HRTF MAGNITUDE MODELING USING A NON-REGULARIZED LEAST-SQUARES FIT OF SPHERICAL HARMONICS COEFFICIENTS ON INCOMPLETE DATA

Jens Ahrens, Mark R. P. Thomas, Ivan Tashev

Microsoft Research, One Microsoft Way, Redmond WA 98052, USA
{jeahrens, markth, ivantash}@microsoft.com

ABSTRACT

Head-related transfer functions (HRTFs) represent the acoustic transfer function from a sound source at a given location to the ear drums of a human. They are typically measured from discrete source positions at a constant distance. Spherical harmonics decompositions have been shown to provide a flexible representation of HRTFs. Practical constraints often prevent the retrieval of measurement data from certain directions, a circumstance that complicates the decomposition of the measured data into spherical harmonics. A least-squares fit of coefficients is a potential approach to determining the coefficients of incomplete data. However, a straightforward non-regularized fit tends to give unrealistic estimates for the region were no measurement data is available. Recently, a regularized least-squares fit was proposed, which yields well-behaved results for the unknown region at the expense of reducing the accuracy of the data representation in the known region. In this paper, we propose using a lower-order non-regularized least-squares fit to achieve a well-behaved estimation of the unknown data. This data then allows for a high-order non-regularized least-squares fit over the entire sphere. We compare the properties of all three approaches applied to modeling the magnitudes of the HRTFs measured from a manikin. The proposed approach reduces the normalized mean-square error by approximately 7 dB in the known region and 11 dB in the unknown region compared to the regularized fit.

***Index Terms*—** head-related transfer functions, spherical harmonics, interpolation, extrapolation, regularization

1. INTRODUCTION

Head-related transfer functions (HRTFs) represent the transfer function from a given sound source position to the ear drums of a human [1]. They can be measured either from a specific grid of discrete positions or along a set of continuous trajectories [2], all typically located on the surface of a notional sphere centered around the midpoint between the ears of a person in an anechoic environment. This paper assumes a discrete grid although the results are directly applicable to continuous trajectories as well.

A direct practical application of HRTFs is the creation of virtual acoustic environments. In this case, the input signal of a sound source is filtered with the HRTF corresponding to a given desired source position. When the resulting ear signals are presented to a listener – typically using headphones – the sensation of a sound source in the desired position is evoked.

Modeling magnitude and phase individually may be assumed to be the most favorable approach [3]. Note though that we consider exclusively the magnitudes of HRTFs in this paper. For convenience, we use the term HRTFs in the remainder even if we refer to the magnitudes only.

Spherical harmonics have been shown to constitute a powerful and flexible basis for the representation of HRTFs since 1) the coefficients are discrete and can therefore be stored conveniently, and 2) the HRTFs for any arbitrary location on the sphere can be retrieved from the coefficients [4]. A frequently faced inconvenience is the fact that practical constraints can prevent the measurement of HRTFs on certain portions of the sphere, typically underneath the subject. The estimation of such unknown data is occasionally termed extrapolation although it is rather an interpolation over missing data.

The coefficients of the spherical harmonics expansion of a given function on a sphere can be derived analytically via an integral over the surface of the notional sphere provided that the data is known anywhere [5]. In the case of either discrete sampling or continuous measurements it is possible to probe an enclosing surface so that the coefficients can only be obtained up to a given expansion order, which restricts the representable amount of detail of the HRTFs. The transformation integral can still be evaluated discretely in this case assuming that the energy of the data above the highest obtainable order is so low that spatial aliasing can be neglected [6].

As mentioned above, it is often not possible to measure locations underneath the subject so that the data is unknown along some portion of the sphere. The transformation integral cannot be used to obtain the corresponding coefficients since the data along the unknown portion is inherently set to zero. Instead a least-squares fit can be applied, which yields a set of coefficients that best describe the known data in a least-squares sense. Unfortunately, reconstructing the data along the unknown region from the resulting coefficients leads to unrealistically high amplitudes, potentially due to measurement noise as well as spatial aliasing. It has been proposed to employ a regularized fit to overcome this blow-up of the data [8]. The drawback is that the regularization can affect the known data. In this paper, we propose a lower-order extrapolation of the measurement data into the unknown region that then supports a non-regularized least-squares fit over the entire sphere. This way we can avoid affecting the representation of the known data while retrieving a well-behaved representation of the HRTF data over the entire sphere.

