Bibliographical reference

Spherical Harmonics Based HRTF Datasets: Implementation and Evaluation for Real-Time Auralization, Jan-Gerrit Richter, Martin Pollow, Frank Wefers and Janina Fels, Acta Acustica **vol. 100** (Number 4), 2014, pp. 667-675

DOI

https://doi.org/10.3813/AAA.918746

Acta Acustica united with Acustica
Vol. 100 (2014) 667 – 675

DOI 10.3813/AAA.918746

Spherical Harmonics Based HRTF Datasets: Implementation and Evaluation for Real-Time Auralization

Jan-Gerrit Richter, Martin Pollow, Frank Wefers, Janina Fels
Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstr. 5, 52072 Aachen, Germany.
jri@akustik.rwth-aachen.de

Summary
Head-related transfer functions (HRTFs) are commonly used to simulate sound events from arbitrary directions in binaural synthesis applications. These HRTFs are usually either measured or simulated with a discrete angular resolution. Using this data directly has several drawbacks, mainly due to the necessary interpolation between these discrete sampling points as well as simplifications in the distance dependence of the near-field HRTF. This paper describes how *spherical harmonics (SH)* can be used to process HRTF data with the use in real-time capable auralization environments. During the generation of the SH coefficients, the optimal transformation center point is used to achieve a compact transformation. Several strategies are introduced to achieve fast filter reconstruction suitable for real-time auralization.

PACS no. 43.60.Sx, 43.66.Pn

1. Introduction

Human beings have the ability to localize the direction of a sound source. This is mainly achieved with the recognition of a combination of different cues from the two ear signals. The *interaural time difference (ITD)* and the *interaural level difference (ILD)* help the localization on the horizontal plane, while sound diffraction from the torso, head and pinna influence the localization ability in the elevation [1]. If a localization environment can be described as a *linear time-invariant (LTI)* environment, then the influence of the head in the wave expansion can be represented as a linear spatial filter called the *head-related transfer function (HRTF)* [2].

These HRTF filters can be obtained either by measurement [3, 4, 5, 6] or by numerical simulation [7, 8, 9, 10] and can be used to render virtual sources using binaural synthesis. In virtual environments, the ability to create virtual sources in any direction relative to the subject is important to achieve a natural sound reproduction and helps with the immersion into the virtual scene. The required HRTF filters for multiple directions are usually saved in a discrete set with a sufficient high spatial resolution. The spatial resolution depends on the *just noticeable difference (jnd)* which is determined by listening tests. These tests usually give jnd values of 1° in the horizontal plane and frontal direction and a value of 5° in back direction [1, 11]. A more comprehensive overview of HRTF representation methods is given in [12]. Note that most HRTF representations with discrete sets have the advantage of very fast filter generation and will perform better in terms of computational effort needed for the filter generation. The filter generation time is the time that is needed, before the filter is loaded and available for further processing. As the data is saved in the same format as it is used, this time is negligible. Despite these advantages, the discrete data format has several downsides that are now discussed in detail.

Between the discrete points no information is available. To generate a filter between points, an interpolation method can be applied [13]. This interpolation will result in a filter that is not a physically correct HRTF representation due to smoothing. Depending on the resolution of measured points, noticeable differences may be heard [14]. It was also shown in [15, 16] that humans use small head movements to improve the localization. These movements are not reproducible with discrete HRTF sets. Furthermore, the discretely measured set represents an HRTF for one specific distance. This distance will most likely represent the HRTF in the far-field. It was shown that the distance dependence in the near-field HRTF is important in near-to-head source imaging [17, 18]. A distance dependent HRTF can be realized with multiple measurements, which will increase the measurement time and the required hard disk space [19]. Additionally, further interpolation between distances is needed.

Contrary to these discrete methods, a representation of HRTF data sets using continuous models can be used [20, 21]. In this paper, a representation using *spherical harmonics (SH)* is used as introduced in [22]. This rep-

Received 10 June 2013,
accepted 15 April 2014.

resentation allows natural reconstruction of the filter for any desired direction and distance [23]. If sampled on a sufficient dense spherical grid a physically correct interpolation of the HRTF is possible, allowing a spatially continuous representation of the HRTF. The distance dependence of the near-field HRTF can be artificially extrapolated from one measurement by the use of the wave propagation term for spherical waves [24]. This enables a correct representation of the HRTF for any source distance without additional measurement effort.

This approach will introduce more computational complexity into the filter generation compared to the discrete set. The feasibility and performance of this approach in a real-time auralization application is studied in this paper and an efficient implementation is introduced.

This efficiency is achieved in two ways during the two transformation steps. The optimal selection of the center point for the transformation can achieve a compact spherical harmonic representation [25] which will lower reconstruction times. The inverse transformation has to be computed as fast as possible to allow a real-time auralization. Several methods to reduce the calculation time are described.

Dataset

In this paper, two HRTF datasets are used to test and evaluate the algorithms. Both sets represent the HRTF of the ITA artificial head [26]; the first set being a BEM simulation of the head, while the second set stems from a measurement. The simulation is sampled on a Gaussian grid [27] with order 89 and 16200 points while the measurement is obtained from a system that samples with maximum order of 47 while neglecting a 30 degree south pole cap for practical reasons [6, 28]. This results in 3800 measurement points. The missing cap will make regularization during the transformation necessary [24].

The used reference coordinate system is depicted in Figure 1.

