

# SPATIAL ENCODING AND DECODING OF FOCUSED VIRTUAL SOUND SOURCES

Jens Ahrens, Sascha Spors

Deutsche Telekom Laboratories, Technische Universität Berlin  
Ernst-Reuter-Platz 7, 10587 Berlin, Germany  
({jens.ahrens,sascha.spors}@telekom.de)

**Abstract:** *Spatial encoding refers to the representation of a sound field which allows storage and transmission of the latter. In the Ambisonics context, sound fields can be spatially encoded when their spherical wave spectrum is band-limited. The process of deriving appropriate loudspeaker driving signals in order to reproduce an encoded sound field is known as spatial decoding. Care has to be taken when virtual sound sources are positioned such that they appear inside a given loudspeaker setup for which they are decoded. The properties of the mathematical formulation make the reproduced sound field deviate strongly from the desired one in certain receiver positions. In this contribution we demonstrate by means of a two-dimensional scenario how the concept of focused virtual sound sources can be applied in order to optimize the reproduction accuracy.*

Key words: sound field reproduction, higher order Ambisonics, circular harmonics, focused sources

## 1 INTRODUCTION

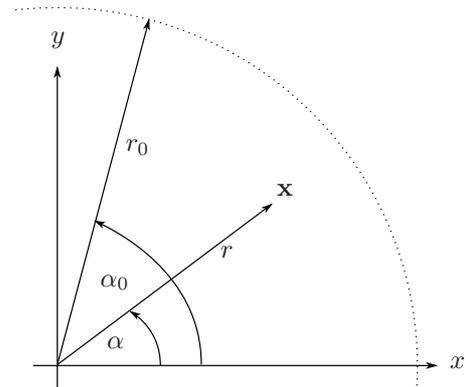
Sound field reproduction techniques like higher order Ambisonics and wave field synthesis employ a large number of loudspeakers to physically reproduce a desired sound field over an extended listening area. Theoretically, these methods are only capable of reproducing virtual sound sources which are positioned outside of the listening area (“behind the loudspeakers”). By reproducing a sound field which converges in one half-space towards a focus point and diverges in the other half-space, the *target half-space*, the perception of virtual sound sources inside the listening area can be elicited for listeners in the diverging part of the sound field. Such a situation is referred to as reproduction of a *focused virtual sound source*.

While being an established technique in wave field synthesis, e.g. [1, 2], focused sources have received far less attention in higher order Ambisonics and related approaches. To our awareness, publicly available work is restricted to [3, 4, 5].

In this contribution, we revisit the published approaches and present a comparison of properties and restrictions. We concentrate on the peculiarities which arise in the spatial encoding and decoding procedures which are widely employed in the Ambisonics-like approaches.

## 2 NOMENCLATURE AND MATHEMATICAL PRELIMINARIES

For convenience, we restrict our considerations to two spatial dimensions. This means in this context that a sound field under consideration is independent from one of the spatial



**Figure 1:** The coordinate system used in this paper. The dashed line indicates the secondary source distribution.

coordinates, i.e.  $P(x, y, z, \omega) = P(x, y, \omega)$ . Refer to Sec. 3 for an outline of the consequences of this assumption.

The two-dimensional position vector in Cartesian coordinates is given as  $\mathbf{x} = [x \ y]^T$ . The Cartesian coordinates are linked to the polar coordinates via  $x = r \cos \alpha$  and  $y = r \sin \alpha$ . Refer to the coordinate system depicted in Fig. 1.

The acoustic wavenumber is denoted by  $k$ . It is related to the temporal frequency by  $k^2 = (\frac{\omega}{c})^2$  with  $\omega$  being the radial frequency and  $c$  the speed of sound. Outgoing monochromatic cylindrical waves are denoted by  $H_0^{(2)}(\frac{\omega}{c}r)$ . The imaginary unit is denoted by  $j$  ( $j = \sqrt{-1}$ ). A propagating two-dimensional sound field  $P(\mathbf{x}, \omega)$  can be

by described by its circular harmonics expansion as [6]

$$P(\mathbf{x}, \omega) = \sum_{\nu=-\infty}^{\infty} \underbrace{\check{P}_{\nu}(\omega) J_{\nu}\left(\frac{\omega}{c}r\right)}_{=\check{P}_{\nu}(r, \omega)} e^{j\nu\alpha}, \quad (1)$$

whereby  $J_{\nu}(\cdot)$  denotes the  $\nu$ -th order Bessel function. The Fourier series expansion coefficients  $\check{P}_{\nu}(r, \omega)$  of  $P(\mathbf{x}, \omega)$  can be obtained via [6]

$$\check{P}_{\nu}(r, \omega) = \frac{1}{2\pi} \int_0^{2\pi} P(\mathbf{x}, \omega) e^{-j\nu\alpha} d\alpha. \quad (2)$$

### 3 SOUND FIELD REPRODUCTION

In this section, we briefly review the general approach presented by the authors in [7, 8]. Its physical fundament is the so-called *simple source approach* and it can be seen as an analytical formulation of what is known as higher order Ambisonics (see e.g. [9]). The simple source approach for interior problems states that the acoustic field generated by events outside a volume can also be generated by a continuous distribution of secondary simple sources enclosing the respective volume [6].

As stated in section 2, we limit our derivations to two-dimensional reproduction for convenience. Furthermore, we assume the distribution of secondary sources to be circular. In order to fulfill the requirements of the simple source approach and therefore for artifact-free reproduction, the sound fields emitted by the secondary sources have to be two-dimensional. We thus have to assume a continuous circular distribution of secondary line sources positioned perpendicular to the target plane (the receiver plane) [6]. Our approach is therefore not directly implementable since loudspeakers exhibiting the properties of line sources are commonly not available. Real-world implementations usually employ loudspeakers with closed cabinets as secondary sources. The properties of these loudspeakers are more accurately modeled by point sources.

The main motivation to focus on two dimensions is to keep the mathematical formulation simple in order to illustrate the general principle of the presented approach. The extension both to three-dimensional reproduction (i.e. spherical arrays of secondary point sources) and to two-dimensional reproduction employing circular arrangements of secondary point sources is straightforward and can be found e.g. in [7].

#### Derivation of the secondary source driving function

The reproduction equation for a continuous circular distribution of secondary line sources and with radius  $r_0$  centered around the origin of the coordinate system is given by

$$P(\mathbf{x}, \omega) = \int_0^{2\pi} D(\alpha_0, \omega) G_{2D}(\alpha - \alpha_0, r, \omega) r_0 d\alpha_0, \quad (3)$$

where  $\mathbf{x}_0 = r_0 \cdot [\cos \alpha_0 \sin \alpha_0]^T$ .  $P(\mathbf{x}, \omega)$  denotes the reproduced sound field,  $D(\alpha_0, \omega)$  the driving function for the secondary source situated at  $\mathbf{x}_0$ , and  $G_{2D}(\alpha - \alpha_0, r, \omega)$  its two-dimensional spatio-temporal transfer function.

