Audio Engineering Society

Convention Paper 6000

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Spatial perception in Wave Field Synthesis rendered sound fields: Distance of real and virtual nearby sources

Helmut Wittek[1,2], Stefan Kerber[1,3], Francis Rumsey[2] and Günther Theile[1]

[1]Institut für Rundfunktechnik, Munich, Germany
wittek@irt.de, theile@irt.de

[2]Institute of Sound Recording, University of Surrey, Guildford, UK

[3]AG Technische Akustik, MMK,
Technische Universität München, Munich, Germany

ABSTRACT

Wave Field Synthesis is capable of reproducing a sound field by means of loudspeaker arrays. It is desired to make a copy of the original sound field in order to create a virtual sound field with the same properties. It has been shown that the properties are similar but not congruent, which leads to the question of what the agreements and differences are with regard to auditory perception. Considerations and experiments are made to illuminate the case of distance perception of virtual sources, and, in particular, the role of the curvature of the wave front for distance perception.

1. INTRODUCTION

Wave Field Synthesis (WFS), a technique introduced and developed by Berkhout et al. [1], uses arrays of loudspeakers to synthesize arbitrarily shaped wave fronts. It offers enhanced spatial reproduction possibilities in comparison to conventional Stereo[1]. This fact is widely accepted although the concrete reasons have not yet been scrutinized sufficiently. It would be worthwhile to identify the properties that actually qualify WFS for spatial reproduction and thus be able to define and quantify the specific advantages of this technique.

[1] The term "Stereo(-phony)" is used in this paper for reproduction techniques which are based on phantom sources.

A first look at the WFS properties reveals a beneficial characteristic: stable virtual source positions are not only possible at the locations of the loudspeakers, but can also be created anywhere in the *horizontal plane*. This characteristic, as illustrated in Figure 1, creates a real *acoustic perspective*, which is based on either constant source directions (in the case of distant sources) or constant source positions within a wide listening area. These positions determine not only the direction in which the virtual source is localized, but also the *spatial decay of the source amplitude* within the listening area. Thus, a quasi-natural variation of the source's amplitude which corresponds to movements within the listening area, enhances the feeling of "immersion".

The aim of this investigation is to look at some possible additional advantages of WFS in the field of spatial reproduction. This task requires the discussion of relevant cues and physical and psycho-acoustical properties.

In natural hearing a source position is determined by two characteristics: *direction* and *distance*.

These two parameters can never exist without each other because no natural sound event can be imagined without distance or direction. In the case of WFS, the perceived virtual source direction (which is synthesized by the WFS array) is in agreement with that of a natural source at the virtual source's position within the entire listening area. The perceived distance, however, is not determined simply by the geometry of virtual source and array, i.e. by the direct sound emitted at the virtual source's position. The straightforward analogy between real and virtual sources fails because the virtual source lacks natural acoustics, i.e. important cues contained in the indirect sound.

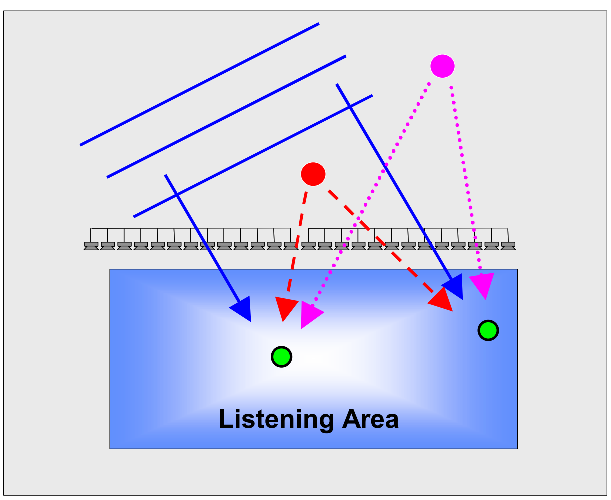

Figure 1: WFS is capable of reproducing both the stable positions of point sources (red and pink, dashed and dotted) and the stable direction of a plane wave (blue, solid), quoted from ([2])

The literature (e.g. [3][4][5][6]) shows various crucial parameters for auditory distance perception as e.g. loudness, direct/reverb ratio, reflection pattern, plausibility, etc. As a consequence, the distance of the WFS virtual source has to be created additionally based on the aforesaid parameters. In other words, as in the case of the real source, the virtual source has to be produced together with the natural acoustic environment if a natural auditory event is desired.

WFS is superior to Stereo in that it can reproduce the curvature of the wave front. Each source distance corresponds to a different wave front curvature at the receiver's position, as it is illustrated in Figure 2. Apparently, these differences are significant only for rather *close* sources. For them, this could lead to the creation of auditory cues related to the direct sound, which is the first wave front. These cues are caused by binaural differences, the so-called "acoustic parallax".

If they were to produce auditory distance perception by overriding the other incorrect cues, a new dimension of sound reproduction would be unveiled: the space in-between the loudspeaker and the listener. In other words, it would be the first time that a loudspeaker technique is capable of producing auditory events closer than the loudspeakers.

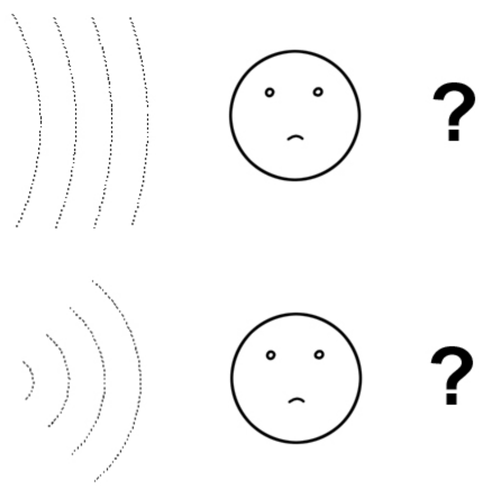

Figure 2: Is distance perception possible from the direct sound only?

2. DEPTH AND DISTANCE

In any natural sound field a sense of spatial depth exists, being the sense of perspective in the (reproduced) acoustic scene ([7]). Depth is regarded as a perceptual attribute of the acoustical scene – in contrast to the source-related attribute distance.

The successful creation of depth in a sound scene is the benchmark of a spatial audio reproduction system. Whereas the perception of distance differences is possible even in Mono, a true sense of depth cannot be achieved by one loudspeaker alone. Stereophonic methods, however, are indeed capable of reproducing spatial depth which is particularly true for the Surround Sound formats. This points to the existence of different auditory cues for the two attributes distance and depth.

In [9] a discussion about the analogy to visual cues emphasizes the importance to separate (sound) scenes with and without depth. The difference between a 2D and a 3D visual scene is the existence of spatial depth: a 2D scene only enables a flat reproduction without depth, although distance cues are present. This is illustrated in Figure 3:

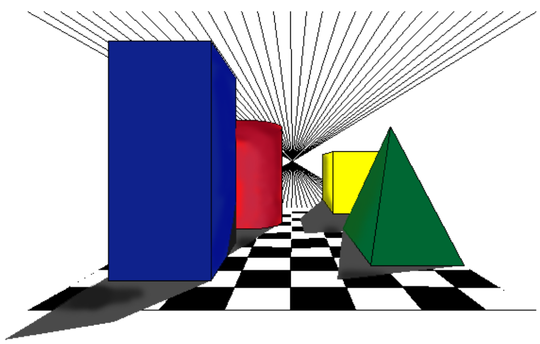

Figure 3: Visual analogy: 2D representation of a 3D scene, taken from [8].

In Figure 3 the relative size of the objects on the 2D picture as well as object occlusion and other distance cues are available and enable a simulation of depth. True depth, however, is possible only through a "stereoscopic" view. A 3D scene contains true depth. The visual analogy knows a third way: the 2½ D representation. Visual 2½ D describes a scene in which you can move but which does not contain true depth cues (like a "stereoscopic" view). Correspondingly, "Acoustical 2½ D" describes a sound scene in which the sense of perspective and dynamic movements are supported, but where true depth cues are lacking.

Auditory cues for the perception of "true" depth, as mentioned, can exist in Stereo as well as in WFS. In contrast to pure distance cues like loudness, direct/reverb-ratio and others, the perception of depth demands a proper acoustic environment.

If the wave front of a WFS virtual source signal contained distance perception cues, which is the topic of interest in this study, the perception of depth would significantly be supported in WFS.

3. DISCUSSION OF WFS PROPERTIES FOR SPATIAL REPRODUCTION

The topic of this paper is an investigation into the distance perception of virtual sources in WFS. This investigation requires a more general discussion of some properties of WFS without which the study would be isolated and could cause misunderstandings.

The box below points out some of the main questions with respect to spatial reproduction through WFS, for which answers are required.

> Which properties of WFS qualify this technique for spatial reproduction?
>
> Is the spatial reproduction achieved with a WFS array superior to Stereo? (and if so, whereby?)

