Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

Interpolation of Head Related Transfer Functions Considering Acoustics

Thibaut Ajdler[1], Christof Faller[1], Luciano Sbaiz[1], Martin Vetterli[1,2]

[1]*Ecole Polytechnique Fédérale de Lausanne (EPFL), Laboratory of Audiovisual Communications, 1015 Lausanne, Switzerland*

[2]*Department of EECS, University of California at Berkeley, Berkeley CA 94720, USA*

Correspondence should be addressed to Thibaut Ajdler (thibaut.ajdler@epfl.ch)

ABSTRACT

We are proposing an interpolation technique for head related transfer functions (HRTFs). For deriving the algorithm we study the dual problem where sound is emitted from the listener's ear and the generated sound field is recorded along a circular array of microphones around the listener. The proposed interpolation algorithm is based on the observation that spatial bandwidth of the measured sound along the circular array is limited (for all practical purposes). Further, we observe that this spatial bandwidth increases linearly with the frequency of the emitted sound. The result of the analysis leads to the conclusion that the necessary angle between HRTFs is about 5 degrees in order to be able to reconstruct all HRTFs up to 44.1 kHz in the horizontal plane.

1. INTRODUCTION

A number of head-related transfer function (HRTF)[1] interpolation methods have previously been proposed. The HRTFs are either interpolated in the time domain, frequency domain, or some other representation such as principle component domain. One of the most simple and straightforward methods applies linear interpolation using the nearest neighbor HRTFs for obtaining HRTFs at any angle in between those [1].

[1]In most cases, we denote with "HRTF" the corresponding time domain impulse response.

More sophisticated methods have been proposed which do not only take into account the two nearest available HRTFs. One of these methods is the inverse distance weighting method [2], where several neighboring HRTFs are weighted by the inverse of the distance to the HRTF to be obtained by interpolation. Also splines [3] have been used for HRTF interpolation.

It has been shown that for interpolation in the time domain or frequency domain performance can be improved by compensating the HRTFs prior to interpolation according to the time of arrival of sound [4]. That is, the

HRTFs are time aligned and interpolation is carried out on the time-aligned HRTFs. Additionally, the time of arrival is interpolated separately.

Comparisons of HRTF interpolation methods have been presented in [5, 2]. Most previous studies have applied and evaluated HRTF interpolation methods with significantly less given measured HRTFs than would be necessary to represent all information contained in the HRTFs at all possible directions. We are deriving the angular spacing between measured HRTFs that is necessary such that HRTFs can be obtained with high precision at all possible azimuthal angles. It will be shown that this angular spacing is directly related to the bandwidth of the HRTFs. We show that with sufficiently small angular spacing all possible HRTFs can be interpolated with high precision.

For deriving the mentioned relations, the dual problem is analyzed: sound is emitted from the left or right ear entrance and the sound field is analyzed on a circle around the subject. Considering a spherical head model with diffraction effects, it is shown that the occurring spatial frequencies on the circle are almost bandlimited. We show that the maximal angular frequency is proportional to the maximal temporal frequency of the emitted signal. Given this insight, spatial interpolation is applied along the circle considering all given HRTFs. The effective bandwidth of spatial frequencies along the circle is such that for a temporal sampling frequency of 44.1 kHz, HRTFs are needed every 5° such that they capture all the information, i.e. such that the HRTFs can be obtained for any angle.

A number of numerical simulations considering models and measured HRTF data are carried out. It is confirmed that when the HRTFs are not closely enough spaced most error occurs above a specific temporal frequency. This frequency is dependent on the spacing between the HRTF measurements.

The paper is organized as follows. Section 2 studies the bandwidth limit of the sound field on a circle in a setup similar to the dual problem of HRTF measurement. Given this result, sampling and interpolation of HRTFs is considered in Section 3. Numerical simulations and experiments are presented in Section 4. Section 5 discusses the proposed technique and some further planned work. Finally, conclusions are drawn in Section 6.

2. INTERPOLATION OF THE SOUND FIELD ALONG A CIRCULAR MICROPHONE ARRAY

In this section we investigate the interpolation of the sound field along a circular array of microphones. The specific setup is presented and the necessary number of microphones needed to interpolate the sound field on the circle is derived. This is done by studying the spatial bandwidth of the sound field along the circle. The results of this analysis will further be applied to the problem of HRTFs interpolation in Section 3.

