Perceptual differences between wavefield synthesis and stereophony

by

Helmut Wittek

Submitted for the degree of Doctor of Philosophy

Department of Music and Sound Recording
School of Arts, Communication and Humanities
University of Surrey

October 2007

Abstract

This thesis describes investigations into differences between wavefield synthesis (WFS), stereophony and natural sound sources with regard to spatial perception. One aim of the investigations was a comparison and a better understanding of the potential of wavefield synthesis and stereophony to reproduce spatial sound fields. A second aim was to find a rationale for observed perceptual differences by examining the general perception mechanisms. The two sound reproduction techniques were discussed and compared with regard to attributes of localisation, sound colour and distance perception at a fixed listening position. A natural source was considered to be the reference for all investigations.

Both for localisation and sound colour, significant differences between the systems were found. Spatial aliasing had a substantial impact on the perceptual performance of WFS, and its reduction or prevention led to significant improvements. Although stereophonic phantom sources were shown not to be localised as certainly as virtual sources in WFS, the sound colour differences between adjacent phantom sources were found to be smaller. It was shown that the wave front curvature does not provide cues for distance perception in WFS for a static listener.

A combination of WFS and stereophony, which is the OPSI method, was proposed to improve the performance of WFS with regard to sound colour reproduction. The results showed that by this technique, which incorporates stereophony into WFS, the sound colour perception can be optimised while the localisation properties do not change.

The investigation aimed at interpreting the observed perceptual differences through a discussion of the applied perception mechanism. It is hypothesised that different mechanisms apply in the perception of stereophonic and WFS sources. The estimation of the subjective colouration grades by numerical predictors supported this idea, because systems incorporating stereophony were graded better than predicted. A full decolouration, however, as proposed by the association model of Theile could not be proven.

Table of contents

List of figures

List of tables

Acknowledgments

The creation of this thesis was enabled and supported by a number of people. I would like to thank Francis Rumsey for his eminently helpful guidance and reliable supervision as well as Günther Theile for having the idea for this venture, for his confidence, personal support and supervision. Furthermore, I am more than grateful for the great experience of collaborating with the following students on several projects related to my thesis: Tobias Augustin, Tillmann Gronert, Klaus Hanselmann, Thomas Huber, Stefan Kerber and Dominik Wegmann. The friendly contact with Prof. Fastl and Prof. Hergesell was very supportive to my work. I am indebted to the subjects of all listening tests for their support and patience.

The work would not have been possible without the friendly support from my employers, which were the IRT in Munich and Schoeps in Karlsruhe. Furthermore, I much appreciate the uncomplicated and cooperative administration at the University of Surrey.

For proof-reading and smoothing of my ‘Germish’, I would like to cordially thank my friend John Oag. The great and reliable encouragement and help from my parents as well as their example was significant for the conduction of this work. I will not forget that this thesis took a lot of time which I could also have spent with my wife and my daughters. Their love and understanding is crucial for my work.

1. Introduction

1.1 Starting point of the thesis

Wavefield synthesis (WFS) is a reproduction technique capable of creating spatial sound fields in an extended area by means of loudspeaker arrays. The properties of the resulting virtual sound field can be close to those of a real sound field. Since the emergence of WFS in the early 1990s, attempts have been made to utilise this technique for various applications including sound reinforcement, auralisation and spatial sound reproduction. Practical demonstrations have indeed shown that WFS offers an enhanced capability of spatial sound reproduction, but a distinct nomination of its advantages and disadvantages and the consequent classification of these in comparison to other existing sound reproduction techniques have been missed. However, it is of vital importance to explore the advantages and disadvantages of the available spatial reproduction techniques in order to judge their applicability in each individual case. Whenever different techniques of sound reproduction could be applied, a choice based on distinct capabilities is then possible.

The perceptual properties of WFS have still not been investigated in the necessary depth and completeness. Hence, a comparison of techniques based on existing research cannot be undertaken reliably. This thesis is intended to be a further step in this direction and should contribute to enabling this comparison. Naturally, a complete comparison of all attributes of spatial perception cannot be aimed at with sufficient thoroughness. The investigation therefore concentrates on aspects about which the most relevant and possibly also the most controversial questions may exist.

A comparison of perceptual attributes is difficult or impossible to achieve solely by theoretical predictions. Thus the investigation had to incorporate practical experiments. For this thesis, three main experiments in the above described fields were undertaken.

1.2 Aims of the research

One aim of this investigation is to explore the differences in the potential of the techniques WFS and stereo[1] for spatial sound reproduction. There are a number of applications which

[1] *"stereo" is the commonly used abbreviation for „stereophony".*

can be realised by both techniques and thus a comparison of perceptual properties has to provide clues about the specific strength and weakness of each technique. Moreover, it is important to detect common properties of both techniques, because only then can the most adequate and efficient system be chosen. In addition to their experimental comparison, a further knowledge of and an improvement of the perceptual properties of each technique, in particular of WFS, is targeted.

The second aim is to help identify the perception mechanisms which apply for WFS and stereo. For stereophonic perception, certain phenomena cannot be explained consistently by existing theories. The existence of some kind of binaural decolouration or a general difference in the perception of WFS and stereo is hypothesised. The basis is the 'association model' of Theile (1980).

1.3 The sound reproduction principles

Pre-existing principles for spatial sound reproduction include stereophony and binaural audio. These and the principle of sound field reconstruction, as realised by the techniques WFS and ambisonics, make up the three fundamental principles of spatial sound reproduction. Theile et al. (2002, 2003) hypothesise that only these three principles exist and that any other system can be traced back to one of them or to a combination of them. The three principles thus were fundamentally different from each other. In the scope of this thesis, only the techniques WFS and stereo are considered, being loudspeaker-based techniques with potentially overlapping applications. A fundamental difference in the perception mechanisms applying to WFS and stereo would imply important consequences for the prediction of their perceptual quality. The investigation discusses the validity of this paradigm with reference to a series of experiments.

1.4 Incorporation of a new technique: OPSI

On the basis of the results obtained from the investigation on perceptual differences between WFS and stereo and their rationales, general changes in the WFS reproduction were considered. A technique was created which aimed to enhance the perceptual properties of WFS based on the assumption of different perception mechanisms. This reproduction technique, named 'OPSI' (Optimised phantom source imaging in wavefield synthesis), is a combination of WFS and stereophonic reproduction. It is included in the experimental investigations and also acts as a tool to validate the mentioned general assumptions. Its applicability for WFS reproduction is discussed.

1.5 Focus of this thesis

In this thesis, a focus is put on the attributes of localisation, sound colour and distance perception.

More precisely, the directional accuracy and the locatedness of the sources synthesised by the different systems, including the OPSI system, are investigated. Furthermore, the sound colour reproduction is compared and a link between perceptual and physical parameters is attempted. In addition, the advantages of WFS for distance perception are discussed by an investigation into the role of the wave front curvature.

The scope of the investigation is narrowed down to the perceived properties on a fixed listening position.

Apart from the detection and discussion of perceptual differences, the general mechanisms of source localisation are discussed in order to find a rationale for the experimental results.

1.6 Movements of the listener in the sound field, listening area

Possibly the most important perceptual difference between WFS and stereo is the variation of the perceived sound field with movements of the listener. As WFS reproduces virtual sources in similar geometry to that of a real source, the synthesised sound field enables movements and a corresponding change of the ‘view angle’, i.e. another perspective to the sound field. This is not intended and not possible in stereo. Furthermore, WFS offers other advantages for a group of listeners regarding the size of the listening area.

As a consequence of these fundamental differences, WFS and stereo are suited for partly different applications. Hence, the application may rule the choice of the system and this does not necessarily demand a preceding comparison of perceptual attributes. This comparison is more interesting when the same application can be realised by both reproduction techniques because only then do the techniques ‘compete’ with each other. Generally, this is the case for applications in which the listener is located at a fixed listening position. Of course, this precondition clearly constricts the number of applications discussed in this thesis. However, for all other applications, the choice of the system is already obvious without a discussion. Furthermore, one should determine whether the application really demands a sound reproduction system capable of recreating geometrical similarity to a real sound field. Often, the desired high fidelity of a reproduction system is meant much more in terms of perceptual than geometrical properties as described below.

1.7 Physical or perceptual agreement?

An important assumption for the comparison of sound reproduction techniques has to be made about the relevance of the physical similarity of synthesised and original sound field. It is often argued that due to the high similarity between the sound fields produced by WFS and natural sources, a high quality of spatial perception is likely to be achieved. This is partly true for some attributes as the investigations in this thesis have also shown. However, it cannot be argued that in general a higher similarity results in a better spatial quality. The aim should rather be to achieve a high accordance of perceptually relevant attributes between original and virtual sound fields. This can be fundamentally different from a high accordance of physical parameters. There may, for instance, be physical differences between original and virtual sound fields that do not have any perceptual consequences. As an example, the geometrical similarity of original and virtual sound fields is irrelevant for a listener at a fixed listening position. Vice versa, even small deficiencies in certain physical parameters might impair perception significantly. An important example is the spatial aliasing in WFS.

In general, the physical properties of the reproduced space do not necessarily have to be realistic or existent in any real situation. This assumption might change the view of WFS because it was often claimed to be superior due to its close physical accordance with real sound fields.

1.8 Structure of the thesis

After this introduction, the perceptual attributes of interest and techniques of their measurement are introduced (chapter 2).

The two sound reproduction techniques stereo (chapter 3) and WFS (chapter 4) are discussed in respect of their physical and psycho-acoustical properties. The hybrid approach OPSI is introduced and discussed in chapter 5.

Setting the scene for the experiments, the differences between WFS and stereo with regard to this investigation are summarised in chapter 6. This comparison establishes the research questions that will be discussed in the experiments.

The three main experiments described in this thesis cover the properties of the different sound reproduction techniques regarding the perceptual attributes of:

- localisation (chapter 7)
- sound colour perception (chapter 8)

- distance perception (chapter 9)

The investigations are tailored to unveil the main perceptual differences between the techniques.

A summary and conclusions are given in chapter 10.

1.9 Original contributions

This thesis describes differences in the perceptual properties between WFS and stereo. Based on the general discussion of existing knowledge, the remaining open questions are thoroughly discussed and considered by experiments. The thesis thus considers questions on perceptual properties that have not been discussed in a sufficient way so far. Furthermore, the proposal of a new WFS reproduction technique, namely the 'OPSI' method, and its utilisation for an investigation of perception principles, are unique in the current research. Finally, the unconventional way of interpreting stereophonic perception departs from usual paths by taking up Theile's association model from 1980.

In addition, the thesis deals with the following open questions about WFS and stereo perception. Answers are provided in the respective chapters and are only generally introduced here:

- How does the localisation performance of WFS compare to real sources and which aliasing frequency is required for a localisation performance similar to that of real sources?

  The results of experiments on localisation and sound colour suggest that a higher aliasing frequency improves significantly the reproduction concerning both attributes.

- How does the localisation performance of WFS compare to phantom sources?

  The performance of stereo in general seems to be underestimated. Nevertheless, it is shown that WFS virtual sources can potentially be localised much better, depending on the spatial aliasing frequency.

- How does the OPSI system compare to the other systems regarding localisation and sound colour perception and which consequences can be derived from that?

  The new system proposal OPSI is shown to offer localisation properties not worse than those of a comparable WFS system while significantly improving sound colour reproduction.

- Is WFS superior to stereo with regard to sound colour perception?

  Again, stereo seems to be underestimated by current research. Results show that indeed stereo is better than WFS regarding the colouration between adjacent sources.

- Which factors are relevant for the sound colour perception of the different sources?

  The influence of aliasing and spectral alterations can be found in the results of the experiments. However, the stereophonic sources were perceived less coloured than predicted. They are hypothesised to gain from decolouration as long as they are successfully localised.

- Can the hypothesised general difference between WFS and stereo perception be verified?

  From the sound colour performance, consequences on the general perception mechanism were derived. The stereophonic sources were perceived less coloured than predicted. However, the experimental results also show that a full decolouration cannot be achieved and a dependence on the ear signal spectra exists.

- Is WFS superior to stereo regarding distance perception? Does the wave front curvature form a distance perception cue?

  Experiments and theoretical investigations show that indeed no cue can be derived from the wave front curvature for a static listener in WFS.

1.10 The employment of signal processing for this thesis

In this thesis, a series of experiments is described which required the use of signal processing for the reproduction of the different systems. The reproduction was realised by a real-time convolution of the source signal with the relevant transfer functions of the system. These transfer functions can be derived from the relevant theory described in sections 3.5.1 (stereo) and 4.2.1 (WFS). The real-time convolution was realised by an employment of the software convolution engine BruteFIR (Torger, 2007). The transfer functions were calculated and processed using the MATLAB® software package including the MATLAB® Signal Processing Toolbox (The Mathworks, 2007). The measurement of room impulse responses was performed using MATLAB® scripts programmed by this author and by Hulsebos (Hulsebos, 2004). All essential and most of the other work regarding signal processing was performed by

this author. The preparation of the binaural room impulse responses for the virtual acoustic system is described in section 8.2.3.

1.11 Publications during and in the field of the PhD

- Wittek, H. (2004) 'Spatial perception in Wave Field Synthesis rendered sound fields: Distance of real and virtual nearby sources'. *Proceedings CFA/DAGA 2004, Strasbourg, France*, March 2004.
- Kerber, S., Wittek, H., Fastl, H., Theile, G. (2004) 'Experimental investigations into the distance perception of nearby sound sources: Real vs. WFS virtual nearby sources'. *Proceedings CFA/DAGA 2004, Strasbourg, France*, March 2004.
- Wittek, H., Kerber, S., Rumsey, F., Theile, G. (2004) 'Spatial perception in Wave Field Synthesis Rendered Sound Fields: Distance of real and virtual nearby sources', *Proceedings 116th AES Convention, Berlin, Germany*, May 2004, Preprint No.6000.
- Theile, G., Wittek, H. (2004) 'Wave field synthesis: a promising spatial audio rendering concept'. Invited Review, *Acoust. Sci. & Tech. (Journal of the Acoustical Society of Japan)*, Vol.25, No.6, 2004, pp.393-399.
- Wittek, H. (2004) 'Räumliche Wahrnehmung von virtuellen Quellen bei Wellenfeldsynthese'. *Proceedings 23rd Tonmeistertagung 2004*, Leipzig, Germany, November 2004.
- Wittek, H. Augustin, T. (2005) 'Räumliche Wahrnehmung von Wellenfeldsynthese: Der Einfluss von Alias-Effekten auf die Klangfarbe'. *Proceedings DAGA 05, Munich, Germany*, March 2005.
- Kerber, S., Wittek, H., Fastl, H. (2005) 'Ein Anzeigeverfahren für psychoakustische Experimente zur Distanzwahrnehmung'. *Proceedings DAGA 05, Munich, Germany*, March 2005.
- Wegmann, D., Theile, G., Wittek, H. (2006) 'Zu Unterschieden in der Hörereigniswahrnehmung bei Wellenfeldsynthese und Stereofonie im Vergleich zum natürlichen Hören'. *Proceedings DAGA 06, Braunschweig, Germany*, March 2006.
- Wittek, H., Rumsey, F., Theile, G. (2007) 'Perceptual enhancement of wavefield synthesis by stereophonic means'. *Journal of the Audio Engineering Society*, Vol.55, No.9, September 2007, pp.723-751.
- Wittek, H. (2007) 'Sound colour properties of WFS and stereo'. Invited Paper, *Workshop Wave Field Synthesis – 1st DEGA Symposium 2007, Ilmenau, Germany*, September 2007.

2. Introduction to the perceptual attributes examined in this thesis

2.1 Introduction

This chapter defines and describes the perceptual attributes on which this thesis will be focussed. It thus prepares for the discussion in the following chapters. The selection of the chosen attributes is justified in section 2.2, before each attribute is introduced in a dedicated section. The discussion of the localisation attributes (section 2.3) is followed by an introduction to the attributes sound colour (section 2.4) and source distance (section 2.5). The chapter is summarised in section 2.6.

2.2 Classification and selection of attributes

This investigation puts an emphasis on the perceptual properties of WFS and in particular on the difference between the perceptual properties of WFS and stereo. Of course, the investigation cannot aim at a complete coverage of all perceptual properties, but rather attempts to detect and consider the few most apparent differences. Moreover, the selection of the specific attributes describing these differences is based on the second aim of this investigation. This aim is to explore the properties of WFS and stereo against the background of their basic perception mechanisms. Therefore, attributes that can give crucial hints on the basic processing are selected for thorough investigation. As will be described in chapter 3.6.2, Theile's association model (1980) is one main basis for a discussion of source perception in this thesis. Hence, an experimental investigation tailored to discuss the hypotheses of this model makes sense. Accordingly, the main focus is put on attributes related to the perception of source location and 'gestalt'[2] (see discussion below). The resulting two attribute groups correspond to the separation in two basic processing stages as hypothesised by Theile.

---

[2] *'Gestalt' is a term used in psychology. A gestalt quality is achieved when "... many groups of stimuli acquire a pattern quality which is over and above the sum of their parts; for example, a square is more than a simple assembly of lines – it has 'squareness'." (Ehrenfeld, 1890 cited in Gross, 1992) The gestalt psychology was mainly introduced by the group around Wertheimer (1880-1943). Possible translations are: 'organised wholes', 'configurations' or 'patterns' (Gross, 1992).*

The attributes of source localisation are those related to the location and the geometry of the source, together with those characterising the quality in which a source is localised. This group could perhaps also be called ‘spatial attributes’. However, this might lead to disagreements regarding the inclusion of attributes related to sound colour. Furthermore, the term ‘spatial attributes’ does not represent the hypothesised separation in the perception process, and therefore lacks an apparent definition. The selected localisation attributes are introduced in section 2.3.

The ‘gestalt’ of the source comprises the acoustical content of the perceived source. After Theile (1980; see chapter 3.6.2), the gestalt perception is independent of the perception of the source location. This means the auditory system is able to analyse the source content independently of the signal characteristics related to the location perception. In other words, signal characteristics could be removed for perception that are not part of the acoustical content of the source, but added by the room and the head and body of the listener. This would be called an ‘inverse filtering’ process. Theile denominates the remaining acoustical content of the source the ‘gestalt’ of the source. By this denomination, the character of the signal is described, being a construct of different sub-characteristics. These add up to a pattern that is detected by the auditory system as a whole, by comparison with known patterns. The term ‘gestalt’ implies that the content of the source is perceived by a logical assembly of the available cues (with the assistance of other cues such as visual cues, the source location, the contextual relationship of the source in the scene etc) and that the created percept can be considered more than just the sum of its parts. In the context of acoustical source perception, an example of a certain source gestalt is the voice of the mother, comprising sub-attributes such as female voice, formants, tune, voice melody, language/dialect etc, together with the visual sensation of the mother. These sub-attributes are meaningful only in their common existence.

The relevant meaning of the perception of the source gestalt is based on the functional principle of Theile’s association model. Through the hypothesised ‘inverse filtering’ of the transfer function between source and inner ears, the gestalt would be completely reconstructed in the case of successful spatial decoding. In other words, a non-successful spatial decoding would create an impaired gestalt of the source, as the inverse filtering process cannot operate. After Theile, the result would be a colouration of the source. This colouration can be measured and thus the success of the gestalt perception can be implicitly estimated. Hence, a measurement of the colouration enables conclusions to be drawn on the localisation process. The attributes sound colour and colouration are introduced in section 2.4.

In addition to the two mentioned groups of attributes related to the perception mechanism, a further attribute was considered within this investigation, due to its high relevance for a com-

parison of perceptual properties. The perceived distance of the source depends on a number of different physical parameters of the sound field. The existence of these parameters in the reproduced sound field and their evaluation decides whether the distance perception is successful. As will be described, during the comparison of wavefield synthesis and stereo, the parameter wave front curvature was discovered as a main cue that could make a significant difference regarding the attribute source distance. This attribute will be introduced in section 2.5.

2.3 Localisation

2.3.1 Collection of attributes found in literature

A sound reproduction system has to be capable of reproducing sources in certain directions at a sufficient quality (as the term '(localisation) quality' is vague and comprises a number of attributes, alternatives will be introduced below). Both the direction and localisation 'quality' of the perceived source can be measured by experimental means. Hence, an investigation is possible which - potentially rather precisely - detects differences between different sound reproduction systems. The detected differences may also give rise to conclusions about basic differences in the reproduction of the systems or the perception of the source.

This section lists attributes of localisation that are used in the context of the evaluation of sound reproduction techniques. It also lists their definition and determines their use for this investigation. In the past, attributes of localisation have often been defined individually for certain investigations, and sometimes they remain unclear in their meaning. The consequence is a lack of consistency between the different definitions (or applied meanings) and a significant difficulty in comparing different results. Some terms, as they are used in the literature, can have different meanings. These ambiguities also stem from the different purposes of investigations, many of them not concentrating on the evaluation of sound reproduction systems as in this case.

The relevant terms found in the literature are listed in Table 2-1 below, together with definitions by this author unless otherwise noted. Attributes written in *italics* are not used further on in this thesis.

Localisation	General mapping law between the location of an auditory event and a certain attribute of the sound source. (Definition according to Blauert, 1997) Mechanism/Process that maps the location of an externalised auditory event to certain characteristics of one or more sound events. (Definition according to Theile, 1980)
Direction	The direction in which the source is perceived
Distance	Perceived range between listener and reproduced source (Definition according to Rumsey's (2002) 'individual source distance')
Depth	Sense of perspective in the reproduced scene as a whole (Definition according to Rumsey's (2002) 'environment depth')
Stability	The degree to which the perceived location of a source changes with time.
Robustness	The degree to which the perceived location of a source changes with movement of the listener.
Accuracy	The degree to which the intended 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 only for the 'directional accuracy', which means the agreement concerning the source direction. The relevant measure for this attribute is the 'directional error' of a source/system.
Resolution	*The achievable precision of the synthesised sound field in terms of direction and/or distance.*
Individual source width ISW, Apparent source width ASW	*Perceived width of the source (Definition according to Rumsey, 2002).*
(Image) focus	The degree to which the energy of the perceived source is focussed in one point.
Definition of the image	*Similar to image focus*

Diffuseness	*Inverse of image focus*
Blur	*Inverse of image focus*
Locatedness	Spatial distinction of a source. (Definition according to Blauert, 1997) The degree to which an auditory event can be said to be clearly perceived in a particular location.
Certainty of source localisation	*Similar to 'locatedness', used by Lund (2000)*
Localisation quality, Localisation performance	These terms describe a mix of attributes. They describe the overall performance of localisation. They should be defined individually, because they can have ambiguous meanings ('quality' of the directional accuracy, sound colour, focus, locatedness or an 'average' quality?).
Externalisation	*The degree to which the auditory event is outside the head*
Spaciousness	*Often used in the same meaning as 'apparent source width' ASW, but also used to describe the perceived size of the environment.*
Presence	Sense of being inside an (enclosed) space or scene. (Definition according to Rumsey, 2002). Often also used as an attribute of sound colour.

Table 2-1: Summary of potential localisation attributes found in the literature

2.3.2 Measurement methods for directional accuracy, image focus and locatedness

Measurement methods for certain attributes of localisation are reviewed and discussed in this section. Based on and derived from these, suitable approaches for the tasks of this investigation can then be chosen. Often, the meaning of a localisation attribute becomes apparent when taking a closer look at the method with which it was measured. The discussion of measurement methods for the attributes directional accuracy, image focus and locatedness will show which approach fits the aims of this investigation best.

Directional accuracy

Experiments of Vogel (1993), Start (1997) and Verheijen (1998) explored the localisation properties of WFS virtual sources. The mean run standard deviation[3] $< \bar{s} >$ of the perceived auditory event directions serves as a measure for the 'overall localisation quality' of the systems. This procedure may be regarded as valid if it is undertaken with respect to a reference, such as a single small loudspeaker, having small source width, sharp focus and good localedness 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 (e.g. Corey and Woszczyk, 2002; Rumsey, 2002) that there can in fact be a difference between the perception of source width, focus and locatedness. This applies particularly for WFS, as believed by this author.
2. It may be possible to judge a change in the overall 'localisation quality' from a change in the standard deviation, but the reverse is not proven: a change of one of the attributes comprised in the term 'localisation quality' does not necessarily lead to a change of the measured standard deviation. This can be observed in the experiment described in chapter 7. Vogel, Start and Verheijen, however, deduced the localisation quality from the measured standard deviations alone, and therefore arrived at different results to the ones described in this thesis. The same problem may also occur for measurements of the minimal audible angle (MAA) which do not necessarily reveal differences in the localisation quality. Start (1997) measured the MAA of virtual sources in WFS.

By mathematical analysis, two other figures can be extracted (see Hartmann, 1983) from the data obtained 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 of the deviation of the perceived directions 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* is

---

[3] *The standard deviation s here is defined as the RMS error 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 $< \bar{s} >$ (see Hartmann, 1983)*

related to a predefined direction x_0 in contrast to the standard deviation *s*, which is related to the mean value x_{mean} of all perceived directions. The relation between the two is $D^2 = s^2 + (x_{mean} - x_0)^2$. Hartmann takes the mean run RMS error $< \overline{D} >$ as the most suitable parameter to describe the 'localisation performance'. Start adopts this definition in his analyses. The mentioned problem of the standard deviation *s* applies to the RMS error *D* even more: 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, shape and quality of the (virtual) source. A system with a directional bias (be it in the system or the measurement procedure) cannot be assessed by this measure. A system which synthesises the desired directions with a directional error of $|x_{mean} - x_0| = +5°$, but is capable of presenting the sources with an optimal focus, will result in a small standard deviation *s*. The RMS error *D* of this measurement, however, would not be less than 5° in spite of that.

The second available measure is the signed error $E = x_{mean} - x_0$, which is a measure for the average deviation of the perceived direction x_{mean} from the predefined direction x_0, in which the sign of the deviation is taken into account. This is a measure for the 'directional accuracy' of a system.

Image focus and locatedness

Apart from the above-mentioned implicit method of measuring the standard deviation and concluding on the attributes image width, focus and locatedness, there are other methods of directly measuring these attributes. Various authors describe measurements of the image focus in the context of the evaluation of sound reproduction systems. Martin et al. (1999b) presented pairs of stimuli, requiring the test subjects to indicate the more focussed of the two stimuli. In his definition, the focus of a (phantom) source is dependent 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 emphasises 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 clear distinctions between the five different systems under investigation in terms of the assessed focus of the sources. He also performed measurements of the IACC (Interaural Cross Correlation) coefficient of the same stimuli using a dummy head, which did not reveal these distinctions.

In (Wittek et al., 2001b) stimuli (phantom sources) were presented in comparison to a reference, this being a single loudspeaker. The subjects were asked to assess the difference in the image focus using a five-grade scale. The results showed clear differences in the perceived

focus which could not have been concluded from the deviations of the perceived directions. The focus data showed a clear trend whereas the measured standard deviations of the perceived directions showed no significant differences. The definition of focus used in this study was similar to Martin's and in both studies a human voice was used as the stimulus.

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 grade on this consistency scale would be given to a source that is localised with no doubt, is very robust and whose image is not diffuse. Corey et al. (2002) made use of Lund's scale and measured the 'certainty of the source location' on a five-grade scale. They 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 source location. However, regarding the spread and the bias of the directional data that Corey et al. call 'accuracy', they state: *"In comparing the localization 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 localization ability"*.

Chapter 7 describes an experiment on the localisation properties focussing on the attributes directional accuracy, image focus and locatedness.

2.4 Sound colour and colouration

The sound colour is one of the important attributes describing a sound or, as in our case, a sound reproduction technique. In this investigation, the term 'timbre' is supposed to have the same meaning as the term 'sound colour'.

The capability of reproducing the correct (or at least a plausible, see chapter 1.7) sound colour of a virtual source is a vital property of a sound reproduction technique. Listeners would rather accept a compromise in the spatial fidelity of the reproduced sound field than a degraded sound colour (Rumsey et al., 2005). It is a challenge for systems providing a spatially enhanced performance, like WFS, not to achieve this enhancement at the cost of a reduced sound colour reproduction quality.

The definition of the attribute sound colour or timbre is difficult. The well-known definition from the American Standards Association (ASA, 1960), is:

”Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.”

However, this definition takes into account neither temporal nor spatial differences, neither of which can be understood as attributes of timbre. Newer descriptions consider the timbre as a multidimensional construct of a number of different sub-attributes which describe the specific sound colour properties of the source (Bloothooft and Plomp, 1988; Zwicker and Fastl, 1990; Bregman, 1994).

Subjective measurements of the sound colour have been somewhat elaborate. An overview of the literature can be found in (Brüggen, 2001a) and (Rubak and Johansen, 2003). Reasons for the difficulty in the measurement of the sound colour include a lack of meaningful objective measures that describe the subjective perception. Furthermore, the sound colour always refers to a reference that is used for comparison, be it in the memory of the listener, or a direct comparison (Brüggen, 2001b). The latter possibility unveils very fine differences: the auditory system is particularly sensitive to changes in the sound colour between two sounds (Bloothooft and Plomp, 1988; Bücklein, 1981). In contrast, the auditory system is able to adapt to a static frequency response which is then regarded as not coloured (Zwicker and Fastl, 1990). When the difference is more interesting than the absolute sound colour, the attribute ‘sound colour difference’ can be used. It is easier to measure the sound colour difference than an absolute measure related with sound colour. Examples for the latter could, for instance, be the naturalness or the degree of distortion of a signal.

The sound colour difference is also called ‘colouration’ (or coloration, AE). Salomons (1995) proposed the following definition for this term:

”The coloration of a signal is the audible distortion which alters the (natural) color of the sound”.

Chapter 8 presents a method of measuring the colouration using a multiple stimulus graphical user interface employing a five-grade colouration scale.

2.5 Distance

This section introduces the general auditory cues for the perception of the source distance. It discusses the specific cues available in the direct sound of a source signal, thus preparing for the discussion in the following chapters. Furthermore, a differentiation between the attributes distance and depth in the context of sound reproduction is performed.

The discussion of the sound reproduction techniques in chapters 3 and 4 and their comparison in chapter 6 as well as the experiment described in chapter 9 consider the specific properties of WFS and stereo regarding distance perception.

2.5.1 Distance cues

The literature (e.g. Nielsen, 1991; Zahorik, 2002; Shinn-Cunningham, 2000; Blauert, 1997) describes various crucial parameters for auditory distance perception. These include:

a) Level (sound pressure)
b) Direct-to-reverberant energy ratio
c) Reflection pattern (timing, level and directions of early reflections)
d) Frequency spectrum (for very near and for far sources)
e) Binaural differences: acoustic parallax as well as intensity/phase differences
f) Motion parallax (changes of perspective with source/listener movements)
g) Interaction with source familiarity and non-acoustical cues such as visual cues

Cues a, b, d, e, g can also be found in above-mentioned literature. Cue c is mentioned by Pellegrini (2001). Cue f is added by this author.

In this investigation, most of these cues are not discussed in detail. A general discussion of distance cues is not intended, but rather a comparison of potentially available cues in the two sound reproduction techniques WFS and stereo. Aiming at this comparison, cues will be identified that only one of these techniques can reproduce, or at least can reproduce much better than the other. The differences between the two techniques require a focus on specific aspects of distance perception cues such as the role of listener movements, the role of an accurate reproduction of the reflection pattern and the role of the wave front curvature. This differentiation is introduced here whereas the comparison is performed in chapter 6.

Cues available for a moving source or listener

The cues mentioned above differ with respect to their relevance for certain listening conditions. There are cues that are available on a fixed listening position and others that are available only with listener or source motions. The latter cues are also called ‘idiothetic’ cues, i.e. cues which are created by the self-motion of the listener and an analysis of resulting changes in the perceived sound field. These cues enable an implicit analysis of the scene geometry (and consequently distance) through moving within the listening area.

The group of cues for static listeners include cues a, b, c, d, e, g. The group of cues for moving listeners in principle only includes cue f. However for cues a, c, e, g listener or source motions are hypothesised to support their perception, as it is known for perception in general that a difference in cues corresponding to known listener or source changes is easier to detect than static cues. Furthermore, certain cues are available only in a relative comparison, be it with other sources or source conditions, or with a stored pattern in the memory of the listener. We can assume that for nearby sources at least, a listener or source movement and the corresponding change of the cues level, direct-to-reverberant energy ratio, reflection pattern and binaural differences provides additional information about the source distance. In particular, the natural change of the reflection pattern is important information. The cue ‘reflection pattern’ is added to the usual list of distance perception cues because it is decisive for the perception of source distance (Pellegrini, 2001) and, moreover, it can be reproduced better in WFS than in stereo. Changes in the reflection pattern that correspond to changes in the source or listener movements theoretically contain unambiguous information about the source location and thus potentially give rise to an evaluation by the auditory system.

Cues available for nearby sources

An indirect distance perception due to listener or source movement can only occur for relatively close sources, as only for these sources are the differences in the cues significant. However, for static listeners or sources, there are also cues that exist only for such close sources. These are the cues based on binaural differences, i.e. cue e of the above list. In the near-field, the so-called ‘acoustic parallax’ serves as an additional auditory cue. The acoustic parallax describes the phenomenon that for nearby sources, the incidence directions are different at the ears. The diffraction at head and ears of the listener also differs due to the curved wave front. Moreover, the binaural differences also include the distortion of binaural level and phase differences for nearby sources.

An estimation of the magnitude of the near-field dependence of the HRTF (head related transfer function) can be made using Figure 2-1. Further measurements are presented in chapter 9.

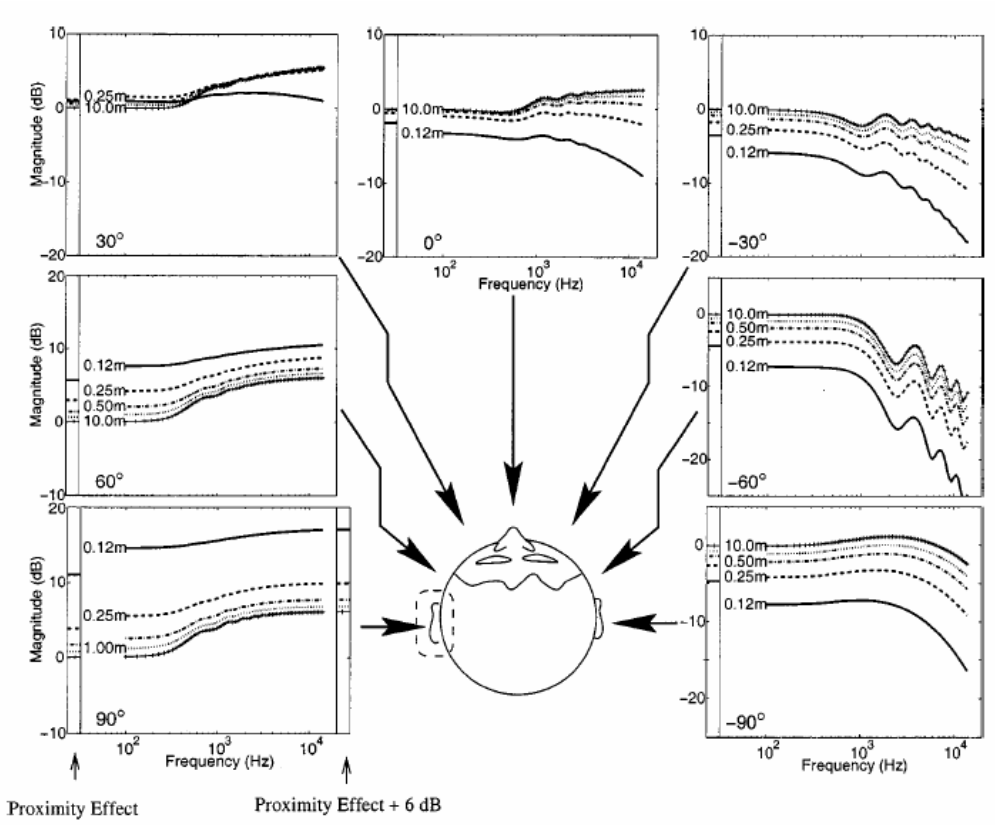


Figure 2-1: from Brungart and Rabinowitz (1999a): The HRTFs for sources in the horizontal plane from 0.125 m to 10 m distance. The head is modelled as a rigid sphere 18 cm in diameter. The HRTFs are calculated by dividing the pressure at the left ear by the free-field pressure at the centre of the head. Results are shown for source locations at 30 degree intervals in azimuth in the front hemisphere.

For Blauert (1997), the distance of sources closer than 3 m can be perceived due to these cues. The results of Nielsen suggest an upper limit of 1 m. Under anechoic conditions, only the cues level and binaural differences remain for fixed listening positions. Brungart and Rabinowitz (1999c), who studied distance perception for sources closer than 1 m, identified the interaural level differences (ILD) at low frequencies (< 3500 Hz) as crucial for distance judgments of these sources in anechoic environments. Although Shinn-Cunningham (2000) notes that by adding reflections, the distance perception of nearby sources improves significantly, the binaural differences are apparently strong enough to override the level cue. Brungart et al. (1999b) showed that proximal-region distance perception with a broadband, random-amplitude source is significantly more accurate for lateral sources than those near the median plane. Their results from distance perception experiments (Brungart and Rabinowitz, 1999c) with anechoic, nearby sources positioned along the interaural axis are shown in Figure 2-2. It can be seen that the low pass and the broadband conditions performed better than the monaural or high pass conditions. They showed that in the near field a strong correlation (as

high as 0.85) can be found between the logarithmically arranged actual and perceived distances.

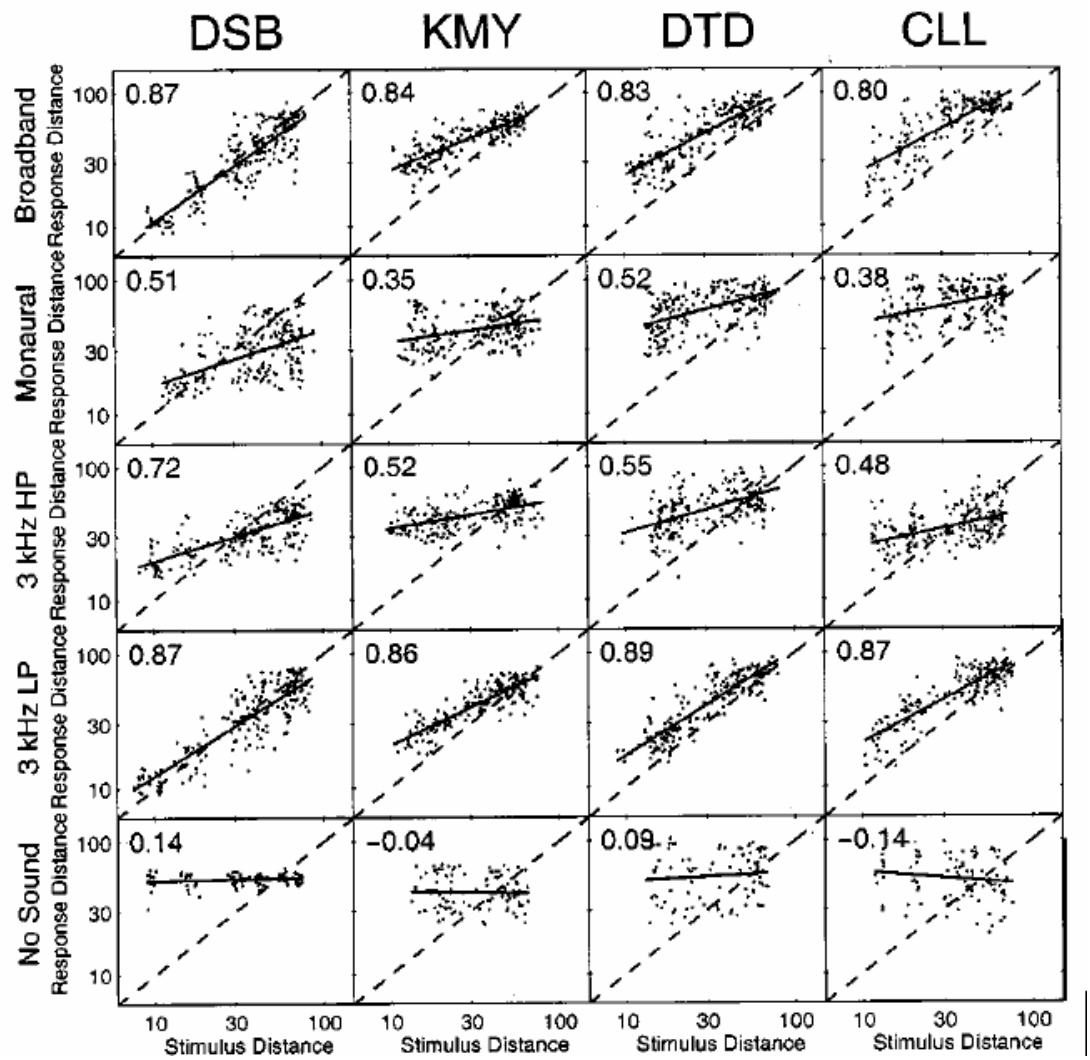


Figure 2-2: from Brungart and Rabinowitz (1999c): Results from the distance perception experiment of nearby dry sources. The sources are positioned along the interaural axis. The dashed lines represent 'correct' responses, while the solid line is the best linear fit of the stimulus data to the response data. The number at the top left of each panel is the linear correlation coefficient. Five different conditions are shown in the rows whereas the columns correspond to single subjects.

In a study by Martens (2003) regarding near-field distance perception, the HRTFs were manipulated so that the gain (in the entire frequency band) at the ipsilateral side was increased compared to the reference HRTF, whilst the gain at the contralateral side was decreased. By this procedure, both the ILD was increased in different steps, and the gain relative to the reference source was varied. The subjects were presented with headphone reproduction of whispered syllables, which were recorded dry and convolved with the manipulated HRTF. The subjects were asked to assess the relative range (distance) of the test items in comparison to a

reference (0 dB increase/decrease) on a 10 point-scale. On this scale, the rating 10 corresponded to no noticeable difference in the range between the test item and the reference, 9 was to be given when the test item was just noticeably closer, the rating 1 corresponded to a very closely perceived auditory event and 0 to an auditory event inside the head. In other words, the subjects were asked to compare the distance of the reference and the test item, the ILD of which was artificially increased.

The results are shown in Figure 2-3. They show the expected perception of closer ranges with increasing ILD. Furthermore, it was discovered that an increase of the ipsilateral gain rather than a decrease of the contralateral gain leads to the perception of a closer range. This investigation once more proved the ability of high ILD to evoke the percept of a close source distance.

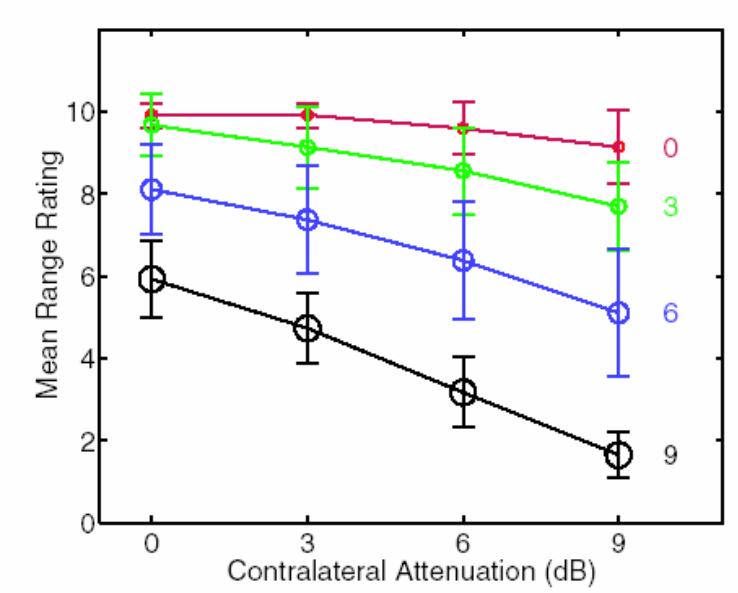


Figure 2-3: Experimental results from Martens (2003): The mean range ratings (y-axis) show whether the auditory event was perceived in the same range (10) as the reference source, closer (<10), extremely close to the listener's ear (1) or inside the head (0). The x-axis shows the contralateral attenuation and the labelled parameters show the ipsilateral gain from 0 (red) to 9 (black). The resulting ILD is the sum of contralateral attenuation and ipsilateral gain.

Distance perception under anechoic conditions

As shown in the literature, the perception of the distance of non-nearby sources is nearly impossible without the presence of room reflections. Investigations by Nielsen (1991) show that in an anechoic chamber, the actual source distance has no influence on the perceived distance, as long as the level at the listening position (receiver level) is kept constant (see Figure 2-4). Level is the most important cue when room acoustics are absent, as shown by Gardner (1969, cited in Blauert, 1997).

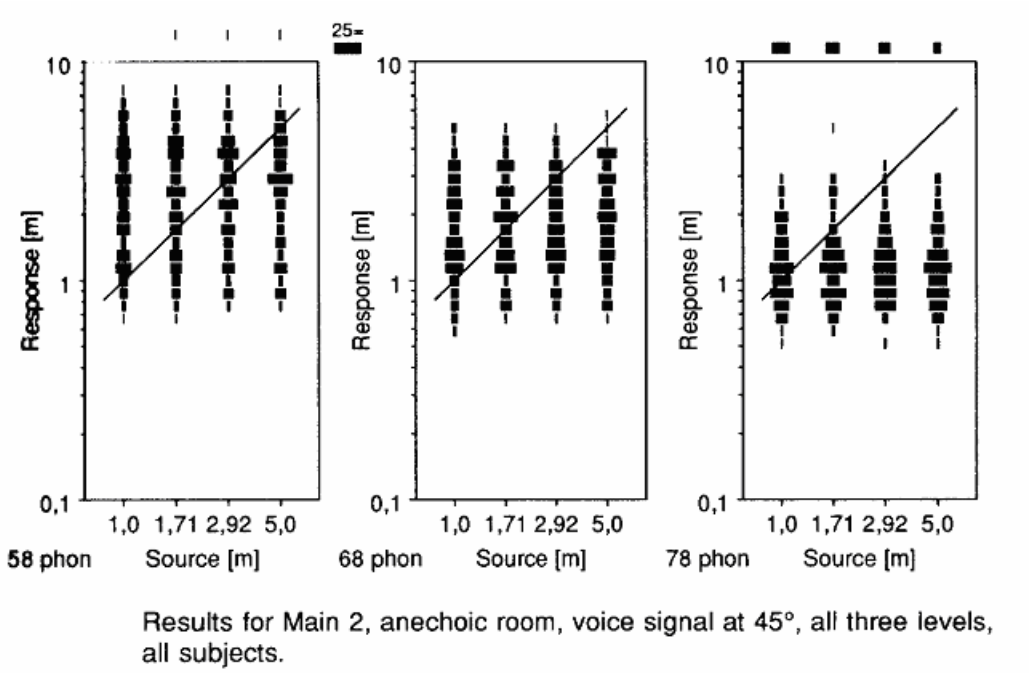


Figure 2-4: Experimental results from Nielsen (1991): There is no correlation between the actual source distance (x-axis) in the anechoic chamber and the perceived distance (y-axis). But: the louder the stimulus the closer it is perceived (the three figures correspond to a different receiver loudness, which is 58, 68 and 78 phon). The solid lines in the diagrams indicate the relation y=x.

2.5.2 Distance and depth

A sound image without depth is unnatural. In any natural sound field a sense of depth is perceivable, being *"the sense of perspective in the reproduced acoustic scene"*, as defined by Rumsey[4] (2002). A sense of depth in a natural environment is given through the perception of sources at different distances together with an analysis of the room reflections, which contain an unambiguous description of the room dimensions. Depth is a scene-related percept that takes into account relationships between multiple auditory events. The listener's successful perception of depth is the benchmark of a spatial audio reproduction system.

Contrary to depth, distance is an attribute of a single source. Although depth supports the perception of the source distance, the latter can also be judged in acoustic scenes without

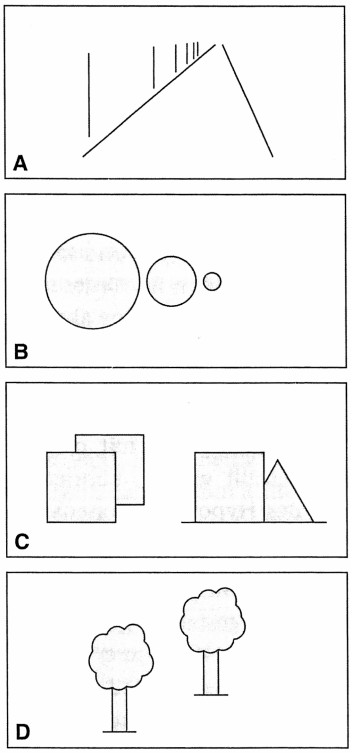
[4] *Rumsey (2002) defines the attribute depth at different levels. He introduces the individual source depth, the ensemble depth, the environment depth and the scene depth, each being the depth defined for the particular level. This investigations will use the attribute depth similar to Rumsey's scene depth.*

depth (e.g. a mono loudspeaker). In this case, the perceived distance may be called ‘pseudo’ distance. It is assumed to be perceived consciously rather than intuitively, as a result of a rational analysis of certain perceived distance cues such as level, direct-to-reverberant energy ratio or frequency spectrum. In those scenes, a ‘pseudo’ depth can also exist, this being the relative perspective of the sources perceived according to a ‘pseudo’ distance.

Examples of ‘pseudo’ distance or depth also exist in visual perception. An interesting parallel can be drawn between the acoustic and the visual perception of depth and distance. By looking at the more apparent visual cues, the acoustic cues can also be illustrated. The visual cues are depicted in (Smith et al., 2003) or (Becker-Carus, 2004). The analogy is hypothesised by this author. Figure 2-5 shows several monocular visual cues to analyse spatial depth. These include the linear perspective, overlay (or interposition, one object is covered by the other), relative size, height-in-field, shadows, amongst others. In spite of these cues, the image does not contain true depth, it is a 2D representation of a 3D visual scene.

When another important cue, the so-called ‘motion parallax’ is added (a corresponding change of the perspective with movements, it is also called an ‘idiothetic’ cue), it is termed a 2½D representation, which enables the perception of perspective due to movements of the viewer. A true 3D representation – and thus the perception of real visual depth - is enabled only through the existence of binocular (or ‘stereoscopic’) cues such as disparity (different signals for the two eyes, see Figure 2-6) or convergence (different axis angles of the two eyes).

The presence (definition by Rumsey, 2002: *“Sense of being inside an (enclosed) space or scene”*) of the listener/viewer can be only achieved when the reproduced scene contains true depth. In contrast, there can be some kind of presence or ‘ensemble envelopment’ (definition by Rumsey, 2002: *“Sense of being enveloped by a group of sound sources”*) which can be created by idiothetic cues. These idiothetic cues create an intra-active scene, i.e. a scene in which the listener/viewer can move. Furthermore, an inter-active scene is capable of creating a strong link between the reproduced scene and the listener/viewer. This kind of presence is sometimes referred to as ‘immersion’, being an expression linked with some form of intra- or inter-activity.



Linear perspective

Relative size

Overlay, interposition

Height-in-field

Figure 2-5: from Becker-Carus (2004): An analogy in visual perception. 'Monocular' cues for depth perception are linear perspective (A), relative size (B), overlay or interposition (C), height-in-field (D), etc.

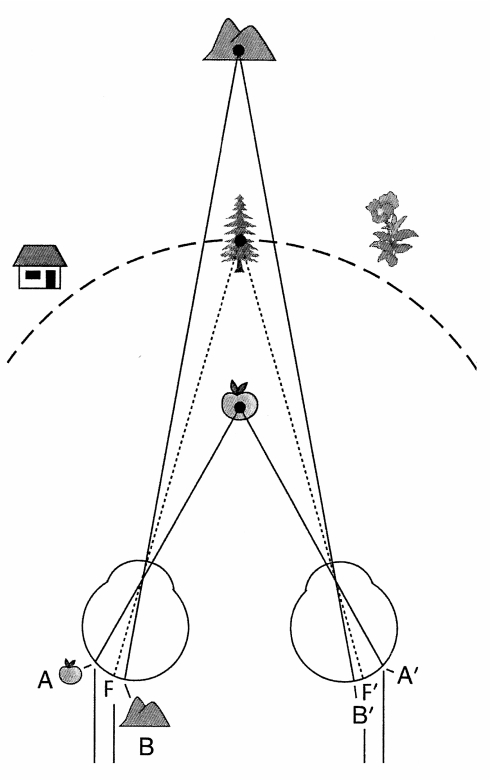


Figure 2-6: from Becker-Carus (2004): An analogy in visual perception. Example for 'binocular' cues for depth perception. Due to the 'binaural disparity', the objects at different distances are imaged at different locations in the eye.

Based on the above discussion, the cues for auditory distance perception listed in section 2.5.1 can be grouped according to their property to enable a representation of 'pseudo' or true depth, and their availability with/without movements of the listener. The terminology of these groups is based on the terminology used to describe the visual cues mentioned above. Hence, an acoustical 2D representation would be one that lacks true depth and does not provide cues for listener movements. An acoustical 2½D representation would be a representation that still lacks true depth, but does provide cues for listener movements, i.e. can be called 'intra-active'. An acoustical 3D representation is a representation that offers cues to perceive true depth, but does not necessarily provide cues for a listener movement. A representation enabling both true depth and listener movements is not defined at this time. Note that in the acoustical context, the terminology of 2D, 2½D and 3D representation does not imply the property to reproduce a certain number of geometrical dimensions. For instance, a 3D representation does not provide the reproduction in three dimensions, which would mean an addi-

tional reproduction of the ‘height’ dimension. Rather, it denotes the quality of the depth reproduction.

These groups of cues for acoustical distance perception can be defined according to this paradigm:

‘2D’ distance cues: Monaural distance cues that enable a ‘pseudo’ distance perception

- Level
- Direct-to-reverberant energy ratio
- Frequency spectrum
- Interaction with other non-acoustical cues

‘2½D’ distance cues: Cues that are available with movements of the listener

- Motion parallax
- Improvement of several 2D cues

‘3D’ distance cues: Binaural distance cues that enable a ‘true’ distance perception

- Reflection pattern
- Binaural differences

The analogy of acoustical and visual cues for distance perception is continued in chapter 6, where a comparison of WFS and stereo regarding the availability of these cues is performed. However, Table 6-1 may be yet observed at this time to find a summary of the above-mentioned analogies regarding the different representations.

2.6 Summary of chapter 2

This thesis approaches the comparison of the perceptual performance of WFS and stereo by an investigation of distinct attributes. The investigation includes the theoretical and experimental comparison of the attributes of localisation, sound colour and distance perception. The basis for a comparison of these attributes was created in this chapter by an introduction of their definition and meaning for this investigation. Furthermore, the rationales of the selection of these particular attributes were depicted.

3. Stereophony and its properties

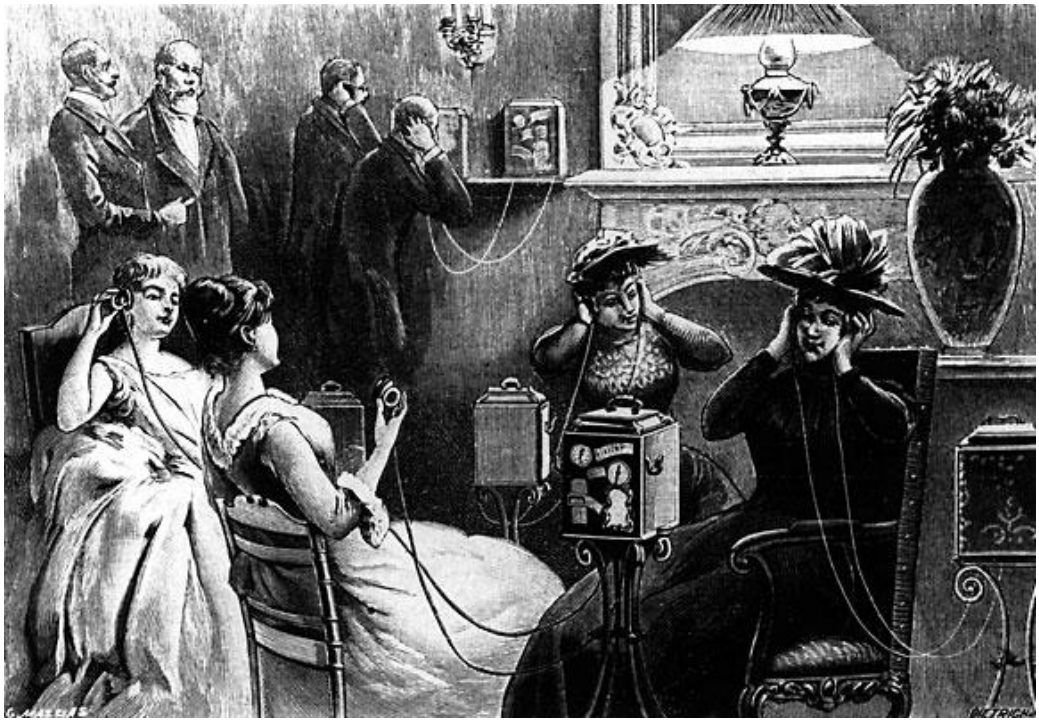


Figure 3-1: The first stereophonic transmission by Clement Ader in the year 1881 (from Daniels, 2002): Listeners enjoy a performance of the Paris opera house transmitted by two telephone lines. Ader patented this stereo telephone.

3.1 Introduction

This chapter introduces stereophonic reproduction. It discusses the properties of stereo which are relevant to the attributes described in this thesis. Furthermore, it presents various approaches that aim to explain the perception of stereophonic sources (which will from now be referred to as 'stereophonic perception'). These approaches differ fundamentally regarding their consequences on the perceptual properties of stereo.

After this introduction, a definition of stereophonic reproduction in section 3.2 will prepare for further discussion. Section 3.3 describes the origin of stereo, anticipating various means of interpreting stereophonic perception, which will be introduced in section 3.4. The properties of stereophonic reproduction, such as its capabilities for directional imaging, sound colour and distance reproduction, are introduced in section 3.5. A discussion of the different possible perception mechanisms is given in section 3.6, before section 3.7 summarises the chapter.

3.2 Definition of stereophony for this investigation

A non-existing, 'virtual' source is localised when two loudspeakers reproduce a coherent signal. The two sound events result in one single auditory event. Level and time differences between the loudspeaker signals determine the perceived direction of this auditory event, which is commonly known as a 'phantom source'. (This denomination already implies a certain perception mechanism, see section 3.4.)

The sound field created by the two loudspeakers is different from that which would be created by a corresponding real source at the location of the phantom source.

This reproduction technique is called stereophony or 'stereo'. It is not restricted to two or any other limited number of loudspeakers. In principle, stereo can mean any spatial sound reproduction system with more than one loudspeaker. However in this thesis, it is assumed that it differs from sound field reconstruction techniques, in that the aim is not to reconstruct a sound field in an expanded listening area. Figure 3-2 shows the standard setup for two-channel stereo.

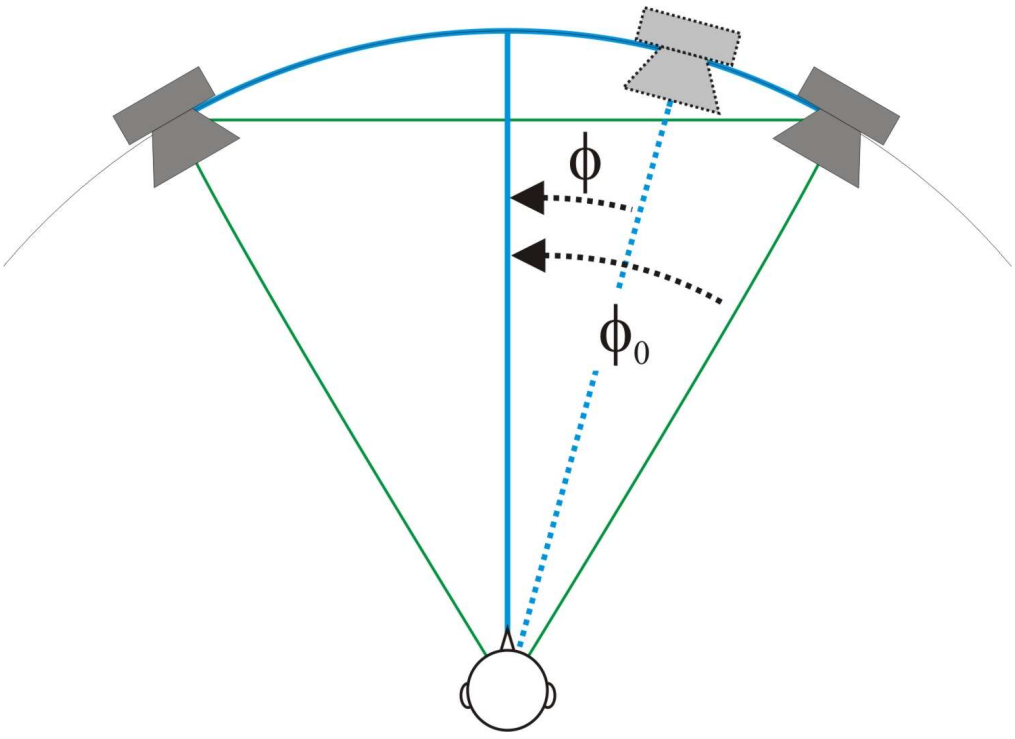


Figure 3-2: Standard setup for two-channel stereophony. One possible 'phantom' source location is illustrated by the dotted loudspeaker. The listener is located at one corner of the equilateral triangle, in the so-called 'sweet spot'. The offset angle of the loudspeakers is 60°. ϕ_0 = 30°, ϕ is the phantom source shift or panning angle.

3.3 Origin of stereophony

Two-channel stereophony marked a major step forward for spatial sound reproduction. From the outset, which can be considered the two-channel telephone transmissions by Ader in 1881 (Hertz, 1981, see Figure 3-1), it was realised that two-channel reproduction offered significantly more than just two simultaneous monophonic channels.

In the 1930s, two parallel developments took place that helped introduce modern stereophony. These developments are examples of two fundamentally different views on stereophonic perception, and together they lead into one of the main discussions of this thesis.

In America Steinberg, Snow and Fletcher (Steinberg and Snow, 1934) from Bell Laboratories explored the 'acoustic curtain', see Figure 3-3.

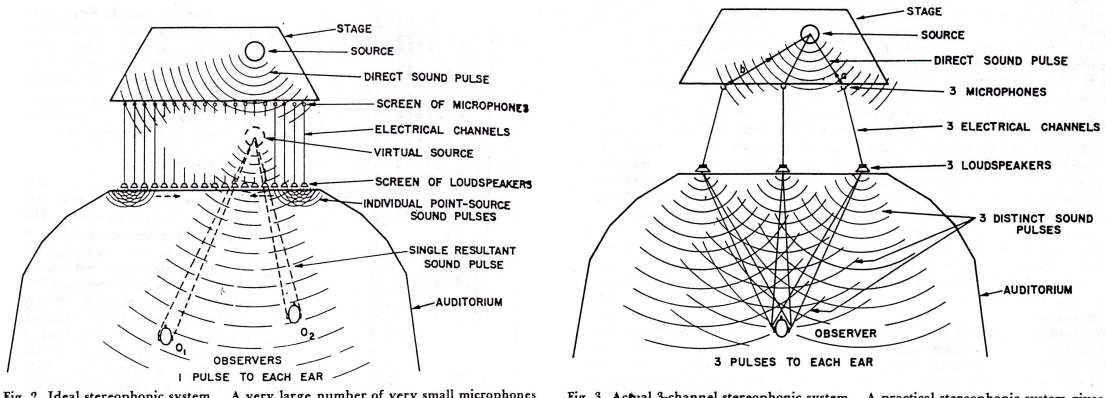


Fig. 2. Ideal stereophonic system. A very large number of very small microphones and loudspeakers would give a perfect reproduction of the original sound.

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 3-3: from (Snow, 1953): Early implementation of stereo (and actually a precursor of wavefield synthesis, see chapter 4.2.1): desired (left) and implemented (right) stereophonic system of Snow, Steinberg and Fletcher.

They aimed to transport the acoustical cues arising from sources in the recording venue to a reproduction room using microphone and loudspeaker arrays. Snow described their ideas 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)

These scientists quickly noticed that, due to technical constraints, it would not be feasible to put their ideas into practice. As a compromise, they limited the practical system to three channels, accepting that the original aim of recreating the real sound field would no longer be ful-

filled. The three-channel stereophony produced in this way was therefore not created as a result of a mathematical analysis of the sound field, but rather as an engineering compromise. Its directional effect is based on perceptual phenomena such as the precedence effect and level and time difference stereophony.

In contrast, Blumlein (1933) aimed at a proportional reproduction of the directional image of the recorded scene by recreating the original physical auditory cues. He found that in a stereophonic setup, the intensity[5] differences between the loudspeakers are converted into phase differences at the listener's ears below a certain limit frequency. Above this frequency, intensity differences between the loudspeakers would translate to similar differences between the ears. Thus both important cues for source localisation would be synthesised correctly: the low frequency phase differences and the high frequency intensity differences.

Blumlein's ideas are the basis of the summing localisation theory, see section 3.6.1. They lead to a computable stereophonic reproduction between the loudspeakers. He proposed a coincident microphone setup for capturing intensity differences, consisting of two bidirectional microphones at an angle of 90°, which nowadays is known as the 'Blumlein pair'.