Surprisingly, there are a number of different opinions about these relatively clear questions. Below, various facts and myths about WFS are discussed, in order to find a road map for the investigation:

1. Degree of Congruence

"The more perfect the copy of the real sound field, the better the reproduction of the space."

If the copy was perfect then of course the properties of the sound field would also be the same.

However, in the practical case of WFS the virtual sound field differs significantly from the reference with respect to several physical parameters (see chapter 7.1). It may show many more similarities to the reference in comparison to other reproduction techniques, but these similarities become irrelevant as soon as one is interested in only some crucial cues. The degree to which these crucial spatial perception cues are reproduced determines the quality of the spatial reproduction. This is a much higher demand on the properties of WFS.

An example for this problem is the auralization ([23]) of real acoustics. Here WFS suffers from the fact that, for practical reasons, it currently only uses the horizontal plane. The relevant perceptual cues are violated when the captured ceiling reflections are reproduced from the horizontal plane. It has already been shown that a perceptually optimized (hyperreal) design of a virtual room can even be superior to the original sound field with respect to spatial quality ([28]). This is thought to be relevant, as any form of audio reproduction will violate several cues (plausibility, visual, etc), and thus the remaining cues have to be strong enough to override those.

2. The dry source

"You can put the source anywhere in the horizontal plane."

A WFS virtual source is very similar to a real source with respect to the perception of its direction - as proven in [15] - and this fact often leads to the conclusion that WFS is capable of producing a perfect copy of any sound field. However, a *dry* virtual source can in the best case have the same properties as a *dry* real source. And one can perceive the distance of a dry real source - as discussed in chapter 5 - only with restrictions. The possibility to reproduce the correct wave front of a virtual source at any position does not lead to the conclusion that the perception of depth is enabled.

3. What about Stereo?

"The superimposed sound field of two stereo loudspeakers has almost nothing to do with the real sound field"

According to Theile, the perception of a phantom source, created by two stereo loudspeakers, is based on a totally different perception pattern. As stated in [16] and [17], the single loudspeaker signals are perceived separately and merged in a *psycho-acoustical* association process. Therefore the *physical* superposition of the loudspeaker signals is not relevant for the auditory system, confirming the mentioned statement to be correct, but irrelevant. Other theories of stereophonic perception which regard the physical superposition as crucial (e.g.[18]) cannot conclusively explain the specific properties of a phantom source. An example is the comb filtering due to the physical superposition of the loudspeaker signals at one ear, which is not perceived by the listener.

In a large number of publications the nature of the phantom source and the reproduction of spaces by stereophonic means is dealt with and its properties are described (e.g. [22][24][25][26]). The spatial perception achieved with a Stereo set-up can be quite convincing. As mentioned in chapter 2 the creation of depth is possible by stereophonic means when special recording techniques or suitable virtual acoustics algorithms are applied. Corresponding to the statement in point 1 above, it is not intended to produce a copy of the real sound field but a reproduction of the relevant cues which leads to a successful spatial perception.

4. The accuracy of the reflections

A WFS virtual source is superior to a phantom source with respect to its stability and its focus ([15]). These properties could give rise to the conclusion that the reproduction of a room which is based on the reproduction of a large number of single sources (reflections) can be advantageous, too. This plausible assumption, however, has not yet been proven scientifically.

5. The listening area

WFS is a volume solution.

It is capable of reproducing the correct direction of virtual sources in the whole listening area and for multiple listeners, and therefore also the correct perspective between different sources. This makes it unique and this possibly is the main advantage in comparison to other techniques like binaural, transaural and Stereo which are sweet-spot or headphone dependent. However, regarding sound quality, handling and complexity, those techniques could, for a fixed listening position, be superior to WFS.

6. The interactivity cue

There is a certain possibility of localizing virtual sources through moving[2] within the listening area.

One the one hand, one can interpret the changes in the source direction and deduce the source position including its distance. One the other hand, realistic changes in the reflection pattern can also reinforce a realistic perception of the space. Furthermore, the perception of distance will be supported if the spatial amplitude decay of a virtual source is similar to that of the original.

These "2½ D" cues are only available when the listener moves within the listening area.

7. The influence of the reproduction room

A WFS array is not capable of producing a perfect copy of the original sound field as soon as the reproduction room reflections disturb the desired reproduced sound field. The distance perception of a virtual source in front of the array (a so-called "focused" source), which is possible in WFS, is avoided because its fragile distance cues - if existent at all - are overridden by the stronger distance cues created by the array speakers themselves (discussed in detail later on). This effect is well-known from Stereo,

[2] Here "moving" means a real change of the listening position and not spontaneous head movements required for solving localization ambiguities

where a source distance closer than the loudspeaker distance cannot be achieved. Special de-reverberation algorithms, which are under investigation ([19][20]) could perhaps be a corrective for that.

The reflections caused by the array itself do not fit a natural reflection pattern that a real source on the virtual source's positions would create. Although one may consider these reflections as having emerged from a virtual mirror source, both incident time, level and sound colour of the mirror sources disagree with the natural case.

4. PROBLEM DEFINITION FOR THE INVESTIGATION

The aim of the investigation is to check the auditory perception of the distance of WFS virtual sources. According to the remarks in the last chapter, there are a lot of different parameters that are relevant for the spatial perception of virtual sound fields.

The most meaningful investigation is expected to be a study about the **spatial perception due to the shape of the wave front**. This study would reveal the meaning of the cues related to the wave front curvature and explain a possible superiority of WFS with respect to these cues. If an influence of these cues cannot be proven, their importance would be reduced and they would play a role only for listener movements. From a perceptual point of view the wave front curvature is considered one of the main differences between WFS and Stereo listened to from a *fixed* position.

Another main difference, the size of the listening area and the possibility to move, is not valid for a fixed listening position. As a volume solution the specific advantages of WFS are apparent and some of its properties are superior to other reproduction techniques. The influence of these properties on the acoustic perception at a *fixed* listening position has not yet been sufficiently investigated.

The specific case of *focused* virtual sources needs to be discussed from a different angle: Here, the weight of the direct sound cues is higher because of the increased direct/reverb ratio. Furthermore, the influence of the reflections of the array speakers themselves is even more disturbing because the array reflections arrive *before* the early reflections of the virtual source and therefore can hardly be masked. Hence, erroneous cues will exist which cannot be overridden by the acoustics of the virtual room, and the direct sound cue will be the only correct cue 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 its reproduction over WFS arrays.

5. DIRECT SOUND CUES FOR DISTANCE PERCEPTION

As shown in the literature, the perception of the distance of a source is nearly impossible without the presence of room reflections. Investigations by Nielsen ([3]) 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 4). Loudness is the most important cue when room acoustics are absent. This was shown in studies cited by Blauert [6]. At least for distances > 1m the variation of the loudness of a source directly leads to a different distance judgment.

This is different in the near-field. The so-called "acoustic parallax", i.e. the fact that the interaural differences vary according to the source distance, serves as an additional auditory cue. For Blauert ([6]) the distance can be perceived for sources nearer than 3 m due to this cue. The results of Nielsen suggest a limit of 1 m. Brungart and Rabinowitz ([10]) 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 ([5]) notes that by adding reflections the distance perception also in the nearby region improves significantly, it is interesting that a "direct sound only" cue is able to override the loudness cue. If ILD were to be sufficiently created in WFS (such that they could override other erroneous cues within a reflective environment), then the mentioned challenge of filling the space in-between the loudspeaker and the listener could be met.

In summary, the literature suggests that any correct perception of the distance of dry sources at a distance of more than roughly 1m is *not* possible if the loudness at the listener's ears is kept constant. This means although WFS is capable of synthesizing the correct wave front for these sources, it is not possible to identify their distance without moving within the listening area.

When the listener moves within the WFS sound field, correct *directional* cues are perceived which can help to indirectly estimate the source position, while quasi-natural loudness cues support the sense of distance.

For source distances < 1m - in the case of WFS this means focused sources - it is not clear if the low frequency ILD cue can override erroneous cues caused by the WFS array's reflections. This was studied in the following investigation.

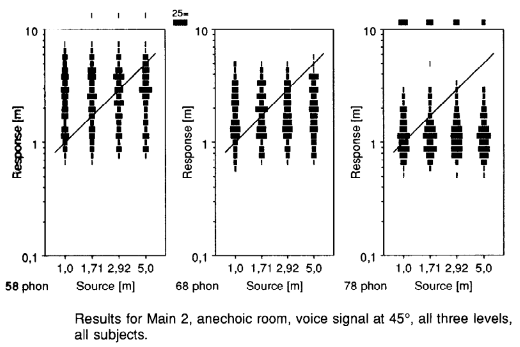

Figure 4: Experimental results from Nielsen ([3]): There is no relationship between the actual source distance (x-axis) in the anechoic chamber and the perceived distance (y-axis). But: the louder the stimulus the closer it is perceived (the 3 figures correspond to a different receiver loudness, which is 58, 68 and 78 phon)

6. EXPERIMENTAL SET-UP

Both real sources (small loudspeakers) and virtual sources were presented in the anechoic chamber. The test panel had to estimate the perceived distance of the source. Different source distances from 0.25 m to 1.9 m were used.

