Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 1pPPa: Binaural Hearing and Binaural Techniques I

1pPPa2. A study of spherical harmonics interpolation for HRTF exchange

Matthieu Aussal*, François Alouges and Brian Katz

***Corresponding author's address: Digital Media Solutions, 45 grande allée du 12 février 1934, Noisiel, 77186, Seine et marne, France, matthieu.aussal@dms-cinema.com**

Today, there exists a growing number of publicly available HRTF datasets, with each set often proposing a unique variation on the spatial discretization measurement grid. These various grids, typically determined by the mechanical system employed, result in datasets which are not directly comparable or exploitable. To alleviate the limitation of incompatible grids and assist in the adaptation of measurements performed on one grid to another, facilitating the inter-exchange of HRTF sets, a fixed radius HRTF interpolation method is proposed. The approach is based on a decomposition of the sound field using spherical harmonics, allowing for a global spatial recomposition. A high spatial density HRTF was used as a test case for evaluating the interpolation method, and a series of measures are employed to quantitatively compare the quality of the interpolation as theoretical laws.

Published by the Acoustical Society of America through the American Institute of Physics

Received 23 Jan 2013; published 2 Jun 2013
Proceedings of Meetings on Acoustics, Vol. 19, 050010 (2013)

INTRODUCTION

Head-Related Transfers Functions (HRTF) are the main acoustical signature of spatial hearing, characterizing firstly interaural information as Interaural Time Difference (ITD) and Interaural Level Difference (ILD), and secondly spectral deformation undergone by a sound wave as it impinges on the listener. In theory, knowledge of those functions offers the ability to synthesize a natural listening experience over headphones, typically called binaural synthesis. In practice, HRTFs are mainly measured in laboratory conditions and each system often has is own unique set-up. Therefore, to allow definition of a universal exchange format, there is a need for automatic spatial interpolation of those functions.

In this study, interpolation algorithms based on spherical harmonic decomposition are the focus. Well known theoretical principles are first introduced before proposing a method for interpolation validation. Finally, theoretical laws are discussed, in agreement with numerical computation, to predict a domain of effectiveness of this type of interpolation.

MATHEMATICAL REMINDERS

As the theoretical background for spherical harmonic interpolation is well introduced in many previous studies, we provide here only several simple mathematical reminders. The curious reader may refer to reference articles as [1, 2, 3, 4, 5].

Theoretical construction of spherical harmonics basis

Spherical harmonics Y_l^m can be introduced as the the trace on the unit sphere S^2 of harmonic homogeneous polynomials of degree l [6]. So let $\forall l \in \mathbb{N}$, $-l \leq m \leq l$, and $-1 < x < 1$, P_l^m the Legendre function associated to Legendre polynomials P_l, defined by:

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x). \quad (1)$$

In spherical coordinates (r, θ, ϕ) with colatitude convention, the family (Y_l^m) defined by:

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos(\phi)) e^{im\theta} \quad (2)$$

form an orthonormal basis of $L^2(S^2)$, which diagonalizes the Laplace Beltrami operator Δ on the sphere S^2:

$$\Delta Y_l^m + l(l+1) Y_l^m = 0. \quad (3)$$

Thus, any function $u \in L^2(S^2)$ can be written as:

$$u(\theta,\phi) = \sum_{l,m} \alpha_l^m Y_l^m(\theta,\phi) \quad (4)$$

where α_l^m are the expansion coefficients.

Theoretical problem resolution

In numerical practice [2], summation is truncated at a fixed value L called the truncation number, in such a way that $\alpha_l^m = 0 \ \forall l > L$. After that, there is a total of $M = (L+1)^2$ terms in spherical decomposition Eq. (4). Then, for a discrete set of measurements $u(\theta,\phi)$ known at N spatial position $\{u_1, ..., u_N\}$ on the sphere S^2, the M unknowns α_l^m are fitted by writing a linear system of equations:

$$\begin{aligned} (u_i)_{i\in[1,N]} = (u(\theta_i,\phi_i))_i &= \left[(\sum_{l,m}^{L} \alpha_l^m Y_l^m(\theta_i,\phi_i) \right]_i \\ &\Updownarrow \\ Y\alpha &= U \end{aligned} \quad (5)$$

where Y is the $N \times M$ left hand side, α the unknowns vector, and U the right hand side.