An iterative approach for the estimation of unknown HRTF data has been presented in [9]. This approach also leads to a blow-up of the data in the presence of measurement noise.

2. MATHEMATICAL PRELIMINARIES

A function $\mathcal{H}(\beta, \alpha, \omega)$ that is square integrable on the surface of a sphere that is centered around the coordinate origin can be represented by the coefficients $\breve{\mathcal{H}}_n^m(\omega)$ of a series of spherical harmonics

$Y_n^m(\beta,\alpha)$ as [5]

$$\mathcal{H}(\beta,\alpha,\omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\breve{\mathcal{H}}_n^m(\omega)Y_n^m(\beta,\alpha)\;. \tag{1}$$

In the present case, we may apply the Helmholtz reciprocity principle and assume that we observe the sound pressure of a sound source in the ear of the subject on the surface of a sphere [8].

The surface spherical harmonics $Y_n^m(\beta,\alpha)$ are a complete and orthonormal set and may be defined as

$$Y_n^m(\beta,\alpha)=(-1)^m\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\beta)e^{\imath m\alpha}\;, \tag{2}$$

where $P_n^m(\cdot)$ denotes mth-order the associated Legendre function of n-th degree, α denotes the azimuth and β the colatitude.

Eq. (1) converges above a certain threshold so that it can be approximated by [5]

$$\mathcal{H}(\beta,\alpha,\omega)\approx\sum_{n=0}^{N}\sum_{m=-n}^{n}\breve{\mathcal{H}}_n^m(\omega)Y_n^m(\beta,\alpha)\;. \tag{3}$$

The coefficients $\breve{\mathcal{H}}_n^m(\omega)$ can be obtained via

$$\breve{\mathcal{H}}_n^m(\omega)=\int_{S^2}\mathcal{H}(\Omega,\omega)Y_n^{-m}(\Omega)\,dA(\Omega)\;, \tag{4}$$

where $dA(\Omega)$ is an infinitessimal surface element on the unit sphere S^2 and Ω a given point on the sphere defined by (β,α).

3. REGULARIZED LEAST-SQUARES FIT

As discussed in Sec. 1, (4) cannot be used to determine the coefficients $\mathcal{H}_n^m(\omega)$ because $\mathcal{H}(\Omega,\omega)$ is usually unknown in some portion of the sphere. An alternative approach is formulating the problem as a system of linear equations for which a least-squares solution can be obtained [8]:

$$\mathbf{Y}\breve{\boldsymbol{\mathcal{H}}}=\boldsymbol{\mathcal{H}}\;, \tag{5}$$

where $\mathbf{Y}$ is an $L\times(N+1)^2$ matrix of basis functions $Y_n^m(\beta,\alpha)$ up to a given order N as

$$\mathbf{Y}=\begin{bmatrix} Y_0^0(\Omega_1) & Y_1^{-1}(\Omega_1) & Y_1^0(\Omega_1) & \dots & Y_N^N(\Omega_1)\\ Y_0^0(\Omega_2) & \ddots & \ddots & \ddots & Y_N^N(\Omega_2)\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ Y_0^0(\Omega_L) & Y_1^{-1}(\Omega_L) & Y_1^0(\Omega_L) & \dots & Y_N^N(\Omega_L)\end{bmatrix}, \tag{6}$$