3.4 Two principle ways of interpreting stereophonic perception

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 3-3] *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."*

Researchers who observe contradictions in the generally accepted summing localisation theory quote this statement by Snow. Indeed, this chapter shows that this theory can only partially explain phantom source perception, and that discrepancies between prediction and actual perception do exist. The most important difference between the summing localisation theory and other approaches is the interpretation of the way in which the loudspeaker signals

---

[5] *Original quote from Blumlein (1933). The term 'intensity' is widely used when describing sound pressure level differences that determine the perceived direction of a phantom source. This usage is questionable since sound pressure is perceived by the auditory system rather than intensity as such; a term such as 'sound pressure level differences' makes more sense (Sengpiel, 2007a).*

are evaluated by the auditory system. There are two fundamentally different approaches, and two fundamentally different types of source being created:

1. ‘*Virtual source* through summing’:
   The loudspeaker signals physically add up at the ears and are evaluated as one sound event. The virtual source can be considered a substitute source.

2. ‘*Phantom source* after separate discrimination’:
   The loudspeaker signals can be evaluated separately. They may form two separate sound events which result in one auditory event.

The summing localisation theory assumes virtual sources (1) through a summing of the loudspeaker signals at the ears (see section 3.6.1).

The association model by Theile (see section 3.6.2) assumes phantom sources (2).

A further third approach cannot be assigned to one of these principles. The hypothesis of a binaural decolouration that is applied to stereophonic signals is discussed in section 3.6.3.

The different perception principles are introduced in section 3.6 after the following discussion of the phantom source’s properties. From this point onwards, the term 'phantom source' will be used for the stereophonic source, regardless of the assumed approach of the perception mechanism.

3.5 Phantom source properties

3.5.1 Directional imaging

Two loudspeakers can create a phantom source. Level and time differences between the loudspeaker signals determine the perceived direction of this source. These interchannel differences can be chosen such that a particular phantom source direction will be created. The mapping law (or ‘panning law’) between level and/or time differences and the resulting phantom source shift can be derived by empirical methods. The lateral displacement of the phantom source due to interchannel level differences ΔL and time differences Δt have been measured by various authors. Typical phantom source shift curves $A_{\Delta L} = f(\Delta L)$ and $A_{\Delta t} = f(\Delta t)$ are plotted in Figure 3-4 and Figure 3-5. The phantom source location is given as the shift from the middle position relative to half the loudspeaker base. 100% phantom source shift corresponds to a phantom source localised in one loudspeaker. These figures also show that with-

out a large error, a phantom source shift from 0% to 75% can be assumed to be proportional to the interchannel difference. Thus, in this central area, a constant shift factor exists.

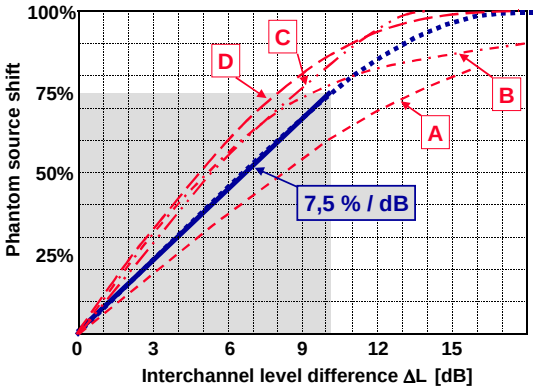


Figure 3-4: from (Wittek and Theile, 2002): Relative phantom source shift $A_{\Delta L}$ = f (ΔL). Dotted thick curve after (Wittek and Theile, 2000b) with shift factor $Z_{\Delta L}$ = 7.5 %/dB. The exact value in Wittek and Theile (2000b) is 7.3%.

curve A: after Leakey (1960)
curve B: after Mertens (1965)
curve C: after Brittain and Leakey (1956)
curve D: after Simonsen cited by Williams (2000)

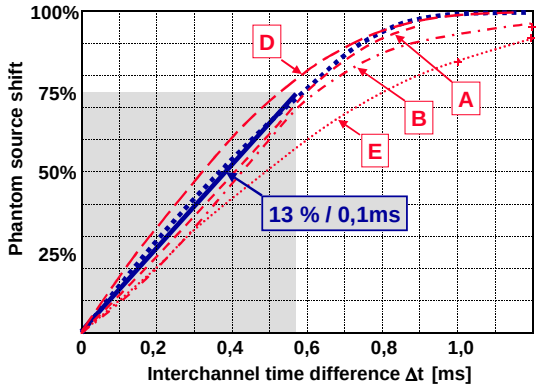


Figure 3-5: from (Wittek and Theile, 2002): Relative phantom source shift $A_{\Delta t}$ = f (Δt). Dotted thick curve after (Wittek and Theile, 2000b) with shift factor $Z_{\Delta t}$ = 13 %/0.1ms. The exact value in Wittek and Theile (2000b) is 12.7%.

curve A: after Leakey (1960)
curve B: after Mertens (1965)
curve D: after Simonsen cited by Williams (2000)
curve E: after Sengpiel (2007b)

It has been shown (Williams, 1984; Theile, 1990) that the phantom source shift can easily be calculated. The calculation can be made independently of the offset angle (see Figure 3-2) of the stereo setup. This is because the relative phantom source shift caused by a certain inter-channel difference is independent of the offset angle (Theile, 2001; see also Martin et al., 1999a). Therefore, it is given as a percent value, where 100% corresponds to full phantom source shift and localisation in one loudspeaker, and 0% corresponds to a localisation in the centre between the two loudspeakers.

The calculation is also possible in cases where a combination of level and time difference is effective. Thus, the phantom source shift A can be calculated as follows:

$$A(\Delta L, \Delta t) = A(\Delta L) + A(\Delta t);$$ (Theile, 1990); valid only for A < 75%

where

$$A(\Delta L) = \Delta L \cdot 7.3\%/dB;$$

$$A(\Delta t) = \Delta t \cdot 12.7\%/0.1ms;$$

This linear approximation fails when the resulting source shift exceeds approximately 75%. A general approximation for the whole phantom source base and all interchannel difference combinations can be given when a multipart approximation formula is applied (Wittek, 2000a, 2001a).

It is important at this point to identify the equivalence of interchannel level and time differences in creating a phantom source shift. Although phantom source properties are potentially different, the directional imaging curves depicted above show that both interchannel level and time differences can produce phantom source shifting. For instance, the process of 'panning' (assigning as mono signal to a certain direction in a stereo mix) can be achieved by using level differences as well as time differences or a combination of level and time differences.

There is an apparent deviation between the data from the different investigations quoted in Figure 3-4 and Figure 3-5. This deviation is mainly caused by the use of different stimuli. It can be shown that audio sources containing transient signals are localised differently to more static signals such as noise (Sengpiel, 2007c). After Pulkki (2001a), this is due to the increased weight given to ITD cues in the frequency region 700-1700 Hz in the case of transient signals, and the resulting localisation due to the direction suggested by these cues.

It should be noted that lateralisation experiments, i.e. those using headphone listening rather than loudspeakers (Blauert, 1997), lead to results contrary to those described here regarding the phantom source shift (see section 3.6.2).

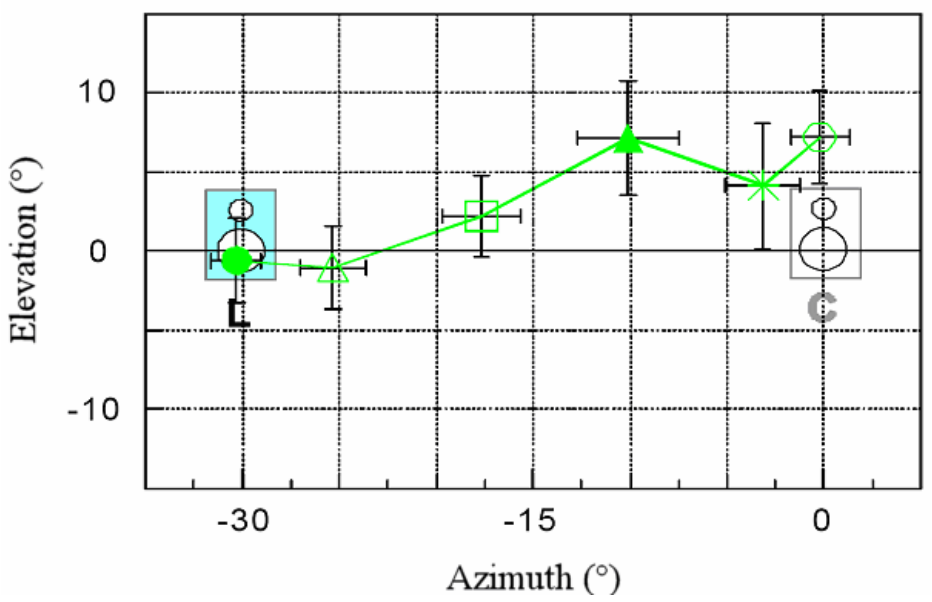


Figure 3-6: from Wittek (2000a): Azimuth and elevation of phantom sources. The diagram shows the mean of the perceived horizontal as well as vertical directions of different phantom sources including the 95% confidence interval of the mean. The two loudspeakers are located at L=(-30°;0°) and R=(30°;0°). The Center loudspeaker (C) is inactive.

A phantom source is located slightly above the line between the two loudspeakers, i.e. with a certain elevation angle relative to the horizontal plane (see e.g. Theile, 1980). This phenomenon cannot be explained by the principles of localisation in the median plane based on the directional bands (Blauert, 1997). Theile (1980) interprets the elevation as a consequence of the association model. In Figure 3-6, measurements of the azimuth and elevation directions of phantom sources from (Wittek, 2000a) are presented, which show the elevation in the median plane.

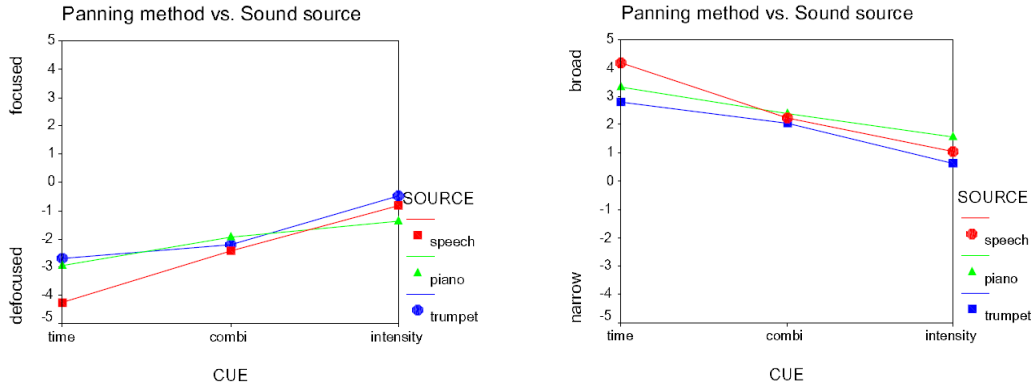
3.5.2 Quality differences between level-panning and time-panning

The preceding chapter has shown that both level- and time-differences lead to accurate phantom source imaging in terms of the reproduced directions. However, it is widely accepted that there are differences in terms of the quality of the reproduced phantom sources. Time-panning and time-difference stereophony are often presumed to produce blurry and unstable phantom sources. It is also accepted that time-difference phantom sources tend to split into several individual images. Theile (1991) argues that the decrease of the phantom source focus applies to time-panning only when a time difference larger than 0.2 ms is utilised. In (Rebscher and Theile, 1990), a reduction in the *"sharpness of the image"* is stated for this case. In his explanation, which is introduced in detail in section 3.6.2, the similarity between ear signals and loudspeaker signals regarding time- and level differences is important. Unnatural signals would give rise to an increased image focus.

An experiment was recently undertaken by Lee and Rumsey (2004), which examined the differences between time, level and combined time/level panning. They investigated both the image focus and width of the phantom source. In their results, a considerable difference between these different panning methods was detected. Figure 3-7a shows that the phantom source was perceived most focussed with level panning, less focussed with combined panning and least focussed with time panning. It should be mentioned that the focus is generally rated rather low for all phantom sources, but this is on a somewhat arbitrary scale. There is no comparison with a single reference source, hence, these results should be considered carefully. It is apparent that a perceptual difference between the panning methods depends on the phantom source shift, because for 0 % phantom source shift there is no physical difference between all panning methods. Lee and Rumsey only show the results for a phantom source at roughly 20° (this means roughly 67 % phantoms source shift), which tends to exaggerate the detected differences when generalised to all phantom sources. Furthermore, the limit of Δt = 0.2 ms for stable time panning (as mentioned by Theile) is exceeded by both the combined

and the time panning phantom sources. Hence, a deterioration of time panning and combined panning in general is not considered evident from this experiment.

The image width results of Lee and Rumsey correspond to that of the image focus.



a) Image focus **b) Image width**

Figure 3-7: from Lee and Rumsey (2004): Subjective comparison of different panning techniques (phantom source at +20°, two-channel loudspeaker setup at +/- 30°). Attributes: a) image focus, b) image width. The mean of the data is shown.

'time': Δt = 0.5 ms; ΔL = 0 dB;
'combi': Δt = 0.25 ms; ΔL = 4 dB;
'intensity': Δt = 0 ms; ΔL = 8 dB;

3.5.3 Sound colour

The properties of a phantom source are different from those of a real source at the same location. This was discussed in a number of investigations (e.g. Silzle and Theile, 1990; Pulkki, 2001b). The perceptual attributes that differ include locatedness, localisation accuracy, image-focus, width, sound colour and robustness. The difference in sound colour is regarded a key parameter for an investigation into the perception mechanism, as the effective ear spectra for sound colour perception differ substantially between different perception mechanisms and thus the perceived sound colour would enable a conclusion on the applied perception mechanism (see 3.5.4).

First, the ear spectra will be analysed, being one important physical basis for the perception of the sound colour.

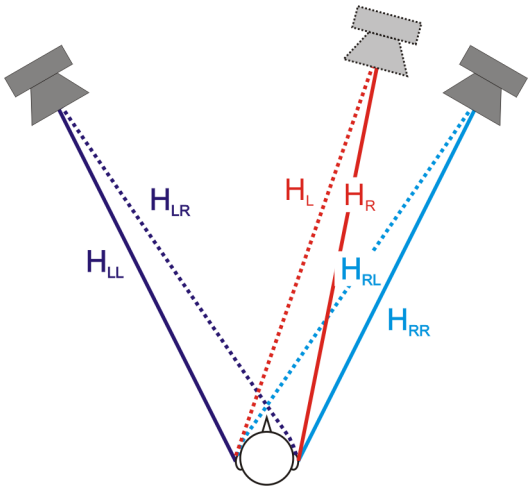


Figure 3-8: Generation of ear signals in a standard stereo setup. H_{LL} and H_{RR} are the ipsilateral[6], H_{LR} and H_{RL} the contralateral[2] ear signals. A virtual source and its ear signals (H_L and H_R) are shown which is at the same location as the phantom source (produced by level panning).

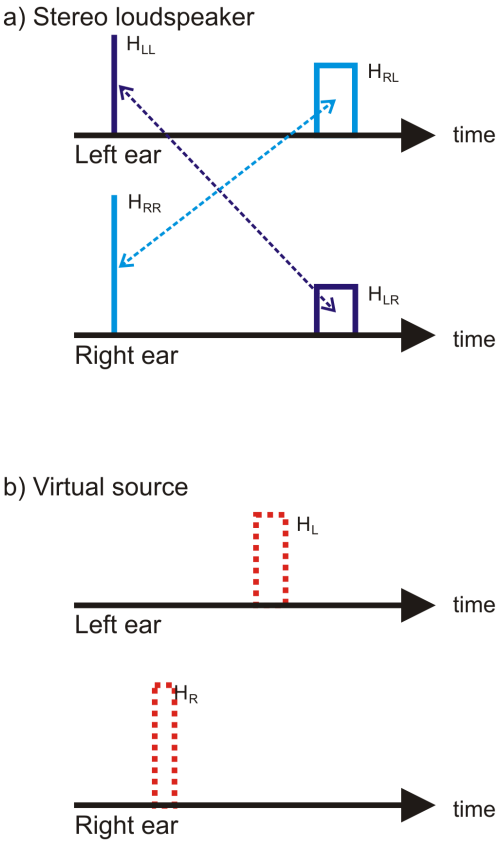


Figure 3-9: Sketches of the ear signals in the time domain:
a) Simplified sketch of the ear signals arising from stereophonic reproduction as in Figure 3-8 (produced by level panning)
b) Simplified sketch of the ear signals arising from the ideal substitute source in Figure 3-8 which is at the same position as the phantom source in a)

Figure 3-8 shows a standard stereo setup similar to Figure 3-2. Here, the signal paths from loudspeakers to ears are drawn such that the ipsilateral (solid lines) and the contralateral (dotted lines) ear signals can be differentiated. An exemplary 'ideal' substitute source would be localised at the illustrated location and create the red source-ear paths. Figure 3-9 shows the corresponding illustration of the ear signals in the time domain. The signals are shown for a level-panned phantom source. The contributions of the loudspeakers can be identified in the time domain illustration. At each ear, the ipsilateral loudspeaker creates the first signal and

[6] *Ipsilateral ear = the ear at the same side as the source. Contralateral ear = the ear at the other side of the source*

the contralateral loudspeaker signal arrives after the corresponding interaural time difference. The level of these contributions depends on the head shadowing as well as the interchannel level difference. The summing of the loudspeaker signals at the ears and the resulting comb filtering can be retraced. The comb filter's properties depend on the time difference between the superimposed signals and their level difference.

The ear signals can further be analysed by a simulation of the binaural room transfer functions (BRTF) which represent the spectrum of the ear signals created by a source or a certain loudspeaker setup in a room (in this case an anechoic chamber). Figure 3-11 (top diagram) shows the resulting BRTF for a level-panned phantom source. Figure 3-10 (top diagram) shows the BRTF of a real source at the same position as the phantom source. The comparison is easier by means of the bottom diagram in Figure 3-11. It shows the difference between these spectra. The strong comb filtering in the contralateral ear signal is apparent (see Figure 3-9a, left ear). The strong first notch is at f = 1.7 kHz.

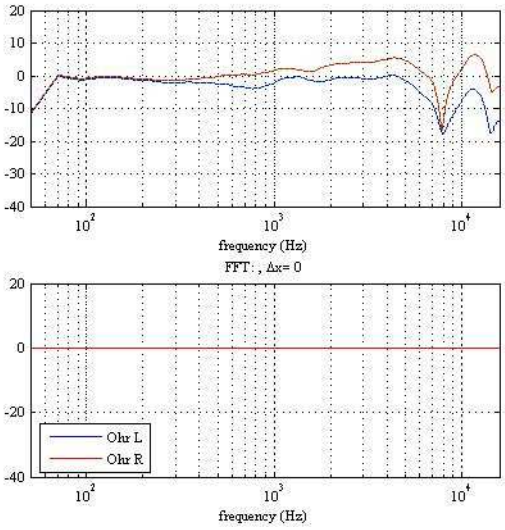


Figure 3-10: Real source 15° on the right.

red: ipsilateral (right) ear signal
blue: contralateral (left) ear signal

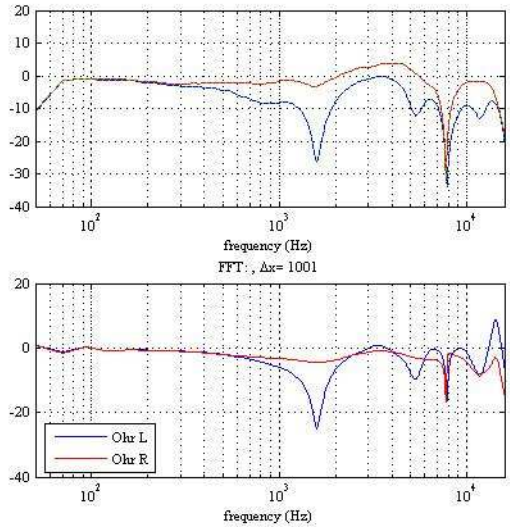
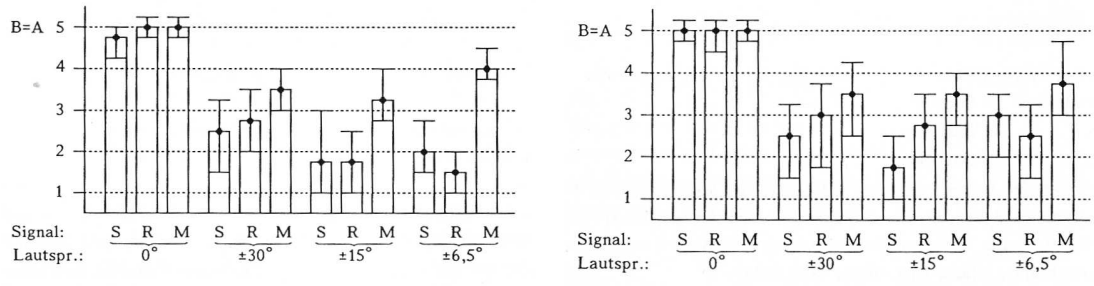


Figure 3-11: Standard stereo setup, ΔL(L/R)= -7dB.
red: ipsilateral (right) ear signal
blue: contralateral (left) ear signal

Top diagrams: binaural room transfer function (BRTF), bottom diagrams: difference between this BRTF and the BRTF of a real source at the same location. The level on the y-axis is given in dB.

Sound colour perception in stereophony has so far not been studied in great detail. Silzle and Theile (1990) compared a real source and phantom sources created by loudspeaker setups of different offset angle in order to study the influence of the offset angle on the attributes sound

colour and image focus. Their results are shown in Figure 3-12a (sound colour) and Figure 3-12b (image focus).



a) Sound colour **b) Image focus**

Figure 3-12: from Silzle and Theile (1990): Comparison of a real source at 0° and a phantom source (Δt, ΔL = 0) reproduced on loudspeaker setups of different offset angle (0°, ±30°, ±15°, ±6.5°). Attributes: a) sound colour, b) image focus. Test signals: voice (S), noise (R), music (M). The bars show the mean and the 95% confidence interval on a 5-grade-scale (5=no difference, 1=very different).

It can be seen that both sound colour and focus of real and phantom sources are significantly different in all test cases. The results are dependent on the test signal. The trend of the data for voice and noise shows that the sound colour seems to decrease in similarity to the real source with a decreasing offset angle of the loudspeaker setup. The focus does not show a significant dependence on the chosen offset angle.

The most precise investigations regarding phantom source colouration were undertaken by Pulkki (2001c) and Ono et al. (2001, 2002). In these investigations, the perceived sound colour was measured by a 'method of adjustment', i.e. the subjects could adjust ('equalise') the spectrum of the virtual source stimulus by narrowband filters such that the perceived timbre matched that of the real source. The timbral difference between real source and phantom source could thus be measured. Simultaneously, the ear signals were recorded by miniature microphones at the ear canal entrance. This analysis of the actual ear signals revealed if the measured timbral differences were caused by the spectral differences in the ear signals.

Figure 3-13 and Figure 3-14 show one result of these studies. The graphs show the 'composite loudness level (CLL) difference' which can be interpreted as the remaining spectral difference between a real and a virtual source after they have been brought in accordance regarding the perceived sound colour. In this example, a phantom source between the loudspeakers at 0° is considered. Both in the case of narrowband (Figure 3-13) and wideband sources (Figure 3-14) a significant spectral difference exists, meaning that the perception of the virtual source differs compared with real source perception. The frequency at which the largest mismatch arises (around 1.7 kHz) matches the detected strongest notch in the ear signal spectrum shown

in Figure 3-11 (bottom diagram). This means that the strong notch produced by comb filtering was not perceived to the same extent as it could be predicted from the ear signals.

Ono's investigations also considered virtual sources created by large time delays (2 and 4 ms) between the stereo loudspeakers. In these cases, the comb filtering was again perceived less than predicted by the ear spectra.

These results underline the need for alternative interpretations of stereophonic perception, as they point to a mismatch of expected/predicted and perceived sound colour of the phantom source.

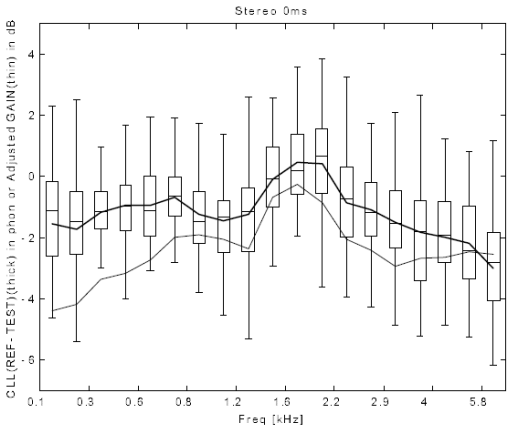


Figure 3-13: from (Ono et al., 2001): 'Composite loudness level (CLL) difference' between real and virtual sources for *bandlimited(narrowband)* noise.

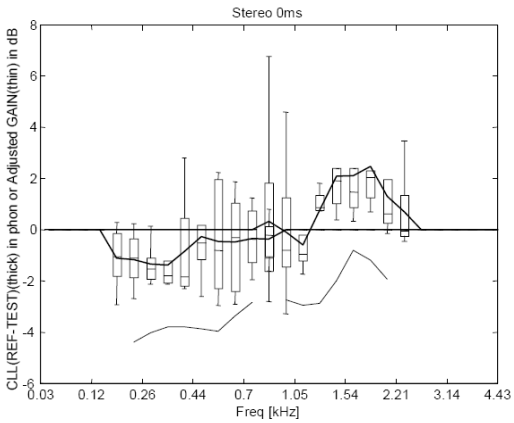


Figure 3-14: from (Ono et al., 2002): 'Composite loudness level (CLL) difference' between real and virtual sources for *wideband* noise.

The CLL is gathered by an adjustment of levels between a real and a virtual source. Thus it can be considered the remaining spectral difference between timbrally equal real and virtual sources. The boxes show the median and the 25% and 75% quartiles of the CLL. The adjusted gain is plotted by the thin line (in dB).

3.5.4 Distance

A stereophonic setup can synthesise the distance of the phantom source as long as the crucial cues for distance perception are correctly reproduced. These were introduced in chapter 2.5. According to the classification of cues described there, the available cues in stereo are 2D, and partly also 3D cues. The 2D cues (monaural cues: level, direct-to-reverberant energy ratio, frequency spectrum, interaction with other non-acoustical cues) can be reproduced in stereophonic and even monophonic reproduction (Theile, 2001). The 3D cues (reflection pattern, binaural differences) can only partially be reproduced in stereo. The binaural differences

due to source distance cannot be created because the loudspeaker distance is constant, and the correct wave front curvature (= wave front curvature according to the desired distance) of the phantom source is not synthesised.

The creation of a reflection pattern according to a natural source is possible as long as lateral reflections can be reproduced (Theile, 2001; Griesinger, 2001). This is possible in multichannel stereo, as shown in Figure 3-15. In this figure, the reflection patterns produced by the two stereo configurations are illustrated. It can be seen that correct lateral reflections can be reproduced as long as loudspeakers are present in this direction.

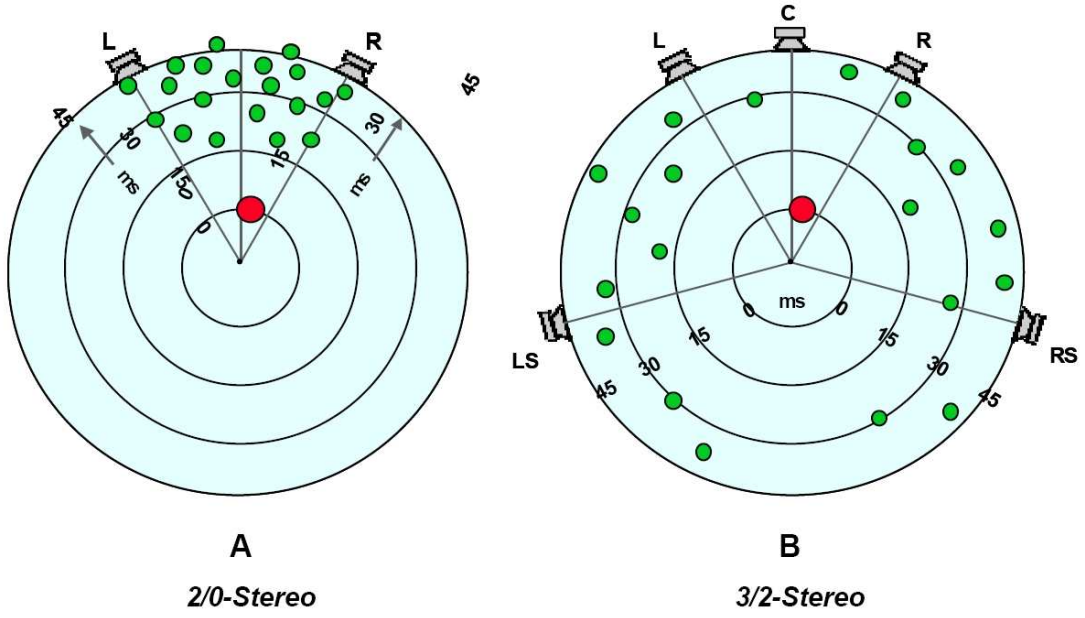


Figure 3-15: from Theile (2001): Spatial and temporal distribution of the reflection pattern. Two different stereophonic standard setups are shown: two-channel stereo (2/0; A) and multichannel surround (3/2 stereo; B). The diagrams show the spatial distribution of the direct sound (red dot) and the reflections (green dots). Moreover, a time axis is installed which proceeds from the centre of the circle outwards. Hence, timing and spacing of the reproduced reflection pattern are illustrated.

The 2½D cues cannot be reproduced in stereo. The directional imaging in stereo is based on level and time differences which depend on the listening position (Wittek, 2001a). This means that stereo creates a 'sweet spot', i.e. a single location at which the desired directional image is perceived. Listening positions outside this location will result in a distorted directional image. Stereophonic reproduction cannot provide correct distance perception cues for a moving listener. When stereophonic reproduction is intended for an extended listening area, certain measures can be taken in order to avoid a disturbing distortion of the directional image at the cost of the spatial fidelity. An example is cinema, where only the centre channel is used for the dialogue in order to prevent an erroneous localisation for the audience.

Distances closer than the distance of the loudspeakers cannot be reproduced in stereo.

3.6 Perception theory

3.6.1 Summing localisation

The generally accepted theory to explain the creation of a phantom source is that of ‘summing localisation’. According to this theory, stereophony works due to a physical summing of the two loudspeaker signals building a virtual source. Summing localisation has been investigated for a considerable amount of time, for example by de Boer (1940) and Wendt (1963), for other sources see Blauert (1997).

The expression ‘summing localisation’ describes the way in which stereophonic perception is considered: the virtual source is understood as a substitute sound source. The sum of the two loudspeaker signals at each ear entrance forms the same binaural localisation cues as would be created by a correspondingly located real loudspeaker. This was first described by Blumlein (1933, see section 3.3), who derived stereophonic panning and microphone recording laws based on theoretical considerations. In addition, the theory of ambisonics, invented and described by Gerzon (1973), is based on the assumption of a perceptually relevant summing of the loudspeaker signals. Consequently, the phantom source in this case would be better called a virtual source, because the sound field of the source is assumed to be reconstructed in a sufficiently accurate way (see section 3.4). In contrast, the term ‘phantom source’ describes a perceived auditory event which is not based on related natural binaural cues. Rather, it is assumed to result from coactions of specific mechanisms of the auditory system which have developed during natural listening, see section 3.6.2.

Physical synthesis through level-panning

It can be shown that the summing of two or more loudspeaker signals can lead to a wave field which is congruent to that of a real source in a certain direction. As described by Blumlein (1933), this is true (though with constraints) for level-panned virtual sources and thus for coincident microphone recording. In the centre of the listening area (the infinitely small ‘sweet spot’), the sound field is correctly synthesised for all frequencies. Due to the dimension of the listener’s head, in real listening the sound field is only correct up to a certain limit frequency of around 1000 Hz. For these low frequencies, the well known ‘stereophonic law of sines’ determines the direction of the created virtual source (see e.g. Lipshitz, 1986). This formula is mathematically derived through a simplified model of the head geometry, without any consideration of head shadowing.

Stereophonic law of sines:

$$\sin\phi = \frac{L-R}{L+R}\sin\phi_0;$$

where:
L, R are the gains of the left and right loudspeakers,
ϕ is the angle of the virtual source,
ϕ_0 is the angle of the loudspeakers, see Figure 3-2.

The meaning of the physical reconstruction of a source's wave field can be illustrated by the snapshots of a 500 Hz sine wave reproduced in a 3m · 2m room. The real source (Figure 3-16) is located at the position $(x;y) = (0.43;0)$. The snapshot shows the resulting sound pressure distribution in the room after the sine wave has emanated from the source. The grey-shading corresponds to the amplitude of the wave. The two loudspeakers of the stereo setup (Figure 3-17) are located at the positions $(-0.86;0)$ and $(0.86;0)$. The virtual source is shifted to the right by an interchannel level difference ΔL(L/R) of -7 dB. Consequently, the location of real source and virtual source are roughly the same.

The comparison of real and virtual source shows the conformity of the wave fields, albeit only for a small area of correct synthesis, and only for low frequencies such as in these simulations.

Another representation helps to analyse frequency-dependent characteristics of the virtual source. The interaural time difference (ITD) can be considered the peak of the interaural cross correlation (IACC[7]). In Figure 3-18 the IACC of a real source at 15° right from the median plane is shown. In this and all other simulations, the listener is located in the sweet spot and his head aims forward. The positive maximum of the IACC, which can be considered the perceived ITD, is highlighted in red. The constant ITD in each of the low and high frequency regions is clearly visible.

---

[7] *IACC: Interaural Cross Correlation. The IACC shows the similarity between the two ear signals for each frequency band. The frequency-dependent time offset of the highest positive peak of the IACC can be interpreted the interaural time difference ITD. Thus, the IACC can help in analysing the goodness of the match of the ITD being an important parameter for source localisation. IACC simulations of WFS virtual sources were proposed by Wegmann (2005).*

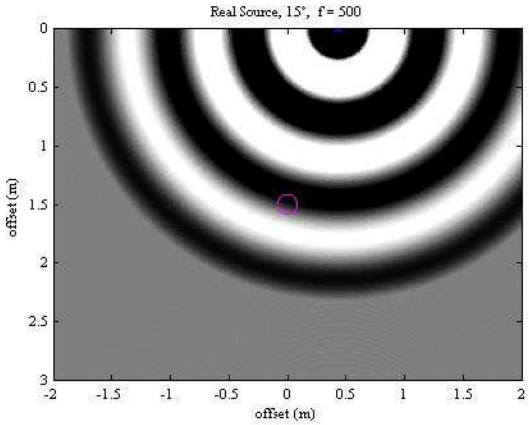


Figure 3-16: Snapshot of the pressure field of a 500 Hz sine wave reproduced by a single, real source at 15°.
The head of the listener in the sweet spot is marked by the purple circle.

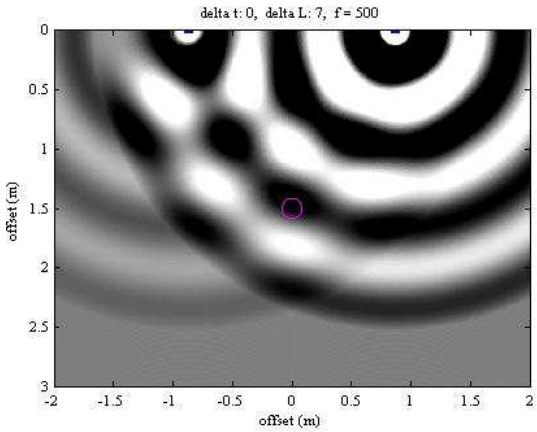


Figure 3-17: Snapshot of the pressure field of a 500 Hz sine wave reproduced by a standard stereo setup (Δt(L/R)= 0 ms; ΔL(L/R)= -7 dB). The head of the listener in the sweet spot is marked by the purple circle.

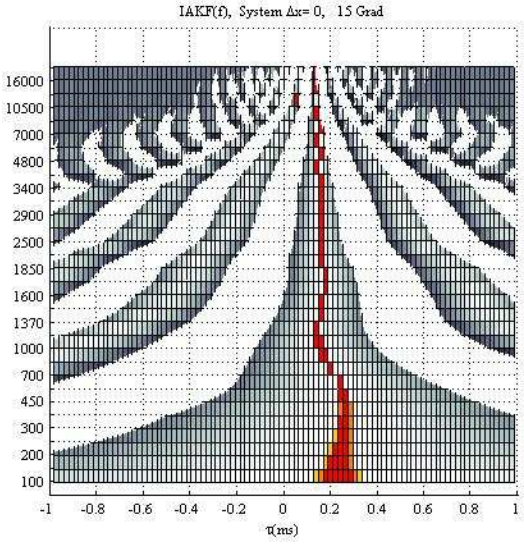


Figure 3-18: IACC of a real source at 15°.

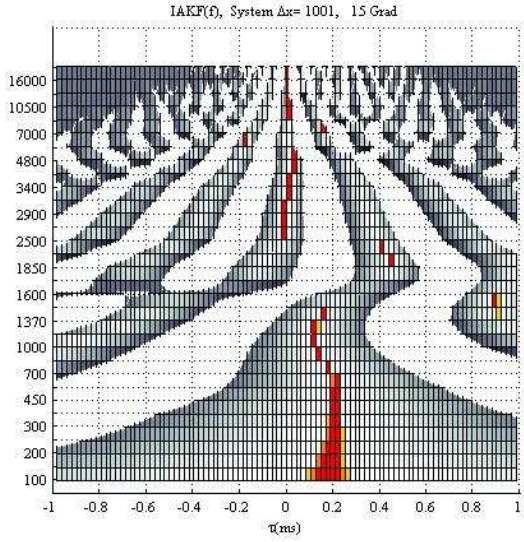


Figure 3-19: IACC of a level-panned phantom source, standard stereo setup. Δt(L/R)= 0 ms. ΔL(L/R)= -7dB.

The positive maximum of the IACC, which can be considered the perceived ITD, is highlighted in red colour. Only the positive values of the IACC are shown for a better transparency of the representation. The grey-shading shows the amplitude of the IACC from dark to white. The IACC was normalised at each frequency band.

In Figure 3-19, the IACC of a level-panned phantom source in the same direction is shown. A defect of the ITD above approximately 1500 Hz is apparent. Due to the influence of head shadowing above this frequency, summing of the ear signals does not occur, and only the ipsilateral loudspeaker signals are valid. Consequently, the ITD is similar to the interchannel time difference Δt. For level-panned sources Δt is 0. Taking into account the known property of the auditory system to rely on interaural level differences (ILD) at higher frequencies this may not have a negative effect.

Figure 3-10 shows the BRTF (ear spectrum) of a real source at 15° to the right of the median plane as a reference. From Figure 3-11, which shows the level-panned phantom source in the same direction, the deviations in the frequency response can be deduced. The simulation of the normalised BRTF (bottom diagrams) shows these deviations. The frequency region below approximately 1500 Hz is synthesised almost perfectly; above 1500 Hz the crucial ILD is still present, even though suffering from modest comb filtering.

These analyses can be complemented by investigations into the resulting directional image of the phantom sources. Pulkki performed several studies applying level-panned phantom sources and various types of narrow- and broadband stimuli (Pulkki et al., 1999a, 1999b). The experimental result from (Pulkki and Karjalainen, 2001c) is depicted in Figure 3-20, which shows the interchannel level difference necessary for a phantom source shift of 15°. The test signals are noise bursts in 1/3-octave bands. The right vertical axis shows level differences, the left vertical axis gives the corresponding theoretical phantom source shift. The theoretical source shift is calculated using the ‘tangent law’, which gives a slightly more precise approximation compared with the mentioned ‘law of sines’. This is due to an incorporation of the circular head shape and the corresponding wave paths. It can be seen that an agreement between theory and experimental results exists for frequencies below 1000 Hz.

For frequencies above 1000 Hz, the tangent law does not lead to good agreement. This could potentially be less relevant due to the increasing importance of ILD for localisation at higher frequencies. It is known that ITD are evaluated also for higher frequencies, by an interpretation of the stimulus envelopes (Blauert, 1997). However, the ILD can override the ITD at higher frequencies, as shown by Wightman and Kistler (1990), even more in the case of a consistent ILD and inconsistent ITD. Figure 3-11 shows some similarity between real source and virtual source transfer functions regarding the high-frequency ILD. Pulkki and Karjalainen (2001c) incorporated an analysis based on the existing directional cues for the real source case. In Figure 3-21, the ITD and ILD are converted such that they represent the corresponding angle at which a real source would be localised when it had these differences. An ideal virtual source would have cues corresponding to those of a real source at a certain angle

at all frequencies. Figure 3-21 shows that the level-panned virtual source does not fulfil this requirement. However, it shows a good agreement of the ITD cues for frequencies below 1000 Hz and also a good agreement of ILD cues for frequencies above 3500 Hz. Consequently, it can be argued that the relevant cues in these two frequency regions are synthesised correctly and therefore provide a successful localisation.

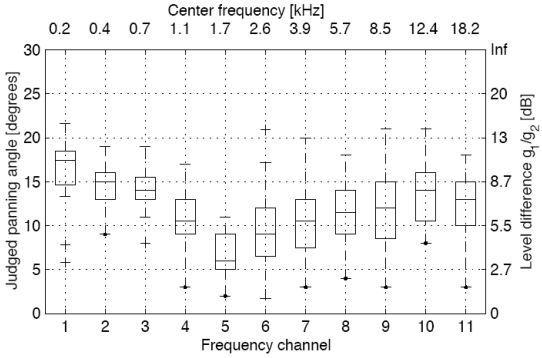


Figure 3-20: from Pulkki and Karjalainen (2001c): Adjusted interchannel level difference (right axis) on a standard stereo setup to match the location of a 15° real source. Left axis gives theoretical panning angle after the tangent law. Test signal: 1/3-octave filtered noise bursts.
The boxes show the median and the 25% and 75% quartiles.

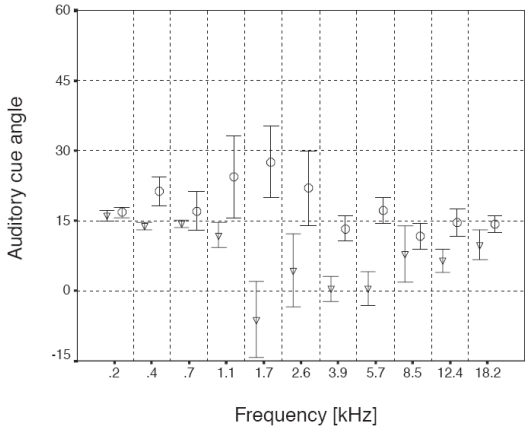


Figure 3-21: from Pulkki and Karjalainen (2001c): The results from Figure 3-20 are converted into 'auditory cue angles' by comparing them with the auditory cues (ITD, ILD) of a real source. Hence, the results show the agreement of the special level-panned virtual source created by the level differences from Figure 3-20 with a real source at 15° regarding ITD cues (triangles) and ILD cues (circles).
Auditory cues simulation was conducted with 20 individual HRTF sets, symmetrically to left and right side. The mean and the 95% confidence interval of the data are shown.

Physical synthesis through time-panning

As depicted in section 3.5.1, both level- and time-panning are suitable for the directional imaging of a phantom source. To a certain extent, this observation disagrees with the theory of summing localisation. Although it is well-known that pure time-panning results in more unstable and blurry phantom sources (for $\Delta t > 0.2$ ms; see Theile, 1991; Rebscher and Theile, 1990), the general functioning of interchannel time differences for stereophonic imaging seems to be apparent. Stereophonic microphone setups creating combined level and time differences are known for their favourable properties regarding the localisation of the resulting phantom sources. The mathematic derivation (see e.g. Lipshitz, 1986), however, does not lead

to this conclusion. A fundamental difference of the type of stereophonic perception may therefore be discussed.

Both combined time- and level-panning as well as pure time-panning are discussed based on the analytical descriptions of the preceding section. The created phantom source still is approximately 15° on the right side of the median plane. The location of the phantom sources is calculated based on the experimental results of Wittek (2001a, see 3.5.1 and Figure 3-4, Figure 3-5). The wavefield snapshots from Figure 3-22 and Figure 3-23 show that with increasing time difference the wave field looses its similarity to the real source case.

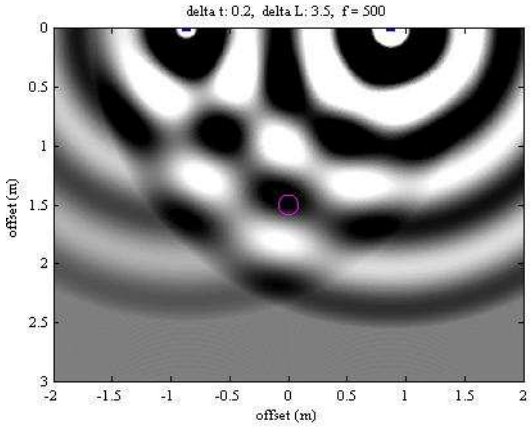


Figure 3-22: Snapshot pressure field of a 500 Hz-Sine wave reproduced on a standard stereo setup.
Δt(L/R)= 0.2 ms; ΔL(L/R)= -3.5 dB.
The head of the listener is marked by the purple circle.

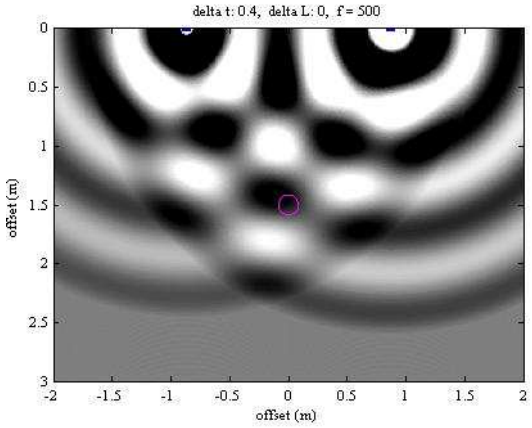


Figure 3-23: Snapshot pressure field of a 500 Hz-Sine wave reproduced on a standard stereo setup.
Δt(L/R)= 0.4 ms; ΔL(L/R)= 0 dB.
The head of the listener is marked by the purple circle.

This failure is even more apparent when Figure 3-24 and Figure 3-25 are analysed: with decreasing interchannel level difference, the low frequency ITD gets lost compared with the real source in Figure 3-18. It can be shown by calculations that the interchannel time difference has virtually no impact on the low frequency ITD.

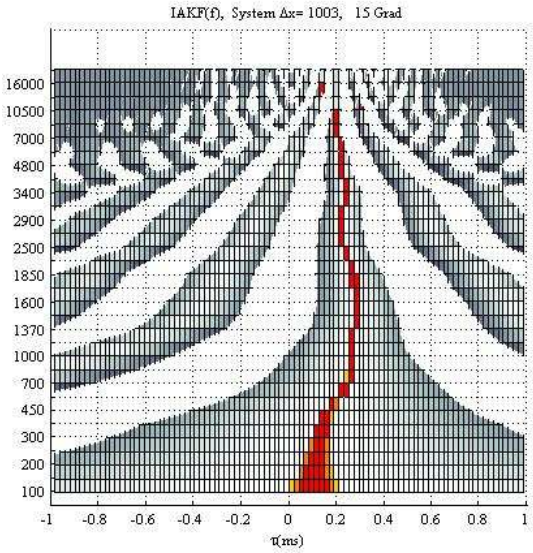


Figure 3-24: IACC of a phantom source produced by combined level- and time-panning, standard stereo setup. Δt(L/R)= 0.2 ms; ΔL(L/R)= -3.5 dB.

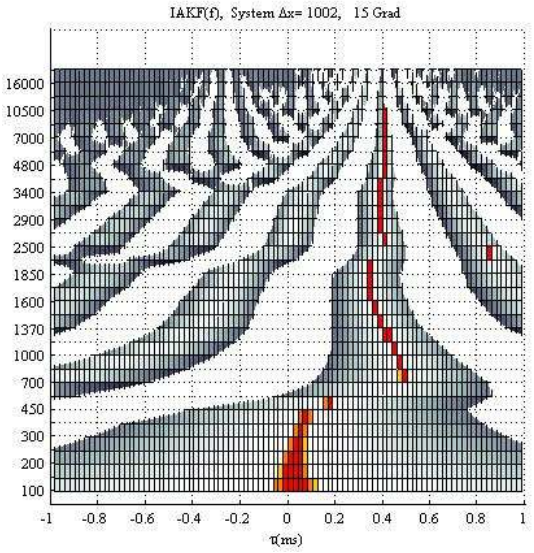


Figure 3-25: IACC of a time-panned phantom source, standard stereo setup. Δt(L/R)= 0.4 ms; ΔL(L/R)= 0 dB.

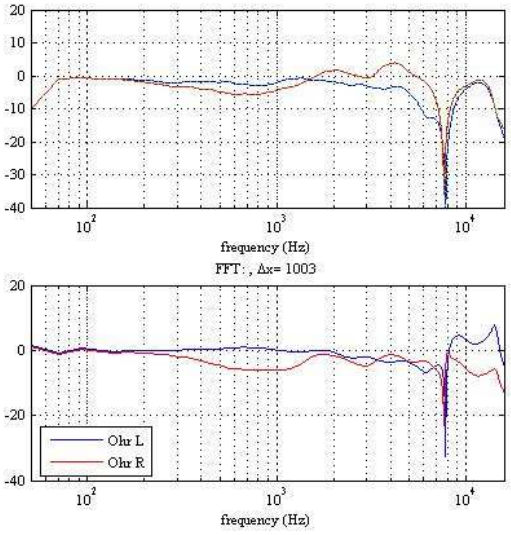


Figure 3-26: Standard stereo setup. Δt(L/R)= 0.2 ms; ΔL(L/R)= -3.5 dB.

red: ipsilateral (right) ear signal
blue: contralateral (left) ear signal

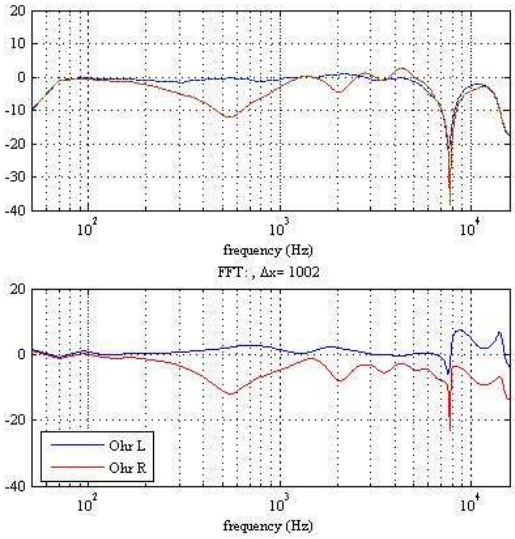


Figure 3-27: Standard stereo setup. Δt(L/R)= 0.4 ms; ΔL(L/R)= 0 dB.

red: ipsilateral (right) ear signal
blue: contralateral (left) ear signal

Top diagram: binaural room transfer function (BRTF), bottom diagram: difference between this BRTF and the BRTF of a real source at the same location. The level on the y-axis is given in dB.

For frequencies above ca. 3 kHz, the ITDs are congruent with the interchannel time differences due to the influence of head shadowing which prevents the summing of the loudspeaker signals. This means that time-panning ($\Delta t_{max} \geq 1$ ms) can also create high frequency ITDs which are larger than in reality ($\Delta t_{max} \sim 600\mu s$). In the case of combined time- and level-panning, a smaller interchannel time difference is created, and therefore, a close to natural high-frequency ITD can be produced. Natural ITDs are produced by means of stereophonic microphone setups that are spaced in ear distance, as for example the ORTF setup (2 cardioids at 110°, spaced by 17 cm). In the same way, at higher frequencies, the interchannel level difference ΔL mainly determines the ILD. Hence, natural ILDs are produced by setups producing ΔL similar to natural listening. Theile (1991) suggests the sphere microphone to produce natural phantom source imaging because of its property of producing ear signal-like interchannel differences, whilst at the same time avoiding the colouration that would be produced by the pinna and ear canal of an actual dummy head.

The analysis of the created ILD is shown in Figure 3-26 and Figure 3-27. The ipsilateral ear signal (right ear, red colour) is weaker than in the case of the real source, and prone to comb filtering. The top pictures show that the contralateral ear (left ear, blue colour) has a higher level than the ipsilateral ear in more or less broad frequency regions.

Both the pure time-panning phantom source as well as the phantom source derived by a combined time- and level-panning show no agreement between physical data and empirical observations regarding directional imaging (see section 3.5.1 and Figure 3-4, Figure 3-5). Neither the low frequency ITD nor the high frequency ILD contain any evidence for the actually perceived phantom source direction. High frequency ITD cues are potentially more adequate than in the case of level-panned phantom sources. However, neither the ITD nor the ILD data observed with pure time-panning or combined time- and level-panning are comparable with corresponding real source data.

In Figure 3-28 and Figure 3-29 the IACC of a phantom source is simulated that actually contains sufficient low frequency ITD cues as well high frequency ILD cues to match the cues of a real source at 15°. In contrast to the expected virtual source location this source is actually localised at 25° (calculated after Wittek, 2001a, see section 3.5.1).

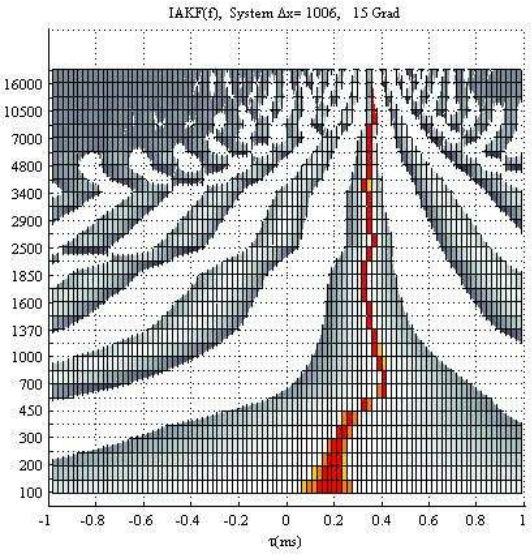


Figure 3-28: Standard stereo setup. Δt(L/R)= 0.35 ms; ΔL(L/R)= -6 dB. The source is localised at 25°.

IACC of a phantom source produced by combined level- and time-panning

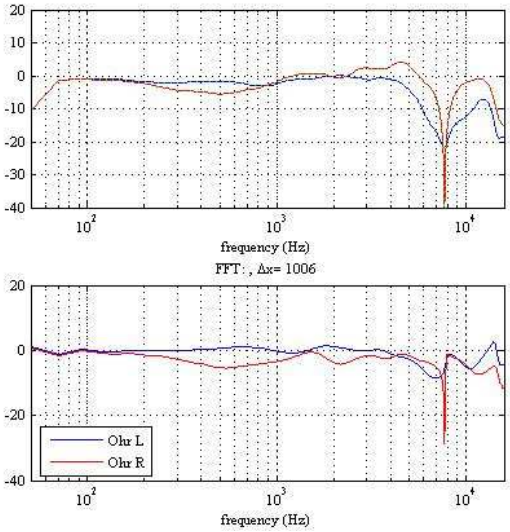


Figure 3-29: Standard stereo setup. Δt(L/R)= 0.35 ms; ΔL(L/R)= -6 dB. The source is localised at 25°.

red: ipsilateral (right) ear signal
blue: contralateral (left) ear signal

top diagram: binaural room transfer function (BRTF), bottom diagram: difference between this BRTF and the BRTF of a real source at the same location. The level on the y-axis is given in dB.

3.6.2 Association model by Theile

Theile suggested an explanation for the described phenomena of stereophonic perception as early as in 1980 with his 'association model' (Theile 1980, 1991). It is assumed that the function of the auditory spatial system is based on two different processing mechanisms, each of them in the form of an associatively guided pattern selection. A stimulus stemming from a sufficiently broadband sound source gives rise to a *location association* in the first, and to a *gestalt*[8] *association* in the second, higher-level processing stage based on auditory experience. These two stages jointly determine in every instance the properties of one or multiple simultaneous auditory events. They can be attributed to the two characteristics of 'location' and

[8] *For a definition of the 'gestalt' see chapter 2.2*

'gestalt', and they are independent of each other but always occur in pairs. Figure 3-30 shows the functional principle of the association model.

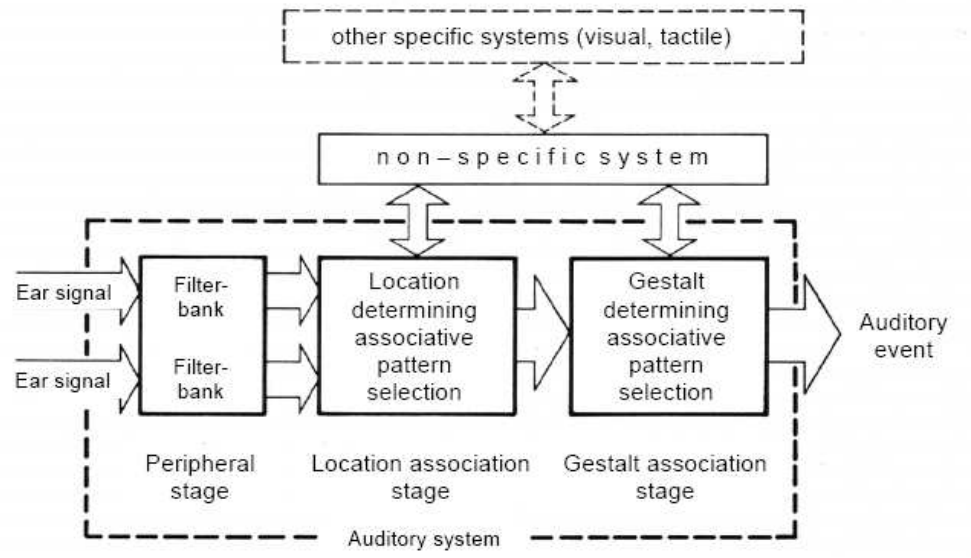


Figure 3-30: from Theile (1980): Functional principle of Theile's association model

The association model interprets localisation as a process of the selection of a localisation stimulus. A localisation stimulus will exist if sufficiently broadband ear signals can be mapped to a single sound event location in terms of their temporal and spectral properties. Under certain conditions, at least two localisation stimuli can be discriminated simultaneously in the superimposed sound field. Two individually identifiable localisation stimuli can lead to a single auditory event location. The fundamental difference between this and summing localisation theories is the suggested ability of the auditory system to separately discriminate the two loudspeaker locations in a stereophonic setup.

The association is divided into two association stages: The *location association stage* detects the source locations ('Where are the loudspeakers?'). Due to a spontaneous analysis for a known 'binaural correlation pattern' of ear signal pairs belonging to a distinct source location, the auditory system discriminates the source locations (effect of the location association stage). When the loudspeakers radiate sufficiently similar signals ('What are the loudspeakers radiating?', effect of the gestalt association stage), they result in a common auditory event.

As a result, the phantom source is not understood as a substitute sound source. The denomination 'phantom source' implies the general difference to a substitute sound source. The phantom source is understood as a fusion process of two different localisation stimuli. Due to the inverse filtering process of the location association stage postulated in (Theile, 1980), the relations between the left and right loudspeaker signals are recognised independent of the

binaural crosstalk outlined in Figure 3-31. Interchannel level and/or time differences determine the lateral displacement of the phantom source in the same way as during headphone listening. This understanding offers new approaches for the explanation of phantom source phenomena, such as perceived direction, distance, elevation, colouration and stability.

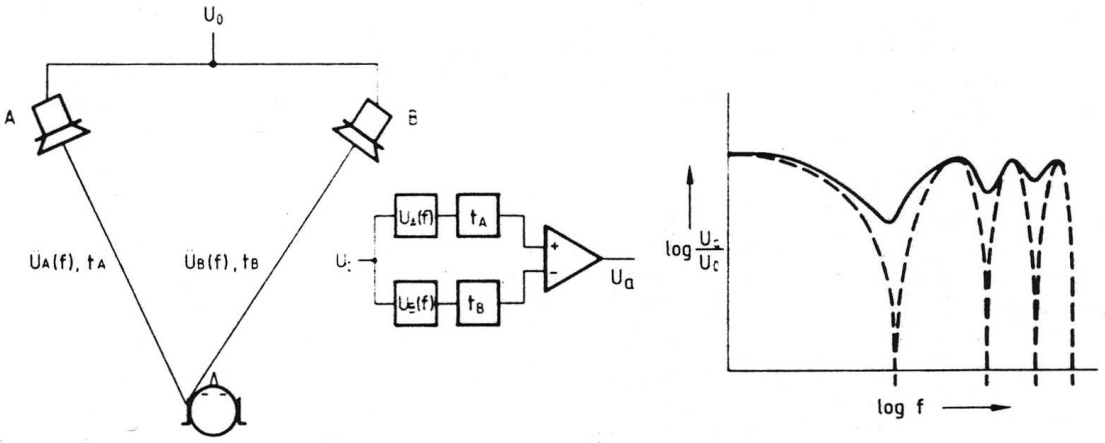


Figure 3-31: from Theile (1980): The summing of the loudspeaker signals at each ear leads to comb filtering which does not result in corresponding colouration of the phantom source. This is one objection against summing localisation theories, compare Figure 3-9 and Figure 3-11.

An important question for this investigation is the perception of colouration due to comb filtering. The comb filtering is created by the summing of the loudspeaker signals at the ears as illustrated in Figure 3-31. Figure 3-9 also shows the creation of comb filtering. In that case, a successful detection of the binaural pattern of each loudspeaker signal would mean an effective grouping of signals according to the dashed arrows. Consequently, for perception, a summing of signals at each ear does not take place. In the case of complete spatial decoding, an inverse HRTF filtering process is effective, and the sound colour is determined by the average spectrum of the two signals that the two loudspeakers individually produce. Thus, the comb filter colouration is suppressed. In his experiments, Theile showed that the suppression of colouration is dependent on the degree of completeness of the localisation stimulus discrimination. If this mechanism is impaired, the resultant spectra will have the effect postulated by the summing localisation principle. The transition between impaired and unimpaired localisation stimulus selection is continuous.

Although Theile's model was discussed (e.g. in Blauert, 1997), it has not yet gained broad acceptance. This is because it offers a more general understanding of a broad scope of phenomena in spatial hearing; however, it does not lead to a direct and possible-to-prove application. It postulates a perceptual mechanism that cannot easily be verified.

Recently, Gernemann-Paulsen et al. (2006) discussed the association model in a neuroscientific regard and collected a number of relevant publications. They came to the interesting conclusion that in spite of a number of open questions, the concept of the association model is valid, and can be related to current neuroscientific research. For instance, they quoted investigations about the so called ‘what’ and ‘where’-channels which are discussed in research about visual perception. The processing in these channels may work similarly to the proposed processing in the location (= where) and gestalt (= what) association stage.

Interpretation of phenomena by the association model

The association model provides a straightforward explanation of several phenomena of auditory perception. Although the following hypothetical reasoning is not evident, it makes a consistent picture and therefore supports the validity of the association model.

After Theile, the differentiation of the two stimulus evaluation stages corresponds entirely to the two elementary areas of auditory experience. The ear signals can be attributed to the two sound source characteristics location and gestalt, which are independent of each other but always occur pairwise. Therefore, the association model is in agreement with many phenomena related to localisation in the superimposed sound field and thus offers approaches for an explanation:

1) Phantom sources

As described in section 3.6.1, the interpretation of the phantom source as a substitute real sound source leads to discrepancies. Rather, it is assumed that due to different source locations, the auditory system can discriminate the source signals (effect of the location association stage). After the spatial decoding the stimuli are fused, because the loudspeakers radiate sufficiently similar signals (effect of the gestalt association stage). This interpretation of the perception of stereophonic source leads to a fundamentally different prediction of sound colour perception. The suppression of comb filtering can be explained as well as the functioning of the directional imaging on a stereo setup.

2) Precedence effect

The precedence effect ($\Delta t > 2$ ms) and ‘summing localisation’ (0 ms $\leq \Delta t < 1$ ms) are defined in different time delay regions. However, both phenomena can be traced back to the same evaluation of stimulus responses in the location association stage. The signals of two sources exhibiting different incidence times result in two non-simultaneous localisation stimuli. The created (separate) localisation stimuli arrive at the gestalt association stage one after the other. In the superimposed sound field, the location association stage acts as a filter enabling the

discrimination of the source signals. Then, the pure source signals can be evaluated in the following gestalt association stage. The signals are interpreted due to their time difference similar to natural hearing. Hence, both the directional imaging for $\Delta t < 1$ ms and the inhibition for $\Delta t > 2$ ms correspond to natural hearing. The precedence effect can be interpreted as a precedence of the first localisation stimulus (Theile, 1980: *"law of the first localisation stimulus"*).

3) Cocktail party effect

The 'cocktail party effect' implies that in binaural hearing a target signal arriving from a certain direction will not be masked by an interfering signal arriving from a different direction to the same degree as in monaural listening. This phenomenon can be explained by the effect of the location association stage. Two different sound sources emanating different signals give rise to two different location associations as well as two different gestalt associations. The two resultant auditory events therefore occur after a two-stage selection from which the largest possible resolution derives. When listening monaurally, the selection effect of the location association stage is reduced significantly. The conjoint effect of the two processing stages, which are determined by the elementary areas of auditory experience, can be particularly well illustrated by the cocktail party effect.

