Interpolation and Range Extrapolation of Head-Related Transfer Functions using Virtual Local Wave Field Synthesis

Sascha Spors, Hagen Wierstorf, and Jens Ahrens

Deutsche Telekom Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

Correspondence should be addressed to Sascha Spors (`Sascha.Spors@telekom.de`)

ABSTRACT

Virtual environments which are based on binaural sound reproduction require datasets of head-related transfer functions (HRTFs). Ideally, these HRTFs are available for every possible position of a virtual sound source. However, in order to reduce measurement efforts, such datasets are typically only available for various source directions but only for one or very few distances. This paper presents a method for extrapolation of measured HRTF datasets from the source distance used in the measurements to other source distances. The method applies the concept of local Wave Field Synthesis to compute extrapolated HRTFs for almost arbitrary source positions with high accuracy. The method is computationally efficient and numerically stable.

1. INTRODUCTION

1.1. Head-Related Transfer Functions

The human auditory system exploits the acoustic characteristics of the outer ear in order to perform spatial scene analysis. The acoustic properties of the outer ear can be captured by measuring the transfer function from an acoustic source to a defined position in the left and right ear canals [1]. These functions are known as head-related transfer functions (HRTFs) or head-related impulse re-

sponses (HRIRs). HRTFs are typically measured in anechoic environments. In this scenario, they depend only on the posture of the listener (e.g. head-orientation) and the position and properties of the acoustic source.

In order to limit the measurement effort, HRTFs are typically measured for a fixed head position and source positions at a fixed distance under varying angles with respect to the head. For ease of illustration, we consider a two dimensional scenario in
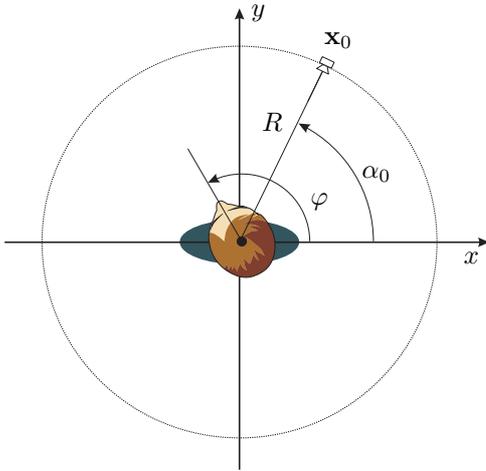
Fig. 1: Geometry used for measurement HRTF datasets.

the sequel were the sources are located in the horizontal plane. Figure 1 shows the considered geometry. The impulse responses from a source at position $\mathbf{x}_0 = R\,[\cos(\alpha_0)\ \sin(\alpha_0)]^T$ to the left and right ear are measured under free-field conditions. The left/right HRIR depends in general on the head orientation φ, the source distance R and source angle α_0. The HRIR is denoted by $h_{\{L,R\}}(\alpha_0, R, \varphi, t)$, the HRTF by $H_{\{L,R\}}(\alpha_0, R, \varphi, \omega)$ were $\omega = 2\pi f$ denotes the radial frequency. The head orientation will be always $\varphi = 90°$ in this study.

It is obvious that the measurement of HRIR datasets becomes a complex task for a densely sampled space of source positions. Therefore, most of the currently available datasets consider only one measurement distance with a limited, sometimes not constant, angular resolution. Typical distances are in the range of 1.5 to 3 meters with an angular resolution of 5 to 10 degrees [2, 3, 4]. It is generally assumed that a source distance exceeding $R = 3\,\text{m}$ does not change the characteristics of the HRIRs essentially [1, 5]. However, it is known that HRIRs change significantly for nearby sources [5, 6, 7, 8, 9]. Such HRIRs are typically termed as near-field or proximal HRIRs. Note, the term near-field is not used in a strict physical sense in this context.

A number of techniques have been proposed in the past to compute HRIRs for arbitrary source positions from HRIR datasets measured at a fixed dis-

tance. Two different classes of techniques are used: (i) interpolation and (ii) extrapolation of HRIRs. While the first class considers computing HRIRs at angles which have not been captured, the second class considers the problem of computing HRIRs at different distances as captured. This paper considers mainly the extrapolation of HRIRs. The most simple method is to modify the amplitude according to the distance law. However, this neglects most of the distance cues present in HRTFs [5]. Besides this simple approach, also work has been conducted to modify the spectrum of measured HRTFs in order to cope for the changes that occur for nearby sources [10, 11]. However, this still neglects the structural changes in HRTFs for different source distances. In order to cope for these a number of approaches have been published that are based upon extrapolation of HRTFs using principles from wave physics. Since our approach is within this class, some relevant approaches in this context are briefly reviewed in the following.

1.2. Range-extrapolation of HRTFs

In [12, 13] two techniques are presented which are based upon the expansion of HRIRs into surface spherical harmonics. These form a orthogonal basis that can be used for angular interpolation and range extrapolation. However, the expansion into spherical harmonics is only valid for a region without sources/scatterers. Hence it can not be applied straightforwardly to measured HRIRs due to the scattering of the head/torso. A solution to this problem has been found by considering the reciprocity theorem of the acoustic wave equation. It is assumed that the transfer paths from the ears to the source positions have been measured. The expansion is then performed with respect to the source positions. However, extrapolating the HRIRs is only possible if no scattering objects are located within the extrapolation region. The application of these techniques to the computation of near-field HRIRs is therefore limited to distances not including reflections from the upper torso. Methods based on spherical harmonics expansions suffer also from numerical issues. Alternatively to these approaches, the distribution of measurement points of a given HRTF dataset can be interpreted as a virtual loudspeaker array. When this array is considered as a unit, were each virtual loudspeaker is driven by an individual signal, the

HRTFs from a desired virtual source to both ears can be synthesized. In the past, various techniques have been published which are based upon the theory of higher-order Ambisonics (HOA) for the computation of the driving signals [14, 15]. The techniques use also an expansion of the HRTFs with respect to spherical harmonics. A general problem is that such expansions are numerical complex and may become numerically instable for sparsely or unequally sampled datasets. Especially the latter two problems may become critical for available datasets. The synthesis of virtual sources at closer distances than the virtual loudspeaker array (so-called focused sources) is also subject to numerical instabilities [16].

1.3. Overview of this paper

The paper is organized as follows: We first briefly review to concept of wave field synthesis (WFS) and local WFS. We then introduce the proposed method for range extrapolation of HRTFs which is based upon the application of local WFS. This is followed by results analyzing the performance of the technique.

2. WAVE FIELD SYNTHESIS

2.1. Basic Principle

The concept of WFS has initially been developed for linear distributions of secondary sources [17] and has later on been extended to arbitrarily shaped convex distributions [18, 19]. WFS is based on the Kirchhoff-Helmholtz integral [20], which states that sound field synthesis (SFS) can be realized by a distribution of secondary monopole and dipole sources located on the boundary ∂V of the listening area V which are driven by the directional gradient and the pressure of the sound field of the virtual source $S(\mathbf{x}, \omega)$, respectively. Figure 2 illustrates the geometry.

WFS is based on a number of reasonable approximations in order to realize this principle by a spatially discrete distribution of secondary monopole sources. The concept of WFS is discussed briefly in the following, the details are outlined in the given references.