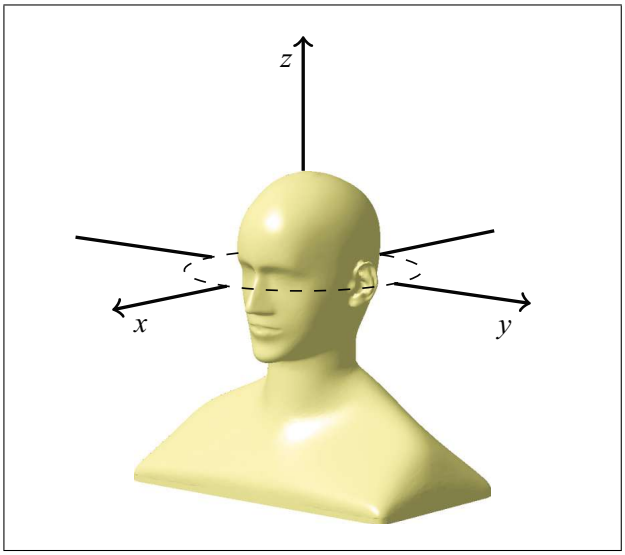

Figure 1. Head-related coordinate system.

2. Spherical Harmonic Transformation

Williams [23] states that any square-integrable function on a sphere $g(\vartheta, \varphi)$ can be described as

$$g(\vartheta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{nm} \mathrm{Y}_n^m(\vartheta, \varphi), \tag{1}$$

with a_{nm} as the *spherical harmonic coefficients* and $\mathrm{Y}_n^m(\vartheta, \varphi)$ as the *spherical harmonic base functions*. Note that the dependencies on frequency and measurement distance are omitted to improve readability.

The *spherical harmonic base functions* are defined as

$$\mathrm{Y}_n^m \equiv \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} \cdot \mathrm{P}_n^m(\cos\vartheta) \cdot \mathrm{e}^{\mathrm{j}m\varphi}, \tag{2}$$

with the *order n* and the *degree m* and the associated Legendre function $\mathrm{P}_n^m(\cos\vartheta)$. Equation (1) is called the *inverse spherical harmonic transformation (ISHT)*. The equation is used to reconstruct filter from the *spherical harmonics* domain. The analogue *spherical harmonic transformation (SHT)*, computes the SH coefficients from the spatial domain.

2.1. Spherical Harmonic representation of HRTF data

In the following, HRTF data sets are regarded as a frequency dependent directivity pattern of a point source inside the blocked ear canal using Helmholtz's reciprocity principle. Assuming that the filter for each direction have the same length, each DFT coefficient per available direction is used to create a directivity pattern for every frequency bin. These directivity patterns can be described as a spherical function $g(f, r_0, \vartheta, \varphi)$ with r_0 as the measurement distance and ϑ and φ as discrete spherical angular parameters.

These functions can then be transformed to the SH domain with the use of the SHT. The SHT will result in frequency dependent coefficients up to a maximum SH order. This way, the HRTF data of each ear is compressed from

$$n_{\mathrm{spatial}} = n_{\mathrm{freq}} \cdot n_{\mathrm{points}} \tag{3}$$

values needed for spatial representation to

$$n_{\mathrm{SH}} = n_{\mathrm{freq}} \cdot (N_{\mathrm{freq}} + 1)^2 \tag{4}$$

values with a spherical harmonic representation. Here, n_{freq} is the number of frequency bins of the filter, n_{points} is the number of measurement points and N_{freq} is a frequency dependent spherical harmonic order limit.

To use the HRTF in an auralization application, the filter have to be reconstructed using the ISHT (see equation 1).

2.2. Transformation Error

During the SHT and the subsequent ISHT of spatial HRTF data, several errors may be introduced into the transformation result. Two of these errors are discussed in the following.

The number of the spatial sampling points are determined by the measurement setup. This limits the correct SH transformation to signals of maximum order N_S [23, 25]. For signals, that exceed this maximum transformation order, aliasing error will be introduced into the resulting SH coefficients [29].

During the HRTF filter reconstruction, a different error can be introduced. If the calculation is done only to a maximum order N_I, with $N_I \leq N_S$, then the energy from all SH-order up to the maximum available order N_S is lost. This error is called truncation error and is defined as $e(\vartheta, \varphi)$ [30]:

$$\hat{g}(\vartheta, \varphi) = \sum_{n=0}^{N_I} \sum_{m=-n}^{n} a_{nm} \mathrm{Y}_n^m(\vartheta, \varphi), \tag{5}$$

$$e(\vartheta, \varphi) = \sum_{n=N_I+1}^{N} \sum_{m=-n}^{n} a_{nm} \mathrm{Y}_n^m(\vartheta, \varphi), \tag{6}$$

$$g(\vartheta, \varphi) = \hat{g}(\vartheta, \varphi) + e(\vartheta, \varphi), \tag{7}$$

with $g(\vartheta, \varphi)$ as a signal with maximum order $N_I \leq N \leq N_S$. A truncation of higher SH orders corresponds to a spatial lowpass filter as it smoothes spatial details.

2.3. Distance Transformation

The range extrapolation allows the correct representation of the HRTF on any desired distance from data corresponding to only one distance.

The SH-coefficients a_{nm} in equation (1) can be subdivided into a pure frequency dependent component b_{nm} multiplied with a term of the outgoing Hankel function (see equation 8) [31]. According to the sign convention, the outgoing Hankel function is of second kind $h_n^{(2)}(kr)$ [31],

$$a_{nm}(r, k) = b_{nm}(k) h_n^{(2)}(kr). \tag{8}$$

The combination of equations (1) and (8) allows a complete isolation of the distance dependence of the HRTF in the Hankel function term. During the reconstruction of HRTF filter from SH coefficients, one additional step has to be taken to accomplish a range extrapolation to any desired distance: The calculated SH coefficients from the HRTF measurement represent the wave expansion at the measurement distance r_0. The corresponding Hankel term has to be replaced with a spherical Hankel term of the desired distance as [31]

$$a_{nm}(r_1, k) = a_{nm}(\mathrm{r}_0, k) \cdot \frac{h_n(kr_1)}{h_n(k\mathrm{r}_0)}. \tag{9}$$

3. Acoustic Centering

As shown in [25], the location of a source inside a surrounding spherical microphone array will influence the energy distribution in the spherical harmonic order. The most *compact* transformation is achieved when the sound source is located in the center of the spherical measurement array. In this case, most of the energy is contained in the lower orders. This will reduce the truncation error described in equation (6) during the filter reconstruction, or will enable to truncate more coefficients while introducing only truncation errors of comparable energy.