A fundamental property of (3) is its inherent non-uniqueness and ill-posedness [10]. I.e. in certain situations, the solution is undefined and so-called *critical* or *forbidden frequencies* arise. The forbidden frequencies are discrete and represent the resonances of the cavity under consideration. However, there are indications that the forbidden frequencies are only of minor relevance when practical implementations are considered [6].

Equation (3) constitutes a circular convolution and therefore the convolution theorem

$$\check{P}_{\nu}(r, \omega) = 2\pi r_0 \check{D}_{\nu}(\omega) \check{G}_{\nu}(r, \omega) \quad (4)$$

applies [11].  $\check{P}_{\nu}(r, \omega)$ ,  $\check{D}_{\nu}(\omega)$ , and  $\check{G}_{\nu}(r, \omega)$  denote the Fourier series expansion coefficients of  $P(\mathbf{x}, \omega)$ ,  $D(\alpha, \omega)$ , and  $G_{2D}(\mathbf{x} - [r_0 \ 0]^T)$ <sup>1</sup>.

From (4) and (1) we can deduce that

$$\check{D}_{\nu}(\omega) = \frac{1}{2\pi r_0} \frac{\check{P}_{\nu}(r, \omega)}{\check{G}_{\nu}(r, \omega)} = \quad (5)$$

$$= \frac{1}{2\pi r_0} \frac{\check{P}_{\nu}(\omega) \cdot J_{\nu}\left(\frac{\omega}{c}r\right)}{\check{G}_{\nu}(\omega) \cdot J_{\nu}\left(\frac{\omega}{c}r\right)}. \quad (6)$$

For  $J_{\nu}\left(\frac{\omega}{c}r\right) \neq 0$  the Bessel functions in (6) cancel out directly. Wherever  $J_{\nu}\left(\frac{\omega}{c}r\right) = 0$  de l'Hôpital's rule [13] can be applied to proof that the Bessel functions also cancel out in these cases, thus making  $\check{D}_{\nu}(\omega)$  and therefore also  $D(\alpha_0, \omega)$  independent from the receiver position.

Introducing the result into (1) finally yields the secondary source driving function  $D(\alpha_0, \omega)$  for a secondary source situated at position  $\mathbf{x}_0$  reproducing a desired sound field with expansion coefficients  $\check{P}_{\nu}(\omega)$  reading

$$D(\alpha, \omega) = \frac{1}{2\pi r_0} \sum_{\nu=-\infty}^{\infty} \frac{\check{P}_{\nu}(\omega)}{\check{G}_{\nu}(\omega)} e^{j\nu\alpha}, \quad (7)$$

whereby we omitted the index 0 in  $\alpha_0$ .

We assume monopole line sources in the remainder of this paper for convenience. The two-dimensional free-field Green's function  $G_{2D}(\mathbf{x} - \mathbf{x}_0, \omega)$  representing the spatio-temporal transfer function of a secondary source at position  $\mathbf{x}_0$  is then the zero-th order Hankel function of second kind  $\frac{i}{4} H_0^{(2)}\left(\frac{\omega}{c} |\mathbf{x} - \mathbf{x}_0|\right)$  [6].

Equation (7) can be verified by inserting it into (3). After introducing the Fourier series expansion of the secondary source sound fields according to (1), exchanging the order of integration and summation, and exploitation of the orthogonality of the circular harmonics  $e^{j\nu\alpha}$  [6] one arrives at the Fourier series expansion of the desired sound field, thus proving perfect reproduction. Note however that the coefficients  $\check{P}_{\nu}(\omega)$  respectively  $\check{G}_{\nu}(\omega)$  are typically derived from interior expansions. This implies that the desired sound field is only correctly reproduced inside the secondary source distribution.

<sup>1</sup>Note that the coefficients  $\check{G}_{\nu}(r, \omega)$  as used throughout this paper assume that the secondary source is situated at the position ( $r = r_0, \alpha = 0$ ) and is orientated towards the coordinate origin [7, 12]

The infinite summation in (7) can of course not be performed in practical implementations. Therefore, a bandlimited approximation  $D_N(\alpha, \omega)$  of  $D(\alpha, \omega)$  is employed reading

$$D_N(\alpha, \omega) = \frac{1}{2\pi r_0} \sum_{\nu=-N}^N \frac{\check{P}_\nu(\omega)}{\check{G}_\nu(\omega)} e^{j\nu\alpha} \approx D(\alpha, \omega). \quad (8)$$

As a consequence of the discretization of the secondary source distribution which has to be performed in practical implementations, the bandwidth  $N$  of  $D_N(\alpha, \omega)$  is chosen according to the desired properties in terms of spatial aliasing [8]. One also speaks of  $N$ -th order reproduction.

For convenience, we exclusively employ continuous secondary source distributions and choose  $N$  either arbitrarily or according to the limitations which arise in the situation under consideration.

#### 4 SPATIAL ENCODING AND DECODING

The spatial sound field encoding and decoding procedure outlined in this section was introduced in the context of Ambisonics [9]. The encoding procedure yields a representation of a sound scene which is independent from the loudspeaker geometry and allows for the storage and transmission of the sound scene. The decoding procedure yields the loudspeaker driving signals for an encoded scene for a given loudspeaker distribution. Note that both model-based and data-based sound scenes can be encoded as explained in Sec. 6 and Sec. 7.

A two-dimensional sound field  $P(\mathbf{x}, \omega)$  to be reproduced can be spatially encoded when it is known on a circle with radius  $r_{\text{ref}}$  and when it is spatially bandlimited:

$$P(\mathbf{x}, \omega) = \sum_{\nu=-N}^N \check{P}_\nu(r_{\text{ref}}, \omega) e^{j\nu\alpha}, \quad (9)$$

$P(\mathbf{x}, \omega)$  has to be free of sound sources for  $r < r_{\text{ref}}$ . The latter is a crucial condition as we will explain in detail in Sec. 5.1.

The correspondences of the coefficients  $\check{P}_\nu(r_{\text{ref}}, \omega)$  in the time domain<sup>2</sup> can be stored and transmitted [9]. Note that  $\check{P}_\nu(r_{\text{ref}}, \omega)$  is also referred to as circular wave spectrum [6]. It is not advisable to store the coefficients  $\check{P}_\nu(\omega)$  (refer to (1)), since they diverge at low frequencies for all expansion orders  $\nu \neq 0$  [14, 15].