The synthesized sound field $P(\mathbf{x}, \omega)$ is given by [19]

$$ P(\mathbf{x}, \omega) = \int_{\partial V} D(\mathbf{x}_0, \omega) \, \frac{1}{4\pi} \frac{e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x} - \mathbf{x}_0|} \, dS_0 \,, \quad (1) $$

where $\mathbf{x}_0$ denotes a position on the boundary ∂V, $D(\mathbf{x}_0, \omega)$ the driving signal for a secondary source at $\mathbf{x}_0$ and dS_0 a surface element for integration. In SFS, the synthesized sound field $P(\mathbf{x}, \omega)$ should match a desired field $S(\mathbf{x}, \omega)$ as close as possible. In WFS, the integral equation (1) is not solved explicitly in order to derive the driving signal $D(\mathbf{x_0}, \omega)$. Instead an interpretation of the underlying physical principle, given by the Kirchhoff-Helmholtz integral, together with reasonable high-frequency/far-field approximations allows for direct calculation of $D(\mathbf{x_0}, \omega)$ via the directional gradient of $S(\mathbf{x}, \omega)$. A special case in this context is so-called *2.5-dimensional synthesis*. Using point sources for the synthesis of a sound field in a plane constitutes a so called 2.5-dimensional problem. The resulting properties are somewhere between two and three dimensional synthesis [19]. The driving signal for a secondary source at position $\mathbf{x}_0$ for 2.5- and three-dimensional WFS is given by [19]

$$ D(\mathbf{x}_0, \omega) = 2 \, a(\mathbf{x}_0) \, c_{2.5\mathrm{D}}(\mathbf{x}_0, \omega) \, \frac{\partial}{\partial \mathbf{n}_0} S(\mathbf{x}_0, \omega) \,, \quad (2) $$

where $a(\mathbf{x}_0)$ denotes a window function, $c_{2.5\mathrm{D}}(\mathbf{x}_0, \omega)$ a filter accounting for 2.5-dimensional synthesis and $\frac{\partial}{\partial \mathbf{n}_0}$ the directional gradient in direction $\mathbf{n}_0$, which is normal to the secondary source contour, evaluated at position $\mathbf{x}_0$. For the three-dimensional synthesis $c_{2.5\mathrm{D}}(\mathbf{x}_0, \omega) = 1$. The window function $a(\mathbf{x}_0)$ takes care that only those secondary sources are active where the local propagation direction of the virtual source $S(\mathbf{x}_0, \omega)$ at the position $\mathbf{x}_0$ has a positive component in direction of the normal vector $\mathbf{n}_0$ of the secondary source [21]. For 2.5-dimensional synthesis, the correction $c_{2.5\mathrm{D}}(\mathbf{x}_0, \omega)$ depends in general on the desired virtual sound field and the geometry of the secondary source distribution [22].

2.2. Virtual Source Models

Instead of prescribing or recording a complex sound field as desired field $S(\mathbf{x}_0, \omega)$, analytic source models are typically used in WFS as basic building blocks for potentially complex acoustic scenes. This is referred to as *model-based rendering*. Frequently used models are virtual point sources, plane waves and focused sources. Model-based rendering has a number of benefits. Most notably in the context of this paper is the low computational complexity and numerical stability when rendering these simple virtual
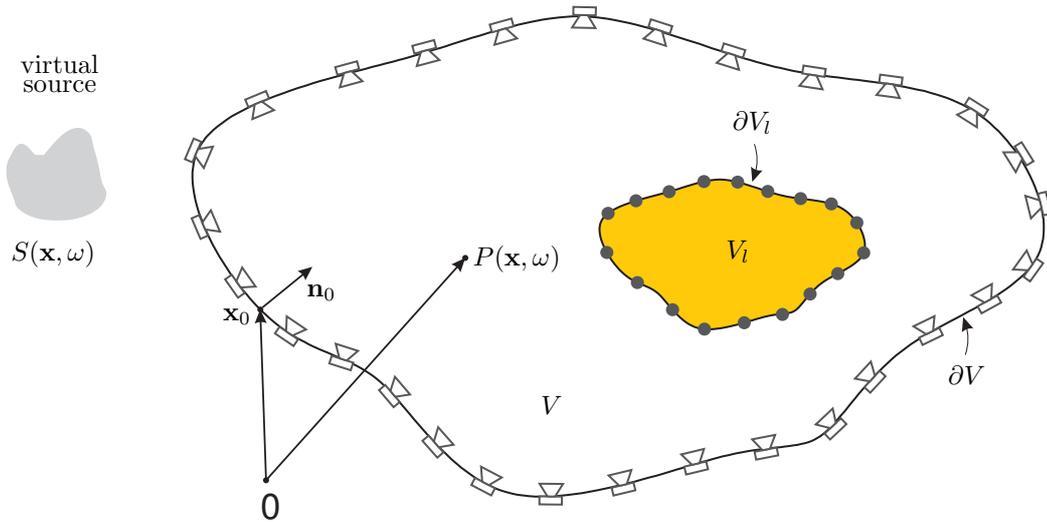
Fig. 2: Geometry and principle of (local) WFS. The loudspeakers are indicated by the loudspeaker symbols placed on the contour ∂V surround the listening area V, while the virtual loudspeakers used by local WFS are indicated by the bullets • placed on the contour ∂V_l surrounding the local listening area V_l.

sources.

For three-dimensional WFS, the driving functions for a virtual plane wave and point source are derived straightforwardly by using (2) with $c_{2.5D}(\mathbf{x}_0, \omega) = 1$. For 2.5-dimensional WFS the synthesis has to be referenced to a fixed listener position/distance. The amplitude decay of the synthesized sound field is incorrect outside of this reference point/line. The derivation of the three- and 2.5-dimensional driving functions for WFS can be found e. g. in [19].

An interesting feature of SFS techniques is the synthesis of *focused sources*. Focused sources aim at creating the impression of a virtual source placed in front of the loudspeakers. This is achieved by synthesizing a sound field which converges towards a focus point and diverges after. The field after the focus point resembles the field of an acoustic point source placed at the focus point. Modified acoustic focusing techniques are used to derive appropriate driving functions. Since the secondary sources emit a sound field that travels towards the listener, one can only expect that the desired sound field of a focused source is correct if the focus point is located in between the active secondary sources and the listener. In the context of WFS, this is a well known limitation of focused sources [23]. The listening area of a focused source in which the wave fronts travel in the correct direction is bounded, in the optimal case, as a half space which is limited by a line through the focus point. The selection of secondary sources has direct influence on the orientation of this line and consequently the nominal direction $\mathbf{n}_{fs}$ of the focused source, as has been shown in [21]. The properties of focused sources have been investigated in detail in [24].

2.3. Spatial Discretization of Secondary Source Distribution

A limited number of discrete loudspeakers is used in practice to realize the secondary source distribution. This constitutes a spatial sampling process which can be modeled by sampling the driving function. The resulting spatial sampling artifacts are well investigated for non-focused [23, 25, 26] and focused sources [24]. For non-focused sources, spatial sampling artifacts become prominent in the synthesized sound field above a geometry and virtual source type specific frequency (also known as spatial aliasing frequency). For WFS, the artifacts are typically spread over the entire listening area. For typical setups, sampling artifacts may be audible as coloration of the virtual source [27].