L denotes the number of measurement points. $\breve{\boldsymbol{\mathcal{H}}}$ is a $(N+1)^2\times 1$ vector of coefficients as

$$\breve{\boldsymbol{\mathcal{H}}}=\begin{bmatrix}\breve{\mathcal{H}}_0^0(\omega)\\ \breve{\mathcal{H}}_1^{-1}(\omega)\\ \breve{\mathcal{H}}_1^0(\omega)\\ \breve{\mathcal{H}}_1^1(\omega)\\ \vdots\\ \breve{\mathcal{H}}_N^N(\omega)\end{bmatrix}, \tag{7}$$

and $\boldsymbol{\mathcal{H}}$ is a $L\times 1$ vector of observations $\mathcal{H}(\Omega_l,\omega)\forall 1\le l\le L$. Eq. (5) can be solved of each frequency bin and for each ear as

$$\breve{\boldsymbol{\mathcal{H}}}=\mathbf{Y}^{\dagger}\boldsymbol{\mathcal{H}}\;, \tag{8}$$

where $\mathbf{Y}^{\dagger}$ denotes the pseudo inverse of $\mathbf{Y}$. In this paper, we compute the Moore-Penrose pseudo inverse, which yields the least-squares solution to (5) [10].

Regularization can be applied in order to give preference to a particular solution to (5) with given properties [10]. We apply Tikhonov regularization as a representative approach. It has been reported in [8] that the results for different regularization methods are similar in the present context.

While the least-squares solution minimizes the energy of the residual as

$$\underset{\breve{\boldsymbol{\mathcal{H}}}}{\operatorname{argmin}}\|\mathbf{Y}\breve{\boldsymbol{\mathcal{H}}}-\boldsymbol{\mathcal{H}}\|_2\;, \tag{9}$$

where $\|\cdot\|_2$ denotes the ℓ_2 vector norm, Tikhonov regularization minimizes

$$\underset{\breve{\boldsymbol{\mathcal{H}}}}{\operatorname{argmin}}\left(\|\mathbf{Y}\breve{\boldsymbol{\mathcal{H}}}-\boldsymbol{\mathcal{H}}\|_2+\lambda\|\mathbf{D}\breve{\boldsymbol{\mathcal{H}}}\|_2\right)\;, \tag{10}$$

where the scalar λ is the regularization parameter that has to be chosen suitably and $\mathbf{D}$ is a diagonal regularization matrix. We use $\mathbf{D}$ as proposed in [11] with its diagonal elements $d_{l,l}$ being

$$d_{l,l}=1+n(n+1)\;, \tag{11}$$

where n denotes the degree of the corresponding basis function Y_n^m. Choosing $\mathbf{D}$ according to (11) penalizes high-degree harmonics more than low-degree ones in order to increase smoothness of the result. The explicit solution to (10) is [10]

$$\breve{\boldsymbol{\mathcal{H}}}=(\mathbf{Y}^T\mathbf{Y}+\lambda\mathbf{D})^{-1}\mathbf{Y}^T\boldsymbol{\mathcal{H}}\;. \tag{12}$$

4. NON-REGULARIZED LEAST-SQUARES FIT USING LOWER ORDER EXTRAPOLATION

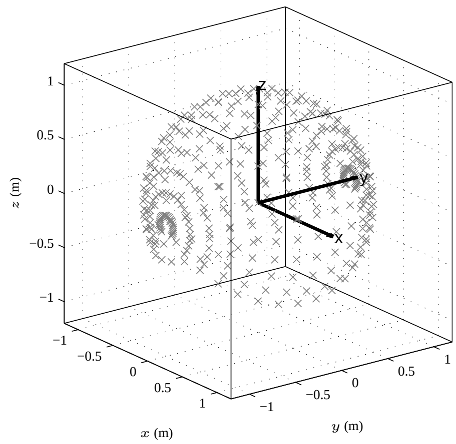
In order to avoid an increase of the modeling error in the known region due to regularization and avoid a blow up in the unknown region, we propose a three-stage approach:

1. Perform a non-regularized low-order least-squares fit according to (8). We have empirically determined $N=3$ as a useful choice. The problem is well-behaved.
2. Use the obtained coefficients to obtain HRTF estimates at a sufficient number of support points in the unknown region. We chose to virtually complement the measurement grid.
3. Perform a non-regularized least-squares fit according to (8) over the entire sphere using the measured data in the known region and the lower-order estimation in the unknown region. The problem is then well-behaved.

