the number of test participants seems to be too low to be able to extract the system bias, because the inter-subject and inter-item deviations prevail. The third experiment reveals a clear system bias which cannot be blamed on Wave Field Synthesis itself.

Huber (2002) found no detectable system bias in his experiments with linear WFS arrays of 4 cm resp. 12 cm loudspeaker interspacing. From these results, it can be concluded that the used array design enables a higher resolution of the localisation (a higher directional accuracy) than the actual resolution of the array. In chapter D.1.4 above, this fact was doubted when applying the large distance (45cm) linear array. For smaller loudspeaker distances and thus increased $\mathbf{f_{alias}}$ these doubts seem to be unjustified.

Verheijen (1998) proved in his experiment that accurate synthesis is not possible for sources which cannot be seen through the "acoustic window" (which is the array). The same holds true for focussed sources (sources in front of the array): Only those sources can be correctly synthesized which are between two lines of sight from the listener to positions near the edges of the array. This zone is further minimized by applying tapering windows (for an explanation of tapering see chapter B.3.2 above).

D.3. Focus and Locatedness

As considered in the last paragraph, the directional accuracy of WFS seems to be satisfying, even on arrays with which a clear degradation of other sound quality attributes is clearly audible. The attributes 'focus' and 'locatedness' of a sound source seem to be much more sensitive to changes in the physical composition of the sound signal. In chapter C above it was discussed how to distinguish between these two attributes and how difficult it is to measure them. The use of the relevant dispersion measures $<\overline{D}>$ and $<\overline{s}>$ was discussed as well.

Investigations of Vogel (1993), Start (1997), Verheijen (1998) and Huber (2002) can be consulted regarding these parameters. All these authors use a similar WFS linear array shape with loudspeaker distances from 11 to 12 cm.

Start, analyzing Vogel's experiment data, finds no significant difference between the stimuli broadband noise and low-passed (<1.5 kHz) noise regarding the measures $<\overline{D}>$ and $<\overline{s}>$ as already mentioned in the previous chapter. Apart from his optimistic conclusion (see above), he draws an interesting parallel to the investigations of Wightman and Kistler (1992), who proved the dominance of low-frequency ITDs (Interaural time differences) for localisation by creating conflicting ITD and ILD cues. WFS, which synthesizes correctly in the lower frequency region only, would thus create correct localisation if $\mathbf{f_{alias}}$ is sufficiently high to enable the ITD cue evaluation by the auditory system. There is no indication for where this limit should be, except for the fact that the auditory system ITD evaluation is valid up to ca. 2,5 - 5 kHz. Wightman and Kistler did not compare the *quality* of the localisation with and without the conflict of cues.

Vogel mentioned, considering an experiment with the large distances (45 cm) array using broadband noise: *„[..] the perceived source consists of a well localised low-frequency image, surrounded by a broader high frequency image [..]"* and *„As listening experiments [..] turned out, the wide frequency image using a broadband noise signal is absent for speech signals."* He explains this phenomenon with the common envelope of high and low frequencies in speech signals. Thus, he included amplitude-modulated (6 Hz) broadband noise in his experiment with the smaller-spaced (12 cm) array, expecting the same effect (but it turned out to be even worse than normal broadband noise – maybe because the limit of ITD dominance was already achieved without the modulation, here is room for further investigations).

Start further tried to evaluate the localisation characteristics by tests applying dummy head recordings in the anechoic chamber. He compared the MAAs (Minimal Audible Angle) of the sound field of real sources and different WFS arrays. According to the results of this experiment there is no difference regarding the MAA between real sources and the WFS array of 11cm loudspeaker interspacing (→$\mathbf{f_{alias}}$ = 1.5 kHz) for both broadband (<8 kHz) and low-passed (<1.5 kHz) noise signals (MAA = 0.8° for broadband and 1.1° for low-passed noise). After reducing $\mathbf{f_{alias}}$ to 750 Hz - by increasing the loudspeaker interspacing to 22 cm - the MAA increased (only 2 subjects, MAA = 1.6°). With that Start provided a first scientific argument for ca. 1.5 kHz as a lower limit for $\mathbf{f_{alias}}$.

Start repeated and expanded his experiments in situ, i.e. the anechoic chamber and two concert halls. In Figure 15 the results are illustrated for the three different rooms. Start found that "*the localisation accuracy of low-frequency noise stimuli is almost identical for synthesized [..] and real sound fields [..]. As expected, localisation performance is seriously degraded for high-frequency noise stimuli".* As depicted in chapter C above, with the standard deviation $<\overline{s}>$, an indication for a change in the localisation quality is given. It can be seen that there is indeed a significant difference between the low-pass and the high-pass condition in room a) and b). In room c), where $\mathbf{f_{alias}}$ is as poor as 750 Hz this effect vanishes. The array's performance obviously gets worse in the "real" rooms b) and c). It cannot be concluded (as Start supposes) that the decreasing $\mathbf{f_{alias}}$ is the only reason for that. There are indications in this author's experience that the undesirable reproduction room influence, i.e. the reflections caused by the array loudspeakers, causes irritations not only for depth perception but as well for localisation (because the latter is in natural listening always linked with depth perception!). These indications are supported by the experiments of Verheijen described below.

Start explored the effect of high-frequency randomisation (see chapter B.3.1 above for an explanation) within his experiments. It turned out to be counterproductive for the measure $<\bar{s}>$ as can be seen in Figure 15.

Verheijen (1998) complemented Start's experiments by comparing different virtual sources (behind and in front of the array) and different array loudspeaker interspacings (11 cm and 22 cm) in his experiments. By applying these two loudspeaker interspacings (accordingly $\mathbf{f_{alias,1}}$ =1.5 kHz and $\mathbf{f_{alias,2}}$ =0.75 kHz) he gave – after Start, as mentioned before - a second indication for the effect of a too low $\mathbf{f_{alias}}$. However, as illustrated in the left part of Figure 16, the increase of the mean standard deviation disappears for the normal listening conditions in the "reproduction room". The reason for that is probably the general decrease of the quality of WFS arrays in real rooms because of the reproduction room influence, as already mentioned in the last paragraph.
Regarding the localisation quality, the reproduction room influence seems to be crucial in comparison to the influence of spatial aliasing. This could be very important regarding the investigation into the perceptual effects of Spatial Aliasing.

Verheijen's experiments applying focussed sources were made, as Verheijen declares, omitting the frequency-equalization factor included in the WFS driving function (3dB/Octave, see chapter B.1 above) and therefore overemphasized the low frequency content of the pink noise bursts[4]. Thus the results may not be that comparable. In spite of that in Figure 16 the results of this second experiment using focussed sources are illustrated as well. From the assessments of the two subjects it may be concluded that the localisation quality is worse for focussed sources than for sources behind the array. This result is supported by the considerations of chapter D.1.4 above as well. The surprisingly good result for the high-pass condition may be a consequence of the overemphasis of the lower frequencies (starting at 2 kHz) which are not yet aliased enough to avoid correct localisation. Spatial Aliasing in the spatial domain can be regarded as a process which starts at $\mathbf{f_{alias}}$ and becomes worse for increasing frequencies. Verheijen explained it slightly more optimistic: *"Apparently, the localisation task is not hindered by the (first-arriving) aliased waves from the outer loudspeakers. Because the dense aliasing tail does not exceed a few milliseconds, an integration mechanism in the auditory system may be held responsible for the reasonable accuracy of localisation for these focussed sources."*

Recent experiments by Huber (2002) can be consulted to test some of the beforementioned findings of the Delft authors. Huber conducted listening tests in the anechoic chamber using a linear WFS array with a minimum loudspeaker distance of 4.2 cm. Applying white noise bursts, he explored the differences between different reproduction techniques including real sources (single loudspeakers), "normal" WFS with a loudspeaker interspacing of Δx=12,7 cm, an enhanced WFS system with Δx=4,2 cm, the normal WFS system driven with the OPSI technique (Wittek, 2002), see chapter B.3.1 above) and conventional 2-channel stereophony. Interestingly, and as already mentioned in chapter C above, the measure run standard deviation $<\bar{s}>$, derived from his experiments' data, would lead to the conclusion that no significant differences exist except for the phantom source (that is the result of the Wilcoxon test). This can be seen from

[4] *Verheijen did not express himself clearly regarding the stimulus: In contrast to the previous experiment in which white noise bursts had been used now pink noise was used and, moreover, the low frequencies of the pink noise (!) were boosted through omitting the equalizing*