The properties of spatial sampling artifacts for focused sources depends on the distance to the focus point. In comparison to non-focused sources, fo-
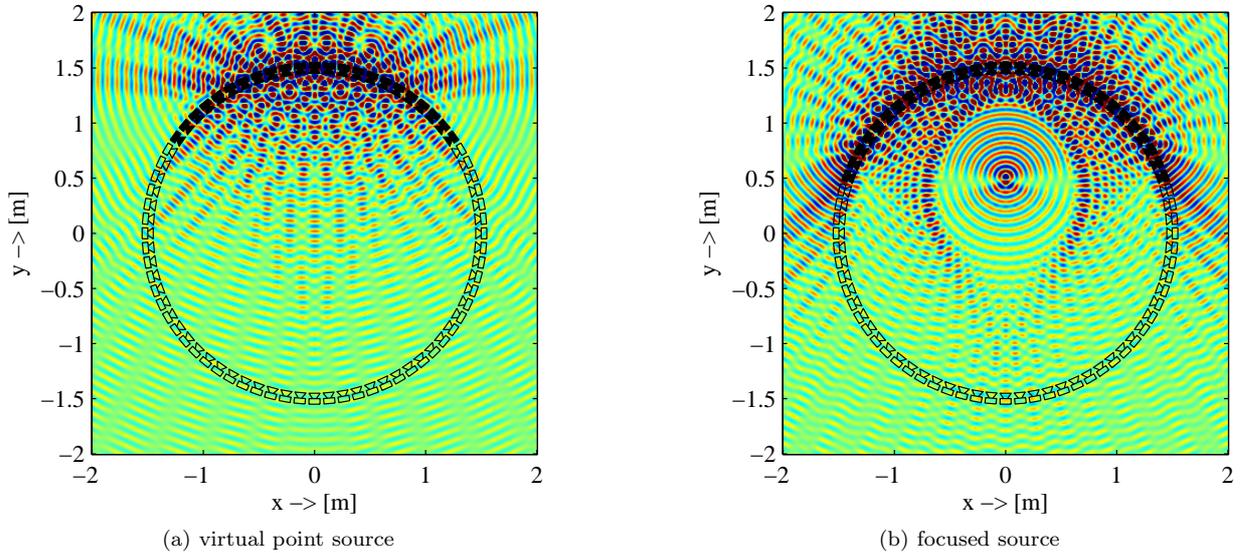
(a) virtual point source

(b) focused source

Fig. 3: Synthesis of a monochromatic virtual point and focused source with WFS using a circular loudspeaker array ($N = 72$, $R = 1.5\,\mathrm{m}$, $f = 4\,\mathrm{kHz}$, $\mathbf{x}_{\mathrm{ps}} = [0\ 3]^T\,\mathrm{m}$, $\mathbf{x}_{\mathrm{fs}} = [0\ 0.5]^T\,\mathrm{m}$, $\mathbf{n}_{\mathrm{fs}} = [0\ -1]^T\,\mathrm{m}$). The active loudspeakers are filled.

cused sources allow the accurate synthesis for much higher frequencies in the vicinity of the focus point. This is illustrated by an example. Figure 3 shows the synthesized sound field for a virtual point and focused source using WFS. For the simulated situation and non-focused sources, sampling artifacts would be present for frequencies above $f_{\mathrm{al}} \approx 1.5\,\mathrm{kHz}$ [25]. The spatial sampling artifacts for the synthesis of a monochromatic virtual point source with $f = 4\,\mathrm{kHz}$ are clearly observable in Fig. 3a. In contrast, it is obvious from Fig. 3b that almost no spatial sampling artifacts are present in a circular region with a radius of about $0.75\,\mathrm{m}$ around the focus point for the synthesis of a monochromatic focused source.

3. LOCAL WAVE FIELD SYNTHESIS
The achievable accuracy of WFS is in general linked to the spatial density of the loudspeakers (number of loudspeakers per length/area). The higher the spatial density is, hence the more loudspeakers are used for a certain region, the less prominent the resulting spatial sampling artifacts will be. The same holds also for other synthesis techniques. However, the number of loudspeakers is restricted, amongst others, by their size and other practical aspects. Lo-

cal WFS aims at an enhanced accuracy in one or more local listening zone(s) which are smaller than the traditional listening zone of WFS [28].

The basic concept of local WFS is to utilize focused sources as virtual secondary sources. Creating a set of focused sources inside the listening area of a given loudspeaker arrangement with a spatially denser distribution than the secondary sources results in a higher accuracy within the area bounded by the positions of the focused sources (local listening area). These virtual secondary sources are then driven like real secondary sources placed at the respective positions. Figure 2 illustrates the principle. The secondary sources are indicated by the loudspeaker symbols on the contour ∂V, while the virtual secondary sources created by the focused sources are indicated by the bullets $\bullet$ on the contour ∂V_l. Focused sources have interesting properties which make them a good candidate as virtual secondary sources [24]: (i) focused sources exhibit an amplitude decay over distance to the focus point which is quite close to that of a point source placed at the focus point, and (ii) in comparison to non-focused virtual sources, focused sources allow the accurate synthesis for much higher frequencies in the vicinity
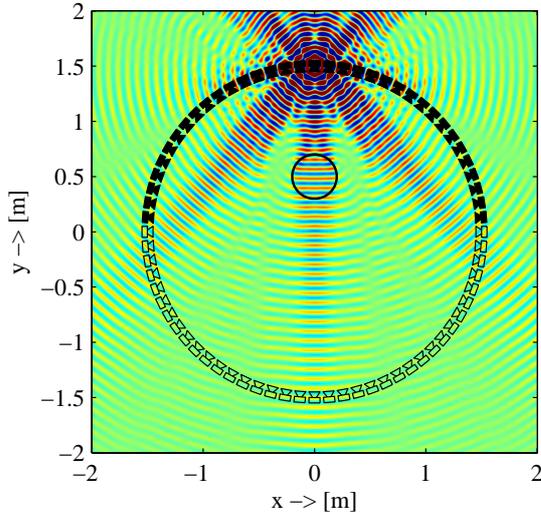
Fig. 4: Local synthesis of a monochromatic virtual point source with local WFS using a circular loudspeaker array ($N = 72$, $R = 1.5\,\text{m}$, $f = 4\,\text{kHz}$, $\mathbf{x}_{\text{ps}} = [0\ 3]^T\,\text{m}$, $M = 72$, $r = 0.2\,\text{m}$). The active loudspeakers are filled.

Fig. 5: Geometry used for extrapolation of HRTF datasets using virtual (local) WFS.

of the focus point (see Fig. 3b).

The local synthesis of a virtual point source is shown exemplarily in Fig. 4. The local listening area was chosen as a circular region with a radius of $r = 0.2\,\text{m}$ whose center is shifted by $\Delta y = 0.5\,\text{m}$ from the center. When comparing Fig. 4 to Fig. 3a it can be observed clearly that no sampling artifacts are present within the local listening area. However, this comes at the cost of stronger deviations outside of the local listening area.