In order to decode (i.e. in order to reproduce) an encoded sound field on a given secondary source distribution, we recall (5) and (6) from Sec. 3. It was shown that calculating the secondary source driving function  $D(\alpha, \omega)$  via the coefficients  $\check{P}_\nu(\omega)$  and  $\check{G}_\nu(\omega)$  in (5) is equivalent to calculating  $D(\alpha, \omega)$  via the Fourier series coefficients  $\check{P}_\nu(r, \omega)$  and  $\check{G}_\nu(r, \omega)$  in (6). It is therefore legitimate to determine

<sup>2</sup>In the Ambisonics context, these signals are termed *Ambisonics signals* and are defined slightly differently [9]. The long evolution of Ambisonics techniques has introduced a number of conventions which we occasionally elide for didactic reasons.

$D(\alpha, \omega)$  (i.e. to decode  $P(\mathbf{x}, \omega)$ ) via

$$D(\alpha, \omega) = \sum_{\nu=-N}^N \frac{1}{2\pi r_0} \frac{\check{P}_\nu(r_{\text{ref}}, \omega)}{\check{G}_\nu(r_{\text{ref}}, \omega)} e^{j\nu\alpha}. \quad (10)$$

The coefficients  $\check{G}_\nu(r_{\text{ref}}, \omega)$  represent the properties of the secondary source distribution under consideration, i.e. its radius  $r_0$  and radiation characteristics [12].

The encoding/decoding procedure presented above differs from the classic higher order Ambisonics procedure as presented e.g. in [9]. The main drawback of the latter is the fact that the secondary sources are assumed to be omnidirectional. As shown in [12], this assumption is an unnecessary restriction.

The decoding operation as represented by (10) is not directly applicable since the denominator  $\check{G}_\nu(r_{\text{ref}}, \omega)$  can exhibit zeros especially when a large frequency range is considered. Elaborating the presented encoding/decoding scheme to reach practicability is beyond the scope of this paper and will be published in the near future. We stick on the presented procedure for the purpose of illustrating the general idea of spatial encoding and decoding.

#### 5 FOCUSED VIRTUAL SOUND SOURCES

##### 5.1. Limitations of the reproduction of non-focused virtual sound sources

Before we introduce the concept of focused virtual sound sources, we review the limitations of conventional (non-focused) virtual sound sources which make the employment of focused virtual sound sources necessary. Exemplarily, we assume the virtual source as well as the secondary sources to be monopole line sources.

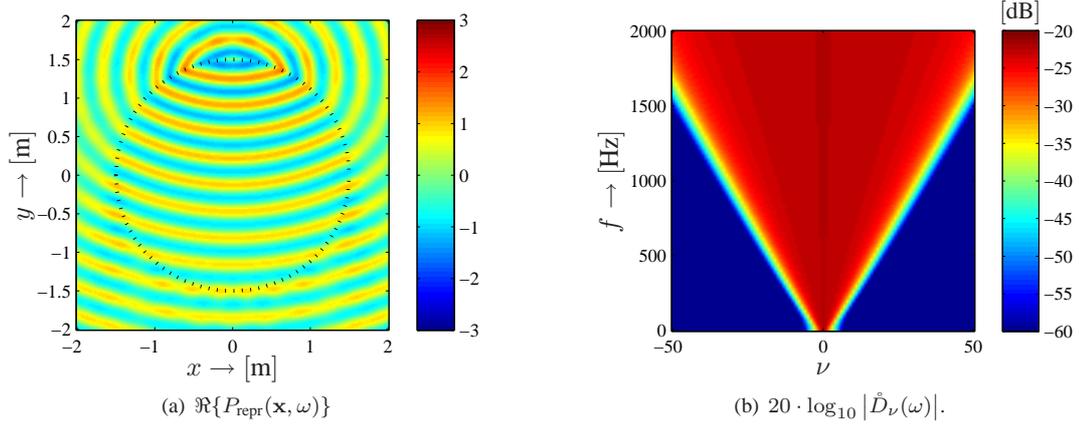
The sound field  $S(\mathbf{x} - \mathbf{x}_s, \omega)$  of such a monopole line source situated at position  $\mathbf{x}_s$  is given by [15]

$$S(\mathbf{x} - \mathbf{x}_s, \omega) = \frac{j}{4} H_0^{(2)} \left( \frac{\omega}{c} |\mathbf{x} - \mathbf{x}_s| \right) = \begin{cases} \underbrace{\sum_{\nu=-\infty}^{\infty} \frac{j}{4} H_\nu^{(2)} \left( \frac{\omega}{c} r_s \right) e^{-j\nu\alpha_s} J_\nu \left( \frac{\omega}{c} r \right) e^{j\nu\alpha}}_{\check{S}_\nu(\omega)} & \text{for } r \leq r_s \\ \sum_{\nu=-\infty}^{\infty} \frac{j}{4} J_\nu \left( \frac{\omega}{c} r_s \right) e^{-j\nu\alpha_s} H_\nu^{(2)} \left( \frac{\omega}{c} r \right) e^{j\nu\alpha} & \text{for } r_s \leq r \end{cases}. \quad (11)$$

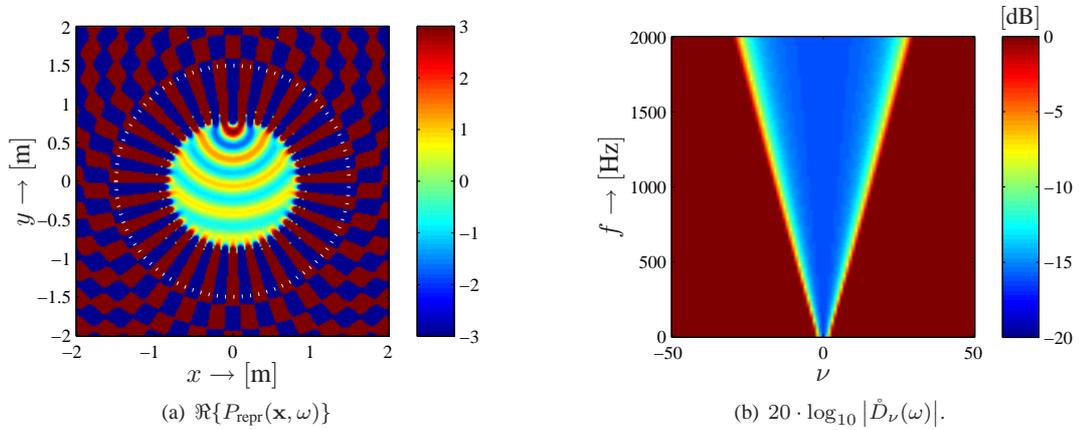
In other words, the coefficients  $\check{S}_\nu(\omega)$  which the driving function features are only valid for  $r \leq r_s$ . As long as the radius  $r_0$  of the secondary source distribution under consideration is smaller than  $r_s$ , the coefficients  $\check{S}_\nu(\omega)$  are valid over the entire receiver area and no problems arise. This situation is illustrated in Fig. 2(a).