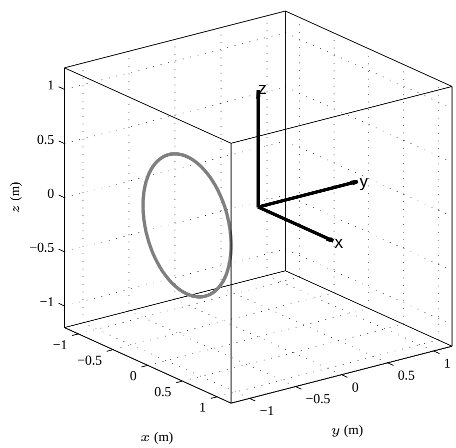
The properties of the presented approach are discussed in Sec. 5 based on sample measurement data.

5. RESULTS

The measurement grid considered in the following is depicted in Fig. 1(a). The physical setup consists of a semicircular arc of radius 1 m equipped with 16 loudspeakers placed equiangularly at intervals of 11.25°. The pivot points of the arc as well as the interaural axis of the subject lie on the y-axis with the midpoint between the ears in the coordinate center. The arc can rotate to 25 different angles in steps of 11.25° between 135° colatitude in the frontal hemisphere and 135° colatitude in the rear hemisphere, which results in $25\cdot 16=400$



(a) Incomplete measurement grid



(b) The great circle of 360 evenly spaced sampling points that are illustrated in Fig. 2.

Fig. 1. Geometric setup; The HATS was placed in the coordinate center and 'looked' along the positive x-axis.

measurement points for each ear of the B&K Head and Torso Simulator (HATS). This measurement grid is referred to as *incomplete*. The ground truth of the data in the unknown region was obtained by mounting the HATS upside down in the rig. The resulting measurement grid covering the entire sphere is referred to as *complete* in the remainder. Note that human HRTFs can deviate significantly from the HATS's HRTFs especially for positions underneath since the HATS does not have a lower body. The obtained data is nevertheless assumed to be qualitatively similar to human HRTFs.

The measurement grid allows for a maximum N of 15. Representing the physical information contained in the HRTFs requires dozens of orders and hence thousands of measurement points [7]. The model will therefore suffer from spatial aliasing [6]. It not clear, however, which order N is required to provide a representation that is perceptually indistinguishable from the true data. Informal listening by the authors of this paper suggests that a 15th-order reconstruction can indeed not be distinguished from the original HRTF.

Note that HRTF data is three-dimensional (frequency, azimuth, colatitude), which makes it inconvenient to illustrate. We will therefore not illustrate the data over the entire sphere but along a given great circle, which is depicted in Fig. 1(b). The properties of the data along this great circle are representative for the entire sphere.

Fig. 2(a) depicts HRTFs reconstructed via (3) from spherical harmonics coefficients that were obtained from a non-regularized least-squares fit according to (5) on data measured on the complete grid at $32 \cdot 16 = 512$ locations. This data is considered the ground truth in the remainder.

Evaluating (5) based on the known data from the setup shown in Fig. 1(a) for the maximum applicable order N results in a badly-conditioned problem. The result is typically still useful in the densely-sampled region but blows up in the unsampled region when the slightest amount of noise is present [8]. This is illustrated in Fig. 2(b). Regularization can avoid the blow up (Fig. 2(c)) at the expense of increasing the error at the known data points. Here, we apply Tikhonov regularization according to (10) with $\lambda =$1e-2.

For reference, Fig. 4 illustrates the match between pure measured data for a sample direction (blue line) and the results of the regularized fit (green line). The deviation is in the order of ± 0.5 dB below 4000 Hz and slightly larger at higher frequencies.

We use

$$\Psi = 10 \log_{10} \frac{\|\hat{\boldsymbol{\mathcal{H}}} - \boldsymbol{\mathcal{H}}\|_2^2}{\|\boldsymbol{\mathcal{H}}\|_2^2} \tag{13}$$

as a global error measure, where $\hat{\boldsymbol{\mathcal{H}}}$ refers to the modeled data, $\boldsymbol{\mathcal{H}}$ refers to the ground truth. Tab. 1 lists Ψ for the different considered methods for the known and unknown region separately.