To smooth solutions and avoid numerical instabilities, a Tikhonov regularization is employed as in [2, 4], carefully chosen for the given problem. In general, this regularization is performed by a minimization of the norm of the standard data or their variations. In our particular case, as α is the weighting vector of spherical harmonics, "low" spatial frequencies have to be emphasize to avoid abrupt variations. The linear system of Eq. (5) then becomes:

$$(Y^T Y + \epsilon D)\alpha = Y^T U \quad (6)$$

with $D_{ij} = (1 + l(l+1))\delta_{ij}$, l the degree of the corresponding spherical harmonic Y_l^m, and $\epsilon \in [0,1]$.

It is important to note that in Eq. (5), the order of truncation L must be chosen in relation to information provided by the right hand side. Without Tikhonov regularization, a high precision given by a high truncation order L will improve the reconstruction of $(u_i)_{i\in[1,N]}$, but with aliasing for u at all others position due to the resolution of an over-determined system. A judicious choice of the value ϵ is required to keep an optimal interpolation accuracy for oversized L. At the same time, L should not be too small, otherwise the resultant spatial average will generate a loss of information by the resolution of an under-determined system.

Application to HRTF interpolation

An HRTF is a complex function in space (ρ,θ,ϕ) and frequency (f). In most cases, HRTFs are measured for a given fixed radius ρ and may be described as a function $H(\theta,\phi,f)$. For each frequency, H is a complex function of $L^2(S^2)$, so spherical harmonic interpolation can be applied, frequency by frequency, to compute HRTF interpolation [2]. Despite the apparent complexity of the algorithm, the main advantage of this approach is the ease of use. It is only necessary to provide the HRTF and corresponding spherical positions of each measurement and the interpolation can be done automatically for any point on the sphere S^2.

Database and Analysis tools

In order to evaluate the effectiveness of the proposed method, a series of analytical validations has been performed to extract an objective domain of validity. In this study, only three parameters are considered:

- number of spatial positions N used to form the linear system (construction points),
- order of truncation L of the spherical harmonics Y_l^m,
- frequency f.

HRTF database

The HRTF of a dummy head was measured in an anechoic chamber. Microphones (DPA 4060) where placed at the entrance of the ear canals (blocked meatus) of a KEMAR mannequin. Three loudspeakers, spaced at 5° intervals, were mounted on a movable arm at a distance of 1.95 m from the center of the head.[7] The mannequin was installed on a rotating turntable and wore a fleece jumper and felt cap in order to provide realistic absorption characteristics.[8] Measurements were made for elevations of –40° to +90°, in 5° steps, with the turntable rotating in 15° increments. Using the three speakers, this resulted in a grid roughly 5°×5°. HRIRs were obtained using a sweepsine excitation signal [9] and recorded at 192 kHz sample rate (RME Fireface 800). In order to obtain a complete full-sphere dataset, the mannequin was then installed in an inverted position and the measurement was repeated. The two datasets were merged by the completion of the upper measurement (–40° to +90°) by the lower one (–90° to –45°). The resulting dataset comprised 1722 positions, represented by black dots in Fig 1.a.