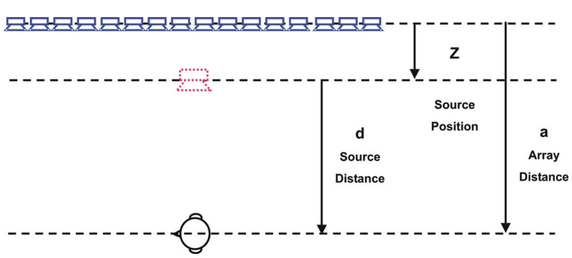

Figure 5: Array - Source – Receiver geometry for the simulations/experiments: (virtual) source (red dotted) is on the ear axis

Figure 5 illustrates the source/array geometry.
The listener's ear axis is perpendicular to the WFS array (synthesizing a virtual source) or a real source. The right ear points to the (virtual) source and the left ear is turned away from the (virtual) source. This was chosen because binaural differences are maximal for this case.

The distance between the source and the center of the listener's head is called the source distance d. The distance between the array and the listener is called array distance a.
Furthermore, the distance between the source and the array is the source position z (= a – d).
The following distances d and corresponding source positions z were chosen. In the case of WFS, *positive* source positions correspond to focused sources, *negative* source positions correspond to sources behind the array.

Source Distance d	**Source Position z**
0.25 m	1 m
0.45 m	0.8 m
0.65 m	0.6 m
0.85 m	0.4 m
1.10 m	0.15 m
1.50 m	*- 0.25 m*
1.90 m	*- 0.65 m*

The non-focused virtual sources behind the array (1.50m and 1.90 m) are indicated by dash-dotted lines in the Figures 10, 11 and 13.

The linear WFS array consists of n = 16 monopole loudspeakers with an interspacing of Δx = 0.17 m. This makes an array length of 2.55 m. (Tapering was done using a spatial window (Hanning), equalization was performed according to the WFS driving functions, which are described e.g. in [11] and [13]).

As a real source a single small loudspeaker of the type ELAC 301 (width = 91 mm) was chosen.
Further details on the experiment design are depicted in chapter 7.3.

7. THEORETICAL ANALYSIS OF THE REAL AND THE SYNTHESIZED WAVE FIELD

Before the experiment is described, some theoretical observations and simulations are presented which will be an important basis for an explanation of the experimental results. Several different parameters are likely to influence this result and therefore a careful separation is necessary. In a number of figures, which will be explained in detail in the following chapters, these influences are illustrated.

For this study, the following measures are important:

1. distribution of the sound pressure level within the listening area for both real and virtual sound field
   → the sound field without Head Shadowing, illustrated in the **black figures** in Table 1

2. binaural signals for both real and virtual sound field
   → the sound field with Head Shadowing, illustrated in the ***red (cursive) figures*** in Table 1

To be able to analyze the reason for differing sound fields in the real and virtual case, the influence of the single parameters are illustrated step by step in the following graphs. Two aspects are analyzed 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. The sound field is analyzed both with and without the influence of the listener's head.

In the following Table 1 the figures in black fonts (columns 1 and 2) show an analysis of the sound field without Head Shadowing. The red figures (column 3) are dummy head simulations, i.e. measures of binaural signals. Column 1 indicates the level spectrum at different distances and columns 2 and 3 indicate the level differences measured for 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 (a = 1.25 m, d as indicated, z = a-d, see Figure 5 for the geometry).

Source	**Level spectra**	**"No-Head ILD", see chapter 7.2**	**ILD**
Real source	Figure 8	Figure 9	*Figure 12*
WFS Virtual Sources	Figure 10	Figure 11	*Figure 13*

Table 1: Assignment of the Figures

Some comments on the technical origin of the figures:

Figure 12 was derived from an HRTF (Head related transfer functions) measurement using the dummy-head Neumann KU 100 and small ELAC 301 (width = 91 mm) loudspeakers.

Figures 8, 9, 10 and 11 are 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 the array loudspeakers as well as the real sources obeys the inverse square law.
The simulations only include simple calculations of travel time and amplitude decay of the involved (secondary) sources.

Figure 13 is a simulation which is based on an HRTF database measured at the IRT using the dummy-head Neumann KU 100 and the loudspeaker K&H O100. These HRTF are available for azimuth directions at a resolution of 6° (that is, 60 measurements in the horizontal plane). The respective HRTF used for the simulations are derived through an interpolation in the frequency domain.

7.1. Physical Deficiencies of Focused Sources in WFS

In theory Wave Field Synthesis is capable of reproducing a perfect copy of an original sound field. However, in practice it is not possible to fulfill all theoretical requirements. Thus, due to a number of reasons, a synthesized WFS sound field differs from the sound field of a real source to some degree. These reasons are the finiteness of the array (*Diffraction Effects*), the discreteness of the array (*Spatial Aliasing*) and the reduction to the horizontal plane (*Amplitude Errors*). Details about these deficiencies as well as their perceptual consequences can be found in [9].

Furthermore, one has to be aware of the special status of a focused source in WFS. It may be regarded as a result of a (virtual) acoustical focusing system behind the WFS-array ([12]). In [13] Boone states: "*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.*"
Hence, focused sources have different properties compared to normal virtual WFS sources behind the array. For example, the correct wave front is synthesized only behind (in the propagation direction) the focus point. This can be seen in Figure 7.

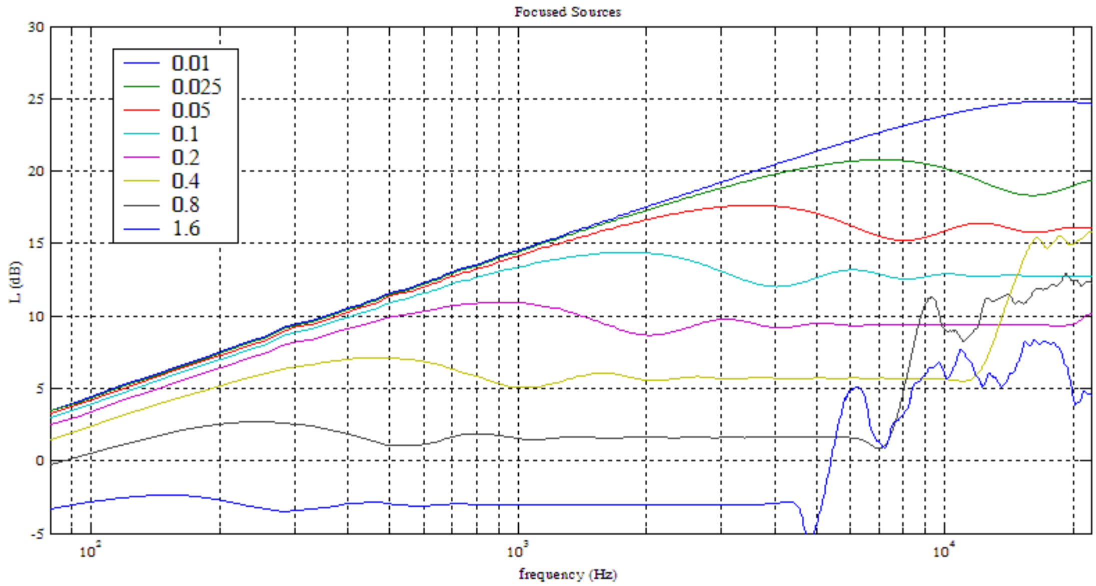

Figure 6: Spectrum of a Focused Source at different Source-Receiver distances d.

(d see legend (in m), source at z=1 m, linear WFS array, n=101 loudspeaker with interspacing of Δx = 0.17 m)

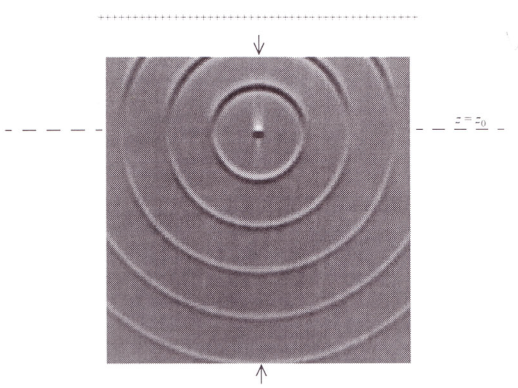

Figure 7: Focused Source, synthesized by a WFS array, taken from [11]. The array is indicated with crosses (+). The sound image is correct for listener positions behind (in propagation direction) the focus point.