In typical HRTF measurements, the center point of the head is used as the measurement center. This leads to a non-optimal sound expansion with respect to the measurement array. To achieve a compact transformation, the HRTF data set of each ear is moved to the frequency dependent optimal center point. To accomplish this, the range extrapolation approach given in equation (9), with a direction dependent radius can be used.

Ben Hagai *et al.* [25] introduced several measures that quantify the energy distribution and can be used in an optimization routine to find the optimal center point. Here, the "center of power" (J_2) function is named the most suitable function and is therefore used in this paper,

$$J_2 = \sum_{n=0}^{N} \sum_{m=-n}^{n} n \left|a_{nm}\right|^2 / L_2, \tag{10}$$

$$\text{with} \quad L_2 = \|a\|_2^2 = \sum_{n=0}^{N} \sum_{m=-n}^{n} \left|a_{nm}\right|^2. \tag{11}$$

The location of the source within a spherical microphone array will not only influence the energy distribution, but also the order limitation of the measured source [25]. The optimized center point should reduce the order limitation of the measured HRTF function. This will reduce aliasing errors that occur when the signals are of higher order than the sampling of the measurement setup allows. Together with the improved truncation error, the transformation error as described in Section 2.2 should be reduced by the centering.

3.1. Optimization Results

To find the optimal center point of the HRTF, an optimization routine with the function J_2 (equation 10) is done. Figures 2 and 3 show the results from an optimization of the left ear of a BEM simulation and measurement of the ITA artificial head. They show the location of the optimized center point relative to the coordinate origin in the head center. Both figures show a similar overall behavior. The y-axis offset shows an optimal acoustic center that is outside of the head itself for low frequencies and converges to the ear position for increasing frequencies. This behaviour is also found in loudspeaker setups where, at low frequencies, the driver moves the air in front of it in phase. This causes the wave to emanate in front of the speaker membrane [32]. The x and z axes show only small

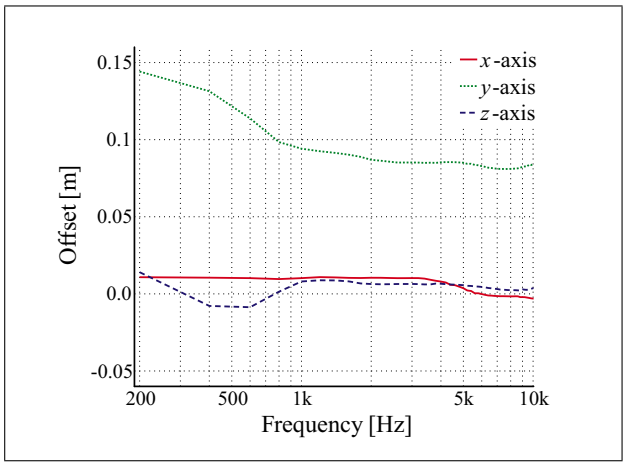

Figure 2. BEM simulated artificial head - Optimized center point for the left ear with criterion J_2 (equation 10).

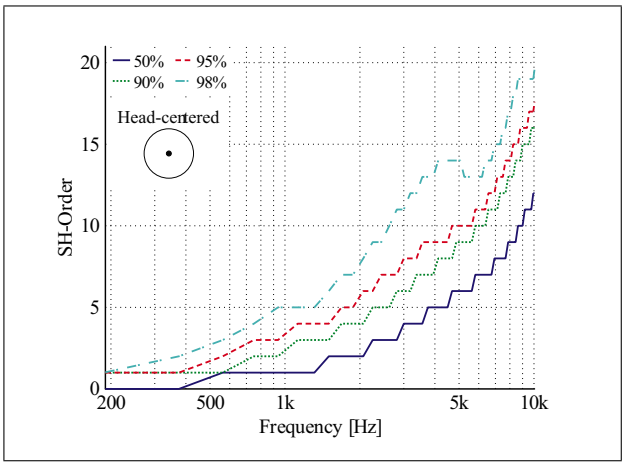

Figure 4. Energy distribution in SH coefficients; head-centered transformation; left ear – BEM simulation data.

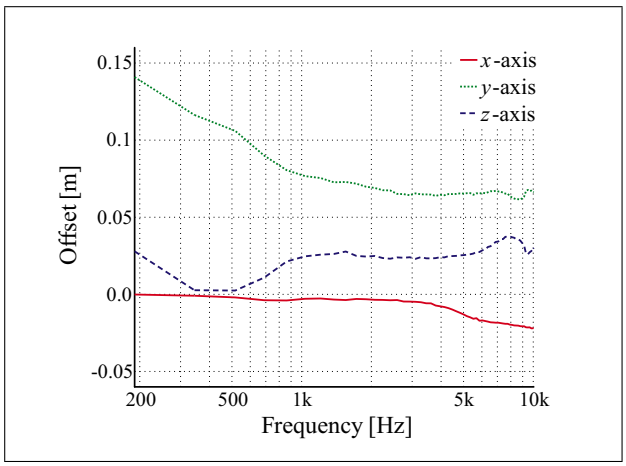

Figure 3. Measured artificial head - Optimized center point for the left ear with criterion J_2 (equation 10).