2.1. Problem setup

Consider the situation depicted in Fig. 1. A circular microphone array of radius r in free field is given. The coordinates of the different microphones are (m_x, m_y, m_z), with $m_x = r\cos\theta$ and $m_y = r\sin\theta$. We assume the sound source to be located inside of the array and to have coordinates $(s_x, 0, s_z)$ (for simplicity of the further derivations, we assume the source to have no y component). The source is located at a distance s from the center of the array. We want to determine the number of microphones that need to be placed on the circle in order to interpolate the sound field at any position of the circle. Note that when the sound field is known at every position on the circle, it can further be extrapolated outside of the circle when no additional source is located in the region of space to be extrapolated [6, 7].

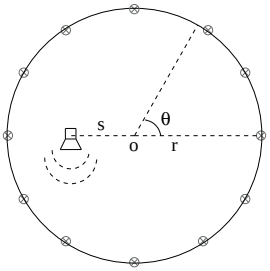

Fig. 1: A loudspeaker emits sound in free field. The sound field is recorded along a circle with equally spaced microphones.

2.2. Spatial bandwidth of the sound signal

In the following, we first show how the sound field is represented in the angular-time domain. Further, an analysis of the sound field is given in both angular and temporal frequency domain. We derive for each temporal frequency the bandwidth along the angular frequency.

2.2.1. Sound field on the microphone array

Consider a source emitting sound in free field. The dif-

ferent time signals recorded at any angle on the circle are gathered in a function $p(\theta, t)$. This function is shown in Fig. 2(a) for the source emitting a Dirac pulse. The top view of the function is presented in Fig. 2(b) and represents the time that the sound has traveled from the source to the microphones. We call this function $h(\theta)$, with

$$h(\theta) = \frac{\sqrt{(s_x - r\cos\theta)^2 + (r\sin\theta)^2 + (s_z - m_z)^2}}{c},$$

where c is the speed of sound propagation.

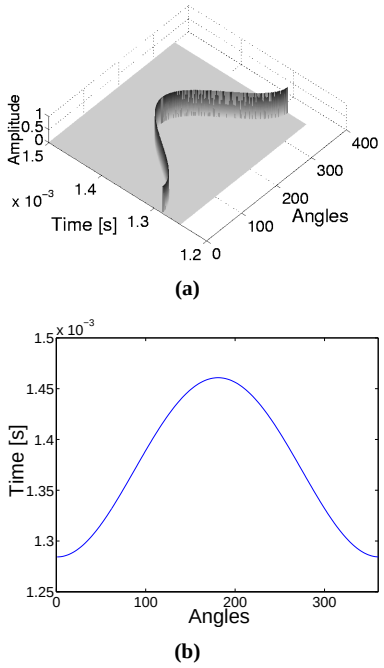

Fig. 2: (a) Sound field measured on the circle. (b) Top view of (a).

2.2.2. Spatial bandwidth

In order to study the spatial bandwidth of the sound field recorded on a circle, variables for the temporal frequency ω and the angular frequency l_θ are used. Note that $l_\theta \in \mathbb{Z}$ due to 2π periodicity of the $h(\theta)$. We consider first the case of a sinusoid of a specific temporal frequency ω emitted by the loudspeaker and recorded on the microphone circle. The angular Fourier transform of the signal gathered on the circle is:

$$Q(l_\theta) = \int_0^{2\pi} e^{-j\omega h(\theta)} e^{-jl_\theta\theta} \mathrm{d}\theta. \qquad (1)$$

Note that in this expression, we do not consider the attenuation depending on the distance traveled. This effect can mostly be considered as negligible as shown in [8].

(1) corresponds also to the Fourier transform of a phase modulation (PM) signal where the carrier frequency would be zero and the modulation function would be $h(\theta)$. The bandwidth of this signal has been studied in the literature [9, 10] and has led to the Carson's rule. In our case, the bandwidth of $Q(l_\theta)$ can be approximated by

$$\text{Bandwidth}(Q(l_\theta)) = \max_\theta \left[\frac{dh}{d\theta}\right] \omega + W, \qquad (2)$$

with W the bandwidth of $h(\theta)$.[2]