Spatial Aliasing limits the correct synthesis of the sound field in the upper frequency range. It depends on the travel time differences between adjacent secondary array loudspeakers. Above the so-called Spatial Aliasing Frequency $\mathbf{f_{alias}}$ the sound field is neither spatially nor spectrally correct.
In the case of WFS focused sources, the travel time differences between adjacent WFS array channels are quite small. The contributions of the secondary array loudspeakers are synthesized so that they focus in one point, this being the virtual source position. That means, their travel times are designed so 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 $\mathbf{f_{alias}}$ very high. For greater distances the travel time differences are bigger, leading to a decreased Spatial Aliasing frequency. This can be seen in Figure 6.

The $\mathbf{f_{alias}}$ of a non-focused source is significantly lower due to the bigger travel time differences.

Diffraction Effects are caused by the limited length of the WFS array. Theoretically, an infinite number of loudspeakers are necessary to correctly synthesize the virtual source. The consequence of a finite array length is the truncation of the exponential function that describes the receiver signal in the time domain. This causes a quasi-comb filter effect (caused by the superposition of two correlated but phase-inverted, time-shifted signals) with a fundamental frequency which depends on the difference between the incident times of the first and the last WFS array loudspeaker signal at the receiver position. Applying a spatial tapering window (by damping the outer array loudspeakers), the ripple can be damped significantly at the expense of a narrowing of the listening area.

However, the low frequencies are still attenuated most and tapering worsens this further.

The *Diffraction Effects* are special in the case of focused sources. The relevant time difference is very small and this makes the fundamental frequency of the resulting quasi-comb filter quite high leading to significant rippling and a loss of low frequencies depending on the source-receiver distance. (Details about the nature of focused sources can be found in [14]).
Figure 6 shows the influence of this effect. Only for larger source distances the low frequencies are synthesized sufficiently. One can of course equalize the frequency response with respect to a reference receiver position at the cost of an over-emphasis of low frequencies for larger distances.

Figure 6 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. Hence, the spatial intensity decay fails to meet the inverse square law. This is due to the reduction of the synthesis to the horizontal plane.
For the linear, correctly synthesized contribution the spatial amplitude decay can be described by the following formula. The amplitudes of the non-linear, incorrectly synthesized contributions of lower and higher frequencies decline even more smoothly.

$$\mathrm{p} \sim \frac{1}{\sqrt{d \cdot a}} \text{ ; (after [11])}$$

with p = sound pressure of a **virtual** source,
d = source distance,
a = array distance.
for the geometry see Figure 5.

To summarize, the desired flat frequency response is achieved only for the mid frequencies, the range and position of this correctly synthesized contribution is depending on the distance d and the array set-up. The spatial intensity decay is smoother than suggested by the inverse square law.

7.2. "No-Head-ILD" as a measure of the sound field without head shadowing

The level of a real source listened to from certain source distances d (for the set-up see Figure 5) were simulated in Figure 8. In the simulations the levels at the two positions of the ears were calculated, i.e. two positions at a

distance = d +/- (ear distance/2).

The ear distance was set to 0.17 m.
Note: As there is no head in this situation, no head shadowing is effective. 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 "No-Head-ILD". It corresponds to an ILD measurement except for the fact that no head shadowing (including pinna effects) occurs.

The inverse square law dictates a level increase with decreasing distance as it can be seen in Figure 8:

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

with p = sound pressure of a **real** source and
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 inverse square law.

Also, the level *differences* between left and right ear positions increase with decreasing distance. These level differences are plotted in Figure 9. In this graph the "No-Head-ILD" are simulated.
It should be remembered that with regard to auditory distance perception, it is important that in particular the low-frequency ILD (<3500 Hz, see [10] and chapter 5) are dependent on the source distance for nearby sources. One sees that for real sources the "No-Head-ILD" are present and that they are indeed dependent on the source distance if we consider distances of roughly d < 1 m.

Now the virtual sources are considered:
In Figure 10 the level spectra of focused sources in different distances are plotted. As mentioned in the last subchapter, the impact of Diffraction Effects and Spatial Aliasing is significant. A flat frequency response and – as a consequence – a significant and consistent "No-Head-ILD" (Figure 11) is present only for a certain mid frequency range. Width and position of this range is dependent on the source distance.
"No-Head-ILD" are indeed present in the important low-frequency range which could be effective, although not in the same quality and quantity as in the case of a real source.

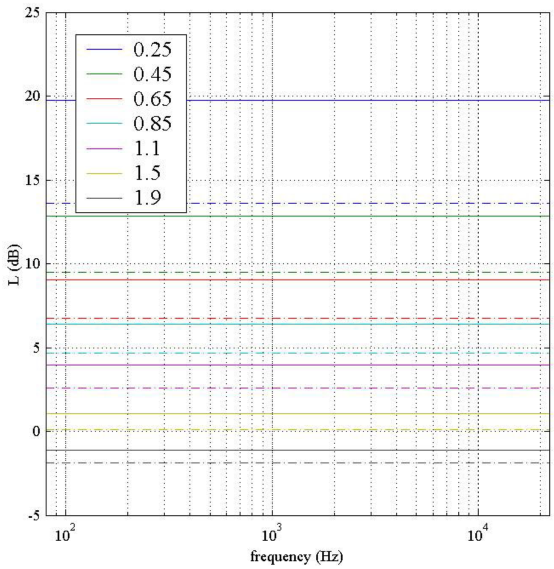

*Figure 8: **Level** of a **Real Source** at distances = d +/- (ear distance/2)*

Solid line: right ear, dashed line: left ear

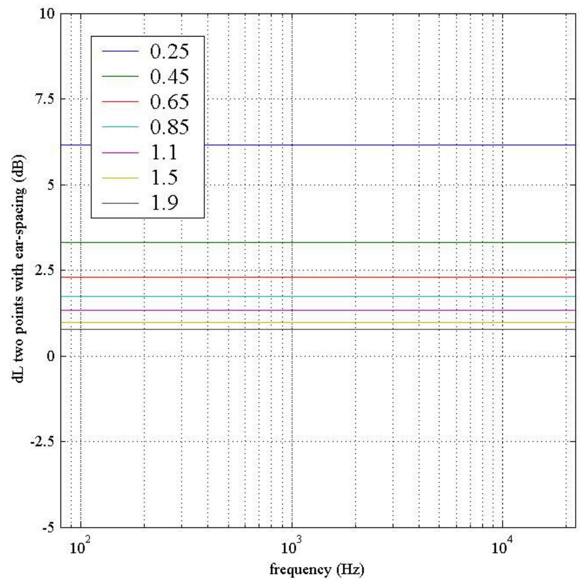

*Figure 9: **"No-Head-ILD": Level differences ΔL** between ear positions in the sound field of a **Real Source** at distance d.*

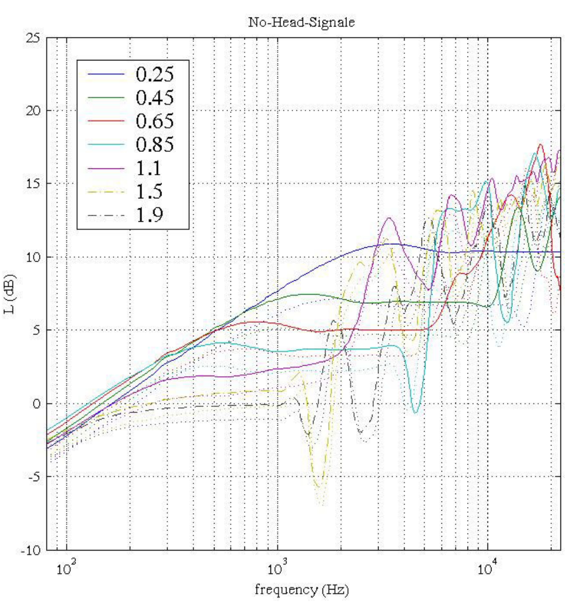

*Figure 10: **Level** of a **focused source** at distances = d +/- (ear distance/2)*

Solid line: right ear, dashed line: left ear

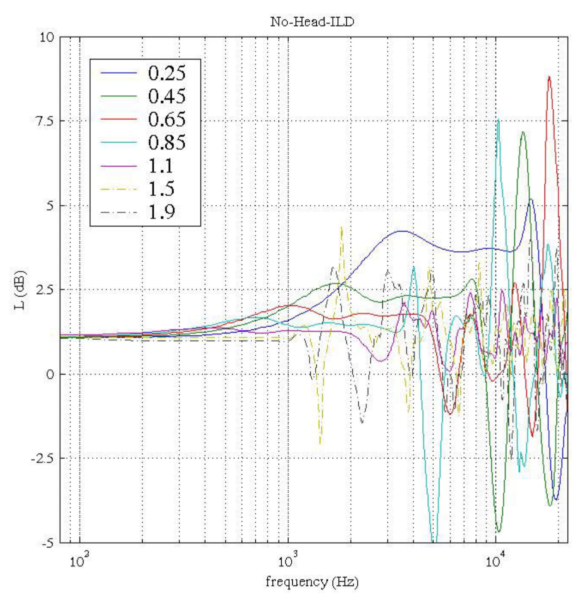

*Figure 11: **"No-Head-ILD": Level differences ΔL** between ear positions in the sound field of a **focused source** at distance d.*