4) Lateralisation

Lateralisation (= the mapping of auditory event and signal in headphone listening) should be regarded as different from localisation in terms of the perception process and the resulting perceptual properties. The localisation of a sound source requires source distance, which equals zero in the case of headphone listening, except when sufficiently complete binaural signals (e.g. dummy head recordings) are reproduced. Therefore, in lateralisation, the created phantom source is inside the head. Furthermore, no substitute sound source can be created, as a natural source with zero distance does not exist. Hence, lateralisation experiments can only provide information about the function of the gestalt association stage. The localisation stage discriminates the two source (ear) signals separately and forwards them to the gestalt association stage.

As a basic principle, lateralisation experiments do not allow any conclusions to be drawn on the functioning of the auditory system when localising a single sound source (because in this case only one localisation stimulus exists). Rather, they illustrate psychoacoustic properties of a 'phantom source inside the head' (loudspeaker distance = 0). In general, the evaluation of different ear signals deriving from one sound source cannot be investigated by means of two

sound sources that are at too small a distance from the ears. Headphone-based listening tests are listening tests with two sound sources, except when dummy head (=binaural) signals are presented (in which case a substitute sound source does exist).

5) Effect of narrowband signals

The functioning of the auditory system with respect to the localisation of a sound source can only be investigated under 'localisation conditions'. One prerequisite is that the sound event exhibits a broadband frequency spectrum. The perceptual process leading to the localisation can only take place if the spectral characteristics allow a mapping of the auditory event distance. When the bandwidth is reduced below a certain limit, a localisation does not exist and the ear signals are interpreted as two localisation stimuli similar to lateralisation.

3.6.3 Binaural decolouration

Other than Theile and his association model, there are other investigations which also address the phenomenon described above. The expression 'binaural decolouration' is often used in this context. It is defined as the 'suppression or reduction of colouration through binaural mechanisms' (after Brüggen 2001a, 2001b; Salomons, 1995). The consequences of this approach are similar to Theile's spatial decoding process and associated HRTF filtering. However, in the case of two coherent sound sources, Theile suggests the segregation of the different streams as a precondition for the decolouration. This is not necessarily assumed by others. It can also be considered a binaural advantage when other aspects of perception improve compared to monaural hearing. Brüggen presumes that one internal spectrum is responsible for timbre perception and that this spectrum is built by the mean of the two ear signals. The mean apparently has the property to smoothen spectral differences because the peaks and notches are different for the two ears in most cases. The idea of an internal spectrum or 'central spectrum' is utilised by Bilsen (1977), Zurek (1979), Kates (1985), Raatgever and Bilsen (1986) and others. These investigations deal with the phenomena binaural echo suppression and repetition pitch rather than stereophonic perception, but it is interesting to investigate their approach in this context. This is performed in the experiment described in chapter 8.

Binaural decolouration is understood as the sound colour improvement when listening with two ears as compared to with only one, see Figure 3-32. The rationale is that two ears can better resolve the incidence angle of reflections. In other words, decolouration means that the negative influence of reflections is suppressed by a certain degree. A sound field consisting of discrete sound signals from different directions as in stereophony could be considered similar and could also gain from decolouration.

In contrast, the colouration caused by spatial aliasing in WFS is not produced by discrete signals from discrete directions, but rather from many signals merging to a dense signal. Hence, decolouration does not apply in the same manner to WFS. Chapter 8 shows experimental results on this topic.

Decolouration therefore is also defined and used in this thesis as the ‘suppression or reduction of colouration through a successful segregation of distinct signals’ as opposed to the colouration caused by a sound field consisting of signals that cannot be segregated

.

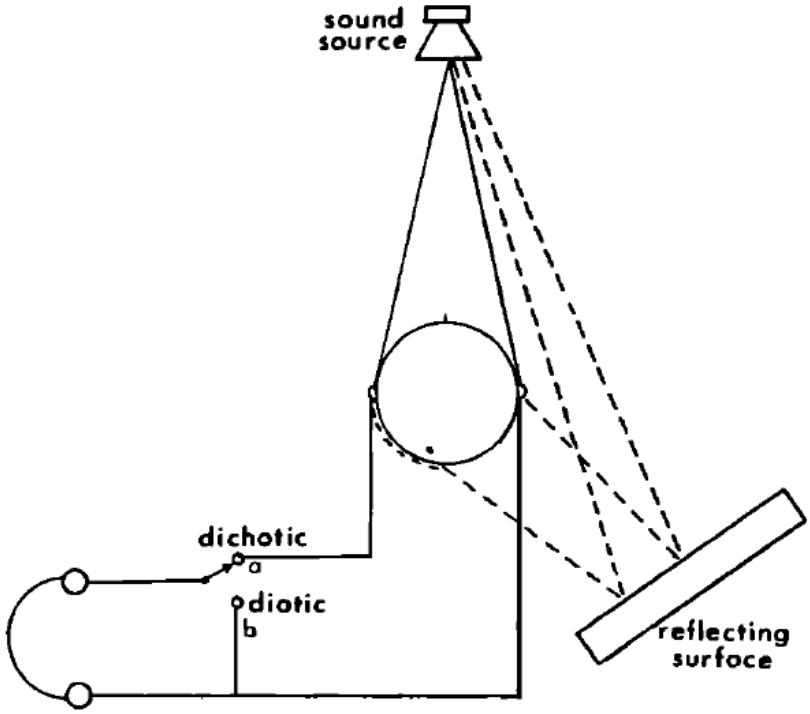


Figure 3-32: from Zurek (1979): Schematic illustration of an experimental setup for testing the binaural advantage for echo suppression.

3.7 Summary of chapter 3

Stereo is apparently more than just a two channel WFS. The unique properties of stereophonic reproduction have been recognised from the very beginning of stereophonic history. They qualify stereo for spatial sound reproduction of a high spatial and timbral fidelity. Hence, stereo can be considered a spatial data reduction system of high efficiency, because a small number of channels contains complex information on spatial properties in a rather high resolution.

The specific properties of stereophonic reproduction regarding the directional imaging, the sound colour reproduction and the distance perception have been discussed.

It was attempted to find a link between the observed properties and the underlying perception mechanism applied by the auditory system. Different approaches for perception theories have been presented and their potential to explain the described phenomena has been discussed. All perception theories still leave open questions and discrepancies in their interpretation of the perceptual properties of phantom sources.

Two main directions in explaining stereophonic perception exist. The first theory assumes a summing of the loudspeaker signals in a way that the summed sound field equals the sound field of real sources with regard to the decisive localisation parameters. This theory is called summing localisation. The other theory implies the ability of the auditory system to segregate between the loudspeaker signals and thus to be able to process the different streams independently. In the case of coherent loudspeaker signals, a fusion takes place that results in the perception of only one auditory event. This theory is called the association model by Theile (1980).

4. Wavefield synthesis and its properties

4.1 Introduction

This chapter introduces the basic principles and properties of sound fields created by wavefield synthesis (WFS). WFS is a sound reproduction technique with great potential as well as inherent reproduction errors. The discussion highlights both by reviewing the existing literature.

The properties are discussed from different perspectives. After this introduction, the physical background and physical properties (including artefacts) are described (section 4.2). The properties with regard to perception are analysed in section 4.3. A focus is put on the attributes investigated in this thesis, namely the attributes of localisation, sound colour and distance perception. The chapter is summarised in section 4.4.

4.2 Basic principles and theoretical background of the wavefield synthesis concept

4.2.1 Theoretical origin

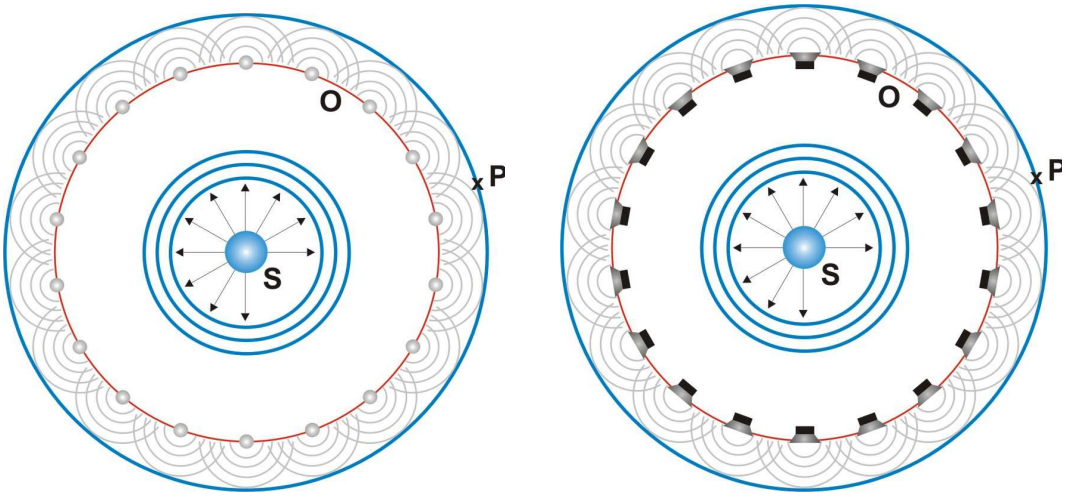
The Huygens[9] principle states that each point of a wave front can be considered the starting point of a new elementary wave. After the Huygens-Fresnel[1] principle, the wave front as a whole can be considered the sum of all the elementary waves that are created on a surface surrounding the source.

An illustration of this principle is presented in Figure 4-1a. The blue wave front arises from the source S. The wave front can also be considered the sum of all elementary waves on the surface O. Through knowledge of the wave front on the surface O, the wave field at every point P can be calculated.

---

[9] *From Born and Wolf (1975): "According to Huygens' construction, every point of a wavefront may be considered as a centre of a secondary disturbance which gives rise to spherical wavelets, and the wave front at any later instant may be regarded as the envelope of these wavelets. Fresnel was able to account for diffraction by supplementing Huygens' construction with the postulate that the secondary wavelets mutually interfere. This combination of Huygens' construction with the principle of interference is called Huygens-Fresnel Principle."*

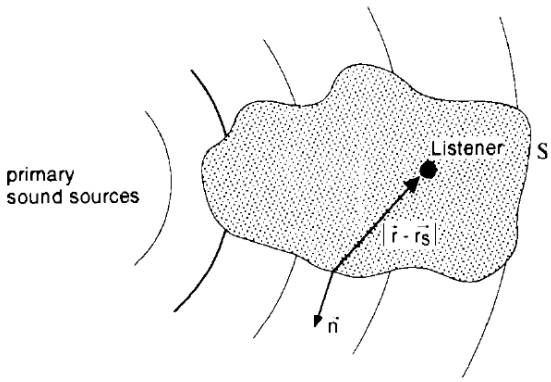
The idea of wavefield synthesis (WFS) is to replace the elementary waves on the surface O by secondary sources, i.e. by single loudspeakers. The wave field in the point P can in principle be synthesised by a superposition of all loudspeaker signals (Figure 4-1b).



a) Huygens-Fresnel principle **b) Wavefield synthesis**

Figure 4-1: from Theile et al. (2003): Illustration of the theoretical origin of wavefield synthesis

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



with

r = point inside S,

ω = angular frequency,

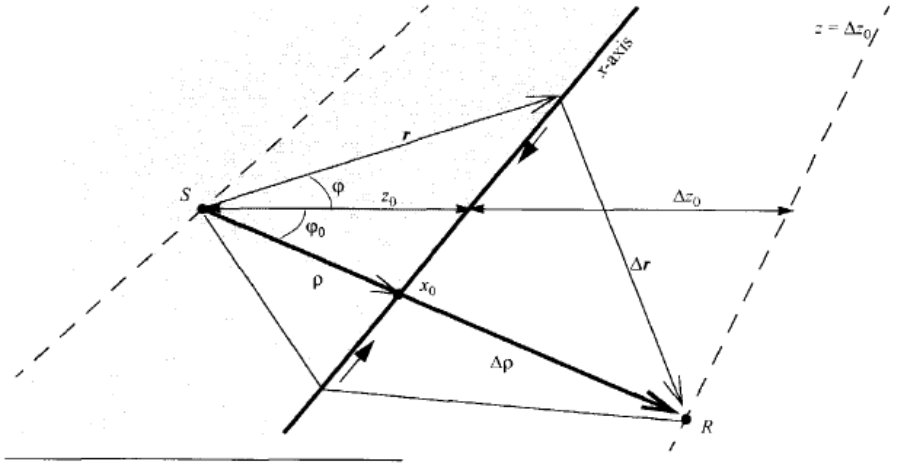
k = angular wavenumber: $k = 2\pi/\lambda$**,**

P(r,ω) = Fourier transformed pressure distribution on S due to primary sources outside S.

Figure 4-2: from Berkhout et al. (1993): Kirchhoff-Helmholtz integral and corresponding geometry. The 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 its gradient (which is proportional to the normal component of the particle velocity) are known on the surface S enclosing V (Berkhout et al., 1993).

The Huygens-Fresnel principle was quantified by Kirchhoff by the so-called Kirchhoff-Helmholtz-Integral (Start, 1997). Figure 4-2 shows the formula and a graphical illustration. The 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 its gradient are known on the surface S enclosing V.

If the surface S degenerates to a plane, separating the listening area from the primary source area, the so-called Rayleigh integrals (Start, 1997) 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 2½D integrals (Start, 1997) 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 2½D synthesis operator' (see Figure 4-3). The synthesis operator can be expressed for sources behind and in front of the array and for arbitrary source directivities. It can be adapted to the actual directivity characteristics of the array loudspeakers (de Vries, 1995).



$$Q_m(r,\varpi) = S(\varpi)\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{\Delta z_0}{z_0+\Delta z_0}}\cos\varphi\frac{\exp(-jkr)}{\sqrt{r}};$$

Figure 4-3: from (Verheijen, 1998): '2½D Synthesis operator' or 'driving function of the array loudspeakers' for a monopole virtual source reproduced by a linear WFS array in the horizontal plane consisting of monopoles. The source S is at $z = z_0$, i.e. behind the array which is on the x-axis. S(ω) is the source signal. The reference line is at $z = \Delta z_0$. r is the vector from virtual source to array loudspeaker, ω is the angular frequency, φ is the angle of incidence of the vector r at the array, k is the angular wavenumber (k = $2\pi/\lambda$).

The synthesis operator is similar to the mathematical formulation of the 'acoustic curtain' (see also chapter 3.3). An acoustic curtain is built by recording at n microphone locations on a line z = 0 and reproducing the recorded signals on the same locations by loudspeakers. Deviating from this simple definition of the acoustic curtain, the described synthesis operator applies a $1/\sqrt{r}$-relationship for the level decrease with distance on the recording side. Furthermore, the sampling includes a $\cos\varphi$-directivity which would equal the use of bidirectional microphones in the case of the acoustic curtain. Last, a position-independent equalisation by the so-called '$\sqrt{jk}$-filter', which equals a 3dB per octave boost is applied. Figure 4-4 illustrates this simple basis of WFS.

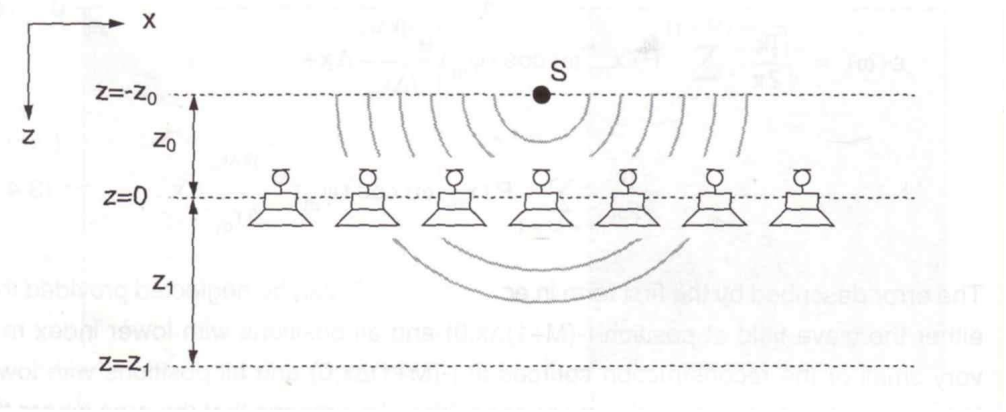


Figure 4-4: from Verheijen (1998): Basic principle of WFS. Sampling and reproduction of the wave field using an 'acoustic curtain': the sound field is sampled at n (microphone) positions and reproduced on the same n positions. Equalising is necessary between sampling and reproduction to ensure a transparent acoustic functioning of the 'acoustic curtain'. The figure shows the setup in the horizontal (x-z) plane. S is the (virtual) source. z_0 and z_1 are the distances of source and receiver line to the array.

This knowledge was derived and first described by authors from the TU Delft in the Netherlands, for instance Berkhout (1987, 1988), Berkhout et al. (1992, 1993), Boone et al. (1995), further in doctoral theses by Vogel (1993), Start (1997), Verheijen (1998), Sonke (2000), Hulsebos (2004) and de Bruijn (2004). In addition to their different scientific approaches, these books in particular give an excellent introduction to and overview of the basic theories of WFS. Early WFS research was also undertaken in Japan by Komiyama et al. (1991) and Ono and Komiyama (1997).

4.2.2 Physical potential of WFS

Figure 4-5 illustrates the basic characteristics of WFS: for the entire listening area, the reproduced 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 also involves a realistic change of the sound pressure level when the distance to the virtual source (=notional source created by the WFS array) is varied. This may be called 'motion parallax', similar to visual perception. The role of motion parallax for acoustic perception was discussed in chapter 2.5.

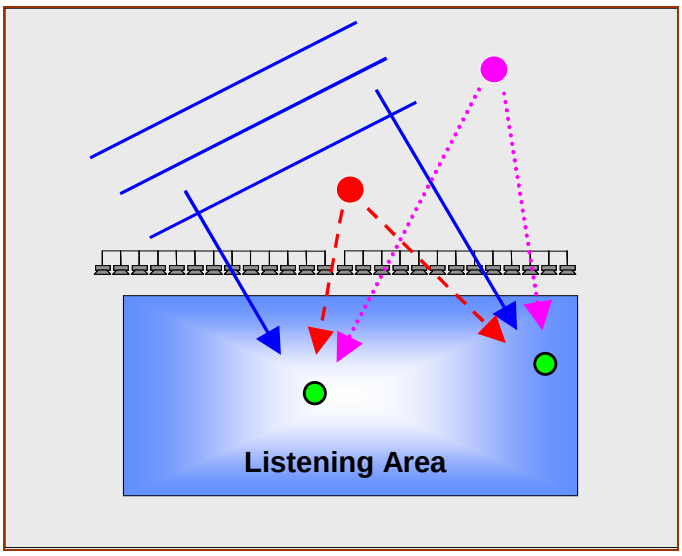


Figure 4-5: from Theile et al. (2003): Illustration of the basic WFS potential. The acoustic perspective of the sound field changes corresponding to the listener position (green) as is the case in a natural sound field. Stable source directions can be achieved with plane waves (blue, solid) and stable source locations with point sources (red and pink, dashed and dotted).

In WFS, when only plane waves are created, it is possible to produce the same acoustical image for many listeners in a large area. Plane waves are perceived from the same relative direction (see blue arrows in Figure 4-5) and, in principle, with the same loudness. Thus, the listening area of WFS is potentially very large. Furthermore, with the incorporation of point sources, it is possible to create an acoustical scene which creates different acoustical images for every listener corresponding to the listener perspective. This acoustical scene can be of an arbitrary size for many listeners in the listening area.

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

acoustic scene can be more plausible. However, in combination with a two-dimensional picture (as in cinema), WFS looses many of its advantages when compared to conventional stereo. The two-dimensional picture contains stable source positions on the canvas and thus, neither the preferences of plane waves nor the creation of a three-dimensional acoustical scene can be utilised. A three-dimensional acoustical scene in combination with a two-dimensional picture gives rise to a localisation mismatch which might be annoying (de Bruijn and Boone, 2003).

The theoretical capabilities of WFS to create a quasi-realistic sound field or to recreate an existing sound field also include source directivity. In theory, it is possible to simulate an arbitrary directivity of the virtual source limited only by the spatial aliasing frequency and the length of the WFS array. Furthermore, the synthesis is only possible in the reproduction plane of the array. This is described by Corteel (2007a) and Jacques et al. (2005).

The location of the 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, see 9.3.1) could make a significant difference to conventional sound reproduction techniques. However, a stable and convincing reproduction of focussed sources is only possible with constraints, see chapters 4.2.8 and 9.

4.2.3 Physical constraints of WFS

For practical reasons, WFS fails to be a perfect realisation of its theoretical basis. In theory, the Kirchhoff-Helmholtz integral can be used to calculate a sound field which is congruent to the desired one. In practice, however, nobody can install arrays with infinitely small transducer spacing, nobody can install an infinitely large loudspeaker array and also a two-dimensional array seems implausible for a realistic application. Furthermore, the array is always positioned in a real room with its individual acoustics which add to the virtual acoustics. In short, compromises have to be made, which, some more than others, decrease the degree of congruence between desired and reproduced sound field. The compromises listed in Table 4-1 can be interpreted as the differences between a practical WFS system and a theoretical, ideal WFS system which is directly deduced from the Kirchhoff-Helmholtz integral.

Compromise	*Rationale of compromise*
Finiteness of loudspeaker spacing	Loudspeaker size, costs
One-dimensionality of WFS array (=array has no vertical dispersion)	Costs, Effort
Finiteness of array length	Costs, Effort
Reproduction room reflections	Reproduction room

Table 4-1: Summary of compromises in WFS

Physical artefact	*Due to which physical compromise*
Spatial aliasing, see section 4.2.5	Loudspeaker spacing
Diffraction effects (truncation effects), see section 4.2.6	Array length
One-dimensionality (3D → 2D error), wave front curvature mismatch, see section 4.2.7	Array size
Reproduction room reflections, see section 4.2.8	Reproduction room reflections

Table 4-2: Physical artefacts of WFS

Perceptual artefact, listed by impaired perceptual attribute, see section 4.3	*Due to which physical artefact(s)*
Image focus	Spatial aliasing, diffraction effects
Locatedness	Spatial aliasing, diffraction effects
Sound colour	Spatial aliasing, diffraction effects
Distance perception and other attributes of spatial perception (perception of depth, envelopment, etc.)	Reproduction room reflections, one-dimensionality, (wave front curvature mismatch)

Table 4-3: Perceptual artefacts of WFS

The setup of a WFS system in practice is a trade-off of these compromises and the available room, budget and other circumstances, such as the available audio infrastructure. The more one knows about the actual consequences of changing a system parameter to a certain degree, the better and faster an optimal system can be designed. This means this trade-off requires a fundamental knowledge of the consequences of the compromises, i.e. the resulting artefacts. Moreover, the costs and efforts for building a WFS system are crucial for its feasibility.

4.2.4 The artefacts of WFS

Both the physical and perceptual artefacts of WFS will be listed below. In this thesis it is shown that WFS can not be interpreted as a logical extension of other technologies with regard to any perceptual parameter. WFS should rather be understood as a different type of technology, with both physically and psycho-acoustically different properties.

The relationship between the physical setup of a WFS system and the resulting artefacts of the reproduced sound field is described in Table 4-2. The physical artefacts can be interpreted as the physical differences between the reproduced sound field and the ideal or original sound field.

The perceptual artefacts of WFS can be interpreted as the differences between the reproduced sound field and the ideal or original sound field with regard to perception. The relationships in Table 4-3 are partly hypothesised from subjective experience as objective investigations do not yet exist. A thorough discussion of WFS can be undertaken by evaluating the quality of the reproduced sound field on the basis of these perceptual attributes.

4.2.5 Spatial aliasing

The finitely small transducer spacing in WFS causes spatial aliasing. The synthesis of the sound field only works perfectly for frequencies below a certain limit frequency, which is dictated by the transducer spacing. This frequency is called the spatial aliasing frequency f_{alias}. Above f_{alias} the sound field is reconstructed erroneously, the aliased wave field contains both spatial and spectral artefacts. The perceptual artefacts arising from these artefacts are listed in Table 4-3. According to subjective experience and physical analysis, spatial aliasing is suspected to cause a degradation of both the localisation performance and the sound colour perception. An illustration of aliased frequency responses is given in Figure 4-6, Figure 4-9 and Figure 4-12.

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 source incidence angle on the microphone side, as described by Sonke (2000). This interrelationship is illustrated in Figure 4-7. Accordingly, the maximum wavelength being *received* correctly without spatial aliasing occurring depends on the maximal reproduction angle on the receiver side.

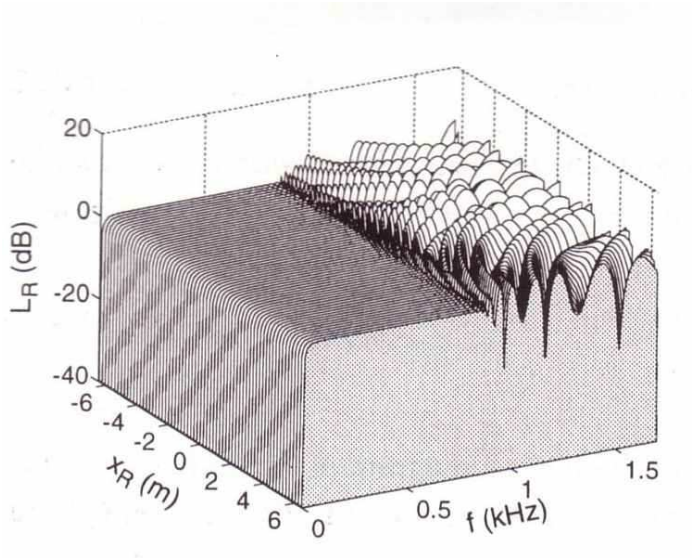


Figure 4-6: from Start (1997): Wave field with spatial aliasing starting at approx. 1 kHz. The x-axis represents a line of receiver positions in the listening area and parallel to the array.

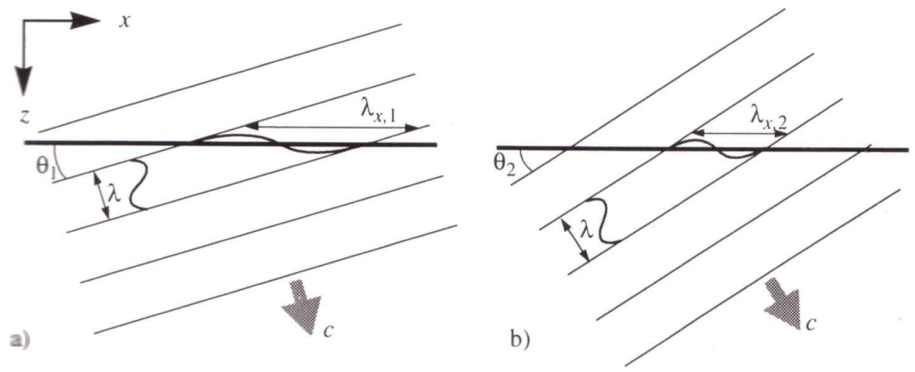


Figure 4-7: from Start (1997): Illustration of interrelationship between sampling (microphone /loudspeaker) distance and maximal wave length. $\lambda_{x,1}$ and $\lambda_{x,2}$ are the relevant components of the wavelength λ in the array-direction x.

a) small incidence angle θ_1: $\lambda_{x,1}$ is relatively large, would potentially be sampled correctly,

b) large incidence angle θ_2: $\lambda_{x,2}$ is relatively small, would potentially be sampled incorrectly.

Figure 4-8 shows how the relevant f_{alias} is determined 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 f_{alias} for the sampled sound field, in which θ^{sec} equals 90° (e.g. Sonke, 2000). This is one reason for differing declarations of f_{alias} in the literature.

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

where

***c* = sound propagation velocity,**

θ^{v} = maximum angle on the sampling side,

θ^{sec} = maximum angle on the reproduction side.

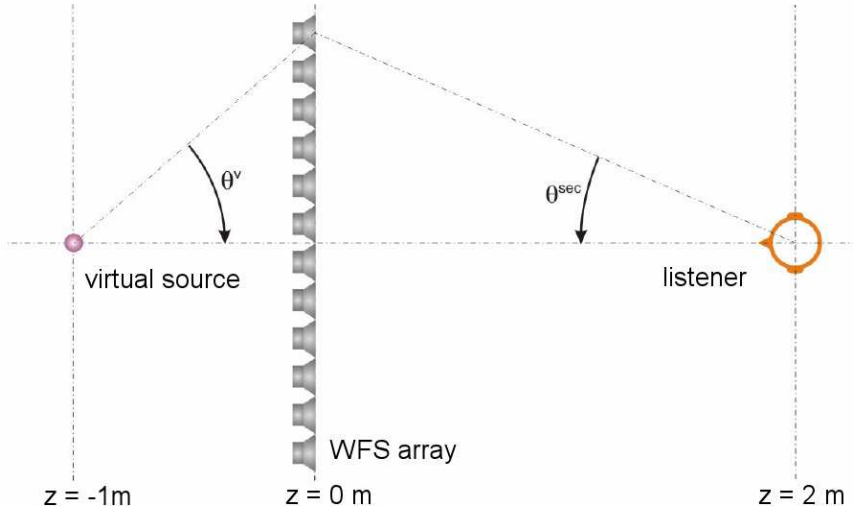


Figure 4-8: Illustration from Huber (2002): Calculation of the spatial aliasing frequency f_{alias} for sources behind the array.

Description of aliasing

Spatial aliasing results in a distortion of the reproduction of frequencies above the spatial aliasing frequency in terms of spatial and timbral fidelity. Not only is the correct direction synthesised but also erroneous (aliased) contributions, and the sound field at the listening position consists of a superposition of different contributions. Hence, the sound field at the listening position suffers from interferences; these depend on frequency and location.

The physical consequences of spatial aliasing can be considered as a distortion of the sound field at the listening position with regard to the spatial distribution of the sound contributions, the frequency response and the response in time domain as well as a distortion that changes with listener and source movements.

The following figures provide illustrations of spatial aliasing:

Figure 4-9 shows spatial aliasing at different positions in the listening area. The radical change of the aliasing with source/receiver movements can be seen. Note also that the peak/notch distances in the aliasing increase with frequency. This is different compared to comb-filtering which has constant peak/notch distances on a linear scale. The figure shows sound field simulations with ideal omni-directional array loudspeakers under ideal, anechoic conditions.

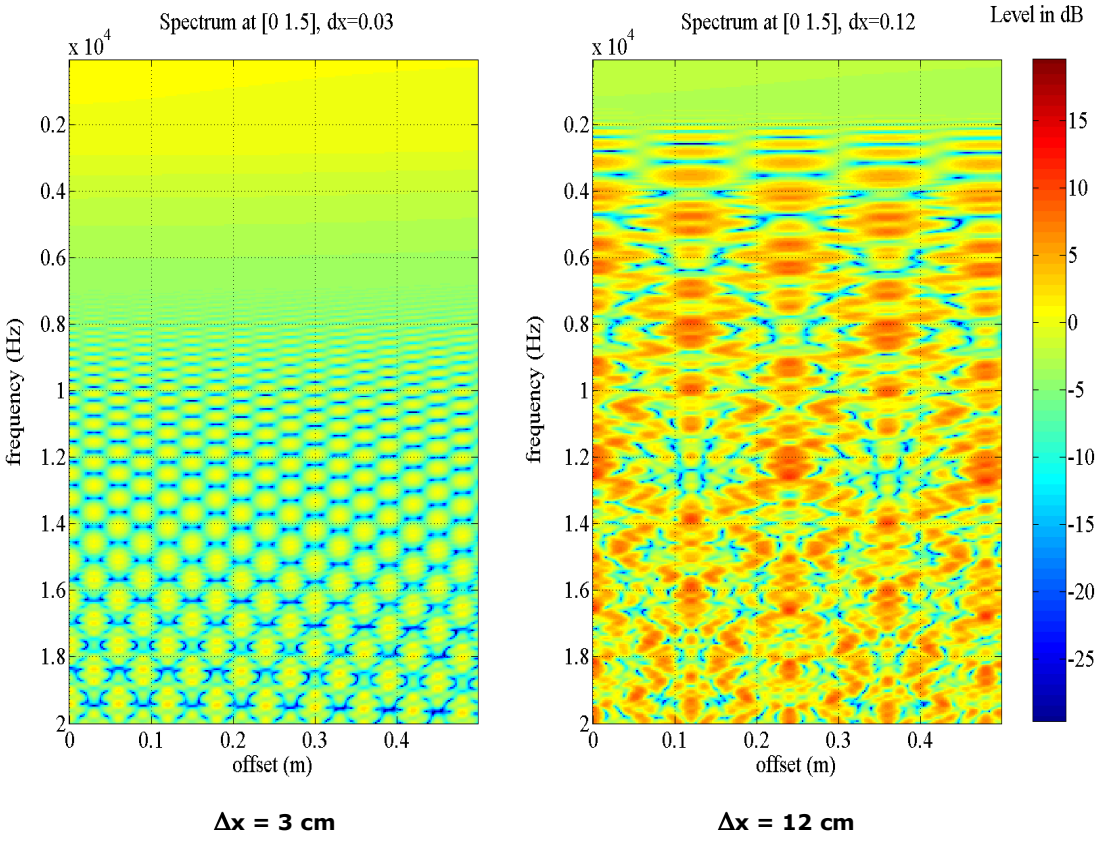


Figure 4-9: Frequency responses (note the frequency axis has a linear scale and proceeds from top to bottom) measured on a line of receiver positions (x=0 ... 0.5 m). The array speaker spacing is Δx = 3 cm in the left figure and Δx = 12 cm in the right figure. The graphs show the rapid change of the spatial aliasing with a listener movement of a few cm.

Figure 4-10 and Figure 4-11 show the reproduction of a sine wave of different frequencies reproduced by the same array. The sound pressure distribution in the horizontal plane is shown and the sound is emanating from a linear array indicated by the small circles. The sources in these figures are focussed sources (75 cm in front of the array). The 1000 Hz sine wave (Figure 4-10) is reproduced correctly in the entire listening area. The 4300 Hz signal (Figure 4-11) produces a sound field that is aliased in almost the entire listening area. This can be seen from the diverse and erroneous phase and directional information in the reproduced wave field. As the spatial aliasing frequency increases with decreasing source-receiver distance, the area close to the source is reproduced correctly in both cases.

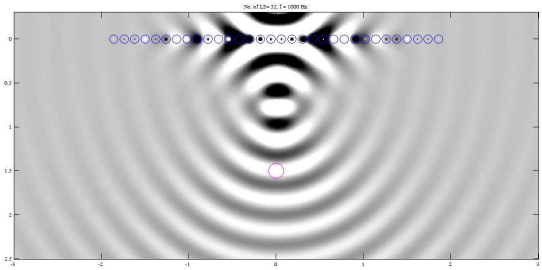


Figure 4-10: Snapshot of the pressure field in the horizontal plane of WFS array of 32 loudspeakers (small circles, Δx=12 cm), focussed source 75 cm in front of the array, sine wave of f=1000 Hz (<f_{alias})

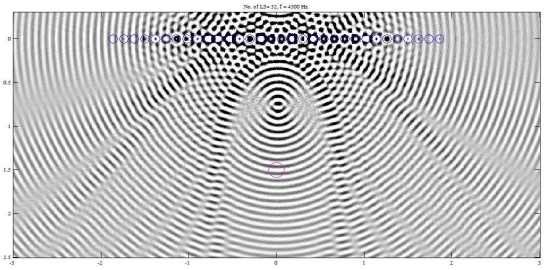
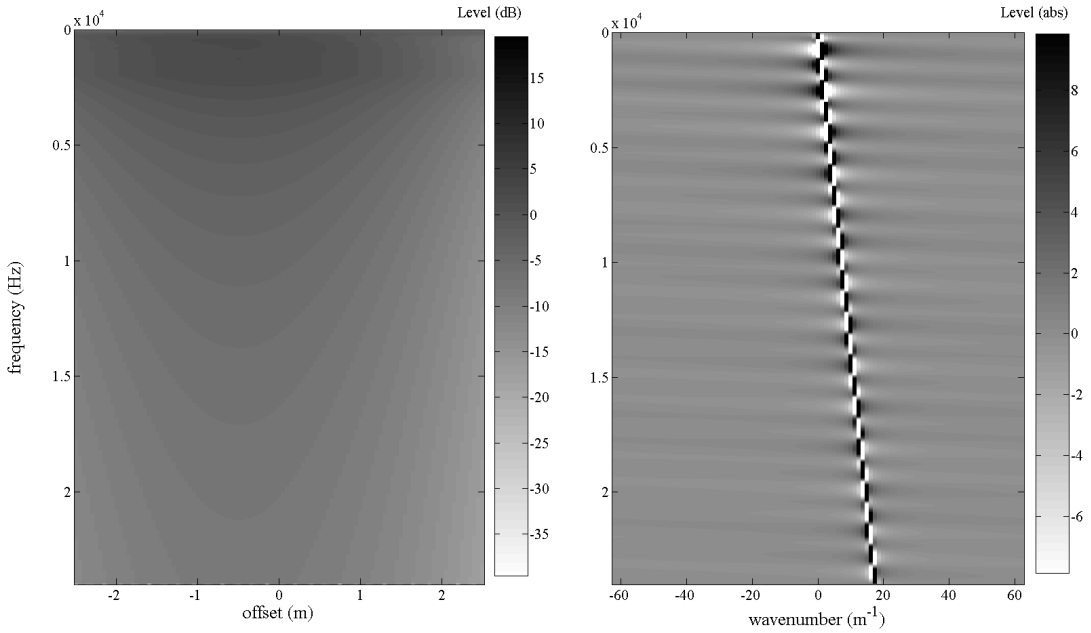


Figure 4-11: Snapshot of the pressure field in the horizontal plane of WFS array of 32 loudspeakers (small circles, Δx=12 cm, focussed source 75 cm in front of the array, sine wave of f=4300 Hz (>f_{alias} except close to the source)

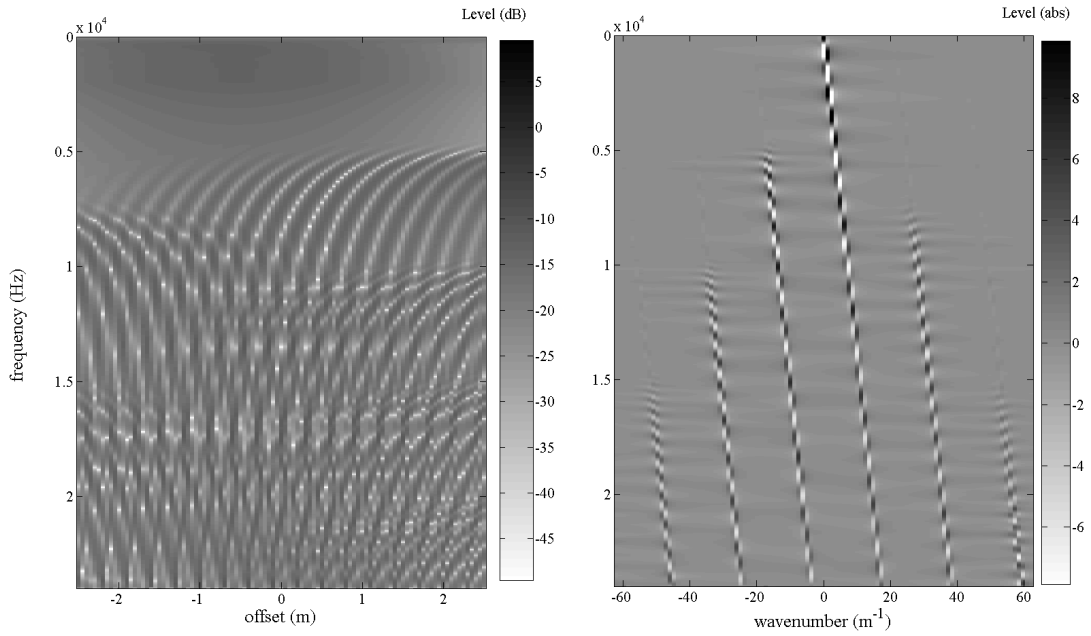
The superposition of correct and incorrect contributions in spatial aliasing can be best observed by a representation of the WFS sound field in the so-called spatial Fourier domain. In this domain each reproduced plane wave is represented by a straight line starting from point (0; 0). The angle of the line relative to the vertical axis depends on the direction of the plane wave. In parallel to aliasing in A/D conversion, aliased contributions are mirrored in the audible area. Figure 4-12 shows an illustration of the spatial Fourier domain representation of a plane wave signal sampled by a linear transducer array. Only the central line shown in sub-Figure 4-12-b1 and Figure 4-12-b2 is the correct contribution. With increasing frequency, additional mirrored contributions give rise to increased spatial aliasing in Figure 4-12-b2. Note that the aliased contributions can not be interpreted simply as plane waves from the wrong direction.

***Without* spatial aliasing:**



a1 **b1**

***With* spatial aliasing:**



a2 **b2**

Figure 4-12: Illustration of spatial aliasing in the spatial Fourier domain. Sub-figures a1 and a2 show the frequency responses on a line of receiver positions of a plane wave synthesised by a linear WFS array. Sub-figures b1 and b2 show the same signal in the spatial Fourier domain. Sub-figures a1 and b1 show a plane wave signal without aliasing, Sub-figures a2 and b2 show the same plane wave containing spatial aliasing. Figures were made using MATLAB scripts from Edo Hulsebos (see Hulsebos, 2004).

Different methods were proposed in the literature to avoid or minimise spatial aliasing. These either aim to reduce the physical deviation of the aliased sound field, or to reduce the perceptibility of aliasing.

De Vries et al. (1994) and Start (1997) suggest to minimise the maximum angle on the sampling side for higher frequencies (which increases f_{alias}, see Figure 4-8). This method can be considered a spatial bandwidth reduction. Another technique described by these authors works similarly: in order to minimise the maximum angle on the receiver side (which increases 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. This would lead to a loss of (spatial) information.

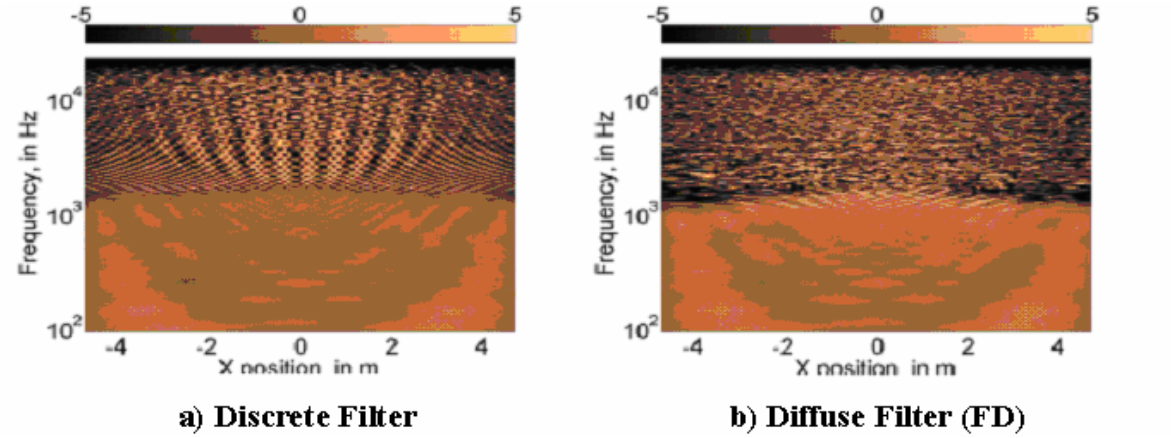


Figure 4-13: from Corteel et al. (2007b): Effect of 'diffusion' of the filters above the aliasing frequency on the frequency spectrum. The two diagrams show the spectrum in a WFS sound field on a line of receiver positions in the listening area. The left diagram (a) shows the spectra created without diffusion, the right diagram (b) shows the spectra created after the diffusion.

Start (1997) tried to avoid the audible periodicity of the aliasing artefacts, and by these means to reduce the quantity as well as the perceptibility of spatial aliasing. He realised this by the method of 'time domain randomisation'. This means that the driving function of each array channel is intentionally delayed by an adequately small random time gap which leads to a diffusion of the array response above the aliasing frequency and thus eliminates extreme dips and peaks in its frequency spectrum. However, at the same time, the correct contributions in the aliased frequency region are suppressed. In this way, the sound field loses the directional information above a certain frequency. This method has both perceptual advantages and disadvantages, as established in experiments by Start (see section 4.3.3). Corteel et al. (2007b) apply a similar method of diffusing the high-frequency responses. As shown in Figure 4-13, the strong deviations do indeed vanish as the result of the diffusion. Corteel et al. give evi-

dence for the positive effect of this diffusion on objective coloration measures (see also chapter 8.6). They also show negative effects of the diffusion (on the variation of the ITD between the frequency bands above the aliasing frequency) which, after their interpretation, could give rise to differences in the perception of source width and distance.

A method to minimise the perceptibility of spatial aliasing is proposed in this thesis. This method will be described in chapter 5. It is called OPSI ('Optimised Phantom Source Imaging in wavefield synthesis') and is a proposal of a hybrid WFS and phantom source reproduction. Spatial aliasing is avoided through the omission of the WFS reproduction of the high-frequency content. Instead, it is proposed to reproduce the high-frequency content with conventional phantom sources that are created by a few loudspeakers within the array. WFS is applied only below $f_{alias,}$ 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 and simulations.

4.2.6 Diffraction 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 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 4-14, and – depending on their level and time-offset at the receiver's location – may give rise to colouration.

A reduction of these diffraction effects (also known as 'truncation 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 magnitude of diffraction effects is substantially reduced, however, this is 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 the diffraction contributions on a fixed reference position, these "*can be interpreted as scaled point sources with a specific directivity pattern radiating*" from the edges of the array. Hence, the compensation (cancellation) of these error signals is possible, albeit with

full cancellation occurring only at the reference position. One important drawback is the accompanying introduction of even stronger colouration outside the listening area.

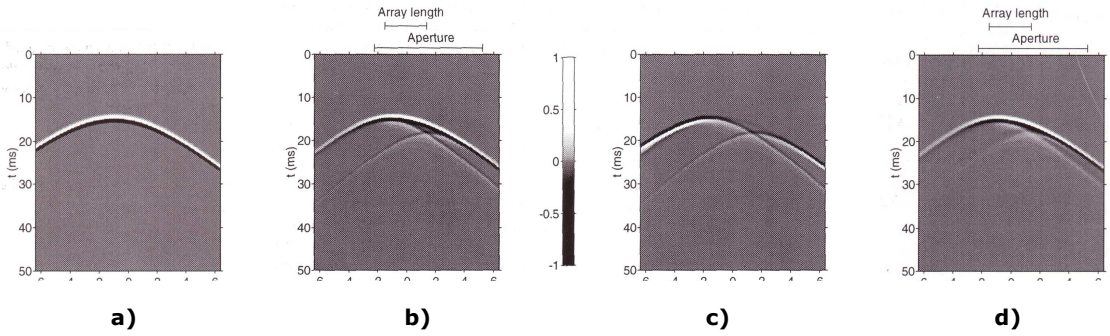


Figure 4-14: from Start (1997): Influence of array truncation. Diffraction effects can be observed. The diagrams show the signal (y-axis is time) in the listening area on a line of receiver positions parallel to the array (x-axis is the offset).

a) Response of an infinite array

b) Response of a truncated array

c) Difference between a) and b)

d) Response of a truncated array after tapering

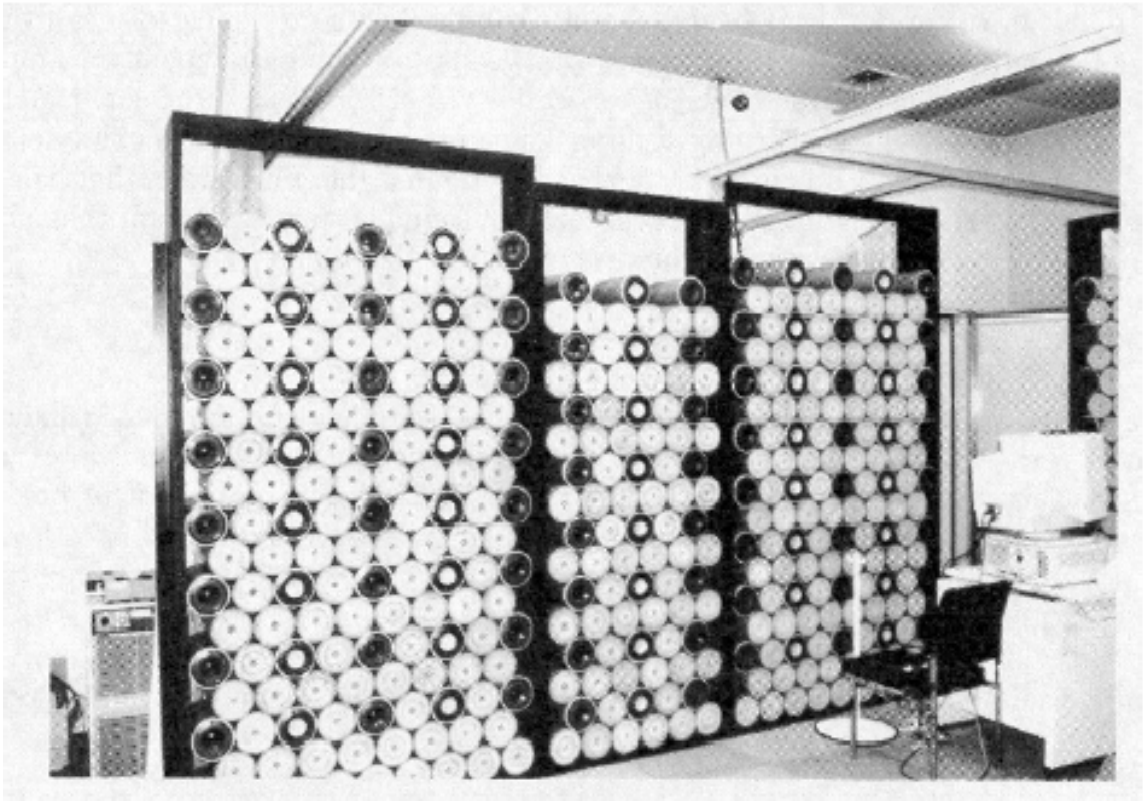


Figure 4-15: 'Loudspeaker wall', used for experiments by Ono and Komiyama (1997)

4.2.7 Reduction of the reproduction dimensions: 3D → 2D

Theory does not restrict WFS to the horizontal plane. Komiyama et al. (1991) and Ono and Komiyama (1997, see Figure 4-15) actually built a two-dimensional loudspeaker array ('loudspeaker wall').

In practice, there are too few convincing arguments for a WFS array to be installed in two dimensions. However, this reduction of the array dimension to a single line and the synthesis dimensions to the horizontal plane does have major consequences.

As a logical consequence, only virtual sources in the horizontal plane can be synthesised. Conventional WFS is capable of creating a correct directional localisation only when sources and receivers are located in the horizontal plane. This is not critical in terms of the directional imaging of the direct sound of sources - which normally are located in the horizontal plane. However, spatial attributes such as listener envelopment and spatial impression could be significantly enhanced by the incorporation of loudspeakers outside the horizontal plane. This is postulated by supporters of new surround sound formats incorporating elevated speakers such as '5.1 with height' or '22.2' (Hamasaki et al., 2006). The reason concerns both the spatial distribution of the reproduced reflections, and the reproduction of the reverberation. These parameters can be enhanced by incorporation of elevated loudspeakers together with the inherent enlargement of the potential reproduction area. Both temporal and spatial reflection density can be reduced and thus reproduced with greater similarity to a real sound field. Furthermore, the temporal and spatial diffuseness can be increased with this addition of a further reproduction dimension.

Every recording made in a room and reproduced through a horizontal only array produces a distorted image of the virtual room. This is due to the fact that the microphones pick up the signals from all directions whereas the reproduction is performed in only one plane. For example, a ceiling reflection will be picked up by the microphones and reproduced from the same direction as the direct sound in the horizontal plane, potentially leading to comb filtering through interference. The reproduced reflection pattern is reproduced distorted in terms of the spatial distribution and the temporal and spatial density.

An easy solution to overcome the restriction of the reproduction dimensions to the horizontal plane is to incorporate single elevated loudspeakers. These may be a sufficient compromise instead of a planar array in order to avoid the mentioned artefacts. 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) do not need the same spatial resolution as required in the vertical and horizontal dimensions.

In addition to these artefacts in the reproduction of the reflection pattern, one must be aware of the fact that in conventional two-dimensional WFS, no real spherical waves are created. Instead, waves with 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 4-16a) the plane wave seems to be perfect, whereas in the vertical section (Figure 4-16b) the cylindrical waveform is represented as a circular waveform emitted from the array.

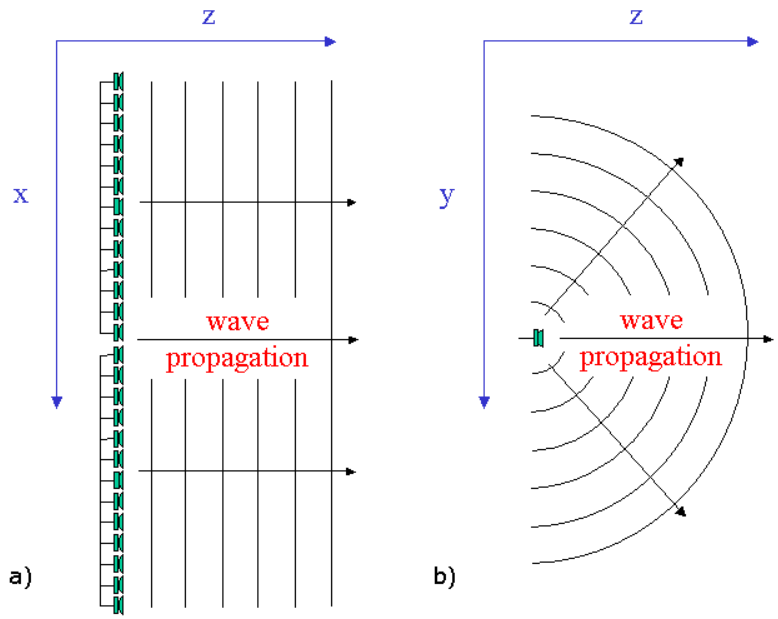


Figure 4-16: 2-dimensional WFS reproducing cylindrical waves: horizontal (a) and vertical (b) section of a linear WFS loudspeaker array reproducing a plane wave. The array is positioned in the horizontal (x-z) plane, the y-axis is the height dimension

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). As a consequence, the levels and the level balance between virtual sources in different distances are reproduced correctly only on one reference line in the listening area (see Figure 4-3).

These ‘amplitude errors’ or ‘spatial decay errors’ are described and quantified by Sonke et al. (1998) and Sonke (2000). Figure 4-17 illustrates the deviation of real and desired sound fields. Sonke describes 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, pp.117ff.).

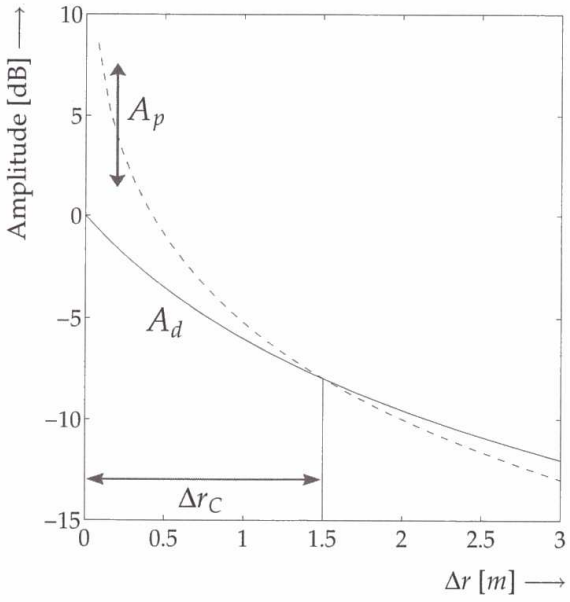


Figure 4-17: from Sonke (2000): Amplitude of a WFS monopole source A_p and a real, desired monopole source A_d along a line defined by the source position at (-1,0) and an array loudspeaker position at (0,0). The amplitudes match at the definable reference receiver line r_c=1.5.

4.2.8 Reproduction room errors

A common misunderstanding exists regarding the influence of the reproduction room in WFS. It is not true that a dry source reproduced by a WFS array automatically produces a natural reflection pattern at the listening position. The virtual sources are not reproduced on stable locations, but rather, depending on the listening position, somewhere on a circle around the array. The general rule is: the virtual source is always located in one line with the array and the listener. This also means that the excited room reflections do not arise from one stable location but from a number of different locations around the array. The consequences are illustrated in Figure 4-18 and Figure 4-19. A virtual focussed source (blue dot and peak) is reproduced by a linear WFS array in a room. Two reflections from back wall and floor are denoted by the corresponding virtual mirror sources. The locations of these virtual mirror sources are not - as in the ideal case (green dots and peaks) - built by mirroring the virtual source, but depend on the relative angle to the listener (orange dots and peaks). In Figure 4-18, the correct virtual mirror sources are denoted by the green circles and the actual virtual

mirror sources are denoted by the orange circles. In Figure 4-19, the correct reflections are denoted by the green peaks and the actual reflections by the orange peaks. Hence, both timing as well as the directions of the virtual mirror sources are clearly erroneous. In the depicted example of this figure, the reflections are produced too early and do not correspond to the actual source position. They do however correspond to the reflection pattern of the array itself, as a comparison of Figure 4-19b and Figure 4-19c shows. This most likely gives rise to the distance perception of a virtual source according to the array distance instead of the synthesised source distance (see section 4.3.5).

Figure 4-18 and Figure 4-19: Illustration of erroneous timing and directions of reproduction room reflections in WFS. A focussed source (blue) is reproduced by a WFS array (grey, solid loudspeaker) and two room reflections from back wall and floor are described by the corresponding mirror sources. The mirror sources and reflections are illustrated for the ideal (green) and the actual (orange) case. The left figure shows the locations of focussed source and mirror sources, the right figure shows the reflection pattern in the time domain.

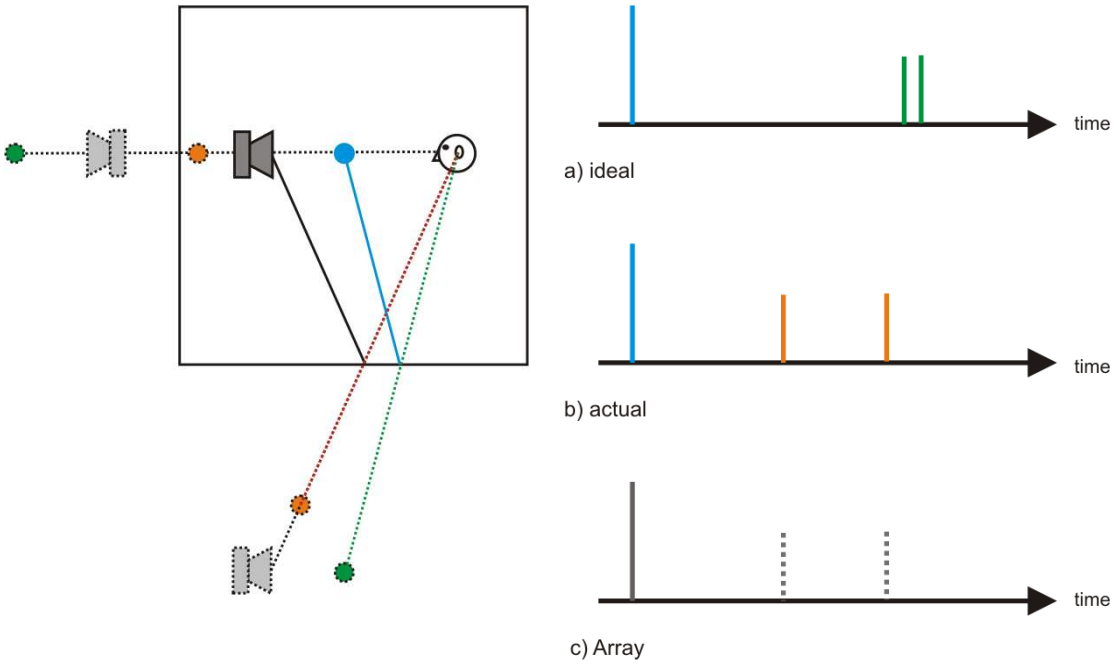


Figure 4-18: Locations of focussed source (blue), ideal mirror sources (green) and actual mirror sources (orange).
Note that the actual mirror source locations are too close compared to the ideal mirror sources.

Figure 4-19: Reflection pattern in the time domain according to the diagram in Figure 4-18:

a) the ideal, correct reflection pattern of a real source at the position of the focussed source

b) the actual reflection pattern of the focussed source

c) the reflection pattern of the array.

It seems to be even more critical when a virtual room is to be reproduced. Now, the reflections of the reproduction room are disturbing in any case. The reflection patterns of reproduction room and virtual room are superimposed and the incidental perception is built based on this superimposed reflection pattern. This situation also exists in stereo where the virtual room is to be imposed on the existing reproduction room. It is known for stereo that no smaller distance than the loudspeaker distance can be synthesised. This corresponds to the aforementioned depictions.

A further serious consequence can be hypothesised in theory for the reproduction of focussed sources. As shown in Figure 4-19, in the case of a focussed source, the correct early reflections arrive later than the erroneous reflections. This means, the incidence time of the early reflections is decreased and the time gap between direct sound and first reflection is smaller than for a natural source at the same distance. The time gap can be considered a potential cue for distance perception (see Pellegrini, 2001). Hence, distance perception would be distorted. In the case of non-focussed sources this problem does not exist. This is because the time gap is too large, which can easily be corrected by the creation of artificial reflections.

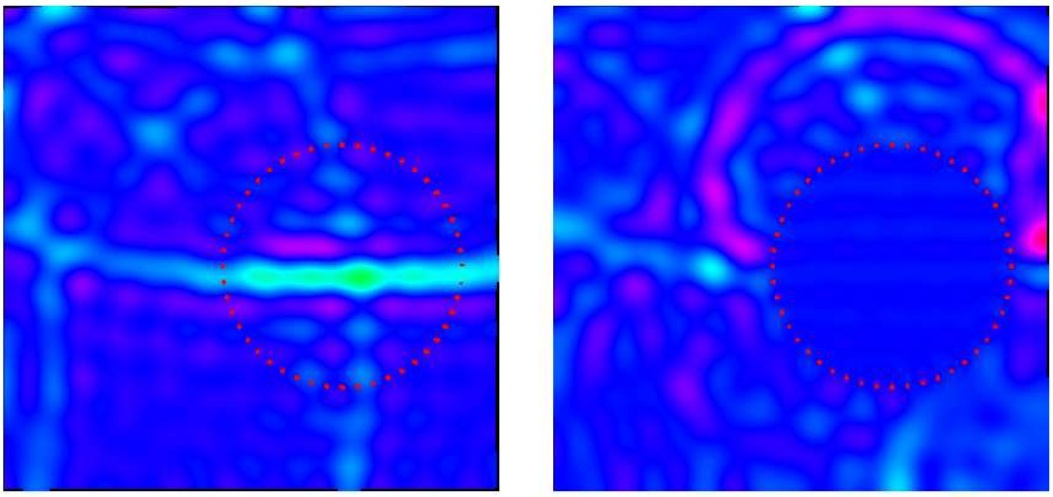


Figure 4-20: from Petrausch et al. (2006): Performance of listening room compensation with WFS: The left figure shows listening room reflections without, the right figure shows the same reflections with compensation performed by a circular WFS array. The figures contain a snapshot of the pressure field produced by a circular WFS array (indicated by red dots) in the horizontal plane. The circular array is located in a room and produces reflections. The snapshot is produced at the time at which a distinct side wall reflection is passing the array area. This side wall reflection can be cancelled well.

WFS primarily aims to recreate an acoustic scene with accurate spatial reproduction. The detrimental impact of the disturbing reproduction room reflections should ideally be suppressed or masked. WFS provides an enhanced possibility to cancel distinct reflections over

an enlarged listening area because of its inherent properties. The use of WFS as a tool to manipulate the reproduction room acoustics was investigated by Corteel and Nicol (2003) and Spors (2006). A 'dereverberation', or the compensation for discrete, annoying reflections is considered. Petrausch et al. (2006) simulates the performance of a WFS array to cancel reproduction room reflections as shown in Figure 4-20.

Apart from the compensation of the reproduction room influence, similar techniques of compensation can also be helpful to avoid other inherent errors of WFS reproduction such as non-ideal loudspeakers, diffraction effects and near-field effects (Corteel, 2007a). Corteel (2006) introduced a technique of multichannel equalisation to WFS. This method aims to equalise the WFS sound field in an enlarged area. The responses of the single WFS loudspeakers are measured by linear microphone arrays in the listening area. The loudspeaker signals and the array microphone signals form a MIMO (multi input, multi output) system which can be optimised by a multichannel inversion process. The result is a reduction of the synthesis errors. Corteel extended this method to reproduce directive sources in WFS (Corteel, 2007a) and to reproduction room compensation (Corteel and Nicol, 2003).

4.2.9 WFS and ambisonics

As shown in Daniel et al. (2003), WFS and ambisonics 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), or more exactly near-field-corrected HOA and WFS *"meet similar limitations"*. Both WFS and HOA and their unavoidable imperfections cause some difference in terms of the process and/or the quality of the perception. In HOA, with a decreasing order of the reproduction, the impaired reconstruction of the sound field will probably result in a blur of the localisation focus and a certain reduction in the size of the listening area.