Figure 17. The differences came to the fore with the direct request for a subjective assessment of the "localisation quality". This attribute was introduced to the test panel similar to what is defined above as locatedness. The results (see Figure 18) make clear that the use of just the dispersion measures as the standard deviation of the directional data are critical, as they do not necessarily reveal a result. These results furthermore give a first valid comparison between the used reproduction techniques. There are three main findings:

1. No reproduction technique can achieve the optimal result of the real source. However, it is shown that a further increase of $\mathbf{f_{alias}}$ beyond 3 kHz leads to a significant improvement regarding the locatedness of the virtual source. This was doubted of by the Delft authors which regarded an array with $\mathbf{f_{alias}}$>1.5 kHz as sufficient.
2. The phantom source is, concerning both the result of the dispersion measure $<\bar{s}>$ and the result of the "localisation quality assessment", clearly inferior to WFS.
3. The hybrid method OPSI, which utilizes a phantom source above $\mathbf{f_{alias}}$ instead of reproducing aliased wave fronts is proven to be equivalent to normal WFS concerning the locatedness. The directional accuracy of the OPSI method for a large listening area was proved in Wittek (2002) and Huber (2002). Hence, this method could be an alternative requiring less technical efforts and having advantages regarding the representation of sound colour by totally avoiding Spatial Aliasing.

Start (1997), as mentioned in chapter C above, investigated the "spaciousness" of the real and the synthesized wave field by comparing the width of a source in relation to a reference source through dummy head recordings. His definition of "spaciousness" is close to what is here defined as source width. Start compared the subjective assessment of the source width (stimuli pairs with a reference) and the objective measure "Interaural Cross Correlation Coefficient" (IACC) and found some (not thoroughly described) correlation. He repeated this experiment with a large-scale DSE (Direct Sound Enhancement, WFS for PA purposes) system in two real big rooms once again through dummy head recordings. He found that the width of all sources was generally much bigger than in the anechoic chamber and that also the differences between the systems vanish, which may be a direct consequence.

D.4. Sound Colour

There are different possible reasons for a degradation of the sound colour of WFS virtual sources compared to real sources.

Physical reasons:

1. Spatial Aliasing distorts the higher frequency spectrum (see chapter B.3.1 above)
2. The finiteness of the array causes distortions through Diffraction Effects (see chapter B.3.2 above)

Psycho-acoustical reasons:

3. A non-optimal localisation process could lead to a distorted perception of the sound colour
4. The synthesis itself does not work, i.e. the auditory event is not determined by the synthesized wave front

In literature, very few considerations have been made on this field. Reasons for this may be the great difficulty to measure the perceived sound colour and similar subjective attributes. One other reason can be a general agreement about the small audibility of negative effects on the sound colour in WFS. An existing bad sound colour consequently often is blamed on the cheap laboratory loudspeakers.

Regarding the perceptual effects of *Spatial Aliasing*, Start (1997) mentions the smoothing of the aliased content due to the finite resolution of the auditory system. Furthermore he explains the audibility of aliasing effects with the fact that the lower limit of aliasing in the spectrum changes quite rapidly when the listener moves. He made an experiment that measures the change of colouration that occurs when the listener (the source) moves. Through this technique, the colouration due to aliasing was clearly proven. Interestingly, the change of colouration diminishes in real rooms; even better is the result for the synthesized signals with the randomized high-frequency content (which do not contain strong aliasing).

It is not clear if the *Diffraction artefacts* cause a colouration of the signal through comb filter effects. These artefacts (after-echoes) could be regarded as reflections that do not lead to audible colouration if they are successfully detected by the auditory system.

After Theile (1980), the perception of sound colour is part of the *localisation process* of the auditory system (as part of the perception of the "source content", see chapter D.1 above). A perceived stimulus is coloured if it cannot be associated to a known sound event. Colouration is therefore also a measure for the "localisation performance" – the 'perfectness' of the audible illusion of WFS (and not only a measure for the *physical* 'perfectness' of the synthesis). Hence, it could be a more important measure as generally believed because it may contribute to the search for the way in which we localize WFS sound fields.

Regarding the fourth item on the list above, this is an untested hypothesis. It is discussed in chapter D.1.4 above. If the synthesized wave front is not actually perceived, but rather single loudspeakers, the possible perceptual consequences are different from the aforesaid items.

D.5. Distance and Depth

A sound image without depth is unnatural. In any natural sound field a sense of depth is existent, being the sense of perspective in the reproduced acoustic scene, as defined by Rumsey (2002). A sense of depth in a natural environment is given through the perception of sources at different distances, or through the analysis of the room reflections, which contain an unambiguous description of the room dimensions.

Depth therefore is a perceptual construct for the analysis of the environment – for the detection of the room dimensions together with an overview of the involved sources and their relative locations. The listener's successful perception of depth is the benchmark of a spatial audio reproduction system.

Contrary to depth, distance is an attribute of the source itself. Although depth supports the perception of the source distance, the latter can also be judged in acoustic scenes without depth (e.g. a mono loudspeaker). In this case, the perceived distance may be called "pseudo"-distance.

There are a number of relevant cues for the perception of depth and distance (→ e.g. Shinn-Cunningham, 2000, Zahorik, 1996, Nielssen, 1991). Some of them are not or only partially existent in some unnatural acoustic scenes (e.g. there are no reverberation cues in the anechoic chamber). The degree to which the relevant cues are existent is crucial for the perception of "true" distance and depth.

An interesting parallel can be drawn between the acoustic and the visual perception of depth and distance. By looking at the rather apparent visual cues, the acoustic cues can also be illustrated. The visual cues are depicted by Ausserhofer et al. (1995). The analogy is hypothesized by this author.

Figure 19 shows an image which contains several visual cues to analyse spatial depth which are for example the existence of a linear perspective, overlay (one object is covered by the other), shadow, relative size (objects appear bigger when they are closer), etc. In spite of that, the image does not contain true depth, it is a 2D representation of a 3D visual scene. Even if another important cue, the so-called "motion parallax" was existent (a corresponding change of the perspective with movements), it would only be a so-called 2½ D-representation, though enabling movements of the viewer.

A true 3D representation – and thus the perception of real visual depth - is enabled through the existence of so-called stereoscopic cues like disparity (different signals for the two eyes) or convergence (different axes angles of the two eyes). The stereoscopic cues are illustrated in Figure 20.

An immersion of the viewer is achieved not until both stereoscopic and motion parallax cues are combined, i.e. the viewer has a real sense of depth *and* can move in the environment.

This differentiation between "true" and "pseudo" depth can illuminate the potential of WFS and Stereo: While a 2D representation of a 3D acoustic scene (a mono loudspeaker) can actually contain distance cues and can give a certain information about the environment, only a 2½ D representation (WFS sound field with dry virtual sources) enables movements of the listener and thus another cue for the perception of depth.

This 2½ D representation, however, lacks the crucial cues for real 3D depth perception, which are cues related with spatially distributed reflections and reverberation. These cues, and therefore real 3D depth perception, can be produced on a fixed listening position as well – this means true depth reproduction is possible also with Stereo!

Only through the combination of motion parallax and reverberation cues is an acoustic scene containing depth and enabling movements created. Table 4 summarizes the mentioned analogies.

Against this background the discussion about the potential of WFS (and Stereo) to simulate depth and distance can be considered with two main questions:

- *Is there any cue of WFS other than motion parallax that is important for the perception of depth and distance?*

- *How important is motion parallax (which means the wave front curvature and the possibility to move) for the perception of depth and distance?*

Cue, visual	**Cue, acoustic**	**enables**	**representation**	**Corresponding spatial audio system being capable of this representation**
Overlay, linear perspective, shadow, relative size, etc.	Loudness, spectral cues, Direct/ Reverberation ratio etc.	Pseudo depth/distance, no movement	2 D representation, non-immersive	Mono
Motion parallax	Mono cues + Motion parallax	Pseudo depth/distance + movement	2½ D representation, immersive	WFS (dry sources)
Stereoscopic Cues: Disparity, convergence	Mono cues + spatially distributed Reflection/ Reverb Cues (Analysis of reflection pattern, side reflections, …)	True depth, no movement	3 D representation, non-immersive	Stereo (reproducing room acoustics)
Stereoscopic Cues and Motion parallax	Mono cues + spatially distributed Reflection/ Reverb Cues + Motion parallax	True depth + movement	3 D representation, immersive	WFS (reproducing room acoustics)

Table 4: Analogy between visual and acoustic cues for the perception of depth and distance

2½ D cues of WFS:

1. Through motion parallax (and the change of the perspective) WFS could create presence. The perspective correspondingly changes with movements of the listener within the listening area. This means: There is a way of implicitely analysing the scene geometry (and consequently distance) through moving within the listening area (see Corteel et el., 2003).
2. Through the realistic presentation of the spatial amplitude decay WFS creates "presence". The amplitudes correspondingly change with movements of the listener within the listening area. This was already mentioned by Start (1997).