4. RANGE EXTRAPOLATION OF HRTFS BY LOCAL WAVE FIELD SYNTHESIS

4.1. Basic Concept

As outlined in Section 1.1, HRTFs are typically measured using an omnidirectional sound source. The basic concept of the proposed technique is to interpret a set of HRTF measurements as emerging from virtual loudspeakers which are driven by local WFS to compute extrapolated HRTFs whereby the local listening area is positioned around the listeners head. This concept is similar to those used in the approaches discussed in Section 1.2. However, they do not apply local SFS to improve the accuracy of the computed HRTFs. As HRTF databases are typically
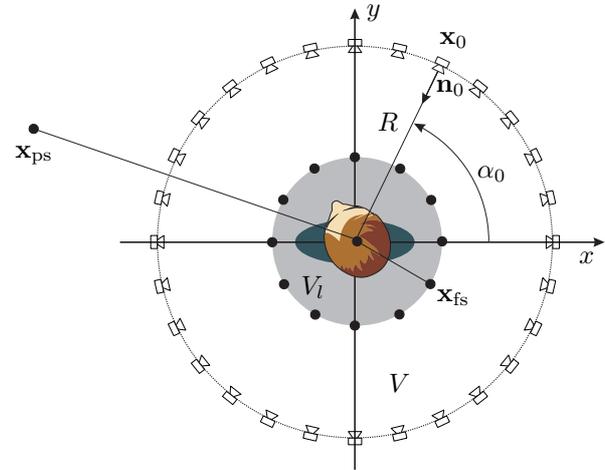
captured from a circular or spherical distribution of source positions, a specialization of Eq. (1) to these geometries is suitable for most cases. Note, WFS allows also to use different secondary source distributions like e.g. a rectangular one. For the ease of illustration we assume in the following the case of HRTFs measured exclusively for source positions in the horizontal plane. This is covered by the theory of 2.5-dimensional WFS. However, the extension to three dimensions is straightforward using the theory of three-dimensional WFS [19]. Figure 5 shows the underlying geometry.

According to (1), the signal $P_{\{\text{L,R}\}}(\varphi, \omega)$ at the left/right ear is given as

$$P_{\{\text{L,R}\}}(\varphi, \omega) =$$
$$\int_0^{2\pi} D(\alpha_0, \omega)\, H_{\{\text{L,R}\}}(\alpha_0, R, \varphi, \omega) R\, d\alpha_0 \ . \quad (3)$$

Hence, the terms describing a monopole point source in (1) have been replaced by the respective HRTFs. Since these include the scattering of the outer ear one may argue that the synthesized ear signals are different from the ones in a natural sound field. In a companion paper [29], the influence of scattering on the sound field synthesized by sound field synthesis has been investigated. It has been shown that the scattering within a synthesized sound field is equal

to the scattering observed in a natural sound field. If the driving signal $D(\alpha_0, \omega)$ of the virtual secondary sources is calculated for a point source with temporal Dirac pulse as source signal the result of (3) is a left and right extrapolated HRTF $\tilde{H}_{\{L,R\}}(\alpha, r, \varphi, \omega)$, respectively. A dataset of extrapolated HRTFs can be computed by repeated calculation for virtual source positions on a circle or sphere. Note, this inherently includes interpolation of HRTFs since the position of the virtual point source is not limited to positions given be the angular sampling interval of the measured HRTFs.

This basic principle has already been proposed by the authors for the calculation of far-field HRTFs [30]. However even for spatially densely sampled HRTFs, spatial aliasing will be present in the extrapolated HRTFs. For instance using a database of horizontal plane HRTFs captured with a source distance $R = 2\,\mathrm{m}$ and an angular resolution of one degree, spatial sampling artifacts will be present for frequencies above $8\,\mathrm{kHz}$ [25]. Although spatial sampling artifacts of WFS result in less severe perceptual degradations as one might assume when inspecting Fig. 3a, it is desirable to reduce them as far as possible.

In [31], the proposed technique has been used for the synthesis of near-field HRTFs. Here, focused sources have been used to compute extrapolated HRTFs for distances closer that the measured HRTFs. Due to the properties of focused sources with respect to spatial sampling (see Fig. 3b) artifacts are not present or negligible for typical HRTF datasets. For the example considered above and the synthesis of a focused source at a distance of $0.5\,\mathrm{m}$ from the center spatial aliasing artifacts become prominent above $20\,\mathrm{kHz}$. Hence, these focused sources can be used as virtual secondary sources for local WFS.

The basic concept of the proposed technique is to use local WFS for the range extrapolation of HRTFs. The next sections will illustrate the application for horizontal plane HRTFs. However, the extension to HRTFs captured on a sphere is straightforward using three dimensional local WFS.

4.2. Range Extrapolation and Interpolation of HRIRs

The driving function for the local synthesis of a virtual point source is derived in the following. This driving function is then used to extrapolate HRTFs.

The driving function for a focused source can be derived by reversing the time in the driving function of a virtual point source together with a sensible selection of the active secondary sources and a pre-delay to ensure causality in practice. The driving function for a focused source with respect to the scenario shown in Fig. 5 is given by [23, 21]

$$D_{\mathrm{fs}}(\alpha_0, \omega) = \sqrt{\frac{2}{\pi}} a_{\mathrm{fs}}(\alpha_0) \sqrt{-j\frac{\omega}{c}} \times$$
$$c_{2.5\mathrm{D}}(\mathbf{x}_0) \frac{(\mathbf{x}_0 - \mathbf{x}_{\mathrm{fs}})^T \mathbf{n}_0}{|\mathbf{x}_0 - \mathbf{x}_{\mathrm{fs}}|} \frac{e^{j\frac{\omega}{c}|\mathbf{x}_0 - \mathbf{x}_{\mathrm{fs}}|}}{\sqrt{|\mathbf{x}_0 - \mathbf{x}_{\mathrm{fs}}|}} \,, \quad (4)$$

where $\mathbf{x}_{\mathrm{fs}} = r_{\mathrm{fs}}[\cos(\alpha_{\mathrm{fs}}) \quad \sin(\alpha_{\mathrm{fs}})]^T$ denotes the position of the focused source, $\mathbf{x}_0 = R[\cos(\alpha_0) \quad \sin(\alpha_0)]^T$ the position of the secondary source as represented by the measured HRTF, $c_{2.5\mathrm{D}}(\mathbf{x}_0)$ a geometry dependent amplitude correction required for 2.5-dimensional synthesis and $\mathbf{n}_0 = -[\cos(\alpha_0) \quad \sin(\alpha_0)]^T$. The amplitude correction is approximately constant for a linear secondary source distribution and a reference line parallel to the secondary source distribution [23]. The pre-delay required for causality of the driving signal has been discarded for simplicity in (4). The secondary source selection criterion for a focused source is given by

$$a_{\mathrm{fs}}(\alpha_0) = \begin{cases} 1 & , \text{ if } \langle \mathbf{x}_{\mathrm{fs}} - \mathbf{x}_0, \mathbf{n}_{\mathrm{fs}} \rangle > 0, \\ 0 & , \text{ otherwise.} \end{cases} \quad (5)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\mathbf{n}_{\mathrm{fs}} = -[\cos(\alpha_{\mathrm{fs}}) \quad \sin(\alpha_{\mathrm{fs}})]^T$ the nominal direction of the focused source for the considered geometry.