4.3 Perceptual properties of WFS

4.3.1 Introduction

Wavefield synthesis is a significant step forward from stereophonic sound reproduction. It offers a noticeable enhancement of the sound field's spatial properties. Nevertheless, there seems to be a broad lack of clarity about the perceptual benefits of WFS. 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 somewhat optimistic assertion refers to the enhanced possibilities of WFS to reconstruct 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 spatial attributes 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 its advantages and drawbacks is apparent. They should be described clearly by means of suitable physical and psychoacoustical 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.

The following presentation of the current knowledge of the perceptual properties of WFS is based on a selection of attributes introduced in chapter 2. A general discussion on the perception principle applied for virtual sources in WFS and the transition to stereophonic perception forms the basis in section 4.3.2. Section 4.3.3 will cover the localisation attributes and section 4.3.4 will introduce the existing knowledge about the sound colour reproduction capabilities of WFS. Finally, a discussion about the specific properties of reproducing the distance of virtual sources is performed in section 4.3.5.

4.3.2 Perception principle for WFS sources

As WFS physically reproduces the natural sound field, the applied method of perceiving the virtual sources is probably similar to natural hearing as long as the reproduction is sufficiently complete. The unavoidable imperfections of this technique cause some differences in terms of the process and/or the quality of the perception, as the experiments in this thesis will show. However, the existence of artefacts most likely does not hinder the auditory system from perceiving the virtual source in the same way as a natural source.

The artefacts of a virtual source cause a degradation of the perceived quality, but it is intended that the perception mechanism is the same as for a natural source. Therefore the relevant localisation cues have to be provided by WFS. It was shown by Wightman and Kistler (1992)

that, as long as the low-frequency interaural time difference (ITD) cues are reproduced correctly, successful localisation is achieved, even if the high-frequency signal part is erroneous, e.g. provides conflicting cues. Hence, the spatial aliasing frequency may be as low as the limit of low-frequency ITD processing, which is between 2.5 to 5 kHz. Wightman and Kistler, however, did not comment on the difference between the quality of the localisation between signals with consistent cues in the entire frequency range and the signals having correct cues only in the lower range. It may be supposed that the latter have a decreased quality in terms of the localisation properties. The experiments described in this thesis will show relevant results.

Transition between WFS and stereo and the precedence effect

Subject to the condition that stereophonic perception and the perception of virtual sources in WFS are based on different principles (as stated), at a certain point there could be a transition between these perception types. In other words, it would depend on the array design whether the loudspeaker signals are perceived as single localisation stimuli (as proposed by the association model, see chapter 3.6.2) or as a whole after the physical synthesis of all array signals.

This can be discussed with reference to an experiment of Vogel (1993, pp.130ff). He conducted the very first experiments into WFS. The linear (all loudspeakers in one line) WFS array setup used for this investigation consisted of 12 loudspeakers, located 45 cm (a very wide spacing for WFS) from each other. This system has quite a poor performance with regard to spatial aliasing, which starts at f_{alias}= 380 Hz. Vogel commented in his experimental results, shown in Figure 4-21: *"… 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 perhaps be doubted. Vogel assumed that the correct synthesis at frequencies below f_{alias} (in this case 380 Hz) is responsible for the correct localisation. However, he did not consider that a simple principle is true for any non-focussed source in any WFS system: the first signal arriving at the listener is always the signal in the source direction. This means that the precedence effect supports the localisation of non-focussed WFS signals. Figure 4-22 illustrates the importance of the precedence effect in WFS localisation: 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 phenomenon creates a correct localisation cue at *all* frequencies.

The precedence effect is hypothesised by this author to be crucial, at least for the situation of Vogel's experiment described above. As a consequence, a larger directional error occurs if not the synthesised wave front – which is perfectly corresponding to the virtual source's di-

rection – but rather the nearest array loudspeaker is localised. In the described experiment by Vogel, the loudspeaker distance was 45 cm. If the precedence effect alone affected source localisation, this system would have a mean directional error of a quarter of the loudspeaker distance, in this case about 11 cm. Vogel's results do not show a smaller mean directional error (see Figure 4-21). Therefore, it cannot be concluded from his results that the system under investigation results in a successful synthesis of the array signals.

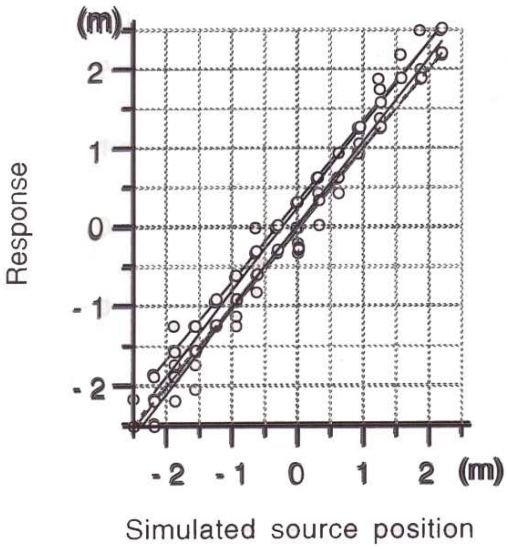


Figure 4-21: from Vogel (1993): Results from experiments with his first linear array setup with spacing 45 cm.
The 4 single graphs were arranged by this author so that all graphs share the same axes.

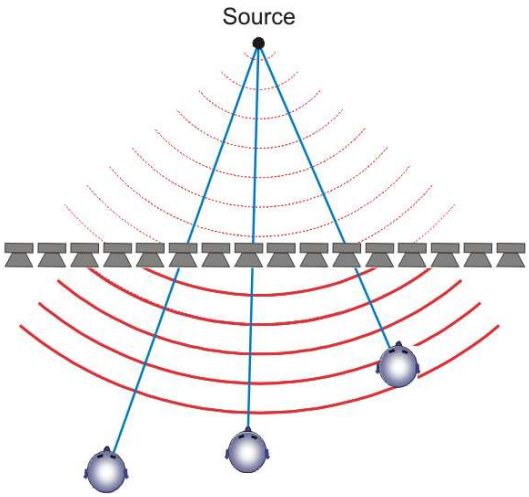


Figure 4-22: Illustration of the signal paths of the first signals arriving at the listener: The shortest path length leads to the shortest delay time and thus the first signal is approaching the listener from the direction of the virtual source. Hence, the precedence effect also may support virtual source localisation.

The precedence effect thus is more valid for the perception of WFS sources 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 sound fields after physical synthesis. The experiments in chapters 7 and 8 will show relevant results on this question.

4.3.3 Localisation properties of WFS

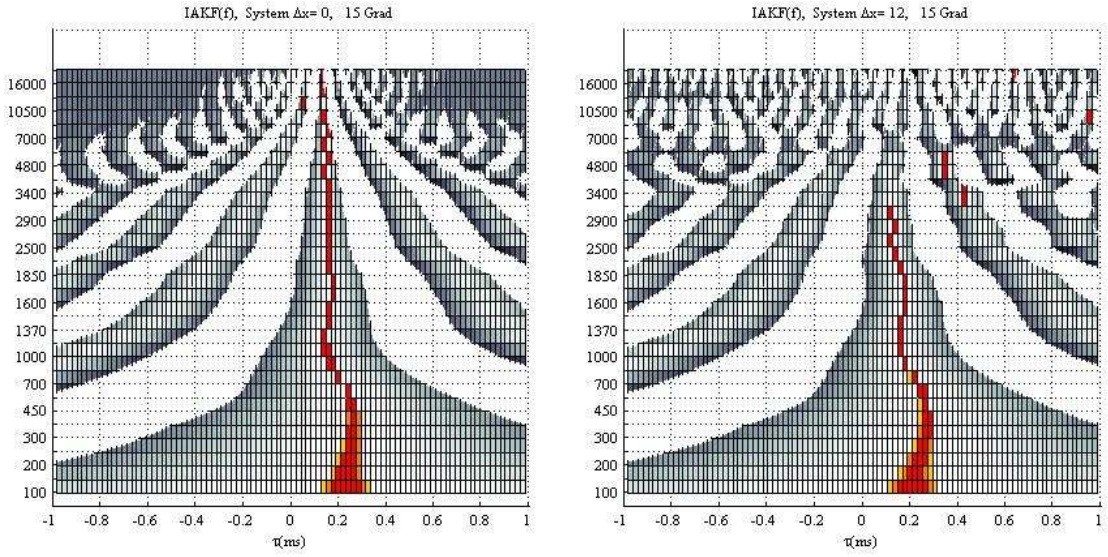
On the basis of the attributes introduced in chapter 2.3, a discussion of existing investigations into the localisation of WFS virtual sources is given. The three attributes of interest are the directional accuracy, the image focus and the locatedness.

Directional accuracy

The directional accuracy of a sound reproduction system is considered good if it is capable of reproducing a source in a certain direction without a significantly large system-caused deviation.

In theory, WFS is capable of creating accurate wave fronts and, as a consequence, accurate virtual source directions below f_{alias}. However, WFS is not capable of synthesising the wave fronts correctly above $f_{alias,}$ and this leads to an incorrect directional representation of these contributions. As these incorrect directional representations do not have a constant directional shift across all frequencies (see Figure 4-12), a direction bias of the perceived source direction probably does not exist. Instead, the various sound incident angles for different frequencies above f_{alias} will probably cause a decrease of locatedness and/or a blur of the virtual source. The influence of the reproduction room (see section 4.2.8) or the design of the test signal may, however, cause a certain bias. For instance, a single sine wave above f_{alias} certainly will be perceived in a wrong direction.

An analysis of the IACC and the ITD can illustrate the degree of similarity for source localisation above f_{alias}. The frequency-dependent ITD can best be analysed by the peak of the interaural cross correlation function (IACC). In Figure 4-23, both analysed sources are located at an azimuth direction of + 15°. The IACC diagrams show the congruence of real source (a) and WFS virtual source (b) below f_{alias}. Above f_{alias}, which equals 1.5 kHz in this example, the maximum of the IACC is correct until roughly 3 kHz. This means that the directional accuracy of a WFS source is correct for substantial frequency regions above the aliasing frequency. Methods of diffusing the WFS driving function above f_{alias} (see section 4.2.5) therefore reduce the physical similarity of real source and WFS virtual source and thus could give rise to a consequent decrease in image focus.



a) real source

b) WFS virtual source, Δx=12 cm, source is 2 m behind the array, listener is 1.5 m in front of the array. f_{alias} = 1.5 kHz

Figure 4-23: IACC (Interaural Cross Correlation) for sources localised at an azimuth of +15°

The directional accuracy as defined above can be measured by the signed error *E* of the data (for an introduction of the statistical measures see chapter 2.3.2. 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 combines both system bias and focus/locatedness into the term accuracy. In this way, he measures if there is any difference between the systems, regardless of whether it is system 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 wavefield synthesis itself.

In his experiments, Verheijen (1998) was the first to prove 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 which are between two lines of sight from the listener to positions near the edges of the array can be correctly synthesised. This zone is further minimised by applying tapering windows (see section 4.2.4).

Image 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 image 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 2.3.2 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 $< \bar{s} >$ was discussed there as well.

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

Start, analysing Vogel's experimental data, finds no significant difference between the broadband and the low-passed (<1.5 kHz) noise stimuli regarding the measures $< \overline{D} >$ and $< \bar{s} >$. Vogel mentioned, considering an experiment on an array with a spacing of 45 cm 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).

Start further tried to evaluate the localisation characteristics with tests using dummy head recordings in an 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 ($\rightarrow f_{alias}$ = 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 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 an initial scientific argument for ca. 1.5 kHz as a lower limit for f_{alias}.

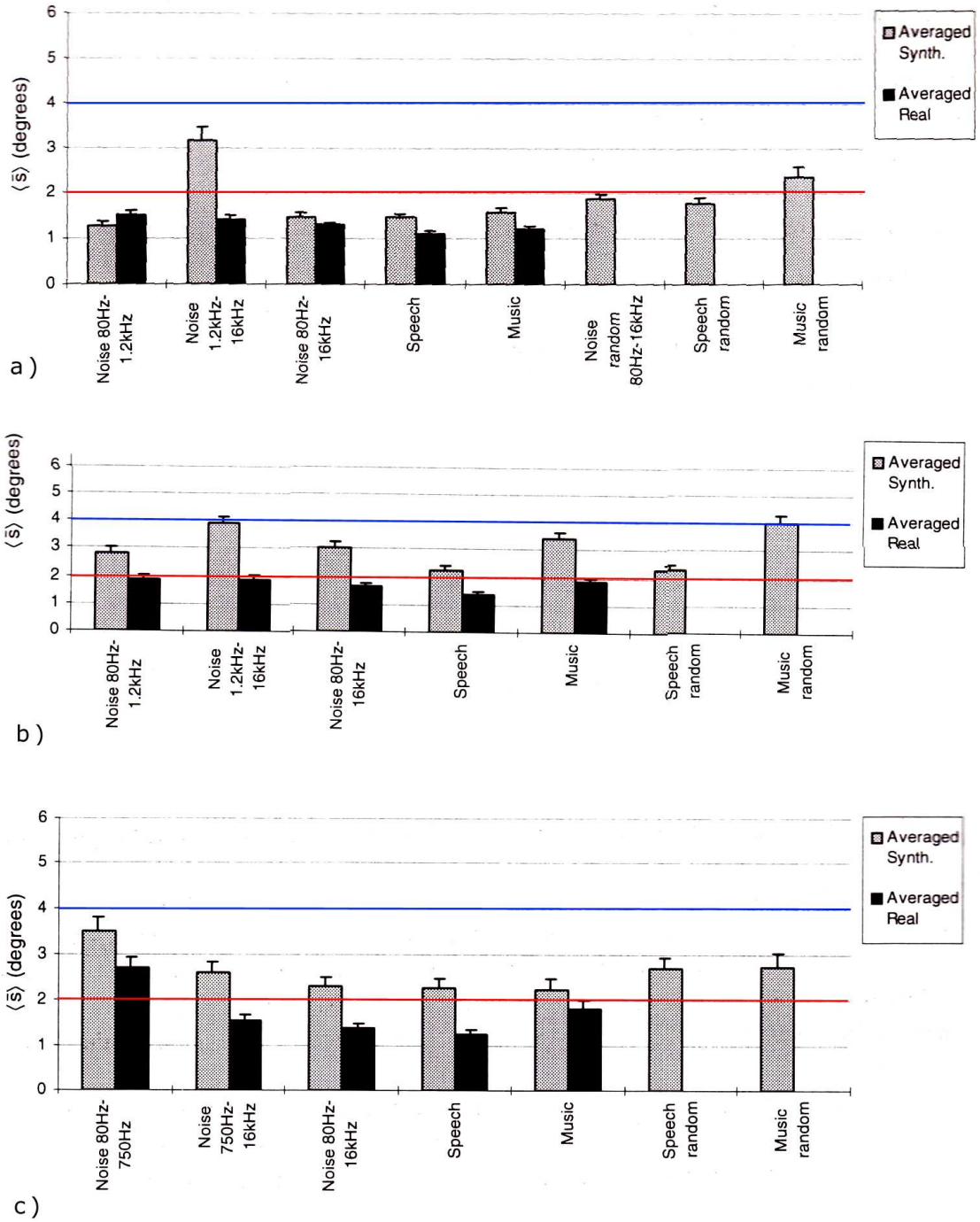


Figure 4-24: from Start (1997): Results of Start's experiments. The run standard deviation $< \bar{S} >$ of measurements of the perceived WFS virtual source directions in the following rooms is shown.

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.

Start repeated and expanded his experiments in an anechoic chamber and two concert halls. In Figure 4-24 the results are illustrated for the three different rooms. Start found that '*the localisation accuracy of low-frequency noise stimuli is almost identical for synthesised … and real sound fields … . As expected, localisation performance is seriously degraded for high-frequency noise stimuli'*. As depicted in chapter 2.3.2, with the standard deviation $< \bar{s} >$ an indicator 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 f_{alias} is as low as 750 Hz, this effect vanishes. The array's performance obviously becomes worse in the 'real' rooms b) and c). It cannot be concluded (as Start supposes) that the decreasing f_{alias} is the only reason for that. In this author's opinion, there are indications that the undesirable reproduction room influence causes irritations not only for depth perception, but also for localisation (see chapter 4.2.4). These indications are supported by the experiments of Verheijen described below.

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 $f_{alias,1}$ =1.5 kHz and $f_{alias,2}$ =0.75 kHz) he gave – after Start, as mentioned above – a second indication for the effect of too low a f_{alias}. However, as illustrated in the left part of Figure 4-25, the increase of the mean standard deviation disappears for the normal listening conditions in the 'reproduction room'. The reason for that is perhaps the general decrease of the localisation quality of WFS arrays in real rooms because of the reproduction room influence.

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-equalisation factor included in the WFS driving function (3dB/octave, see section 4.2.1) and therefore overemphasised the low frequency content of the pink noise bursts[10]. Thus the results may not be that comparable. In spite of that, the results of this second experiment using focussed sources are added to Figure 4-25. From the assessments of the

---

[10] *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 equalising*

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 section 4.3.2 as well. The surprisingly good result for the high-pass condition may have been a consequence of the existence of mid range frequencies (above 2 kHz) which were not aliased, and thus provided correct localisation. Verheijen explained it slightly more optimistically: *"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."*

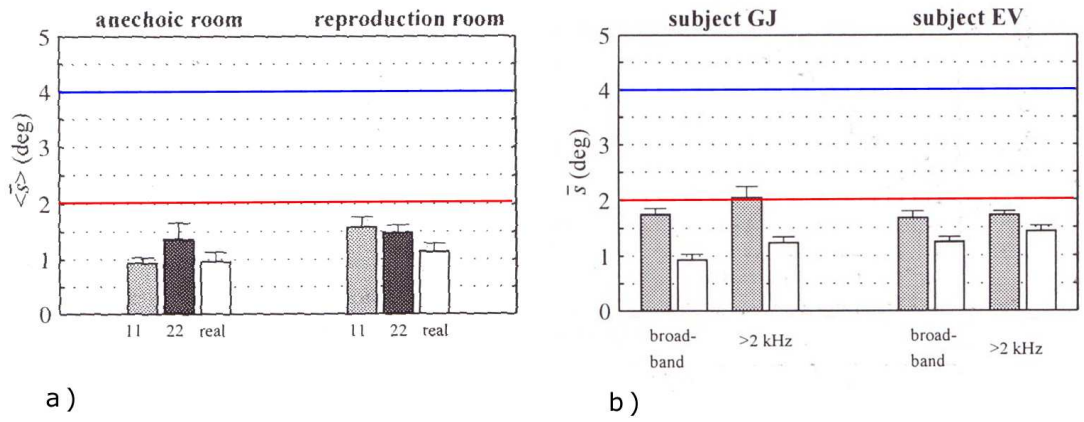


Figure 4-25: from Verheijen (1998): Results from Verheijen's experiments. (grey bars: virtual sources Δx=11 cm; black bars: virtual sources Δx=22 cm; white bars: real sources)

- **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)**

Figures were arranged and customised by this author.

Start (1997) investigated the 'spaciousness' of the real and the synthesised 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 was defined as source width (see chapter 2.3.1). Start compared the subjective assessment of the source width and the objective measure interaural cross correlation (IACC) coefficient 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 large rooms, once again through dummy head recordings. He found that the width of all sources was generally much larger than in the anechoic chamber, and also that the differences between the systems vanished, which may be a direct consequence.

Summarising, investigations on the localisation properties of WFS exist, but there is a necessity to further investigate in this direction. Up to now it is not clear where the lower limit of the aliasing frequency for adequate source localisation is, and whether a further increase of the aliasing frequency will result in a corresponding improvement of the localisation properties. Furthermore, the consequences of a decrease of the aliasing frequency need to be investigated thoroughly, because in practice a low aliasing frequency is likely to occur.

4.3.4 Sound colour and colouration

The sound colour of virtual sources in WFS is apparently impaired in comparison to natural sources. However, from the practical experience of WFS researchers, a reliable judgement can often not be derived, due to compromises that had to be undertaken in the laboratory systems regarding the size, quality and spacing of the utilised array loudspeakers. Also, the incorporation of DML panels (‘distributed mode loudspeakers’) or MAPs (‘multi actuator panel’; Boone, 2004) led to system-inherent sound colour degradations caused by the loudspeakers themselves. Existing scientific investigations on sound colour and colouration in WFS are rare.

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

Physical reasons:

a) Spatial aliasing distorts the higher frequency spectrum.
b) Finiteness of the array causes distortions through diffraction effects.

Psycho-acoustical reasons:

i. A non-optimal localisation process could lead to a distorted perception of the sound colour.
ii. The auditory event is not or not only determined by the synthesised wave front.

In the literature, very few investigations have been made in the field of WFS and the perceived sound colour. Reasons for this may be the great difficulty to measure the perceived sound colour and similar subjective attributes. One other reason could be a general agreement about the small audibility of negative effects on the sound colour in WFS. Bad sound colour is consequently often blamed on the degraded reproduction quality of the loudspeakers.

It is an important challenge to create a link between the physical and the perceptual artefacts of Table 4-3, because only then does it become possible to improve the performance of WFS

with respect to aliasing. However, the literature does not yet thoroughly describe how spatial aliasing is perceived. The perception of aliasing may be assumed in several possible ways. The question is how prominently the correct contribution is perceived – particularly in the frequency region just above the aliasing frequency.

Start (1997) alleges that a smoothing of the aliased content due to the finite resolution of the auditory system in the time domain does exist. Furthermore, he explains the audibility of aliasing effects with the fact that the lower limit of aliasing in the spectrum changes rapidly when the listener moves. His experiments revealed that by avoiding the periodicity of aliasing by time domain randomisation (see section 4.2.5) the colouration decreased (Figure 4-26), but the localisation performance apparently worsened (Figure 4-24). Furthermore, Start found that the colouration significantly decreased when the test signals were produced in a real room instead of the anechoic room. However, in his experiment on colouration, he utilised fixed dummy head recordings of WFS virtual sources at different positions, which can be regarded as non-optimal. His colouration measure was computed from paired comparisons.

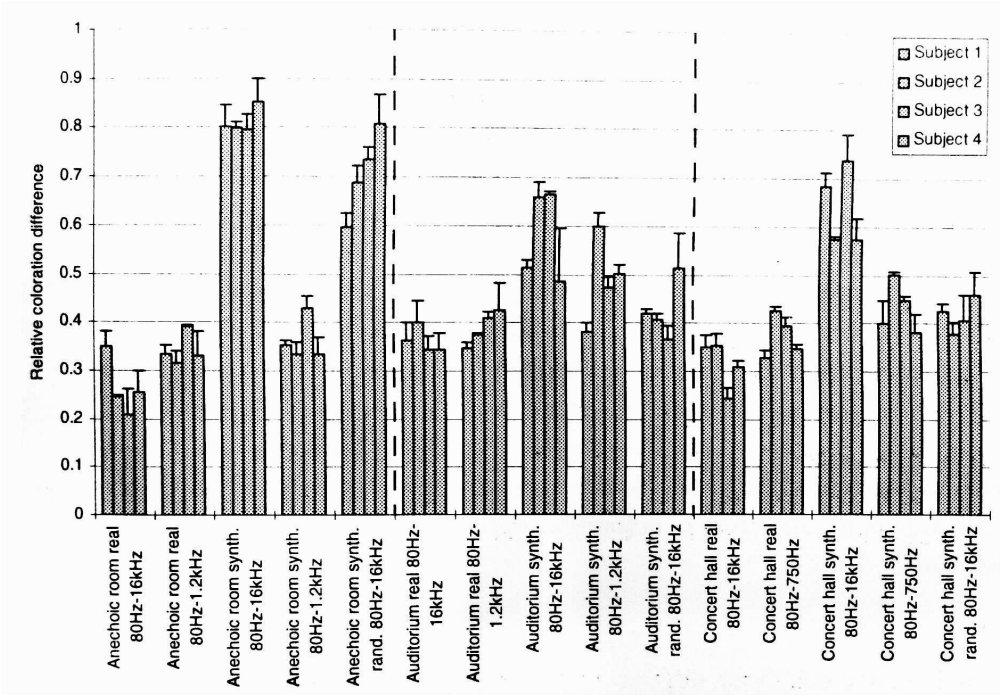


Figure 4-26: Colouration experiment by Start (1997): The colouration was measured by a difference between the sound colour of different virtual sources. Relative to the results in the anechoic room, the colouration decreased in the auditorium and the concert hall. The high frequency randomisation further decreased the colouration.

It is not clear whether diffraction artefacts cause a colouration of the signal through comb filter effects. These artefacts (after-echoes) could be regarded as reflections, and as such they

do not necessarily lead to audible colouration if they are successfully detected by the auditory system (see also Corteel, 2007a).

De Bruijn (2004) compared the perceived colouration of simulated WFS arrays with loudspeaker spacings from 12.5 cm to 50 cm with male and female speech signals. He used a dichotic reproduction of signals which were simulated at positions spaced in between-ears distance (0.2 m), but not binaural signals. He found significant differences between the perceived coloration for the different loudspeakers spacings and concluded that a spacing below 25 cm is sufficient for sound colour reproduction in the specific application of video conferencing.

After Theile (1980), the perception of the sound colour is part of the localisation process of the auditory system (as part of the 'gestalt' perception, see chapter 3.6.2). If a stimulus cannot be associated to a particular direction, it is perceived as coloured because the auditory system is not able to apply adequate inverse filtering. Colouration is therefore also a measure for the success of the localisation (and not only a measure for the *physical* 'perfectness' of the synthesis). Hence, it could be a more important measure than generally believed because it may provide information on the perception mechanism.

Further investigations on the sound colour properties of WFS will have to determine the lower limit of the aliasing frequency for different applications. Furthermore, the general way in which the aliased contributions in the WFS sound field are perceived has to be investigated before improvements can be proposed.

4.3.5 Distance

The discussion in chapter 2.5 introduced the attribute distance with regard to a comparison of WFS and stereo. The particular potential of WFS to simulate distance and depth will be considered here. The capability to adequately reproduce the source distance is vital for the success of this reproduction technique. Implementations of WFS which lack this capability, diminish the advantages, which WFS achieves by the creation of perspective, significantly. Examples of non-optimal WFS systems with regard to distance perception exist, which simply reproduce dry sources and thus cannot create an adequate distance perception. The specific cues available in WFS for distance perception have to be discussed, and system-inherent weak points unveiled.

For the discussion, the distance cues of WFS are organised into the different categories defined in chapter 2.5.2:

- 2D distance cues: monaural distance cues that enable a ‘pseudo’ distance perception (level, direct-to-reverberant energy ratio, frequency spectrum, interaction with other non-acoustical cues)

- 2½D distance cues: cues that are available with movements of the listener (motion parallax, improvement of several 2D cues)

- 3D distance cues: binaural distance cues that enable a ‘true’ distance perception (reflection pattern, binaural differences)

2D cues of WFS:

The 2D cues correspond to the cues available for WFS as well as for natural and stereophonic sources. It is hypothesised that there is no difference regarding the existence of these cues between WFS and stereo. Of course, WFS offers these cues in a large listening area, whereas in stereo the listening area is smaller.

2½D cues of WFS:

Through motion parallax (the change of the perspective with listener movements), WFS is able to create presence. Furthermore, WFS offers a realistic presentation of the spatial amplitude decay of a virtual source (Start, 1997; see section 4.2.4). The perspective of the acoustic scene and the amplitudes of the sources correspondingly change with movements of the listener within the listening area. However, the spatial amplitude decay does not fully agree with the real case.

These cues enable an implicit analysis of the scene geometry (and consequently distance) through moving within the listening area. They are called ‘idiothetic’ cues (see chapter 2.5). In WFS, the idiothetic auditory cues are rather similar to natural hearing. Figure 4-5 illustrates the change of source distances, source direction and perspective, which belongs to a change of the receiver position.

It is an interesting question whether the self-motion cues are able to override other, potentially wrong cues. Furthermore, it is of vital importance whether spontaneous, unconscious head movements would then be sufficient for depth perception or if conscious movements are required which give rise to a significant change of the perceived scene perspective in the sound scene.

A relevant investigation on these 2½D properties of WFS was performed by Noguès et al. (2003). Figure 4-27 shows their experiment setup. They asked the subjects to control the distance of a WFS virtual test source (green) so that the source distance matched the distance of

two other simultaneously reproduced virtual reference sources (orange). The subjects were asked to move around in the listening area. In the first experiment, the subjects could control the WFS test source distance through a change of the WFS test source position only. The other parameters such as direct/reverberant energy ratio and assumingly also receiver level were kept constant. In other words, by this procedure the subjects were asked to put the middle guitar in between the other two using the perspective cue of the WFS sound field. The results show that the subjects were indeed able to do this with a good agreement between adjusted and correct positions.

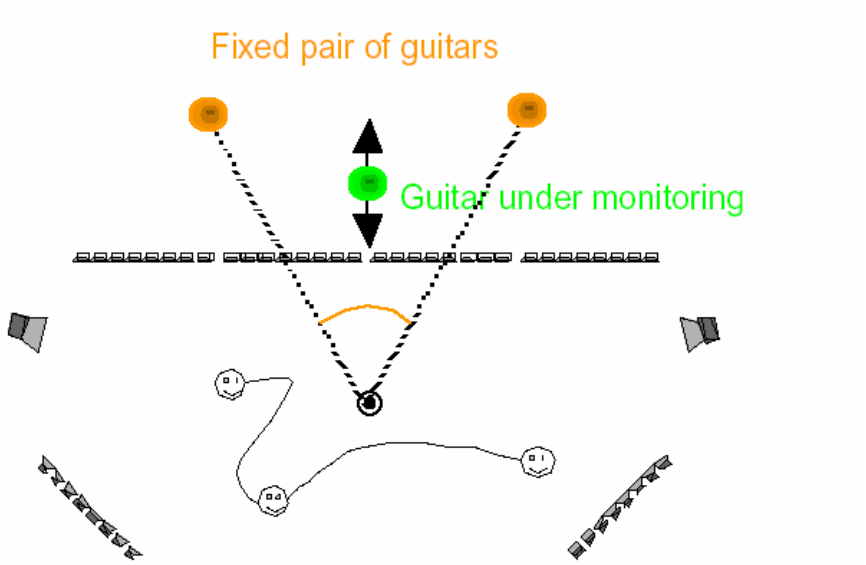


Figure 4-27: Experiment of Noguès et al. (2003): The subjects could move around in the listening area and so indirectly adjusted the guitar distance correctly.

In the other experiment in Noguès et al. (2003) the attribute 'source presence' (meaning the energy ratio between direct sound and reverberation) was included in the experiment. The subjects were asked to adjust the presence of sources synthesised at different distances in order to match the perceived distance with reference sources. The results showed that the adjusted presence was not dependent on the synthesises distance but only on the presence of the reference sources. This means that the perception of source distance due to the wave front curvature did not exist for far-field (> 1.5 m) sources, and even the perspective cue was overridden by other cues like the source presence. Their study only covered non-near-field virtual sources. Distance perception in the near-field will be considered in chapter 9.

3D cues of WFS:

WFS creates the correct shape of the curvature of the wave fronts and thus theoretically offers a cue to acquire source distance as illustrated in Figure 4-28. However, it is arguable that just by the correct wave front, a correct or even any distance perception at a static listening posi-

tion will be enabled. Literature does not support evidence for this assertion either, see chapter 2.5.1. Only at very near distances (< 1 m) is the wave front curvature a reliable cue for natural sources (Brungart and Rabinowitz, 1999c). However, in spite of not being a crucial cue, a correct wave front curvature may support the distance perception.

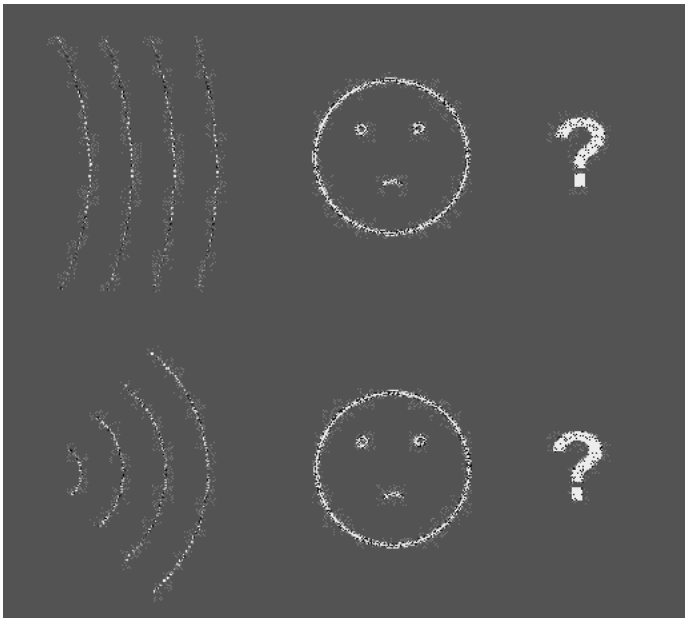


Figure 4-28: Is distance perception possible due to the wave front curvature?

The role of the dry wave front curvature in WFS is discussed in an experiment in chapter 9. Pre-existing investigations do not prove the existence of any effect of the wave front curvature on a static listening position. In de Bruijn et al. (2001) the distance perception of dry sources is commented: *"Although it is already well known that with WFS the source location is extremely stable when observed from different listener locations, depth or distance perception can only be obtained in combination with reflections and reverberation."*

In an experiment by Usher et al. (2004), the subjects were told that they were free to move within the inner curtain, which can be seen in the experiment setup in Figure 4-29. Thus, it can be deduced that - similar to the experiment by Nogués et al. described above - the movement of the subjects significantly improved the achieved distance perception. They presented dry focussed sources in distances of 1, 2 and 3 m from the listening point in the centre of the inner circle. The loudness was kept constant at the listening position, which means that loudness differences due to the different source distances were not available for distance perception. The sources were actually perceived at different distances, albeit apparently not very consistently. The distinction between closer and farther sources was possible only when the closest source (1 m) was incorporated, which could suggest the aforementioned conclusion about the role of listener movements.

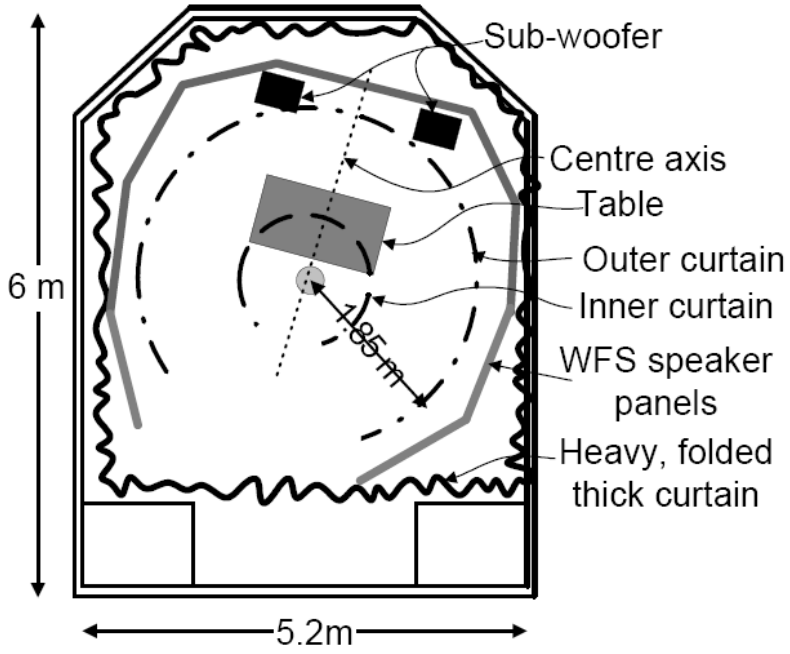


Figure 4-29: from Usher et al. (2004): Plan view of the experiment. The subjects were free to move within the inner curtain. Hence, some kind of distance perception was achieved.

Other investigations about distance perception in WFS which did not isolate the wave front curvature cue from the level cue can be disregarded, because the level cue is dominant and can lead to distance perception, as discussed in chapter 2.5.1. In this way, any reproduction technique, even mono, enables distance perception as proved in literature (Gardner, 1969 cited in Blauert, 1997).

There is another possible property of WFS that might give rise to an enhanced distance perception: WFS can very accurately simulate position and level of the early reflections. Hence, WFS may provide an enhanced possibility to discriminate these. In this author's opinion, it has not been proven that WFS is superior to 'sweet spot' stereophony (i.e. a stereo setup whereby the listener is located in the middle of a circle on which the loudspeakers are placed) concerning this point. Of course, the listening area in which correct (in this case distance-) perception is enabled is significantly larger in the case of WFS. But this argument is not regarded specific for distance perception. In a study by Neher et al. (2003), it was found that the listener could not 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, however, give rise to a better spatial perception.

Boone and de Bruijn (2003) investigated speech intelligibility using a comparison between two different WFS virtual sources. The two virtual sources were driven with different signals (a speech signal and a broadband noise as a masker) and compared with one loudspeaker, driven with both signals at the same time. They found that even when both virtual sources are in one and the same direction but synthesised 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 some perceptual difference between two virtual sources which only differ regarding their synthesised distance. However, this perceptual difference is not necessarily due to an analysis of the wave front curvature, but it could also be caused by other criteria that differ between sources at different distances. These are for example the response of the reproduction room or the energy distribution within the array. At least for frequencies above f_{alias} these criteria could lead to a difference in the width and the timbre of the virtual source. The conclusion in (Usher et al., 2004) supports this assumption: *"… in the absence of any indirect sound, when a source is positioned beyond a certain distance using a WFS system the curvature of the wavefront seems not used to determine the distance of the virtual source, but rather the timbre of the perceived source dominates."*

4.4 Summary of chapter 4

WFS is a spatial sound reproduction technique with unique potential. The performance of WFS has been described and discussed both with regard to physical as well as perceptual properties.

The unique properties with regard to directional imaging, in particular the enhanced acoustical perspective and the size of the listening area, qualify WFS and contrast it to other sound reproduction techniques. WFS is a solution for multiple listeners and is capable of avoiding a 'sweet spot' – which means the same acoustical image or the same acoustical scene can be created for many listeners. WFS is capable of producing a three-dimensional acoustical scene and the listener can move within that scene.

However, WFS is a compromise in its practical realisation. Certain physical parameters are impaired due to practical limitations and this influences the performance of WFS significantly. Both spatial aliasing as well as the limitation of the array to the horizontal plane avoid a perfect reproduction of the sound field as originally aimed by WFS. Hence, WFS does not have the potential of recreating a true copy of the sound field. Furthermore, the influence of the reproduction room makes the reproduction of a virtual room more difficult.

The perceptual attributes discussed include the attributes of localisation, sound colour and distance perception. The impairments of each attribute were analysed and relevant literature discussed.

The further discussion in this thesis and further experience will show whether the drawbacks limit the applicability of WFS. WFS can be an ideal reproduction technique for certain applications which demand its unique features.

5. The ‘OPSI’ method

5.1 Introduction

A new method of WFS reproduction is presented and discussed in this chapter. This method, known as OPSI (‘Optimised Phantom Source Imaging in wavefield synthesis’), aims at avoiding spatial aliasing artefacts, whilst also reducing the costs of the loudspeaker array. Furthermore, in the experiments of this thesis, it acts as a tool to investigate the perception principles of WFS and stereo reproduction.

OPSI is a hybrid approach of WFS and stereo. It uses a WFS virtual source below f_{alias} and a stereophonic phantom source above f_{alias}. As their individual auditory event directions match, they are perceived as one common source. The theoretical properties of OPSI are discussed and presented by simulations. The validity of these predictions is verified by the experiments in the subsequent chapters.

Section 5.2 introduces the OPSI method and section 5.3 describes a method of deriving OPSI signals. A pilot experiment for a determination of the so called OPSI localisation error is depicted in section 5.4. Section 5.5 discusses the size of the listening area for OPSI sources by simulations before section 5.6 summarises the chapter.

5.2 Substitution of the high-frequency contributions

Spatial aliasing is an artefact of WFS which is produced by the inaccurate synthesis of the WFS sound field above the spatial aliasing frequency f_{alias} (see chapter 4.2.5). Regarding perception, the reproduced sound field above f_{alias} is potentially incorrect in terms of the quality of the localisation and the timbral fidelity. Movements of source and/or receiver lead to an audible colouration of the source.

A solution is presented which makes use of well known facts about the quality of stereophonic perception. It is still not sufficiently understood why the comb filter effect in stereophonic listening is less audible than could be predicted from the ear signal spectra (see chapter 3.5.3). The sound colour is known to change only slightly between phantom sources from different directions, and it hardly changes at all while moving the head (provided that natural signals are reproduced).

In general, a phantom source is not considered superior to a WFS virtual source. However, above f_{alias}, the WFS virtual source no longer has better physical properties than the phantom

source. The idea is to combine phantom source and WFS virtual source to reduce the negative effects of spatial aliasing. This results in a hybrid system that uses a WFS source below f_{alias}, and a phantom source above. The arising hybrid source is expected to suffer less from sound colour degradations and negative effects with respect to localisation in comparison with the pure WFS virtual source.

The dominant cue for localisation is provided by the low frequency part of the source signal (see chapter 4.3.2). This means that the directional accuracy of the high frequency part of a virtual source could be lower than that of a real source without substantially degrading the localisation performance. This also applies for pure WFS: the low frequency part controls the perceived direction of the virtual source; the high frequency part plays a minor role – though it may increase the perceived image width to a certain degree (see chapter 4.3.3). It is assumed here that the consistency between high and low frequency part is important for the localisation quality of the virtual source. Though Wightman and Kistler (1992) generally proved the dominance of the low frequency ITD cue, it seems a plausible assumption that inconsistent cues might give rise to a decreased locatedness and an increased image width.

It is a challenge for WFS research to gain insight into the impact of the aliased high frequency part of a WFS virtual source on localisation and sound colour perception. Its substitution by another technique could be a suitable method to test its role for localisation and its influence on the perceived sound colour. Should it be true that the WFS reproduction totally fails above f_{alias}, there is no reason for maintaining the same reproduction technique for this frequency range. On the other hand, spatial aliasing can be regarded as a superposition of both correct and incorrect contributions; therefore it is possible that above f_{alias} the correct contribution may be salient (see chapter 4.3.3). Hence, a replacement of WFS above f_{alias} may lead to a degraded localisation performance.

5.3 Generation of OPSI signals

Figure 5-1 illustrates how a substitution of the high frequency part of WFS can be realised. The array is fed with low-passed ($< f_{cross}$) WFS signals to reproduce the accurate wave front. The high frequency source components are generated by several loudspeakers (solid, blue) which are fed with high-passed ($\geq f_{cross}$) signals. These loudspeakers need to be spaced significantly wider than the loudspeakers of the WFS array and thus produce a stereophonic image. The low and the high frequency source are expected to merge unless the difference in their incident angles is too large. The cut-off frequency f_{cross} of low-pass and high-pass is identical, such that the contributions in the two frequency bands add up to give a flat fre-

quency response similar to a two-way loudspeaker system. This frequency is called the cross-over frequency f_{cross}. In the optimal case, f_{cross} is chosen such that $f_{cross} = f_{alias}$.

This new technique is called 'OPSI' (Optimised Phantom Source Imaging in wavefield synthesis).

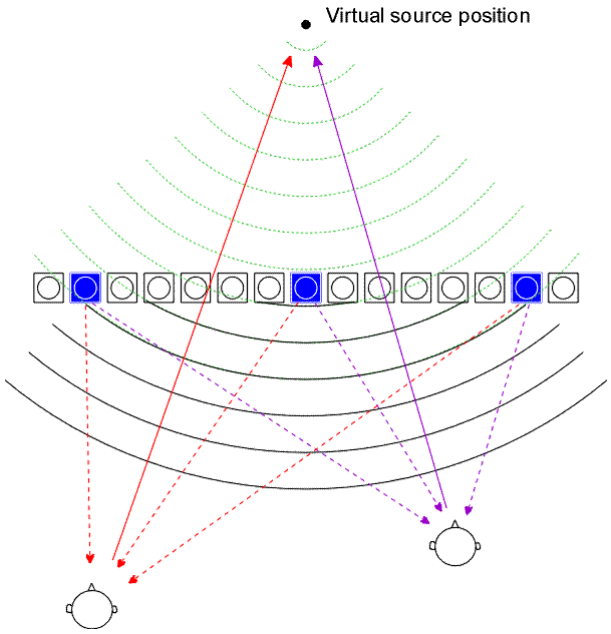


Figure 5-1: Example of an OPSI system: Three loudspeakers (blue) replace the WFS array for reproducing the high frequency part

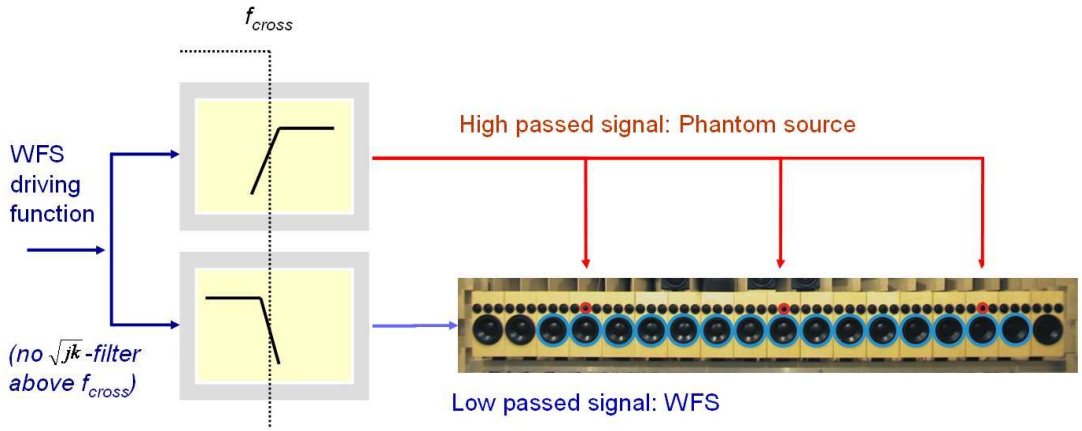


Figure 5-2: Generation of OPSI signals: The WFS array of woofers (marked in blue) is fed with the low-passed WFS signals only. The tweeters (marked in red) are fed with the high-passed WFS signals after a level adjustment. A small number of tweeters installed in the array is sufficient for a reproduction due to the OPSI method. It is important that the $\sqrt{jk}$ -filter (3dB/octave boost) in the WFS driving function is applied only for frequencies below f_{cross}.

Figure 5-2 illustrates the method of generating an OPSI signal for the WFS array. The array is provided with the conventional WFS signals after the low-pass at f_{cross}. The tweeters are fed with the high-passed WFS signals after a level adjustment. It is mandatory that the $\sqrt{jk}$ -filter (3dB/octave boost) in the WFS driving function is applied only for those frequencies below f_{cross}. In conventional WFS, this filter is generally applied below f_{alias}. A level adjustment (due to the smaller number of drivers) is necessary for the phantom source speakers in order to ensure a flat frequency response at the listening position. This level adjustment equals the proportion of array and phantom source speakers.

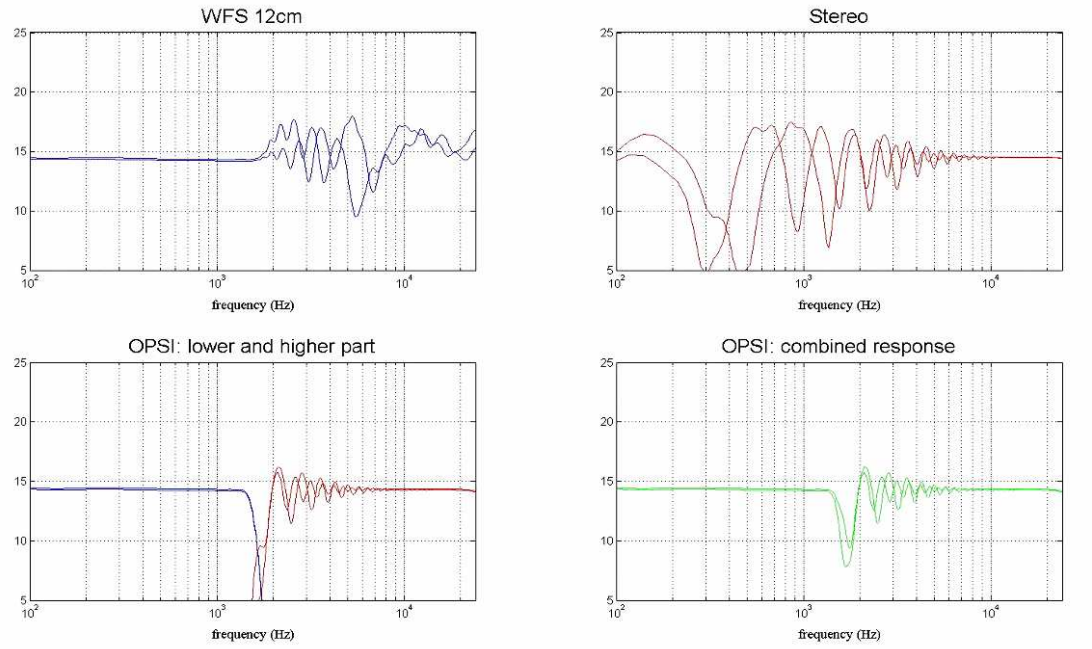


Figure 5-3: Frequency spectra illustrating the principle of OPSI. Spectra are simulated in the reproduced sound field with omni-directional microphones at ear positions. The simulation assumes ideal omni-directional loudspeakers and anechoic conditions. Spectra are smoothed according to critical bands by Patterson filters.
The WFS sound field (top-left) contains spatial aliasing above f_{alias}, in this example f_{alias}=1800 Hz. The stereophonic reproduction (top-right) contains strong comb-filtering. The peak-notch level difference of the comb filtering decreases with frequency because of the increasing integration of the critical band filtering. The idea of OPSI is to combine the unaliased low frequency part of WFS with the high frequency part of stereo (bottom-left). They are added up to give a newly created OPSI signal (bottom-right).

It is expected that the perceptual artefacts of spatial aliasing will vanish under such circumstances. However, it is not obvious what will happen to the perceived sound colour. The stereo-like reproduction generated by OPSI most likely utilises more than two loudspeakers in the array. When more than two of these loudspeakers create a similar level at the listening position, a severe comb filtering could occur. It is known that a reproduction of coherent signals on more than two loudspeakers gives rise to audible colouration (Theile, 2001). It is hy-

pothesised that the OPSI solution does indeed improve the sound colour reproduction of a WFS source, whilst not degrading the localisation performance, at least not significantly. These hypotheses are examined in chapters 7 and 8.

Figure 5-3 shows an analysis of frequency responses for WFS, stereo and OPSI. The frequency responses show that OPSI also reduces the spectral deviations in the ear signals that could give rise to colourations. The reason is the increasing integration of the critical band filters: with increasing frequency, the number of peaks in the comb filter that fall into a single critical band increases. The result is an increasing smoothing of the comb filter until the spectrum becomes totally flat at high frequency bands (see top-right diagram in Figure 5-3). These analyses are discussed further in chapter 8.5.

5.4 Pilot experiment: maximum OPSI localisation error

The WFS virtual source and the phantom source are expected to merge and to be perceived as one auditory event. This requires their individual auditory event directions to be in sufficient agreement. In other words, the directions of low-frequency virtual source and high-frequency phantom source must not differ excessively, otherwise they will be perceived separately, or give rise to an increased image width.

The difference between the two individually perceived source directions, namely the directions of low-frequency WFS source and high-frequency phantom source, is known as the *OPSI localisation error*. In order to avoid negative effects, the OPSI localisation error should be smaller than a specified maximum value, which has yet to be determined. For the applicability of the OPSI method, it is crucial that the localisation performance of an OPSI source is not noticeably worse than that of a normal WFS virtual source.

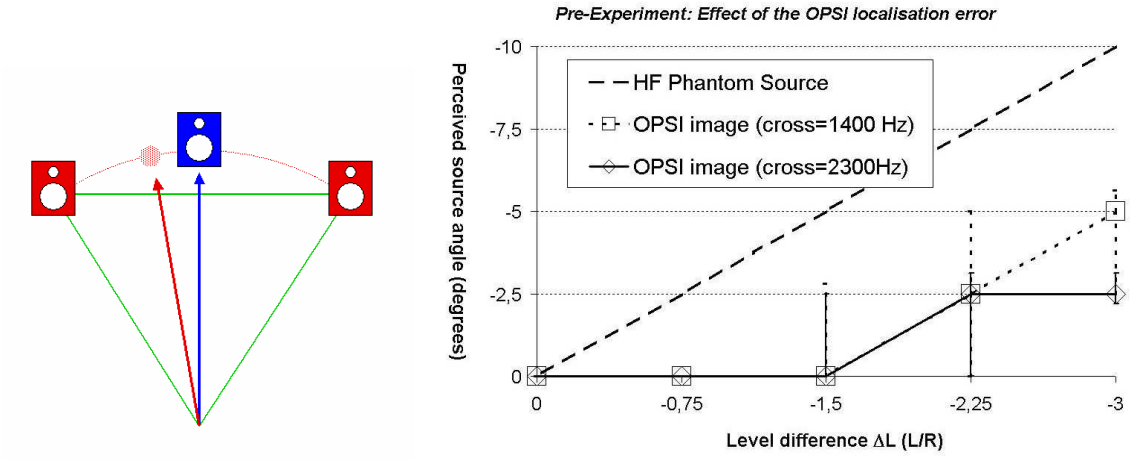
A pilot study helped determine the maximum allowed OPSI localisation error. The experiment was performed using the three frontal speakers of a conventional 3/2 stereo setup (Left, Center, Right). The test signal (anechoic female speech), was split into a low-frequency part and a high-frequency part at the crossover frequency. As illustrated in Figure 5-4a, the Center loudspeaker (blue) reproduced the low-frequency part (LF) of the test signal, simulating the WFS array. The high-frequency part (HF) of the test signal was superimposed by the two loudspeakers Left and Right.

The phantom source was created in five different directions using inter-channel level differences. These level differences together with the according phantom source directions are

given in Table 5-1. The dependence between level difference and phantom source shift was estimated using data of an informal test described in section 5.5.

1) ϕ (LF) = Center speaker	**0°**	**0°**	**0°**	**0°**	**0°**
2) ΔL(L/R)	**0 dB**	**- 0.75 dB**	**- 1.5 dB**	**- 2.25 dB**	**- 3 dB**
3) ϕ (HF) = Phantom source (L/R)	**0°**	**- 2.5°**	**- 5°**	**- 7.5°**	**- 10°**
4) OPSI Localisation Error = 3) – 1)	**0°**	**- 2.5°**	**- 5°**	**- 7.5°**	**- 10°**

Table 5-1: The dependence between level difference and phantom source shift was estimated using data of an informal test described in section 5.5.



a) **b)**

Figure 5-4: Pilot experiment: Determination of the maximum OPSI localisation error.

Figure 5-4a shows the applied loudspeaker setup. The Center loudspeaker (blue) emanates the low frequency part (LF) simulating the WFS contribution to the OPSI source and the red loudspeakers emanate the high frequency (HF) phantom source. The blue and red arrows illustrate the perceived angles ϕ(LF) and ϕ(HF) of the respective individual contributions.

Table 5-1 shows the parameters for the different stimuli. An OPSI localisation error was created through the superposition of LF contribution at 0° and a phantom source HF contribution at five different directions. The phantom source was shifted by a level difference ΔL(L/R).

Figure 5-4b shows the result of the pilot experiment. The median of the perceived angles of the merged source is shown with upper and lower quartiles. The results are shown for two different crossover frequencies $f_{cross,1}$ = 1400 Hz and $f_{cross,2}$ = 2300 Hz. The dashed line shows the OPSI localisation error.

Nine subjects took part in the pilot experiment. The subjects were able to toggle between two cases: case 1 was the situation in which HF and LF images matched perfectly. This case corresponds to the first column of Table 5-1 with an OPSI localisation error of 0°. Case 2 was one randomly chosen situation in which HF and LF images did not match. It was asked

whether the perceived source direction changed while toggling between the cases. If it did change, the perceived direction of the second source had to be noted. For this purpose, a degree scale was visible to the subjects.

Two different crossover frequencies were used in the experiment. The source content was split at $f_{cross,1}$ = 1400 Hz and $f_{cross,2}$ = 2300 Hz respectively. Through this procedure, it was checked whether the maximum allowed OPSI localisation error depended on the crossover frequency.

The results are presented in Figure 5-4b, the x-axis of which corresponds to Table 5-1. It can be seen that a change in the perceived source direction is virtually non-existent for the cases where the OPSI localisation error does not exceed 5°. For an OPSI localisation error of 10°, the perceived direction changes by roughly 2.5° for f_{cross} = 2300 Hz and by roughly 5° for f_{cross} = 1400 Hz. The subjects did not report on any split images. The locatedness and the image focus of the OPSI source were subject of the experiment described in chapter 7.

5.5 Size of the listening area

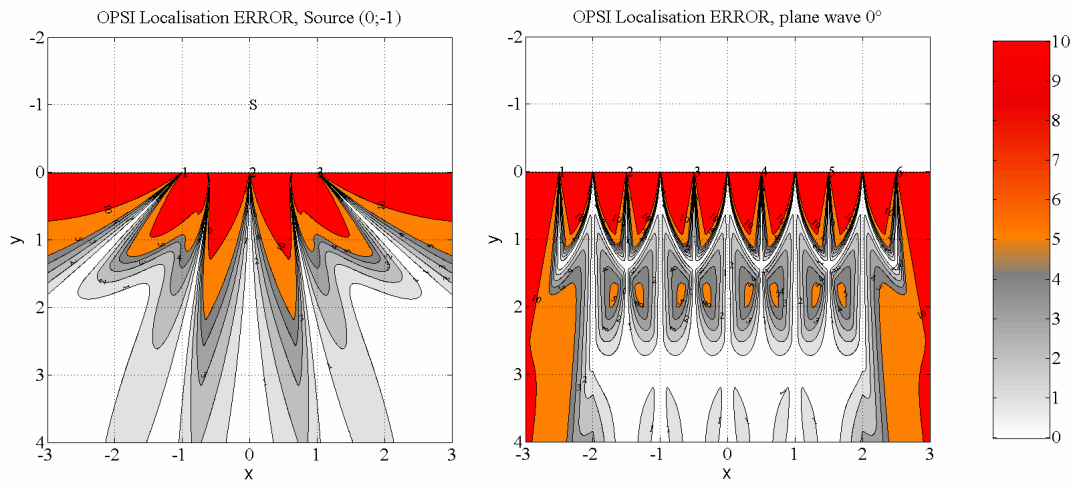
It is desired to design the OPSI source such that it is localised similarly at various receiver positions in the listening area. The advantage that WFS offers, of having an extended listening area, should not be compromised by the OPSI method. However, it is well known that the perceived phantom source direction changes when the listener moves within the listening area. As the OPSI method utilises phantom sources for the high-frequency part of the source, this effect must be taken into consideration, and evaluated. On the basis of known psychoacoustical data, simulations can be performed in order to estimate the OPSI localisation error. As found in the preceding section, the OPSI localisation error has to be lower than 5° for optimal reproduction.

In principle, the level differences that are necessary for a certain phantom source shift can be calculated according to phantom source theory (Williams, 1984; Theile, 2001; Wittek, 2001a; Wittek and Theile, 2002). However, these general theories cannot be applied when using high frequency phantom source content alone. This is because the phantom source shift is dependent on frequency (Pulkki, 1999a). Hence, a preliminary test had to be performed to find the approximate dependence of the phantom source shift on the level difference for high frequency (> 2 kHz) source content. According to this test, the shift factor of 7.3%/dB (Wittek and Theile, 2000b) changes to 12%/dB for high frequency content only. This mapping rule was also used for the creation of the phantom source directions in the pilot experiment de-

scribed in the preceding section. The shift factor for time differences does not change, it is 12.7%/0.1 ms (Wittek and Theile, 2000b).

These rules can be used to make simulations of the OPSI localisation error. At each point within the entire listening area, the corresponding incidence times and levels of the signals of the phantom-source-generating loudspeakers were calculated. By a mathematical procedure (not specified here) that translates the relationships of level and time difference stereophony to vector theory, an incorporation of more than two loudspeakers was possible. By this procedure, an estimate of the apparent phantom source direction could be given. The OPSI localisation error is calculated as the difference between this direction and the direction of the WFS virtual source.

Figure 5-5 shows the OPSI localisation error in degrees on a scale from 0° to 10°. An impairment of the localisation can be expected for the red areas which correspond to an OPSI localisation error of 10° or more. The diagrams show a listening area of the size $4 \cdot 6\ m^2$. The virtual source (S) and the phantom source speakers (denoted by numbers) are indicated in the figures. The WFS virtual source direction is assumed to be in ideal agreement with the direction of the source position. The WFS array (reproducing the virtual source) is on the line y=0. The distance between the phantom source speakers is 1 m.



a) Virtual source 1 m behind the array ('S' is at y=-1, x=0).

b) Plane wave = virtual source infinitely far behind the array.

Figure 5-5: Simulations of the OPSI localisation error (in degrees) for two different virtual source positions. The array is located at y = 0 m. The phantom source loudspeakers are marked with numbers. The unit of x- and y-axis is meters. The maximum allowed OPSI localisation error of 5° is indicated in orange colour.

Figure 5-5 shows that the OPSI localisation error depends on the position of the virtual source. It has also been observed that the size and position of the stereo loudspeaker setup influences the performance, although it is not simulated here for the sake of brevity. In order to optimise for a minimum OPSI localisation error, the suitable stereo loudspeakers should be chosen depending on the synthesised virtual source distance. In the case of plane waves or sources with larger distances, more than just a few stereo loudspeakers have to be used. In practice it would be optimal to automatically and dynamically allocate a suitable stereo setup to the virtual source. The stereo loudspeakers are usually part of the WFS array. They should make a grid of loudspeakers spaced by not more than 1 m. In Figure 5-5a, only the three nearest stereo loudspeakers are chosen in order to keep the number of active stereo loudspeakers small, whilst also minimising the OPSI localisation error at the boundaries of the listening area. For the plane wave in Figure 5-5b, the whole set of stereo loudspeakers of a certain grid is active leading to correct localisation due to the precedence effect for most parts of the listening area. The area close to the array shows larger OPSI localisation errors in both figures, but it should be noted that pure WFS also performs non-optimally in this area due to near-field errors (Corteel, 2007a).

The fewer OPSI loudspeakers are used, the better the perceived sound colour is expected to be. Hence, a position and distance-dependent control of the stereo loudspeaker signals needs to be performed. It is optimal when the number and position of the stereo loudspeakers are chosen depending on the virtual source position. An optimum coverage of most parts of the listening area is possible for all virtual source positions when certain rules are applied. From the simulation shown above it may be assumed that a grid of stereo loudspeakers with an inter-loudspeaker distance of not more than 1 m is required to keep the OPSI localisation error sufficiently small in a large listening area.

5.6 Summary of chapter 5

OPSI is a new method proposed to avoid spatial aliasing. It makes use of the minimal sound colour differences between phantom sources. Its idea and implementation has been described in this chapter.

The OPSI idea is intended to improve the performance of WFS without introducing new artefacts. Based on the results of a pilot experiment, the area in which no negative effect on the directional accuracy occurred was calculated. This area can be maximised by an intelligent selection of OPSI loudspeakers depending on the virtual source distance. Consequently, the listening area of an OPSI source was found to be not substantially smaller that that of a conventional WFS virtual source.

The approach of substituting the high-frequency contributions of WFS with phantom sources has been verified for its applicability. The OPSI method is only likely to be beneficial if the sound colour reproduction is improved whilst the quality of spatial reproduction is kept at least stable. Hence, the OPSI approach needs to be investigated regarding its sound colour performance. In addition, the merging of low- and high-frequency contribution needs to be examined with regard to the properties of localisation. Further experimental investigations designed to address these points are described in the following chapters 7 and 8.

6. Rationales for the experimental comparison between WFS and stereophony

6.1 Introduction

The sound reproduction techniques WFS and stereophony were introduced in chapters 3 and 4. Their physical and perceptual properties have been described to the extent to which they have been so far investigated in literature. This thesis aims at comparing their perceptual properties for listeners at a fixed listening position. Therefore, this chapter summarises the established properties in order to analyse the remaining open questions and differences regarding the most important attributes of spatial perception. The fields considered as most apparent and relevant for practical sound reproduction were examined by investigations described in the subsequent chapters. The discussion in this chapter will prepare a foundation for these investigations and will explain their validity and intention.

This chapter is organised by the perceptual attributes on the basis of which a comparison will be made. After this introduction, the attributes of localisation (section 6.2) and sound colour perception (section 6.3) will be discussed. The general advantage of WFS over stereo is believed by this author to be its capability of supporting listener movements. Section 6.4 explains why this capability is not considered by investigations in this thesis. Finally, section 6.5 deals with the potential of WFS and stereo to synthesise the impression of source distance, before section 6.6 summarises the chapter.

6.2 Directional imaging, image focus, locatedness

Both in WFS and stereo, it is possible to create virtual source directions in the entire horizontal plane. A prerequisite in the case of WFS is a surrounding loudspeaker array, while stereo requires a surrounding setup of loudspeakers (a ‘multichannel surround’ setup). However, the properties of the directional image are apparently different. Although one can expect that both systems are capable of reproducing a source in a certain direction, the attributes directional accuracy, image focus and locatedness of the source have not yet been investigated in a direct comparison. Furthermore, the OPSI system has never been incorporated in practical experiments.