3D cues of WFS:

1. WFS creates the correct shape of the curvature of the wave fronts. This is a topic under investigation. Just by creating dry correct wave fronts, it is expected that

no correct distance perception will be enabled. This point of view regarding WFS is supported by de Bruijn (2001). Literature does not support evidence for this assertion either. Only at very near distances (<1m) is this a reliable cue, as found by Brungart (1999) and Shinn-Cunningham (2000) emphasizes "drastic improvements" of the distance perception for nearby sources when reverberation is added.
However, in spite of not being a crucial cue, a correct wave front curvature (and thus a consistency between curvature and actual distance) may *support* the distance perception.

2. WFS can very accurately simulate position and level of the first reflections. WFS may provide an enhanced discriminability of reflections.

   In this author's opinion it is not proven that WFS is superior to sweet spot stereophony (i.e. stereo listening having the same distance to all loudspeakers) concerning this point. Of course the listening area in which correct (in this case distance-) perception is enabled is significantly bigger in the case of WFS. But this argument is not regarded specific for distance perception. In a study by Neher (2003) it was found that the listener cannot distinguish between stimuli which were different in terms of the direction of the early reflections.

   The enhanced possibility for the auditory system to distinguish between distinct reflections in comparison to stereophony may indeed give rise to a better spatial perception.

Boone and de Bruijn (2003) investigated speech intelligibility using a comparison between two different WFS virtual sources on the one hand, driven with two different signals (one speech signal, one broadband noise as a masker) and one loudspeaker on the other hand, driven with both signals at the same time. He found that even when both virtual sources are in one and the same direction but synthesized at different distances, the speech intelligibility threshold for the WFS virtual sources was lower (ca. 0.5 - 1 dB) than for the single loudspeaker. In this way, an enhanced segregation ability of the auditory system was measured. This is an argument for the existence of at least *any* perceptual difference between two virtual sources which only differ regarding their synthesized distance. However, it could as well be caused by the (for different virtual sources different) influence of the reproduction room or by a difference in the energy distribution within the array caused by the different distances of the sources behind the array. At least for the frequencies above $\mathbf{f_{alias}}$ this would lead to a difference in the width and the focus of the virtual source.

Some investigations, both on Stereophony and WFS, in which the distance of sources was measured, are less meaningful due to the dominance of the loudness cue: they were made keeping the *source* levels constant for all source positions and thus offering relative amplitude differences (due to the variation of the source-receiver distance) to the listener. In this way any reproduction technique, even Mono, enables distance perception as often proved in literature (see Blauert, 1997). A special capability of a spatial audio reproduction technique to create true distance perception can therefore not be measured in this way.

E. References

Ausserhofer, A. et al. (1995)

"Human Depth Perception", website, http://www2.iicm.edu/stereograms

Berkhout, A.J. (1987)

"Applied Seismic Wave theory", Elsevier, Amsterdam

Berkhout, A.J. (1988)

"A holographic approach to acoustic control", Journal of the AES, Vol.36, pp.977-995

Berkhout, A.J. and de Vries, D. (1992)

"Wave front synthesis: a new direction in electro-acoustics", Preprint No.3379, 93th AES Convention, San Francisco

Berkhout, A.J., de Vries, D. and Vogel, P. (1993)

"Acoustic control by wave field synthesis", Journal of the Acoustical Society of America, Vol.93, pp 2764-2778

Berkhout, A.J., de Vries, D. and Sonke, J.-J. (1997)

"Array technology for acoustic wave field analysis in enclosures", Journal of the Acoustical Society of America, Vol.102 pp 2757-2770

Blauert, J. (1997)

"Spatial Hearing", MIT Press, Cambridge, Massachusetts

Bleistein, N. (1984)

"Mathematical methods for wave phenomena", Academic Press Inc., Orlando

Boone, M.M., Verheijen, E.N.G. and van Tol, P.F. (1995)

"Spatial sound field reproduction by wave field synthesis", Journal of the AES, Vol. 43(12), p.1003-1012

Boone, M.M., de Bruijn, W. and Horbach, U. (1999)

"Virtual Surround speakers with wave field synthesis", Preprint No.4928, 106th AES Convention, Munich

Boone, M.M. and de Bruijn, W. (2000)

"On the applicability of DML panels for wave field synthesis based sound reproduction", Preprint No.5165, 108th AES Convention, Paris

Boone, M.M. (2001)

"Acoustic rendering with wave field synthesis", ACM Siggraph and Eurographics Campfire, 2001

Boone, M.M. and de Bruijn, W. (2003)

"Improved speech intelligibility in teleconferencing using Wave Field Synthesis", Preprint No.5800, 114th AES Convention, Amsterdam

Brix, S., Sporer, T. and Plogsties, J. (2001)

"CARROUSO - An European approach to 3D-Audio", Preprint No.5314, 110th AES Convention, Amsterdam

Brungart and Rabinowitz (1999)

"Auditory Localization of nearby sources (I, II und III)", Journal of the Acoustical Society of America, Vol.106 (3,4,6)

Corey and Woszczyk, W. (2002)

"Localization of lateral phantom images in a 5-channel system with and without simulated early reflections", Preprint, 113th AES Convention, Los Angeles

Corteel, E., Horbach, U. and Pellegrini, R. (2002)

"Multichannel Inverse Filtering of Multiexciter Distributed Mode Loudspeakers for Wave Field Synthesis", Preprint No.5611, 112th AES Convention, Munich

Corteel, E., Noguès M. and Warusfel, O. (2003)

"Monitoring Distance Effect with Wave Field Synthesis", Proc. of the 6th Int. Conference on DAFX, London

Corteel, E. and Nicol, R. (2003)

"Listening room compensation for Wave Field Synthesis. What can be done?", AES, 23rd Int. Conference, Copenhagen, Denmark

Daniel, J., Moreau, S. and Nicol, R. (2003)

"Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging", Preprint No. 5788, 114th AES Convention, Amsterdam

de Boer, K. (1940)

"Plastische Klangwiedergabe", Philips Technische Rundschau, No.5

de Bruijn, W., Piccolo, T. and Boone, M.M. (1998)

"Sound recording techniques for wave field synthesis and other multichannel sound systems", Preprint No.4690, 105th AES Convention, Amsterdam

de Bruijn, W., Rooijen and Boone, M.M. (2001)

"Recent developments on WFS for high quality spatial sound reproduction", Preprint No.5370, 110th AES Convention, Amsterdam

de Vries, D., Start, E. and Valstar (1994)

"The wave field synthesis concept applied to sound reinforcement: restrictions and solutions", Preprint No.3812, 96th AES Convention, Amsterdam

de Vries, D. (1995)

"Sound Enhancement by Wave Field Synthesis: Adaption of the synthesis operator to the loudspeaker directivity characteristics", Preprint No.3971, 98th AES Convention, Paris

de Vries, D., Berkhout, A.J. and Sonke, J.-J. (1996)

"Array technology for measurement and analysis of sound fields in enclosures", Preprint No.4266, 100th AES Convention, Kopenhagen

Gabrielsson, A and Lindström (1985)

"Perceived Sound quality of High-Fidelity Loudspeakers", Journal of the AES, Vol.33, No.1/2

Griesinger, D. (2001)

"The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique", AES, 19th Int. Conference on Surround Sound, Elmau, Germany

Hartmann, W.M. (1983)

"Localization of Sound in Rooms", Journal of the Acoustical Society of America, Vol.74(5)

Horbach, U. et al. (1999)

"Design and Applications of a Data-based Auralization System for Surround Sound", Preprint No. 4976, 106th AES Convention, Munich, Germany

Huber, T. (2002)

"Zur Lokalisation akustischer Objekte bei Wellenfeldsynthese", Master's thesis IRT/ TU Munich, Germany, www.hauptmikrofon.de/wfs.html