For local WFS these focused sources will act as virtual secondary sources which are driven by the driving function of the desired virtual point source. For extrapolation of HRTFs virtual point sources are used. The traditional driving function for a virtual point source placed at position $\mathbf{x}_{\mathrm{ps}} \notin V_l$ is given by [23, 19]

$$D_{\mathrm{ps}}(\alpha_0, \omega) = \sqrt{\frac{2}{\pi}} a_{\mathrm{ps}}(\alpha_0) \sqrt{j\frac{\omega}{c}} \times$$
$$c_{2.5\mathrm{D}}(\mathbf{x}_0) \frac{(\mathbf{x}_0 - \mathbf{x}_{\mathrm{ps}})^T \mathbf{n}_0}{|\mathbf{x}_0 - \mathbf{x}_{\mathrm{ps}}|} \frac{e^{-j\frac{\omega}{c}|\mathbf{x}_0 - \mathbf{x}_{\mathrm{ps}}|}}{\sqrt{|\mathbf{x}_0 - \mathbf{x}_{\mathrm{ps}}|}} \,, \quad (6)$$

where $\mathbf{x}_{\mathrm{ps}} = r_{\mathrm{ps}}[\cos(\alpha_{\mathrm{ps}}) \quad \sin(\alpha_{\mathrm{ps}})]^T$ denotes the position of the virtual source. The window function

$a_{ps}(\mathbf{x}_0)$ can be expressed as

$$a_{ps}(\mathbf{x}_0) = \begin{cases} 1 & \text{, if } \langle \mathbf{x}_0 - \mathbf{x}_{ps}, \mathbf{n}_0 \rangle > 0, \\ 0 & \text{, otherwise.} \end{cases} \quad (7)$$

The driving function for the local synthesis of a virtual point source follows by combining (6) with (4) for spatially discrete distribution of virtual secondary sources. Here, a total number of M virtual secondary sources and N measured HRTFs is assumed. For ease of illustration it is further assumed that the amplitude corrections $c_{2.5D}$ are neglected. The driving signal of the n-th secondary source is given by

$$D(\alpha_{0,n}, \omega) = \frac{2}{\pi} a_{fs}(\alpha_{0,n}) \frac{\omega}{c} \sum_{m=1}^{M} a_{ps}(\alpha_{fs,m}) \times$$

$$\frac{(\mathbf{x}_{fs,m} - \mathbf{x}_{ps})^T \mathbf{n}_{fs,m}}{|\mathbf{x}_{fs,m} - \mathbf{x}_{ps}|^{3/2}} \frac{(\mathbf{x}_{0,n} - \mathbf{x}_{fs,m})^T \mathbf{n}_{0,n}}{|\mathbf{x}_{0,n} - \mathbf{x}_{fs,m}|^{3/2}} \times$$

$$e^{-j\frac{\omega}{c}(|\mathbf{x}_{fs,m} - \mathbf{x}_{ps}| - |\mathbf{x}_{0,n} - \mathbf{x}_{fs,m}|)}, \quad (8)$$

where $\mathbf{x}_{0,n} \in \partial V$ denotes the position of the n-th secondary source and $\mathbf{x}_{fs,m} \in \partial V_l$ the position of the m-th focused source (virtual secondary source) with main propagation direction $\mathbf{n}_{fs,m}$. The extrapolated HRTFs are given by discretizing (3) as

$$\tilde{H}_{\{L,R\}}(\alpha_{ps}, r_{ps}, \varphi, \omega) =$$

$$\sum_{n=1}^{N} D(\alpha_{0,n}, \omega) H_{\{L,R\}}(\alpha_{0,n}, R, \varphi, \omega), \quad (9)$$

where weights due to the discretization of (3) have been omitted. For equiangular sampling these weights are constant and hence do not play a major role in the context of HRTFs. For practical application the number of virtual secondary sources should not exceed the number of measured HRTFs [28].

Note, the proposed technique inherently performs an angular interpolation of the HRTFs. The position of the virtual point source is not restricted to the angles given by the angular discretization of the (virtual) secondary sources.

Local WFS is not restricted to the synthesis of virtual point sources. The concept holds also for all other source types that can be synthesized by conventional WFS. Hence, it is also possible to compute far-field HRTFs or HRTFs for directive sources using

virtual plane waves or virtual sources with complex directivity [32, 33].

4.3. Spatial Sampling

For focused sources the radius r_{al} of the region with negligible sampling artifacts can be estimated as [31]

$$r_{al} < \frac{N_a' c}{\pi e f}, \quad (10)$$

where $N_a' = 2N_a + 1$ with N_a denoting the number of active secondary sources as given by the secondary source selection criteria (5). Equation (10) can be used to estimate the required angular sampling of a measured HRTF dataset. Approximately half of the measured HRTFs is used for the synthesis of a focused source if $r_{fs} \ll R$. Hence, the following lower limit for the number N of equiangular HRTF measurements in the horizontal plane holds

$$N > \frac{\pi e f r_a l}{c}. \quad (11)$$

For a typical head diameter of $20\,\text{cm}$ and full audio bandwidth of $20\,\text{kHz}$, Eq. (11) states that $N = 100$ is sufficient. However, this would require that the secondary sources are placed at a quite close distance of $r_{fs} = 0.1\,\text{m}$ around the center of the virtual head. In practice, r_{fs} should be chosen larger and consequently more measured HRTFs are required.

Practice has shown that the condition (10) holds also quite well for the local synthesis of a virtual point source using virtual secondary sources. Hence, the extrapolated HRTFs are free of major secondary source sampling artifacts below the frequencies predicted by (10). The authors will further investigate on this topic.

4.4. Implementation

A number of aspects have to be considered for a practical implementation of the HRTF extrapolation technique using (8) and (9). These are discussed briefly in the following.

2.5-dimensional synthesis techniques suffer from systematic amplitude deviations in the synthesized sound field. These deviations are consequently also present in the extrapolated HRTFs if no countermeasures are taken. They will be most prominent for lateral sources.

For focused sources spatial weighting of the driving signal (also known as tapering) is important to

reduce artifacts due to truncation [21]. For illustration of the basic principle, this tapering has been neglected in (8). However, its straightforward to incorporate tapering into the window function $a(\alpha_0)$ used for secondary source selection For the practical implementation, the traditional tapering techniques of WFS [23] have been used.

WFS allows for a very efficient computation of the loudspeaker driving functions [23]. The same holds also for local WFS [28]. It can be concluded from an inverse Fourier transformation of (8) and (9) that HRTFs can be extrapolated using local WFS by superposition of weighted and delayed versions of these HRTFs followed by a filtering with a filter having the high-pass characteristic ω. For delays that are restricted to integers of the temporal sampling interval this scheme can be implemented very efficiently by delay lines. However, fractional delay lines may improve the results. Alternatively, it is also possible to compute a set of near-field HRTFs as described in [31]. This set of HRTFs can then be interpreted as a virtual loudspeaker array which is driven by traditional WFS.

Another practical issue has to be considered with respect to the filtering of the HRTFs (as described above). It is known from the theory of WFS [22] that this filter should only be applied in a frequency range were no major sampling artifacts are present. Consequently, this filter should be flattened out above a certain frequency when the extrapolation is performed for the case that spatial aliasing is present.