However, for  $r_0 > r_s$  the sound field

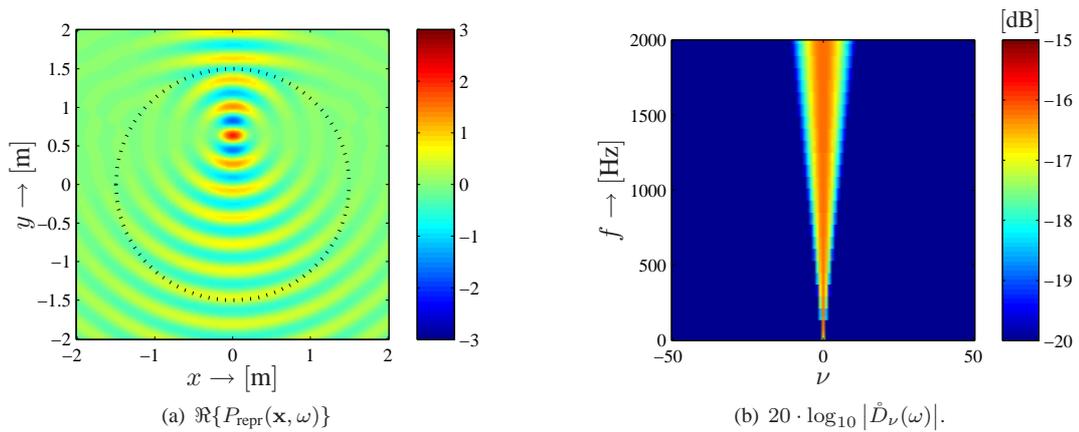
$$P_{\text{repr}}(\mathbf{x}, \omega) = \sum_{\nu=-\infty}^{\infty} \check{S}_\nu(\omega) J_\nu \left( \frac{\omega}{c} r \right) e^{j\nu\alpha} \quad (12)$$



**Figure 2:** Illustration of a continuous circular distribution of secondary line sources with a radius of  $r_0 = 1.5$  m reproducing a virtual line source at position  $(r_s = 3$  m,  $\alpha_s = \frac{\pi}{2}$ ). The emitted frequency is  $f = 1000$  Hz. The values are clipped as indicated by the colorbar. The dotted line indicates the secondary source distribution.  $N = 27$ .



**Figure 3:** Illustration of a continuous circular distribution of secondary line sources with a radius of  $r_0 = 1.5$  m reproducing a virtual line source at position  $(r_s = 0.7$  m,  $\alpha_s = \frac{\pi}{2}$ ). The emitted frequency is  $f = 1000$  Hz. The values are clipped as indicated by the colorbar. The dotted line indicates the secondary source distribution.  $N = 27$ .



**Figure 4:** Illustration of a continuous circular distribution of secondary line sources with a radius of  $r_0 = 1.5$  m reproducing a virtual line source at position  $(r_s = 0.7$  m,  $\alpha_s = \frac{\pi}{2}$ ). The emitted frequency is  $f = 1000$  Hz. The values are clipped as indicated by the colorbar. The dotted line indicates the secondary source distribution.  $N = 16$  and angular weighting is applied.

is reproduced at all locations where  $r < r_0$ . This situation is illustrated in Fig. 3(a). For  $r \leq r_s$ ,  $P_{\text{repr}}(\mathbf{x}, \omega)$  does coincide with the desired sound field and no problems arise at first sight.

However, at  $r_s < r < r_0$ ,  $P_{\text{repr}}(\mathbf{x}, \omega)$  does not coincide with the desired sound field [3, 14]. Actually, it is not such that the chosen method fails. The method is driven to reproduce (12) and this is what it does. The crucial point is that (12) does not perfectly represent what is desired. It does only represent a point source at  $\mathbf{x}_s$  for  $r \leq r_s$ . For  $r_s < r$ , (12) is mathematically well defined but its physical meaning is very different from the sound field of a point source.

Of course, the smaller  $r_s$ , the smaller is the region of accurate reproduction. Clearly, for  $r_s = 0$  reproduction fails entirely.

It has to be noted that in the sound field depicted in Fig. 3(a), the secondary sources are driven at extremely high levels and a significant amount of destructive interference takes place at locations  $r \leq r_s$ . The reproduced sound field is therefore very sensitive towards misplacement and mismatch of loudspeakers when a practical implementation is considered.

## 5.2. Focused virtual sources by angular weighting

The results presented in this section have partly been derived in [3] whereby the investigation was driven by aspects of implementation. In personal correspondence the author of [5] announced the outline of an approach employing a modification of the encoding procedure in order to avoid excessive energy components in the decoded signals leading to similar properties of the reproduced sound field. The latter approach will focus on filter design aspects in order to achieve an efficient implementation.

In this section, we revisit the subject of [3, 5] and treat it from a physical perspective in order to illustrate the basic properties of the reproduced sound field.

A closer look at the properties of (12) for  $r > r_s$  shows that it is actually the higher orders which introduce a high amount of energy at low frequencies into the driving function [14]. Compare Fig. 2(b) and Fig. 3(b).

From Fig. 2(b) it is evident that for sources outside the listening area, the energy of the driving function fades out towards high orders. On the contrary, for sources inside the secondary source distribution the energy is lowest at low orders and steadily rises towards high orders. Refer to Fig. 3(b).

Fortunately, only the lower orders are required in order to achieve accurate reproduction around the center of the secondary source distribution. It was therefore proposed in [3] to use only these lower orders in order to get rid of the high energy components in the driving function of virtual sources inside the secondary source distribution.

Sharply bandlimiting the driving function is not optimal because it leads to an uneven amplitude distribution of the reproduced sound field. Better results are obtained when higher orders are smoothly faded out. We refer to the procedure described above as *angular weighting*.

Fig. 4 is the correspondence to Fig. 3 whereas in the former, a cosine shaped window was applied in the angular frequency domain. This window is given by

$$w_\nu = \begin{cases} \frac{1}{2} \left( \cos \left( \frac{\nu}{N(f)} \pi \right) + 1 \right) & \forall |\nu| \leq N(f) \\ 0 & \text{elsewhere} \end{cases} . \quad (13)$$

The bandwidth  $N(f)$  of the driving function was chosen for each frequency  $f$  under consideration such that those orders exhibiting high energy are not contained. Note, that for didactic reasons, we applied the angular window  $w_\nu$  on the driving function. It may as well be applied in the encoding procedure represented by (9). Refer to Fig. 6 described in Sec. 5.4.