Hulsebos, E. and de Vries, D. (2001)

"Spatial decomposition and data reduction of sound fields measured using microphone array technology", Proceedings of the 17th ICA, Rome

Hulsebos, E., de Vries, D. and Bourdillat (2002)

"Improved microphone array for auralization of sound fields by wave field synthesis", Journal of the AES, Vol.50, No.10, pp.779-790

Hulsebos, E. (2002)

"Auralization of room acoustics by Wave Field Synthesis", Proceedings of the 22nd Tonmeistertagung, Hannover

Komiyama, S. et al. (1991)

"Distance Control System for a Sound Image", AES, 9th Int. Conference, p.233ff

Kyriakakis, C. and Lim (2001)

"Adaptive Synthesis of Immersive Audio Rendering Filters", Preprint No.5422, 111th AES Convention, New York

Lund, T. (2000)

"Enhanced Localization in 5.1 Production", Preprint No.5243, 109th AES Convention, Los Angeles

Martin, G. et al. (1999)

"Controlling Phantom Image Focus in a Multichannel Reproduction System", Preprint No.4996, 107th AES Convention, September 1999, New York

Neher, T., Brookes, T. and Rumsey, F. (2003)

"Unidimensional Simulation of the Spatial Attribute "Ensemble Depth" for training purposes, Part 1: Pilot Study Into Early Reflection Pattern Characteristics", AES, 24th Int. Conference, Banff, Canada

Nielsen, S. (1991)

"Distance Perception in Hearing", Aalborg University Press, ISBN 87-7307-447-0

Ono and Komiyama (1997)

"A Method of reproducing Concert Hall Sounds by "Loudspeaker Walls"", Preprint No.4490, 102nd AES Convention, Munich

Ono and Pulkki, V. (2002)

"Binaural Modeling of Multiple Sound Source Perception: Coloration of Wideband sound", Preprint No.5550, 112th AES Convention Munich

Pulkki, V. (1999)

"Analyzing virtual Source Attributes using a binaural auditory Model", Journal of the AES, Vol.47/4

Pulkki, V. (2001)

"Spatial Sound Generation and Perception by Amplitude Panning techniques", Helsinki University of Technology, ISBN 951-22-5531-6

Reisinger, G. (2003)

"Einsatz stereophoner Aufnahmetechniken bei Wiedergabe räumlich ausgedehnter Schallquellen mittels der Wellenfeldsynthese", Master's thesis IRT/ FH Düsseldorf, Germany, www.hauptmikrofon.de/wfs.html

Reisinger, M. (2002)

"Neue Konzepte der Tondarstellung bei Wiedergabe mittels Wellenfeldsynthese", Master's thesis IRT/ FH Düsseldorf, Germany, www.hauptmikrofon.de/wfs.html

Rumsey, F. (2002)

"Spatial Quality Evaluation for reproduced Sound: Term., Mean., and a Scene-based parad.", Journal of the AES, Vol.50, No.9

Shinn-Cunningham, B. (2000)

"Distance Cues for virtual auditory space", Proceedings of the IEEE 2000, International Symposium on Multimedia Information Processing, Sydney, Australia, p.227-230.

Snow (1953)

"Basic principles of stereophonic sound", SMPTE Journal, Vol.61, pp.567-589

Sonke, J.-J. and de Vries, D. (1997)

"Generation of diffuse reverberation by plane wave synthesis", Preprint No.4455, 102nd AES Convention, Munich

Sonke, J.-J., de Vries, D. and Labeeuw (1998)

"Variable acoustics by wave field synthesis: a closer look to amplitude effects", Preprint No.4712, 104th AES Convention, Amsterdam

Sonke, J.-J. (2000)

"Variable Acoustics by wave field synthesis", Thela Thesis, Amsterdam, Netherlands, ISBN 90-9014138-3

Spors, S., Kuntz, A. and Rabenstein, R. (2003)

"An Approach to Listening Room Compensation with Wave Field Synthesis", AES, 24th Int.Conferenceon Mutlichannel Sound, Banff, Canada

Start,E. (1997)

"Direct Sound Enhancement by Wave Field Synthesis", Dissertation, Technical University Delft, ISBN 90-9010708-8

Steinberg and Snow (1934)

"Physical Factors", Bell Systems technical Journal, Vol.132, pp.245-258,1934

Theile, G. (1980)

"Über die Lokalisation im überlagerten Schallfeld", Dissertation, Technische Universität Berlin, www.hauptmikrofon.de/wfs.html

Theile, G., Wittek, H. and Reisinger, M. (2002)

"Wellenfeldsynthese-Verfahren: Ein Weg für neue Möglichkeiten der räumlichen Tongestaltung", Proceedings of the Tonmeistertagung, 21.TMT 2002, Hannover

Theile, G., Wittek, H. and Reisinger, M. (2003)

"Potential Wavefield Synthesis Applications in the Multichannel Stereophonic World", AES, 24th Int.Conference on Mutlichannel Sound, Banff, Canada

Verheijen, E. (1998)

"Sound reproduction by Wave Field Synthesis", Dissertation, Technical University Delft

Vogel, P. (1993)

"Application of wave field synthesis in room acoustics", Dissertation, Technical University Delft

Wendt, K. (1963)

"Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie", Dissertation, Technische Hochschule Aachen

Wightman and Kistler (1992)

"The dominant role of low-frequency interaural time differences in sound localization", Journal of the Acoustical Society of America, Vol. 91(3)

Wittek, H. et al. (2001)

"Studies on Main and Room Microphone Optimization", AES, 19th Int. Conference on Surround Sound, Elmau, Germany

Wittek, H. (2002)

"OPSI", website, www.hauptmikrofon.de/wfs.html

Zahorik, P. (1996)

"Auditory distance perception: A literature review", website, http://www.waisman.wisc.edu

F. Figures:

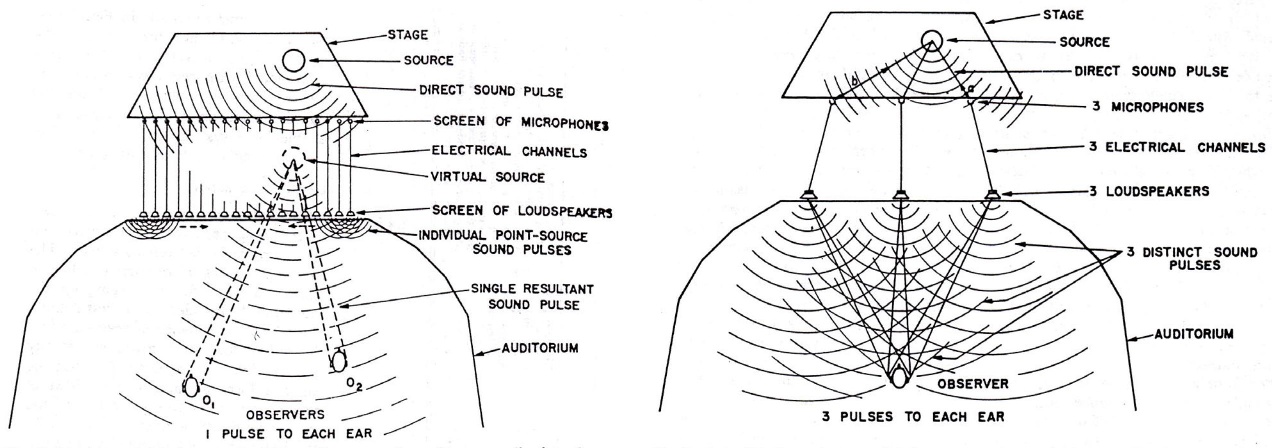


Fig. 2. Ideal stereophonic system. A very large number of very small microphones and loudspeakers would give a perfect reproduction of the original sound.

Fig. 3. Actual 3-channel stereophonic system. A practical stereophonic system gives a multiple reproduction of the original sound which the observer interprets as coming from a single source.