The first derivative of $h(\theta)$ with respect to θ is given by

$$\frac{dh}{d\theta} = \frac{s_x r \sin\theta}{c\sqrt{(s_x - r\cos\theta)^2 + (r\sin\theta)^2 + (s_z - m_z)^2}}. \qquad (3)$$

To know where the maxima of this function occur, we calculate the second derivative of $h(\theta)$ with respect to θ. By setting the obtained expression to zero, we obtain the values of the angle θ that maximize the first derivative of $h(\theta)$. Replacing these values of θ in (3), we obtain:

$$\frac{dh}{d\theta} = \pm\frac{\sqrt{A + 2s_x r} - \sqrt{A - 2s_x r}}{2c}, \qquad (4)$$

with $A = s_x^2 + r^2 + (s_z - m_z)^2$. When the source is located in the same plane as the circular array, the expression of the derivative gets simpler: for a source located inside of the circular array, the maximal derivative of $h(\theta)$ is $\frac{dh}{d\theta} = \pm\frac{s}{c}$; for a source located outside of the circular array, the maximal derivative of $h(\theta)$ becomes $\frac{dh}{d\theta} = \pm\frac{r}{c}$. Further we can also prove that the maximal derivative of $h(\theta)$ associated to a source outside of the plane of the array is always smaller than the derivative for a source on the plane:

$$\left|\frac{\sqrt{A + 2s_x r} - \sqrt{A - 2s_x r}}{2c}\right| \leq \left|\frac{min(s, r)}{c}\right|.$$

The maximum derivative is thus only dependent on the minimum between the distance from the center of the array to the source and the radius of the array,

$$\left|\frac{dh}{d\theta}\right| \leq \frac{min(s, r)}{c}. \qquad (5)$$

[2] To apply Carson's rule, we need to satisfy either $W \ll \max_\theta \left[\frac{dh}{d\theta}\right] \omega$ or $W \gg \max_\theta \left[\frac{dh}{d\theta}\right] \omega$.

The signal $h(\theta)$ is a very smooth signal and therefore its bandwidth W can be shown to be approximately zero (unless in specific cases where the source is very close to microphones). Therefore, for a source located inside the circular array, (2) becomes

$$\text{Bandwidth}(Q(l_\theta)) \approx \frac{\omega s}{c}. \tag{6}$$

In case of a Dirac emitted from the source, the sound field can be represented by its 2-dimensional Fourier transform (2D-FT) $P(l_\theta, \omega)$. This spectrum has a bow-tie shape following the rule:

$$|l_\theta| \leq |\omega| \frac{s}{c}. \tag{7}$$

This spectrum is shown in Fig. 3.

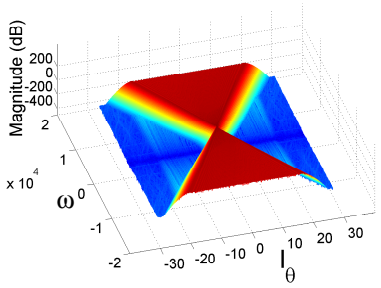

Fig. 3: 2D-FT of the sound field recorded on a circular array.

3. HRTF SAMPLING AND INTERPOLATION

In Section 2, we have presented theoretical results for the sampling of the sound field using a circular microphone array. The same theory can be applied for the dual problem, the case when the sound field is measured at one position and sound is emitted by a circular loudspeaker array. A very interesting application of this dual problem is the sampling of HRTFs in an anechoic chamber to measure the characteristics of the pinnae, head, and torso of a person [11]. The typical setup for HRTF measurement is shown in Fig. 4. The loudspeakers are located along a circle around the person. The microphone is located in the ear of the listener to capture the sound field at the entrance of the ear. One can place all the different HRTFs (time domain impulse responses) next to each other and take the 2D-FT of this data. To understand the shape of the spectrum, we consider the same theory as the one presented in Section 2. When the head is well

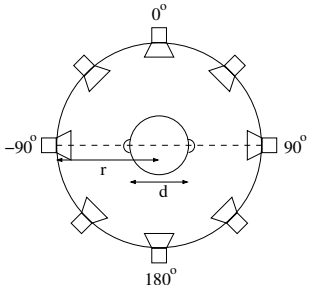

Fig. 4: Setup for the recording of HRTFs.