Errors measurements

In this study two signals have to be compared, a reference H_{ref} and a computed one H_{com}, both functions of time and space. We propose as a suitable error measurement the global Mean Square Error (MSE) in frequency domain [10, 11], defined as:

$$E_{\mathbf{x},f}(H_{com},H_{ref}) = 20 * \log 10\left(\frac{||H_{com}(\mathbf{x},f)-H_{ref}(\mathbf{x},f)||_{\mathbf{x},f}}{||H_{ref}(\mathbf{x},f)||_{\mathbf{x},f}}\right) \quad (7)$$

where $\mathbf{x} = (\rho,\theta,\phi)$ is the spatial position, f HRTF temporal frequencies, and $||.||_{\mathbf{x},f}$ is the euclidean norm along $\mathbf{x}$ and f. It is important to note that this norm can be applied along only $\mathbf{x}$ to define the spatial MSE $E_{\mathbf{x}}$ and along only f to define the frequency MSE E_f.

Secondly, in order to meet the assumptions of theoretical considerations detailed in the previous section, two kinds of points have to be considered (as per [1]):

- The first type consists of reconstructed

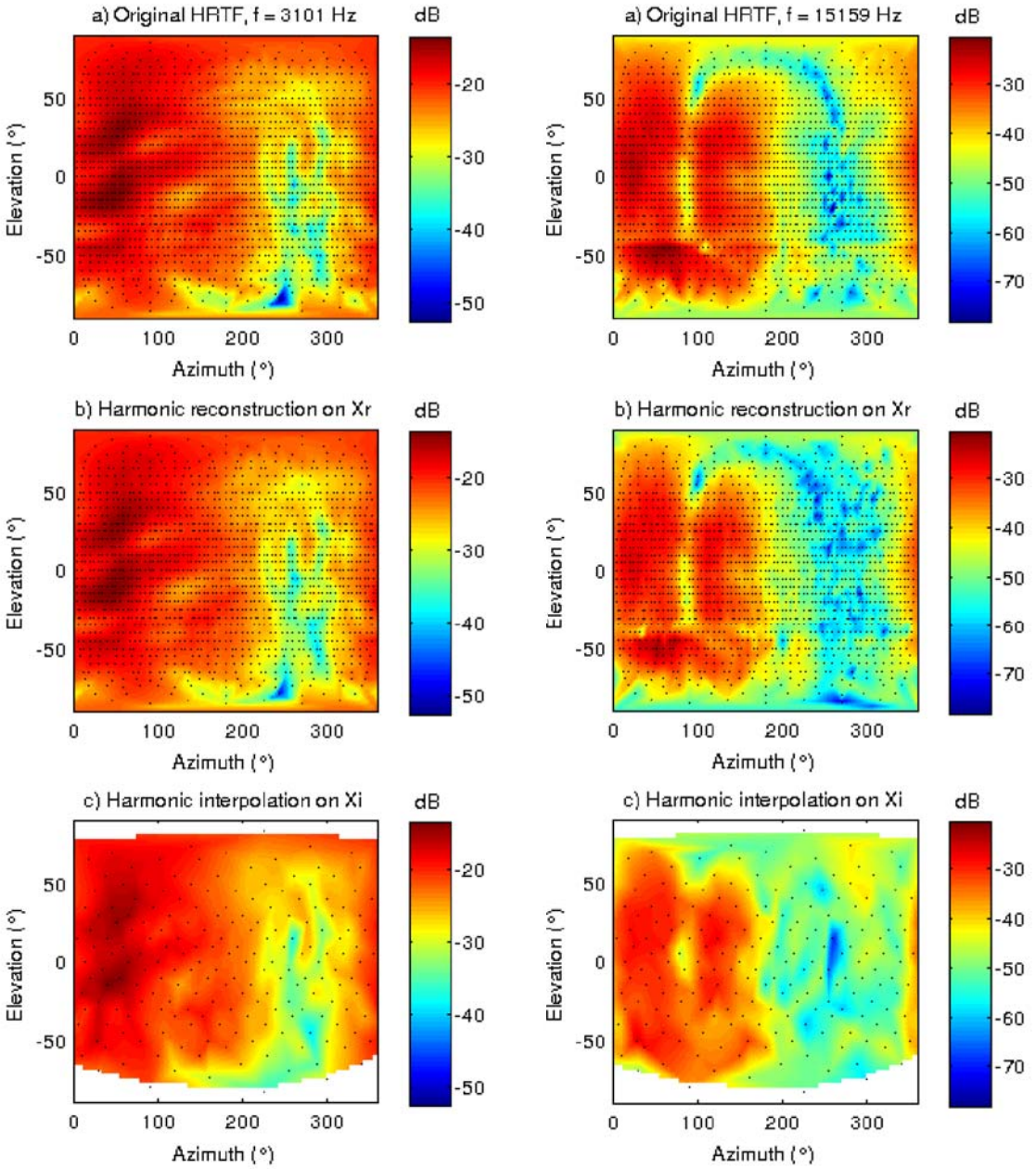

FIGURE 1: Representation of HRTF at fixed parameters N, L, and f for $f = 3101$ Hz (left column) and $f = 15159$ Hz (right column) showing (a) original, (b) reconstructed, and (c) interpolated HRTFs for the KEMAR left ear. Black dots represent the data grid.

points $\mathbf{x}_r$. This set of point they used for the spherical harmonics construction in order to compute the expansion coefficients a_l^m in Eq. (5). It is expected that the measured error of these points always decreases with truncation order L.

- The second type consists of interpolated points $\mathbf{x}_i$, the set of points not used for the computation of a_l^m in Eq. (5). These reflect the effectiveness of the spherical harmonic interpolation.

Error measures $E_{\mathbf{x},f}$, $E_{\mathbf{x}}$, and E_f were calculated for both sets of points, $\mathbf{x}_r$ and $\mathbf{x}_i$.

Meshes coarsening

In order to compute the two kinds of points $\mathbf{x}_r$ and $\mathbf{x}_i$, in order to evaluate the interpola-