Figure 1:
Desired (left) and implemented (right) stereophonic system of Snow, Steinberg (taken out of Snow, 1953)

$$P(\mathbf{r},\omega)=\frac{1}{4\pi}\iint_S\left[P(\mathbf{r}_s,\omega)\frac{\partial}{\partial n}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|}\right)-\frac{\partial P(\mathbf{r}_s,\omega)}{\partial n}\frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|}\right]dS$$

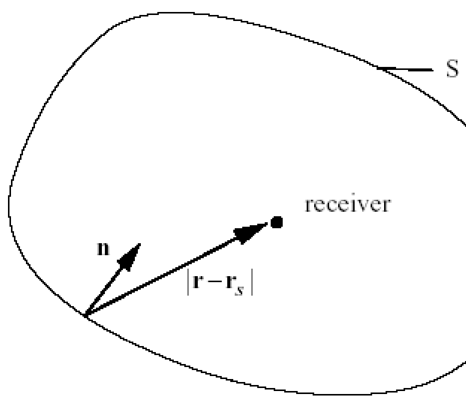


Figure 2:
Geometry for the Kirchhoff-Helmholtz Integral (taken out of Boone, 2001)

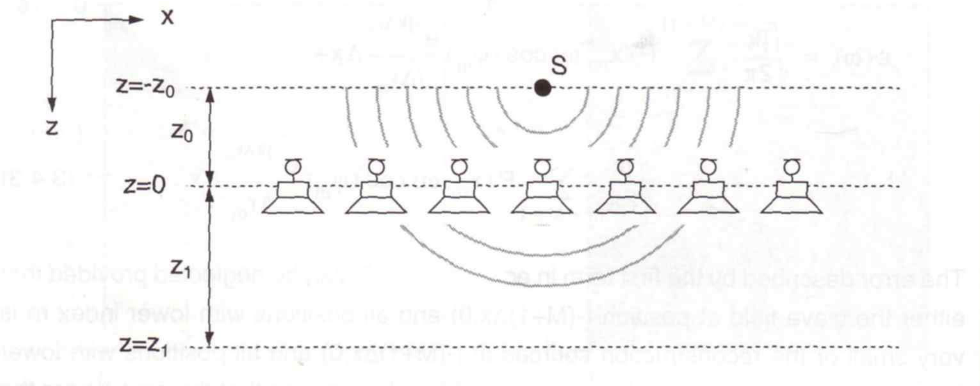


Figure 3: Basic Principle of WFS:

Sampling/ Reproduction using an "Acoustic Curtain"
(taken out of Verheijen, 1998): Sound field is sampled at n (microphone) positions, equalized and reproduced on n positions

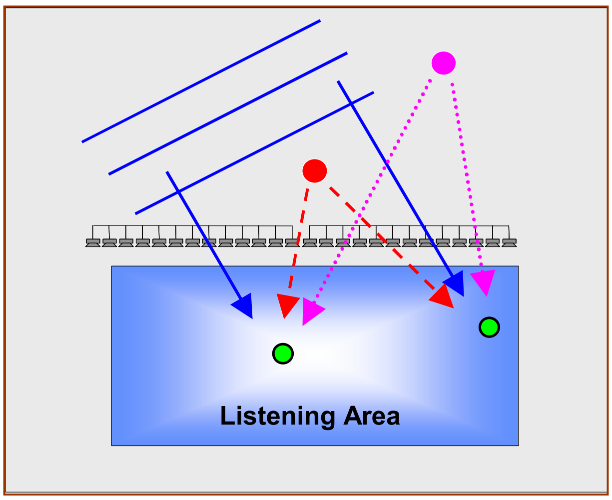


Figure 4:

Illustration of the WFS principles:
Acoustic perspective of the sound field.
Stable directions with plane waves (blue, solid) and
Stable locations with point sources (red and pink, dashed and dotted)
taken from Theile et al. (2003)

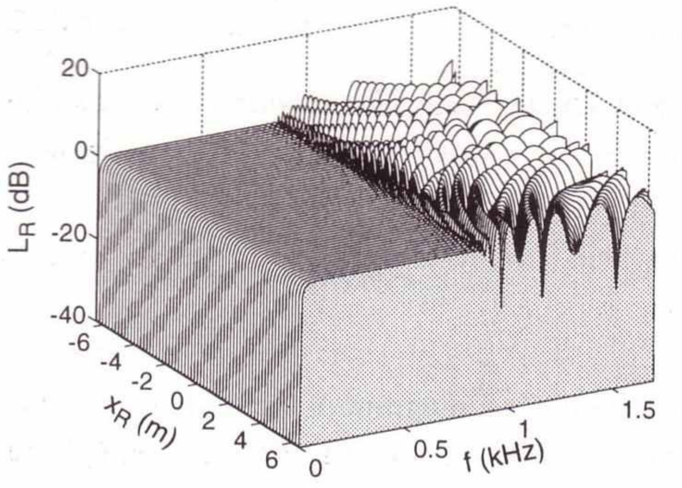


Figure 5:
Wave field with aliasing starting at approx. 1 kHz (taken out of Start, 1997). The x-axis is within the listening area and parallel to the array

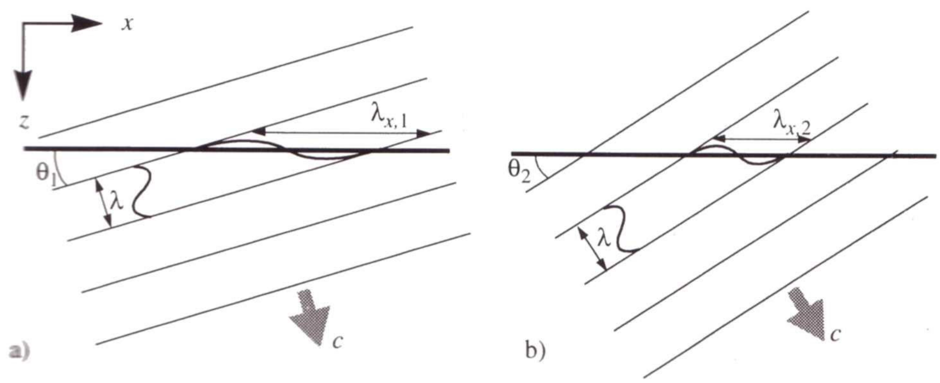


Figure 6:
Illustration of interrelationship between sampling (microphone /loudspeaker) distance and maximal wave length
(taken out of Start, 1997):

$\lambda_{x,1}$ and $\lambda_{x,2}$ are the relevant components of the wavelength λ in the array-direction x.

a) small incidence angle Φ_1: $\lambda_{x,1}$ is relatively big, would be sampled correctly

b) big incidence angle Φ_2: $\lambda_{x,2}$ is relatively small, would be sampled incorrectly

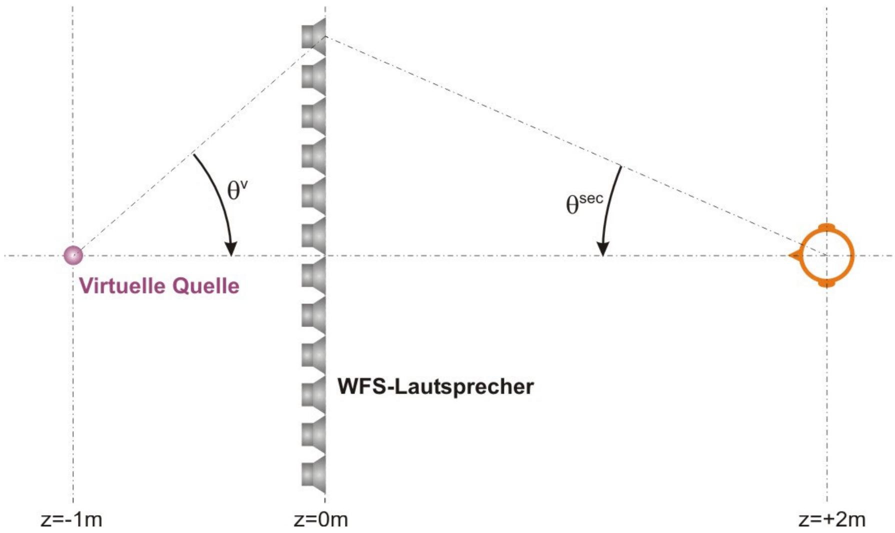


Figure 7:

Proposed calculation of the *effective* Spatial Aliasing Frequency f_{alias} (taken out of Huber, 2002)

	maximum angle on the sampling side
θ^{sec}	**maximum angle on the reproduction side**

$$f_{Alias} = \frac{c}{\Delta x \left| \sin\theta^{sec} - \sin\theta^{v} \right|}$$