In stereo, the phantom image is known to be less focussed than for natural sources. This applies in particular for lateral phantom source reproduction. The image focus and the robust-

ness of the phantom source can be improved with an increasing number of loudspeakers, leading to a narrower grid of stable loudspeaker positions (Zieglmeier and Theile, 1996). Furthermore, the focus depends on the way in which the phantom sources are produced, whether this be time or level panning, or a combination of both. The listening position is also a decisive factor, as an off-centre listening position produces deviations between the incidence times of the loudspeaker signals (refer to chapter 3.5 for references). With multichannel surround in particular, the difference in focus between a phantom source and a single loudspeaker signal is used creatively by the sound engineer. For example, the Center channel alone creates a small focus and a high locatedness whereas a phantom source between the Left and Right channel, which is localised in the same direction, is perceived less focussed (Silzle and Theile, 1990).

In WFS, both the focus and locatedness depend on the spatial aliasing frequency (Start, 1997). Spatial aliasing is considered the main reason for a decreased focus in WFS. OPSI sources, which contain no spatial aliasing but on the other hand employ stereophonic imaging, cannot be predicted at all with regard to the localisation quality. The experimental results in chapter 7 will show how they perform.

WFS sources in front of the array (‘focussed’ sources) have to be treated separately. The image focus of a source depends on its locatedness, which is generally considered lower for such sources. There are several reasons for this. Firstly, the precedence effect does not support the localisation of focussed sources, because the first contributions do not arrive from the virtual source direction (see chapter 4.3.2). Secondly, distance perception is potentially problematic for focussed sources, because the perceived reflection pattern generally does not fit the synthesised source distance (see chapter 4.2.8). A pilot experiment investigating the locatedness of focussed sources is described in chapter 8.7.

The investigation described in chapter 7 was centred on exploring differences between the sound reproduction techniques WFS, OPSI and stereo with regard to the attributes of localisation. These attributes were the directional accuracy, the image focus and the locatedness. It was aimed to answer the following research questions:

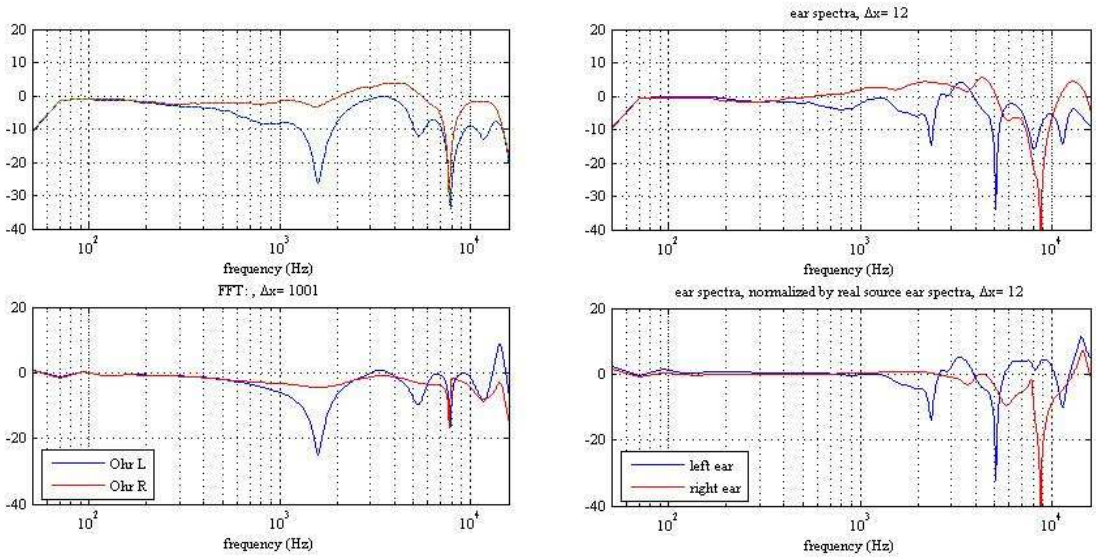
- How does the localisation performance of WFS compare to real sources and what aliasing frequency is required for a sufficient localisation? Does an increase of the aliasing frequency above this limit still improve the localisation performance?

- How do WFS and stereo differ with regard to the attributes of localisation?

- Can the OPSI method be validated in terms of its directional imaging?

6.3 Sound colour, colouration

Figure 6-1 shows an analysis of the ear signal spectra of a stereo and a WFS source. In the lower diagrams, the difference between the respective frequency spectrum and the spectrum of a real source at the same location as the virtual source is given. The properties of the two spectra are not dissimilar, both showing deviations above a certain frequency. It can be seen that the WFS virtual source (b) is an exact copy of the real source below f_{alias}. The level-panned phantom source (a) also shows some similarity to the real source below 1000 Hz. An analysis of these spectral differences together with an analysis and measurement of their effect on the perceived sound colour is detailed in chapter 8.



a) Two-channel stereo **b) WFS virtual source**

Figure 6-1: Frequency spectra of the ear signals for sources at an azimuth of +15° created by different reproduction techniques. red: ipsilateral (right) ear signal, blue: contralateral (left) ear signal. Top diagram: binaural room transfer function (BRTF), bottom diagram: difference between this BRTF and the BRTF of a real source at the same location.

- **a) Two-channel standard stereo setup, L(L/R)= -7dB.**
- **b) WFS virtual source, Δx=12 cm, source is 2 m behind the array, listener is 1.5 m in front of the array. $f_{alias} \approx$ 1.5 kHz.**

According to the discussions in chapters 3.5.3 and 4.3.4, neither stereo nor WFS offers a transparent reproduction. In both cases, the physical spectra of the ear signals show prominent deviations from that of a reference source. Furthermore, the sources at different locations or from different directions clearly differ from each other, for each reproduction system. In the case of stereo, the spectral deviations are caused by comb filtering through a summing of the loudspeaker signals; in the case of WFS they are mainly caused by spatial aliasing.

Hence, both the perceived sound colour and the colouration of the systems are likely to be non-optimal. Their respective performance has not yet been sufficiently investigated in literature, neither for each system independently, nor in a direct comparison. The comparison of stereophonic reproduction and a sound field reconstruction technique such as WFS is expected to challenge established views on these techniques and specify the respective best areas of application.

The second aim of the thesis is to help identify the perception mechanisms applied when listening to WFS and stereo reproductions. The prediction of the colouration performance is a suitable tool for this task, because it checks the role of the physical ear spectra on the perceived colouration. In this way, the presence of any kind of decolouration is checked.

The comparison of the sound reproduction systems WFS, OPSI and stereo was the focus of the investigation described in chapter 7. The research questions regarding the attributes sound colour and colouration were as follows:

- Can the prefigured advantage of OPSI regarding sound colour reproduction be confirmed?

- What effect does spatial aliasing have on the colouration in WFS? Does the colouration differ with different spatial aliasing frequencies?

- Can colouration in WFS be predicted by the frequency spectrum of the ear signals? What are suitable predictors?

- What is the relationship between predicted and actually perceived colouration for the WFS, the OPSI and the stereo sources? Can a decolouration in the perception of the signals be identified?

- Can conclusions thus be drawn from these relationships that help to explain the perception mechanism?

6.4 Size of the listening area, robustness, listener movements

The listening area is defined as the maximum area over which the sound field is reproduced without distortions regarding perceptual attributes. The typical aim of a reproduction system is to have a large listening area so that a number of people can listen simultaneously. The robustness of a reproduction system describes its ability to maintain the same sound image despite movements of the listener.

There are two types of listening area. With the first type (A), it is intended that all listeners share the same acoustical perspective, i.e. a scene is reproduced (e.g. with accompanying two-dimensional picture) that should create the same acoustical sensation for all listeners. No listener movements are supported in this case.

The second type (B) supports listener movement, with each listener having his own acoustical perspective on the scene. A movement of the listener leads to an according change of the acoustical perspective. The listeners then do not have the same acoustical sensation at different locations.

The choice between these two types of listening area influences the potential of the reproduction systems to produce a large listening area. This choice is further influenced by the existence of a picture. A comparison of the size of the listening area is not possible without considering these alternatives.

Stereo is a reproduction system that is optimal at a single listening point, this being the so-called ‘sweet spot’. As the directional image is produced by interchannel level and time differences, a distortion of these inevitably leads to a distortion of the directional image. Only at the sweet spot will the original incident time relations between all loudspeakers be kept. The angular fidelity of the stereophonic image will therefore be correct only within the sweet spot. Small movements of the listener out of the sweet spot will strongly influence the perceived sound image. In principle, the robustness of stereo is small. Only in the case of two-channel stereo, i.e. 2/0 stereo, is the distortion of the incidence times of the loudspeaker signals zero on an entire axis, this being the line of symmetry between the loudspeakers.. The level relations between the loudspeakers are distorted on all locations except this point or line as well. However, the level distortions are much smaller than the time distortions and thus do not distort the directional image to the same extent (Wittek, 2001a).

The listening area depends on the panning method or the technique applied for the stereophonic recording. Although the distortions with listener movements are equal for all panning methods, in a level-panned image, the extreme left and right sources stay constant and thus level difference stereo is regarded more stable (Wittek, 2001a). The listening area is bigger at the price of a reduced spatial complexity of the stereo mix, for instance in cinema mixes that are optimised for a large audience. In this example, the dialogue is reproduced monophonically in the Center channel so that the sound image matches the visual image for the whole audience. Using such methods, both the robustness and size of a stereo listening area can be increased. However, in stereo, only a listening area of type A can ever be created, which means that listener movements are not supported. Any listener movement results in an un-

natural change in the acoustic image. The only constant source locations for an enlarged listening area are the loudspeaker positions.

In WFS, the listening area is determined by the size of the array. Sources can only be reproduced in directions in which an array is located. The WFS virtual sources are robust within the listening area. WFS can produce a large listening area for both types of listening areas. This means that WFS can create the same acoustical image for all listeners, as well as an acoustical scene enabling listener movements.

With constraints, the spatial quality of the reproduced scene does not depend on the distance of the listener from the array. However, in the region very close (< 1 m) to the array, the synthesis fails due to the far-field approximation of the WFS theory (Corteel, 2007a). Furthermore, due to the 'amplitude errors' (see chapter 4.2.7), the level balance between sources at different distances is distorted at all locations except for a reference line of receiver positions. The reason is the distortion of the level decay with distance for conventional one-dimensional WFS arrays. In the case of reproduced focussed sources, the listening area is reduced because the source must always be between the listener and the array.

When WFS is applied with an accompanying two-dimensional picture, for example in cinema applications, the advantage of an enlarged listening area is reduced. The flat picture and the sound image containing true perspective causes a mismatch of localised directions as soon as sound sources are synthesised in front or behind the screen (de Bruijn and Boone, 2003). Also, when a conventional stereo mix is reproduced in a cinema by a WFS array, the advantage of WFS is rather small, as the front loudspeakers have to be reproduced at the same positions as the original loudspeakers in order to fit the visual image. Only the surround channels can be positioned further away leading to an improved level distribution within the listening area.

Relevance for this research

One of the main advantages of WFS over stereo is generally considered to be the capability of supporting listener movements. The effect of listener movements, the size of the listening area and the robustness of the systems differ significantly between WFS and stereo. In spite of this, the scope of this thesis only covers the perceptual differences at a fixed listening position. Hence, these attributes are not considered in this investigation. This does not mean that these differences are ignored or considered unimportant. Rather, the differences at a fixed listening position are considered more relevant for the discussion about the general perception mechanisms. The perception mechanism most probably does not depend on conscious listener movements, because it is regarded as a spontaneous process (Theile, 1980).

The discussion in this section is necessary in order to put the scope of this thesis into the general context of the field. This thesis does not target a general comparison between WFS and stereo, nor does it aim at defining a superior reproduction technique. Rather, it covers specific properties that are important for certain applications. The results of this investigation have to be considered in combination with the above-mentioned general differences.

6.5 Depth, distance, immersion

In this section, the capabilities of WFS and stereo to reproduce depth and distance will be compared. This will be achieved through an assignment of available cues according to the scheme and terminology introduced in chapter 2.5.2. A differentiation was made between the 2D distance cues which are the monaural distance cues that enable a 'pseudo' distance perception, the 2½D distance cues which are cues that are available with movements of the listener and the 3D distance cues which are binaural distance cues that enable a 'true' distance perception. The 2D cues include level, direct-to-reverberant energy ratio, frequency spectrum and interaction with other non-acoustical cues. The 2½D cues include motion parallax and the 3D cues consist of the reflection pattern and binaural differences.

A true, intuitive depth and distance perception ('3D') relies on the natural existence of reflections and reverb and, at least for non-nearby sources, it does not depend on the wave front curvature of the direct sound. The 3D cues, and therefore real 3D depth perception, do not rely on listener movements and thus can be produced at a fixed listening position as well. Only through the combination of motion parallax and reflection pattern cues can an acoustic scene be created containing depth, which also enables movements. Table 6-1 summarises the cues for distance perception and the analogies to visual perception as mentioned in chapter 2.5.2.

	2D representation, non-intra-active	**2½D representation, intra-active**	**3D representation, non-intra-active**	**3D representation, intra-active**
Cue, visual	Monocular cues (overlay, linear perspective, shadow, relative size, etc)	Monocular cues + motion parallax	Monocular cues + binocular cues (disparity, convergence)	Monocular cues + motion parallax + binocular cues
Cue, acoustical	Monaural cues (loudness, spectral cues, direct/reverb ratio, etc)	Monaural cues + motion parallax	Monaural cues + reflection pattern	Monaural cues + motion parallax + reflection pattern
Enables	Pseudo depth/ distance, no movement	Pseudo depth/ distance + movement	True depth, no movement	True depth + movement
Spatial audio system being capable of this representation	Mono	WFS (dry sources)	Stereo (with adequate spatial reproduction)	WFS (with adequate spatial reproduction)

Table 6-1: Analogy between visual and acoustic cues for the perception of depth and distance

Table 6-1

This means that true depth and distance reproduction is possible also in stereo. There is no essential difference in distance reproduction except for the existence of motion parallax and binaural differences (the latter are not included in the table because it first has to be checked whether they represent an existing cue at all in WFS, see chapter 9). The precondition for natural distance and depth perception in stereo is the existence of lateral reflections (Theile, 2001). This means, only multichannel stereo offers the required cues to a sufficient degree. Furthermore, systems such as 3/2 stereo suffer from certain shortcomings that arise from a too low number of loudspeakers and the resulting overly large loudspeaker spacings. The robustness of lateral phantom sources and the reduced possibility to place sources in the rear and on the sides are negative consequences. These shortcomings will not be concentrated upon here because they are not a failure of the system itself, but only of its realisation. Furthermore, the solution is simple, namely the incorporation of additional loudspeakers (Zieglmeier and Theile, 1996). It is also assumed here that a stereo system is able to reproduce an adequate reflection pattern with sufficiently high resolution of reflections. The accuracy in which the reflection pattern is reproduced is different for WFS and stereo in the case of a stereo system with a low number of loudspeakers. This difference may give rise to perceptual differences.

The only remaining difference that might be valid for distance perception on a fixed listening position is the wave front curvature cue. This cue is valid for distances below roughly 1 m in real source distance perception.

The investigation described in chapter 9 explored the cues for distance perception of nearby sources at a fixed listening position. The research questions were as follows:

- Is WFS capable of reproducing cues related to the wave front curvature at a fixed listening position? How do the sound fields of real and virtual sources differ with regard to these cues?

- Is distance perception in WFS possible due to these cues?

6.6 Summary of chapter 6

As a preparation for the investigations in this thesis, this chapter defined the main research questions. At the end of each section, distinct open questions were expressed, which formed the basis for theoretical and practical examinations described in the subsequent chapters. A summarising comparison of properties was presented, which were introduced in the chapters 3, 4 and 5. Furthermore, those fields were selected in which demands for further insights exist. The motivation for the selection is a more complete comparison of the sound reproduction techniques WFS and stereo at a fixed listening position and the validation of the OPSI method. In addition, further knowledge about the effect of spatial aliasing in WFS on perception is aimed at.

Regarding the localisation attributes, an experimental comparison of directional accuracy, image focus and locatedness will be described in chapter 7. Furthermore, a study concerning the colouration in the respective systems was undertaken (chapter 8). This study was targeted to unveil basic properties in the reproduction and perception principles of the systems. Lastly, an investigation on distance perception explored the effect of the wave front curvature in WFS for a static listener (chapter 9).

Due to the attention of this thesis on perceptual differences at a fixed listening position, some main differences in the properties between WFS and stereo were not considered. It was emphasised that this does not mean that these differences were ignored or considered unimportant. Rather, when defining the topical focus of this thesis, the differences for a fixed listening position were considered more relevant for the discussion about the general perception mechanisms.

7. Experiment 1: Localisation properties of WFS, OPSI and stereo

7.1 Introduction

This chapter describes an experiment comparing the localisation performance of several systems, including natural reference sources, stereophony, different WFS systems, and an OPSI system. The experiment consisted of a measurement of perceived auditory event direction and an elicitation of the locatedness of these auditory events. The evaluation of the measurements allows a comparison of the localisation properties of the different sound reproduction techniques.

The research questions for this investigation were developed in chapter 6.2. The general aim of this experiment was a direct comparison of the imaging capabilities of the systems under investigation. Furthermore, by incorporating WFS systems of different quality, an estimation of the effect of spatial aliasing on localisation could be made. Finally, it was investigated whether the OPSI concept could be validated regarding its imaging properties.

Section 7.2 summarises the contents of the experiment before the experimental procedure is described in section 7.3. The systems under assessment are introduced in section 7.4. Section 7.5 depicts the results of the experiment which are discussed in section 7.6. A summary is given in section 7.7.

7.2 Contents of the experiment

In the experiment, a number of systems produced (virtual) sources in various directions. For each source, the following measures were recorded:

- the perceived direction of the auditory event
- the subjective locatedness.

The measurement of the perceived direction allows an analysis of the directional accuracy of a system. By statistical analysis of the subjective direction measurements, the image focus may be indirectly inferred. These attributes of localisation were introduced in chapter 2.3.

These systems participated in the experiment:

- Natural sources (single loudspeakers)
- Stereophony
- WFS (Δx_1=4.2 cm)
- WFS (Δx_2=12.7 cm)
- OPSI

The systems are described in section 7.4.

Meaning of the 'locatedness' for this study

The locatedness is defined as the spatial distinction of the perceived source. It is hypothesised that from this measure, conclusions about the quality or success of the perception process can be drawn. The better a source can be perceived, the better the locatedness will be. This attribute can not be estimated by physical measures of the sound field alone, as complex psychoacoustical processes underlie the perception process of a source. Any ambiguities in the process are believed to lead to a degradation of the locatedness (Theile, 1980).

Hence, the relevance of certain physical parameters of the sound field for spatial perception can be studied by an assessment of the locatedness. After the relevant physical parameters for spatial perception have been established, a virtual source can be designed which fulfils the requirements for optimal spatial perception. The physical parameters which were identified to be irrelevant, may then be arbitrary. The virtual source is no longer required to be a perfect copy of the natural source. The question of which relevant physical parameters are reproduced insufficiently by a WFS virtual source is vital, and still unresolved. The same applies for stereophonic sources – the locatedness of a phantom source is better than one could expect by analysing its physical properties (see chapter 3.6).

7.3 Experimental procedure

7.3.1 Acquisition of auditory event directions and locatedness grades

In prior experiments concerning the localisation of WFS sources (Start, 1997; Verheijen, 1998), the auditory event directions were measured using numbered array loudspeakers. However, this technique may cause the measured directions to be 'pulled' towards the visible anchors ('ventriloquism effect'; Blauert, 1997). To avoid this, an acoustically transparent canvas was used to hide the loudspeakers. There was no scale whatsoever on the curtain visi-

ble to the subjects. Figure 7-1 shows the curtain and a subject using a laser-pointer to indicate the perceived auditory event direction. The subjects were filmed by a camera in order that their results could be collected. The directions were measured on a two-dimensional scale comprising both azimuth and elevation of the perceived source direction as can be seen in Figure 7-2. The validity of this procedure was examined in a pilot test.

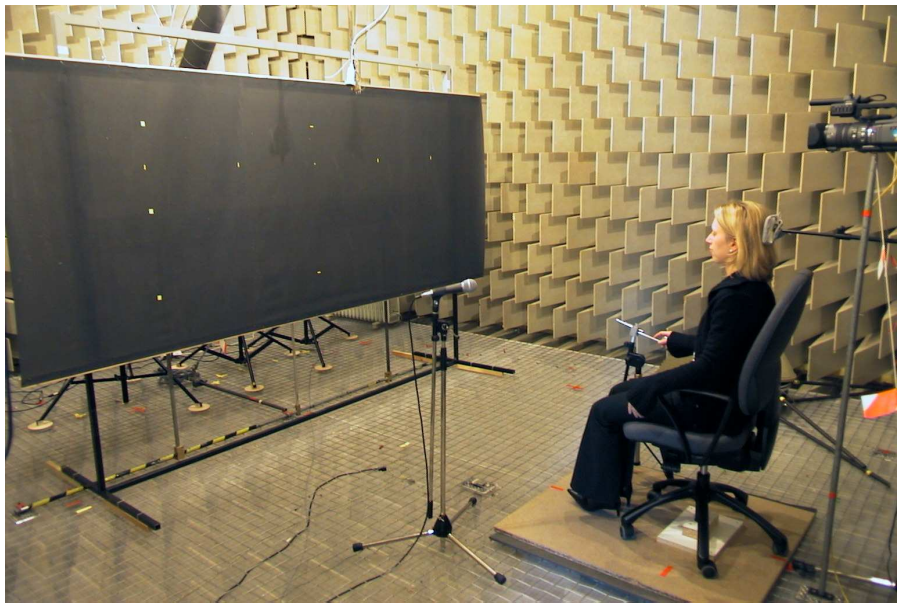


Figure 7-1: Experimental procedure: The subjects could target the position of the auditory event through a laser pointer. The laser pointer was mechanically supported by a flexible joint.

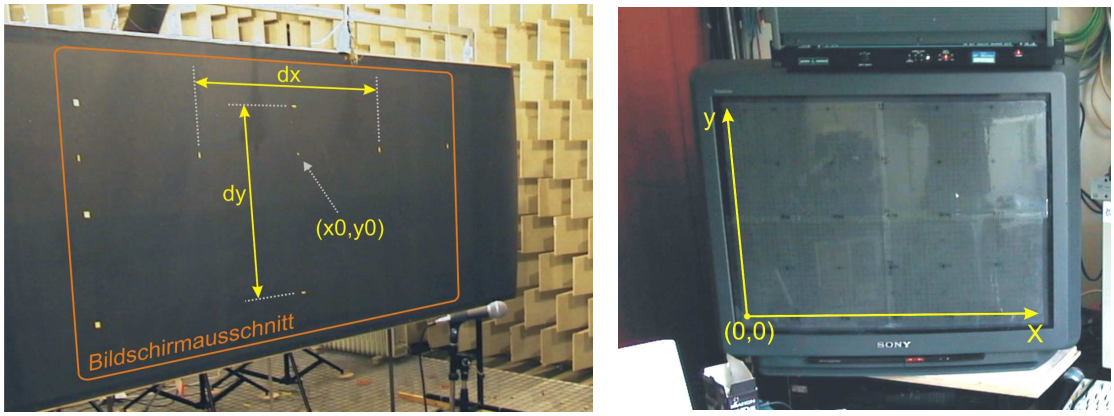


Figure 7-2: The scale on the acoustically transparent curtain (left picture) was invisible to the subjects. The experimenter viewed the results on a screen in the neighbouring room (right picture) which showed the picture from the video camera. Pictures from Huber (2002).

After pointing at the perceived source direction, the subjects were asked to assess the locatedness of the reproduced source. As this attribute is not self-explanatory, it was extensively explained in a preceding talk with the experimenter, and defined in the test sheet given to the subjects prior to the experiment. The grading of the locatedness was made on a 5-grade scale, ranging from 1 (very good) to 5 (bad):

How well can you localise the source? How well can you assign a particular direction to the perceived source?

1 – very good

2 – good

3 – fair

4 – bad

5 – very bad

In advance of the experiment, three example test items were reproduced to introduce the test procedure to the subjects. This allowed the subjects to familiarise themselves with the range in which the perceived auditory event directions and perceived locatedness would lie.

7.3.2 Stimuli

It is known that the quality of the localisation depends on the type of the source signal. To achieve an optimal localisation performance, a broadband signal is required (chapter 3.6.2). Furthermore, signal onsets can serve as an important cue (Blauert, 1997). Therefore, pink noise bursts were chosen as a suitable stimulus. However, noise bursts are not a natural source content. This is a crucial point considering that the perception of image focus, distance and other spatial attributes is likely to depend on the prior knowledge of the source content. In this experiment, as the difference between the systems was the key point, this prior knowledge was regarded as disturbing because it would conceal the system-caused differences. In other words, for instance the expected character of a human voice would probably influence the subjects' perception. A source signal such as pink noise bursts, which is close to natural signals regarding the frequency spectrum and neutral to the subjects concerning the prior knowledge, served best for the task of this experiment.

Figure 7-3 shows the envelope of the pink noise bursts. The burst had a length of 1 sec and was repeated three times. The rise and the fall time was 100 ms. A break of 400 ms was introduced between each of the bursts. The parameters of the noise bursts were chosen deliberately after preliminary tests. They were found to provide optimal conditions for localisation.

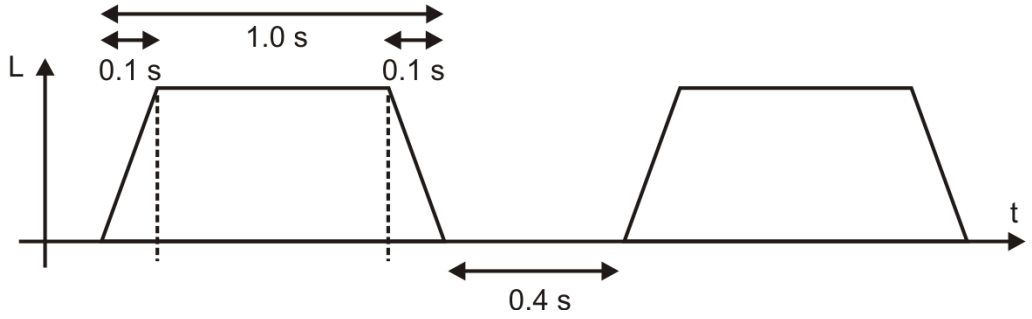


Figure 7-3: Envelope of the stimulus: Pink noise bursts with a duration of one second each

All reproduced test items were calibrated to have an equal level at the receiver position. A constant receiver level of 68 db(A) was created using computational simulations and measurements at the receiver position with a measurement microphone.

7.3.3 Test panel

18 subjects participated in the experiment, their ages ranging from 21 to 58 years. None of the subjects had known hearing impairments. The experience of the subjects in audio differed considerably. A post-screening (chapter 7.5.1) resulted in the rejection of five subjects.

7.3.4 Experimental setup

Figure 7-4 shows the setup of the sources used in the experiment. Three different source directions were reproduced, these being –10°, 0° and 5° relative to the array normal. The subjects (a head indicates the subjects' position) were positioned 2 m in front of the array. The loudspeakers representing the natural sources (red) were installed 1 m behind the array in three different positions. The phantom sources were produced through two (blue) of the array speakers. The subjects could move their head freely.

The experiment was undertaken in the anechoic chamber of the IRT, Munich. Its volume is 80m^3, with a floor size of 4.5 · 6 m^2. Its lower limit frequency is 80 Hz.

Figure 7-5 shows an overview of all systems of the experiment. The active loudspeakers/drivers for each system are marked in colour and with arrows.

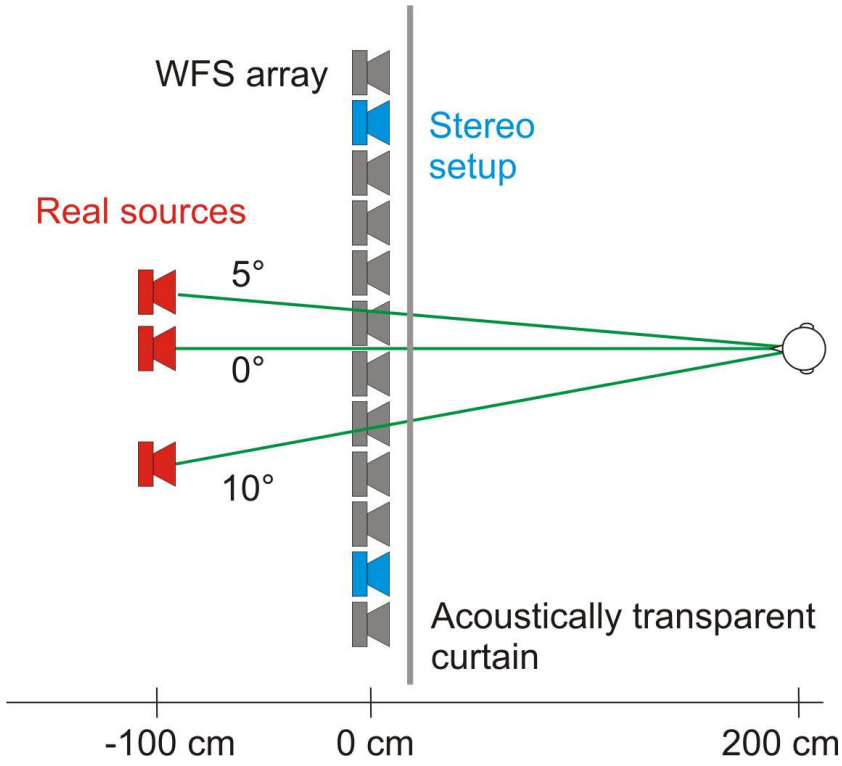


Figure 7-4: Test source setup for the experiment: The subject is denoted by the head at the right side. The WFS array including the stereo setup (blue), single loudspeakers (red), and acoustically transparent canvas can be seen on the left side.

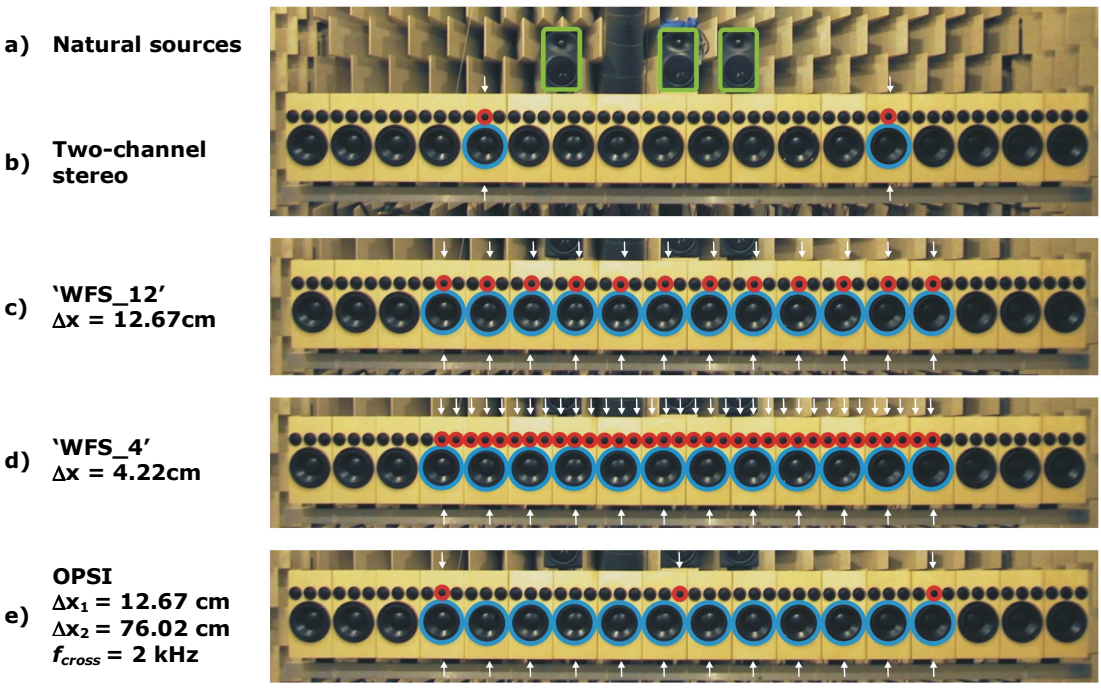


Figure 7-5: Illustration of all systems of the experiments. Pictures from Huber (2002).

The array consisted of custom-made two-way loudspeakers, with one woofer and three tweeters in each loudspeaker box. This setup enabled a very small distance of secondary sources for WFS. The woofer distance was 12.67 cm; the tweeter distance was 4.22 cm. The chosen cross-over frequency was 2 kHz, and was applied by means of digital filtering.

7.4 Systems under assessment

Natural Sources

Single loudspeakers were used as a reference for this experiment. As can be seen in Figure 7-4, they were positioned 1m behind the WFS array at three different positions. To avoid shadowing by the WFS array, they were installed slightly higher than the array, which can be seen from Figure 7-5a. The loudspeakers were Klein&Hummel O100, being active monitor loudspeakers of studio quality. They were digitally equalised to match the frequency response of the array loudspeakers.

Two-Channel stereo

The 5th and 14th array loudspeakers were used to reproduce phantom sources, according to stereophonic principles (see Figure 7-5). The resulting offset angle of the two loudspeaker stereo setup was 31.8°. In (Wittek and Theile, 2000b; see chapter 3.5.1), methods were presented to calculate the necessary signal differences for a phantom source at a certain azimuth. According to these methods, three different phantom sources were produced at different azimuths, these being -10°, 0° and 5° (see Figure 7-4). The phantom source shift was achieved through an equivalent use of both time and level differences.

WFS_12 and WFS_4

The virtual WFS sources were produced using the so-called 2½D operator, i.e. the driving function for a WFS array, as mentioned in chapter 4.2.1. This essentially means that each loudspeaker reproduces a signal similar to that which a microphone at the loudspeaker position would pick up from a source at the virtual source's position. A different level roll-off and a frequency-dependent filtering compensates for the 2D design of the array. This frequency compensation (the so-called '$\sqrt{jk}$ -filter') is to be applied below f_{alias} only, meaning that each source requires a unique, individually designed filter.

In this experiment, the way in which WFS configurations differ with respect to their localisation performance was of particular interest. Two different WFS setups were created. The first setup had a secondary source (array loudspeaker) spacing of Δx = 12.67 cm. For simplicity, it will be referred to as WFS_12 from this point onwards. The second setup had a secondary source spacing which was three times smaller, Δx = 4.22 cm. It will be referred to as WFS_4.

As explained in chapter 4.2.6, tapering should be applied to the array signals in order to reduce truncation effects. For the WFS arrays used in the experiment, cos^2-windows were applied at the array edges. The tapering window is schematically illustrated in Figure 7-6.

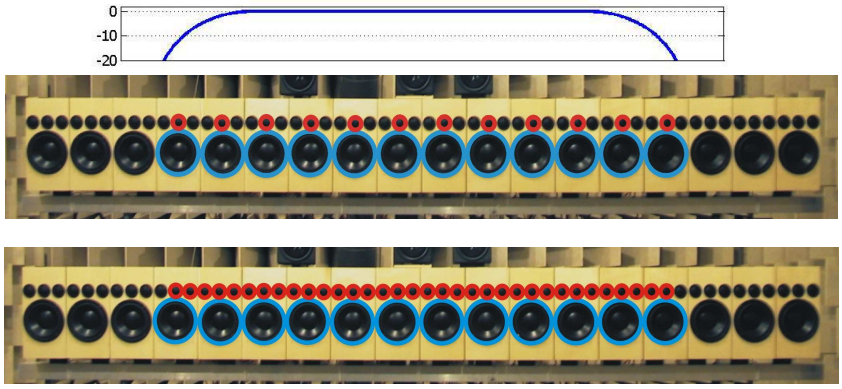


Figure 7-6: Tapering window for the WFS arrays:
The diagram on top of the photos shows a schematic view on the tapering window. It illustrates the level attenuations which are applied to the corresponding array loudspeakers.

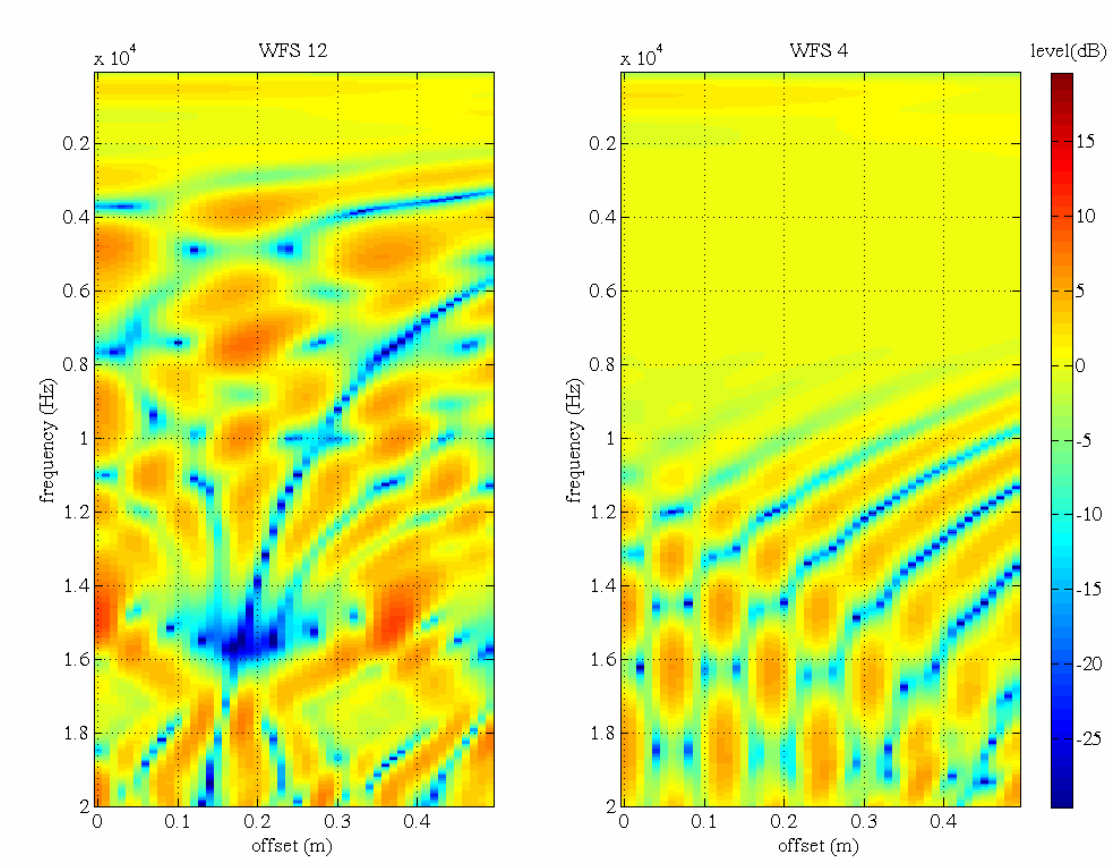
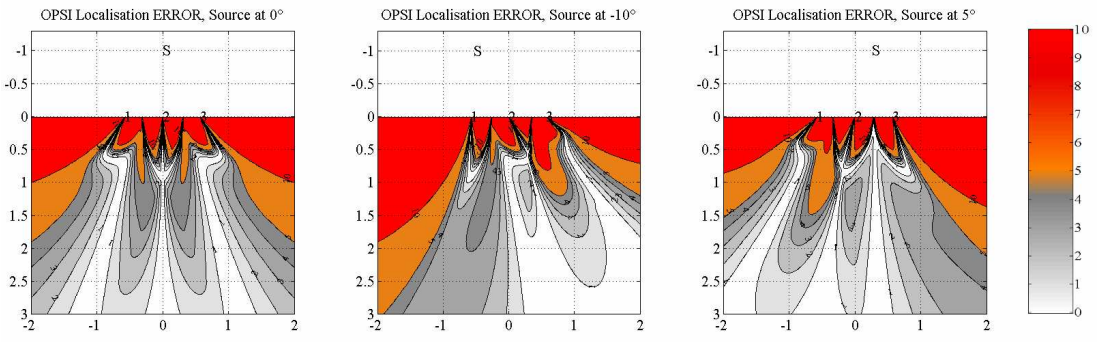


Figure 7-7: Frequency spectra of a WFS virtual source (1 m behind the array) at a line of receiver positions z=2 m. Left diagram: WFS_12; Δx=12.67 cm; f_{alias}=2.5 kHz. Right side: WFS_4; Δx= 4.22 cm; f_{alias}=7.5 kHz.

As the aliasing frequency depends on the secondary source distance, it is different for the two WFS configurations. It can be estimated from Figure 7-7. Simulations and measurements were applied to determine an exact value of the aliasing frequency. For WFS_12, with a virtual source positioned 1 m behind the array, f_{alias} will be 2.5 kHz. WFS_4 results in f_{alias} being 7.5 kHz.

OPSI

The OPSI signals were produced using low-passed WFS_12 signals and three phantom source loudspeakers at spacings of 76 cm, as shown in Figure 7-5. The crossover frequency was f_{cross} = 2 kHz. The level of the phantom source loudspeakers was adjusted by listening to match the level of the low frequency contributions. The OPSI localisation error ε (see chapter 5.4) of the three sources at the listening position was ε = 0° for the source at 0°, ε = 1.9° for the source at -10° and ε = 2.4° for the source at 5°. This can also be seen in the contour plots in Figure 7-8, which show simulations of the OPSI localisation error for the three sources of the experiment.



a) 0° **b) -10°** **c) 5°**

Figure 7-8: Contour plots showing simulations of the OPSI localisation error for the three sources of the experiment. The array is located at y = 0. The position of the virtual source is denoted by the letter 'S' behind the array. A listening area of 4m · 3m is shown, the actual listening position in the experiment was (x;y)=(0;2). The contours show the OPSI localisation error ε = {0, 1, 2, 3, 4, 5, 10}°.

7.5 Results

7.5.1 Screening of the listening panel

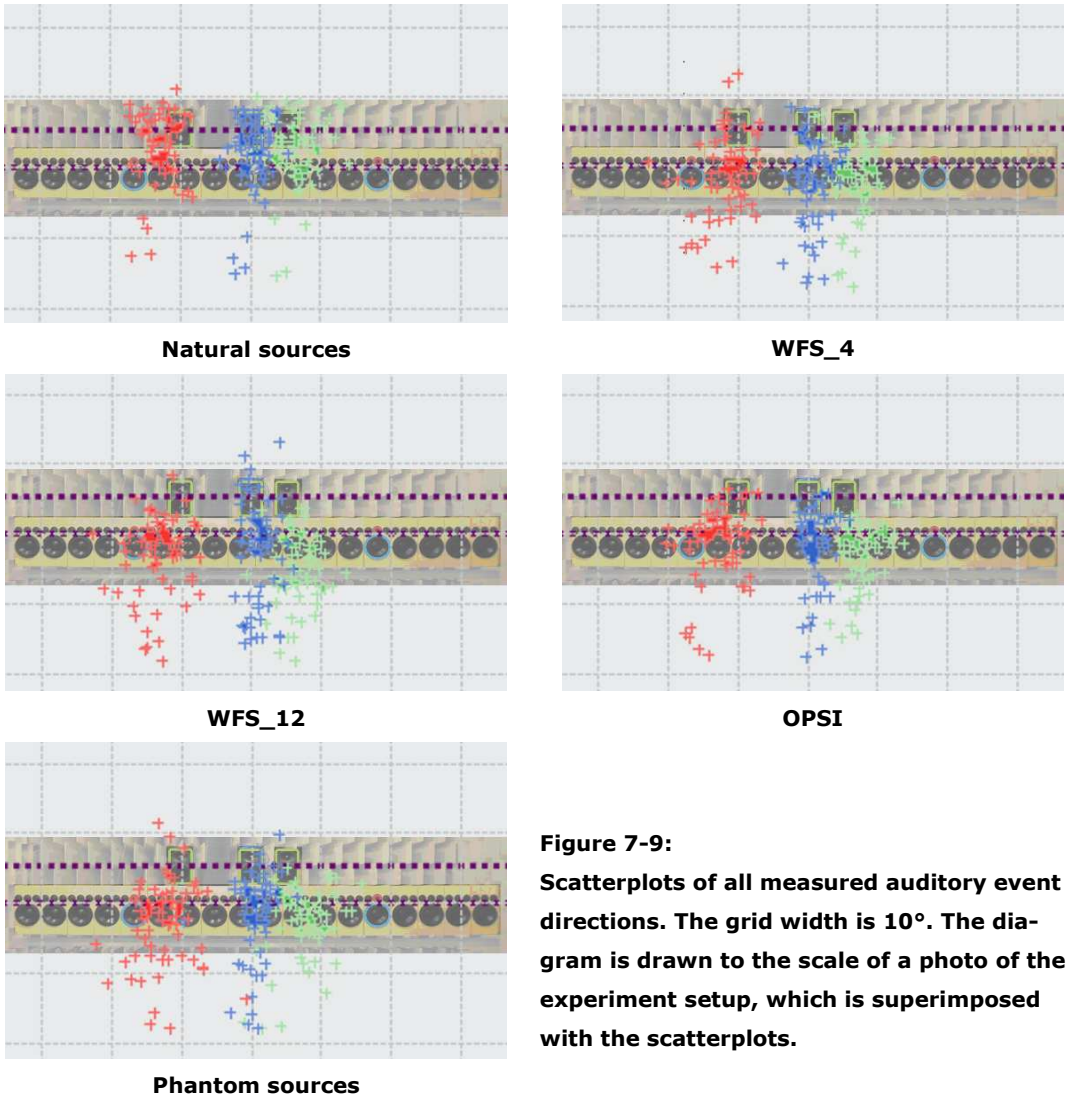


Figure 7-9:
Scatterplots of all measured auditory event directions. The grid width is 10°. The diagram is drawn to the scale of a photo of the experiment setup, which is superimposed with the scatterplots.

Figure 7-9 shows scatterplots of all measured auditory event directions. It can be seen that the subjects' responses were spread considerably for all systems. The task of localising sources in an anechoic environment without visible anchors seems to have been challenging for a number of subjects, some of whom were inexperienced listeners. A post-screening of the obtained results was performed to identify those listeners found to be unable to detect the subtle differences in the source properties. Therefore, the run standard deviations $\bar{s}$ were calculated to measure the consistency of localising repeatedly reproduced test items.

The run standard deviation $\bar{s}$ is the averaged standard deviation of all runs of one subject (for further information on the statistics for localisation experiments see, Hartmann, 1983). It is defined as follows:

Run standard deviation: $\bar{s} = \sqrt{\frac{1}{L}\sum_{k=1}^{L} s^2(k)}$;

where

s = standard deviation of the responses of one subject and one test condition;

L = number of test conditions per subject;

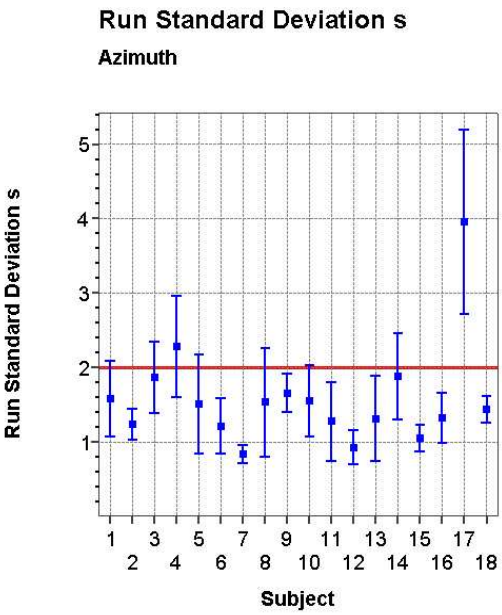


Figure 7-10: Run standard deviation of the azimuth angles averaged over all test conditions. The red line corresponds to the limit for the exclusion of subjects. The mean is shown with the 95% confidence interval.

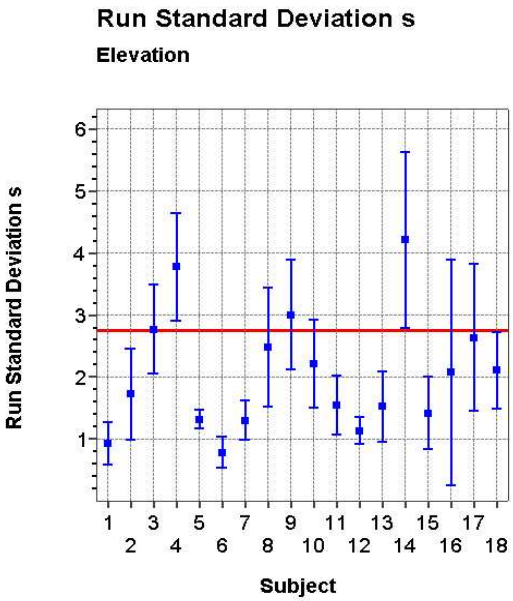


Figure 7-11: Run standard deviation of the elevation angles averaged over all test conditions. The red line corresponds to the limit for the exclusion of subjects. The mean is shown with the 95% confidence interval.

The data of Figure 7-10 and Figure 7-11 show the run standard deviations for each subject in both dimensions of azimuth and elevation, averaged over all stimuli. These data reveal a localisation inconsistency for some subjects. Two limits were chosen such that those subjects who had a run standard deviation above these limits were excluded from the listening panel. These limits (highlighted with red lines in the figures) were defined as a maximum run standard deviation of azimuth angles of 2°, and a maximum run standard deviation of elevation

angles of 2.75°. Based on the inspection of the above figures, subjects 3, 4, 9, 14 and 17 did not fulfil these requirements and were rejected for the further analyses.

7.5.2 Directional accuracy and focus

The directional accuracy, i.e. the ability of a system to reproduce a source in the intended direction can be measured using certain perceived location error values and the standard deviation (see chapter 2.3.2). The signed error E is a suitable measure to estimate the directional accuracy of a system. The mean run signed error $\overline{E}$ measures the averaged error of all runs, directions and subjects.

Run signed error $\overline{E}$: $\overline{E} = \frac{1}{L}\sum_{k=1}^{L} E(k)$;

where

L = number of test conditions per subject;

The mean run signed error $<\overline{E}>$ is defined as the mean of the run signed errors of all subjects.

In Figure 7-12, Figure 7-13, Figure 7-14 and Figure 7-15 one can see the mean azimuth and elevation angles as well as their deviations from the expected reference direction, i.e. the mean run signed error $<\overline{E}>$. The results are arranged by system ('Real' means the single loudspeakers, 'PSQ' means stereo phantom sources) and labelled by the reproduced direction of the test signal ('Reference Direction').

For almost all systems, the azimuth angles were overestimated for the reference directions −10° and 5°. This can be seen from Figure 7-12 and Figure 7-13. There is no identifiable dependence on the system.

The elevation results show that there is a constant bias of roughly −2° to −3° for all systems from the expected elevation angle which was assumed to be at a position between the woofer and the tweeter of the two-way drivers. Indeed, the woofers were positioned roughly 2° below the reference line. The results indicate that the woofer positions were crucial for the perceived elevation angle. Moreover, perception of the elevation under anechoic conditions is known to be difficult, and this could easily have been biased by aspects such as source spectrum or visible anchors. The expected position of the loudspeakers (which were not visible to the subjects) was lower, which could also bias the results.

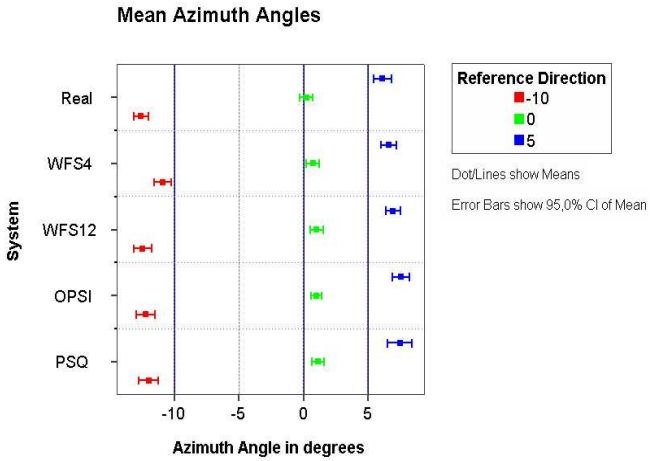


Figure 7-12: Mean azimuth angles

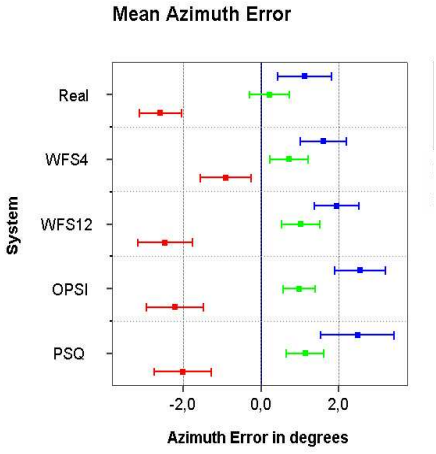


Figure 7-13: Mean run signed error < $\overline{E}$ > of the perceived azimuth angles

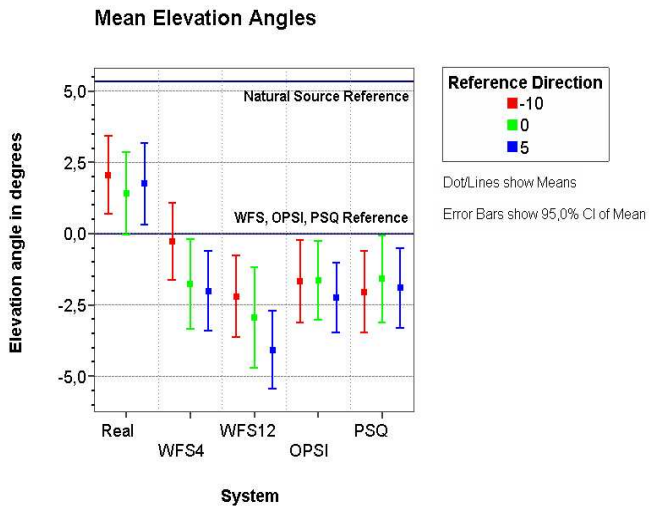


Figure 7-14: Mean elevation angles

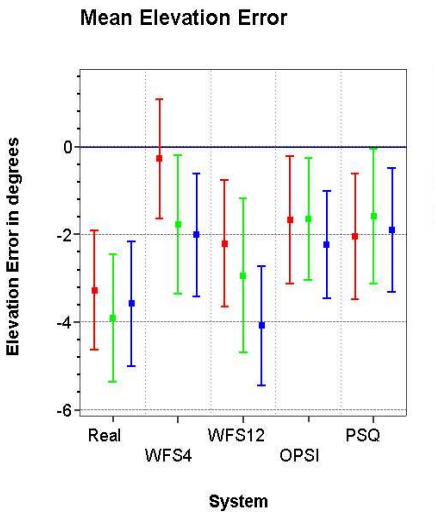
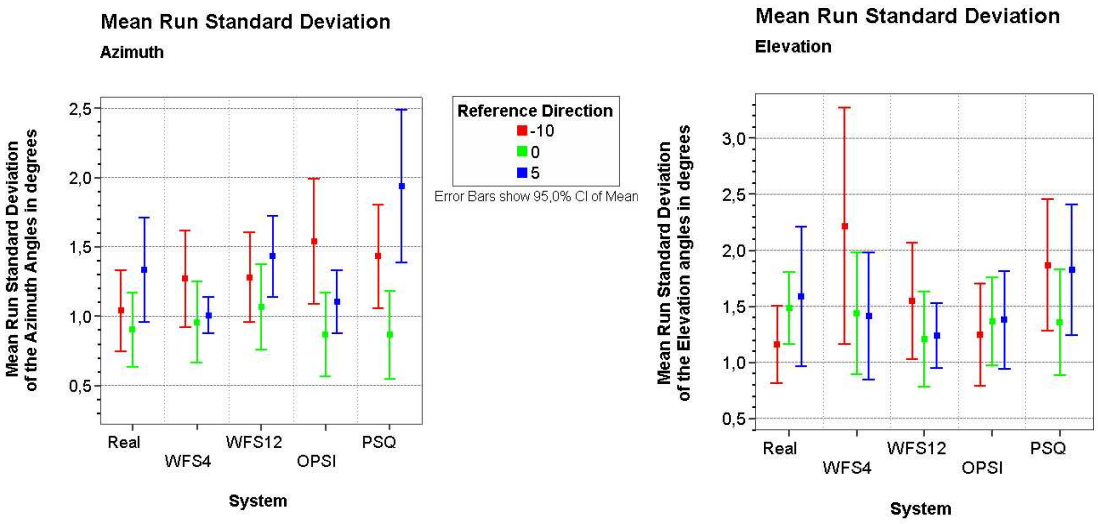


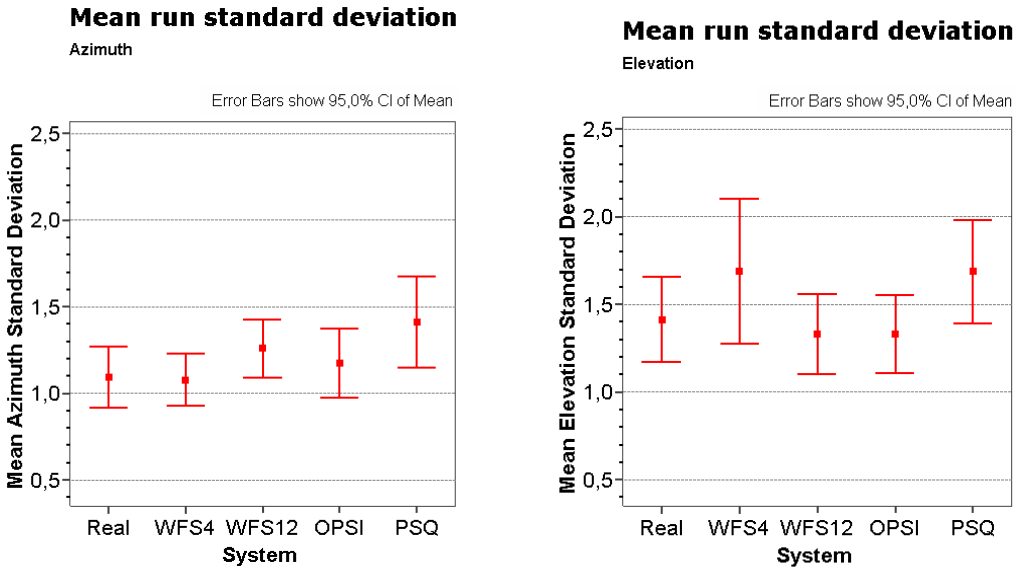
Figure 7-15: Mean run signed error < $\overline{E}$ > of the perceived elevation angles

The confidence intervals indicate that the spread in the measured angles is generally higher for the elevation angles than for the azimuth angles. This can also be seen in Figure 7-16 and Figure 7-17, both of which show the mean run standard deviation < $\overline{s}$ > (mean of the run standard deviations $\overline{s}$ over all subjects) of the azimuth and the elevation angles. The mean run standard deviation is a measure of the intra-subject deviations, and therefore – within the limitations discussed in chapter 2.3.2 – it may also serve as an indicator for the image focus.



a) Azimuth **b) Elevation**

Figure 7-16: Mean run standard deviation $< \overline{s} >$ of the perceived angles for each reference direction. a) azimuth, b) elevation.



a) Azimuth **b) Elevation**

Figure 7-17: Mean run standard deviation $< \overline{s} >$ of the perceived azimuth angles, mean of all directions. a) azimuth, b) elevation.

Through multiple comparison analyses, the differences between the systems with regard to the mean run standard deviation were checked for their significance (Tukey HSD, significance level 95%). The mean run standard deviation $< \overline{s} >$ of the azimuth angles, shown in Figure 7-16a, exhibit no significant differences for the 0° direction. A significant difference (at a significance level above 95%) was only detected for the 5° direction between phantom

sources and the WFS_4 system as well as the OPSI system. For the mean run standard deviation $< \bar{s} >$ of the elevation angles, shown in Figure 7-16b, no significant differences between the different systems were detected.

Figure 7-17 shows the mean run standard deviation $< \bar{s} >$ averaged over all reference directions. The Tukey HSD test again detects no significant differences with regard to the azimuth angle. Only the LSD test detects a significant difference (at a level above 95%) between $< \bar{s} >$ of the phantom sources and that of the real sources as well as the WFS_4 system. No significant differences between the different systems were found in the mean run standard deviations of the elevation angles.

The perceived azimuth and elevation angles are highly dependent on the subject. The mean individual errors, shown in Figure 7-18 and Figure 7-19, indicate that the subjects consistently stay with their individual bias - independently of the reproduction system. A certain individual bias may also influence the overall result, as for example the strong underestimation of elevation angles of subject 8.

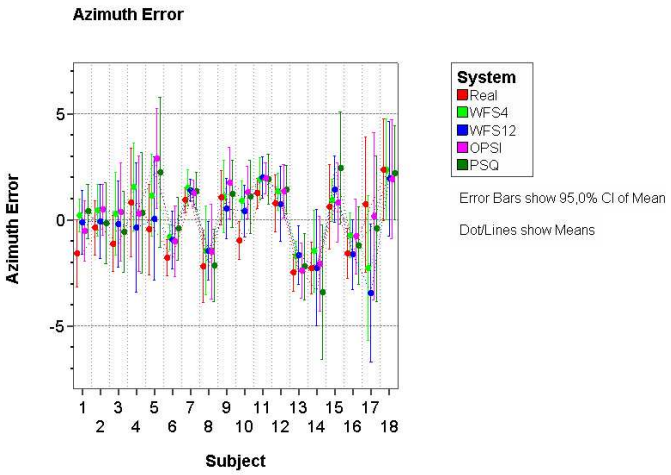


Figure 7-18: Mean individual azimuth angle errors E

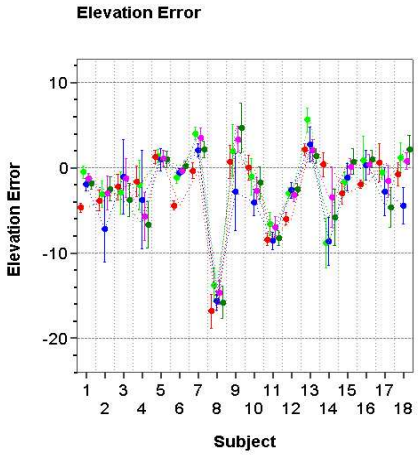


Figure 7-19: Mean individual elevation angle errors E

7.5.3 Locatedness

As discussed in chapter 2.3.2, the locatedness does not necessarily correlate with the measured standard deviation of the perceived directions. Therefore, it needed to be investigated separately. The relevant data is presented in Figure 7-20.

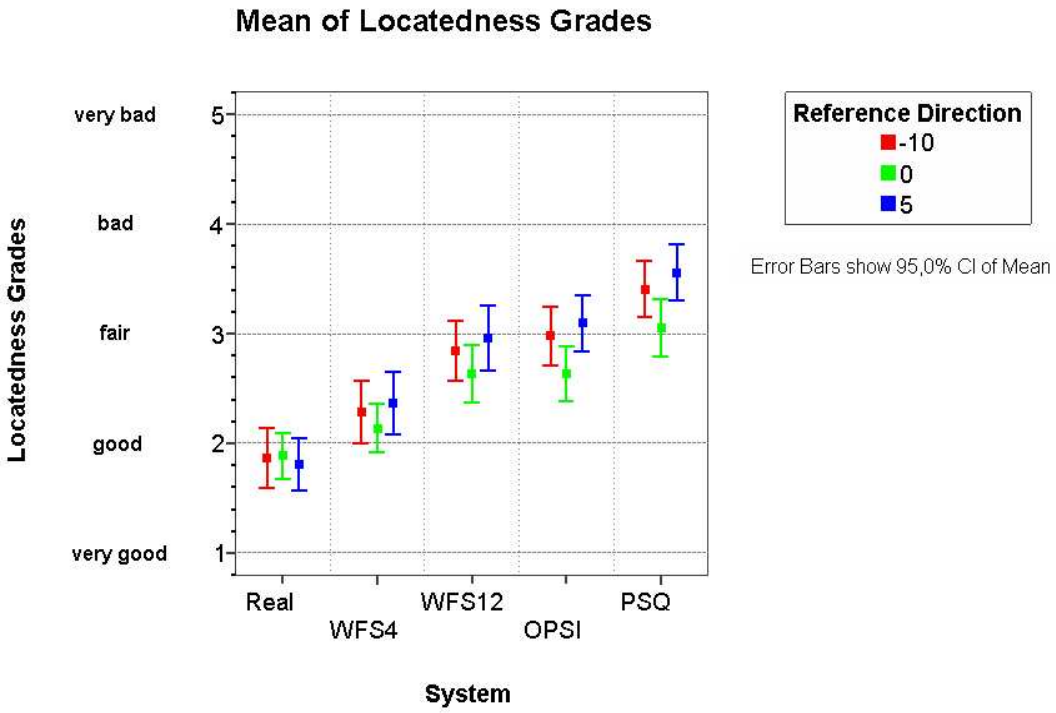


Figure 7-20: Subjective assessment of the locatedness. The mean of all test items of one system and reference direction is shown.

In order to analyse the significance of the differences between the systems with respect to the means of the locatedness, multiple comparison tests were performed. They showed that the differences between any two of the systems are significant (Tukey HSD, at a significance level of 95%), except for the pair WFS_12-OPSI. This applies for the data averaged over all reference directions. The test results are shown in Table 7-1. In the tests using the data of each source direction individually, several differences were not detected as significant.