tion method, no interpolations are permitted for $\mathbf{x}_r$. Mesh coarsening based on a Fibonacci grid [12] was used to extract $\mathbf{x}_i$ without any preferred spatial direction. To do this, the measurement point closest to the exact grid is extracted from the original mesh to form an approximated uniform grid. The remaining points form the $\mathbf{x}_r$ set. In this study, 200 interpolated points were selected, represented by black dots in Fig 1.c, leaving up to 1522 reconstructed points to be considered for Spherical Harmonic construction (black dots in Fig 1.b). It is important to note that both poles, North and South, are always taken as reconstructed points, in order to keep the full sphere hypothesis, according to theory.

To represent HRTF measurements and errors MSE in space, the sphere S^2 was represented in two dimensions with azimuth and elevation. Triangular linear interpolation was carried out for data plotting using the barycentric mean of the three closest neighbors.

RESULTS AND DISCUSSIONS

In this section, MSE colormap in Fig 2 and Fig 3 is from less than -20 dB in blue to 0 dB in red. The corresponding linear scale is from less than 10% to 100% in the relative sense.

HRTF and error representation

In the next two subsections, all three considered parameters are fixed in order to draw spatial representations. Values are chosen such that there are $N = 1522$ reconstructed points $\mathbf{x}_r$, at frequencies $f \simeq 3000$ and 15000 Hz, with a truncation order $L = 30$ to avoid harmonic averaging. Fig 1.a presents measured left HRTF amplitude on original measurement positions for each frequency. There is more spatial variation for the larger frequency f due to the fact that the HRTF is the result of a scattering problem, so it is expected that spatial variations will increase with frequency. Secondly, following the measurement protocol, the HRTF database has an apparent discontinuity at low elevations, between –40° and –45°, which is more pronounced at higher frequencies.

Fig 1.b represents the left HRTF amplitude values on reconstructed points $\mathbf{x}_r$, used to compute Eq. (5). As HRTFs present greater variability at high frequencies, spatial aliasing appears due to insufficient size of N and L.