In order to illustrate the deviation of the data reconstructed from the incomplete measurements to the ground truth, Fig. 3 depicts the ratio $\Delta(\beta_l, \alpha_l)$ of the modeled data $\hat{\mathcal{H}}(\beta_l, \alpha_l)$ under consideration and the ground truth $\mathcal{H}(\beta_l, \alpha_l)$ on a logarithmic scale as

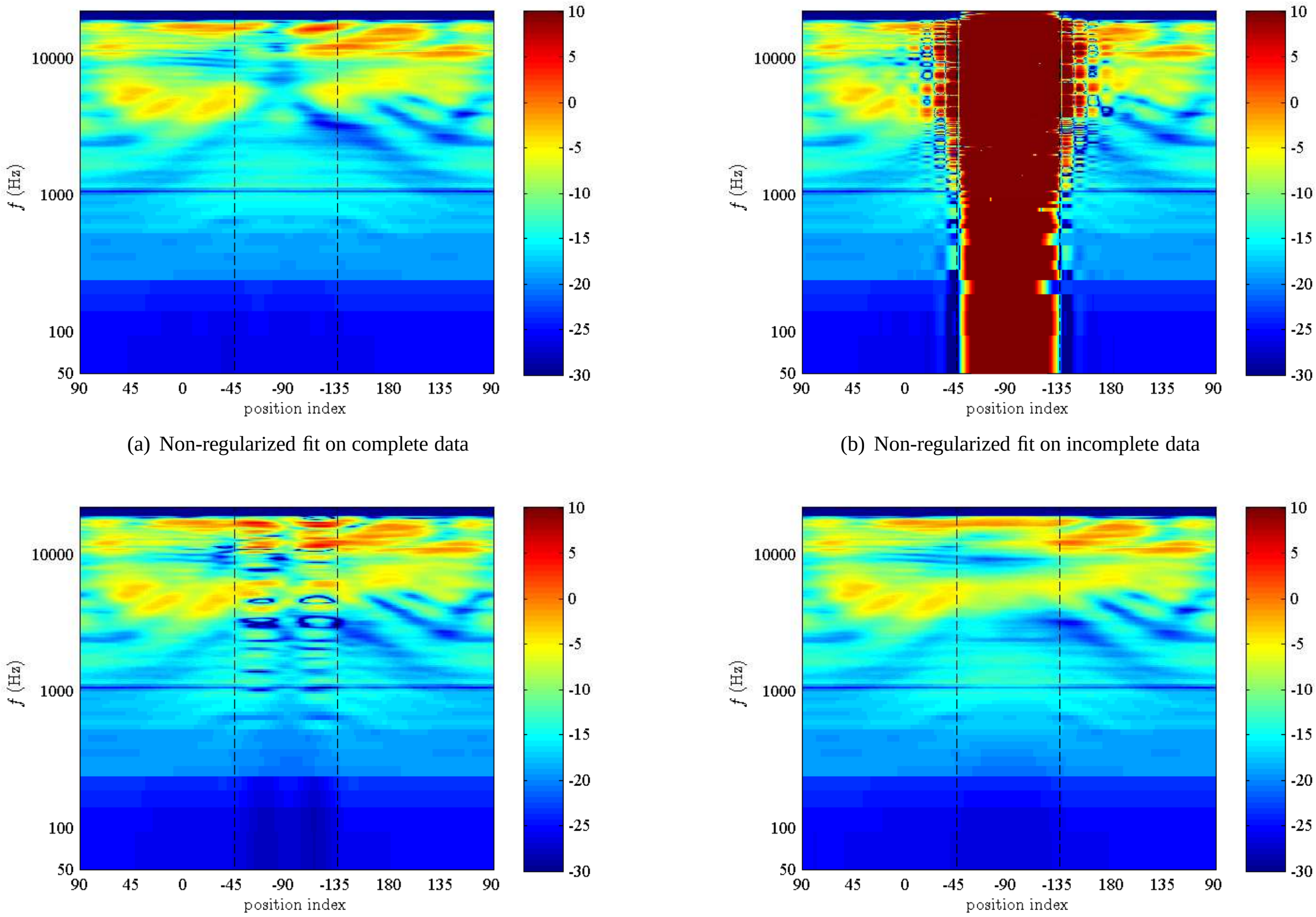
$$\Delta(\beta_l, \alpha_l) = 20 \log_{10} \frac{\hat{\mathcal{H}}(\beta_l, \alpha_l)}{\mathcal{H}(\beta_l, \alpha_l)} \,. \tag{14}$$

(β_l, α_l) denotes the direction of interest. Fig. 3(a) illustrates $\Delta(\beta_l, \alpha_l)$ for the reconstruction via the regularized fit.