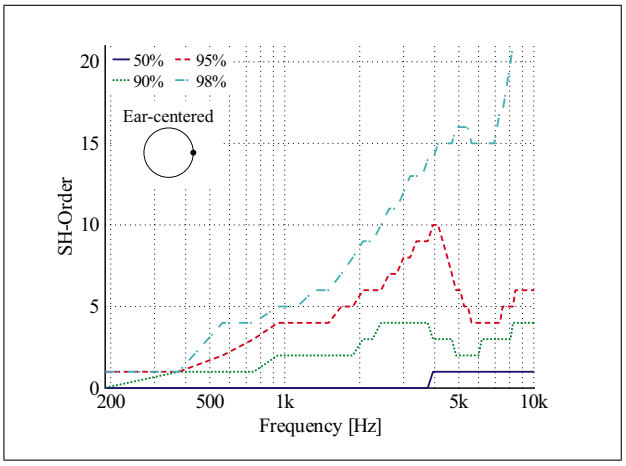

Figure 5. Energy distribution in SH coefficients; ear-centered transformation; left ear – BEM simulation data.

variations for the BEM simulated HRTFs. The z-axis offset of the measurement data indicates a slight offset in the height of the head within the grid during the measurement.

3.2. Energy Distribution

To validate the optimization results, the normalized energy distribution in the SH order is plotted in Figures 4 and 5 for a head-centered transformation of the BEM HRTF and the optimized centered transformation, respectively.

The figures show the minimum order needed to obtain 50 %, 90 %, 95 % and 98 % of the signals energy. For the head-centered transformation, two effects can be observed: First, the majority of the energy is located in increasingly high SH order with increasing frequency. It can be observed that the energy is also spread out over a bigger range, as the 50 % line increases.

When looking at the energy distribution for the ear-centered transformation, it is clear that the energy gets concentrated into low orders with a smaller spread. 50 % of the energy are always located in the first orders. 90 % is achieved with orders below 5, even at higher frequencies. This should improve the truncation error during the filter reconstruction. Interestingly, the line for the 98 % energy does not show improvement.

3.3. Reconstruction Error

Both the reduction of aliasing errors during the SHT and the reduction of the truncation error during the ISHT that results from the optimized center point should lower the overall transformation error. To quantify this error and the improvements, the following error criterion is used: The original spherical HRTF data in the frequency domain $g(f, \Omega)$ is transformed into SH coefficients with a maximum order of 89. Subsequently all HRTF filters for every direction are reconstructed with a maximum transformation order of N_I. The resulting filter $h_{N_I}(f, \Omega)$ potentially contain both aliasing and truncation error. A squared difference between the original filter and the reconstructed filter is calculated which is integrated over every direction of the sphere. The resulting error energy $r_{N_I}(f)$ is then normalized to the energy of the original signal.

$$r_{N_I}(f) = \oint_{S^2} \left| g(f, \Omega) - h_{N_I}(f, \Omega) \right|^2 \mathrm{d}\Omega, \qquad (12)$$

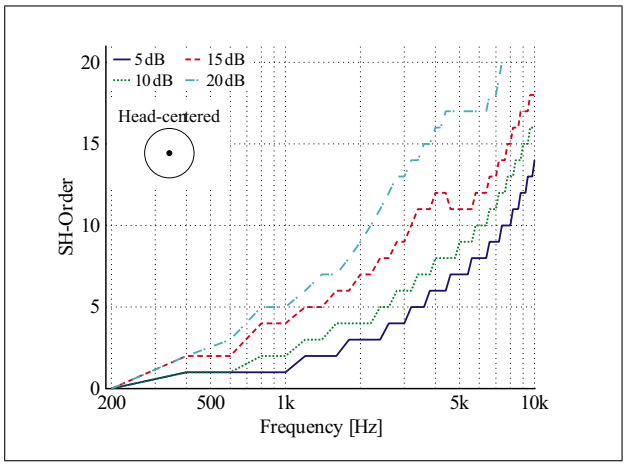

Figure 6. Logarithmic error distance for head-centered transformation; left ear; BEM simulation data.

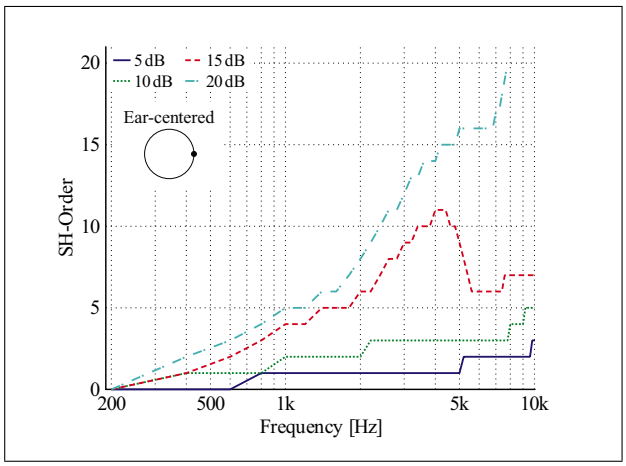

Figure 7. Logarithmic error distance for ear-centered transformation; left ear; BEM simulation data.

$$d(f, N_I) = 10 \log \frac{s(f)}{r_{N_I}(f)} \text{dB}, \tag{13}$$

$$s(f) = \oint_{S^2} |g(f, \Omega)|^2 \, \mathrm{d}\Omega. \tag{14}$$

Figures 6 and 7 show the logarithmic error distance $d(f, N)$ (see equation 13) for maximum transformation order N_I (see equation 5) for the head-centered and ear-centered transformation.