Fig 1.c represents the left HRTF amplitude values on interpolated points $\mathbf{x}_i$, computed by spherical harmonic interpolation from previously selected $\mathbf{x}_r$. This small mesh of 200 pts is sufficiently uniform to be objective for spatial and frequency error measurements.

Reconstructed and interpolated HRTF MSE

Fig 2.a shows the Mean Square Error (MSE) on $\mathbf{x}_r$, and Fig 2.b shows the MSE on $\mathbf{x}_i$, for each spatial position for two select frequencies. As expected, both depend on frequency, with a notable difference between accuracy on $\mathbf{x}_i$, poorer for $\mathbf{x}_r$. First, despite the relative metric induced by MSE, there is a large difference on the effectiveness of the algorithm between left and right sound source positions relative to the left ear HRTF. This is probably due to the fact that ipsilateral sources are more scattered by the head than contralateral ones. Secondly, for the two frequencies, errors induced by the full-sphere measurement protocol at –40° elevation are visible.

In order to better understand the frequency dependence of MSE, Fig 2.c provides the spatial MSE $E_{\mathbf{x}}$ as a function of frequency. As expected, the higher the frequency, the greater the spatial averaged error.

Finally, to have a more precise idea of the L dependence of MSE in space and frequency $E_{\mathbf{x},f}$, Fig 2.d shows the convergence of results as a function of order of truncation. $\mathbf{x}_r$ values always converge for higher order, more quickly than for $\mathbf{x}_i$, according to theory. Only interpolation points are considered, because a convergence of $\mathbf{x}_i$ values induces a convergence for $\mathbf{x}_r$, which is not reciprocal (see above).

Theoretical laws and results

For scattering problems there are two theoretical laws providing good error agreement. The first one comes from Fast Multipole liter-

ature [2, 13], and expresses the frequency dependence of the truncation order:

$$L \geq E\left(\frac{2\pi f}{c}a\right)+1 \tag{8}$$

where E is the integer part function, a is the radius of the smallest sphere surrounding the scattering object (here the dummy head), and c the speed of sound.

The second one is a classical law taken from boundary element method (BEM) [6, 14], and expresses the frequency dependence of the mesh used to compute spherical harmonics. For a uniform mesh of N construction points on the sphere S^2:

$$N \geq E\left(4\pi a^2\left(\frac{nf}{c}\right)^2\right) \tag{9}$$

where n is the number of points used to intercept an elementary signal (generally small, $n = 4$ for example).

For these two laws, superiority relation is allowed by the Tikhonov regularization in Eq. (6), which smooth solutions for overdetermined problems. In the equality case, a third law could be extracted from the two previous ones, expressed by the relation $N \approx (L)^2 \approx M$. This relation reflects well the need to solve a square system in Eq. (5).

In Fig 3.a, Fig 3.b, and Fig 3.c, spatial MSE $E_{\mathbf{x}}$ is represented. For each figure, one of the three parameters N, L, and f is fixed. Laws in equality cases are represented by the dark curves, with the first law in Fig 3.a, the second law in Fig 3.b, and both laws in Fig 3.c. In each case, dark curves delineate areas of good and bad numerical interpolations, as expected. Also, these analytical laws give an efficient estimation of the value of N and L needed to compute the chosen frequency.

Conclusion and future works

In summary, a validation method for harmonic interpolation of full-sphere HRTFs was proposed and evaluated. Two theoretical laws have been proposed and empirically validated. For the highest frequency to be reproduced, those laws give a good idea of the size of the mesh N, with the associated truncation order L necessary to maintain good accuracy. For example, to correctly interpolate HRTFs from 20 Hz to 20 kHz, truncation order $L \simeq 37$ for an uniform grid of $N \simeq 7000$ points are needed, which is very difficult to measure in practice. But, in accordance with measurement datasets, a prediction of an objective frequency domain of validity for this kind of interpolation is possible. So, subject to the limitations imposed by databases themselves, the exchange of HRTF databases is possible.