After applying an angular window, it is indeed such that the reproduced sound field converges towards a focus point at the position of the intended virtual source and then diverges into the target half-space whose boundary contains the center of the secondary source distribution. In the context of wave field synthesis such a focus point between a converging and diverging wave field is termed *focused virtual sound source*. Note that the focused source achieved by angular weighting always radiates towards the center of the secondary source distribution.

The reproduction of focused virtual sound sources is actually a special case of *acoustic focusing* [16]. The latter is a technique which aims at the concentration of acoustical energy in a small spot. Unlike with focused sources no attention is paid to the fact in what regions the synthesized sound field converges and diverges.

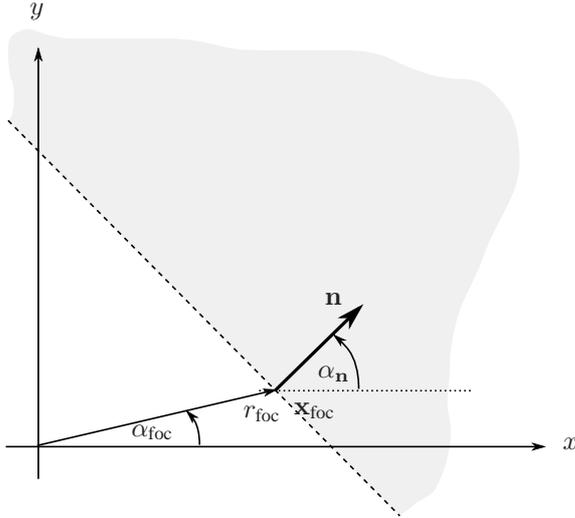
We emphasize the fact that focused virtual sources achieved by angular weighting can only be encoded/reproduced up to a given order which depends on the position of the source. This circumstance is independent from the properties of the secondary source distribution employed. Refer to Sec. 5.4 for a discussion of the physical justification to achieve focused virtual sound sources by applying angular weighting.

## 5.3. Focused virtual sources by explicit modeling

In this section we briefly review the concept of *explicitly modeling* focused virtual sound sources as presented by the authors in [4]. The basic idea is to model the desired sound field to be reproduced such that its closed-form description is valid over the entire receiver area (i.e. the region bounded by the secondary source distribution) in order to make the properties of the reproduced sound field predictable.

As a consequence of causality, it is impossible to reproduce a virtual sound source which is positioned inside the secondary source distribution. What can be achieved is the reproduction of a sound field which converges in one half-space towards a focus point and diverges in the other half-space (the *target half-space*). The diverging part of the reproduced sound field can be controlled such that it resembles the sound field of a sound source at the position of the focus point.

The concept of explicitly modeling focused virtual sound sources has been presented in [4] for two-dimensional reproduction of a focused point source and it has been restricted to the purely propagating part of the latter. Both



**Figure 5:** Definition of the quantities apparent in (14). The focus point is situated at position  $\mathbf{x}_{\text{foc}}$ . The normal vector  $\mathbf{n}$  points into the target half-space which is indicated by the grey-shaded area and bounded by the dashed line.

restrictions do not constitute principle limitations of the method but are rather a matter of convenience: Focused point sources are mathematically and conceptually the simplest case, and the fact whether evanescent components are perceptually significant and even audible is not clear. Finally, there are indications that according evanescent component of a point source can not be recreated over an extended area.

The modeling of such a propagating sound field with a focus point involves three major steps: a) Decompose a monopole (or other) source at the intended position of the focus point into a continuum of plane waves, b) identify those plane wave components propagating into the desired target half-space, c) construct the desired sound field from the plane wave representation of step b). As a consequence of causality, the modeled sound field converges towards the focus point in the half-space other than the target half-space.

Note that the time-reversal approach applied in wave field synthesis [1] actually implicitly models such a sound field converging and diverging in different half-spaces [2].

The explicitly modeled sound field of a focused source at position  $\mathbf{x}_{\text{foc}}$  with nominal orientation  $\alpha_{\mathbf{n}}$  is given by [4]

$$\begin{aligned}
 P(\mathbf{x}, \omega) &= \sum_{\nu=-\infty}^{\infty} J_{\nu} \left( \frac{\omega}{c} r \right) e^{j\nu\alpha} \times \\
 &\times \underbrace{\sum_{\eta=-\infty}^{\infty} w_{\eta} j^{-\eta} e^{-j\eta\alpha_{\mathbf{n}}} J_{\nu-\eta} \left( \frac{\omega}{c} r_{\text{foc}} \right) e^{-j(\nu-\eta)\alpha_{\text{foc}}}}_{= \check{S}(\omega)}.
 \end{aligned} \tag{14}$$

Refer to Fig. 5 for a graphical illustration of the setup.

$\check{w}_{\eta}$  represent the coefficients of a window function applied on the plane wave representation in step b). In [4] a rectangular window is employed. For convenience we use a

cosine-shaped window in this contribution since it causes a smoother amplitude distribution of the reproduced sound field over space by the cost of a marginally smaller target area. In this case

$$\check{w}_{\eta} = \begin{cases} \frac{\pi}{2} & \text{for } |\eta| = 1 \\ \frac{1}{2} \left( \frac{1}{1-\eta} + \frac{1}{1+\eta} \right) (j^{-\eta} + j^{\eta}) & \text{elsewhere} \end{cases} \tag{15}$$

Note that the window with coefficients  $\check{w}_{\eta}$  in (14) is applied in a different domain than  $w_{\nu}$  (13).

We want to emphasize that in (14), no restriction is posed neither on  $\mathbf{x}_{\text{foc}}$  nor on  $\alpha_{\mathbf{n}}$  nor on the maximum order  $N$  which can be employed.

#### 5.4. Comparison of the two approaches

Fig. 6 and Fig. 7 illustrate the two alternative approaches - angular weighting (Fig. 6) and explicit modeling (Fig. 7) - independent from the secondary source distributions. In other words, the two figures depict the sound fields which are encoded for a regular decoding process as given by (10). In both Fig. 6(a) and Fig. 7(a), the depicted sound fields converge in the half-space where  $y > 0.7$  m towards the focus point and diverge in the target half-space where  $y < 0.7$  m. For the explicitly modeled focused source, this behavior is not surprising since it has been modeled alike. This approach is physically perfectly justified.