Between the Figures, an improvement of 10–15 dB in low orders and high frequencies is visible. While this error is averaged over the whole sphere, the Figures show a overall significant improvement to the reconstruction error.

To visualize this improvement, the influence of the center point to the error distance is shown in Figure 8. The Figure shows the minimum order needed to achieve 15 dB error with different transformation center points: The original head-centered transformation, one center point halfway between head-center and ear and the optimized

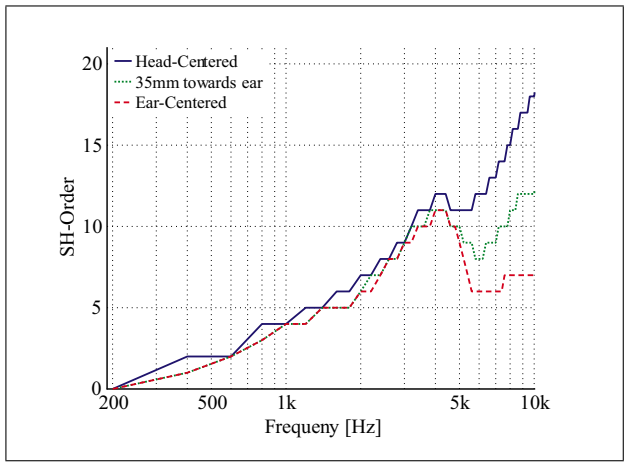

Figure 8. Minimum order to achieve 15 dB error distance with different center points; BEM simulation data.

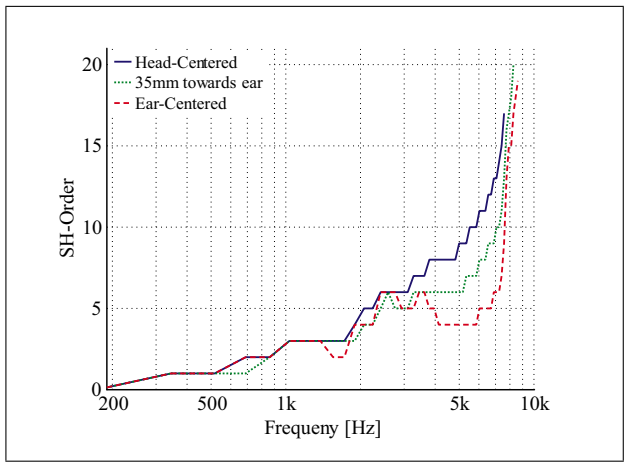

Figure 9. Minimum order to achieve 15 dB error distance with different center points; measurement data.

center. The comparison between the original and the optimized center shows a big improvement in the required order. At frequencies between 5 kHz and 10 kHz the orders can be reduced by a factor of up to three. Figure 9 shows the same graphs for the measured HRTF. Very similar effects are visible here.

4. Implementation

To analyze the feasibility of the SH approach for HRTF datasets in real time auralization application, a C++ implementation into an existing virtual acoustic environment [33] was done.

Here, the improvements accomplished by the acoustic centering in the reconstruction error (see Section 3) are used as one important factor to lower the computation time: As the calculation time will increase quadratically with the order (linear with the number of coefficients), the order reduction quantified as up to a factor of three will significantly improve the calculation time.

Real-time capable auralization is limited by the delay that humans tolerate between visual and auditive stimuli

and is usually stated as 20–30 ms [34, 35]. If the additional latency from the filter generation is too large there will be a noticeable delay between head movement and auralization or delayed auralization of multiple sources. As the overall latency of the auralization system cannot be definitely specified, the goal of the implementation was to achieve shortest calculation time and thus introducting as little additional delay, while remaining as exact as possible.

The filter reconstruction, as shown in equation (1), can be subdivided into three parts: The creation of the SH base function for the desired direction, the calculation of the Hankel values for the range extrapolation (equation 8) and the multiplication of all components. These three components are optimized to achieve low calculation times:

SH Base Function Generation

During the generation of the SH base function, a symmetric property of spherical harmonic can be used to reduce computation effort [23],

$$\mathrm{Y}_n^{-m}(\vartheta,\varphi) = (-1)^m \overline{\mathrm{Y}_n^m(\vartheta,\varphi)}. \tag{15}$$

Here $\overline{x}$ denotes the complex conjugate of the variable x. The use of this symmetry instead of a complete summation, leads to a performance increase by the factor two, as only half the coefficients need to be generated through summation while the other half can easily be calculated.

As seen in Figure 2, the optimal center point is frequency dependent. During the reconstruction, the offset to the head center must be taken into account to compensate this shift. This leads to a frequency dependent sampling of the SH base functions, as every frequency has an independent source direction. As the generation of the SH coefficients for one direction is relatively slow, only one frequency independent translation is done during the SHT. This translation is done with the mean value of the offset seen in Figures 2 and 3. This influences the improvement from the centering only marginally, as the highest error to the optimized position is introduced at low frequencies, where the centering has the least effect

Multiplication

The complex valued multiplication execution time can be reduced with the use of vectorization. Here, Streaming SIMD Extension (SSE) instructions are used to simultaneously compute multiple values. The Intel Performance Primitive (IPP) library supplies a high level interface to these instructions and is therefore used in this paper. This increases the execution time of complex multiplication by a factor of approximately 3. Figure 10 shows measured calculation times for filter generation without range extrapolation for both a pure C++ implementation and an implementation using the *IPP* library.

Range Extrapolation

As the calculation of a single Hankel function value is computationally expensive, the range extrapolation method has to be simplified. This is possible in two different ways.