Works on different spherical databases of HRTFs should now be carried out to completely demonstrate the robustness of the presented approach. This study could also be extended to a set of HRTFs measured at different distances in order to evaluate spherical harmonic extrapolation of HRTFs.

ACKNOWLEDGMENTS

This work was performed in the context of a CIFRE industry-linked research funding program with Digital Media Solutions (DMS). The authors would like to thank Gaëtan Parseihian, LIMSI-CNRS, for the HRTF measurement, and IRCAM for use of their measurement facilities, regarding the acquisition of the KEMAR HRTF. Finally, thanks to Omnihead (www.univ-brest.fr/mstis/omnihead/accueil.htm) for the use of the KEMAR dummy head.

References

[1] M. Aussal, F. Alouges, and B. F. Katz, “ITD interpolation and personalization for binaural synthesis using spherical harmonics”, in *Audio Engineering Society UK Conference*, 04_01–10 (York, UK) (2012).

[2] R. Duraiswaini, D. Zotkin, and N. Gumerov, "Interpolation and range extrapolation of hrtfs [head related transfer functions]", in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 4, 45–48 (2004).

[3] A. Laborie, R. Bruno, and S. Montoya, "A new comprehensive approach of surround sound recording", in *Audio Engineering Society Convention 114*, 5717:1–20 (Amsterdam) (2003).

[4] K.-V. Nguyen, T. Carpentier, M. Noisternig, and O. Warusfel, "Calculation of the head related transfer unction in the proximity region using spherical harmonics decomposition: comparison with measurements and evaluation", in *2nd Intl. Symp. on Ambisonics and Spherical Acoustics*, 1–4 (Paris) (2010).

[5] D. Zotkin, R. Duraiswami, and N. Gumerov, "Regularized HRTF fitting using spherical harmonics", in *Applications of Signal Processing to Audio and Acoustics (WASPAA '09)*, 257–260 (2009).

[6] J.-C. Nedelec, *Acoustics and Electromagnetic equations. Integral representations for harmonic problems*, number 144 in Applied Mathematics Science (Springer Verlag, New York) (2001).

[7] "LISTEN HRTF database", (2003), URL `recherche.ircam.fr/equipes/salles/listen`.

[8] B. F. Katz, "Acoustic absorption measurement of human hair and skin within the audible frequency range", J. Acoust. Soc. America **108(5)**, 2238–2242 (2000).

[9] P. Fausti and A. Farina, "Acoustic measurements in opera houses: comparison between different techniques and equipment", J. Sound and Vibration **232**, 213–229 (2000).

[10] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Sound field analysis along a circle and its applications to HRTF interpolation", J. Aeronaut. Sci. **56**, 156–175 (2008).

[11] W. Zhang, P. Abhayapala, and R. Kennedy, "Horizontal plane HRTF reproduction using continuous Fourier-Bessel functions", in *31st International Conference: New Directions in High Resolution Audio*, 4, 1–9 (London) (2007).

[12] R. Swinbank and R. James Purser, "Fibonacci grids: A novel approach to global modelling", Quarterly J. Royal Meteorological Soc. **132**, 1769–1793 (2006).

[13] N. Gumerov, A. O'Donovan, R. Duraiswami, and D. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation", J. Acoust. Soc. America **127**, 1370–386 (2010).

[14] B. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation", J. Acoust. Soc. America **110**, 2440–2448 (2001).