5. RESULTS

The properties of the proposed extrapolation technique are illustrated in the following section. Since local WFS relies on the accurate synthesis of focused sources, we first investigate the properties of extrapolated near-field HRTFs using traditional WFS. This is followed by results of the proposed technique employing local WFS for range extrapolation of HRTFs.

5.1. Experimental Setup

The proposed technique has been implemented in a two-stage structure. In a first stage, a set of near-field HRTFs is computed using the concept of focused sources in WFS. These are then used in a second step as virtual secondary sources for the application of local WFS to range extrapolation of

HRTFs. The two-stage structure has the benefit that the properties of the extrapolated near-field HRTFs can be investigated separately from the extrapolated HRTFs derived by local WFS. However, it has been ensured by the authors that the results are equal to a direct implementation of the proposed technique using (8) and (9).

For computation and evaluation of the extrapolated HRTFs a custom dataset of horizontal-plane HRTFs captured for different source distances is used. The dataset has been captured with the KEMAR mannequin at a angular resolution of one degree for the four source distances $R = \{3, 2, 1, 0.5\}$ m. Refer to [34] for a detailed description of the freely available dataset. The measurements taken at a distance of 3 m are used for the extrapolation to other distances.

5.2. Properties of HRTFs for Nearby Sources

The perceptually relevant properties of HRTFs with respect to the incidence angle of a source have been investigated extensively in the past decades [1]. The properties of HRTFs depend also on the distance of the source to the listener. This holds especially for nearby sources, which are often referred to as *near-field sources* in the literature. The characteristics of near-field HRTFs have been investigated in a number of publications, for instance in [5, 6, 7, 8, 9]. We briefly review the results from the study [5], were HRTF datasets have been measured with a KEMAR mannequin for source distances ranging from 0.12 m to 1 m. The analysis of the datasets revealed that the most relevant changes take place with respect to the interaural level difference (ILD). The ILD increases substantially for lateral sources as the source distance decreases below 1 m. This holds also at low frequencies were the ILD for distant (far-field) sources is small. The interaural time difference (ITD) was found to be roughly independent of the source distance, even for nearby sources. The study also revealed a number of characteristic structures in the magnitude spectrum of the HRTFs, that can be identified when plotting the spectra of the HRTFs over the incidence angle of the source for different distances.

Due to the above discussed properties of near-field HRTFs we will investigate the ILD and magnitude spectrum of the extrapolated HRTFs for evaluation of the proposed approach.

5.3. Properties of Extrapolated Near-Field HRTFs

Figure 6e shows a 6 ms snapshot of the left-ear HRIRs for a source distance of $R = 3$ m. The HRIRs have been temporally aligned with respect to the first wavefront to facilitate comparison. The HRIRs shown in Fig. 6e are used to calculate extrapolated HRIRs using (local) WFS.

Considering the number of active secondary sources for a focused source in the given scenario, condition (10) predicts that the region which is free of spatial sampling artifacts is $r_{al} < 64$ cm assuming full audio bandwidth of 20 kHz. For a typical head diameter the extrapolation of the measured HRIRs from $R = 3$ m to $r_{fs} = 0.5$ m is just within this limit. Hence, this case will be regarded in the following.

Figure 6a shows the measured HRIRs for a source distance of $R = 0.5$ m. These will serve as a reference for the extrapolated HRIRs at the same distance. Figure 6b shows the extrapolated HRIRs for a source distance of $r_{fs} = 0.5$ m. Inspection of Fig. 6b and comparison with Fig. 6a reveals that the extrapolated HRIRs are free of sampling and other artifacts as predicted by (10). It can be observed furthermore, the extrapolated HRIRs are very close to the measured HRIRs at the same distance.

As discussed in Section 5.2, the ILD changes for nearby sources and is probably a strong cue for humans to estimate the distance of a source. Hence, the proposed method should accurately reproduce the ILD. In the left of Fig. 7 the ILD for the two measured data sets with $R = 3$ m and $R = 0.5$ m (gray lines) is compared to the ILD of the extrapolated data set for $r_{fs} = 0.5$ m. As can be seen, the ILD is quite close to the measured one for the nearby source. Minor deviations are only visible at the lateral positions of the head with $\alpha_0 = 0°$ and $\alpha_0 = -180°$. These may be accounted to the characteristics of the loudspeaker used for the HRTF measurements and to the amplitude deviations of 2.5-dimensional synthesis.

Another cue for the perception of nearby sources is an amplitude increase for low frequencies. This can be seen in the right of Fig. 7. Below 500 Hz the amplitude is higher for both the measured and the extrapolated near field HRTF. For frequencies below 200 Hz deviations can be seen for the extrapolated HRIR. This might be accounted to the high frequency approximations of WFS. However, this could be corrected by a modification of the used pre-equalization filter.

5.4. Range Extrapolation by Local Wave Field Synthesis

The previous section has shown that it is possible to extrapolate HRIRs to distances that are closer than the measured one using focused sources of traditional WFS. The evaluation of the sampling condition (10) for the considered scenario, as given above, shows that the extrapolation to distances larger than r_{fs} but smaller than the measurement distance will not be possible without artifacts. Figure 6d shows the results when extrapolating a measured HRIR dataset from $R = 3$ m to $r_{fs} = 2$ m. For comparison, Fig. 6c shows the measured HRIRs for a source distance of $R = 2$ m. Comparing both figures, spatial sampling artifacts can be observed as additional wave fronts occurring before the first wavefront of the desired focused source for $t < 0$ ms [24]. This is a consequence of the time-reversal nature of acoustic focusing. Since the human auditory system is quite sensitive to pre-echos, one can expect audible artifacts in this case. Hence, traditional WFS cannot be used for extrapolation in this case.

The extrapolated HRIRs for a distance of $r_{fs} = 0.5$ m have now been used as virtual secondary sources for extrapolation using local WFS. Note that in the present scenario, the virtual secondary sources synthesize a virtual source which is non-focused with respect to the virtual secondary sources. The results for a source distance of $r_{ps} = 2$ m are shown in Fig. 6f. Comparison of with Fig. 6d reveals that no sampling artifacts are present using the proposed technique. The local synthesis technique allows sampling artifact free extrapolation of HRTFs. Comparing Fig. 6f with 6c shows also that the extrapolation results are accurate. This will also hold for other source distances, since the spatial sampling artifacts of virtual point sources are approximately independent from the source position. The results indicate that the proposed technique provide accurate extrapolation results. Due to the minor changes with respect to the ILD for distant sources, the authors have not compared the ILDs of measured and synthesized HRIRs for $r_{ps} = 2$ m.