centered in the middle of the loudspeaker array, we consider the position of the microphone to be $\frac{d}{2} = 9$ cm away from the center of the circle (half the spacing between the two ears) [11]. We therefore can use (7) to predict that the 2D spectrum of the HRTFs will also have a bow-tie shape spectrum following the relation:

$$|l_\theta| \leq |\omega| \frac{d}{2c} \approx |\omega| \frac{0.09}{c}. \tag{8}$$

This relation gives us the angular frequency support corresponding to any temporal frequency. Therefore, the Nyquist theorem will give us for any temporal frequency the necessary angular spacing between consecutive loudspeaker positions. We can define an angular sampling frequency, $l_{\theta_S} = \frac{2\pi}{\Delta\theta}$, where $\Delta\theta$ corresponds to the spacing between two consecutive loudspeaker positions. To satisfy Nyquist we need to have

$$|l_{\theta_S}| \geq 2|\omega_{\max}| \frac{d}{2c} \approx 2|\omega_{\max}| \frac{0.09}{c}, \tag{9}$$

with $\omega_{\max}$ the maximal temporal frequency present in the signal. In particular, (9) indicates that in order to sample HRTFs for a average adult human ($d \approx 0.18$ m) with a temporal sampling rate of 44.1 kHz, a spacing of $4.9°$ is necessary. Sampling the HRTFs with a larger angular spacing will lead to errors in the interpolation due to aliasing.

3.1. Head shadowing

The theory expressed above is valid in the case of HRTFs when the effect of the head shadowing is not considered. In practice, the wave is diffracted by the head [11]. This diffraction has to be taken in account and modifies $h(\theta)$. A model is given by [12]. The HRTFs are expressed as:

$$H(\rho, \mu, \theta) = -\frac{\rho}{\mu} e^{-i\mu\rho\Psi}, \tag{10}$$

with

$$\Psi(\rho,\mu,\theta)=\sum_{m=0}^{\infty}(2m+1)P_m(\cos\theta)\frac{h_m(\mu\rho)}{h'_m(\mu)},$$

where μ is the normalized temporal frequency, P_m is a Legendre polynomial of degree m, and h_m is an m^{th} order spherical Hankel function. Taking the Fourier transform of (10) we also observe a bow-tie spectrum satisfying (8), as shown in Fig. 5(a). The obtained spectrum considering HRTFs measured on a Kemar head [13] sampled every 5° in an anechoic chamber is shown in Fig. 5(b). We can observe that (9) is satisfied since a spacing of 5° results in almost no aliasing at 44.1 kHz.

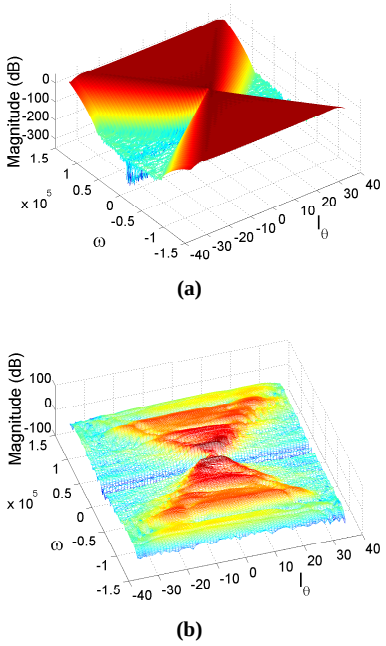

Fig. 5: 2D Spectra of HRTF: (a) using a diffraction model. (b) Using measured data.

3.2. HRTF Interpolation

The interpolation of the dataset is obtained using a sinc interpolator in time domain (or more efficiently a zero-padding in frequency domain). This interpolation is very suitable in the case of a circular array since the Fourier transform is applied on a circular array that is 2π periodic. In the case of interpolation along a linear array, interpolation performance decreases due to the finite length of the array which introduces border effects [14].

Depending on the angular sampling of the database to be considered, interpolation is only applied for frequencies satisfying (9). Higher frequencies will not be correctly interpolated due to the spectral repetitions appearing because of the angular sampling.

4. SIMULATIONS AND EXPERIMENTS

In this section, we present numerical simulations and experimental results to study the error obtained by interpolation of the HRTFs.

4.1. Interpolation error

To verify the performance of the interpolation, we used 36 measurements spaced every 10° and interpolated HRTFs every 5°. The interpolated HRTFs were compared with corresponding measurements which were available and the normalized mean square error (MSE) was calculated.