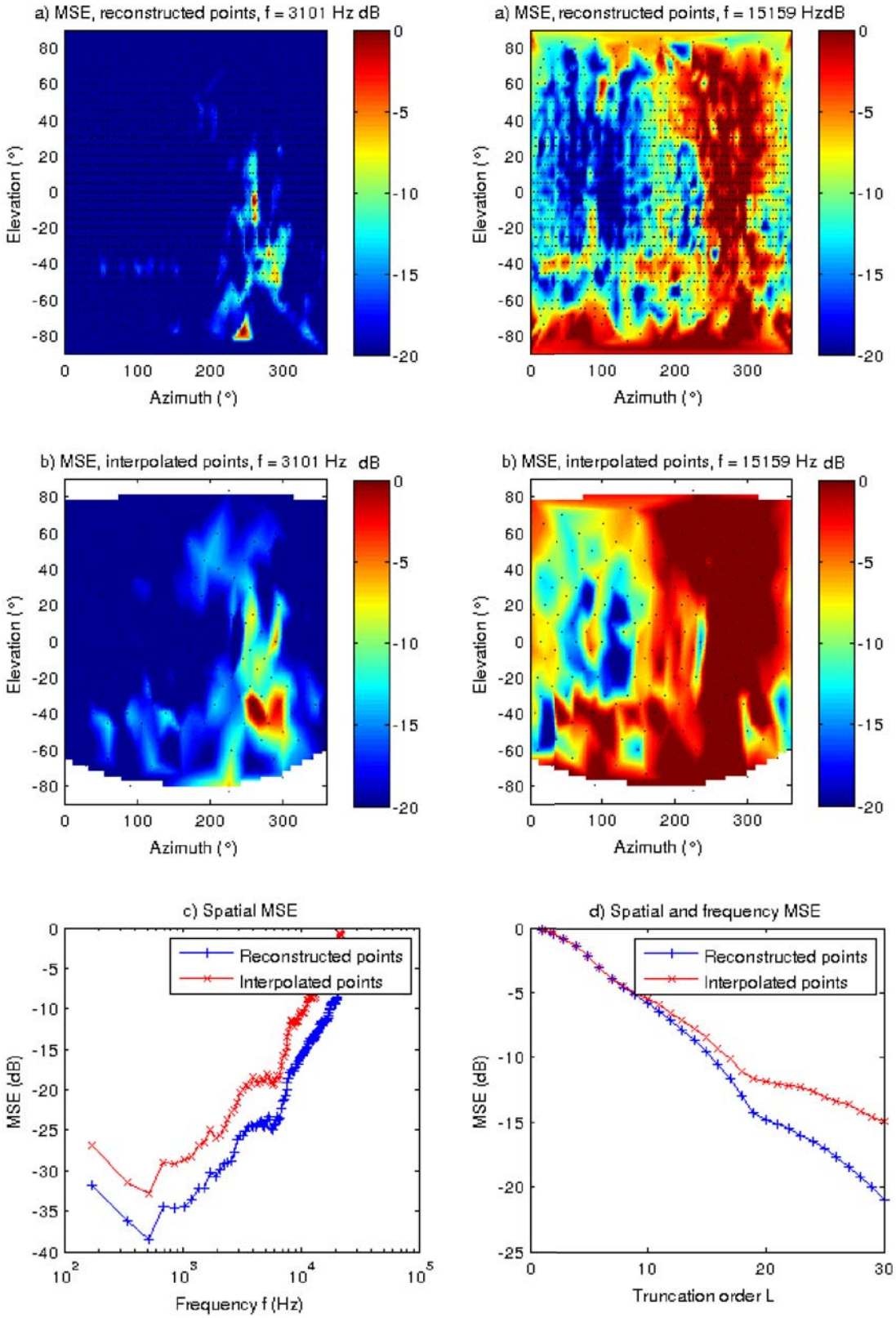

FIGURE 2: Representation of Mean Square Errors (MSE) at fixed parameters N, L, and f for $f = 3101$ Hz (left) and $f = 15159$ Hz (right) for the KEMAR left ear HRTF showing (a) reconstructed and (b) interpolated points. Black dots represent the data grid. (c) Spatial and (d) spatial-frequency mean MSE.