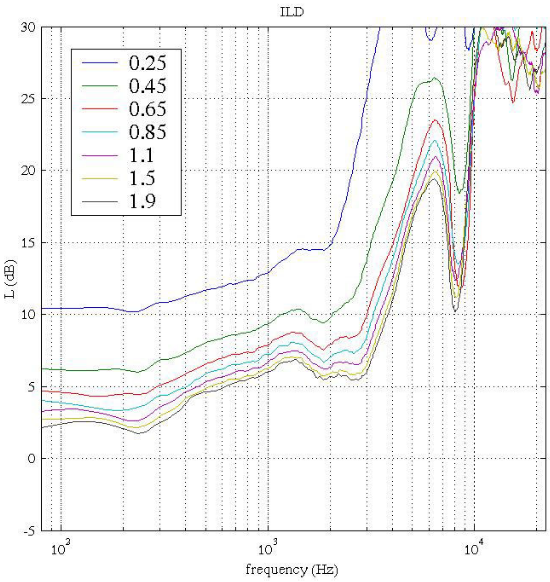

Figure 12: ***Interaural Level differences ILD*** *in the sound field of a* ***Real source*** *at distance d*

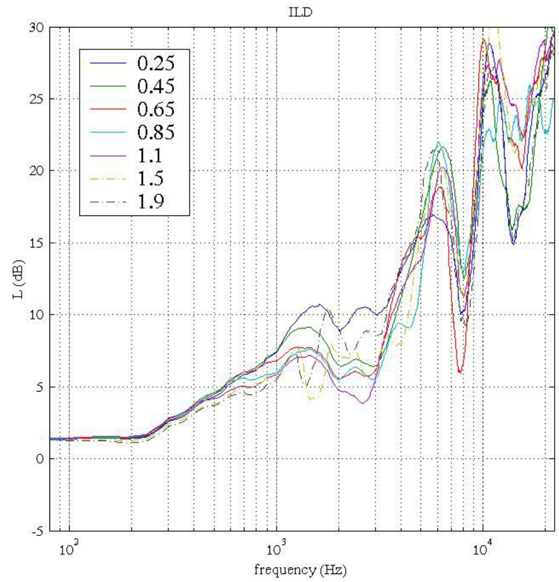

Figure 13: ***Interaural Level differences ILD in*** *the sound field of a* ***focused source*** *at distance d*

7.3. The Head in the WFS sound field

The simulations in Figure 12 and Figure 13 show the ILD of real and virtual sound field.
As depicted in the last chapter the focused sources can only partially produce significant differences between the ILD corresponding to different distances. This can be seen from Figure 13. ILD remain only in a small frequency range. The ILD of real sources are plotted in Figure 12. From these graphs it may be concluded that for distances below roughly 1 m the ILD differ significantly and thus one may gather the source distance from these ILD only.
It will be subject to a further simulation in chapter 11 to investigate which role Diffraction effects play and, furthermore, how non-correct Head-Shadowing influences the ILD.

8. LISTENING TESTS: EXPERIMENTAL DESIGN

1. Test panel selection

The perception of the distance of dry sources in the anechoic chamber is a very difficult task for the 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.
Therefore only experienced audio researchers participated in the experiment. Some results of naïve listeners were collected for test reasons. They showed no relationship between reference distance and perceived distance and were ignored.
The data of 7 persons who performed both experiments are shown here.

2. Two separate listening tests

For each type of source (real, virtual) a separate test was performed. There were two reasons for that: Firstly, the different installations would have disturbed each other. Furthermore, the sound colour of the two different sources were, although equalized, 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*20 min each.

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. 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 14.
For each distance the assessment was repeated 4 times, except for the distances of 45, 85 and 150 cm for which it was repeated 7 times. This makes a total number of 37 test signals. Due to the existence of two different test signals (depicted in the next paragraph) 74 signals in total were presented in a random order

which was the same for all participants in both experiments.

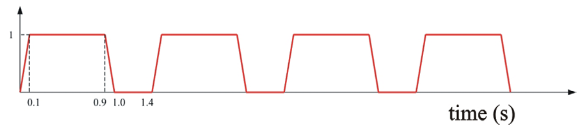

Figure 14: Envelope of the pink noise bursts used in the experiment

4. Method of "Conflicting cues"

It is known from the literature and has been informally verified by the authors that the relative loudness 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 only the perceived loudness and to check the validity of binaural cues a special method of randomly varying the receiver loudness was applied:

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. The test signals with the *randomly* chosen source level consequently had *no* natural variation of the receiver level at the listening position. Hence, the different cues used for distance perception (loudness, 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.

"Non-conflicting cues"- signals		
distance in cm	Source level in db_{rel}	Receiver level in db(A)
25	0	69.2
45	0	65.4
65	0	62.0
85	0	60.0
110	0	57.9
150	0	55.4
190	0	52.9

Table 2: "Non-conflicting cues"- signals: Source and receiver levels

"Conflicting cues"- signals		
distance in cm	Source level in db_{rel}	Receiver level in db(A)
25	- 9.2	60.0
45	- 3.4	62.0
65	- 9.1	52.9
85	- 4.6	55.4
110	+ 7.5	65.4
150	+ 13.8	69.2
190	+ 5	57.9

Table 3: "Conflicting cues"- signals: Source and receiver levels

Through this method it was possible to judge which role loudness and binaural cues play in the listener judgment of the respective sources.

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 the following tables the relevant source and receiver levels for both types of signals are shown:

5. Test geometry, room

The test was performed in the anechoic chamber of the IRT. For the test geometry see chapter 6 and Figure 5 as well as Figure 15 and Figure 16. 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 synthesized *behind* the array.

6. Elicitation of responses

Different methods for the elicitation of distance judgments from listeners are used in literature (see [3][4][5][10]). This difficult task is realized e.g. 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. To avoid this effect, a graphical elicitation method is sometimes used and the test panel is requested to draw the perceived source position on the response sheet. However, the relationship between the perceived acoustical event and the drawn figure is not straightforward.

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 15 and Figure 16. After each test signal the

Figure 15: The left side of the curtain:

The single real loudspeaker at a distance d

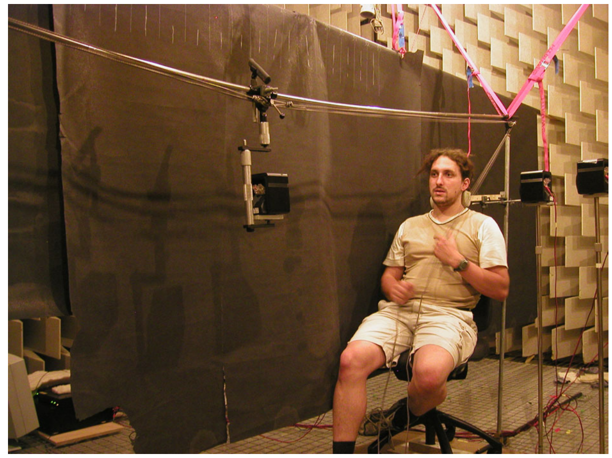

Figure 16: The right side of the curtain:

The dummy loudspeaker "cableway" is used to indicate the perceived distance

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. In a preliminary test this method was successfully checked for its validity.

7. Training of participants

As mentioned in point 1 of this chapter the required task was quite difficult for the listeners. 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 they were presented with a small set-up of three loudspeakers, located at distances of 50, 80 and 110 cm and visible to the participant. He/she was requested to toggle between the three loudspeaker 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 randomized 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.

In the beginning the participants were fairly confused by the fact that the visually perceived distance of the loudspeakers did not correspond to their auditory perception. 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 reported that they were able to perceive non-loudness cues for a distance judgment.

9. LISTENING TEST 1: DISTANCE PERCEPTION OF NEARBY REAL SOURCES

Figure 17 and Figure 18 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 indicates the number of results combined in a certain distance range. The red graph shows the mean of these results and the blue graph (which corresponds to the blue y-axis on the right) gives 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 17 that the natural test signals are perceived quite consistently, containing a 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.

Figure 18 shows the result for the "conflicting cues"-test signals. The blue curve indicates the permuted receiver level values.

The results 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 quite similar to the results of Brungart and Rabinowitz (who measured the region of d < 1 m).

The results for the "conflicting cues"-signals suggest that the direct sound cues are indeed relevant and that e.g. possible remaining reflections in the anechoic chamber play no decisive role.

10. LISTENING TEST 2: DISTANCE PERCEPTION OF NEARBY VIRTUAL SOURCES

Figure 19 and Figure 20 show the results for the virtual sources. In Figure 19 the reproduced receiver level corresponds to the reference source distance. Now, in contrast to the real sources (see Figure 17), 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 20. Once again, the range of the perceived distances is quite 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.