The extrapolated HRIRs have also been used in a virtual auditory environment (VAE). Results from informal listening supports the properties found
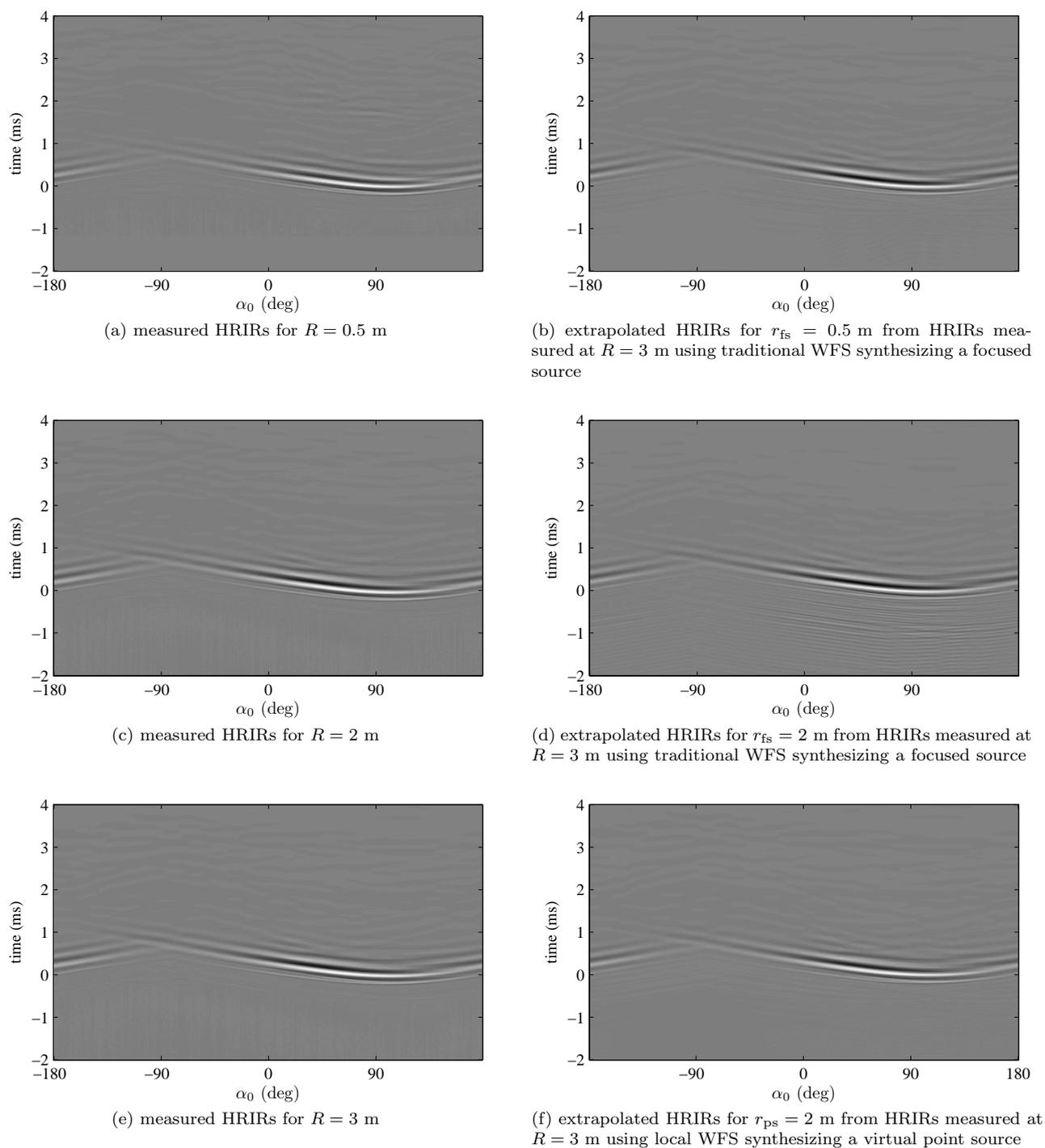
(a) measured HRIRs for $R = 0.5$ m

(b) extrapolated HRIRs for $r_{\text{fs}} = 0.5$ m from HRIRs measured at $R = 3$ m using traditional WFS synthesizing a focused source

(c) measured HRIRs for $R = 2$ m

(d) extrapolated HRIRs for $r_{\text{fs}} = 2$ m from HRIRs measured at $R = 3$ m using traditional WFS synthesizing a focused source

(e) measured HRIRs for $R = 3$ m

(f) extrapolated HRIRs for $r_{\text{ps}} = 2$ m from HRIRs measured at $R = 3$ m using local WFS synthesizing a virtual point source

Fig. 6: Snapshots of HRIRs (left-ear) for $\varphi = 90°$ and different source distances. The left row shows measured HRIRs, the right row extrapolated HRIRs using (local) WFS. The temporal axis has been shifted.
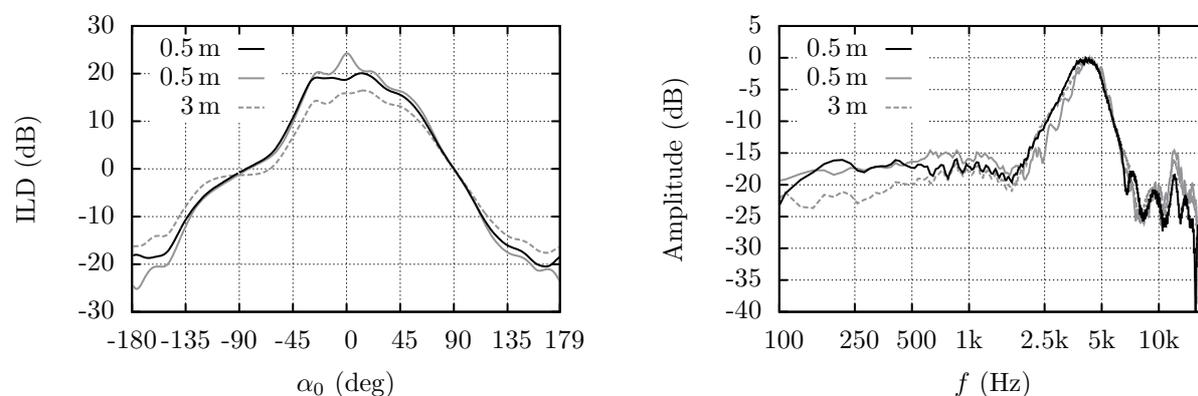
Fig. 7: Interaural level differnces (ILD) for all angles α_0 (left). Magnitude spectrum for the left HRTF for an angle of $\alpha_0 = 0°$ (right). The dotted gray lines show the measurement results for the HRTF dataset with $R = 3\,\text{m}$. The gray lines for the measured dataset with $R = 0.5\,\text{m}$ and the black lines the extrapolated dataset with $r_\text{fs} = 0.5\,\text{m}$.

above. Sound samples may be downloaded at `http:\\audio.qu.tu-berlin.de\?p=570`.

6. SUMMARY AND CONCLUSIONS

This paper presented a method for interpolation and range extrapolation of free-field HRTFs to distances other than the measurement distance. The HRTFs are interpreted as a virtual loudspeaker array which is driven by local WFS. In contrast to traditional WFS, local WFS aims at the accurate synthesis within a local listening area instead of the entire area surrounded by the loudspeakers. This comes at the cost of more prominent artifacts outside of the local listening area. However, this is not a problem for range extrapolation since here the synthesized sound field has only to be correct around the head of the listener.

The application of local WFS for range extrapolation of HRTFs has a number of benefits in contrast to other methods:

- accurate extrapolation results over the entire frequency range for typical spatial sampling intervals of measured HRTFs,

- inherent interpolation capabilities for extrapolated HRTFs,

- the method is not prone to numerical instabilities,

- low computational complexity due to weighted delay line structure,

- HRTFs measured on non circular/spherical contours can be used, and

- the possibility to calculate synthetic HRTFs for non point sources.