The observations can be summarised as follows:

1. The locatedness of the natural source (label: 'Real') is best. No system, including the WFS system with a loudspeaker interspacing of 4 cm, can achieve this high grade. The difference between the natural source and WFS_4 is small, but still significant.
2. The locatedness of WFS_4 is significantly better than that of WFS_12.
3. The locatedness of an OPSI source is not worse than that of a normal WFS source with the same WFS array loudspeaker interspacing (WFS_12).
4. The locatedness of the phantom source is worse than that of all other sources. In other words, all WFS systems are better than the conventional phantom source with respect to the locatedness.

5. The results are generally similar for all directions. However, a significant increase of the locatedness compared to the 0° direction can be found for the OPSI source at 5° and the phantom sources at 5° and -10°.

Multiple Comparisons

Dependent Variable: Locatedness

	(I) System	(J) System	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	OPSI	PSQ	-,44*	,106	,000	-,73	-,15
		Real	1,05*	,106	,000	,76	1,34
		WFS12	,09	,106	,916	-,20	,38
		WFS4	,64*	,106	,000	,35	,93
	PSQ	OPSI	,44*	,106	,000	,15	,73
		Real	1,49*	,106	,000	1,20	1,78
		WFS12	,53*	,106	,000	,24	,82
		WFS4	1,08*	,106	,000	,79	1,37
	Real	OPSI	-1,05*	,106	,000	-1,34	-,76
		PSQ	-1,49*	,106	,000	-1,78	-1,20
		WFS12	-,96*	,106	,000	-1,25	-,67
		WFS4	-,41*	,106	,001	-,70	-,12
	WFS12	OPSI	-,09	,106	,916	-,38	,20
		PSQ	-,53*	,106	,000	-,82	-,24
		Real	,96*	,106	,000	,67	1,25
		WFS4	,55*	,106	,000	,26	,84
	WFS4	OPSI	-,64*	,106	,000	-,93	-,35
		PSQ	-1,08*	,106	,000	-1,37	-,79
		Real	,41*	,106	,001	,12	,70
		WFS12	-,55*	,106	,000	-,84	-,26
LSD	OPSI	PSQ	-,44*	,106	,000	-,64	-,23
		Real	1,05*	,106	,000	,84	1,26
		WFS12	,09	,106	,398	-,12	,30
		WFS4	,64*	,106	,000	,43	,85
	PSQ	OPSI	,44*	,106	,000	,23	,64
		Real	1,49*	,106	,000	1,28	1,70
		WFS12	,53*	,106	,000	,32	,73
		WFS4	1,08*	,106	,000	,87	1,29
	Real	OPSI	-1,05*	,106	,000	-1,26	-,84
		PSQ	-1,49*	,106	,000	-1,70	-1,28
		WFS12	-,96*	,106	,000	-1,17	-,75
		WFS4	-,41*	,106	,000	-,62	-,20
	WFS12	OPSI	-,09	,106	,398	-,30	,12
		PSQ	-,53*	,106	,000	-,73	-,32
		Real	,96*	,106	,000	,75	1,17
		WFS4	,55*	,106	,000	,34	,76
	WFS4	OPSI	-,64*	,106	,000	-,85	-,43
		PSQ	-1,08*	,106	,000	-1,29	-,87
		Real	,41*	,106	,000	,20	,62
		WFS12	-,55*	,106	,000	-,76	-,34

Based on observed means.

*. The mean difference is significant at the ,05 level.

Table 7-1: SPSS chart showing the results of multiple comparison tests (Tukey HSD, LSD) performed on the locatedness grades averaged over all source directions. The mean difference for each pair is shown as well as its standard error, the level of significance for that difference and the 95% confidence interval. Note the level of significance corresponds to the probability with which the means are equal. Thus the level of significance for the means to be different is 1-Sig.

7.6 Discussion

7.6.1 Directional accuracy and focus

The measurement of the perceived directions leads to several conclusions regarding the imaging characteristics of the different systems.

All systems can be assumed to be accurate regarding the creation of a certain source direction. In the experiment, mean azimuth errors $< \overline{E} >$ between –3° and +3° were measured. For calculating *E*, the synthesised direction was used as a reference. However, when the perceived direction of the real source is taken as a reference, the azimuth errors were in general substantially smaller. Hence, a systematic deviation in the measurements can be assumed as a reason for the high errors instead of a random inaccuracy. The standard deviations were comparable to other investigations (Blauert, 1997). The mean elevation error showed that the woofers, which were located below the reference line, determined the perceived direction. The mean elevation errors, corrected by this constant shift, were rather small and revealed no system-specific inaccuracy.

The mean run standard deviation $< \overline{s} >$ of the perceived directions may, with limitations (chapter 2.3.2), serve as an indicator for the image focus of the source. The described results show that only for the phantom sources a larger image focus may be concluded in that case. It should be noted that the offset angle of the stereo setup was rather small (31.8° compared to 60° in standard stereo) which led to higher necessary interchannel differences for the shift of the phantom source ($\Delta L/\Delta t$ = 4.3 dB/–0.25 ms for the –10° source and –2.15 dB/0.125 ms for the 5° source). Thus, a larger focus of the source could be caused by the increased time differences (chapter 3.5.2).

For all other measured sources, no significant differences were detected in the standard deviations. For the OPSI source at –10°, a slightly but not significantly larger standard deviation compared to the real source was measured. This could lead to the conclusion that the focus of the OPSI image depends on the source direction. A general dependence of localisation quality on the OPSI localisation error is plausible. However, calculations show that the OPSI localisation errors for all sources in the experiment were below 2.5°. This is also apparent from Figure 7-8, which shows simulations of the OPSI localisation error for the sources of the experiment. Apart from the OPSI localisation error, the localisation focus of the individual phantom source is likely to influence the localisation performance of the merged OPSI source. It may be assumed that the phantom source derived from the source at –10° has a larger focus due to the significant contribution of two stereo loudspeakers compared to the 0°

phantom source. The latter is produced primarily by the middle stereo speaker, which is perceived with a substantial localisation dominance, i.e. with a higher level and earlier than the other two. Thus, it may be concluded that the focus of OPSI sources is dependent on the localisation performance of the individual phantom source contribution.

7.6.2 Locatedness

The locatedness data show significant differences between the reproduction systems. These differences were not obtained by measurements of the standard deviation of the perceived directions. The results identified the real sources (the reference) as having the best grades. The significant difference between WFS_4 and WFS_12 is evidence for the positive effect of further increasing the spatial aliasing frequency from 2.5 kHz to 7.5 kHz. The OPSI sources were localised with the same locatedness compared to the corresponding WFS_12 system. It may be hypothesised that the locatedness of an OPSI source is in general equal to its underlying WFS system. This would mean that by using an OPSI system with a smaller WFS array spacing (and a corresponding increased crossover frequency), the localisation performance could be improved. The experiment on sound colour perception described in chapter 8 shows relevant results on the optimal array spacing and crossover frequency. The results show that the OPSI concept could be validated, because it did not lead to worse results in comparison with the corresponding WFS system.

The results of the locatedness obtained for the phantom source focus confirmed the conclusions derived from the standard deviation data mentioned above. They were localised with the least locatedness.

Both the locatedness data and the standard deviations of the perceived azimuth directions above seem to show a difference between the results for the 0° direction and the other two directions. This difference is significant for the phantom sources and one OPSI source, and can be explained by the reasons mentioned above. Only the significant difference in $< \bar{s} >$ of the real sources at 0° and 5° cannot be explained in that way. An inhomogeneous reproduction room may be a possible reason.

There is a significant difference between the locatedness of the natural sources and the WFS virtual sources. However, this investigation was not designed to find the reason behind this, but rather to detect it. There are a number of possible reasons which can be found by considering the remaining differences between natural listening and listening to a virtual source. Among them are: spatial aliasing, diffraction effects, distance perception conflict (the source distance is not correctly reproduced by a dry virtual source, see chapter 9).

The differences between natural and WFS sources are not expected to vanish in a real room, because the reflections of the reproduction room may even facilitate the localisation of the natural source. On the contrary, the reflections of the WFS array can theoretically even disturb the localisation of the virtual source due to the fact that they are not corresponding to the reflection pattern of a natural source (see chapter 4.2.8).

7.7 Summary of chapter 7

Following the experiments of Start (1997) and Verheijen (1998), an investigation into the localisation performance of WFS together with a comparison of natural sources, OPSI and phantom sources was performed.

A classification of the WFS virtual source's localisation performance in comparison to the other reproduction techniques could be made. The experiment's results show that the locatedness of a WFS virtual source is significantly better than that of a phantom source. However, results as good as those found when listening to real sources can not be achieved.

Increasing the spatial aliasing frequency from 2.5 kHz to 7.5 kHz significantly improves the locatedness of a virtual source in WFS. This difference is not evident when only the standard deviations of the measured auditory event directions are considered.

The OPSI concept proved to be perceptually valid with respect to localisation. In comparison to WFS no significant quality degradations were measured.

The localisation focus, assumed to be derived from the standard deviations of the perceived directions, of the WFS sources did not show an identifiable difference to that of the natural sources. Only for the phantom sources were significantly greater standard deviations found.

8. Experiment 2: Sound colour properties of WFS, OPSI and Stereo

8.1 Introduction

The investigation described in this chapter focuses on the sound colour reproduction capabilities of the systems WFS, stereo and OPSI. The research questions were developed in chapter 6.3. The aims of the experiment were a comparison of the different reproduction systems regarding their colouration and an evaluation of the physical causes of the colouration.

An experiment was performed using a special test method that targeted the determination of sound colour differences within one system. The reproduction systems were acoustically simulated by a virtual acoustic system using head-tracking. These simulations also enabled a theoretical comparison of the spectral alterations of the ear signals and the subjective colouration grades. A prediction of the colouration was attempted based on these spectral alterations. The prediction examined the relationship between the physical and perceptual measures and thus aimed to reveal general differences between the reproduction techniques. Furthermore, the relevance of spatial aliasing for sound colour perception in WFS together with the perceptibility of comb filtering in stereo could be observed. By an incorporation of the OPSI method, the hypotheses regarding the different perception mechanisms were investigated.

The experiment is described by its setup in section 8.2, its results in section 8.3, and a discussion of the results in section 8.4. An analysis of objective physical measures is performed in section 8.5. Based on this analysis, a prediction of the colouration perception could be performed. This is described in section 8.6 and discussed in section 8.7. Finally, section 8.8 summarises the chapter.

8.2 Experimental setup

8.2.1 Colouration

In this investigation, the sound colour reproduction capabilities of different sound reproduction techniques are under test. It is known that the auditory system is able to adapt quickly to the static frequency response of the transmitting system which is then regarded as not coloured (Zwicker and Fastl, 1990). Therefore, the absolute sound colour of an individual test system, which could be caused by the loudspeaker type, the reproduction room or various

other parameters, is of no interest in this investigation. Good sound colour reproduction is not necessarily dependent on the absolute sound colour.

These postulations led to a method of exploring the system colouration by a measurement of the perceivable sound colour differences between different sources within the same system (=intra-system colouration). This measurement is believed to correspond to the actually perceived colouration during the reproduction as soon as more than one source is reproduced or the source moves or the listener moves. The absolute sound colour difference between the virtual source and a reference source (‘real source’) was not considered relevant. This difference was comparably large due to system-inherent spectral properties, which are not considered decisive for the timbral fidelity of the system. For example, the pronounced reproduction of lower frequencies which is known in stereo does probably not give rise to a degraded sound colour reproduction (Theile, 1980).

For this investigation, an extended definition of the attribute colouration will be used based on the definition by Salomons (1995), which is given in chapter 2.4.

Definition of the attribute colouration for this investigation:

The perceived colouration of a system is the audible distortion, which alters the sound colour when switching between different source or receiver positions.

In contrast to the attribute sound colour, colouration can be said to either exist, or not to exist. This means, compared with the exploration of the sound colour, a much simpler experiment paradigm can be used which only aims to detect the existence of any colouration and to quantify the degree of colouration.

The differences between the colourations appearing in the experiment were expected to be subtle. In spite of this, these differences could lead to fundamental discoveries regarding perception properties and aliasing perception. Hence, a method had to be found that was capable of detecting a small colouration in a reliable and reproducible way.

8.2.2 Test method

For the sound colour experiment, a modified MUSHRA method was used. The MUSHRA method (multi stimulus test with hidden reference and anchor, ITU-R BS.1534) is known to give comparable and reproducible results, as the same anchors are used in each trial. With this procedure, the grades of the different trials are made comparable, and the influence of time-variant aspects such as the order of the stimuli or listener fatigue is avoided. Furthermore, the anchors span a recurrent scale which defines the range of possible differences of the experi-

ment and thus optimises the subjects' use of the scale. For the sound colour experiment, this aspect necessitates anchors that provide a good span of possible colouration. The colouration is largely due to spatial aliasing, meaning that spectral inadequacies in the upper frequency range are the main cause (see chapter 4.2.5). Consequently, the relevant anchor has to be of a similar nature, with different degrees of spectral deviation.

In pilot experiments a suitable anchor was found, this being the reference signal, processed with sine-ripple spectra of different ripple depths (for a discussion of sine-ripple spectra see Supin et al., 1999). The ripple depth (amplitude of the sine) is defined as the difference between the maximum of the first half wave and zero. Five anchors were used; the ripple depth was 0, 1, 2, 3 and 4 dB. The utilised anchors are shown in Figure 8-1:

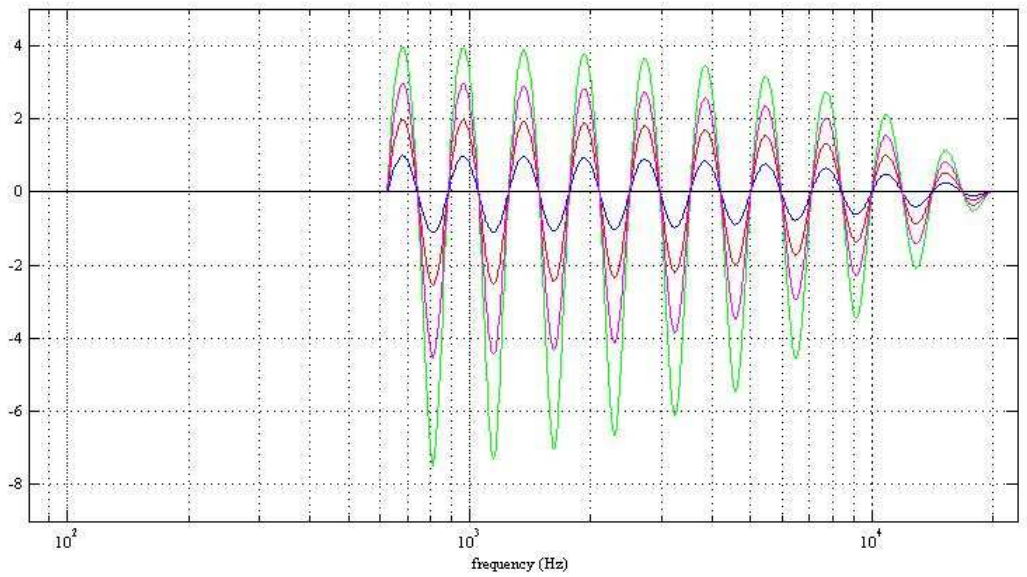


Figure 8-1: Anchors of experiment 2: sine-ripple spectra from 625 to 20.000 Hz. The ripple density is 2 (ripples per octave). The ripple depth is 0, 1, 2, 3 and 4 dB. The anchors were designed in order to sound not dissimilar to spatial aliasing.

As described, it was decided to assess the colouration, which was defined as an intra-system parameter. This means that only one system has to be assessed in each trial, which makes the experiment design easier. In each trial 9 different signals were reproduced. These were:

- the reference (direction -5°)
- three stimuli in other directions than the reference (-10°, 3°, 15°). The directions were chosen such that the differences between reference and stimulus direction were unequal for all stimuli.
- the hidden reference
- four anchors (see Figure 8-1)

Figure 8-2 shows the multiple stimulus graphical user interface of the experiment. The subjects could switch between the 9 different stimuli by clicking on the buttons 'REF' or 'A'-'H'. A click in the grey area switched between the last two chosen stimuli, so that the subjects were not distracted looking for the relevant buttons. The order in which the 9 stimuli were arranged was chosen randomly by the MATLAB control script. The trial could not be finished before all stimuli had been replayed once. Only one system was reproduced in each trial. The order in which the different reproduction systems were assessed was chosen randomly by the script.

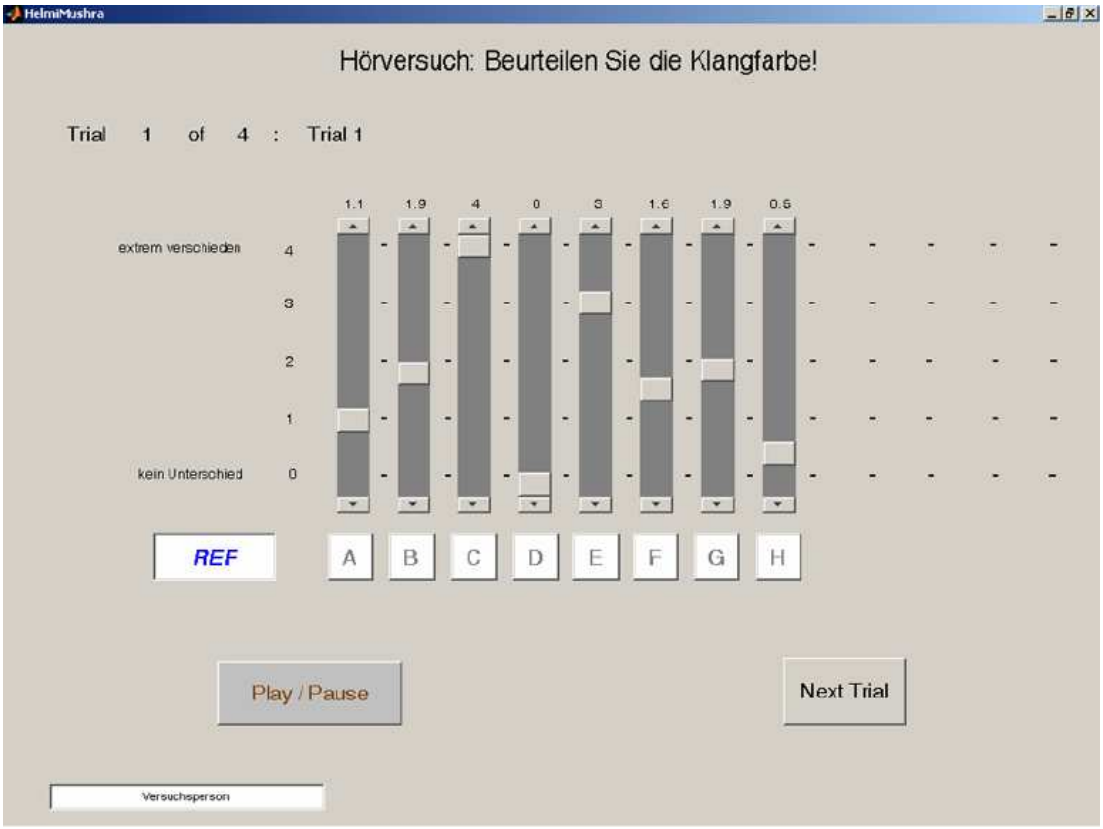


Figure 8-2: Screenshot of the multiple stimulus graphical user interface display of the experiment. The software was programmed in MATLAB. The software recorded the grades and controlled the outputs of the test PC which was connected with the audio PC running Nuendo and the BRS auralisation plug-in. The German labels denote the task of the experiment ('Hörversuch: Beurteilen Sie die Klangfarbe' means 'Listening test: Assess the sound colour') and the two grade descriptions ('extrem verschieden' means 'extremely different' and 'kein Unterschied' means 'no difference').

8.2.3 Virtual acoustics system: BRS

For practical reasons, a virtual reproduction based on a binaural system utilising headphones and head-tracking was used. This system, known as BRS (Binaural Room Synthesis, see e.g. Horbach et al., 1998, 1999; Rathbone et al., 2000) was developed at the IRT (Institut für Rundfunktechnik, Munich) and was realised as a VST plug-in (IRT, 2007) that can be run within the host software Steinberg Nuendo. The test PC, running the MATLAB-based multiple stimulus GUI, sent audio signals to the audio PC, running Nuendo and the BRS plug-in, see Figure 8-3.

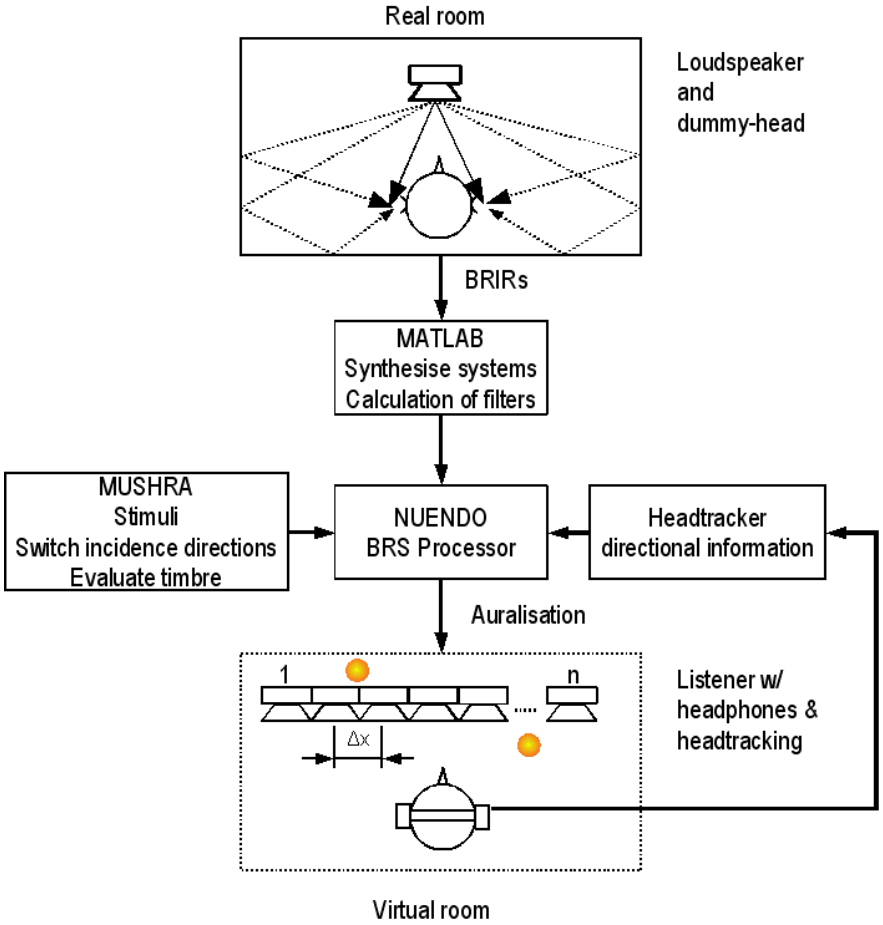


Figure 8-3: Experimental system architecture, diagram from Hanselmann (2006)

The BRS filters were changed after each trial in order to assess the next reproduction system.

BRS is a convolution-based system, i.e. it needs binaural measurements of BRIRs (Binaural Room Impulse Responses) to correctly reproduce the virtual sources. The filters used in the BRS system were produced by utilising a database of BRIRs measured in the listening room of the IRT. By using this database, a natural BRIR of an arbitrary system consisting of single loudspeakers in different directions can be produced. The resulting BRIRs are produced by

superimposing the BRIRs of the single loudspeakers of the system according to the driving function of the system. A larger distance of one individual loudspeaker is simulated simply by a time delay and an according decrease of the level of the BRIR. A disadvantage of this procedure is that the reverb level is erroneously decreased. Hence, systems containing large differences between the distances of the individual loudspeakers would not be auralised correctly. Furthermore, all individual loudspeakers are aligned such that their 0°-axes are aiming the listener. This is a consequence of the measurement procedure of the database. In this procedure, the dummy head was installed on a turntable and the measurements for the different directions were derived by rotating the dummy head (and not the loudspeaker). Both compromises have no effect on the direct sound except for the assumption that the individual loudspeakers are ideal monopoles. The compromises are not expected to be relevant for this experiment.

Hence, the resulting BRIRs may be called quasi-natural, because they do not exactly equal BRIRs that would have been measured from the existing system, but do contain all relevant acoustical information without a meaningful compromise. The BRS system produces out-of-head-localisation, which is a prerequisite of any meaningful experiment about spatial and timbral attributes. Only successful out-of-head-localisation avoids a system-inherent colouration of the binaural reproduction (Rathbone et al., 2000). This is a very important aspect, because the timbral fidelity of the test system itself has to be better than the best source in order to create correct grades. The results described in section 8.3 show that this is difficult to achieve. In reality, some system-inherent colouration remains even for head-tracked binaural systems, which is presumably due to the use of non-individualised HRTFs or tracking errors (Rathbone et al., 2000). Nevertheless, the system is considered suitable for the task of the experiment because the timbral differences appearing in the experiment are mostly larger than the assumed timbral differences due to the BRS system. Furthermore, the negative influence of the BRS system is equal for all systems of the experiment and should therefore not bias the differences between the measured colourations.

A pilot test of this experiment was repeated on a real system to prove the applicability of the BRS system by one subject, see Figure 8-4. The two different stereophonic techniques, level-panning and time-panning, were assessed for their colouration. The results show that the mean assessed coloration is slightly bigger and the variation is higher for the BRS system. It cannot be proven but it is assumed that this deviation from real reproduction does not change the colouration differences between the reproduction systems.

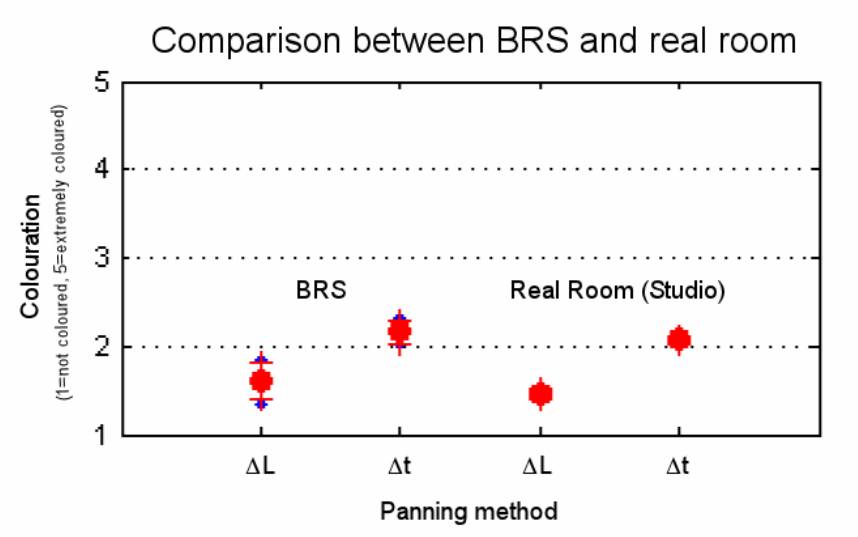


Figure 8-4: Pilot test, diagram from Wegmann (2005): Validation test of the virtual acoustic system BRS. The colouration grade is assessed for the two stereo systems ΔL (level panning) and Δt (time panning), in both the virtual BRS environment and the real room environment. Mean and 95% confidence intervals are shown for one subject and four trials. The single blue dots denote individual results.

8.2.4 Stimuli

The stimuli of the experiment were dry pink noise bursts of length 800 ms, with a fade-in and fade-out time each of 50 ms (see also chapter 7.3.2). These were regarded to be most sensitive to changes in the sound colour. The noise burst was repeated after a break of 500 ms until the 'Pause'-button was pressed. The parameters of the noise bursts were chosen after a validation in preliminary tests. The multiple stimulus GUI software was programmed such that the stimulus was always completely reproduced without changes. Pressing the buttons 'REF' and 'A' – 'H' or the grey area (see Figure 8-2) led to a switch between sources/anchors during the subsequent break between the stimuli.

The dry stimuli were convolved in real-time with the corresponding BRIRs of the current system and azimuth to result in the binaural stimuli assessed by the subject.

8.2.5 Test procedure

The experiment took place in the same room as the measurement of the BRIR database, this being the ITU-R BS.1116 compliant listening room of the IRT. The subjects were sitting such that the visual scenery matched as closely as possible the acoustic scenery produced by the virtual acoustics system. This means that they were positioned in front of the real WFS test array which was built of 32 small broadband speakers. This array was not active in the ex-

periment, but it served as a visual anchor to support the acoustic illusion. In this author's experience it is mandatory for any virtual acoustic system to provide this visual anchor (even if it was perfect), because the visual cues are important for distance perception. A successful perception of source distance was hypothesised to be crucial for sound colour perception (see chapter 3.6.2).

The subjects were asked to grade the perceived colouration of the stimuli in the multiple stimulus GUI window (Figure 8-2) on a 5-grade scale. Only the extremes of the scale were labelled with verbal descriptions:

Is there a timbral difference between the reference and the stimulus?

0='no difference' (German: 'kein Unterschied')
4='extremely different' (German: 'extrem verschieden')

The attribute 'colouration/timbral difference' was defined as the perceived sound colour difference between the reference and the chosen stimulus. A small training phase introducing possible colourations was performed before the experiment. Furthermore, the first two trials were not used in the data, as the subjects needed some time to accommodate to the experimental conditions.

The test was split into a minimum of three sessions in order to avoid listener fatigue. The subjects were allowed to interrupt the test and continue it at a later time.

8.2.6 Systems under test

Table 8-1 lists the systems assessed in the experiment together with the relevant system parameters. These are the name of the system, the number of utilised loudspeakers (in order to estimate its complexity), the spatial aliasing frequency f_{alias}, the OPSI crossover frequency f_{cross} and the depiction of the colour code in the table. The colour code is explained in Table 8-2 below.

The length of the array was 6 m. The OPSI systems utilised phantom source speakers which were spaced by 168 cm. Thus, in total 4 phantom source speakers were used. They were placed symmetrically to the middle of the array. The listening position was at a distance of 1.5 m from the middle of the array. In other words, the BRS system synthesised the systems at this distance.

System	Spacing of loudspeakers	No. of loudspeakers	f_{alias} [Hz]	f_{cross} [Hz]	$f_{alias} > f_{cross}$
Real Sources	-	1	-	- (24000)	
Stereo	173 cm	2	-	- (0)	
OPSI_3	3 cm	200	9600	750	↑
OPSI_3	3 cm	200	9600	1500	↑
OPSI_3	3 cm	200	9600	3000	↑
OPSI_3	3 cm	200	9600	6000	→
WFS_3	3 cm	200	9600	- (24000)	↓
OPSI_12	12 cm	50	2400	750	↑
OPSI_12	12 cm	50	2400	1500	→
OPSI_12	12 cm	50	2400	3000	↓
OPSI_12	12 cm	50	2400	6000	↓
WFS_12	12 cm	50	2400	- (24000)	↓
OPSI_24	24 cm	24	1200	750	→
OPSI_24	24 cm	24	1200	1500	↓
OPSI_24	24 cm	24	1200	3000	↓
OPSI_24	24 cm	24	1200	6000	↓
WFS_24	24 cm	24	1200	- (24000)	↓
WFS_48	48 cm	12	600	- (24000)	↓

Table 8-1: Systems under test. The WFS systems are labelled by the loudspeaker spacing which is 3, 12, 24 or 48 cm. Systems containing significant aliasing are marked red (arrow downwards). Systems with unnecessarily low crossover frequency are marked blue (arrow upwards). Systems containing no aliasing and an as high as possible crossover frequency are marked green (arrow right).

$f_{alias} > f_{cross}$; the crossover frequency is unnecessarily low	↑
$f_{alias} \approx f_{cross}$; the crossover frequency is optimal	→
$f_{alias} < f_{cross}$; the crossover frequency is too high, there is aliasing in the signal	↓

Table 8-2: Three categories of OPSI systems, the same colour code is used in Table 8-1, Figure 8-8 and Figure 8-14.

The frequency responses of the OPSI systems at two positions according to the ear positions of the subjects are listed in Figure 8-8. There, the different contributions in the OPSI signals are shown separately: the frequency response of the WFS signal, its contribution to the OPSI signal and the frequency response of the high-passed phantom source response. Using these figures, the OPSI systems can be classified into three categories, marked in the same colours as in Table 8-1. This classification is described in Table 8-2.

8.3 Results

Figure 8-5 shows the colouration grades of the four anchors and the hidden reference. It can be confirmed that the anchors span the whole scale of colouration grades. The results of the four anchors are similar for all systems, which confirms that the anchors do indeed act as a recurrent scale that makes the results comparable.

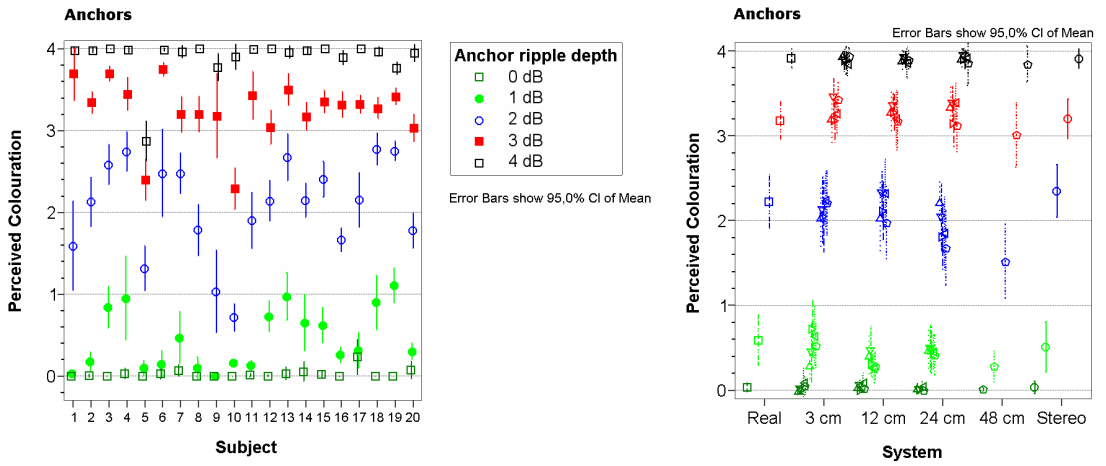


Figure 8-5: Results of the experiment: Colouration grades of the four anchors and the hidden reference. The mean of all trials is shown for each subject (left diagram) and each system (right diagram).

Figure 8-6 and Figure 8-7 show the results of the experiment. In Figure 8-6 the perceived colouration is shown for each system and OPSI crossover frequency f_{cross}. Figure 8-7 sorts the results of the OPSI and WFS sources by f_{cross}. In these figures the labels of the WFS and OPSI systems describe the spacing and the crossover frequency f_{cross}, as shown in Table 8-2.

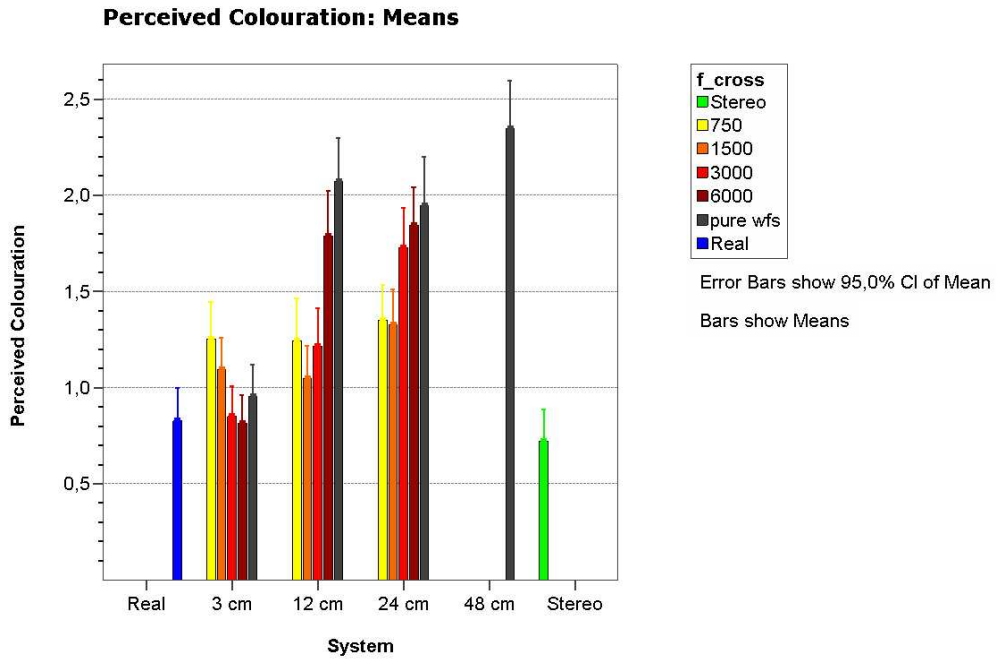


Figure 8-6: Results of the experiment: the perceived colouration is shown for all systems of the test. The category is the OPSI crossover frequency in Hz.

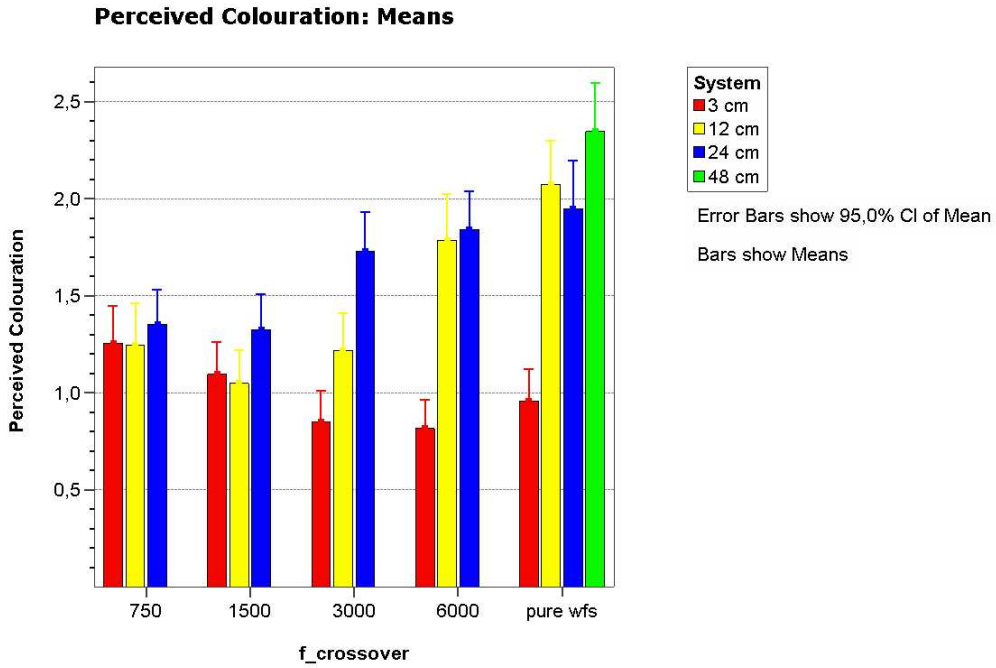


Figure 8-7: Results of the experiment: the perceived colouration is shown against the OPSI crossover frequency in Hz. The category is the WFS loudspeaker spacing in cm.

The following tables show whether the measured differences are significant. A significant difference due to the LSD test at a significance level of 95% is denoted by an asterisk (*).

a) WFS_3:

f_{cross}	750	1500	3000	6000
1500				
3000	*	*		
6000	*	*		
pure WFS	*			

b) WFS_12:

f_{cross}	750	1500	3000	6000
1500				
3000				
6000	*	*	*	
pure WFS	*	*	*	

c) WFS_24:

f_{cross}	750	1500	3000	6000
1500				
3000	*	*		
6000	*	*		
pure WFS	*	*		

d) f_{cross} = 750 Hz:

System	WFS_3	WFS_12
WFS_12		
WFS_24		

e) f_{cross} = 1500 Hz:

System	WFS_3	WFS_12
WFS_12		
WFS_24		*

f) f_{cross} = 3000 Hz:

System	WFS_3	WFS_12
WFS_12	*	
WFS_24	*	*

g) f_{cross} = 6000 Hz:

System	WFS_3	WFS_12
WFS_12	*	
WFS_24	*	

h) Pure WFS:

System	WFS_3	WFS_12	WFS_24
WFS_12	*		
WFS_24	*		
WFS_48	*		*

Tables 8-3: Significance tests of the differences in Figure 8-6 and Figure 8-7. A significant difference due to the LSD test at a significance level of 95% is denoted by an asterisk (*).

The main results of the experiment can be depicted as follows:

1. The reference, i.e. the real source does not have an optimal colouration grade (=0).
2. The real source and the phantom source, as well as the optimal WFS_3/OPSI_3 systems have the best colouration grades.
3. The perceived colourations of the WFS and OPSI systems generally increase with increasing array loudspeaker spacings.
4. An OPSI system can significantly reduce the perceived colouration in comparison to the corresponding WFS system.
5. The more the aliasing frequency exceeds the crossover frequency, the larger the colouration. In other words: the colourations increase with the amount of aliasing in the signals.
6. When the crossover frequency is reduced below 3000 Hz, the colourations also increase.
7. The achieved results are plausible, can be sufficiently explained and thus are considered to be reliable.

8.4 Discussion of the subjective results

The discussion is organised by the list of observations above.

Observation 1:
The reference, i.e. the real source does not have an optimal colouration grade (=0).

The least noticeable colouration is achieved with the reference sources of the experiment. The chosen HRTFs of the different reference sources are taken from the same database, they are congruent except for a rotation that corresponds to the difference in sound incidence angles. However, the grade for the reference source is not 0, which would mean that no colouration had been perceived. There are several reasons for the remaining colouration:

Firstly, no perfect real conditions exist in the experiment. The difference between the experimental system and the acoustic reality is primarily in the HRTFs, which were non-individual. Other possible parameters leading to a non-perfect function of the virtual acoustics system may be latency, headphone calibration and the absence of a visual support for the acoustically perceived sources (Rathbone et al., 2000).

Secondly, there is an inherent problem with this type of measuring procedure: the change of position of the virtual source not only leads to a difference in sound colour, but also to a noticeable difference in localisation attributes. The source direction or distance, (or both) change, which might influence the sound colour perception of the listener. The general postulation that none other than the parameter under test should change in an experimental setup cannot be fulfilled. Under ideal conditions this change of localisation attributes can be perceived as such, and it should be possible to differentiate between these and the colouration attribute.

Thirdly, it cannot be generally assumed that real sources at different locations sound the same. The spatial decoding process associates inverse HRTF filtering (Theile, 1980), this however does not imply that colouration due to the position-dependent effect of the HRTF can be fully avoided. This is particularly true if the localisation stimulus discrimination may be impaired because of imperfect rendering tools (e.g. no individual HRTFs).

Observation 2:

The real source and the phantom source, as well as the optimal WFS_3/OPSI_3 systems have the best colouration grades.

As expected, the real sources and WFS_3 sources have the best colouration grades. The spatial aliasing in the WFS_3 system leads to a small degradation in its sound colour performance, whereas the OPSI_3 with crossover frequencies of 3000 Hz and 6000 Hz have slightly better grades. It may also be surprising that the phantom sources (red bar) achieve the same optimal grades.

Observation 3:

The perceived colourations of the WFS and OPSI systems generally increase with increasing loudspeaker distances.

The sound colour performance deteriorates with increasing WFS loudspeaker distance and thus with an increased aliasing in the signal. The experimental results show that only the WFS_24 system achieves slightly better grades than the WFS_12 systems, which cannot be explained at this point. The objective analysis of the next sections leads to a possible explanation of this result.

Observation 4, 5 and 6:

An OPSI system can significantly reduce the perceived colouration in comparison to the corresponding WFS system.

The more the aliasing frequency exceeds the crossover frequency, the larger the colouration. In other words: the colourations increase with the amount of aliasing in the signals.

When the crossover frequency is reduced below 3000 Hz, the colourations also increase.

Figure 8-8 shows an overview of the used systems. The green, blue and red frames illustrate how good the chosen crossover frequency matches the spatial aliasing frequency. Corresponding to Table 8-1 and Table 8-2, the green (dotted) boxes show the conditions in which the crossover frequency is chosen 'optimally', i.e. there is no aliasing and the crossover frequency is not too far below the aliasing frequency. The red (dashed) boxes show the conditions in which there is aliasing in the signal and the blue (solid) boxes show conditions in which there is more phantom source contribution than would have been necessary.

A possible interpretation of the results is: an OPSI system applying the optimal crossover frequency (green box) always achieves the best possible result for any WFS loudspeaker distance. When the crossover frequency is too high (→ aliasing in the signal) or too low (→ too little WFS information) the sound colour performance is degraded. Note that the OPSI systems OPSI_3/750 Hz, OPSI_12/750 Hz and OPSI_24/750 Hz produce a similar signal, as there is a perfect WFS signal below the same crossover frequency and similar stereo bases for the stereo part. It makes no difference whether this unaliased WFS signal is produced by a 24 cm-spaced, a 12 cm-spaced or a 3 cm-spaced array. The same is true for the systems OPSI_3/1500 Hz and OPSI_12/1500 Hz.

A reduction of the crossover frequency below 3000 Hz leads to a degraded sound colour performance. This can be read from the results of the OPSI_3 system. The same rule is valid for the other systems, only that the existence of aliasing in the signal gives even worse results. This also means that for best results, the crossover frequency cannot be reduced further than the aliasing frequency.

Observation 7:

The achieved results are plausible, can be well explained and thus are considered to be reliable.

The plausibility of the observed results is high. They can be suitably explained by the presented interpretations. The mentioned differences are significant (see Tables 8-3). The grades for repeated stimuli show a good match. The colouration grades of the anchors used were rather similar for all trials, see Figure 8-5.

8.5 Objective analysis

An objective analysis of the experiment stimuli may lead to a better understanding of the obtained results of the listening test. The perceived colouration may for instance be ruled by the actual differences in the resulting ear signals. This would make the analysis considerably easier and would enable a solid basis for quality prediction and optimisation. The existence of this dependence is also the key to the fundamental question of the applied perception mechanisms.

The objective analysis includes an analysis of these system parameters:

- The free-field transfer functions
- The free-field *binaural* transfer functions

The free-field transfer functions are analysed in order to confirm the correct simulation of the reproduction systems and to evaluate the degree of correctness of the sound field. They are plotted here to provide an overview of the systems of the experiment. Figure 8-8 shows the free-field transfer functions of the 15 OPSI systems simulated at a distance of 1.5 m from the array. The free-field transfer functions can be interpreted as the simulated response of two omni-directional microphones spaced at ear distance in the listening area. The red and green graphs show the WFS and the stereophonic part of the combined OPSI signal. The figure is organised as a table in which the x-axis contains the five different crossover frequencies f_{cross} and the y-axis three different WFS loudspeaker spacings Δx.

The OPSI principle can be easily retraced by this figure. The perfectly flat response below f_{alias}, the aliasing and the stereophonic comb filtering are clearly visible. The different systems are classified by coloured frames that indicate how well the crossover frequency f_{cross} matches the spatial aliasing frequency f_{alias}. The colour code corresponds to Table 8-2.

The frequency responses of Figure 8-8 can help in analysing the sound field and its physical parameters. However, for this investigation, these transfer functions are not as interesting as the actual ear signals, and more precisely the difference between the ear signals of the sources. Hence, the differences between the binaural transfer functions from different source positions is the decisive physical measure.

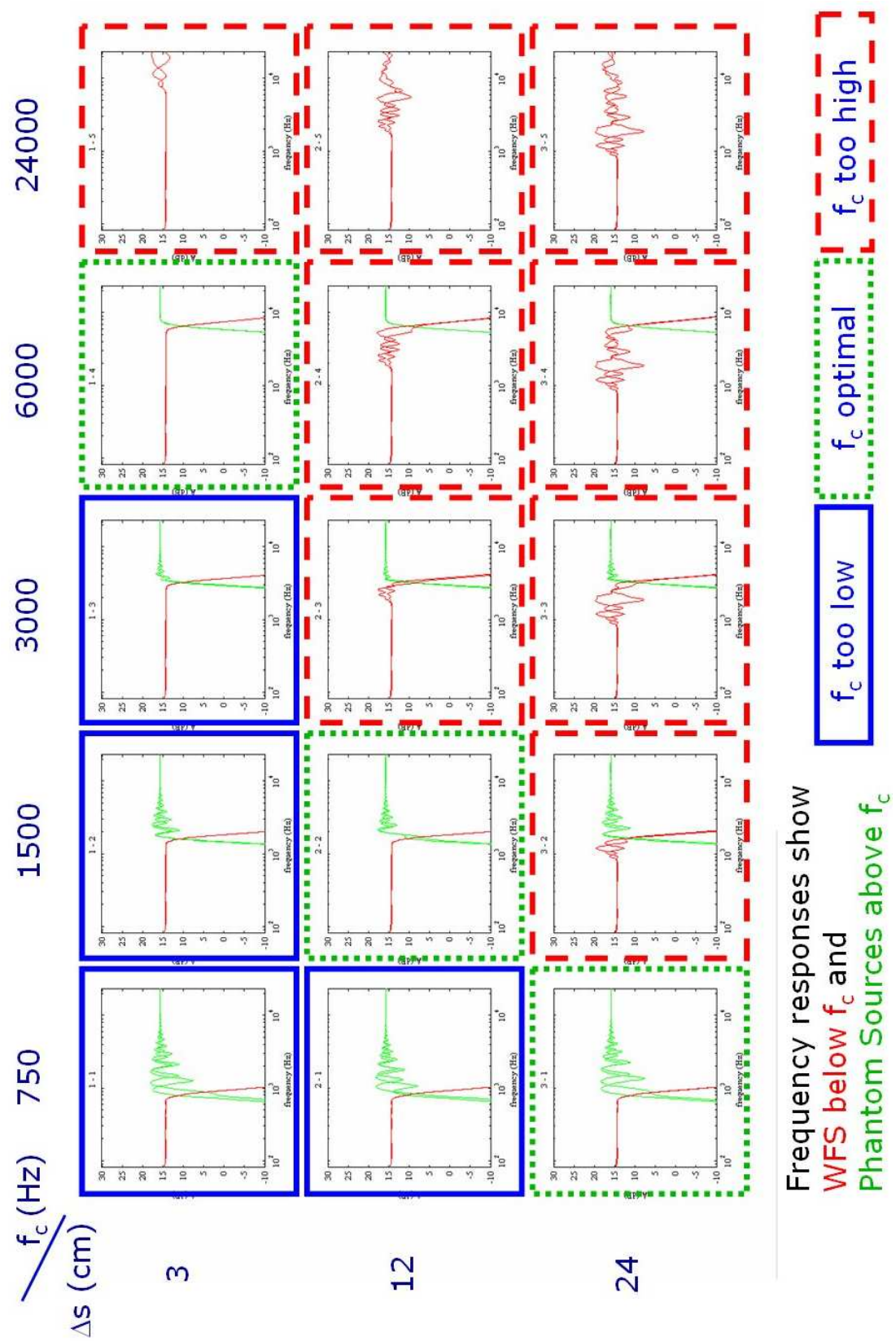


Figure 8-8: Free-field transfer functions (omni-directional microphones at ear distance). Frequency responses are smoothed with critical band filters. The figure shows an arrangement of the OPSI systems by loudspeaker spacing (y-axis) and crossover frequency f_{cross} (x-axis). The colour code of the frames corresponds to Table 8-2.

The BRS system utilises ear signals that must contain the room acoustics in order to produce out-of-head-localisation. For the analysis, only the direct sound without the reflections and the reverb tail is used, as the differences between the BRIRs are mainly in the direct sound. Furthermore, the direct sound in particular is assumed to be salient for the perception of colouration in this experiment.

'Internal spectrum', 'central spectrum'

This investigation adopts an easy approach to predicting the perceived colouration. The prediction utilises a combination of the left and right ear signals, and literature suggests the 'internal spectrum' or 'central spectrum' for this task (Bilsen, 1977; Zurek, 1979; Kates, 1985; Brüggen, 2001a, 2001b; Salomons, 1995). The central spectrum is generally considered as the spectrum evaluated by the auditory system after the localisation of the source. It is generated by averaging the left and the right power density spectra. Raatgever and Bilsen (1986) introduce the CAP (central activity pattern) function, which is a realisation of the central spectrum model including weighting functions for frequency dominance and reflection delay time. As the input signals for the prediction described in this chapter were binaural signals, a further weighting could be omitted. The weighting of delay times is not adequate for WFS due to the dense sequence of the individual loudspeaker signals in the response of a WFS signal. Zurek proposes a compression of the amplitude spectra before the central averaging – it should be noted that his motivation was the best fit of the A_0-criterion (see below) and not a fundamental explanation of the meaning of this compression. Such a compression is not applied in this investigation.

Furthermore, a critical-band filtering is performed before the averaging to simulate the frequency analysis properties of the auditory system. This critical band filtering is realised through a Patterson filter bank.

The central spectrum is the starting point for a prediction of the colouration. The next section will introduce this aspect.

'Spectral alterations'

The intra-system spectral differences in the binaural transfer functions between the reproduced sources, processed after the central spectrum theory, will be referred to as '*spectral alterations*'. The spectral alterations of the experiment's systems are given in Figure 8-9.

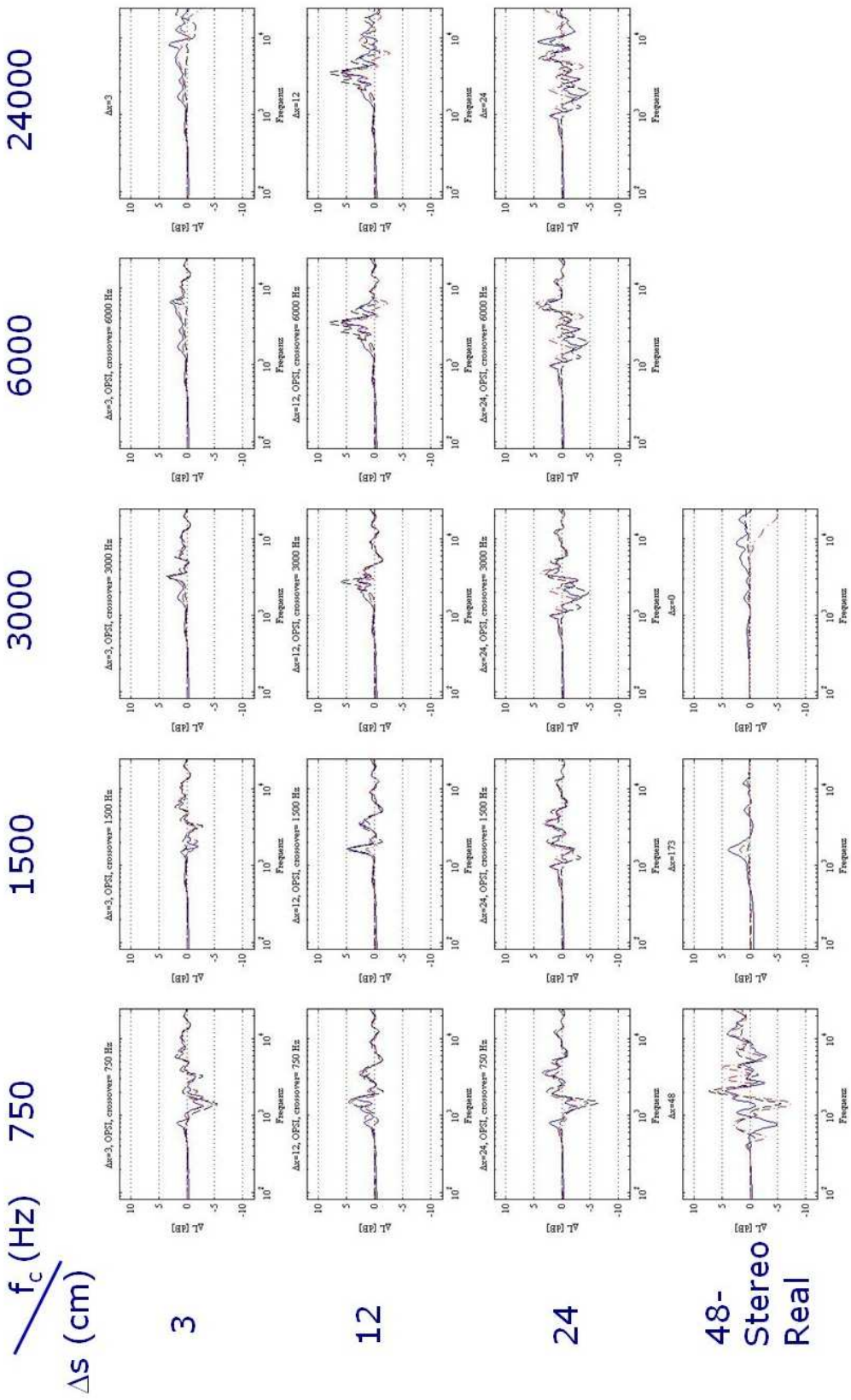


Figure 8-9: Spectral alterations = intra-system binaural transfer function differences between the reference direction and the other source directions, processed after the 'central spectrum' theory.

8.6 Prediction of the colouration perception

The perception mechanisms described in chapter 3.6 differ fundamentally. One of their main differences is the relevance of the actual existing ear signals for perception. It is hypothesised that the superimposed signals existing in stereophonic listening are not directly processed as is the case for a single source. A certain procedure of signal stream segregation or decolouration may be effective and change the listener's perception. A prediction of the perceived colouration based solely on the spectral alterations may therefore lead to different results for the different system types because it does not take into account the hypothesised listener's ability to segregate or decolour.

The prediction is attempted by performing a regression analysis. It is based on the measured spectral alterations and the perceived colouration gathered in the experiment.

Predictors

The spectral alterations are a frequency-dependent measure. Thus, in order to enable a prediction based on the spectral alterations, the relevant information has to be extracted and converted into a single value. The goal is that the new parameter correlates with the perceived colouration. In the literature, different proposals for this task exist:

- A_0-criterion: measure defined by Atal et al. (1962). Renewed definition by Salomons (1995): *"Coloration is perceptible if the maximum modulation depth (i.e. the level difference between maxima and minima) of the spectrum convolved with auditory filters) exceeds a certain threshold"*. The A_0-criterion was used as a quantitative measure in this investigation. This will be called the A_0-measure.
- Peak-to-trough ratio: measure that is similar to A_0-measure, used in (Krumbholz, 2004) and (Zurek, 1979).
- B_0-criterion: measure defined by Atal et al. (1962). Renewed definition by Salomons (1995): *"Coloration is perceptible if the area of the autocorrelation peak belonging to the delay time of the most dominant reflection exceeds a certain threshold B_0, normalised on the area of the peak at zero delay time."* The B_0-criterion measures colouration due to a distinct reflection and is not suitable for the task of this experiment.
- Spectral deviation (SD): standard deviation of the spectral alterations on a logarithmical frequency scale. It measures the mean deviation of the spectrum from its mean value. The standard deviation has to be calculated from the graph in logarithmical

frequency representation to correspond to auditory perception. This measure was used by Berkley (1980, cited in Brüggen, 2001a). It is used in this investigation as well.

- D_0-criterion: measure that is similar to SD, used in (Salomons, 1995)

These predictors are comparably crude measures that are defined without a precise psychoacoustic justification. It cannot be expected that they fully agree with actual perception. One problem is: the difference between two spectra does not incorporate the absolute spectrum which is likely to have a significant influence on perception. For instance, it is known that peaks are more prone to be perceived than notches, see Figure 8-10 and (Bücklein, 1981). Furthermore, there is no weighting applied to consider the unequal influence of different frequency bands on the colouration perception (Tsakiris et al., 2005). Brüggen (2001b) detected a 'dominant contribution to colouration' of frequencies below 2 kHz. The mentioned measures will provide a rough prediction and the results will show if this is sufficient for the aim of this investigation.

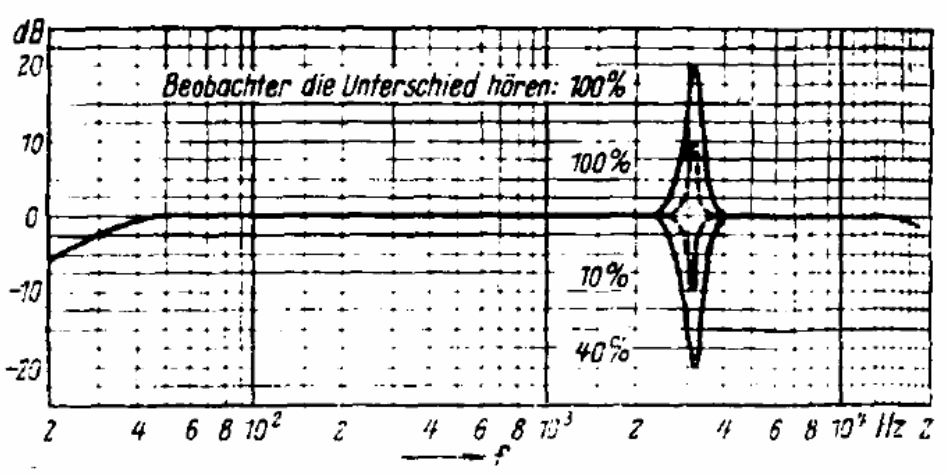


Figure 8-10: from (Bücklein, 1981): Audibility of peaks and notches at 3.2 kHz. 100% of the subjects detected the peaks of 10 and 20dB, whereas only 10% detected the 10dB notch and 40% of the subjects detected the 20dB notch.

Regression

It can be seen from Figure 8-11, Figure 8-12 and Figure 8-14 that the regression based on predictors A_0-measure and spectral deviation produce rather good results in terms of the qualitative distribution of the results. The quality of the regression can be read from the R^2 values (R^2 = squared correlation coefficient):

- predictor A_0-measure: $R^2 = 0.69$
- predictor spectral deviation SD: $R^2 = 0.69$

- multiple regression using predictors A_0-measure and spectral deviation SD: $R^2 = 0.76$

The multiple regression based on both predictors A_0-measure and spectral deviation SD leads to a better performance in spite of the high correlation of the predictors of 0.81. The relevant statistics do not confirm collinearity.

Spectral alterations of stereo and WFS

The differences in the frequency responses which can be deduced from Figure 8-8 and Figure 8-9 show that the stereophonic contribution is different from the WFS contribution regarding shape and amplitude. It would seem that the stereophonic signals generally have less spectral alterations in the higher frequencies. The predictors shown in Figure 8-11 and Figure 8-14 confirm this observation. The rationale can be found by analysing the time domain signal for the two cases stereo and WFS. The superposition of the two (or more) loudspeakers creates a comb filter, which is defined by the path difference between the signals at the receiver's position. This path difference defines the peak-to-peak distance in the frequency response which is constant throughout the whole frequency range. The greater this peak-to-peak distance, the smaller the fundamental frequency of the comb filter, and the lower the frequency for which more than one peak falls into one critical band. This means that stereophonic signals with a large path difference between the two loudspeakers do not have significant comb filtering in the higher frequency ranges due to the smoothing by the critical band filtering. The minimum path difference is significantly larger for stereophonic setups than for WFS. This is particularly true for a two-channel stereo setup in which the two loudspeakers are positioned in sufficiently different directions (usually $\pm 30°$). Also, head shadowing further decreases the comb filter effect for frequencies above 2 kHz. Stereophonic setups that do not correspond to these descriptions are indeed known for worse colouration properties. Examples include a stereo setup at the side of the listener, or a stereo setup consisting of more than two loudspeakers which reproduce coherent signals.

A surprising result of the subjective data described in the last subchapter was the difference in the colouration grades of the systems WFS_12 and WFS_24 (see Figure 8-6). It was questioned why the system WFS_24 was graded better than the system WFS_12. Indeed, both chosen predictors show the same property (Figure 8-11): The prediction grades WFS_24 better than WFS_12. This confirms the quality of the prediction for the pure WFS systems.

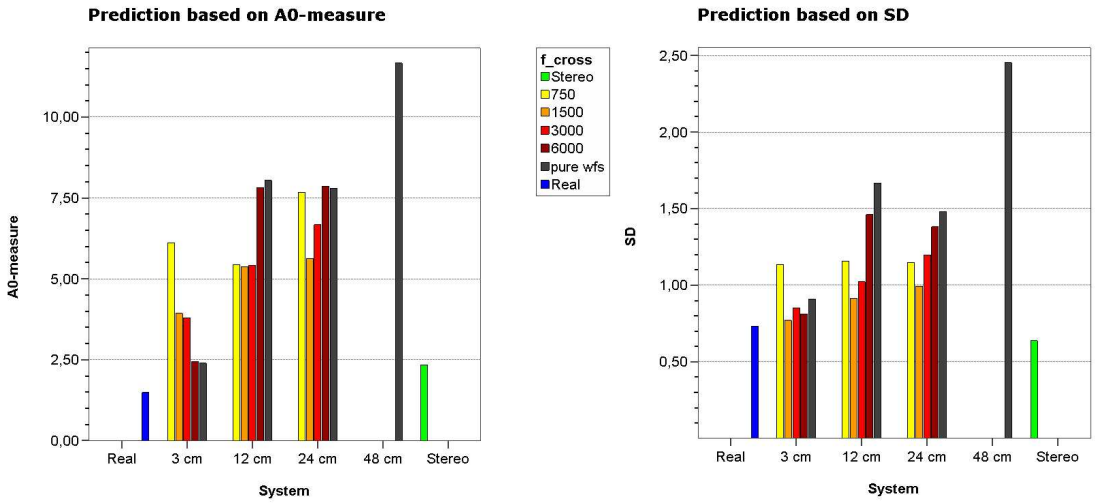


Figure 8-11: Results of the experiment as predicted by different measures based on the spectral alterations. Left figure: A_0-measure, right figure: spectral deviation SD. Compare with Figure 8-6.