This is illustrated in Figure 21 and Figure 22, where the data of Figures 18 and 20 are plotted once again. Now 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 19 and Figure 22 look very much the same. This indicates that there exists no auditory distance perception cue 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 generalized to other array and signal conditions as long as the conditions which cause this fact (which are analyzed in chapter 6) do not change.

In chapter 11 it is investigated 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 optimized (see chapter 7.1) and the preconditions for auditory distance perception could be improved. The investigation by Komiyama et al. ([27]) uses such an array-design for an investigation into distance perception, its very positive conclusions, however, are not deduced from experiments under the same rigorous conditions.

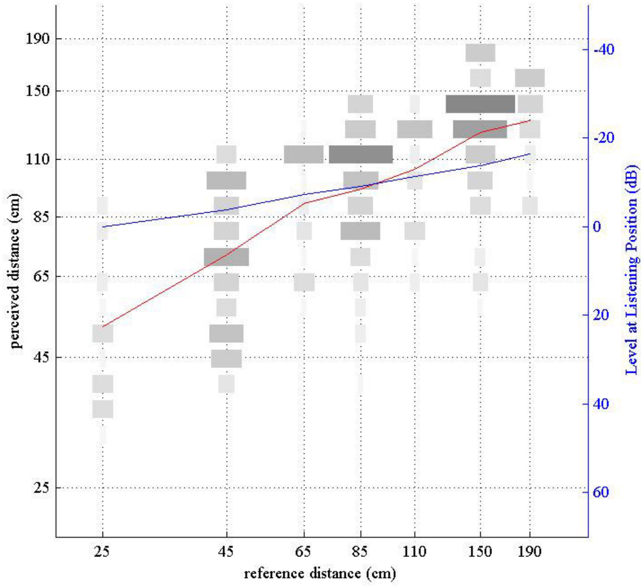

*Figure 17: **Real Sources**, natural cues*

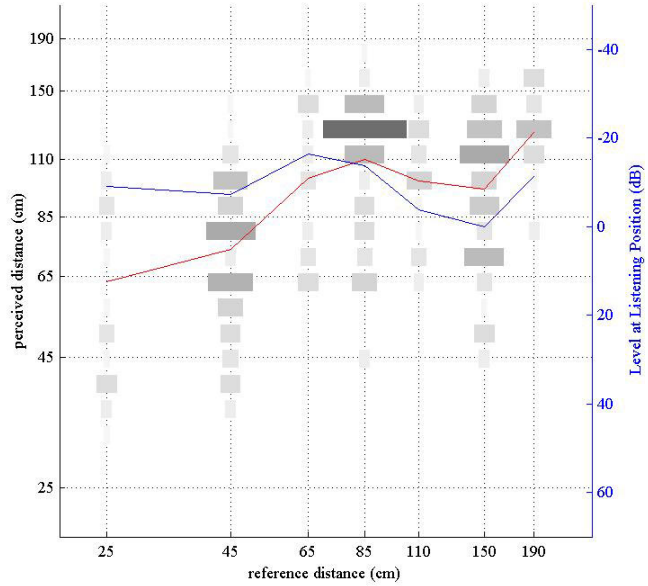

*Figure 18: **Real Sources**, conflicting cues*

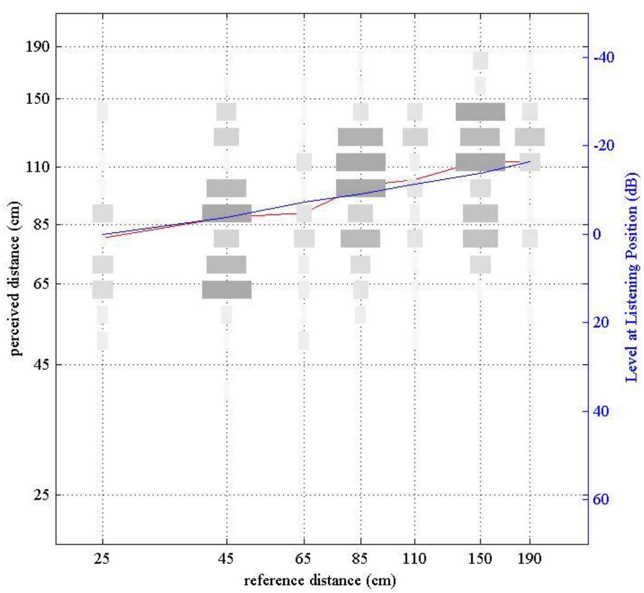

*Figure 19: **Virtual Sources**, natural cues*

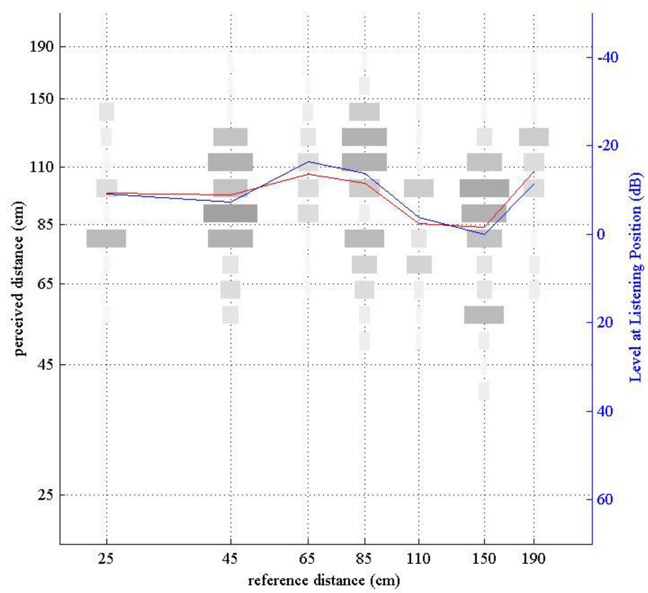

*Figure 20: **Virtual Sources**, conflicting cues*

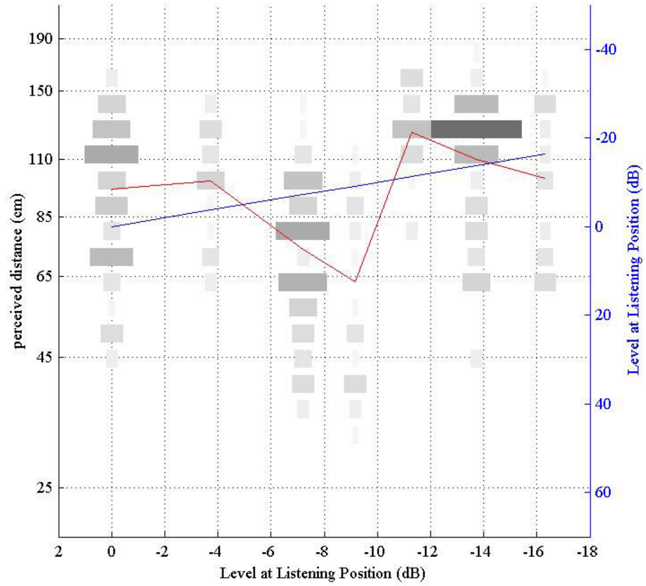

*Figure 21: **Real Sources**, conflicting cues, sorted by Level at listening position (same data as in Figure 18)*

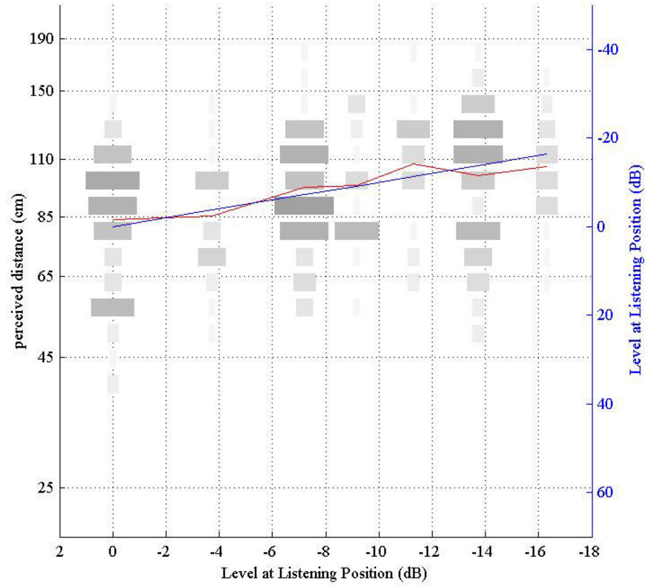

*Figure 22: **Virtual Sources**, conflicting cues, sorted by Level at listening position (same data as in Figure 20)*

11. AN APPROACH TO SIMULATE HEAD-SHADOWING EFFECTS IN WFS

As discussed in chapter 5, the ILD (Interaural Level Differences) can be a cue for the distance perception of close sources. However, as shown in the last chapters, a "normal" linear WFS array does not create sufficient ILD for the listener.