Fig. 2(d) depicts the result of the proposed non-regularized least-squares approach based on the data from Fig. 2(a). The match between the measured and modeled data is depicted in Fig. 4 for a sample direction, the deviation $\Delta(\beta_l, \alpha_l)$ from the ground truth is shown in Fig. 3(b), and the error Ψ is listed in Tab. 1.

The apparent benefits over the regularized approach as described in Sec. 3 are:

- The amount of detail of the estimated data in the unknown region is only moderate. Comparing all results for the incomplete data to the ground truth in Fig. 2(a) suggests that a reliable estimate of the unknown data is not possible. The error will usually be significant. Recall Tab. 1 for numbers. A lower-order estimation from the non-regularized coefficients fit assures that no prominent spikes and dips occur in the unknown region, a circumstance that is favorable in perceptual terms. Note that peaks are especially prominent with respect to perception [12]. It can therefore be assured that no harmful data is created. This is not assured by the regularized fit, which does not restrict the order as consequently so that the amount of detail of the data estimates is higher.
- The negative impact on the known data is marginal. Theoretically, the sharp transition from higher-order data to lower-order data at the boundary of the known region requires a very high-order if not infinite-order expansion. An additional small amount of spatial aliasing is consequently added. The non-regularized least-squares fit smoothes this transition by limiting the resulting expansion to order N by the cost of an error increase in the known region by approx. 2 dB (Tab. 1).



(a) Non-regularized fit on complete data

(b) Non-regularized fit on incomplete data

(c) Regularized fit on incomplete data

(d) Proposed approach applied to incomplete data

Fig. 2. HRTF magnitudes in dB along the great circle shown in Fig. 1(b) reconstructed from different types of least-squares fit. Position index 0 represents the corresponding direction in the frontal horizontal plane and positive indices represent the upper hemisphere, negative indices represent the lower hemisphere. The left half of the plot thus represents the frontal hemisphere; the right half represents the rear one. The area between the black vertical lines represents unknown data.

	simple non-reg. fit	regularized fit	proposed approach
known region	-20.67 dB	-11.41 dB	-18.60 dB
unknown region	132.08 dB	8.93 dB	-2.56 dB

Table 1. Error Ψ as defined in (13)

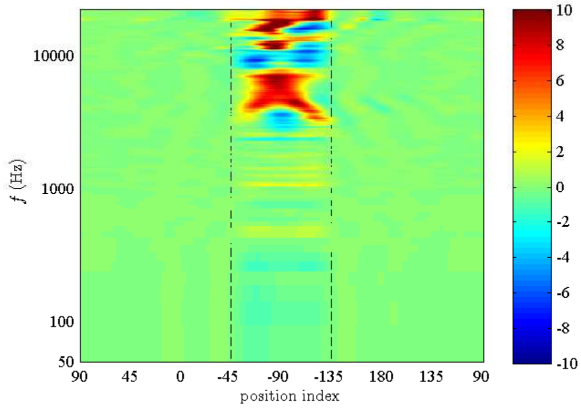
6. FURTHER OBSERVATIONS

As mentioned in Sec. 3, the complete measurement grid theoretically allows for retrieving data for $N = 15$. The least-squares fit will deliver a solution also for $N > 15$. The non-regularized fit – be it on the incomplete data or on the full-sphere data – will make sure that the data at the measurement locations is well represented. However, the reconstructed data tends to blow up between the measurement points and the model is therefore not useful in general. A regularized fit avoids the blow-up but represents the measurement data less well. Additionally, the estimation of the unknown data gets more and more spiky the higher N. As mentioned above, this circumstance may be considered unfavorable from a perceptual point of view.