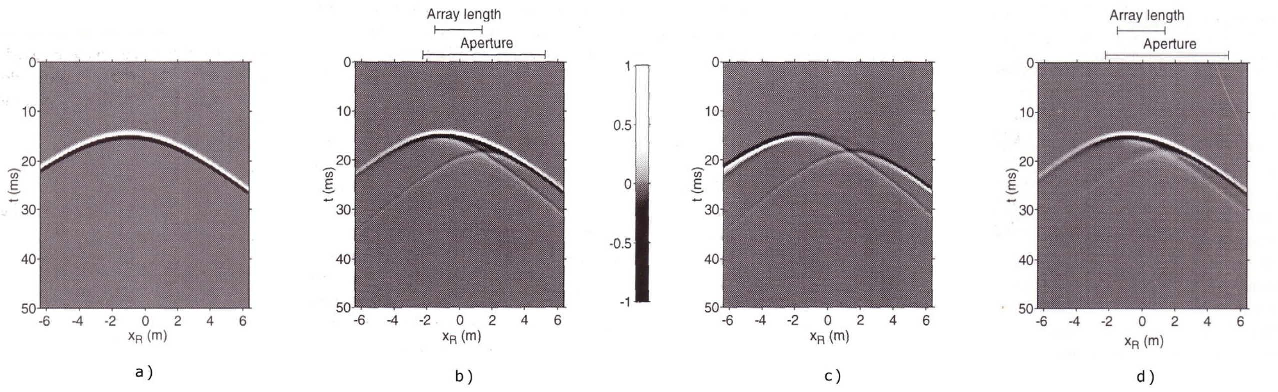


Figure 8: Influence of Array truncation: Diffraction effects (taken out of Start 1997):

- **a) Response of an infinite array**
- **b) Response of the truncated array**
- **c) Difference between a) and b)**
- **d) Response of the truncated array after tapering**

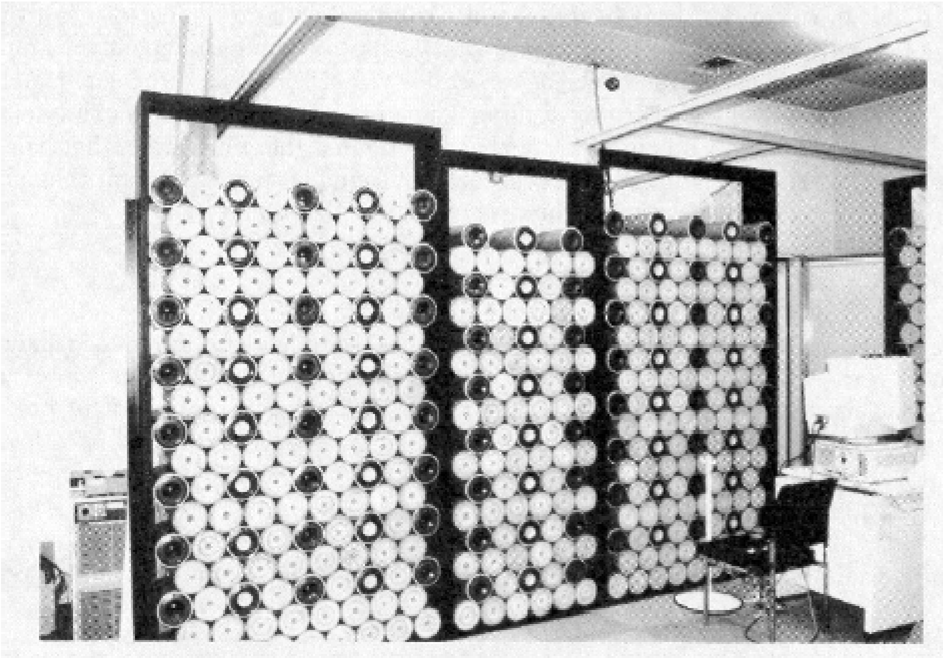


Figure 9: „Loudspeaker Wall", used for experiments by Ono et al. (1997)

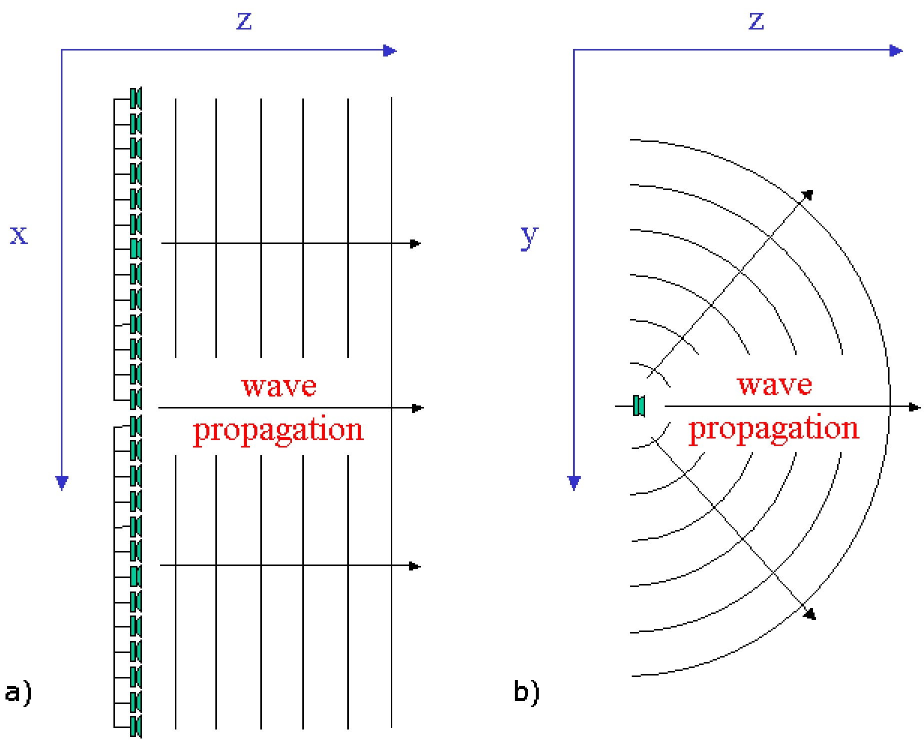


Figure 10:

Horizontal (a) and vertical (b) section of a linear WFS loudspeaker array reproducing a plane wave

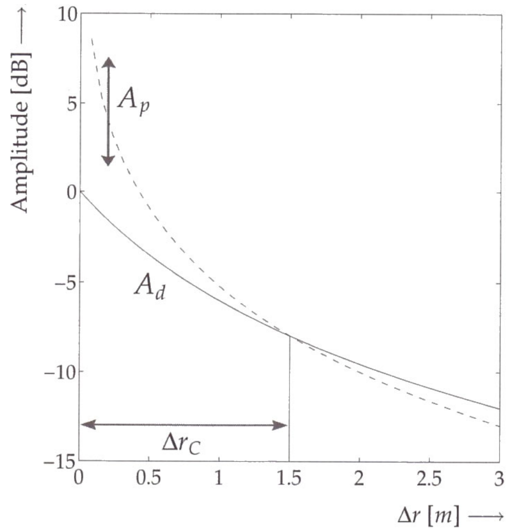


Figure 11:
Amplitude of a WFS monopole source A_p and a real, desired monopole source A_d
along a line defined by the source position at (-1,0) and an array loud-speaker position at (0,0).
The Amplitudes match at the definable optimal receiver line r_c=1.5.
(taken out of Sonke, 2000)

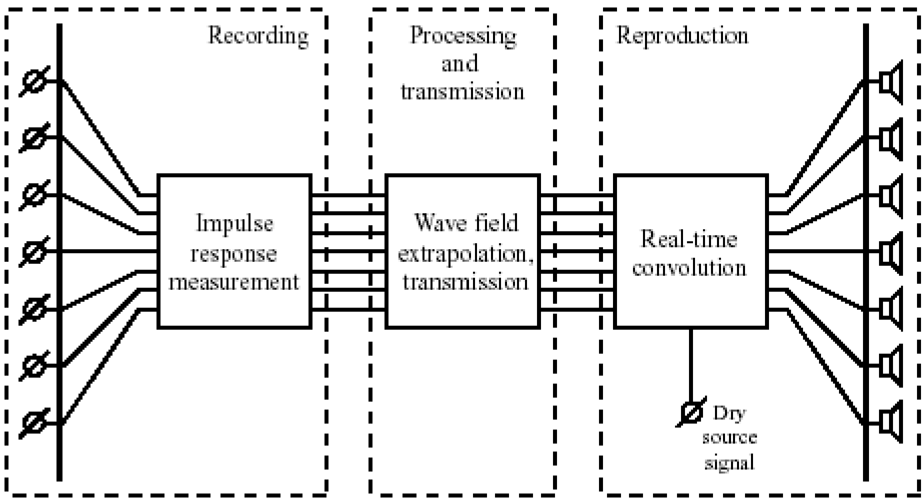


Figure 12:
Overview of the Auralization process (taken out of Hulsebos, 2001)