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 7.1 and as can be seen in Figure 10. Furthermore, for low frequencies, both the level and 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 significant ILD could be synthesized by a longer array.

Hence, another simulation setup was created using a array of the length 21.3 m and a decreased interspacing of (0.17/4) m = 4.25 cm. This results in a number of array loudspeakers of n=501.

The new simulation setup enables a closer view on the characteristics of the WFS sound field.

This "Super-Array" shows low frequency artefacts as well but it is capable of reproducing a flat frequency response of focused sources for frequencies above ca 1 kHz. This can be seen from Figure 23. This Figure can be well compared with Figure 10, where the responses of the normal short WFS array are shown. With a longer array, at the price of additional ripples, a reproduction of lower frequencies is achieved.

As a result, also significant level differences can be produced at lower frequencies (Figure 24). According to the theoretical considerations of chapter 7.1., the level differences are smaller than those caused by real sources (see Figure 9).

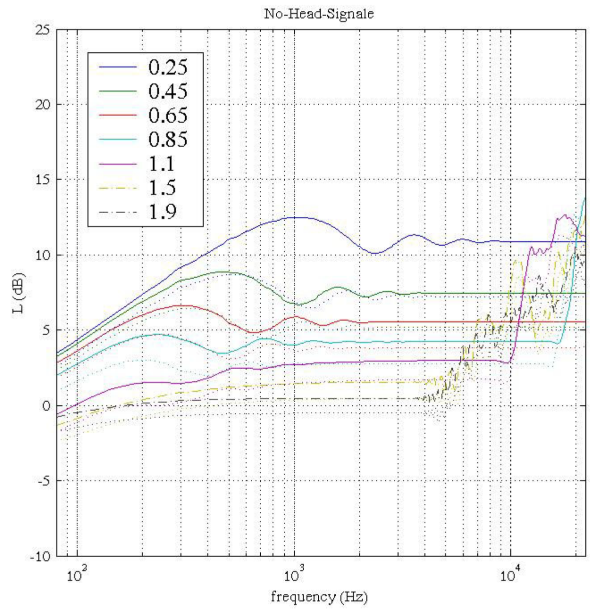

*Figure 23: Super-Array: **Level** of a **focused WFS virtual source** at distances = d +/- (ear distance/2)*

Solid line: right ear, dotted line: left ear

Dash-dotted lines: non-focused sources

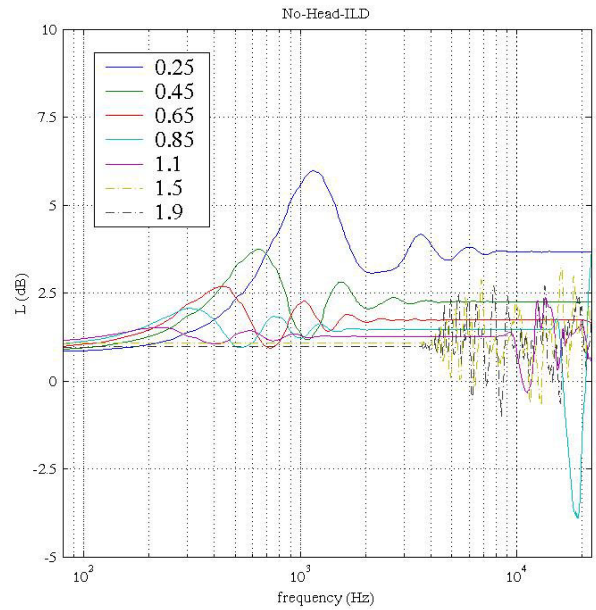

*Figure 24: Super-Array: **"No-Head-ILD": Level differences ΔL** between ear positions in the sound field of a **focused WFS virtual source** at distance d.*

Dash-dotted lines: non-focused sources

11.1. "Head-Shadowing"

In this paper, the ILD is considered as a result of two different parameters:

1. The level differences due to the different distances of the two ears to the source (max. 17 cm). This level difference is called "No-Head-ILD" in this paper.
2. Pinna effects as well as "Head Shadowing", leading to an increase of the "No-Head-ILD". They are, for the sake of simplicity, called "Head-Shadowing effects".

Depending on the source distance, each has a different influence on the ILD.
The "No-Head-ILD" can be calculated easily, as shown in chapter 7.2. The influence of the "Head-Shadowing", on the other side, can be deduced from measurements of ILD and "No-Head-ILD". In the approach of this paper, 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 gives an opportunity to compare real and virtual sound field.

11.2. Comparison of Real and Virtual Source

With the help of the parameter "Head-Shadowing effect" the influence of the head being in the sound field can be studied. ILD derived from measurements with a dummy head being in the sound field of a real source in an anechoic chamber are shown in Figure 25 (This is the same as Figure 12, the head is turned -90° to the source).

The "Head-Shadowing effect", derived by calculating the difference between the ILD and the "No-Head-ILD" is shown in Figure 27.

It can be seen that the Head-Shadowing effect, similar to the ILD in general, significantly differs only for very close sources (< 65 cm). Presumably, when the source is close the head, head diffraction differs significantly compared with that of a more distant source.

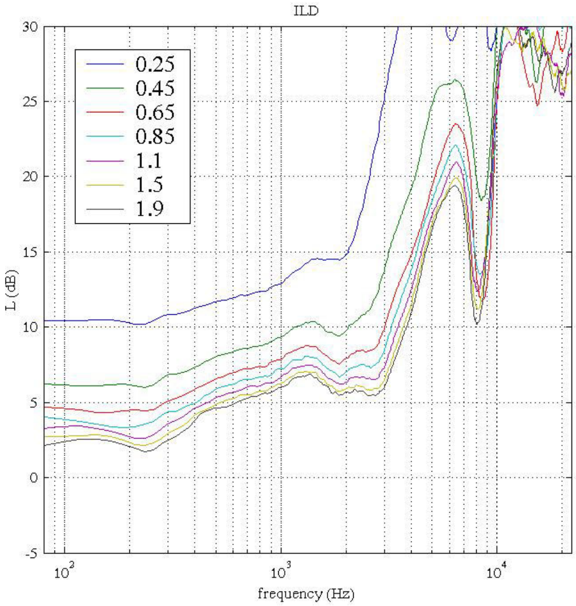

*Figure 25: **Interaural Level differences ILD** in the sound field of a **real source** at distance d.*

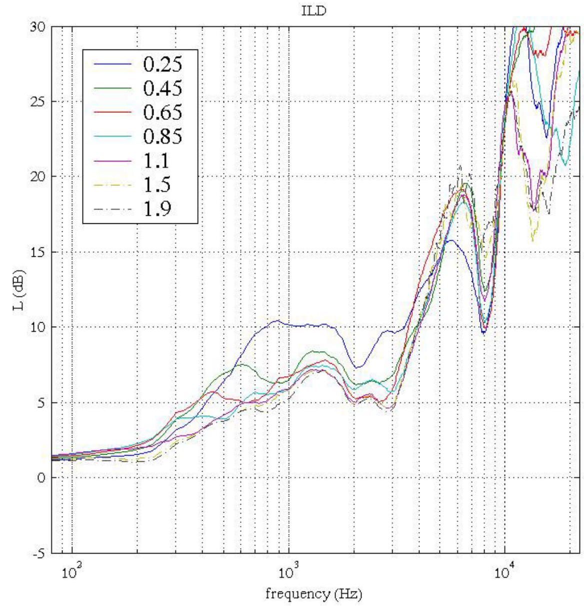

*Figure 26: Super-Array: **Interaural Level differences ILD** in the sound field of a **focused source** at distance d.*

Dash-dotted lines: non-focused sources

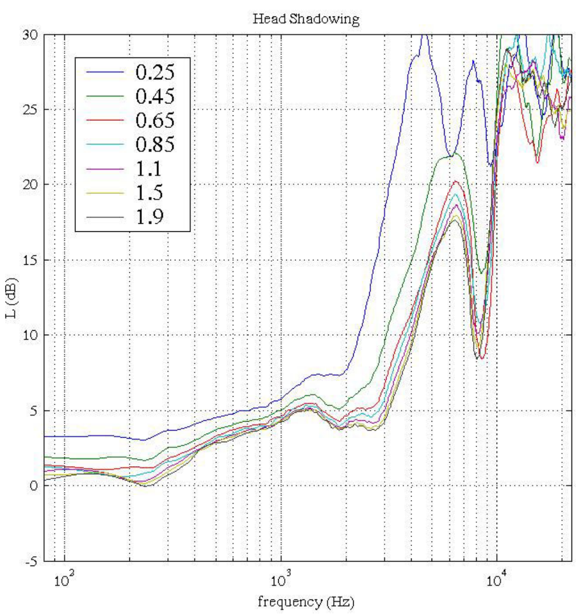

*Figure 27: **Head-Shadowing effect** in the sound field of a **real source** at distance d*