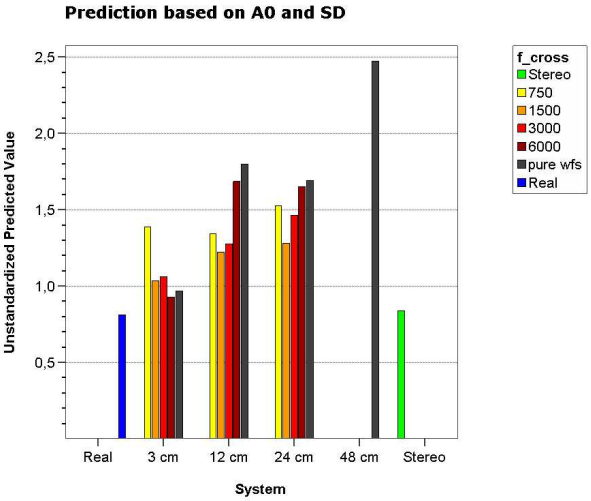


Figure 8-12: Results of the experiment as predicted by combined predictors SD and A_0-measure .

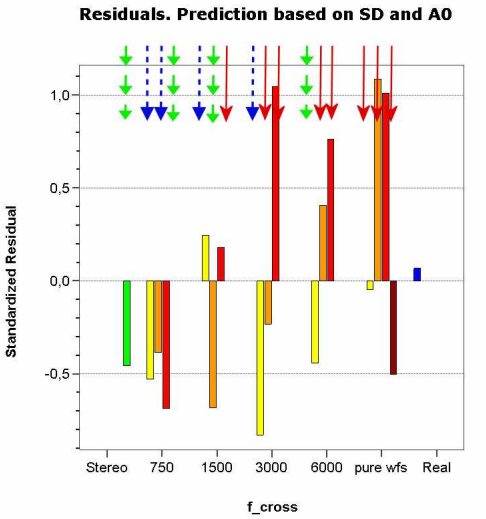


Figure 8-13: Standardised residuals of the regression based on SD and A_0-measure. The colour code of the arrows above the diagram corresponds to Table 8-1 and Table 8-2: green, multiple arrows: $f_{cross}=f_{alias}$; blue, dashed arrows: $f_{cross}<f_{alias}$; red, solid arrows: $f_{cross}>f_{alias}$.

Residuals

Two figures can be observed, which are related to the accuracy of the prediction. Figure 8-14 shows the prediction against the mean colouration grades of the experiment. The standardised residuals of the regression based on predictors SD and A_0-measure are shown in Figure 8-13, where positive residuals correspond to an underestimation of the colouration and negative residuals to an overestimation of the colouration.

The results show that the prediction quality is dependent on the type of system. The systems containing aliasing (red, solid arrows in Figure 8-13; red colour in Figure 8-14) are mostly underestimated in their colouration, whereas the systems without aliasing but with a stereophonic contribution (blue and green arrows in Figure 8-13; blue and green colour in Figure 8-14) are mostly overestimated. This means that the perceived colouration of the stereophonic systems is lower than was predicted by the spectral alterations. This leads in to confirming the hypothesis that stereophonic perception is different from conventional auditory perception.

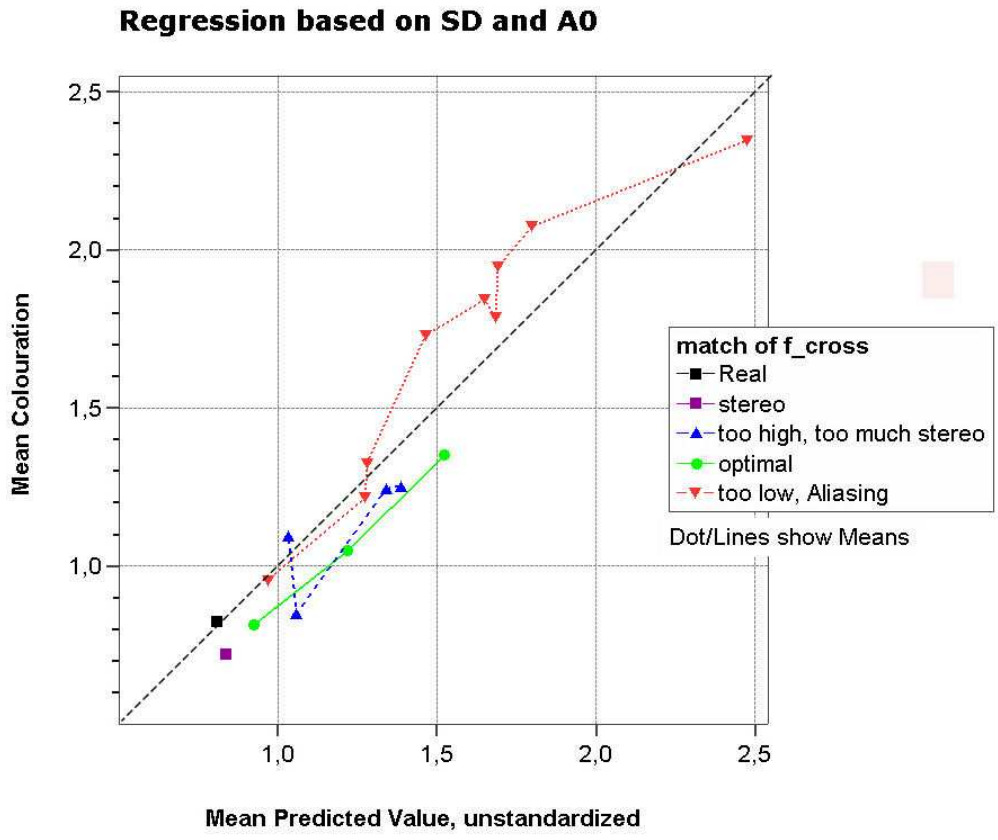


Figure 8-14: Regression analysis: The mean colouration grades of the experiment are drawn against the mean predicted values. Green and blue systems (no aliasing) are predicted and graded better than red systems (with aliasing). These colours correspond to Table 8-1 and Table 8-2. Systems containing stereo (green, blue, purple) are rather overestimated, aliased systems (red) are rather underestimated in their colouration by the prediction.

8.7 Discussion of the prediction

Prediction accuracy

The prediction described in the preceding section is somewhat vague. It cannot be the aim of this investigation to find a more accurate prediction for several reasons. Firstly, the achieved results already show a certain difference between the systems, and indeed this difference is more interesting rather than absolutely accurate. Secondly, a more accurate prediction would result in an enormous increase of complexity, exceeding the scope of this task within the investigation. Lastly, even if this was done and the complexity was increased, it is not expected that the perception would be simulated significantly better. Literature shows that colouration perception is a difficult task and that a reliable objective measure does not yet exist as long as binaural perception is considered (Brüggen, 2001b; Tsakiris et al., 2005).

In spite of the crudeness of the chosen measures, the prediction results in an unexpectedly high correlation. This means, the results of the listening test described in this section correlate to the relevant predictors to a surprisingly high degree. The R^2-value of the prediction based on all test signals is $R^2=0.76$ whereas the prediction based only on the pure WFS signals even result in $R^2=0.81$.

Spectral alterations

The results of both the objective and subjective measures show some basic properties of the OPSI principle and therefore may help reveal some basic properties of stereophonic perception.

The prediction leads to the conclusion that the OPSI and stereo sources show less colouration. Keeping in mind that the prediction is based on colouration perception as suggested by the summing localisation theory, this theory seems to be confirmed by the results. The spectral deviation SD and the A_0-measure of the OPSI sources are in most cases smaller than those of the pure WFS sources as described in the preceding section. Also the predicted colouration of the pure phantom sources is much smaller than that of the pure WFS sources.

This leads to the conclusion that the existing spectral alterations most likely influence colouration perception also for stereophonic sources. It seems that a stringent function of the association model cannot be derived from these results, as this would result in a higher independence of the colouration from the spectral deviations.

On the other hand, the analysis of the residuals in Figure 8-13 and the regression in Figure 8-14 showed that the sources containing stereophonic signals were overestimated in their colouration. This supports the idea of an existing decolouration of these signals. Though the

decolouration of the perceived signals exists, it does not seem as effective as would be suggested by the association model. It can be hypothesised that decolouration due to stream segregation as suggested by the association model does not lead to a full separation regarding the colouration perception. Consequently, this would mean that the auditory system is only able to some degree to recognise the original sound colour. Experiments by Brüggen (2001b) showed that the perceived sound colour of a binaural signal often equals the predicted sound colour as perceived by the less coloured ear. In other cases, however, this 'binaural advantage' was larger.

Relationship between locatedness and colouration

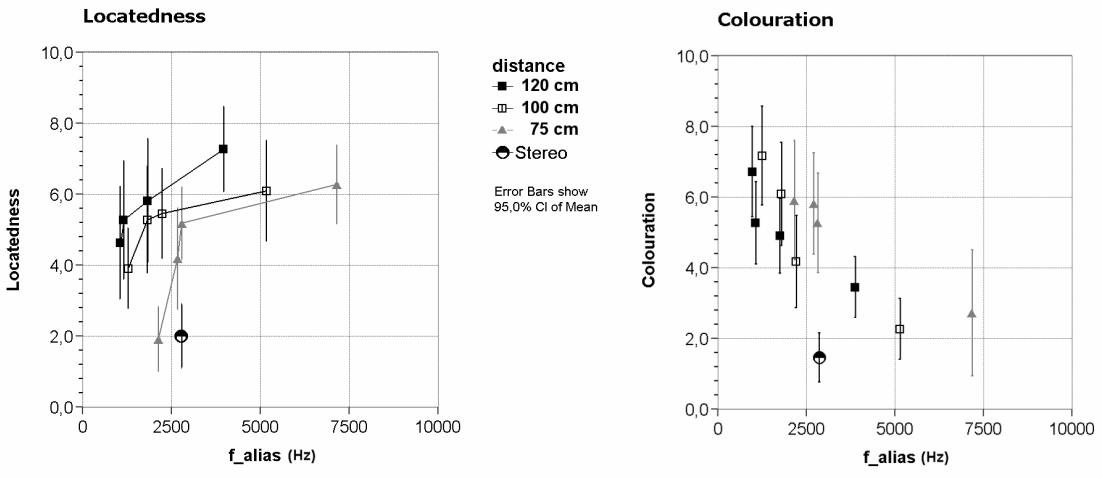
In a preparatory pilot experiment, different stereophonic techniques were used to create the phantom source shift. A phantom source shifted by level differences was compared to a phantom source shifted solely by time differences. The results of this comparison showed that colouration was higher for the time difference phantom source compared with the level difference phantom source, see Figure 8-4. Furthermore, the locatedness was found to be inferior. The prediction based on the spectral alterations would suggest the same result. Again, the conclusion is: the colouration performance of the stereophonic reproduction depends on the spectral alterations to a certain degree.

A second conclusion may be even more important. It was observed that the attributes locatedness and colouration are related. The time-panned phantom sources were inferior in terms of both locatedness and colouration. It would be interesting to know whether the imperfection of one attribute implies the imperfection of the other .

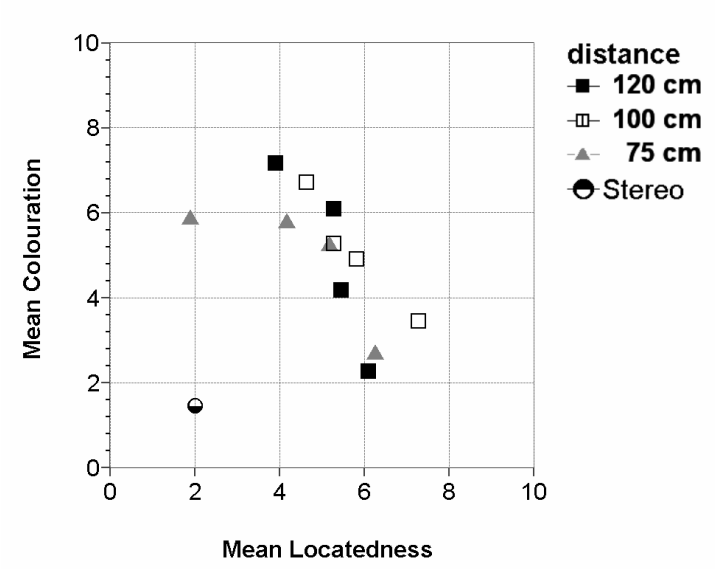
A second pilot experiment was performed in order to reveal the relationship between colouration and aliasing frequency. As it was planned to perform the experiment on a real WFS setup, the aliasing frequency f_{alias} had to be increased with the help of a special procedure. It is known (see chapter 9.3.1) that for focussed sources a decrease in the source-listener distance leads to an increase of f_{alias}. Therefore, the source-listener distance was varied by positioning a focussed source at different distances. Hence, an f_{alias} as high as 7 kHz could be produced.

The expected dependency between aliasing and colouration was detected, but more importantly, an anti-correlation between locatedness and colouration was found. It turned out that the source-receiver distance was a parameter influencing the results significantly. Figure 8-15 shows plots with the results of this pilot experiment. The sources with the decreased focussed source distance were more difficult to localise and had an increased colouration. Thus, the decreased locatedness apparently also influenced the sound colour properties of the virtual source. The relationship between these two attributes is shown in Figure 8-15c. The result of

the stereo setup (semi-solid circle) identified a significant difference between stereo and WFS reproduction. The experiment showed that only the stereo source can show good colouration grades along with rather bad locatedness grades. This result can be interpreted by discriminating between the localisation of the separate loudspeaker signals – which is successful – and the fusion of the phantom source which creates the diffusion of the localisation.



a) Locatedness vs. aliasing frequency **b) Colouration vs. aliasing frequency**



c) Mean colouration vs. mean locatedness

Figure 8-15: Second pilot experiment (after Augustin, 2004) applying focussed WFS sources at different source-listener distances. The attributes colouration and locatedness are gathered independently. They are plotted against the aliasing frequency and against each other. High figures mean good locatedness and large colouration respectively. The parameter is the source-listener distance.

Theile hypothesised that the success of the location association determines the success of the decolouration (Theile referred to it as ‘inverse filtering’). He explained that in the case of an increased colouration in stereophonic reproduction this is due to a decreased success of the location association stage. In his experiment, Theile artificially deteriorated the success of the location association stage, for example by introducing delays to one of the ears. These signals have the same central spectrum as natural signals but they are not natural. This indeed caused an increase of the perceived colouration (Theile, 1980).

8.8 Summary of chapter 8

WFS properties

Spatial aliasing introduces colouration to the virtual sources in WFS. Aliasing is particularly disturbing because it changes rapidly with source or listener movements and considerably large amplitude peaks and notches are noticeable up to the highest frequencies. The colouration can be predicted with comparably good accuracy by an analysis of the spectral alterations of the ear signals. Colouration generally is dependent on the aliasing frequency and the shape of the aliasing, i.e. the peak/notch distance and the spectral deviation. In the experiment, a significant difference in the sound colour reproduction between the different WFS systems was found.

Perception of stereo

Stereo showed the least colouration of all systems. In addition, the spectral alterations were smaller for the stereo and OPSI sources. This means that the spectrum of the source is often flatter in the case of stereo compared with typical WFS. Though comb filtering exists, the better spectrum can be explained by the head shadowing and the smoothing due to critical band filtering.

However, both the stereo and OPSI sources were graded better than predicted from the spectral alterations of the ear signals. Hence, the experimental results show that the perception of stereo is to be regarded differently from that of WFS. This suggests the existence of some kind of decolouration which leads to an improvement of sound colour perception in stereo reproduction, similar to the decolouration of a sound source in a reflective environment. It is hypothesised that the decolouration is successful as soon as the auditory system is able to segregate the interfering sound contributions. This was similarly described by Theile in his ‘association model’.

A stereophonic reproduction therefore has advantages over WFS, because in WFS the discrete loudspeaker signals cannot be segregated due to their density in arrival time and incidence angles.

The segregation cannot be considered as leading to an ideal separation with regard to localisation and sound colour perception. The described and other experiments have shown that a certain dependency of the perceived sound colour on the superimposed sound field exists. This leads to the hypothesis of a partial decolouration in stereophonic perception. It can be considered an interpretation of Theile's association model.

OPSI performance

The sound colour properties of a WFS virtual source can be optimised by avoiding aliasing. Both theoretical and practical investigations have shown how an introduction of stereophonic techniques to WFS can help avoid colouration artefacts. The results show that both the spectral alterations are minimised and the colouration perception is improved by the OPSI method. The theoretical assumptions and the principle of OPSI, being a hybrid reproduction technique of WFS and stereo, are confirmed in the experiment in relation to colouration.

9. Experiment 3: Relevance of the wave front curvature for distance perception in WFS

9.1 Introduction

The aim of the investigation described in this chapter is to examine the auditory perception of the distance of WFS virtual sources. The study is restricted to an evaluation of the validity of the wave front curvature for distance perception on a fixed (or static) listening position. Theoretical investigations aim to find evidence for the existence of distance-dependent differences in the ear signals. The experiments examine the distance perception of nearby virtual WFS sources as well as natural sources under anechoic conditions.

The chapter starts with this introduction, presenting the objectives of the investigation. Section 9.2 introduces the experimental setup on which the simulations in section 9.3 are also based. The experiments are presented with an introduction of their design (section 9.4) and a description of the results for real sources (section 9.5), and for the virtual sources (section 9.6). Further considerations on the reproduction of head shadowing in WFS are presented in section 9.7. A summary and the conclusions are given in section 9.8.

This investigation is based upon the discussions in preceding chapters. The basic cues for distance perception were described in chapter 2.5.1. The specific properties of WFS for the reproduction of source distance were discussed in chapter 4.3.5. Finally, a comparison between WFS and stereo regarding this attribute was performed in chapter 6.5, where the rationales for the investigation described in this chapter were developed and the objectives defined.

This investigation is restricted to the possible differences between WFS and stereo at a fixed listening position only. This means that cues available to both reproduction techniques, or available only with listener movements, are ignored. The investigation concentrates on the remaining differences regarding the physical properties of the sound field that could provide a cue for distance perception. These differences consist only in the reproduction of the wave front curvature. This study reveals the meaning of the cues related to the wave front curvature in WFS on a fixed listening position.

As a solution for multiple or moving listeners, the specific advantages of WFS are apparent, and some of its properties are superior to other reproduction techniques. However, the role of these properties on the acoustic perception at a fixed listening position has not yet been suffi-

ciently investigated. On a fixed listening position, neither the size of the listening area nor the possibility to move therein are relevant, and thus a superiority of WFS regarding distance perception cannot be claimed based on existing knowledge.

Due to the focus of this investigation on the wave front curvature, the virtual sources have to be at a close distance to the listener because only there are the potential cues valid. Hence, the virtual sources used in this experiment are focussed sources. The specific perceptual properties of focussed virtual sources in WFS need to be discussed separately: the weight of the direct sound cues is apparently higher compared to other virtual sources due to the increased direct-to-reverberant energy ratio. Furthermore, the influence of any reflections from the array speakers themselves is considered particularly disturbing because the array reflections arrive before the early reflections of the virtual source and therefore can hardly be masked (see Figure 4-19 in chapter 4.2.8). Hence, erroneous cues for distance perception exist which cannot be overridden by the acoustics of the virtual room and the direct sound cues would be the only correct cues to be interpreted by the auditory system.

For these reasons, the (elsewhere very important) issue of reflections and reverberation is left aside in this investigation. Concentration is put upon the direct sound cues for the perception of (nearby) sources and their reproduction over WFS arrays. These can be examined under anechoic conditions.

9.2 Setup for experiment and simulations

The experiment aimed at measuring the perceived distance of nearby virtual and natural (real) sources. An experimental setup was created in the anechoic chamber of the IRT, Munich. Its volume is 80m^3 and its floor size is $4.5 \cdot 6\ \text{m}^2$. Its lower limit frequency is 80 Hz.

Figure 9-1 shows the array/source/listener geometry. The listener's ear axis is perpendicular to the WFS array (synthesising a virtual source) and the real source respectively. The right ear points to the (virtual) source; the left ear is turned away from the (virtual) source. This head orientation was chosen because the binaural differences are maximum for this case. The virtual or real source is denoted by the dotted loudspeaker in Figure 9-1. The subjects could move their head freely.

The distance between the source and the centre of the listener's head is called the source distance *d*. The distance between the array and the listener is called the array distance *a*. Furthermore, the distance between the source and the array is the source position *z* (= *a* – *d*).

Table 9-1 and Figure 9-2 show which distances *d* and corresponding source positions *z* were chosen. In the case of WFS, positive source positions correspond to focussed sources, negative source positions correspond to sources behind the array.

The non-focussed virtual sources behind the array (1.50 m and 1.90 m) will be indicated by dash-dotted lines in the figures of the next section.

The linear WFS array consists of n = 16 loudspeakers with an interspacing of Δx = 17 cm. This makes an array length of 2.55 m. Tapering was performed using a spatial window (Hanning window), equalisation was performed according to the WFS driving functions (see chapter 4.2.1).

As a real source, a single small loudspeaker of the type ELAC 301 (width = 9.1 cm) was chosen. Further details on the experiment design are depicted in section 9.4.

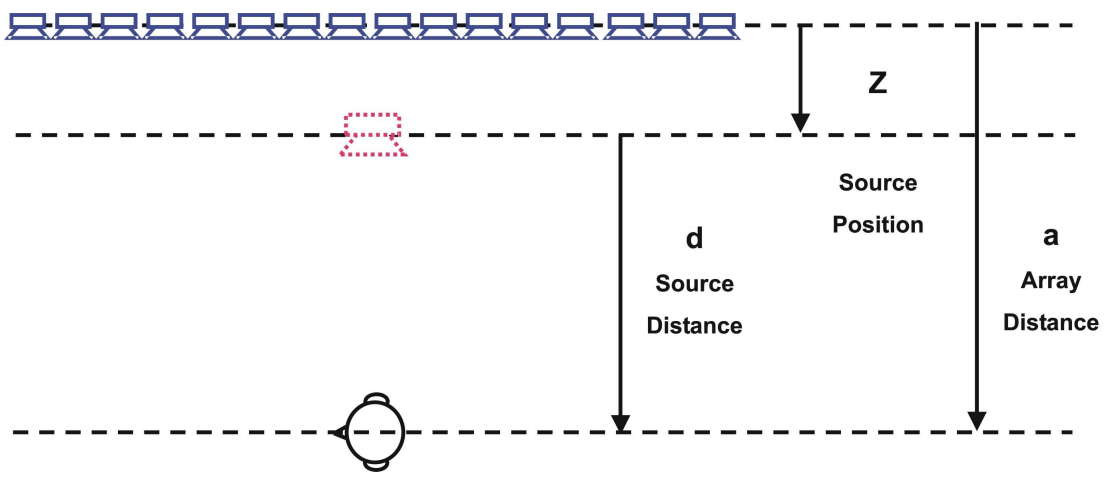


Figure 9-1: Array-source–listener geometry for the simulations/experiments: the (virtual) source (red dotted loudspeaker) is located on the ear axis which means the sound propagates perpendicular to the listener's line of sight.

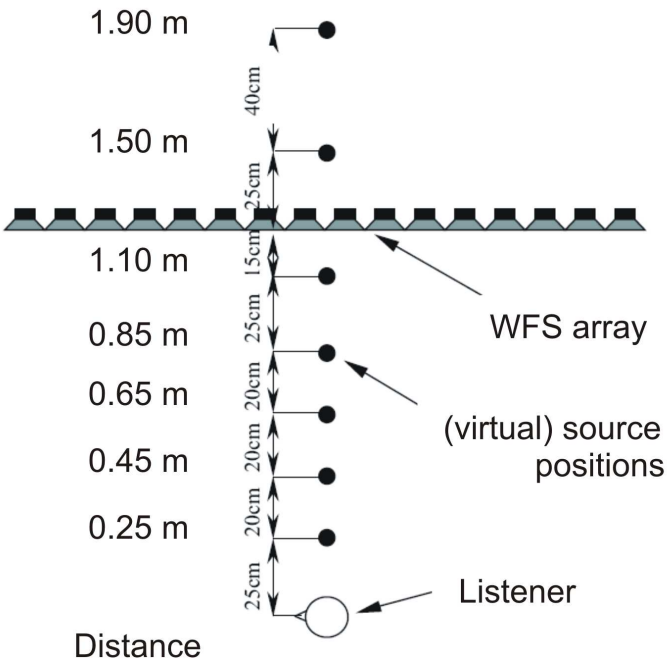


Figure 9-2: Illustration of the experiment geometry with all source positions of the experiment. Illustration from (Kerber, 2003).

Source distance *d*	Source position *z*
1.90 m	*- 0.65 m*
1.50 m	*- 0.25 m*
1.10 m	0.15 m
0.85 m	0.40 m
0.65 m	0.60 m
0.45 m	0.80 m
0.25 m	1.00 m

Table 9-1: Source distances *d* and corresponding source positions *z* used in the experiment. Two of the virtual sources (written in *italic*) were synthesised behind the array, the others in front of the array.

9.3 Theoretical analysis of the real and the synthesised wave field

Before the experiment is described, theoretical observations and simulations are presented in this section which will be an important basis for an explanation of the experimental results.

9.3.1 Physical deficiencies of focussed sources in WFS

As discussed in chapter 4.2.3, a WFS virtual source is not a perfect copy of the real source for several reasons. This applies to virtual sources behind the array, as well as to focussed sources. Focussed sources furthermore have a special status in WFS. They are not derived from basic WFS theory. Boone et al. (1996) state: *"One might argue that the situation with a virtual source in front of the array is not in agreement with the Kirchhoff theory, which states that the source must be behind the array. However, our synthesized virtual source is not a true source and could also be present due to a focussing transducer behind the array, indicating that the theory is applicable indeed."*

Focussed sources have different properties compared to normal virtual WFS sources behind the array. These properties will be summarised in this section.

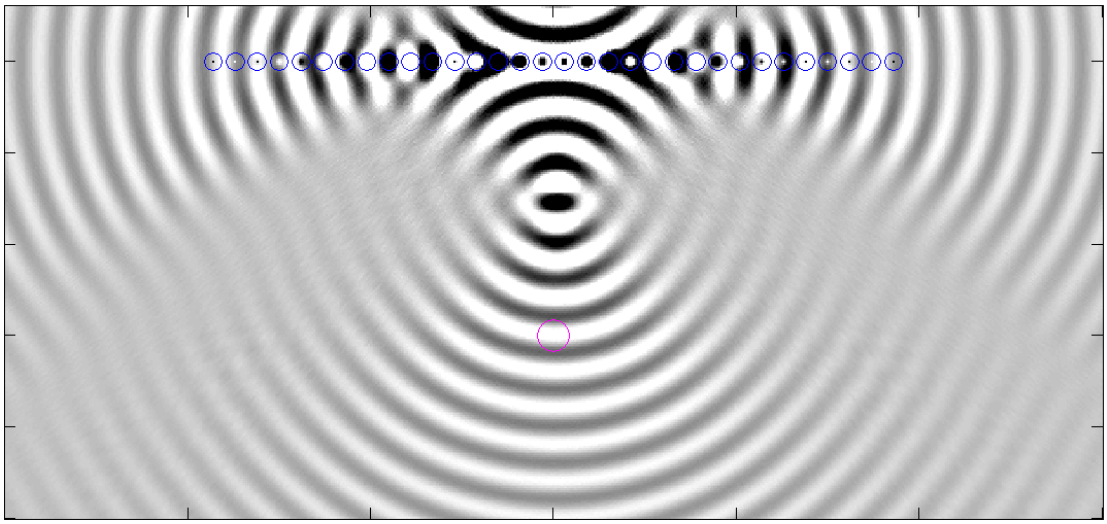


Figure 9-3: Snapshot of the pressure field of a focussed source, synthesised by a WFS array. The virtual source emanates a sine wave of a frequency well below f_{alias}. The array speakers are indicated by blue circles. Tapering is applied. The sound image is not correct for listening positions between the focus point and the array. The pink circle indicates a possible listening position.

Listening Area

The correct wave front is synthesised only behind (in the propagation direction of the array) the focus point. In other words, for correct perception, the source has to be between the listener and the array. In fact, the wave front curvature is synthesised correctly also between the source and the array. However, in this area, the propagation direction of the sound is reversed, meaning that the sound emanates from the array and not, as desired, from the source. A snapshot of the pressure field of a focussed source is given in Figure 9-3.

Spatial aliasing

Spatial aliasing limits the correct synthesis of the sound field in the upper frequency range. It depends on the travel time differences between adjacent array loudspeakers. Above the spatial aliasing frequency $f_{alias,}$ the sound field is neither spatially nor spectrally correct. In the case of focussed sources in WFS, the travel time differences between adjacent WFS array loudspeakers are rather small. The contributions of the array loudspeakers are synthesised such that they focus in one point, this being the virtual source position. That means that their travel times are designed such that they arrive at the source position at the same time. Consequently, at a small distance d from the source, the travel time differences between adjacent array loudspeakers are very small. This makes f_{alias} very high. For greater source distances, the travel time differences are bigger, leading to a decreased f_{alias}. This can be seen in Figure 9-4 in which the frequency spectra of focussed sources at different distances are shown. The increase of f_{alias} with decreasing source distances can be deduced from these graphs. In the example in Figure 9-4, f_{alias} is about 3 kHz for a source-receiver distance of 1.6 m. A decrease of the source-receiver distance by a factor of 2 leads to a doubling of f_{alias} .

The spatial aliasing frequency f_{alias} of a non-focussed source is significantly lower due to the bigger travel time differences.

Low frequency level

For a focussed source, the level of the lower frequencies does not increase with decreasing virtual source distance in the same way as for higher frequencies (which follow the $1/r$ -law). This leads to a distance-dependent level roll-off in the lower frequency bands. This failure can be explained by the nature of acoustic focussing. In the context of 'time-reversal acoustics' (or TRM = 'time reversal mirror'), Fink (2002) describes that focal points always have a minimum size of $\lambda/2$. This also applies to focussed sources in WFS. From Figure 9-4 it can be seen that only for distances roughly above $\lambda/2$, the frequency spectrum is flat. With further decreasing distances the focal point size is under-run, leading to a lack of level increase for the lower frequencies, i.e. a loss of low frequencies. One can of course equalise the frequency

response with respect to a reference receiver position at the cost of an over-emphasis of low frequencies for larger distances.

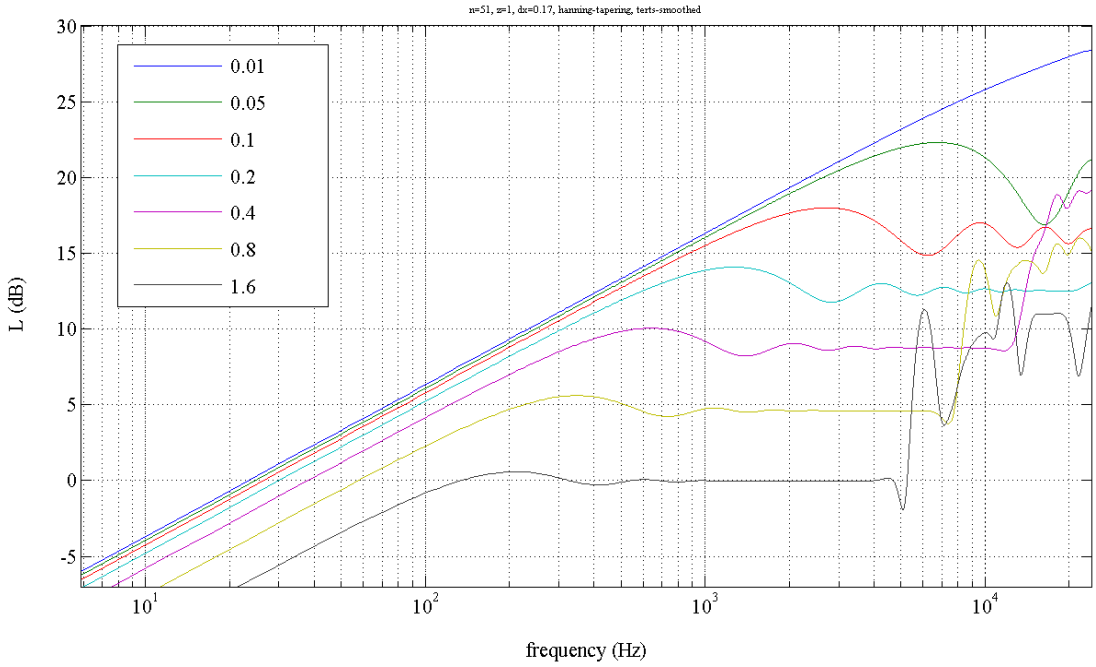


Figure 9-4: Spectrum of a focussed source at different source-receiver distances *d*. The spatial aliasing frequency can be deduced from the graphs as well as the low frequency roll-off. (*d* see legend (in m), source at *z*=1 m, linear WFS array, *n*=51 loudspeakers with an interspacing of Δx=0.17 m, tapering by Hanning window)

Diffraction effects

The diffraction effects have to be considered differently in the case of focussed sources. The relevant time difference of the truncation that underlies the creation of diffraction effects is very small. This makes the fundamental frequency of the resulting quasi-comb filter rather high leading to significant rippling depending on the source-receiver distance (see also de Vries and Berkhout, 1981). The ripples can be easily observed in Figure 9-4. In the frequency spectrum they are located between the low frequency level decay and the unaliased, flat frequency area.

Figure 9-4 also illustrates the amplitude distribution, i.e. the relationship between source distance and the sound pressure level of the source. Each doubling of the distance leads to a decrease of less than 6 dB as it would be in the case of natural sources. Hence, the spatial intensity decay fails to meet the $1/r$-law. This is due to the reduction of the synthesis dimensions to the horizontal plane (see chapter 4.2.7). For the correctly synthesised contribution, the spatial amplitude decay can be described by the following formula. The amplitudes of the incor-

rectly synthesised contributions in lower and higher frequency bands decline differently, which in general means less strongly.

$$p \sim \frac{1}{\sqrt{d \cdot a}}$$; (after Verheijen, 1998)

Level of a virtual source measured at a certain distance from the source

with p = sound pressure,

d = source distance,

a = array distance.

for the geometry see Figure 9-1.

To summarise, the desired flat frequency response is achieved only for the mid frequencies, the range and position of this correctly synthesised contribution depends on the distance d and the array set-up. The spatial intensity decay is smaller than suggested by the $1/r$-law.

9.3.2 Origin of the simulations

Different aspects are likely to influence the experimental result and therefore a careful separation of influential parameters is necessary. These parameters are analysed in a number of figures in the following subsections. Two aspects are analysed separately: on the one hand the influence of the properties of the sound field itself, i.e. the amplitude distribution, spatial aliasing, etc, and on the other hand the impact of head shadowing. Hence, the sound field is analysed both with and without the influence of the listener's head.

The following measures are important for this analysis. Figures containing analyses of these measures are summarised in Table 9-2.

a) The distribution of the sound pressure level for real and virtual sound field → this is the sound field without head shadowing. Figures showing this measure are written in **bold** fonts in Table 9-2.

b) The ear signals for real and virtual sound field → this is the sound field with head shadowing. Figures showing this measure are written in *italic and red* fonts in Table 9-2.

In the following Table 9-2, column 1 lists figures containing the level spectra at different distances, while columns 2 and 3 list figures containing the level *difference* spectra at different source distances.

Row 1 corresponds to the reference source (‘real source’), a single loudspeaker. Row 2 gives the plots for the WFS virtual sources.

Source	**Level spectra**	**‘No-Head ILD’,** *see chapter 9.3.3*	**ILD**
Real source	**Figure 9-5**	**Figure 9-6**	*Figure 9-9*
WFS Virtual Sources	**Figure 9-7**	**Figure 9-8**	*Figure 9-10*

Table 9-2: Summary of the diagrams in sections 9.3.3 and 9.3.4

The derivation of these figures needs some explanation.

Figure 9-9 was derived from an HRTF (Head related transfer function) measurement using the dummy-head Neumann KU 100i and small ELAC 301 (width = 91 mm) loudspeakers.

Figure 9-5, Figure 9-6, Figure 9-7 and Figure 9-8 were derived from simulations. These simulations are based on the following assumption: The WFS array consists of ideal monopoles and the real source is an ideal monopole, too. This means the intensity decay of both the single array loudspeakers and the real sources obeys the $1/r$-law. The simulations only include simple calculations of travel time and amplitude decay of the involved (secondary) sources.

Figure 9-10 is a simulation based on an HRTF database measured in the IRT listening room using the dummy-head Neumann KU 100i and the loudspeaker K&H O100. These HRTF data are available for azimuth directions at a resolution of 6° (that is, 60 measurements in the horizontal plane). The respective HRTFs used for the simulations necessitate a finer resolution and therefore are derived through an interpolation in the frequency domain. Due to the fact that the HRTF database was measured using a rotating dummy head recording the 0° axis of a loudspeaker, this simulation is also based on the assumption that the array consists of ideal monopoles.

9.3.3 ‘No-Head-ILD’ as a measure of the sound field without head shadowing

The levels of a real source observed from certain source distances *d* (for the setup see Figure 9-1 and Figure 9-2) were simulated in Figure 9-5. In the simulations the levels at the two positions of the ears were calculated, i.e. two positions at a

$$\text{distance} = \text{d} \pm (\text{ear distance}/2);$$

The ear distance was set to 17 cm.

As there is no head in this situation, no head shadowing occurs. However, due to the similar geometry, i.e. the same distance between the two measurement positions, the level difference between these two signals is called the 'No-Head-ILD'. It corresponds to an ILD measurement except for the fact that no head shadowing (including pinna and ear canal effects) occurs.

The $1/r$ -law dictates a level increase with decreasing distance as can be seen in Figure 9-5.

$$p \sim \frac{1}{d} ;$$

Level of a real source measured at a certain distance from the source

with p = sound pressure,

d = source distance.

The spatial intensity decay of the reference loudspeaker ELAC 301, measured on the central axis, was experimentally proven to perfectly meet the $1/r$ -law.

Also, the level *difference* between left and right ear positions increases with decreasing distance. This level difference is plotted in Figure 9-6. In this graph the 'No-Head-ILD' is simulated.

It should be remembered that with regard to auditory distance perception, it is important that in particular the low-frequency ILD (f < 3500 Hz; Brungart and Rabinowitz, 1999c; see chapter 2.5.1) depends on the source distance for nearby sources. It can be seen that for real sources the 'No-Head-ILD' is present, and that it indeed depends on the source distance if distances of roughly d < 1 m are considered.

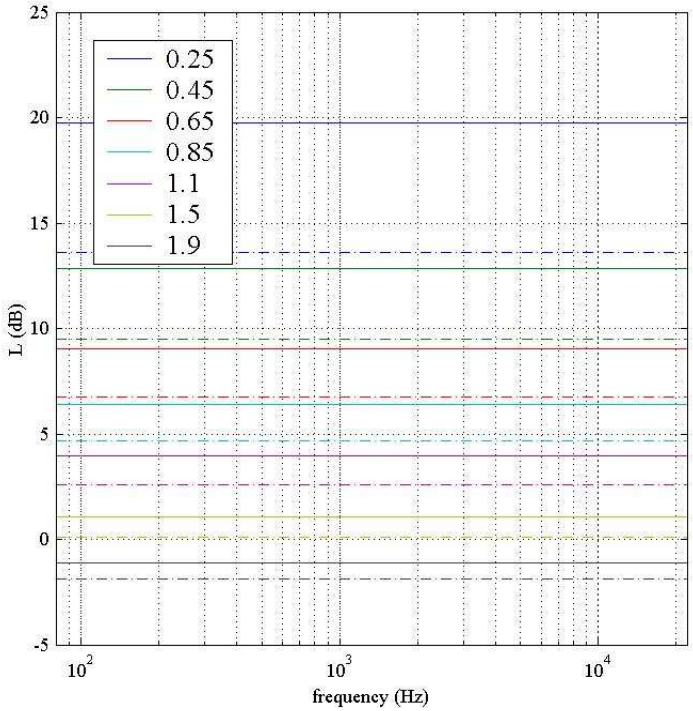


Figure 9-5: Level of a real source at distances = *d* ± (ear distance/2). Solid line: right 'ear', dashed line: left 'ear'. Source at 90°.

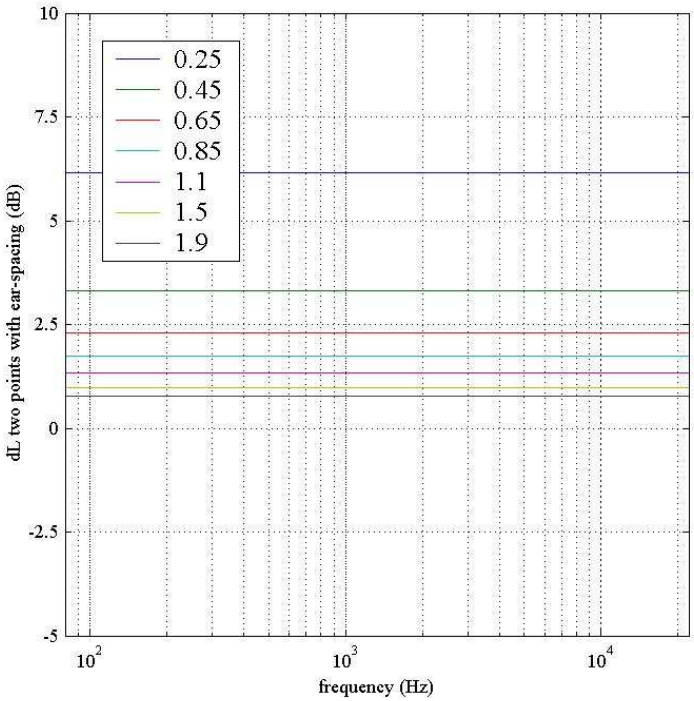


Figure 9-6: 'No-Head-ILD': level difference ΔL between ear positions in the sound field of a real source at distance *d*. Source at 90°.

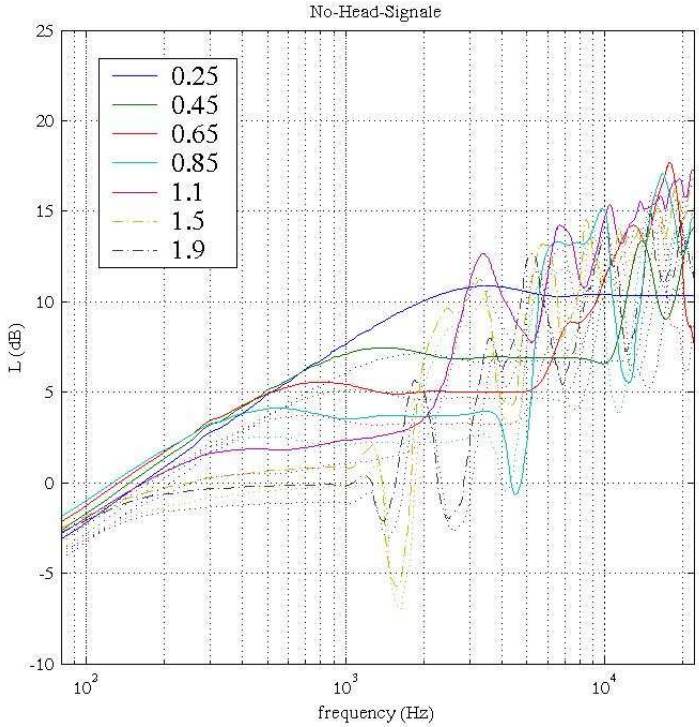


Figure 9-7: Level **of a** focussed source **at distances = *d* ± (ear distance/2). Solid line: right 'ear', dashed line: left 'ear'. Source at 90°.**

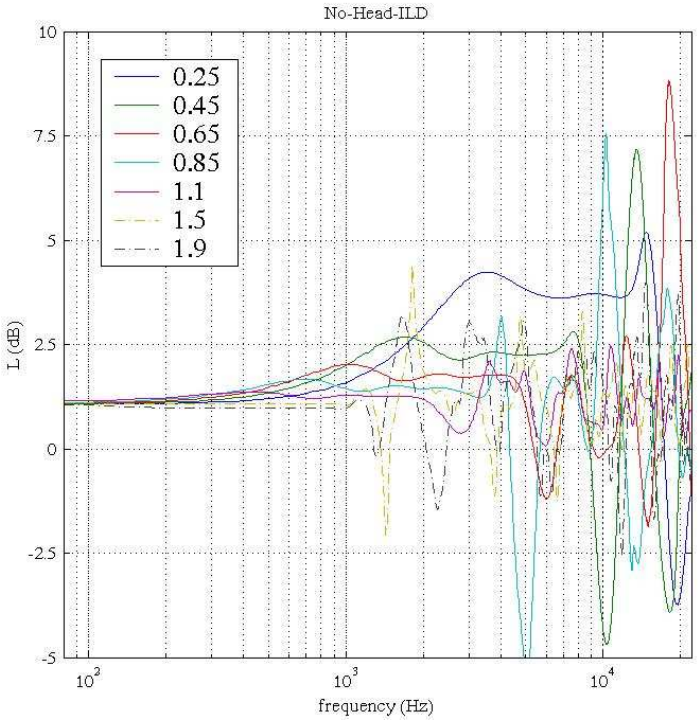


Figure 9-8: 'No-Head-ILD': level difference ΔL **between ear positions in the sound field of a** focussed source **at distance *d*. Source at 90°.**

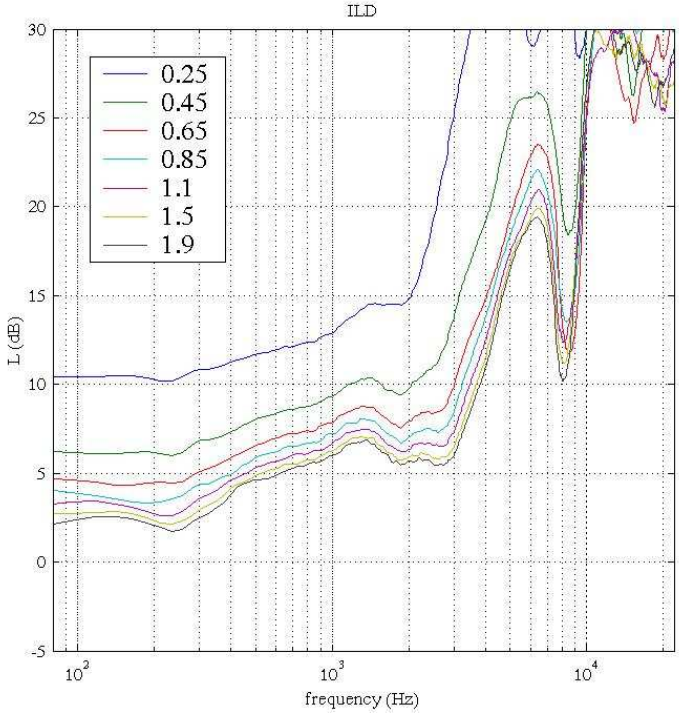


Figure 9-9: Interaural level difference ILD in the sound field of a real source at distance *d*. Source at 90°.

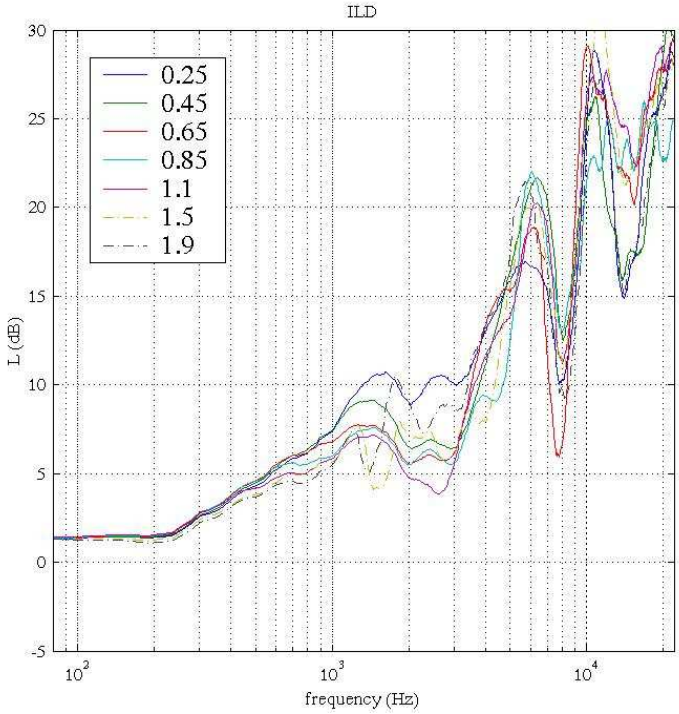


Figure 9-10: Interaural level difference ILD in the sound field of a focussed source at distance *d*. Source at 90°.

Now the virtual sources are considered:

In Figure 9-7 the level spectra of focussed sources at different distances are plotted. As mentioned in section 9.3.1, the impact of diffraction effects, low frequency roll-off and spatial aliasing is significant. A flat frequency response and – as a consequence – a significant and consistent 'No-Head-ILD' (Figure 9-8) is present only for a certain mid frequency range. The width and position of this range depends on the source distance.

A 'No-Head-ILD' is indeed present in the important low-frequency range ($\Delta L \approx 1.2$ dB) although it is not as large as in the case of a real source (compare Figure 9-6). Furthermore, the 'No-Head-ILD' does not change with different distances and thus cannot be used as a cue for distance perception.

9.3.4 The head in the WFS sound field

The simulations in Figure 9-9 and Figure 9-10 show the ILD in the real and virtual sound field. Similar to the 'No-Head-ILD' depicted in the last section, the focussed sources can only partially produce significant differences between the ILD corresponding to different distances. This can be seen from Figure 9-10. Differences in the ILD remain only in a small frequency range. The ILD of real sources is plotted in Figure 9-9. From the graphs in this figure, it may be concluded that for distances below roughly 1 m the ILD differs significantly and thus one may deduce the source distance from this ILD alone.

A further simulation, described in section 9.7, was designed to unveil how head-shadowing is reproduced in WFS and whether a longer WFS array can reproduce a more adequate ILD.

9.4 Listening tests: experimental design

Continuing the descriptions of the experiment setup in section 9.2, the experimental design is depicted in the following subsections.

9.4.1 Test panel selection

The perception of the distance of dry sources under anechoic conditions is a very difficult task for a test panel. Although a certain validity of the direct sound cues for the nearby region is expected, these cues are fragile and their detection is not simple (Brungart and Rabinowitz, 1999a, 1999c; Brungart et al., 1999b). Therefore only experienced audio researchers (exclusively from the staff of the audio systems department of the IRT) participated in the experi-

ment. Some results of naïve listeners were collected for comparison. They showed no relationship between reference distance and perceived distance, and were therefore ignored.

The data of the seven persons who performed both experiments are shown here. All subjects were male and their age was between 20 and 60. None of them had any known hearing impairment.

9.4.2 Two separate listening tests

For each type of source (real, virtual) a separate test was performed. There were two reasons for the separation: firstly, the different installations would have disturbed each other. Furthermore, the sound colours of the two different sources were, although equalised, noticeably different and it could not be excluded that the change of sound colour between the examples would play a role in the distance judgments of the listeners. The tests had a duration of $2 \cdot 20$ minutes each.

9.4.3 Test signals

Pink noise bursts were chosen as the test signal. The duration was 1000 ms including 100 ms onset and offset. This signal was tested as suitable for an optimal detection of source distance changes as found in a pilot test. This burst was repeated 6 times with an interval of 400 ms. The envelope of the first 4 (of 6) bursts of the test signal is plotted in Figure 9-11.

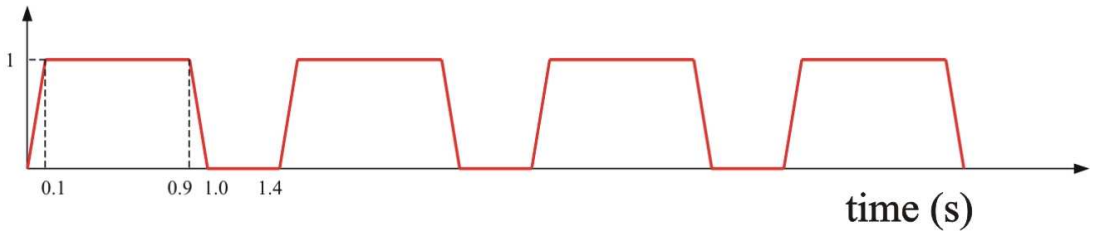


Figure 9-11: Envelope of the pink noise bursts used in the experiment. Diagram from (Kerber, 2003).

For each distance the assessment was repeated four times, except for the distances d = 45, 85 and 150 cm for which it was repeated seven times. This makes a total number of 37 test signals. Due to the existence of two different test signal groups (described in the next subsection) 74 signals in total were presented in a random order which was the same for all participants in both experiments.

9.4.4 Method of 'conflicting cues'

It is known from literature (see chapter 2.5.1) and it has been informally verified by the author that the relative level of the test signals serves as a crucial distance cue when no other cue is available. In order to avoid a distance judgment due to the perceived level alone, and also to check the validity of binaural cues, a special method of randomly varying the receiver level was applied. This method aimed to isolate the level factor in the assessment of the results.

Both test signals with *constant* source level and signals with a *random* source level were reproduced. The test signals with *constant* source level consequently had a natural variation of the receiver level at the listening position due to variations in distance, i.e. a variation of the receiver level after the $1/r$ -law. The test signals with the *randomly* chosen source level consequently had *no* natural variation of the receiver level at the listening position. Hence, the two different cues used for distance perception (level, binaural cues) were either conflicting or non-conflicting. The signals with constant source level are referred to as *'non-conflicting cues'*-signals, the signals with randomly chosen source level as *'conflicting cues'*-signals. By this method it was possible to judge which role level and binaural cues play in the listener judgment of the respective sources.

	'non-conflicting cues'- signals		*'conflicting cues'- signals*	
Distance *d* in cm	Source level in db_{rel}	Receiver level in db(A)	Source level in db_{rel}	Receiver level in db(A)
25	0	69.2	- 9.2	60.0
45	0	65.4	- 3.4	62.0
65	0	62.0	- 9.1	52.9
85	0	60.0	- 4.6	55.4
110	0	57.9	+ 7.5	65.4
150	0	55.4	+ 13.8	69.2
190	0	52.9	+ 5	57.9

Table 9-3: Source and receiver levels of the experiment stimuli. 'Non-conflicting cues' and 'conflicting cues'-signals are listed separately. Note that in the case of the 'non-conflicting'-signals the source level is constant and the receiver level decreases with increasing distance.

The levels for the 'conflicting cues'-signals were assigned according to a special scheme. Thus, they were not truly random, but arose from a permutation of all receiver levels. In Table 9-3 the relevant source and receiver levels for both types of signals are shown.

9.4.5 Experimental setup

The test geometry has already been introduced in section 9.2. The experimental setup for the test with the real sources is shown in Figure 9-12. The experimental setup with the WFS array can be seen in Figure 9-14. The picture is taken from another perspective. However, the geometry for both test setups was the same except for the reproduction systems. A curtain was set up to hide the active loudspeakers and their position. The curtain consisted of an acoustically transparent material. As the listeners were seated at a distance of 1.25 m from the array, the 6^{th} and 7^{th} test source (d = 1.5 and 1.9 m) were synthesised *behind* the array. The subject was located at a seat behind the curtain, see Figure 9-13.

9.4.6 Elicitation of responses

Various methods for the elicitation of distance judgments from listeners are used in the literature (see Nielsen, 1991; Shinn-Cunningham, 2000; Brungart and Rabinowitz, 1999a; Zahorik, 2002). This difficult task has been realised in the past, for example with some visible dummy loudspeakers which have to be selected by the test panel after the test sound is heard. Through this method the test results are shifted towards these loudspeaker positions ('ventriloquism effect'). To avoid this effect, a graphical elicitation method is sometimes used and the test panel is requested to draw the perceived source position on a response sheet. However, the relationship between the perceived acoustical event and the drawn figure is not straightforward.

Another specific problem in the investigation of this chapter is the orientation of the head, which was chosen to be perpendicular to the source direction (see Figure 9-1). Thus, a measurement scale cannot be installed in the direction of the notional sources. Their visual representation in the listener's line of sight is necessary.

For these reasons it was decided to apply a different method. A custom-built cableway equipped with a movable, silent dummy loudspeaker in *front* of the listener was used. Pictures of this set-up are shown in Figure 9-13 and Figure 9-14 above. After each test signal, the listener had to adjust the distance of the dummy loudspeaker so that it matched the apparent distance of the auditory event. A laser beamer installed on the dummy loudspeaker indicated

the adjusted distance on the curtain which could then be recorded by the experiment supervisor. In a preliminary test the validity of this method was verified.

Figure 9-12: The left side of the curtain: The single real loudspeaker at a distance *d*

Figure 9-13: The right side of the curtain: The dummy loudspeaker 'cableway' is used to indicate the perceived distance

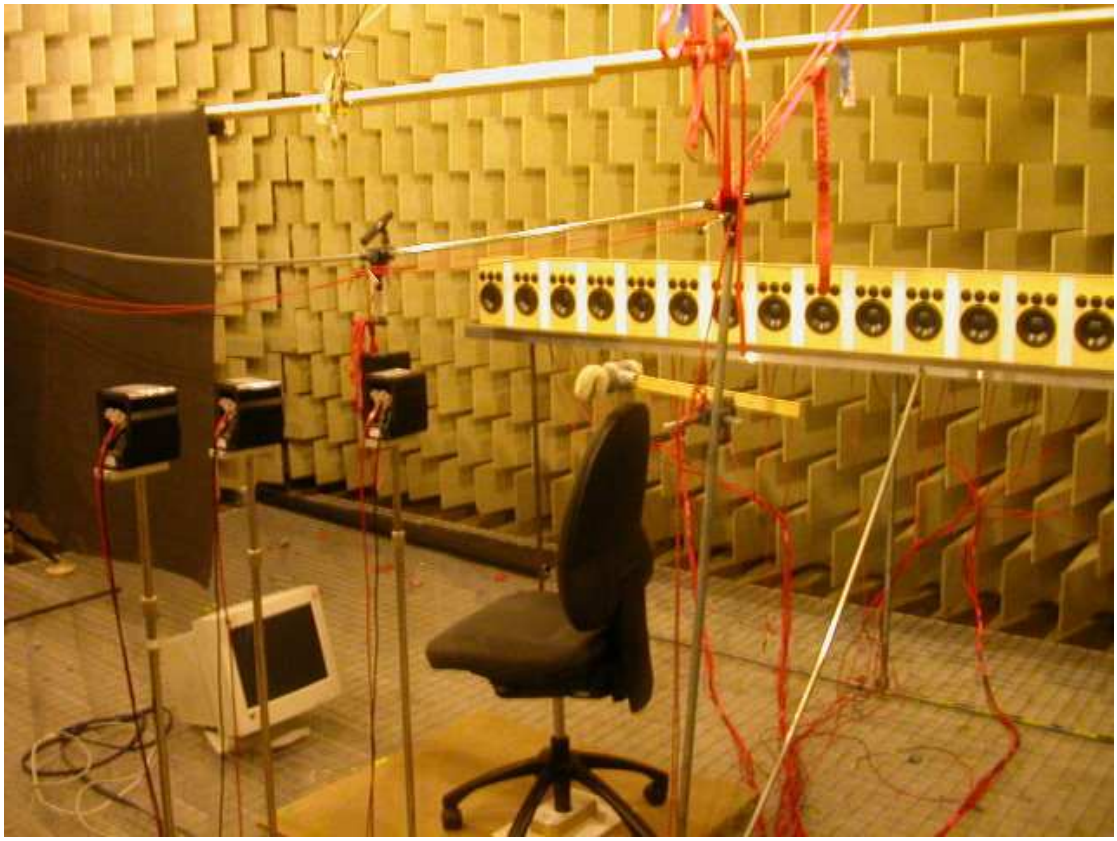


Figure 9-14: View of the experimental setup with the WFS array installed. The curtain was shifted aside for the photo in order to enable the view of the array. In the foreground three loudspeakers are visible with which the preliminary training phase was performed. Pictures from (Kerber, 2003).

9.4.7 Training of participants

The task required of the listeners was challenging. Therefore it was necessary to make them sensitive to the audible changes as caused by varying the distance of a source. In a short training session before both listening tests, the subjects were presented with a small set-up of three visible loudspeakers, located at distances d = 50, 80 and 110 cm (see Figure 9-14). They were requested to toggle between the three loudspeakers by pressing one of three keys on a keyboard. When one loudspeaker was selected, the test sound (dry orchestral music) was reproduced only through this loudspeaker. The reproduction level was randomised each time the key was pressed. The range in which the random level was chosen was adjusted for all loudspeakers so that the different distances could not lead to different receiver levels.

Initially, the participants were fairly confused by the fact that the visually perceived distance of the loudspeakers did not correspond to their auditory perception regarding the source level. The levels seemed to change randomly and could not be used for a distance judgment. In this way the listeners learned to listen for other existent acoustic cues. After some time (2-3 minutes) all participants who were included in the experimental evaluation reported that they were able to use non-level cues for distance judgment.

9.5 Listening test 1: distance perception of nearby real sources

Figure 9-15 and Figure 9-16 show the results of the first distance perception experiment. The results of all selected participants are plotted in the form of a histogram. The darkness and size of the grey boxes indicate the number of results combined in a certain distance range. The red graph shows the mean of these results and the blue graph (which belongs to the blue y-axis on the right) indicates the relevant receiver level of the reference sources in a reverse axis style. The distances are plotted on a log-log scale according to the properties of the auditory system.

It can be see from Figure 9-15 that the natural test signals are perceived quite consistently, containing an overestimation of source distances $d < 1$ m and an underestimation of distances $d > 1$ m. This under and overestimation of distances is well known from literature (Zahorik, 2002).

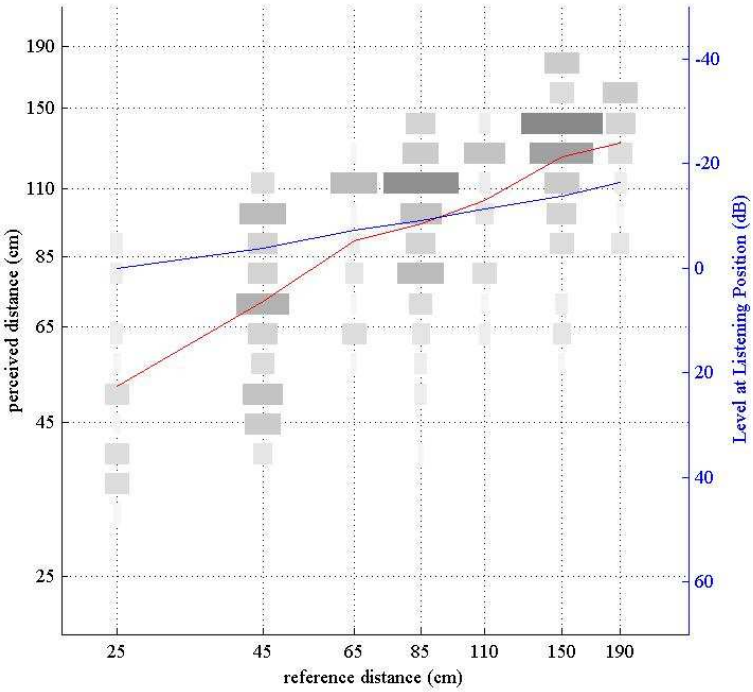


Figure 9-15: Real sources, natural cues

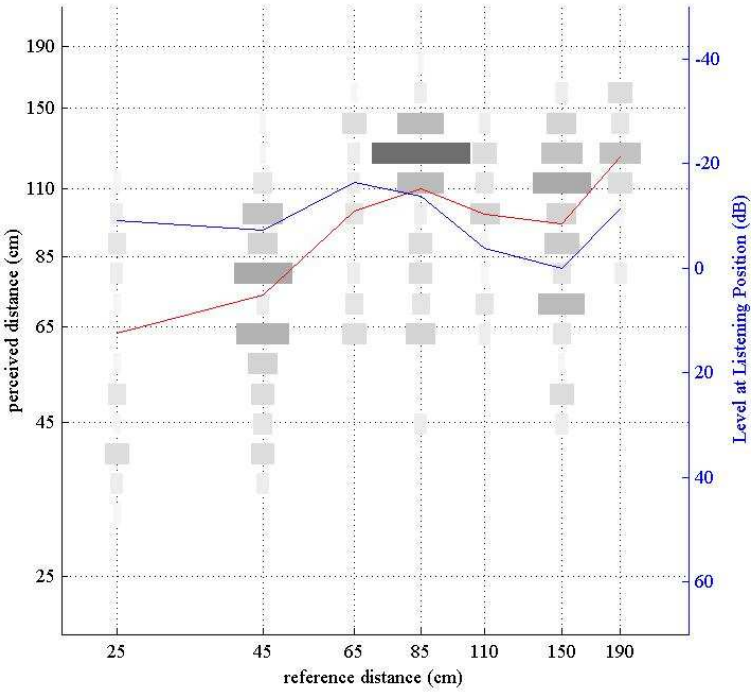


Figure 9-16: Real sources, conflicting cues

Figure 9-16 shows the result for the ‘conflicting cues’ test signals. The blue curve indicates the permuted receiver level values. The results in this figure can be split into two regions:

For distances $d > 1$ m there is virtually no relationship between the perceived and the reference source distance. Instead, the perception is determined by the respective receiver level as can be deduced from the similarity of the blue and red curve.