For the angularly weighted focused source, the above described behavior is surprising at first sight. Mostly because the mathematical expression describing the sound field, i.e.

$$S(\mathbf{x}, \omega) = \sum_{\nu=-N}^N w_{\nu} \check{S}_{\nu}(\omega) J_{\nu} \left( \frac{\omega}{c} r \right) e^{j\nu\alpha} \tag{16}$$

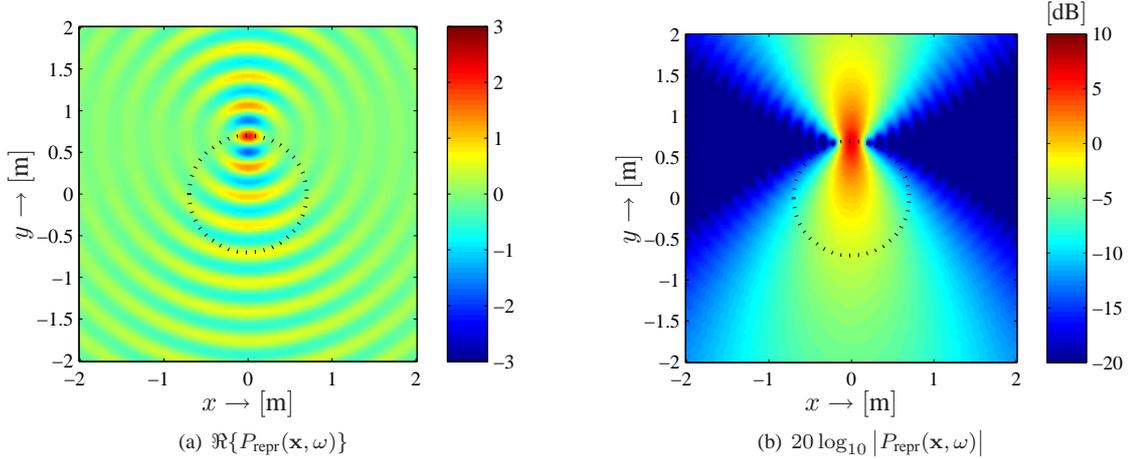
with  $w_{\nu}$  given by (13), is derived from an interior expansion (refer to (11) case 1) which is only valid for  $r < r_s$ . As a consequence, (16) is valid only there in a strict sense [6].

On the other hand, (16) describes by definition a sound field which is physically possible [6]. It is derived from a sound field which diverges at  $r < r_s$ . Again as a matter of causality, there seems to be no other way than (16) being a sound field which diverges also at all other locations in the half-space containing the coordinate origin bounded by the tangent on the strict validity region through the focus point.

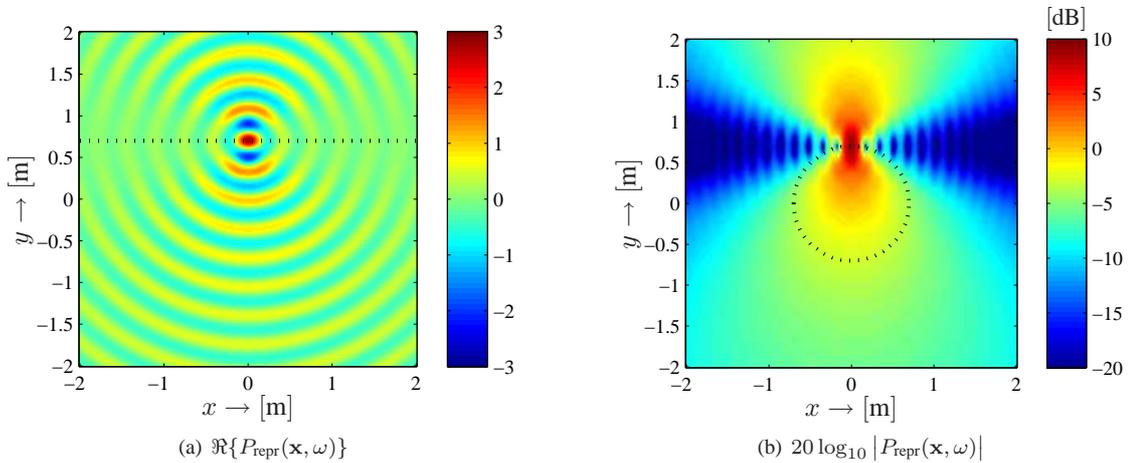
A thorough comparison of Fig. 6(b) and Fig. 7(a) suggests that the amplitude distribution is slightly better balanced over the entire target-half space for the explicitly modeled focused source. However, it is likely that the application of more sophisticated angular windows in the angular weighting approach is able to even out the short comings apparent in Fig. 6(b).

The amplitude decays in both approaches are depicted in Fig. 8 along the  $y$ -axis together with the amplitude decay of a monopole line source which represents the desired properties. Fig. 8 suggests that the amplitude decays at the positions shown are accurate enough to be perceptually convincing.

The most important difference between the explicit modeling and the angular weighting approach is the fact that



**Figure 6:** Focused source by angular weight at position with  $N = 16$ . The emitted frequency is  $f = 1000$  Hz. The values are clipped as indicated by the colorbar. The dotted line encloses the strict region of physical validity.



**Figure 7:** Focused source with  $N = 16$ . The emitted frequency is  $f = 1000$  Hz. The values are clipped as indicated by the colorbar. The dotted line represents the boundary of the target half space.

the nominal orientation of the focused source can be freely chosen in the explicit modeling approach whereas it always points towards the center of the secondary source distribution in the angular weighting approach. The cost for this freedom in the explicit modeling approach is a significantly higher computational complexity.

The fact that the spatial bandwidth (i.e. the maximum order) of the focused source is inherently limited in the angular weighting approach has not been found to be a restriction. The properties of the created sound field in the angular weighting approach can not be well controlled for virtual sources close to the center of the secondary source distribution since only very few orders can be employed. For virtual sources at the center, the approach fails. As a workaround, virtual sources can be reproduced at a location at some distance from the center which can be assumed to be perceptually acceptable.

Note that it can not be judged at this stage whether the angular weighting approach can be applied on any type of sound source since to ultimate physical justification has not been

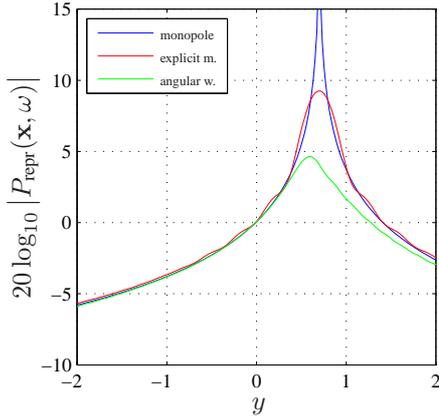
found. However, it is likely that no fundamental restrictions arise.