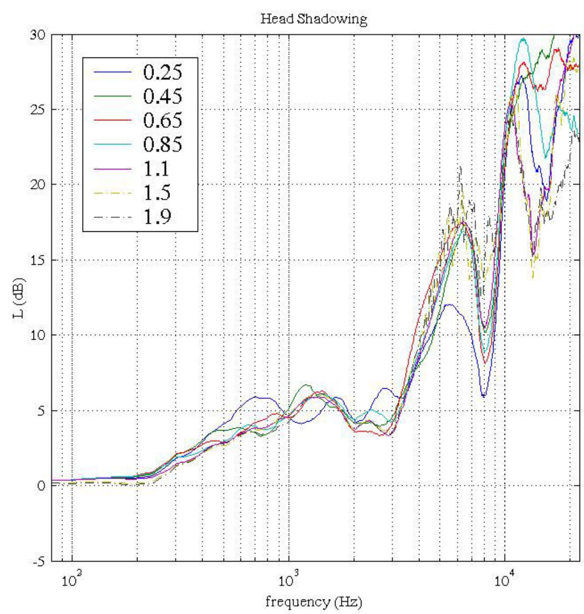

*Figure 28: Super-Array: **Head-Shadowing effect** in the sound field of a **focused source** at distance d.*

Dash-dotted lines: non-focused sources

In the frequency band from 1 to 5 kHz the Head-Shadowing effect can create an additional level difference of circa 5 dB for a source in 45 cm distance. It is plausible that this difference can serve as an auditory cue.

The virtual sources are analyzed in Figure 26 and Figure 28.

Figure 26 shows the ILD for various source distances derived from virtual sources. It can be seen that indeed the ILD increase with decreasing distance, albeit not as strong as for the real sources, which are analyzed in Figure 25. This corresponds to the results of the previous figures. The differences from Figure 24 can be identified in Figure 26 rather well.
A further view on the virtual source's characteristics offers Figure 28. Here the Head-Shadowing effect is presented. In contrast to the real source, the Head-Shadowing effects of the virtual sources show nearly no dependency 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 be a consequence of inexact measurements as well, but this cannot be deduced from these simulations.

It is interesting that the Head-Shadowing effect is, apart from the small ripples, the same for all source distances. Although derived from different test setups (e.g. different loudspeakers at the measurements) a comparison of Figures 27 and 28 suggests that all virtual sources create the same Head-Shadowing, which is the Head-Shadowing of a real source in a distance of > 1m.
A possible deduction from this observation is: The Head-Shadowing effect of a focused virtual source is only dependant on the distance of the *reproduction array*. As this deduction is quite audacious, the author requests (other) explanations for the outcome of these simulations from other sides.
This does not mean that the ILD are the same for all virtual source distances. It can be seen from Figure 26 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.

12. CONCLUSION

The spatial properties of WFS focused sources, in particular with regard to distance perception, were considered in theory and through listening tests. Although a number of conditions mentioned in chapters 3 and 4 influence these properties, this investigation could illuminate parameters which are relevant, and parameters which are irrelevant for the perception of distance in WFS.
The effects of Head Shadowing, together with the physics of WFS were analyzed to gain an insight into the potential of WFS virtual (esp. focused) sources to create a sense of distance.

It was found for the dry virtual sources of the experiment that both theory and experiments suggest that certain relevant cues for the perception of the distance are not existent. A further simulation (chapter 11), optimizing low frequency reproduction, brought up the question of how Head Shadowing is synthesized in WFS.

The study concentrated on dry sources which are in this case considered to explain the meaning of the direct sound (the first wave front) of a virtual source.
As discussed in chapter 5 the direct sound only partially offers auditory cues for distance perception. These cues are existent and relevant for dry real sources as shown in theory and practice.

A way to overcome 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 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.

13. ACKNOWLEDGMENTS

We would like to thank Prof. Fastl from the AG Technische Akustik, MMK, Technische Universität München, Munich, Germany, for his support. Furthermore, the help and the discussions with Tobias Neher and John Oag were essential for this paper.

14. REFERENCES

[1] Berkhout, A. J.; de Vries, D., 1993; Vogel, P: "Acoustic Control by Wave Field Synthesis," J. Acoust. Soc. Am., Vol. 93, pp. 2764–2778, 1993.

[2] Theile, G., Wittek, H., Reisinger, M., 2003: "Potential Wavefield Synthesis Applications in the Multichannel World", Proc. AES 24th Int. Conference on Multichannel Audio, Banff, Canada

[3] Nielsen, S., 1991: "Distance Perception in Hearing", Dissertation Aalborg University, Aalborg University Press, Aalborg, Denmark, ISBN 87-7307-447-0

[4] Zahorik, P., 1996: “Auditory distance perception: A literature review”, Paper for fulfilment of PhD, Wisconsin University

[5] Shinn-Cunningham, B.G., 2000: “Distance cues for virtual auditory spaces”, Pacific-Rim on Multimedia, Sydney, Australia, 2000

[6] Blauert, J., 1997: “Spatial Hearing”, MIT Press, ISBN 0-262-02413-6

[7] Rumsey, F., 2002: “Spatial Quality Evaluation for reproduced Sound: Terminology, Meaning, and a Scene-based paradigm”, Journal of the AES, Vol.50, No.9

[8] Ausserhofer, A. et al., 1995: “Human depth perception”, website, http://www2.iicm.edu/stereograms

[9] Wittek, H., 2003: “Perception of Spatially synthesized sound fields – Literature review about WFS”, available online at: http://www.hauptmikrofon.de/wfs.htm

[10] Brungart and Rabinowitz, 1999: „Auditory localization of nearby sources, I, II and III”, J. Acoust. Soc. Am., Vol. 106, No. 3,4 and 6.

[11] Verheijen, E., 1998: “Sound reproduction by Wave Field Synthesis”, PhD Thesis, TU Delft

[12] Boone, M.M., 2004: personal communication.

[13] Boone, M.M., Verheijen, E., Jansen, G., 1996: “Virtual reality by sound reproduction based on Wave Field Synthesis”, Preprint No. 4145, AES 100th convention, Copenhagen, Denmark

[14] de Vries, D. and Berkhout, A.J., 1981: “Wave theoretical approach to acoustical focusing”, J. Acoust. Soc. Am. 70(3), pp.740-748

[15] Huber, T.: „Untersuchungen zur Lokalisation virtueller Quellen bei Wellenfeldsynthese”. Draft of the Diploma work IRT/ TU München, 2001, available at: http://www.hauptmikrofon.de/wfs.htm

[16] Theile, G.: "On the Naturalness of Two-Channel Stereo Sound". J. Audio Eng. Soc. 39, No. 10, 1991, S. 761 – 767

[17] Theile, G.: “Über die Lokalisation im überlagerten Schallfeld” (German), Dissertation, TU Berlin, 1980, available at: http://www.hauptmikrofon.de/theile.htm

[18] Pulkki, V., 2001: “Spatial Sound Generation and Perception by Amplitude Panning Techniques”, Dissertation Helsinki University of Technology, ISBN 951-22-5531-6

[19] Spors, S., Kuntz, A., Rabenstein, R., 2003: „An Approach to Listening Room Compensation with Wavefield Synthesis”, Proc. AES 24th Int. Conference on Multichannel Audio, Banff, Canada

[20] Corteel E., Nicol R., “Listening room compensation for Wave Field Synthesis. What can be done?” AES 23rd Int. Conf., Copenhagen, Denmark, May 2003.

[21] Corteel, E., Kuhn, C., 2003, expected at AES New YORK

[22] Theile, G.: "Multichannel natural music recording based on psychoacoustic principles". AES 19th Int. Conf. on Surround Sound, Elmau, Germany, June 2001, new version: http://www.hauptmikrofon.de/theile.htm

[23] de Vries, D.; Hulsebos, E.; Bourdillat, E.: “Auralization by Wave Field Synthesis”. 21. Tonmeistertagung Hannover, 2000, S. 121-128

[24] Hamasaki, K., Hiyama, K., 2003: ”Reproducing Spatial Impression by Multichannel Audio”, AES 24th Int. Conf. On Multichannel Audio, Banff, Canada, June 2003

[25] Griesinger, D., 2001: “The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique”, AES 19th Int. Conf. On Surround Sound, Elmau, Germany, June 2001

[26] Wittek, H., Neumann, O., Schäffler, M. and Millet, C.: “Studies on Main and Room Microphone Optimization”, AES 19th Int. Conf. on Surround Sound, Elmau, Germany, June 2001

[27] Komiyama, S., 1997: “Distance Control System for a Sound Image”, AES 9th Int. Conf., February 1991, p.233 f

[28] Pellegrini, R., van Zon, R., 2002: “Comparing measured room-impulse responses to virtual rooms for sound reproduction using wave field synthesis”, 22. Tonmeistertagung Hannover, 2002