Using a spacing of 10°, interpolation of the HRTFs can only be correctly done up to 10.8 kHz as follows from (9). Therefore, prior to interpolation the HRTFs (impulse responses) are lowpass filtered using a lowpass filter with a cutoff frequency corresponding to the maximum temporal frequency associated to the specific angular spacing.

The first data set was a simple simulation in free field with a set of loudspeakers and a microphone mimicking a setup for HRTF measurement without considering head shadowing. The MSE of the interpolation is shown as solid line in Fig. 6(a). The same simulation was carried out using the model of HRTFs from [12] and the MSE is shown as a dashed line in Fig. 6(a). We see that these two experiments show a MSE varying from -35 to -65 dB. Best interpolation is obtained at the positions around -90 and 90° while slightly worse interpolation is achieved at 0 and 180° (the angles are referenced in Fig. 4).

Finally, the MSE on the interpolation of the measured HRTFs from [13] is shown as a solid line in Fig. 6(b). The MSE was slightly higher than in the simulations but still of the order of -40 dB.

For purpose of comparison, we applied a technique proposed in the literature [4] in order to compare the results. HRTFs were first aligned and the time aligned version of the HRTFs were linearly interpolated. The time of arrivals were also linearly interpolated. The MSE on the interpolation is presented in dashed line in Fig. 6(b). It

can be observed that this method is clearly worse than the proposed technique.

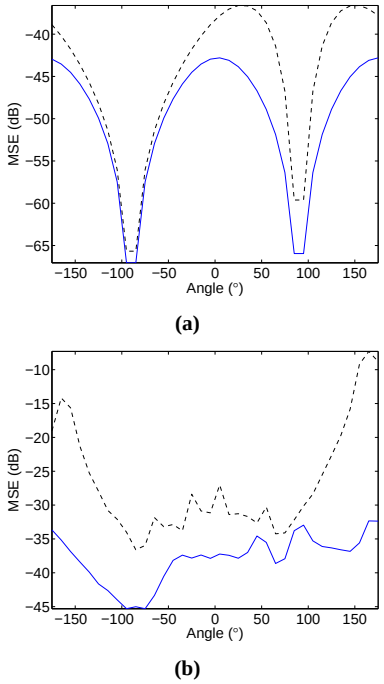

Fig. 6: Interpolation MSE. (a) Comparison of MSE for interpolation of simulated HRTFs with (dashed) or without head shadowing (solid). (b) MSE of the interpolation applied to measured HRTFs (solid) and nearest neighbor linear interpolation technique applied to measured data (dashed).

4.2. Frequency dependence of interpolation error

In this section, we present results on the frequency dependent character of the error obtained by interpolating the HRTFs. As was explained in Section 3.2, correct interpolation can only be achieved when (9) is satisfied. This relation can be verified by observing the frequency dependent error on the interpolation when applying the algorithm on HRTFs too coarsely spaced without prior lowpass filtering. The HRTFs considered in this section were sampled at a temporal sampling frequency of 44.1 kHz.

Fig. 7(a) presents the frequency dependent error on HRTFs averaged over all spatial positions. Interpolation was applied in the case of 10° (curves A and B) and 20° (curves C and D) angular spacing between consecutive loudspeakers. Curves A and C (solid lines) represent the simulated HRTFs without head shadowing effect, while the effect is taken into account in curves B and D (dotted lines). It can be observed that the curves considering the head shadowing are very close to the simpler model without considering head diffraction. Therefore it can be concluded that taking this effect in account does not modify the error significantly.

Fig. 7(b) presents the same results in case of measured HRTFs. The interpolation error when using HRTFs every 10° is shown as solid line and as dotted line for an angular spacing of 20°. In both figures, the two vertical lines correspond to the maximal values of the temporal frequencies corresponding to angular samplings of 10° and 20° as given by (9). The two figures allow us to conclude that interpolating HRTFs for higher frequencies than the ones predicted by (9) leads to large errors while the interpolation error stays limited when obeying (9).

5. DISCUSSION

The proposed technique explores the bandlimited character of the sound field along the circle in an HRTF measuring scenario. Conventional interpolation technique is applied considering all the HRTFs along the circle. Our method does not require any time alignment as is often done in previous techniques.