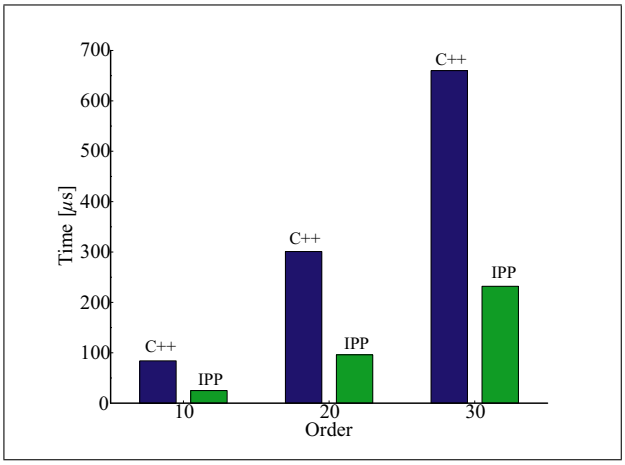

Figure 10. Comparison between C++ and IPP implementation filter generation times without range extrapolation. 256 sample filter.

As a first simplification, a look-up-table (LUT) is precomputed that saves values for every kr value in 5 cm steps. Between the discrete sampling points, a fast linear interpolation is done. This reduced the Hankel calculation time to a inexpensive memory look-up while introducing a small error.

As a second step, the Hankel function calculation is reduced to the near-field of the head. In the far-field, a simple $1/r$-decay can be assumed without significant additional error. The cutoff frequency, where the far-field assumption becomes viable, depends on the maximum transformation order (the higher the order, the higher the cutoff frequency) and on the distance to the source (the closer to the source, the higher the cutoff frequency). Through simulation of the error, a linear dependency with the order and the distance is assumed.

To show the additional error introduced by this approximation, the logarithmic error distance described in equation (13) is calculated for filter generations with complete Hankel function values and with the far-field approximation. The difference between these two values is plotted in Figure 11 for a source distance of 0.8 m. The plot also shows the near-field limit that is used for the given distance and maximum order. Here the assumption of a linear error dependency is validated. The Figure shows that the near-field limit can also be modeled as a low pass filter [36] where the Hankel function is used at low frequencies, while an approximation can be done above a cutoff frequency. Only approximately 0.25 dB additional error is introduced into the filter transformation on average.

4.1. Test System

The presented calculation times stem from runtimes averaged over 10,000 times and measured with *Windows Performance Counters* on an *Intel Core i7 860* processor. All calculations are done with single precision. Between the repeats, random angles are chosen. The distance to

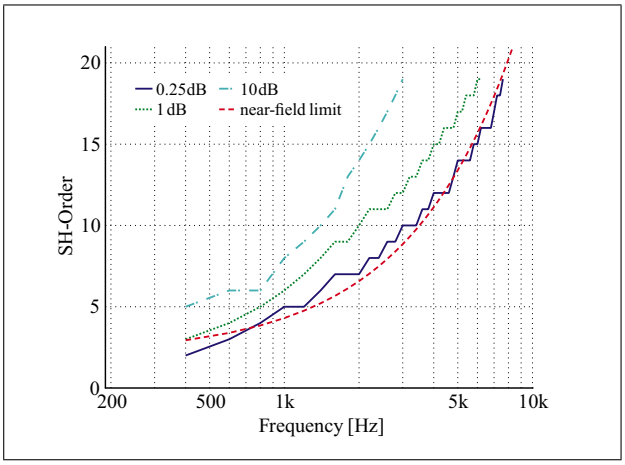

Figure 11. Additional overall error from far-field approximation during range extrapolation. Near-field limit signifies the limit where near-field approximation becomes valid.

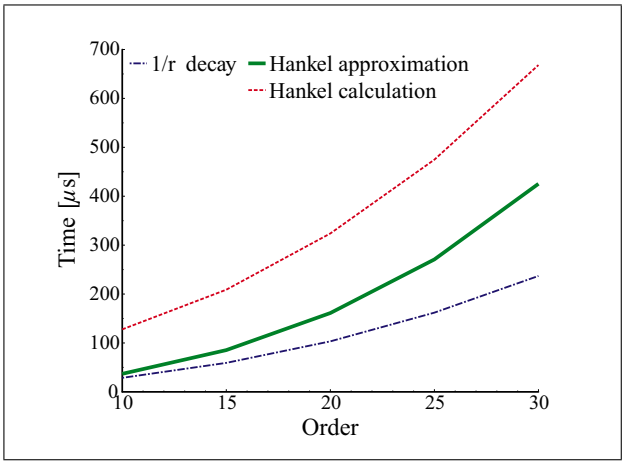

Figure 12. Reconstruction calculation time over different order limits with different range extrapolation methods (filter length 256). "Hankel calculation": full Hankel range extrapolation; "Hankel approximation": preferred method with far-field approximation of the Hankel function; 1/r-decay: reference curve.

the source is set to 0.8 m with a measurement distance of 0.5 m. The used software includes

- Microsoft Windows 7 64bit
- Microsoft Visual Studio 2010 SP1,
- Microsoft C++ Compiler XE Version 12
- IPP Version 7.0.1,
- Boost Version 1.46 (Legendre/Hankel function).

4.2. Runtimes

To compare the optimized methods, full filter generation times where measured for each method.
Figure 12 shows the filter calculation time with different range extrapolation methods over different transformation order and a common filter length of 256 samples. Here, a quadratic dependency with the order based from the number of coefficients is clear in all methods. To show the calculation time improvement from the Hankel function approximation during the range extrapolation, two bounding conditions are shown. Depending on the distance to the listener, the Hankel approximation calculation times vary: For filters with a receiver closely positioned to the source, the calculation time is bounded by the pure Hankel calculation time as the function is in the nearfield of the head even for high frequencies. For very large distances, the time is bounded by the 1/r-decay function as most frequency points are in the far-field. Figure 12 shows that even with a relatively small distance to the source, the improvement from the approximation is still approximately a factor of two.