For distances $d < 1$ m a certain correlation between perceived and reference source distance is observed whereas the receiver level is less relevant.

These observations lead to the following conclusions:

- Apparently a certain perception of distance is possible due to the binaural cues contained in the direct sound only.
- It appears that the upper limit of distance perception due to binaural cues is at about 1 m.
- The results are similar to the results of Brungart and Rabinowitz (1999c), who measured the region of $d < 1$ m.

9.6 Listening test 2: distance perception of nearby virtual sources

Figure 9-17 and Figure 9-18 show the results for the virtual sources. In Figure 9-17 the reproduced receiver level corresponds to the reference source distance. Now, in contrast to the real sources (see Figure 9-15), the differences between all perceived distances are much smaller. The degree of over- and underestimation respectively is significantly higher. Although the graph increases monotonically, its gradient is smaller, indicating a loss of auditory cues for distance perception. Additionally, the actual distance of the WFS array (1.25 m) could play a certain role.

The results of the test using the ‘conflicting cues’-signals are plotted in Figure 9-18. Once again, the range of the perceived distances is fairly small. The results make clear that the receiver level (level at the listening position) is crucial for the perceived distance. There is no relationship between perceived and reference source distance. Instead, the correlation between perceived distance and receiver level is high.

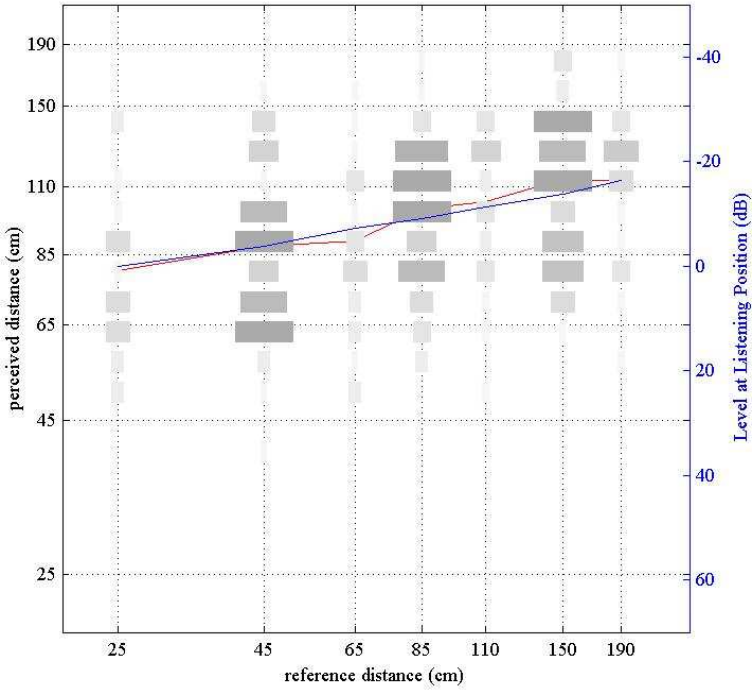


Figure 9-17: Virtual sources, natural cues

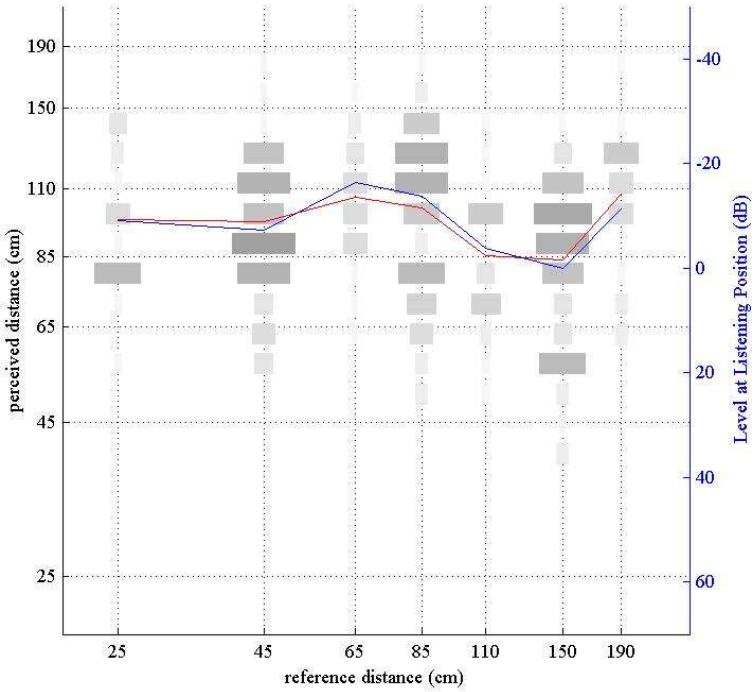


Figure 9-18: Virtual sources, conflicting cues

In Figure 9-19 and Figure 9-20, the data of Figure 9-16 and Figure 9-18 are plotted once again. This time, the results are sorted according to the receiver level to check the correlation. Obviously, the correlation in the case of the WFS virtual sources is high. Note that Figure 9-17 and Figure 9-20 look very much the same. This indicates that no auditory distance perception cue exists that conveys the actual distance of the virtual source.

This means that at a fixed listening position, the curvature of the wave front of dry WFS virtual sources is irrelevant for distance perception. This is true for the virtual sources created in the experiment and may be generalised to other array and signal conditions as long as the conditions that cause this (analysed in chapter 9.3) do not change. The discussion in section 9.7 will show whether the length of the array plays a role for the creation of realistic ILD.

A solution for the problem of reduced spatial amplitude decay with linear WFS arrays could be an extension of the WFS array into two dimensions, such that it covers a whole plane. In that case, the amplitude distribution could be optimised (see chapter 9.3.1) and the preconditions for auditory distance perception could be improved. The investigation by Komiyama et al. (1991) uses such an array-design for an investigation into distance perception. They indeed detected a successful distance perception of virtual focussed sources. Their conclusions, however, are not deduced from experiments under the same rigorous conditions (no isolation of the level cue).

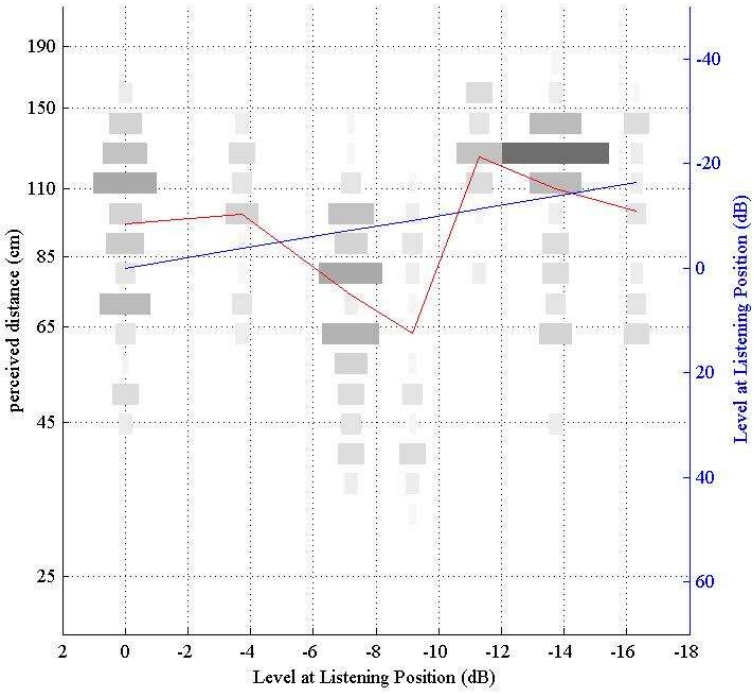


Figure 9-19: Real sources, conflicting cues, sorted by the level at the listening position (same data as in Figure 9-16)

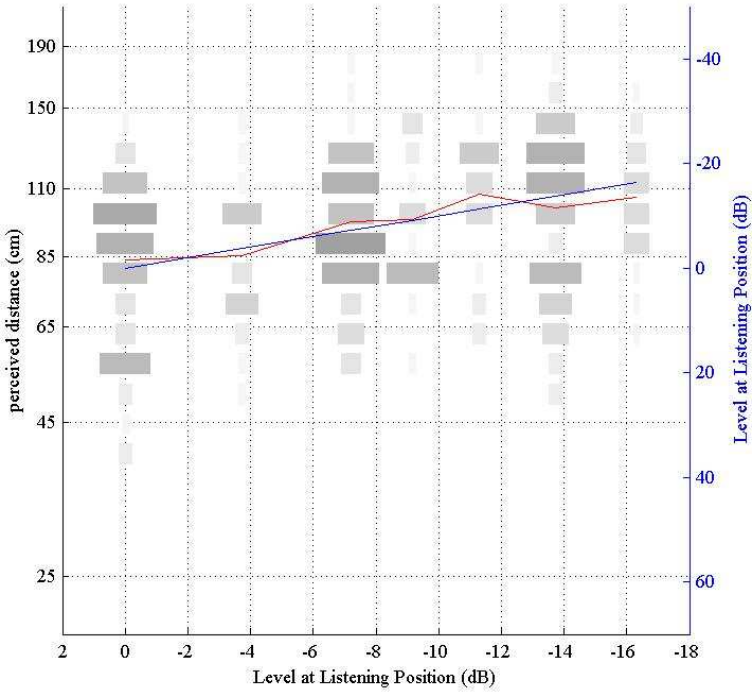


Figure 9-20: Virtual sources, conflicting cues, sorted by the level at the listening position (same data as in Figure 9-18)

9.7 Head shadowing effects in WFS

The simulations and experiments described in the preceding sections showed a lack of cues for distance perception cues in the sound field of focussed sources in WFS. In this section, further focus is put on the effect of the head in the WFS sound field. The following simulations give further opportunities to analyse the failures described.

9.7.1 Isolation of the impact of head shadowing

In the following simulations, the ILD is considered to be a result of two different influential aspects:

a) Level differences due to different distances of the two ears to the source (max. 17 cm). This level difference is called 'No-Head-ILD'.

b) Pinna effects, influence of the ear canals and 'head shadowing'. They are, for the sake of simplicity, called 'head shadowing effects'.

The 'No-Head-ILD' can be calculated easily, as shown in chapter 9.3.3. The influence of the head shadowing on the other hand, can be deduced from measurements of ILD and 'No-Head-ILD'. In the approach of this investigation, it is mathematically derived from simply subtracting the 'No-Head-ILD' from the ILD. In other words, the 'No-Head-ILD' and the head shadowing effect add up to the ILD.

Although this is a very simple approach, it offers an opportunity to compare real and virtual sound field.

9.7.2 Simulation of a long array

As shown in the previous sections, a linear WFS array does not create sufficient ILD for the listener to create a cue for distance perception. One of the main shortcomings of the test setup is the limited array length of 2.55 m, leading to a significant loss of level for lower frequencies as mentioned in chapter 9.3.1, and as can be seen in Figure 9-4. Furthermore, for low frequencies, not only the level decreases, but also the level *differences* between different positions in the sound field vanish. Although the array size of the experiment setup is quite typical, it remains interesting whether a significant ILD could be synthesised by a longer array.

To prepare for the discussion in the next section, a simulation of a long array, the so-called 'super-array', is performed in this section. A simulation setup was created using an array of

length 21.3 m and a decreased interspacing of 17 cm / 4 = 4.25 cm. This results in a number of array loudspeakers of n = 501.

This new simulation setup enables a further view of the characteristics of the WFS sound field. The ‘super-array’ shows low frequency artefacts as well but it is capable of reproducing a flat frequency response for focussed sources for frequencies above approximately 1 kHz. This can be seen from Figure 9-21. This figure can be well compared with Figure 9-7, where the responses of the normal short WFS array are shown. With a longer array, at the price of additional ripples, a better reproduction of lower frequencies is achieved.

As a result, significant level differences can also be produced at lower frequencies (Figure 9-22, compare with Figure 9-8). According to the theoretical considerations of chapter 9.3.1., the level differences are smaller than those caused by real sources (see Figure 9-6).

This ‘super-array’ is the basis for the discussion of the next subsection.

9.7.3 Simulations of head shadowing

With the help of the parameter ‘head shadowing effect’, the influence of the head being present in the sound field can be studied. ILDs derived from measurements with a dummy head being in the sound field of a real source in an anechoic chamber are shown in Figure 9-23 (this figure is a repetition of Figure 9-9; the head is oriented perpendicular to the direction of the source).

The head shadowing effect, derived by calculating the difference between the ILD and the ‘No-Head-ILD’ is shown in Figure 9-25. It can be seen that the head shadowing effect, similar to the ILD in general, differs significantly only for very close sources (< 65 cm). Presumably, when the source is close to the head, head *diffraction* differs significantly compared with that of a more distant source.

In the frequency band from 1 to 5 kHz, the head shadowing effect can create an additional level difference of approximately 5 dB for a source in 45 cm distance. It is plausible that this level difference can serve as an auditory cue.

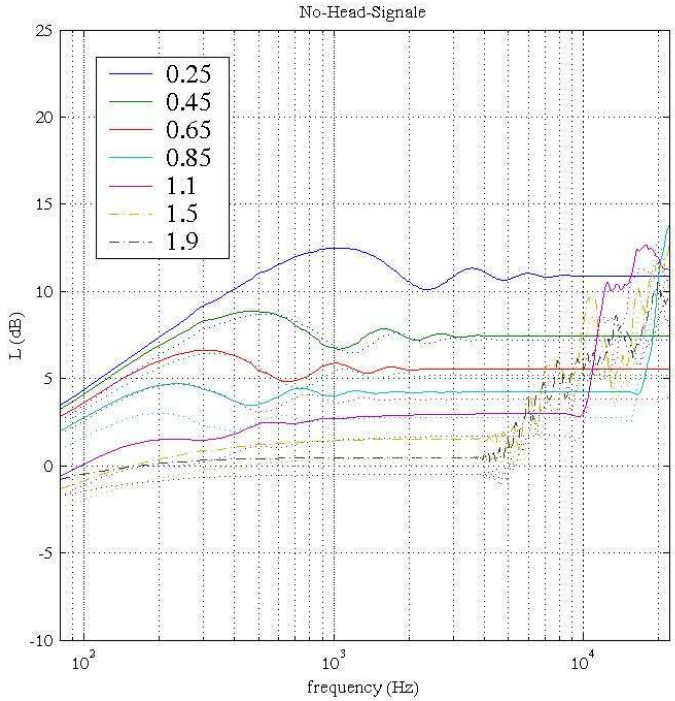


Figure 9-21: Super-array: Level of a focussed WFS virtual source at distances = *d* ± (ear distance/2). Solid line: right 'ear', dotted line: left 'ear'. Dash-dotted lines: non-focussed sources. Source at 90°.

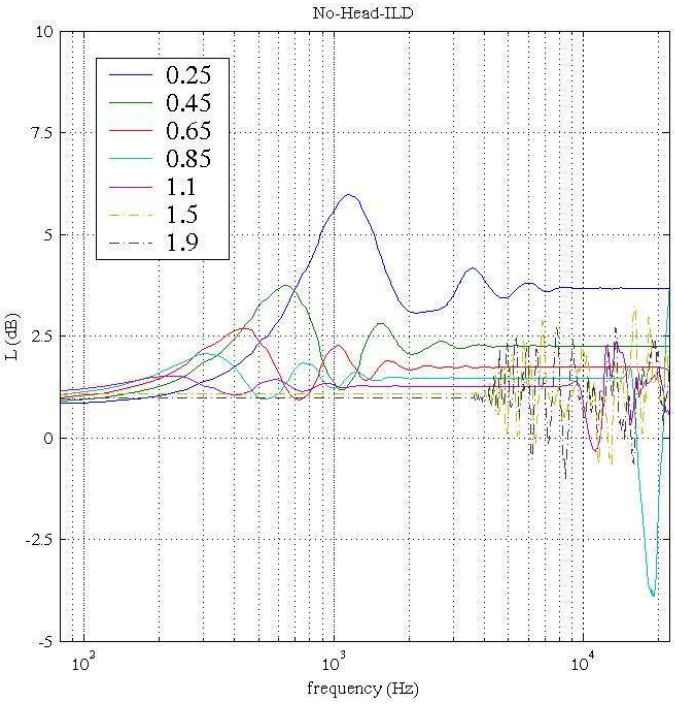


Figure 9-22: Super-array: 'No-Head-ILD': level difference ΔL between ear positions in the sound field of a focussed WFS virtual source at distance *d*. Dash-dotted lines: non-focussed sources. Source at 90°.

The virtual sources are analysed in Figure 9-24 and Figure 9-26. Figure 9-24 shows the ILD for various source distances derived from virtual sources. It can indeed be seen that the ILD increases with decreasing distance, albeit not as strongly as for the real sources, which are analysed in Figure 9-23. This corresponds to the results of the previous figures. The differences from Figure 9-22 can be identified in Figure 9-24 rather well.

A further view on the virtual source's characteristics is offered in Figure 9-26. Here, the head shadowing effect is presented. In contrast to the real source, the head shadowing effects of the virtual sources show nearly no dependence on the source distance. The ripples, which could already be seen in the ILD and the 'No-Head-ILD', are still existent, albeit significantly damped. They could perhaps be a consequence of inexact measurements, but this cannot be deduced from these simulations.

It is surprising that the head shadowing effect, apart from the small ripples, is the same for all source distances. Although derived from different test setups (e.g. different loudspeakers at the measurements), a comparison of Figure 9-25 and Figure 9-26 suggests that all virtual sources create the same head shadowing, which is the head shadowing of a real source at a distance of $d > 1$ m. This leads to the following possible assumption: the head shadowing effect of a focussed virtual source depends only on the distance of the *reproduction array*.

This does not mean that the ILD are the same for all virtual source distances. It can be seen from Figure 9-24 that indeed certain differences due to the source distance are present. It may be doubted whether these differences are big enough to serve as an auditory cue. However, they are bigger than in the setup that was used in the experiments of this chapter.

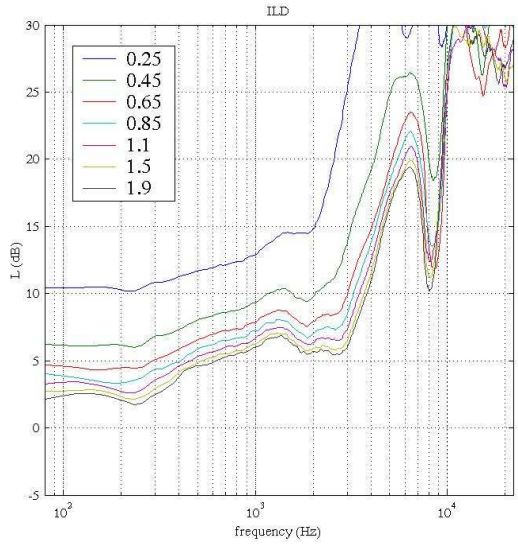


Figure 9-23: Interaural level difference ILD in the sound field of a real source at distance *d*. Source at 90°.

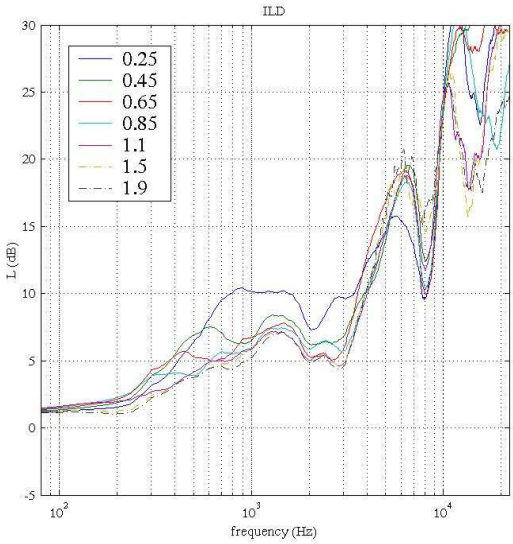


Figure 9-24 : Super-array: interaural level difference ILD in the sound field of a focussed source at distance *d*. Source at 90°. Dash-dotted lines: non-focussed source

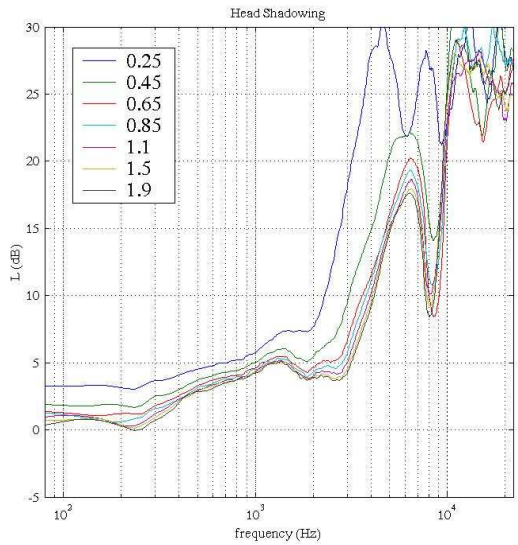


Figure 9-25: Head shadowing effect in the sound field of a real source at distance *d*. Source at 90°.

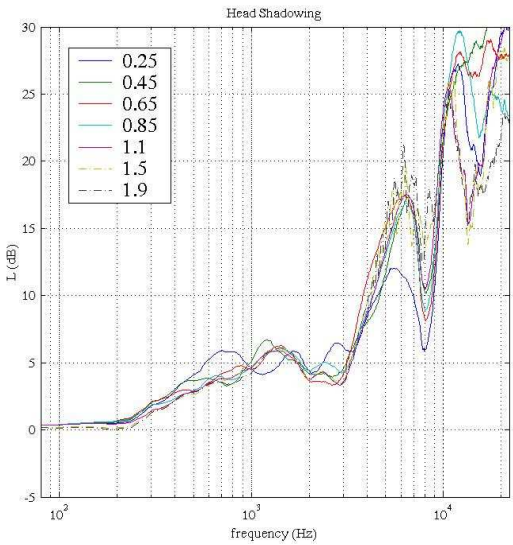


Figure 9-26: Super-Array: head shadowing effect in the sound field of a focussed source at distance *d*. Source at 90°. Dash-dotted lines: non-focussed sources

9.8 Summary of chapter 9

The properties of WFS focussed sources with regard to distance perception were considered in theory and through listening tests. Although a number of cues mentioned in chapter 9.1 influence distance perception, this investigation focussed on highlighting the existing differences between WFS and stereo at a fixed listening position, namely the cues related to the wave front curvature. Both theoretical and practical investigations showed that WFS fails to reproduce these cues. Nearby sources, synthesised as dry focussed sources in WFS, do not give rise to correct distance perception when the level cue is factored out.

The rationale for this failure is on the one hand an insufficient focussing of low frequencies which causes the absence of crucial low-frequency ILD. On the other hand, head shadowing was insufficiently synthesised in the simulations which could point to a general failure of WFS.

One way to balance the described deficiency of WFS to produce ILD for distance perception is to apply natural acoustics to the virtual source. These additional cues can possibly make up for the lack of binaural direct sound cues. However, disturbing reflections caused by the WFS array itself may hinder the perception of the distance of virtual sources in front of the array. Another way to reduce faulty distance perception is to allow the listeners to move within the listening area where idiothetic cues could improve distance perception (see chapter 6.5).

10. Summary, conclusions and outlook

10.1 Introduction

This research was motivated by the need to compare WFS and stereo in general, and for certain applications in particular. WFS has already been used in various applications, and the need for a sound reconstruction technique such as WFS has not always been apparent. Thus, a direct comparison between WFS and stereo seemed necessary in order to provide a solid basis for the choice of one of the two alternatives. This thesis has addressed some important aspects of such a comparison.

The most important differences between WFS and stereo were identified and discussed. These were found in the quality of the directional imaging, in sound colour reproduction and distance perception. Furthermore, only WFS has the capability of reproducing a sound field with a real acoustical perspective, a property which is important for a moving listener. This difference is so clear from a theoretical examination of the physical and perceptual principles that it does not need to be discussed in the same depth, in the opinion of this author. This capability opens WFS for applications in which stereo cannot be used at all. Thus, the comparison in this thesis is focussed on the perceptual properties for a non-moving listener.

The thesis started with an introduction chapter 1, which set the scene for the theoretical and experimental comparison of the two systems. In chapter 2 the main attributes of localisation, sound colour and distance perception were discussed and a nomenclature was defined in order to prepare the ground for the discussion in the other chapters.

In the following sections a summary and conclusions of the research is provided for each main topic of the thesis. Section 10.7 gives an outlook on possible further work in the field.

10.2 Pre-existing knowledge on the perceptual properties of WFS and stereo

The discussion in chapters 3 and 4 summarised the current status of knowledge on the perceptual properties of these reproduction techniques. The focus was put on the attributes that were examined in the course of the thesis, more precisely the attributes of localisation, sound colour and distance perception. In addition, the physical properties were reviewed, as well as the general perception principles which are the basis for a discussion of perceptual capabilities.

By exploring similar and dissimilar properties, which was continued in chapter 6, a direct and structured comparison was enabled.

Perception mechanisms

A fundamental difference between the two sound reproduction techniques may be found in the general perception mechanism (chapter 3.6). Two different approaches to explaining stereophonic perception were introduced, more precisely the theory of 'summing localisation' (e.g. Blauert, 1997 and the 'association model' by Theile, 1980). The summing localisation theory assumes a physical synthesis of the loudspeaker signals in order to create a substitute source that physically resembles a real source regarding the essential localisation cues. This theory assumes the principle of perception to be the same for WFS and stereo. It was shown to what extent a synthesis as proposed by this theory succeeds, and on the other hand what contradictory results are obtained. The latter include, in short, the missing explanation for the suppression of perception of comb filtering in stereophonic hearing. Furthermore, the perceived phantom source direction, in particular in the case of interchannel time differences, cannot be predicted nor explained sufficiently by summing localisation. The association model can offer an explanation for these phenomena, because it assumes the presence of two different processing mechanisms. The first stage is able to separately locate the individual loudspeaker signals by a comparison of the ear signals with a known pattern. The second stage will fuse their coherent signals after a location dependent inverse filtering. Hence, the ear signals are not directly evaluated for localisation and sound colour perception.

The fundamental perception mechanism has important consequences for the prediction of the perceptual properties of stereophonic listening. If summing localisation were assumed, stereo would not perform as well as WFS. If a complete functioning of source localisation after the association model were assumed, stereo would partly perform even better than WFS, because with regard to perception it would not suffer from interferences of the loudspeaker signals as in WFS.

Localisation

Existing literature suggests that the localisation properties of stereo differ significantly compared to those of natural sources. The phantom source was reported to be less focussed and the locatedness was clearly lower (e.g. Martin et al., 1999b; Silzle and Theile, 1990). In spite of that, directional imaging between the loudspeakers is possible with sufficient accuracy for many applications. The localisation properties of WFS, more precisely its directional accuracy and perceived image focus, were reported to be nearly as good as for real sources as long as the spatial aliasing frequency was above a limit of 1.5 kHz (Start, 1997).

Sound colour

The perceived sound colour of a phantom source is known to be significantly different from that of a natural source (Silzle and Theile, 1990). Despite that, in the experience of sound engineers the timbral difference between adjacent source directions on the same stereo loudspeaker setup is rather small (Wittek, 2000a). This is different in WFS, where movements of source or listener cause an audible change in the spatial aliasing (Start, 1997). Both the spatial aliasing frequency and the reproduction room influence the perception of colourations in WFS (Start, 1997; de Bruijn, 2004).

Distance and depth, acoustical perspective and listener movements

The capabilities of reproducing a sound field with depth were discussed on the basis of a general differentiation between two different types of listening area (chapter 6.4). These two types correspond to a reproduction with and without a real acoustical perspective which enables the listener to move in the sound field and which creates an accurate sound image for listeners at different locations. WFS has the ability to create this real acoustical perspective and thus enables listener movements and a corresponding change of the individual 'view' angle on the acoustical scene. Stereo can produce only one 'view' angle of the acoustical scene. Different types of distance perception cues were also differentiated. A successful distance perception was asserted to be produced by cues that are essentially available in stereo as well, more precisely the monaural cues (level, direct-to-reverberant energy ratio, etc) and the binaural cues (reflection pattern). A difference in these cues potentially exists only due to the 'binaural differences' cue, which is a cue related to the wave front curvature. That is produced only in WFS.

10.3 The OPSI method

A new method of avoiding spatial aliasing in WFS was proposed in chapter 5. The idea of OPSI ('Optimised Phantom Source Imaging in wavefield synthesis') is the substitution of the high-frequency contributions of the WFS array by stereophonic reproduction. The OPSI method avoids the effects of aliasing and thus was assumed to result in advantages regarding the perceived colouration of the reproduced sound field. A pilot experiment determined the maximum allowed deviation between the perceived directions of the low and the high frequency contribution in an OPSI signal, which was called the OPSI localisation error. Practical results on the perceptual performance of the OPSI method were obtained in the experiments summarised below.

10.4 Experiment 1 on localisation properties

After defining a set of research questions, which were discussed in chapter 6.2, the subsequent chapter 7 described an experiment on the localisation properties of WFS, OPSI and stereo. The experiment included measurements of the perceived azimuth and elevation directions of various sources which were reproduced by single loudspeakers, different WFS arrays, and a stereo setup in an anechoic chamber. In addition, the attribute locatedness (=the spatial distinction of a perceived source) was elicited by direct subjective assessments in the listening test.

The directional accuracy of all systems was confirmed. Only small differences in the standard deviations of the perceived directions could be found from which differences in the image focus could have been deduced. A significantly larger standard deviation was found only for the phantom sources.

Clearer differences between the systems were found with regard to the attribute locatedness. These differences led to the conclusion that the localisation performance of natural sources can be considered as a reference which cannot even be achieved by the WFS system with an aliasing frequency as high as 7.5 kHz (loudspeaker spacing $\Delta x = 4.2$ cm). In spite of that, this WFS system was still significantly better than the WFS system with an aliasing frequency of 2.5 kHz ($\Delta x = 12.7$ cm) with regard to the assessed locatedness. This confirms the assumption that an increase of the aliasing frequency above the previously proposed limit of 1.5 kHz (Start, 1997) leads to significant improvements in the localisation performance. Spatial aliasing seemingly has an impact on the localisation performance although the dominant contributions for localisation were identified to be in the lower frequency bands (Wightman and Kistler, 1990).

The locatedness of the OPSI system was graded similar to its corresponding WFS system with ($\Delta x = 12.7$ cm). It seems that omitting the aliased contributions and replacing them with phantom sources has no negative effect on the localisation performance. This validates the OPSI concept, at least regarding the properties of directional imaging. The locatedness of the phantom sources was graded worst, which also corresponds to the larger standard deviations of the perceived directions mentioned above.

10.5 Experiment 2 on sound colour properties

The experimental investigations on the sound colour properties of WFS, OPSI and stereo were described in chapter 8. Its research questions were defined in chapter 6.3. The experi-

ment was performed using BRS, which is a virtual acoustic system including head-tracking. In this way it was possible to simulate arbitrary WFS systems, even an ideal WFS system exhibiting a loudspeaker spacing as low as 3 cm. The subjects were able to switch between sources at different locations to evaluate the perceived sound colour difference between these sources. In this way, the perceived colouration occurring within one system was obtained. The colouration was graded with the help of a multiple stimulus graphical user interface employing reoccurring anchors which spanned a range of different colourations. Thus, the measured colouration was referred to the same scale, and a comparison could be made between the results for the different systems.

Again, the OPSI concept could be validated, because the colouration of the OPSI system was smaller than that of the conventional WFS systems. The choice of the crossover frequency was shown to be essential because both too high a crossover frequency, causing spatial aliasing, as well as too low a crossover frequency, unnecessarily omitting correct WFS contributions, causes an increase of the perceived colouration. As a conclusion, it was shown that the crossover frequency must be above the aliasing frequency. Furthermore, with regard to colouration, the crossover frequency may be as low as 3000 Hz without causing negative consequences as long as the aliasing frequency is above 3000 Hz.

The effect of spatial aliasing on the perceived colouration could clearly be shown. A decrease of the spatial aliasing frequency led to an increase of the colouration. A theoretical prediction of the colouration based on the spectral alterations of the ear signals was successful. This means that the frequency spectra of the ear signals govern the perception of the sound colour. For the WFS systems a good prediction of the colouration was achieved whereas the systems containing stereophonic contributions and lacking spatial aliasing were mostly overestimated in their colouration. In other words, the perceived colouration was smaller than predicted. Indeed, the best OPSI systems and pure stereo were perceived with the least colouration, which was as low as the colouration of the ‘ideal’ WFS system ($\Delta x = 3$ cm) and the natural reference sources.

The good results for the stereophonic systems led to the conclusion that some kind of partial decolouration is active in stereophonic perception, be it a decolouration as proposed in literature in the context of early reflections (Salomons, 1995; Brüggen, 2001a, 2001b) or based on the association model of Theile (1980). The comparison of perceived and predicted colouration showed that decolouration could improve the perceived sound colour substantially. In spite of that, some dependence on the spectral alterations of the ear was still found. Hence, a stringent functioning of a decolouration process based on the association model could not be proven.

10.6 Experiment 3 on the effect of the wave front curvature in WFS

The experiment described in chapter 9 investigated the role of the wave front curvature in WFS for a listener at a fixed listening position. A listening test was performed employing both real as well as WFS virtual sources at close distances ($d \geq 25$ cm) from the listener. In WFS, focussed sources (=sources in front of the array) were used. The subjects were trained to use the 'binaural differences' cue for distance perception in advance of the experiment. The level cue could be isolated from this cue by utilising a test method of 'conflicting cues'.

The experiment was performed in an anechoic chamber. The head of the subjects was oriented perpendicular to the direction of the (hidden) source in order to provide maximal binaural differences. A special cableway equipped with a dummy loudspeaker was available in front of the listener to be able to elicit distance judgements from the subjects. The subjects could pull wires to position the dummy loudspeaker at a continuously variable distance which was to correspond to the perceived distance of the reproduced source.

It was shown that - as reported in literature (Brungart and Rabinowitz, 1999c) - distance perception is possible for nearby real sources due to binaural differences. The low-frequency ILD significantly changes with decreasing source distance and can be evaluated. Simulations showed that the ILD depends on the distance-dependent head shadowing as well as on the distance-dependent level difference due to the $1/r$ -law.

The focussed WFS sources of the experiment could not create a correct distance perception. The distance judgements were created solely by the level cue regardless of the reproduced source distance. Simulations showed that indeed the crucial cues for distance perception of nearby sources do not exist in the sound field produced by focussed sources. The reason is that acoustical focussing is restricted to a minimum size of the focal point of half the wavelength. Hence, the important low frequency ILD cues cannot be reproduced sufficiently. Furthermore, it is suspected that in WFS the head shadowing is ruled by the distance of the array and not by the actual source distance.

10.7 Outlook on possible further work in the field

Perception mechanism

This thesis aimed at identifying and investigating aspects in the comparison of the perceptual properties of WFS and stereo that had not been investigated sufficiently before. A number of disagreeing properties were indeed found and were investigated in a direct comparison. Starting from this comparison of perceptual properties, the research also tried to find rationales for

the observed phenomena. However, the proposed explanations could not sufficiently be proven and can only be regarded as hypotheses. A functioning of Theile's association model or any other mechanism that results in a decolouration of stereophonic signals seems rather plausible after studying the results of this thesis. This research, however, was not primarily aimed at exploring the fundamental perception mechanisms. A specific investigation in that direction would be better suited to illuminate the validity of the mentioned hypotheses. The association model is a rather broadly defined concept which has to be translated to the more specific processes in spatial perception and their consequences on the perceptual properties. As another option, other rationales could exist for an improvement of the perceived sound colour. The hypothesis of a binaural decolouration was mentioned in the literature in the context of early reflections (Salomons, 1995; Brüggen, 2001a, 2001b). An investigation incorporating stereophonic reproduction in that context could give rise to similar results regarding the binaural advantage in sound colour perception.

Perceptual properties of WFS

This investigation dealt with the properties of WFS with regard to localisation and sound colour at the same time. These two groups of properties are strongly related to each other, because any change in the reproduction of WFS, be it a method such as OPSI or the diffusion of the WFS driving functions (chapter 4.2.5), has an impact on both the perceived attributes of localisation as well as the sound colour. Hence, an improvement of WFS may result from a trade-off regarding the quality of these attributes. Further investigations that may have to balance different alternatives should bear in mind the general priority which is given to sound colour in spatial perception (Rumsey et al., 2005). Hence, further attempts at avoiding the perceptual artefacts of spatial aliasing may be most promising. The OPSI method is one approach in that direction, but it is not the only option.

The literature has considered the artefacts of WFS (spatial aliasing, diffraction effects, reduction of the reproduction dimensions to the horizontal plane) in order to find improvements for practical applications. However, the influence of the reproduction room on spatial perception has been dealt with less thoroughly, although it is considered by this author to be one of the main reasons for an impaired spatial perception in WFS. Some attempts have been made to use the WFS array to cancel distinct reflections in the horizontal plane, but the possibilities are limited and the general problem of faulty reproduction room reflections from floor and ceiling cannot be solved in this way (e.g. Spors, 2006; Corteel and Nicol, 2003). Hence, as a result, the virtual source is often perceived at the distance of the array regardless of its synthesised physical distance. As mentioned in this thesis (chapters 2.5 and 6.5) the difference between WFS and stereo regarding the capabilities of reproducing source distance are smaller

than generally assumed due to this problem. WFS, however, has been targeting an essential enhancement of the spatial reproduction. Solutions for this inherent problem have to be found when the full potential of WFS for the reproduction of depth, distance and acoustical perspective is to be demonstrated.

Comparison of WFS and stereo

New stereophonic formats such as 7.1 or 22.2 (Hamasaki et al., 2006) are being discussed at present, which employ an increased number of loudspeakers that are partly located in a second, elevated plane above the listener. These formats allegedly have the potential to produce a sound field exhibiting an enhanced spatial quality for multiple listeners. These formats suggest themselves as an alternative to WFS as long as no real acoustical perspective is desired.

As discussed in this thesis, in principle no disadvantages can be identified in stereophony for the creation of a spatial sound field exhibiting accurate depth and distance as well as accurate properties of directional imaging and sound colour reproduction. However, investigations on suitable methods for a practical implementation of an enhanced spatial sound field based on these formats and on the related bottlenecks are still lacking. These investigations can not only be performed by sound engineers during practical work, but rather psycho-acoustic principles have to be applied and explored and thus research has to be carried out in this field. In a number of applications, these formats appear to have a greater potential than WFS to efficiently reproduce a spatial sound field in an extended listening area.

References

American Standard Association ASA (1960) *'American Standard Acoustical Terminology'*. New York, Definition 12.9, Timbre, 45.

Atal, B.S., Schroeder, M.R., Kuttroff, K.H. (1962) 'Perception of coloration in filtered gaussian noise-short-time spectral analysis by the ear'. *Proceedings 4th Int. Congress on Acoustics ICA 1962, Copenhagen, Denmark*, August 1962.

Augustin, T. (2004) *'Zur Wahrnehmbarkeit von Klangfarbenveränderungen bei Wellenfeldsynthese'*. Diploma thesis, Technical University Munich, Germany, August 2004.

Becker-Carus, C. (2004) *'Allgemeine Psychologie'*. Munich, Germany: Spektrum Akademischer Verlag, Elsevier.

Berkhout, A.J. (1987) *'Applied Seismic Wave theory'*. Amsterdam, The Netherlands: Elsevier, 1987.

Berkhout, A.J. (1988) 'A holographic approach to acoustic control'. *Journal of the Audio Engineering Society*, Vol.36, No.12, December 1988, pp.977-995.

Berkhout, A.J., de Vries, D., Vogel, P. (1992) 'Wave front synthesis: a new direction in electro-acoustics'. *Proceedings 93rd AES Convention, San Francisco, California*, September 1992, Preprint No.3379.

Berkhout, A.J., de Vries, D., Vogel, P. (1993) 'Acoustic control by wave field synthesis'. *Journal of the Acoustical Society of America*, Vol.93, No.5, May 1993, pp.2764-2778.

Bilsen, F.A. (1977) 'Pitch of noise signals: Evidence for a "central spectrum" '. *Journal of the Acoustical Society of America*, Vol.61, No.1, January 1977, pp.150-161.

Blauert, J. (1997) *'Spatial Hearing'*. Cambridge, Massachusetts: MIT Press, 1997.

Bleistein, N. (1984) *'Mathematical methods for wave phenomena'*. New York: Academic Press Inc., 1984.

Bloothooft, G. and Plomp, R. (1988) 'The timbre of sung vowels'. *Journal of the Acoustical Society of America*, Vol.84, No.3, September 1988, pp.847-860.

Blumlein, A.D. (1933) 'Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems'. British Patent Specification 394,325, reprint in the *Journal of the Audio Engineering Society*, Vol.6, No.2, April 1958, pp.91-98.

Boone, M.M., Verheijen, E.N.G., van Tol, P.F. (1995) 'Spatial Sound Reproduction by Wave Field Synthesis'. *Journal of the Audio Engineering Society*, Vol.43, No.12, December 1995, pp.1003-1012.

Boone, M.M., Verheijen, E.N.G., Jansen, G. (1996) 'Virtual Reality by Sound Reproduction Based on Wave Field Synthesis'. *Proceedings 100th AES Convention, Amsterdam, The Netherlands*, April 1996, Preprint No.4145.

Boone, M.M., Horbach, U., de Bruijn, W.P.J. (1999) 'Virtual Surround speakers with wave field synthesis'. *Proceedings 106th AES Convention, Munich, Germany*, April 1999, Preprint No.4928.

Boone, M.M. and de Bruijn, W.P.J. (2003) 'Improving Speech Intelligibility in Teleconferencing by using Wave Field Synthesis'. *Proceedings 114th AES Convention, Amsterdam, The Netherlands*, February 2003, Preprint No.5800.

Boone, M.M. (2004) 'Multi-Actuator Panels (MAPs) as Loudspeaker Arrays for Wave Field Synthesis'. *Journal of the Audio Engineering Society*, Vol.52, No.7/8, July/August 2004, pp.712-723.

Born, M., Wolf, E. (1975) '*Principles of Optics'*. New York: Pergamon Press, 5th edition, 1975, pp.370ff.

Bregman, A.S. (1994) *'Auditory Scene Analysis'*. Cambridge, MA: MIT Press.

Brittain, F.H. and Leakey, D.M. (1956) 'Two Channel Stereophonic Sound Systems'. *Wireless World*, Vol.62, May 1956, pp.206-210.

Brix, S., Sporer, T., Plogsties, J. (2001) 'CARROUSO - An European approach to 3D-Audio'. *Proceedings 110th AES Convention, Amsterdam, The Netherlands*, April 2001, Preprint No.5314.

Brüggen, M. (2001a) '*Sound coloration due to reflections and its auditory and instrumental compensation'*. Dissertation Ruhr-Universität Bochum. Berlin, Germany: dissertation.de - Verlag im Internet.

Brüggen, M. (2001b) *'*Coloration and binaural decoloration in natural environments*'*. *Acta Acustica united with Acustica*, Vol.87, No.3, May/June 2001, pp.400-406.

Brungart D.S. and Rabinowitz, W.M. (1999a) 'Auditory localization of nearby sources. Head-related transfer functions'. *Journal of the Acoustical Society of America*, Vol.106, No.3, September 1999, pp.1465-1479.

Brungart D.S., Durlach, N.I., Rabinowitz, W.M. (1999b) 'Auditory localization of nearby sources. II. Localization of a broadband source'. *Journal of the Acoustical Society of America*, Vol.106, No.4, October 1999, pp.1956-1968.

Brungart D.S. and Rabinowitz, W.M. (1999c) 'Auditory localization of nearby sources. III. Stimulus effects'. *Journal of the Acoustical Society of America*, Vol.106, No.6, December 1999, pp.3589-3602.

Bücklein, R. (1981) *'*The Audibility of Frequency Response Irregularities*'*. *Journal of the Audio Engineering Society*, Vol. 29, No.3, March 1981, pp.126-131.

Corey, J. and Woszczyk, W. (2002) 'Localization of Lateral Phantom Images in a 5-channel System with and without Simulated Early Reflections'. *Proceedings 113th AES Convention, Los Angeles, CA*, September 2002, Preprint No.5673.

Corteel, E. and Nicol, R. (2003) 'Listening Room Compensation for Wave Field Synthesis. What Can Be Done?' *Proceedings 23rd AES Int. Conference on Signal Processing in Audio Recording and Reproduction*, Helsingør, Denmark, May 2003.

Corteel, E. (2006) 'Equalization in an Extended Area Using Multichannel Inversion and Wave Field Synthesis'. *Journal of the Audio Engineering Society*, Vol.54, No.12, December 2006, pp.1140-1161.

Corteel, E. (2007a) 'Synthesis of Directional Sources Using Wave Field Synthesis, Possibilities, and Limitations'. *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, Article ID 90509, 18 pages, 2007.

Corteel, E., Nguyen, K., Warusfel, O., Caulkins, T., Pellegrini, R. (2007b) 'Objective and Subjective Comparison of Electrodynamic and MAP Loudspeakers for Wave Field Synthesis'. *Proceedings 30th AES Int. Conference on Intelligent Audio Environments, Saariselkä, Finland*, March 2007.

Daniel, J., Moreau, S., Nicol, R. (2003) 'Further Investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging'. *Proceedings 114th AES Convention, Amsterdam, The Netherlands*, February 2003, Preprint No.5788.

Daniels, D. (2002) *'Kunst als Sendung, Von der Telegrafie zum Internet'*. Munich, Germany: Verlag C.H. Beck.

de Boer, K. (1940) 'Plastische Klangwiedergabe'. *Philips Technische Rundschau*, 5. Jahrgang, Heft 4, pp.108-115, 1940.

de Bruijn, W.P.J., van Rooijen, W., Boone, M.M. (2001) 'Recent developments on WFS for high quality spatial sound reproduction'. *Proceedings 110th AES Convention, Amsterdam, The Netherlands*, April 2001, Preprint No.5370.

de Bruijn, W.P.J. and Boone, M.M. (2003) 'Application of Wave Field Synthesis in Life-size Videoconferencing'. *Proceedings 114th AES Convention, Amsterdam, The Netherlands*, February 2003, Preprint No.5801.

de Bruijn, W.P.J. (2004) '*Application of Wave Field Synthesis in Videoconferencing'*. PhD thesis, Technical University Delft, The Netherlands, ISBN 90-9018438-4, October 2004.

de Vries, D. and Berkhout, A.J. (1981) 'Wave theoretical approach to acoustical focusing'. *Journal of the Acoustical Society of America*, Vol.70, No.3, September 1981, pp.740-748.

de Vries, D., Start, E.W., Valstar, V.G. (1994) 'The Wave-Field Synthesis Concept Applied to Sound Reinforcement Restriction and Solutions'. *Proceedings 96th AES Convention, Amsterdam, The Netherlands*, January 1994, Preprint No.3812.

de Vries, D. (1995) 'Sound Enhancement by Wave Field Synthesis: Adaptation of the Synthesis Operator to the Loudspeaker Directivity Characteristics'. *Proceedings 98th AES Convention, Paris, France*, January 1995, Preprint No.3971.

Fink, M. (2002) 'Acoustic Time-Reversal Mirrors'. In: Fink, M. et al. (eds.): '*Imaging of Complex Media with Acoustic and Seismic Waves'*. Topics in Applied Physics, Vol.84, pp.17-43. Berlin, Germany: Springer-Verlag.

Gernemann-Paulsen, A., Neubarth, K., Schmidt, L., Seifert, U. (2006) 'Zu den Stufen im Assoziationsmodell'. *Proceedings 24th Tonmeistertagung 2006, Leipzig, Germany*, November 2006.

Gerzon, M.A. (1973) 'Periphony: With-height Sound Reproduction'. *Journal of the Audio Engineering Society*, Vol. 21, No.1, January/February 1973, pp.2-10.

Griesinger, D. (2001) 'The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique'. *Proceedings 19th AES Int. Conference on Surround Sound, Elmau, Germany*, May 2001.

Gross, R.D. (1992) *'Psychology: Science of Mind and Behaviour'*. 2nd edition, London, UK: Hodder & Stoughton, pp.239ff.

Hamasaki, K., Iwaki, M., Nakayama, Y., Nishiguchi, T., Okubo, H., Okumura, R. (2006) 'Natural Reproduction of Symphony Orchestra Music by an Advanced Multichannel Live Sound System'. *Proceedings AES 121st Convention, San Francisco, California*, October 2006, Preprint No.6966.

Hanselmann, K. (2006) *'Timbre perception of WFS-OPSI synthesised sound fields'*. Diploma thesis, Hochschule der Medien, Stuttgart, Germany, June 2006.

Hartmann, W.M. (1983) 'Localization of sound in rooms'. *Journal of the Acoustical Society of America*, Vol. 74, No.5, November 1983, pp.1380-1391.

Hertz, B.F. (1981) '100 Years with Stereo: The Beginning'. *Journal of the Audio Engineering Society*, Vol.29, No.5, May 1981, pp.368-370.

Horbach, U., Pellegrini, R., Felderhoff, U., Theile, G. (1998) 'Ein virtueller Surround Sound Abhörraum im Ü-Wagen'. *Proceedings 20th Tonmeistertagung 1998*, Karlsruhe, Germany, November 1998.

Horbach, U., Karamustafaoglu, A., Pellegrini, R., Mackensen, P., Theile, G. (1999) 'Design and Applications of a Data-Based Auralization System for Surround Sound'. *Proceedings 106th AES Convention, Munich, Germany*, April 1999, Preprint No.4976.

Huber, T. (2002) *'Zur Lokalisation akustischer Objekte bei Wellenfeldsynthese'*. Diploma thesis, Technical University Munich, Germany, July 2002.

Hulsebos, E.M. (2004) *'Auralization using Wave Field Synthesis'*. PhD thesis, Technical University Delft, The Netherlands, Oktober 2004.

IRT (2007) '*Binaural Room Synthesis'* [online]. http://www.irt.de/fileadmin/media/downloads/produkte/englisch/IRT_Binaural_Room_Synthesis_VST-Plugin_e.pdf [Accessed 28 August 2007].

Jacques, R., Albrecht, B., Melchior, F., de Vries, D. (2005) 'An Approach for Multichannel Recording and Reproduction of Sound Source Directivity'. *Proceedings 119th AES Convention, New York*, October 2005, Preprint No.6554.

Kates, J.M. (1985) 'A central spectrum model for the perception of coloration in filtered Gaussian noise'. *Journal of the Acoustical Society of America*, Vol. 77, No.4, April 1985, pp.1529-1534.

Kerber, S. (2003) *'Zur Wahrnehmung virtueller Quellen bei Wellenfeldsynthese'*. Diploma thesis, Technical University Munich, Germany, August 2003.

Komiyama, S., Morita, A., Kurozumi, K., Nakabayashi, K. (1991) 'Distance Control System for a Sound Image'. *Proceedings 9th AES Int. Conference on Television Sound Today and Tomorrow, Detroit, Michigan*, February 1991.

Krumbholz, K. (2004) 'Mechanisms determining the salience of colouration in echoed sound: Influence of interaural time and level differences'. *Journal of the Acoustical Society of America*, Vol.115, No.4, April 2004, pp.1696-1704.

Leakey, D. M. (1960) 'Further Thoughts on Stereophonic Sound Systems'. *Wireless World*, Vol.66, April 1960, pp.154-160.

Lee, H.-K. and Rumsey, F. (2004) 'Elicitation and Grading of Subjective Attributes of 2-Channel Phantom Images'. *Proceedings 116th AES Convention, Berlin, Germany*, May 2004, Preprint No.6142.

Lipshitz, S. (1986) 'Stereo Microphone Techniques: Are the Purists Wrong?' *Journal of the Audio Engineering Society*, Vol.34, No.9, September 1986, pp.716-744.

Lund, T. (2000) 'Enhanced Localization in 5.1 Production'. *Proceedings 109th AES Convention, Los Angeles, California*, August 2000, Preprint No.5243.

Martens, W.L. (2003) 'Perceptual Evaluation of Filters Controlling Source Direction: Customized and Generalized HRTFs for Binaural Synthesis'. *Acoustical Science and Technology*, Vol.24, No.5, September 2003, pp.220-232.

Martin, G., Woszczyk, W., Corey, J., Quesnel, R. (1999a) 'Sound Source Localization in a Five-Channel Surround Sound Reproduction System'. *Proceedings 107th AES Convention, New York*, August 1999, Preprint No.4994.

Martin, G., Woszczyk, W., Corey, J., Quesnel, R. (1999b) 'Controlling Phantom Image Focus in a Multichannel Reproduction System'. *Proceedings 107th AES Convention, New York*, August 1999, Preprint No.4996.

Mertens, H. (1965) 'Directional Hearing in Stereophony - Theory and Experimental Verification'. *EBU Review*, Part A: Technical report 92, August 1965.

Neher, T., Brookes, T., Rumsey, F. (2003) 'Unidimensional Simulation of the Spatial Attribute "Ensemble Depth" for training purposes, Part 1: Pilot Study Into Early Reflection Pattern Characteristics'. *Proceedings 24th AES Int. Conference on Multichannel Sound, Banff, Canada*, May 2003.

Nielsen, S. (1991) *'Distance Perception in Hearing'*. Aalborg, Denmark: Aalborg University Press.

Noguès, M., Corteel, E., Warusfel, O. (2003) 'Monitoring Distance Effect with Wave Field Synthesis'. *Proceedings 6th Int. Conference on Digital Audio Effects DAFX-03, London, UK*, September 2003.

Ono, K. and Komiyama, S. (1997) 'A Method of Reproducing Concert Hall Sounds by "Loudspeaker Walls"'. *Proceedings 102nd AES Convention, Munich, Germany*, February 1997, Preprint No.4490.

Ono, K., Pulkki, V., Karjalainen, M. (2001) 'Binaural Modelling of Multiple Sound Source Perception: Methodology and Coloration Experiments'. *Proceedings 111th AES Convention, New York*, November 2001, Preprint No.5446

Ono, K., Pulkki, V., Karjalainen, M. (2002) 'Binaural Modelling of Multiple Sound Source Perception: Coloration of Wideband Sound'. *Proceedings 112th AES Convention, Munich, Germany*, April 2002, Preprint No.5550.

Pellegrini, R. (2001) *'A virtual reference listening room as an application of auditory virtual environments'*. Dissertation, Ruhr-Universität Bochum, Germany.

Petrausch, S., Rabenstein, R., Spors, S. (2006) 'Simulation and Visualization of Room Compensation for Wave Field Synthesis with the Functional Transformation Method'. *Proceedings 119th AES Convention, New York*, October 2005, Preprint No.6547.

Pulkki, V., Karjalainen, M., Välimäki, V. (1999a) 'Localization, Coloration, and Enhancement of Amplitude-Panned Virtual Sources'. *Proceedings 16th AES Int. Conference on Spatial Sound Reproduction, Rovaniemi, Finnland*, March 1999.

Pulkki, V., Karjalainen, M., Huopaniemi, J. (1999b) 'Analyzing virtual Source Attributes using a binaural auditory Model'. *Journal of the Audio Engineering Society*, Vol.47, No.4, April 1999, pp.203-217.

Pulkki, V. (2001a) 'Coloration of Amplitude-Panned Virtual Sources'. *Proceedings 110th AES Convention, Amsterdam, The Netherlands*, May 2001, Preprint No.5402.

Pulkki, V. (2001b) *'Spatial Sound Generation and Perception by Amplitude Panning techniques'*. PhD thesis, Helsinki University of Technology, Helsinki, Finland, August 2001.

Pulkki, V., Karjalainen, M. (2001c) 'Localization of Amplitude-Panned Virtual Sources, I: Stereophonic Panning'. *Journal of the Audio Engineering Society*, Vol.49, No. 9, September 2001, pp.739–752.

Raatgever, J. and Bilsen, F.A. (1986) 'A central spectrum theory of binaural processing. Evidence from dichotic pitch'. *Journal of the Acoustical Society of America*, Vol.80, No.2, August 1986, p.429-441.

Rathbone, B., Fruhmann, M., Spikofski, G., Theile, G. (2000) 'Untersuchungen zur Optimierung des BRS-Verfahrens'. *Proceedings 21st Tonmeistertagung 2002*, Hannover, Germany, November 2002, pp.92-106.

Rebscher, R., Theile, G. (1990) 'Enlarging the Listening Area by Increasing the Number of Loudspeakers'. *Proceedings 88th AES Convention, Montreaux, Switzerland*, March 1990, Preprint No.2932.

Rubak, P., Johansen, L.G. (2003) 'Colouration in Natural and Artificial Room Impulse Responses'. *Proceedings 23rd AES Int. Conference on Signal Processing in Audio Recording and Reproduction, Helsingør, Denmark*, May 2003.

Rumsey, F. (2002) 'Spatial Quality Evaluation for reproduced Sound: Terminology, Meaning, and a Scene-based paradigm'. *Journal of the Audio Engineering Society*, Vol.50, No.9, September 2002, pp.651-666.

Rumsey, F., Zielinski, S., Kassier, R., Bech, S. (2005) 'On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality'. *Journal of the Acoustical Society of America*, Vol. 118, No.2, August 2005, pp.968-976.

Salomons, A.M. (1995) *'Coloration and binaural decoloration of sound due to reflections'*. PhD Thesis, Technical University Delft, The Netherlands.

Sengpiel, E. (2007a) *'Intensitätsstereofonie ist keine Intensitätsstereofonie…'* [online]. Available: www.sengpielaudio.com/IntensitaetsStereofonieIstKeine1.pdf, [Accessed 28 August 2007].

Sengpiel, E. (2007b) *'Hörereignisrichtung b2 in Abhängigkeit von der Interchannel-Laufzeitdifferenz Δt'* [online]. Available: http://www.sengpielaudio.com/HoerereignRichtungDt.pdf, [Accessed 28 August 2007].

Sengpiel, E. (2007c) *'Hörereignisrichtung b1 in Abhängigkeit von der Interchannel-Pegeldifferenz ΔL'* [online]. Available: http://www.sengpielaudio.com/HoerereignRichtungDL.pdf, [Accessed 28 August 2007].

Shinn-Cunningham, B. (2000) 'Distance Cues for virtual auditory space'. *Proceedings of the IEEE 2000 Int. Symposium on Multimedia Information Processing, Sydney, Australia*, December 2000, p.227-230.

Silzle, A., Theile, G. (1990) 'HDTV-Mehrkanalton: Untersuchungen zur Abbildungsqualität beim Einsatz zusätzlicher Mittenlautsprecher'. *Proceedings 16th Tonmeistertagung 1990*, Karlsruhe, Germany, November 1990, pp.208ff.

Smith, E.E., Nolen-Hoeksema, S., Frederickson, B.L., Loftus, G.R. (2003) *'Atkinson & Hilgard's Introduction to Psychology'*. 14th edition, Belmont, California: Wadsworth Thompson Learning.

Snow, W.B. (1953) 'Basic principles of stereophonic sound'. *Journal of the SMPTE*, Vol.61, November 1953, pp.567-589.

Sonke, J.-J., Labeeuw, J., de Vries, D. (1998) 'Variable Acoustics by Wavefield Synthesis: A Closer Look at Amplitude Effects'. *Proceedings 104th AES Convention, Amsterdam, The Netherlands*, April 1998, Preprint No.4712.

Sonke, J.-J. (2000) *'Variable Acoustics by wave field synthesis'*. PhD thesis, Technical University Delft, The Netherlands. Amsterdam, The Netherlands: Thela Thesis, ISBN 90-9014138-3.

Spors, S. (2006) *'Active Listening Room Compensation for Spatial Sound Reproduction Systems'*. Dissertation, University of Erlangen-Nuremberg, Germany.

Start, E.W. (1997) *'Direct Sound Enhancement by Wave Field Synthesis'*. PhD thesis, Technical University Delft, ISBN 90-9010708-8.

Steinberg, J.C. and Snow, W.B. (1934) 'Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective: Physical Factors'. *Bell Systems Technical Journal*, Vol.XIII, No.2, April 1934, pp.245.

Supin, A.Ya., Popov, V.V., Milekhina, O.N., Tarakanov, M.B. (1999) 'Ripple depth and density resolution of rippled noise'. *Journal of the Acoustical Society of America*, Vol.106, No.5, November 1999, pp.2800-2804.

The Mathworks (2007) MATLAB® [online]. Available: www.matlab.com [Accessed 7 December 2007].

Theile, G. (1980) 'Über die Lokalisation im überlagerten Schallfeld / On the localisation in the superimposed sound field'. Dissertation, Technical University Berlin, Germany, 1980. Available: www.hauptmikrofon.de/theile.htm [Accessed 28 August 2007].

Theile, G. (1990) 'On the Performance of Two-Channel and Multi-Channel Stereophony'. *Proceedings 88th AES Convention, Montreux, Switzerland*, February 1990, Preprint No.2887.

Theile, G. (1991) 'On the Naturalness of Two-Channel Stereo Sound'. *Journal of the Audio Engineering Society*, Vol.39, No.10, October 1991, pp.761-767.

Theile, G. (2001) 'Multichannel natural music recording based on psycho-acoustic principles'. *Proceedings 19th AES Int. Conference* on *Surround Sound: Techniques, Technology and Perception*, Elmau, Germany, June 2001, pp. 201-229.

Theile, G., Wittek, H., Reisinger, M. (2002) 'Wellenfeldsynthese-Verfahren: Ein Weg für neue Möglichkeiten der räumlichen Tongestaltung'. *Proceedings 21st Tonmeistertagung 2002*, Hannover, Germany, November 2002.

Theile, G., Wittek, H., Reisinger, M. (2003) 'Potential Wavefield Synthesis Applications in the Multichannel Stereophonic World'. *Proceedings 24th AES Int. Conference on Multichannel Sound, Banff, Canada*, May 2003.

Torger, A. (2007) *'BruteFIR: software convolution engine'* [online]. Available: http://www.ludd.luth.se/~torger/brutefir.html [Accessed 7 December 2007].

Tsakiris, V., Orinos, C. , Laskaris, K. (2005) 'Objective and Subjective Evaluation of Digital Equalization Systems - Measurements of Resonances and Colorations'. *Proceedings AES 118th Convention, Barcelona, Spain*, May 2005, Preprint No.6463.

Usher, J., Martens, W.L., Woszczyk, W. (2004) 'The influence of the presence of multiple sources on auditory spatial imagery as indicated by a graphical response technique'. *Proceedings 18th Int. Congress on Acoustics ICA 2004, Kyoto, Japan*, April 2004.

Verheijen, E.N.G. (1998) *'Sound reproduction by Wave Field Synthesis'*. PhD thesis, Technical University Delft, The Netherlands.

Vogel, P. (1993) *'Application of wave field synthesis in room acoustics'*. PhD thesis, Technical University Delft, The Netherlands.

Wegmann, D. (2005) *'Zu Unterschieden in der Hörereigniswahrnehmung bei Wellenfeldsynthese und Stereofonie im Vergleich zum natürlichen Hören'*, Diploma thesis, Fachhochschule Oldenburg/ Ostfriesland/ Wilhelmshaven, Germany, August 2005.

Wendt, K. (1963) *'Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie'*. Dissertation, Technische Hochschule Aachen, Germany.

Wightman, F.L. and Kistler, D.J. (1992) 'The dominant role of low-frequency interaural time differences in sound localization'. *Journal of the Acoustical Society of America*, Vol. 91, No.3, pp.1648-1661.

Williams, M. (1984) 'The Stereophonic Zoom: A Practical Approach to Determining the Characteristics of a Spaced Pair of Directional Microphones'. *Proceedings 75th AES Convention, Paris, France*, March 1984, Preprint No.2072.

Williams, M. (2000) *'Microphone Arrays for Stereo and Multichannel Sound recording'*. Milan, Italy: Il Rostro.

Wittek, H. (2000a) *'Untersuchungen zur Richtungsabbildung mit L-C-R-Hauptmikrofonen'*. Diploma thesis, Fachhochschule Düsseldorf, Germany. Available: www.hauptmikrofon.de/wittek.htm [Accessed 28 August 2007].

Wittek, H. and Theile, G. (2000b) 'Untersuchungen zur Richtungsabbildung mit L-C-R-Hauptmikrofonen'. *Proceedings 21st Tonmeistertagung 2000*, Hannover, Germany, November 2000, pp.432-454.

Wittek, H. (2001a) *'Image Assistant 2.0* and documentation' [online]. Available: www.hauptmikrofon.de/ima2.html [Accessed 28 August 2007].

Wittek, H., Neumann, O., Schaeffler, M., Millet, C. (2001b) 'Studies on Main and Room Microphone Optimization'. *Proceedings 19th AES Int. Conference* on *Surround Sound*, Elmau, Germany, June 2001.

Wittek, H. and Theile, G. (2002) 'The Recording Angle – Based on Localisation Curves'. *Proceedings 112th AES Convention, Munich, Germany*, May 2002, Preprint No.5568.

Zahorik, P. (2002) 'Auditory display of sound source distance'. *Proceedings of the Int. Conference on Auditory Display 2002, Kyoto, Japan*, July 2002.

Zieglmeier, W., Theile, G. (1996) 'Darstellung seitlicher Schallquellen bei Anwendung des 3/2-Stereo-Formates'. *Proceedings 19th Tonmeistertagung 1996*, Karlsruhe, Germany, November 1996, pp.159–169.

Zurek, P.M. (1979) 'Measurements of binaural echo suppression'. *Journal of the Acoustical Society of America*, Vol.66, No.6, December 1979, pp.1750–1757.

Zwicker, E., Fastl, H. (1990) *'Psychoacoustics. Facts and Models'*. Heidelberg, Germany: Springer Verlag.