When HRTFs with high bandwidth are required the angular spacing between HRTF measurements needs to be quite small (e.g. 5° for a sampling rate of 44.1 kHz). We have shown that when the angular spacing between HRTF measurements is too large, the error occurs mostly at high frequencies. For example, when HRTFs need to be interpolated up to 2 kHz, it is sufficient to measure them only every 40°. This motivates further work: we plan to apply the proposed technique only at frequencies where according to the analysis of this paper no aliasing occurs. At higher frequencies we plan to devise a different interpolation algorithm, possibly taking advantage of the "phase deaf" character of the auditory system at higher frequencies.

6. CONCLUSION

We proposed an HRTF interpolation algorithm based on spatial interpolation of the sound field on a circle. For a given audio bandwidth we derived the necessary angular spacing between HRTF measurements such that HRTFs

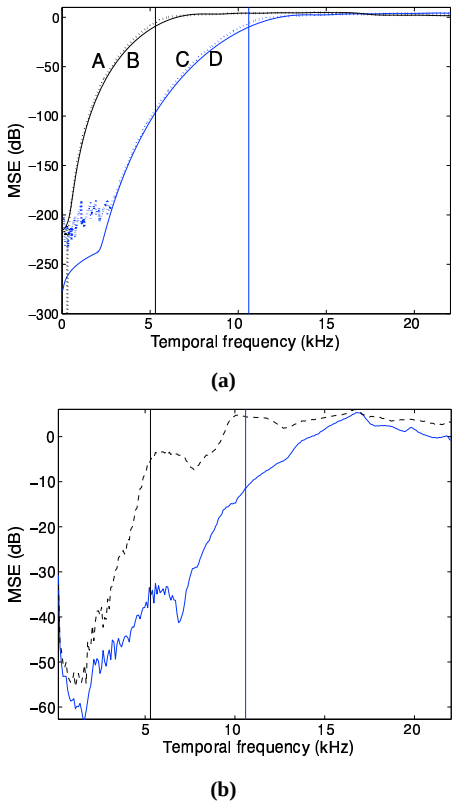

Fig. 7: Frequency dependent MSE for the interpolation of HRTFs positions. (a) Simulations without considering the head (dotted lines, curves A and C) and with spherical head model (solid, curves B and D). Curves A and B are obtained with an angular spacing of 10° and curves C and D with 20°. (b) Measured HRTFs for an angular spacing of 10° (solid) and for 20° (dashed). In both figures, the two vertical lines correspond to the maximum values of the temporal frequencies corresponding to angular samplings of 10° and 20° as given by (9).

at any angle can be interpolated accurately. Numerical results carried out with head models and measured data, indicate that indeed HRTFs can be interpolated very precisely if the angular spacing between measurements is small enough relative to the required audio bandwidth.

7. REFERENCES

[1] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA, 1994.

[2] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in *Proc. AES 16th Int. Conf.*, 1999, pp. 319–329.

[3] S. M. Robeson, "Spherical methods for spatial interpolation: Review and evaluation," *Cartography and Geographic Information Systems*, vol. 24, no. 1, pp. 3–20, 1997.

[4] M. Matsumoto, S. Yamanaka, M. Tohyama, and H. Nomura, "Effect of time arrival correction on the accuracy of binaural room impulse response interpolation," *J. Audio Eng. Soc.*, pp. 56–61, Jan./Feb. 2004.

[5] P. R. Runkle, M. A. Blommer, and G. H. Wakefield, "Comparison of head-related transfer function interpolation methods," in *Proc. IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust.*, 1995.

[6] A.J. Berkhout, *Applied Seismic Wave Theory*, Elsevier Science, 1987.

[7] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by wave field synthesis," in *110th Conv. of the AES*, 2001.

[8] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function on the circle with application to HRTF interpolation," *In proceedings of IEEE ICASSP*, 2005.

[9] A.B. Carlson, *Communication systems*, McGraw-Hill, 1986.

[10] M. Do and M. Vetterli, "On the bandlimitedness of the plenoptic function," *submitted to IEEE ICIP*, 2005.

[11] J. Blauert, *Spatial hearing*, MIT press, 2001.

[12] R. O. Duda and W.M. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, pp. 3048–3058, 1998.

[13] B. Cardner and K. Martin, "HRTF measurements of a Kemar dummy-head microphone," Tech. Rep., MIT Media Lab, 1994.

[14] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *Submitted to IEEE transactions on Signal Processing*, 2005.