However, the are also limits of presented approach and study that should be mentioned:

- low frequency properties are not optimal due to the approximations used in WFS, and

- a formal perceptual evaluation of the method has not been performed so far.

For improvement of the former limit, the combination of local WFS for higher frequencies with traditional WFS for lower frequencies seems to be a promising solution. Another possibility would be to use near-field compensated higher-order Ambisonics (NFC-HOA) for the synthesis of the focused sources which are then driven by traditional WFS. The concept of local SFS using virtual secondary sources allows for such combinations [28]. A formal listening experiment investigating the perceptual properties of the proposed method is in progress.

The proposed technique shares some interesting properties with NFC-HOA. The synthesis of a sound field by NFC-HOA using a circular loudspeaker array is most accurate in a circular zone around the center. The size of this zone decreases as frequency increases. According to [35] limits can be given for the reconstruction of band-limited sound field in a limited area with limited error ($\epsilon < 16\%$) from N weighted point sources. These results should also hold for range extrapolation of HRTFs using NFC-HOA. For two-dimensional synthesis the required number of measured HRTFs is given by $N = 100$ for a typical head diameter and full audio bandwidth. Interestingly, a similar result has been derived in Section 4.3 for the presented approach. However, the application of local WFS to extrapolation of HRTFs has the benefits of low computational complexity and numerical stability unlike NFC-HOA.

These two factors are of special interest for VAEs were a database of HRTFs is required that covers all potential source positions. Using the proposed technique the required HRTFs could be computed online with low computational complexity.

7. REFERENCES

[1] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, 1996.

[2] B. Gardner and M. Keith. HRTF measurements of a KEMAR dummy-head microphone. *MIT Media Lab Perceptual Computing - Technical Report #280*, May 1994.

[3] J. Blauert, M. Brueggen, A.W. Bronkhorst, R. Drullman, G. Reynaud, L. Pellieux, W. Krebber, and R. Sottek. The AUDIS catalog of human HRTFs. *The Journal of the Acoustical Society of America*, 103(5):3082–3082, 1998.

[4] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano. The CIPIC HRTF database. In *Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, October 2001.

[5] D.S. Brungart and W.M. Rabinowitz. Auditory localization of nearby sources. Head-related transfer functions. *Journal of the Acous-tical Society of America*, 106(3):1465–1479, September 1999.

[6] D.S. Brungart, N.I. Durlach, and W.M. Rabinowitz. Auditory localization of nearby sources. II. Localization of a broadband source. *Journal of the Acoustical Society of America*, 106(4):1956–1968, October 1999.

[7] B.G. Shinn-Cunningham, N. Kopco, and T.J. Martin. Localizing nearby sound sources in a classroom: Binaural room impulse responses. *Journal of the Acoustical Society of America*, 117(5):3100–3115, May 2005.

[8] T. Lentz. *Binaural Technology for Virtual Reality.* PhD thesis, RWTH Aachen, 2007.

[9] J. Huopaniemi and K. Riederer. Measuring and modeling the effect of source distance in head-related transfer functions. In *International Congress on Acoustics*, Seattle, USA, June 1998.

[10] D. Romblom and B. Cook. Near-field compensation for HRTF processing. In *125th AES Convention*, San Francisco, USA, October 2008. Audio Engineering Society (AES).

[11] D. Brungart. Near-field virtual audio displays. In *IMAGE 2000*, Scottsdale, USA, July 2000.

[12] R. Duraiswami, D.N. Zotkin, and N.A. Gumerov. Interpolation and range extrapolation of HRTFs. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[13] W. Zhang, T.D. Abhayapala, and R.A. Kennedy. Modal expansion of HRTFs: continuous representation in frequency-range-angle. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Japan, 2009.

[14] D. Menzies and M. Al-Akaidi. Nearfield binaural synthesis and Ambisonics. *Journal of the Acoustic Society of America*, 121(3):1559–1563, March 2007.

[15] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich. A 3D ambisonic based binaural

sound reproduction system. In *AES 24th International Conference on Multichannel Audio*, Banff, Canada, June 2003. Audio Engineering Society (AES).

[16] J. Ahrens and S. Spors. Focusing of virtual sound sources in higher order ambisonics. In *124th AES Convention*, Amsterdam, The Netherlands, May 2008. Audio Engineering Society (AES).

[17] A.J. Berkhout. A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, 36:977–995, December 1988.

[18] E.W. Start. Application of curved arrays in wave field synthesis. In *110th AES Convention*, Copenhagen, Denmark, May 1996. Audio Engineering Society (AES).

[19] S. Spors, R. Rabenstein, and J. Ahrens. The theory of wave field synthesis revisited. In *124th AES Convention*. Audio Engineering Society (AES), May 2008.

[20] E.G. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[21] S. Spors. Extension of an analytic secondary source selection criterion for wave field synthesis. In *123th AES Convention*, New York, USA, October 2007. Audio Engineering Society (AES).

[22] S. Spors and J. Ahrens. Analysis and improvement of pre-equalization in 2.5-dimensional wave field synthesis. In *128th AES Convention*, pages 1–17, London, UK, May 2010. Audio Engineering Society (AES).

[23] E.N.G. Verheijen. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1997.

[24] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens. Physical and perceptual properties of focused sources in wave field synthesis. In *127th AES Convention*. Audio Engineering Society (AES), October 2009.

[25] S. Spors and R. Rabenstein. Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis. In *120th AES Convention*, Paris, France, May 2006. Audio Engineering Society (AES).

[26] J. Ahrens, H. Wierstorf, and S. Spors. Comparison of higher order ambisonics and wave field synthesis with respect to spatial discretization artifacts in time domain. In *AES 40th International Conference on Spatial Audio*, Tokyo, Japan, October 2010. Audio Engineering Society (AES).

[27] H. Wittek. *Perceptual differences between wavefield synthesis and stereophony*. PhD thesis, University of Surrey, 2007.

[28] S. Spors and J. Ahrens. Local sound reproduction by virtual secondary sources. In *AES 40th International Conference on Spatial Audio*, pages 1–8, Tokyo, Japan, October 2010. Audio Engineering Society (AES).

[29] J. Ahrens and S. Spors. On the scattering of synthetic sound fields. In *130th AES Convention*. Audio Engineering Society (AES), May 2011.

[30] S. Spors and J. Ahrens. Generation of far-field head-related transfer functions using sound field synthesis. In *German Annual Conference on Acoustics (DAGA)*, March 2011.

[31] S. Spors and J. Ahrens. Efficient range extrapolation of head-related impulse responses by wave field synthesis techniques. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011.

[32] E. Corteel. Synthesis of directional sources using wave field synthesis, possibilities, and limitations. *EURASIP Journal on Advances in Signal Processing*, 2007, 2007. Article ID 90509.

[33] J. Ahrens and S. Spors. Implementation of directional sources in wave field synthesis. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, October 2007.

[34] H. Wierstorf, M. Geier, A. Raake, and S. Spors. A free database of head related impulse response measurements in the horizontal plane with multiple distances. In *130th AES Convention.* Audio Engineering Society (AES), May 2011.

[35] R.A. Kennedy, P. Sadeghi, T.D. Abhayapala, and H.M. Jones. Intrinsic limits of dimensionality and richness in random multipath fields. *IEEE Transactions on Signal Processing*, 55(6):2542–2556, 2007.