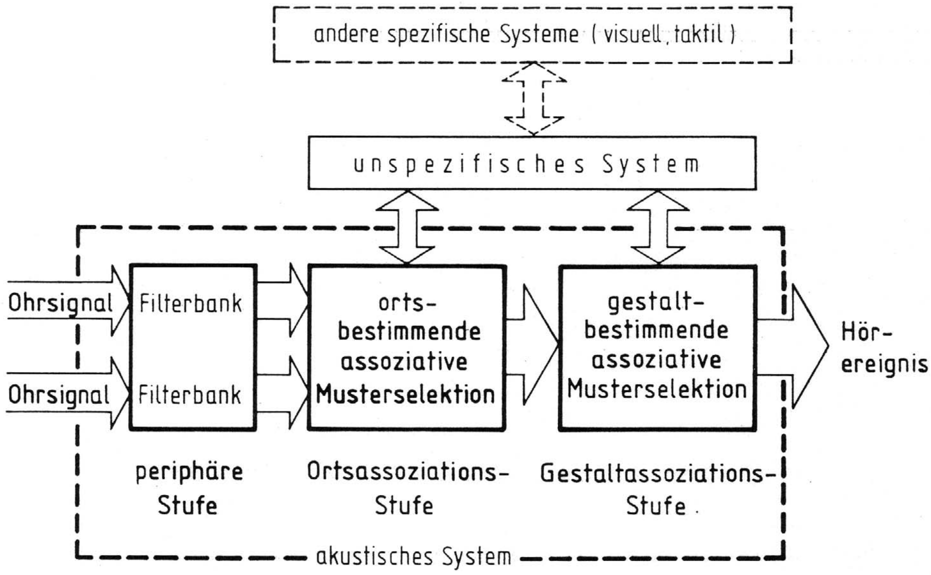


Figure 13: Overview of the "Association model" by Theile:

The two ear signals (left side of the picture) caused by a source signal are received, filtered and associated with a direction ("Ortsassoziation" = association of a source location). After an inverse filter process the obtained pure source signal is associated to a known pattern ("Gestaltassoziation" = association of the Gestalt).

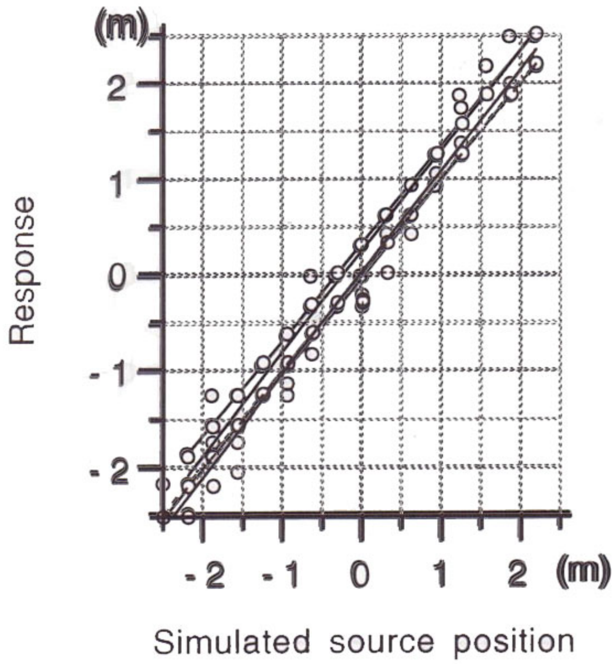


Figure 14:
Results from Vogel's experiment with his first linear array setup (The 4 single graphs, taken out of Vogel (1993), were arranged by this author so that all graphs share the same axes)

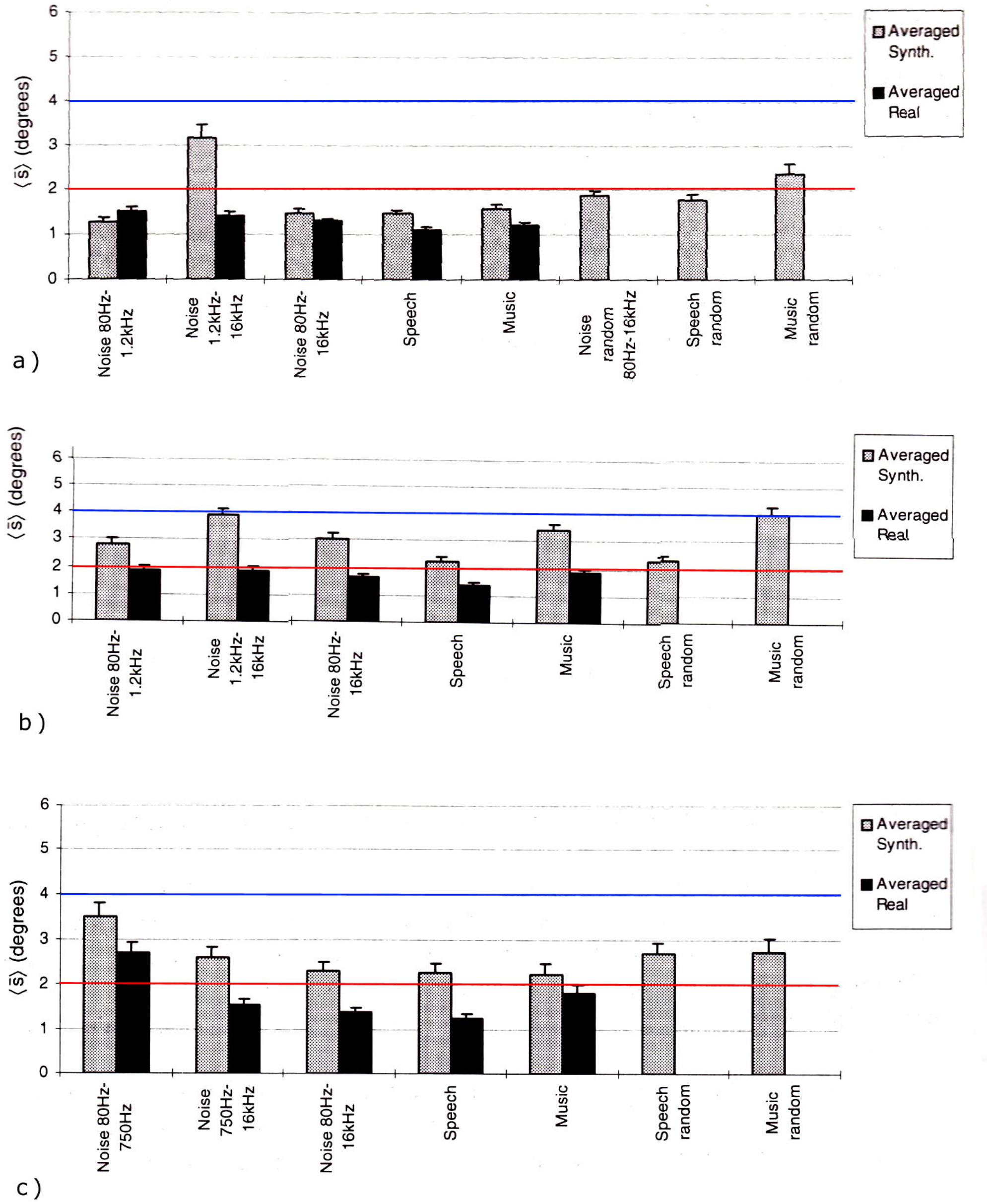


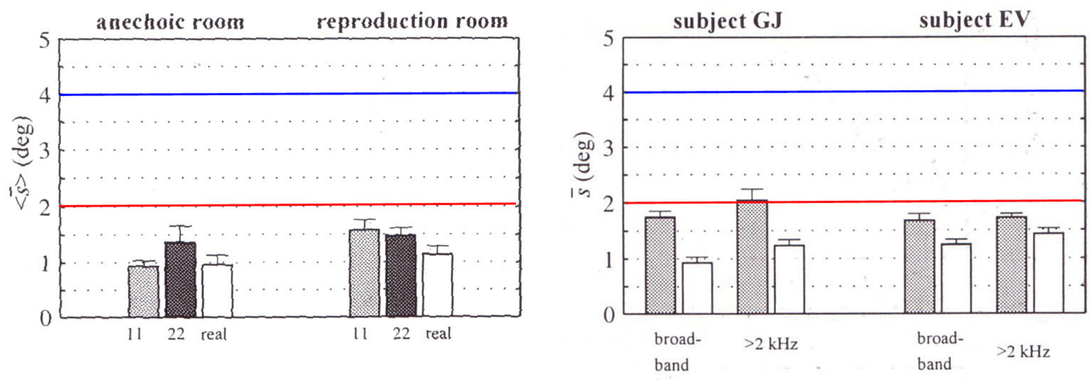
Figure 15: Results of Start's experiments (taken out of Start 1997): Run standard deviation < $\overline{s}$ >