With the approximation, a complete filter generation takes approximatly 100 μs for a filter length of 256 samples and order 20. This stands in contrast with approximately 8 μs that is needed for a filter generation from a discrete data set. Figure 13 shows each components share in the overall filter generation time. Each component takes up approximately 33 % of the generation time and therefore no significant bottleneck in the filter generation can be found. To further increase the transformation speed, several approaches are possible. First, the implementation

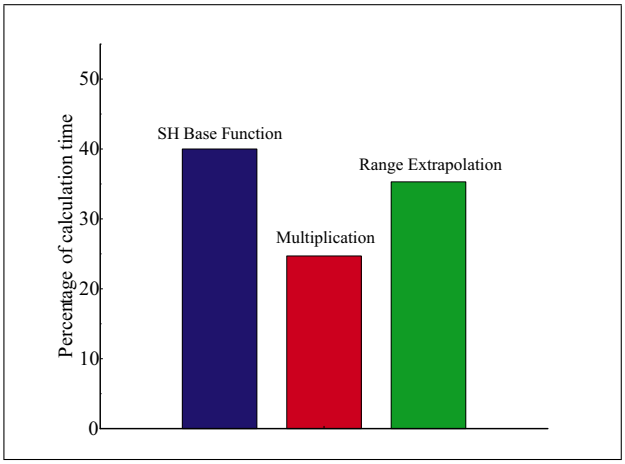

Figure 13. Percentage of each filter generation component onto the whole calculation time.

of the SH base function and Hankel function generation could be optimized for speed. Furthermore, a LUT for the SH base functions will increase performance by approximately 30 %. Here, the filter directions would no longer be spatially continuous as they would be limited to the precomputed directions.

5. Conclusion

In this paper an efficient implementation of HRTF data sets using the spherical harmonic transformation is introduced. The SH model allows representation even between sampling points without additional interpolation. It is shown that with the use of range extrapolation, a HRTF representation on any distance from one measurement is possible. This allows to reduce the measurement effort significantly, as only one distance has to be measured per direction. A more compact file size is possible as less information has to be stored. This leads to a reduction of file sizes from approximately 100 MB for discrete sets to sizes of, for ex-

ample, about 1 MB for filter length of 256 samples and maximum SH order of 20.

The downsides of this model have been studied and presented in detail. The most notable downsides are the increased filter reconstruction times and the introduced errors from truncation and spatial aliasing. As the increased filter reconstruction time will increase the latency of the auralization system, the goal of the efficient implementation is to achieve smallest possible reconstruction times while remaining as exact as possible. Several steps, both during the creation of the SH coefficients and the filter reconstruction, have been taken into account to achieve these goals.

First, the reconstruction error of the HRTF with respect to the SH order was analyzed. Here the influence of the center point of the transformation was examined. It was shown that the reconstruction error is dependent on the center point of the spherical harmonic transformation. An optimized center point was calculated that minimizes both aliasing and truncation errors and achieves a transformation with an reduced reconstruction error.

This reduction prooved to be a first important step in achieving a real-time capable implementation, as less coefficients where needed to achieve a result with comparable reconstruction error. With this optimized generation of the SH coefficients, several steps where taken during every calculation step of the filter reconstruction to further improve the transformation time.

The generation time of the SH base functions was improved using symmetries in the coefficients. The complex valued multiplication could be accelerated using libraries that use vectorization of values and thus computing multiple values simultaneously. These improvements do not impact the calculation's accuracy and exact results are still reproduced. The biggest speed improvement could be achieved during the range extrapolation method. A set of Hankel function values are precomputed in a look-up table. Additionally a far-field approximation of the Hankel function is used which, dependent on the distance was used further improve the range extrapolation time.

The increased computational complexity introduced by the SHT increased the filter generation times despite the improvements. A typical filter generation takes approximately 100 μs for a filter length of 256 samples and order 20. Compared to a discrete data set that takes approximately 8 μs the SH approach still requires more computational effort, but the achieved filter generation times are still a factor of over 1000 smaller than the real-time auralization limit of 20–30 ms. While this is true for auralization with one source, the increased complexity will prove to be problematic if many sources are auralized. In room acoustics simulations, where reflected sound from one source will come from multiple directions, this method should not be used as it will not audibly improve the simulation while increasing the simulation time.

5.1. Outlook

The most interesting questions that comes from this work can be answered using intensive listening tests. The biggest question remaining is to quantify the maximum order needed for a good localization with different signals. As shown, even with optimized center points, the reconstruction error grows substantially to higher frequencies and lower orders. An interesting question would be, how significant these errors are to the localization.

More listening tests could be done to study the influence of small head movements on the localization. If subjects are unsure about the exact location of a sound event, it was shown that small head movements are used to pinpoint the source. As the spatial sampled HRTF grids are limited by design in their spatial resolution, these very small head movements could not be auralized correctly. With a sufficiently high maximum SH order however, this should be possible and thus should allow better localization especially on the cones of confusion.

References

[1] J. Blauert: Spatial hearing: The psychophysics of human sound localization. 2nd ed. MIT Press Cambridge, 1997.

[2] F. L. Wightman, D. J. Kistler: Headphone simulation of free-field listening. I: Stimulus synthesis. The Journal of the Acoustical Society of America **85** (Feb. 1989) 858–867.

[3] B. Gardner, K. Martin: HRTF measurements of a KEMAR. The Journal of the Acoustical Society of America **97** (1995) 3907–3908.