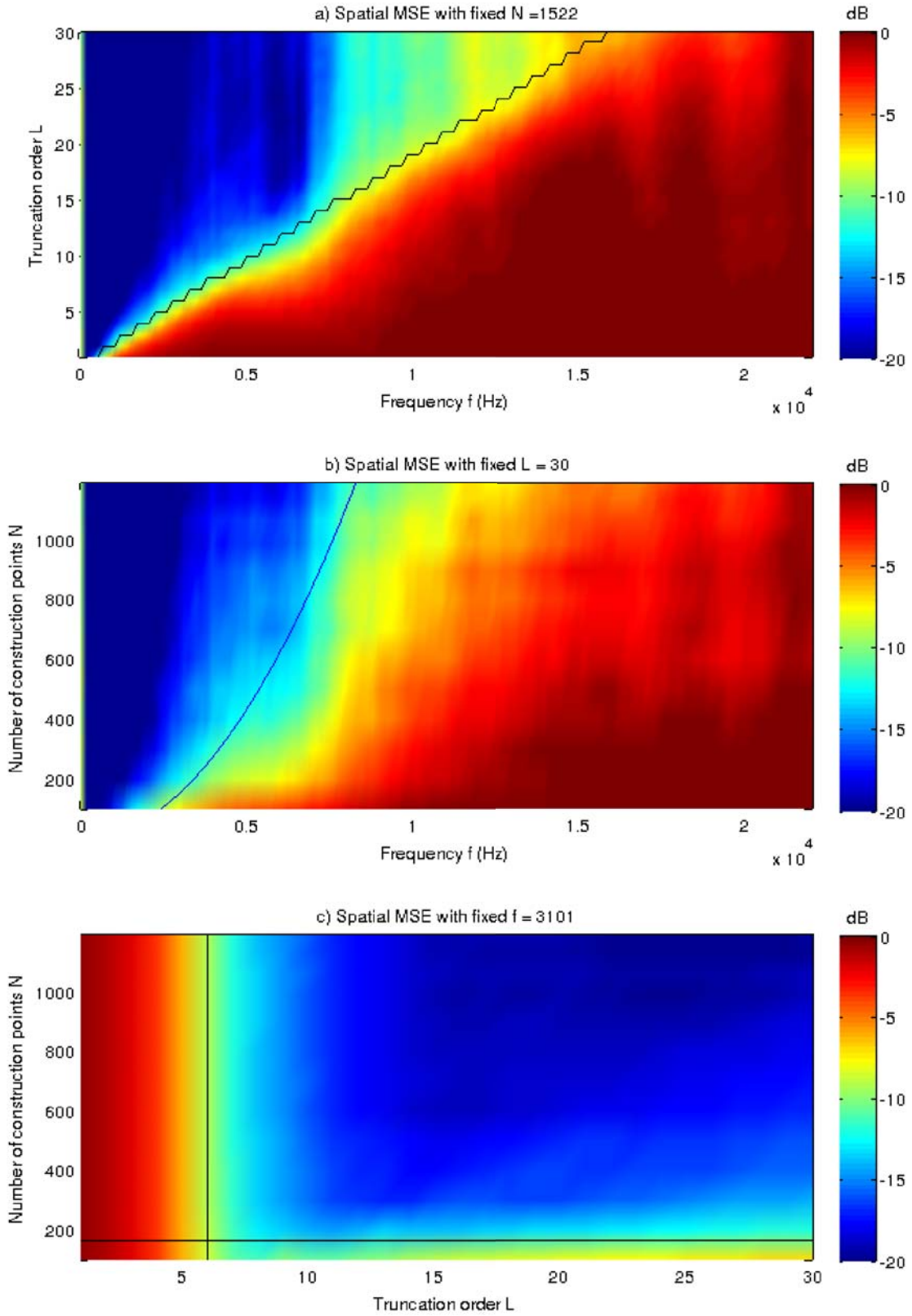

FIGURE 3: Representation of Mean Square Errors (MSE) as a function of different parameters: (a) L & f, (b) N & f, and (c) N & L. Dark curves represent the corresponding theoretical law.