## 6 DATA-BASED REPRODUCTION

Data-based reproduction refers to the reproduction of recorded signals [17]. In the Ambisonics context, recordings are typically made using a spherical microphone array with which the spherical harmonics coefficients of the recorded wave field can be extracted [18]. For the two-dimensional scenario employed in this paper, the microphone array is circular and the coefficients  $P_\nu(r_{\text{ref}}, \omega)$  (refer to (9)) are extracted.  $r_{\text{ref}}$  represents the radius of the microphone distribution. The coefficients  $\hat{P}_\nu(r_{\text{ref}}, \omega)$  can be directly inserted in (9) and therefore, encoding is straightforward.

It has to be noted that practical constraints like the discrete property of the employed microphone distributions and measurement noise restrict the recording to a few lower spatial modes  $\nu$  [18, 19]. For convenience, we limit the in-

## 7 MODEL-BASED REPRODUCTION



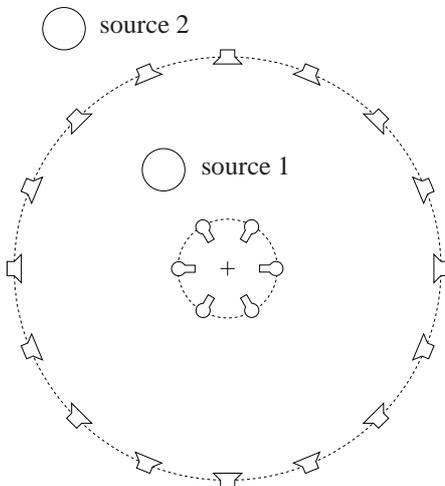
**Figure 8:** Cross sections through Fig. 6(b) (green) and Fig. 7(b) (red) along the  $y$ -axis. The blue line indicates the amplitude decay of a monopole line source whose position coincides with that of the focus points.

investigation to physical aspects. Practical challenges in the Ambisonics context are discussed e.g. in [20].

The coefficients  $\hat{P}_\nu(r_{\text{ref}}, \omega)$  which are extracted represent an expansion of the recorded wave fields around the center of the microphone array. When such a recording is reproduced, the center of the microphone array virtually coincides with the center of the secondary source distribution. Refer to Fig. 9 for an illustration.

If it happens that a sound source is recorded which is closer to the center of the microphone array than the secondary sources (like source 1 in Fig. 9), the same issues arise that are discussed in Sec. 5.1. There is no way to employ the explicit modeling of focused virtual sound sources in this situation and angular weighting as described in section Sec. 5.2 has to be applied. Note that recorded sound sources can be direct sound sources or indirect ones like reflecting surfaces, e.g. the floor underneath the microphone array.

Sources farther away from the microphone array than the secondary sources (like source 2 in Fig. 9) can be straightforwardly reproduced without modification.



**Figure 9:** Data-based reproduction.

Model-based reproduction refers to the reproduction of virtual scenarios which are composed of a number of sound objects which are described via analytical source models like plane and spherical waves [17]. Of course, model-based and data-based reproduction can be combined to e.g. the reproduction of a virtual scene in an acoustical environment whose properties are extracted from measurements.

Model-based reproduction is the situation where focused virtual sound sources unfold their full potential. If the description of the scenario to be reproduced is object-based such as in [21], the scenario can be reproduced directly without explicit encoding and decoding. It can be straightforwardly detected in real-time whether a virtual source has to be reproduced as a focused or non-focused source. When the explicit modeling approach is employed the target half-space of the focused sources can be chosen according to the instantaneous position of the receiver(s).

When a scenario is encoded prior to reproduction, care has to be taken. If the dimensions of the loudspeaker distribution on which the scenario will be decoded are known at the time of encoding, virtual sources closer to the center of the loudspeaker distribution than the loudspeakers can be encoded as focused sources employing any of the presented approaches.

When the dimensions of the loudspeaker system employed are not known at the time of encoding, one can go the safe way and decode any present sound source as focused source with nominal orientation towards the center. Then the reproduction can be accomplished no matter what loudspeaker system is employed.

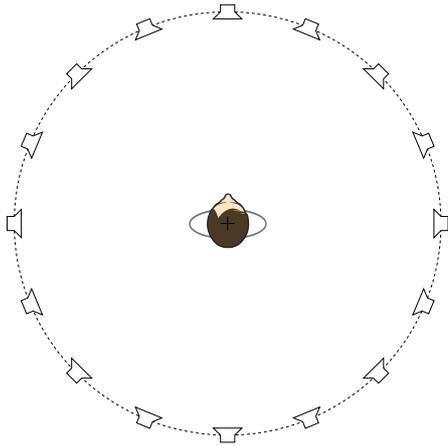
Note that model-based reproduction provides the possibility to encode different virtual sources at different orders.

## 8 APPLICATION EXAMPLE: VIRTUAL SOUND FIELD REPRODUCTION

Apart from the straight-forward employment of focused virtual sound sources in loudspeaker-based reproduction, we want to present another method of high potential which typically employs model-based rendering: Virtual sound field reproduction via headphones.

This type of virtual sound field reproduction employs the spatio-temporal transfer functions from specific positions to the ear drums of either a human or a mannequin. These transfer functions are referred to as head-related transfer functions (HRTFs). In conventional HRTF-based reproduction the input signal of a virtual sound source to be reproduced is filtered with that pair of HRTFs representing the intended position of the source. This requires a vast amount of data to be measured.

The key idea is to interpret the measurement positions as virtual loudspeakers. If the measurement point are distributed on a circle or a sphere centered around the listener, Ambisonics or related techniques (like the presented one) can be employed [22]. The most important property is the fact, that the listener is always in the virtual sweet spot [8].



**Figure 10:** The geometric setup in virtual sound field reproduction. The loudspeaker symbols represent measurement points of the HRTFs.

With a modest amount of measurement points, a sweet spot of the size of a human head can be achieved even for the highest considered temporal frequency.

Measurements should be accomplished under anechoic conditions in order to avoid implausible reverberation caused by the virtual loudspeakers. Room information has to be reproduced additionally in an appropriate manner. In this anechoic situation head rotations of the listener can be realized as rotations of the sound field which can be comfortably implemented [9].

Since the curvature of the wave front has been shown to be not significant for the distance perception of sources farther than a few meters [23], it might be sufficient to place the virtual sources on the virtual loudspeaker contour for far intended virtual source locations. This is computationally very efficient. The perceived source distance has to be controlled via the level and reverberation anyway [24].

For close sources, diffraction and scattering on the body have an essential perceptual impact. For close virtual sources, e.g. around the head of the listener, the limited validity of the interior expansion represented by (11) suggests that scattering and diffraction of the torso are not appropriately modeled since the latter is positioned in the region where the source field is not correctly modeled.