[4] D. N. Zotkin, R. Duraiswami, E. Grassi: Fast head-related transfer function measurement via reciprocity. The Journal of the Acoustical Society of America **120** (2006) 2202–2215.

[5] P. Majdak, P. Balazs, B. Laback: Multiple exponential sweep method for fast measurement of head-related transfer functions. J. Audio Eng. Soc **55** (2007) 623—-637.

[6] M. Pollow, B. Masiero, P. Dietrich, J. Fels, M. Vorländer: Fast measurement system for spatially continuous individual HRTFs. 4th Int. Symposium on Ambisonics and Spherical Acoustics, University of York, UK, 2012.

[7] B. F. G. Katz: Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. The Journal of the Acoustical Society of America **110** (2001) 2440–2448.

[8] J. Fels, P. Buthmann, M. Vorlander: Head-related transfer functions of children. Acta Acustica united with Acustica **90** (2004) 918–927.

[9] R. Paulsen: Statistical shape analysis of the human ear canal with application to in-the-ear hearing aid design. Dissertation. Technical University of Denmark, 2004.

[10] T. Walsh, L. Demkowicz, R. Charles: Boundary element modeling of the external human auditory system. The Journal of the Acoustical Society of America **115** (2004) 1033–1043.

[11] C. L. Searle, L. D. Braida, M. F. Davis, H. S. Colburn: Model for auditory localization. The Journal of the Acoustical Society of America **60** (Nov. 1976) 1164–1175.

[12] C. Cheng, G. Wakefield: Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. Audio Engineering Society Convention 107 (1999).

[13] K. Hartung, J. Braasch, S. Sterbing: Comparison of different methods for the interpolation of head-related transfer functions. Journal of the 16th Audio Engineering Society (2012).

[14] P. Minnaar, J. Plogsties, F. Christensen: Directional resolution of head-related transfer functions required in binaural synthesis. J. Audio Eng. Soc (2005) 919–929.

[15] W. R. Thurlow, P. S. Runge: Effect of induced head movements on localization of direction of sounds. The Journal of the Acoustical Society of America **42** (1967) 480–488.

[16] L. B. W. Jongkees, R. A. Veer: On directional sound localization in unilateral deafness and its explanation. Acta Oto-Laryngologica (1958).

[17] T. Lentz: Binaural technology for virtual reality. Dissertation. RWTH Aachen University, 2008.

[18] D. S. Brungart, W. M. Rabinowitz: Auditory localization of nearby sources. Head-related transfer functions. The Journal of the Acoustical Society of America **106** (Sep. 1999) 1465–1479.

[19] T. Lentz, D. Schröder, M. Vorländer, I. Assenmacher: Virtual Reality System with Integrated Sound Field Simulation and Reproduction. EURASIP Journal on Advances in Signal Processing (2007).

[20] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, D. N. Zotkin: Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. The Journal of the Acoustical Society of America **127** (Jan. 2010) 370–86.

[21] W. Zhang: Efficient continuous HRTF model using data independent basis functions: Experimentally guided approach. Audio, Speech, and Language Processing, IEEE Transactions on **17** (2009) 819–829.

[22] M. Evans, J. Angus, A. Tew: Analyzing head-related transfer function measurements using surface spherical harmonics. The Journal of the Acoustical Society of America **104** (1998) 2400–2411.

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

[24] R. Duraiswami, D. N. Zotkin, N. A. Gumerov: Interpolation and range extrapolation of HRTFs. IEEE International Conference on Acoustics, Speech, and Signal Processing (2004).

[25] I. Ben Hagai, M. Pollow, M. Vorländer, B. Rafaely: Acoustic centering of sources measured by surrounding spherical microphone arrays. The Journal of the Acoustical Society of America **130** (Okt. 2011) 2003–2015.

[26] A. Schmitz: Ein neues digitales Kunstkopfmeßsystem. Acustica **81** (1995) 416–420.

[27] F. Zotter: Sampling strategies for acoustic holography / holophony on the sphere. NAG-DAGA, Rotterdam (2009).

[28] B. Masiero, M. Pollow, J. Fels: Design of a Fast Broadband Individual Head-Related Transfer Function Measurement System. Proceedings of Forum Acusticum 2011, Aalborg, Denmark, 2011, ACTA ACUSTICA united with ACUSTICA, Hirzel, 136.

[29] B. Rafaely, B. Weiss, E. Bachmat: Spatial aliasing in spherical microphone arrays. Signal Processing, IEEE Transactions on (2007).

[30] D. N. Zotkin, R. Duraiswami, N. A. Gumerov: Regularized HRTF fitting using spherical harmonics. Applications of Signal Processing to Audio and Acoustics (2009).

[31] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, M. Noisternig: Calculation of Head-Related Transfer Functions for Arbitrary Field Points Using Spherical Harmonics Decomposition. Acta Acustica united with Acustica **98** (Jan. 2012) 72–82.

[32] J. Vanderkooy: The acoustic center: A new concept for loudspeakers at low frequencies. Audio Engineering Society Convention 121, 2006.

[33] D. Schröder, F. Wefers, S. Pelzer: Virtual Reality System at RWTH Aachen University. Proceedings of the Inernational Symposium on Room Acousics (2010).

[34] A. Kohlrausch, S. van de Par, R. van Eijk, J. F. Juola: Human performance in detecting audio-visual asynchrony. Journal of the Acoustic Society of America **120** (2006) 3084–3085.

[35] M. Vorländer: Auralization: Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. Springer-Verlag, 2007.

[36] C. Borß: An Improved Parametric Model for the Design of Virtual Acoustics and its Applications. Dissertation. 2011.