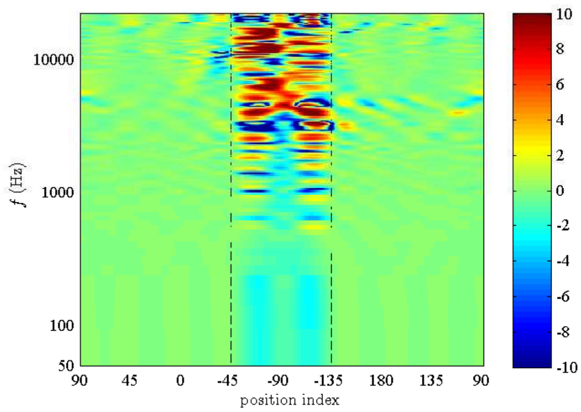
The tendency of blowing up between support points is also apparent in Fig. 2(b) where the modeled data shows high amplitudes even outside the unknown region. The measurement points themselves are properly modeled.

7. CONCLUSIONS

We have presented a three-stage approach for non-regularized least-squares fitting of spherical harmonics coefficients on incomplete



(a) Regularized fit on incomplete data



(b) Proposed approach applied to incomplete data

Fig. 3. Deviation $\Delta(\beta_l, \alpha_l)$ of the data reconstructed from the incomplete measurements to the ground truth as defined in (14).

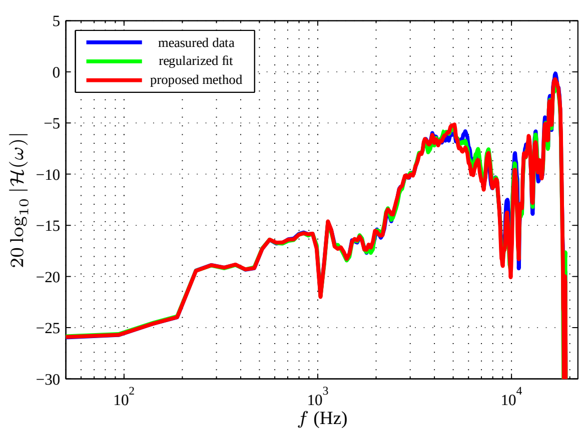


Fig. 4. Comparison of measured and modeled HRTF magnitudes on a logarithmic scale for a representative sample direction.

HRTF magnitude data. The three stages are: 1) Perform a lower-order least-squares fit, 2) calculate a sufficient number of support points in the unknown region from the lower-order data, and 3) perform a higher-order least-squares fit on the full-sphere data set. Unlike with previously published methods, the negative impact of the presented approach on the modeled data in the known region is minimal while obtaining a well-behaved model for the unknown region.

This investigation suggests that it is not possible to reconstruct the unknown data from the known one. However, the data that the presented approach estimates for the unknown region is at least not harmful. This is a consequence of the fact that the created data is of low order and its amount of detail is limited so that no prominent peaks or dips arise. There are indications that such smoothly varying data is perceptually preferable. Previous approaches do not guarantee such smoothness.

8. REFERENCES

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

[2] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2009, IEEE, pp. 325–328.

[3] J. Zaar, "Phase unwrapping for spherical interpolation of head-related transfer functions," Diploma thesis, University of Music and Performing Arts, Graz, 2011.

[4] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2400–2411, 2010.

[5] N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*, Elsevier, Oxford, 2004.

[6] J. R. Driscoll and D. M. Healy, "Computing fourier transform and convolutions on the 2-sphere," *Advances in Applied Mathematics*, vol. 15, pp. 202–250, 1994.

[7] W. Zhang, T. D. Abhayapala, and R. Kennedy, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 257–260, Apr. 2010.

[8] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized HRTF fitting using spherical harmonics," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2009, IEEE, pp. 257–260.

[9] W. Zhang, R. Kennedy, and T. D. Abhayapala, "Iterative extrapolation algorithm for data reconstruction over sphere," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, NV, USA, Apr. 2008, IEEE, pp. 3733–3736.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[11] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004, vol. 4, pp. iv–45–iv–48.

[12] D. Moulton, "Golden ears," Moulton Laboratories, 1992.