A lot of effort was put in [25, 26] in order to translate expansion centers such that valid regions contain all parts of the listener's body which have significant influence in terms of scattering and diffraction. It can be argued whether such a strict interpretation of the regions of validity is necessary. The simulations presented in Sec. 5.2 indicate that a looser interpretation might be acceptable. It might therefore be sufficient to reproduce close virtual source positions employing one of the presented focusing techniques. Although we have not found an analytic proof it can be assumed that appropriate angular weighting does indeed produce a sound field which exclusively diverges in a half-space. In this half-space, diffraction and scattering on the virtual listener might indeed be acceptably modeled. Whether or not the employment of explicitly modeled focused sources provides benefit

in this respect is not clear.

A still unresolved problem is the presence of the listener in the reproduced wave field, be it virtual or real. Sound field reproduction methods like the presented approach typically assume free-field conditions. Especially in situations of virtual sound sources in prominent positions like under the listener's chin, it has to be assumed that the listener's presence has a significant effect. Initial work to clarify this aspect includes [27] but general results are not yet available.

## 9 CONCLUSIONS

We have presented an investigation of the reproduction of focused virtual sound sources in the Ambisonics context. Two approaches can be employed in order to model a focused source: 1) manipulation (angular weighting) of the interior spatial expansion of a sound source in order to extend the region of physical validity, or 2) the explicit modeling of a sound field diverging from a focus point into the target half-space.

The physical properties of both approaches differ in some aspects of minor importance although the angular weighting approach lacks the ultimate physical justification. The latter approach exhibits significant advantages with respect to the practical handling. The main advantages are: i) angular weighting is expected to be applicable on any arbitrary sound fields, and ii) it is computationally significantly more efficient than the explicit modeling of focused sources.

The explicit modeling approach provides the freedom to arbitrarily choose the nominal radiation direction, i.e. it enables to freely rotate the target half-space around the position of the focused source. This constitutes a major benefit for environments with user tracking [28]. Finally, the explicit modeling approach is a powerful tool for investigation of the fundamental properties and limitations of the reproduction of focused sources.

## REFERENCES

- [1] E.N.G. Verheijen. Sound reproduction by wave field synthesis. PhD thesis, Delft University of Technology, 1997.
- [2] J. Ahrens and S. Spors. Notes on rendering focused directional virtual sound sources in wave field synthesis. In *34. Jahrestagung der Deutschen Gesellschaft für Akustik*, Dresden, Germany, March 10–13 2008.
- [3] J. Daniel and S. Moreau. Further study of sound field coding with higher order ambisonics. In *116th Convention*, Berlin, Germany, 2004. Audio Engineering Society (AES).
- [4] J. Ahrens and S. Spors. Focusing of virtual sound sources in higher order Ambisonics. In *124th Convention of the AES*, Amsterdam, The Netherlands, May 17–20 2008.
- [5] F. Adriaensen. HOA focussing examples. <http://www.kokkinizita.net/ambisonics/focussing/index.html>. Retrieved May 09.

- [6] E. G. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. Academic Press, London, 1999.
- [7] J. Ahrens and S. Spors. An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions. *Acta Acustica utd. with Acustica*, 94(6):988–999, Nov./Dec. 2008.
- [8] S. Spors and J. Ahrens. A comparison of wave field synthesis and higher-order Ambisonics with respect to physical properties and spatial sampling. In *125th Conv. of the AES*, San Francisco, CA, Oct. 2–5 2008.
- [9] J. Daniel. Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. PhD thesis Université Paris 6, 2001.
- [10] L. G. Copley. Fundamental results concerning integral representations in acoustic radiation. *JASA*, 44:28Ü–32, 1968.
- [11] B. Girod, R. Rabenstein, and A. Stenger. *Signals and Systems*. J.Wiley & Sons, 2001.
- [12] J. Ahrens and S. Spors. Sound field reproduction employing non-omnidirectional loudspeakers. In *126th Conv. of the AES*, Munich, Germany, May. 7–10 2009.
- [13] E.W. Weisstein. L’Hospital’s Rule. Math-World – A Wolfram Web Resource. <http://mathworld.wolfram.com/LHospitalsRule.html>, Retrieved Jan. 08.
- [14] J. Daniel. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new Ambisonic format. In *23rd International Conference of the AES*, Copenhagen, Denmark, May 23–25 2003.
- [15] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications Inc., New York, 1968.
- [16] S. Yon, M. Tanter, and M. Fink. Sound focusing in rooms: The time-reversal approach. *JASA*, 113(3):1533Ü–1543, March 2003.
- [17] R. Rabenstein and S. Spors. Multichannel sound field reproduction. In Benesty, J., Sondhi, M., Huang, Y, (Eds.), *Springer Handbook on Speech Processing and Speech Communication*, Springer Verlag, 2007.
- [18] B. Rafaely. Analysis and design of spherical microphone arrays. *IEEE Trans. On Speech and Audio Process.*, 13(1):135–143, Jan. 2005.
- [19] B. Rafaely, B. Weiss, and E. Bachmat. Spatial aliasing in spherical microphone arrays. *IEEE Trans. on Signal Proc.*, 55(3):1003–1010, March 2007.
- [20] S. Moreau, J. Daniel, and S. Bertet. 3D sound field recording with higher order ambisonics – Objective measurements and validation of a 4th order spherical microphone. In *120th Convention of the AES*, Paris, France, May 20–23 2006.
- [21] M. Geier, J. Ahrens, and S. Spors. ASDF: Ein XML format zur beschreibung von virtuellen 3D-audioszenen. In *34. Jahrestagung der Deutschen Gesellschaft für Akustik*, Dresden, Germany, March 10–13 2008.
- [22] A. McKeage and D. McGrath. Sound field format to binaural decoder with head tracking. In *AES Convention 6r*, Melbourne, Australia, 1996.
- [23] J. Blauert. *Spatial Hearing*. Springer, New York, 1997.
- [24] A. W. Bronkhorst and T. Houtgast. Auditory distance perception in rooms. *Nature*, 397(6719):517–520, February 1999.
- [25] D. Menzies. Nearfield binaural synthesis and ambisonics. *JASA*, 121(3):1559–1563, March 2007.
- [26] Dylan Menzies. Nearfield binaural synthesis report. In *Acoustics 08*, Paris, France, 2008.
- [27] T. Betlehem and M. A. Poletti. Sound field reproduction around a scatterer in reverberation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, April 19–24 2009.
- [28] F. Melchior, C. Sladeczek, D. de Vries, and B. Fröhlich. User-dependent optimization of wave field synthesis reproduction for directive sound fields. In *124th Conv. of the AES*, Amsterdam, The Netherlands, May 17–20 2008.