a) anechoic chamber, f_{alias} = 1.4 kHz

b) Auditorium, Delft University of Technology, f_{alias} = 1.2 kHz

c) Concert hall "De Doelen", Rotterdam, f_{alias} = 0.75 kHz

Figures are arranged and customised by this author



a)

b)

Figure 16: Results from Verheijen's experiments:

gray bars: virtual sources Δx=11 cm

black bars: virtual sources Δx=22 cm

white bars: real sources.

a) virtual source behind the array, assessments in two different rooms as indicated (stimulus: white noise bursts)

b) virtual source in front of the array ("focussed"), test signals broadband noise and high-passed noise(>2 kHz), individual results of two subjects (stimulus: noise bursts with energy concentrated in low frequency region)

Figure is arranged and customised by this author

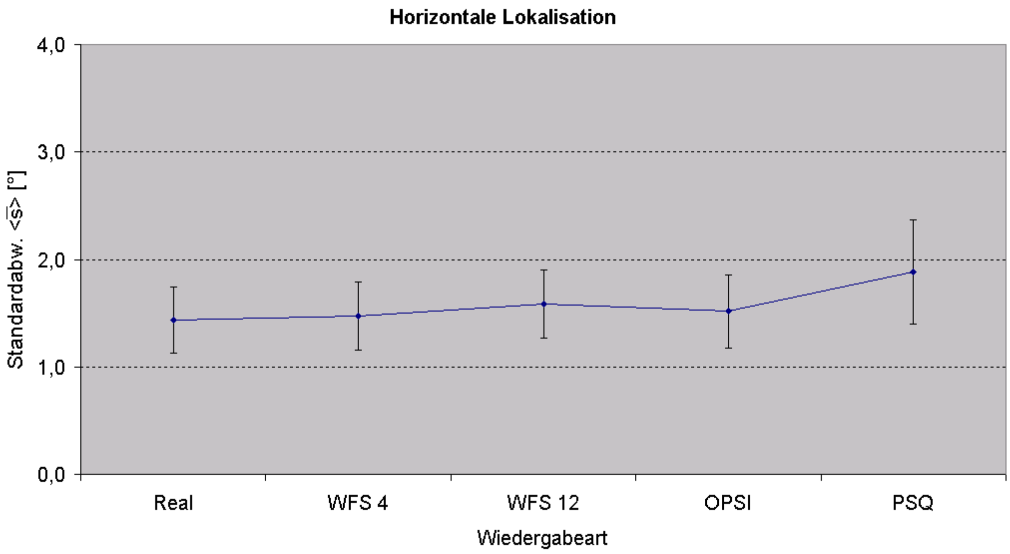


Figure 17: Results of Huber's experiments (taken out of Huber, 2002): Run standard deviation < $\overline{S}$ > for five different reproduction systems:

Real:	**real sources, Δx=11 cm**
WFS 4:	**virtual sources, Δx=4 cm**
WFS 12:	**virtual sources, Δx=12 cm**
OPSI:	**hybrid virtual/phantom sources, Δx=12 cm/70 cm**
PSQ:	**phantom sources, Δx=114 cm**

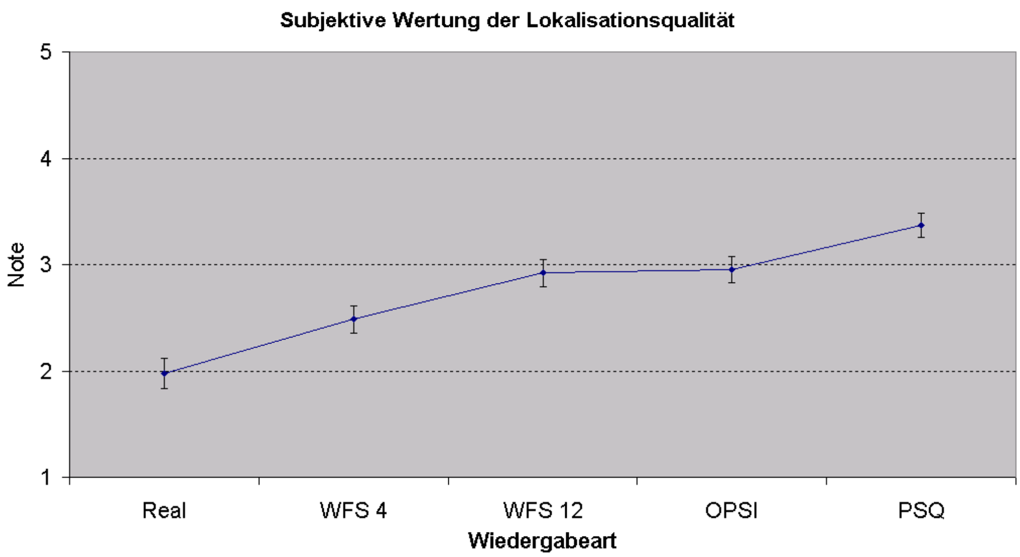


Figure 18: Result of Huber's experiment (taken out of Huber, 2002): Assessment of the "localisation quality" (interpreted and explained as locatedness): Systems as in Figure 17 above.

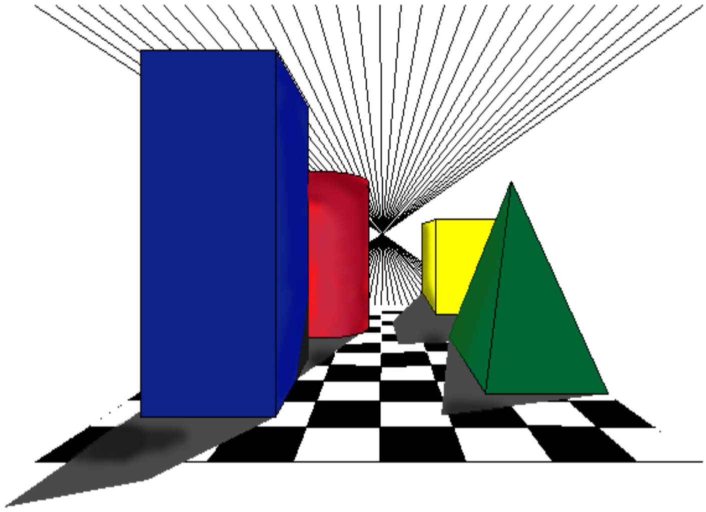


Figure 19: A visual analogy: 2 D representation of a 3D scene,

taken from Ausserhofer et al. (1995)

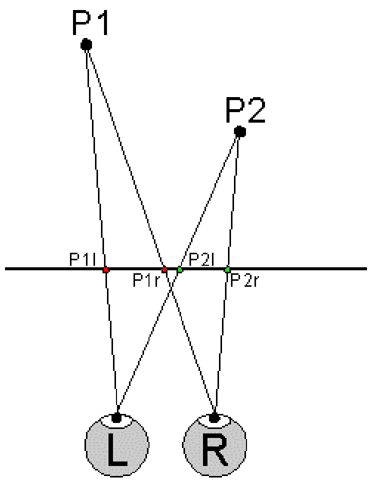


Figure 20: A visual analogy: "Stereoscopic" depth cues, L and R are the two eyes of the viewer, the different signals on the screen are indicated by the red and green dots,

taken from Ausserhofer et al. (1995)