Acoustic based rendering by interpolation of the Plenacoustic function

Thibaut Ajdler[a] and Martin Vetterli[a,b]

[a] Laboratory for Audio-Visual Communications (LCAV), Swiss Federal Institute of Technology Lausanne (EPFL), 1015 Lausanne, Switzerland
[b]Department of EECS, University of California at Berkeley, Berkeley CA 94720, USA

ABSTRACT

In the present paper, we study the spatialization of the sound field in a room, in particular the evolution of room impulse responses as function of their spatial positions. The presented technique allows us to completely characterize the sound field in any arbitrary location if the sound field is known in a certain finite number of positions. Our technique simply starts from the measurements of impulse responses in a finite number of positions and with this information the total sound field can be recreated. An analytical solution of the problem is given for any rectangular room. Further, we determine the number and the spacing between the microphones needed to perfectly reconstruct the sound field up to a certain temporal frequency.

1. INTRODUCTION

The name of the plenacoustic function has been for the first time mentioned in.[1] The first analysis of the function has been given in.[2,3] In the present paper, we review some aspects explained in[3] and we develop some new mathematical aspects of the plenacoustic function representation in the frequency domain.

The plenacoustic function characterizes the sound field in space, e.g. inside a room. The idea of the plenacoustic function is similar to the one introduced by the plenoptic function already studied in the image processing world.[4] There, the idea is to consider the light field in the neighborhood of a scene. The plenoptic function records the light intensity of all the rays of light in this neighborhood. This function requires a huge amount of memory to be stored. Sampling of the plenoptic function is studied in.[5]

We study a similar idea in the area of audio. We focus on the plenacoustic function in rooms. Therefore, we are interested in the room impulse responses to characterize what one would hear at any point in the room. Knowing this information, we can simply calculate the convolution of the sound produced by the source at some point in the room with the room impulse response from the source's position to the listener's position. This gives precisely what the sound is at the listener's position. The plenacoustic function is thus parameterized by the following factors:

$$P_a(S\ and\ R\ position,\ characteristics\ of\ room,\ time)$$

with S and R being the source and receiver and the characteristics of the room being for example the reflection factors of the walls of the room.

The plenacoustic study answers the following question: "How many microphones do we need to place in the space in order to completely reconstruct the sound field at any position in the space?"

The outline of the paper is the following. We briefly summarize the description of the plenacoustic function in Section 2. Section 2.1 introduces the model for the room impulse responses. We then explain the construction of the plenacoustic function in Section 2.2. In Section 2.3, we study the sampling of this function. An analytical solution of the plenacoustic function is then given in Section 3. Different sampling methods are compared to achieve better reconstruction in Section 4. Section 5 shows the experimental results obtained by measuring the room impulse responses in a room. The conclusions are finally drawn in Section 6.

Further author information:
Thibaut Ajdler: E-mail: thibaut.ajdler@epfl.ch, Telephone: +41 21 693 12 07
Martin Vetterli: E-mail: martin.vetterli@epfl.ch

2. PLENACOUSTIC FUNCTION

In this section, we review different theoretical aspects of the plenacoustic function. More information can be found in.[2,3]

2.1. Room impulse response

In order to calculate the plenacoustic function in a room, we need to know the room impulse response at any point in the room. We use the image method discussed in.[6] The method is based on the creation of virtual sources in order to simulate the effect of the reflections on the walls. In the case of a rectangular rigid-walls room of size (L_x, L_y, L_z), we get the following expression:

$$p(t, S, R) = \sum_{p=0}^{7} \sum_{v=-\infty}^{\infty} \frac{\delta[t - |d_p + d_v|/c]}{4\pi |d_p + d_v|}, \tag{1}$$

where $d_p = (x_s \pm x_r, y_s \pm y_r, z_s \pm z_r)$, $d_v = (2nL_x, 2lL_y, 2mL_z)$ and (n, l, m) being an integer vector triplet. The room impulse response is a function of time and is dependent on the source's position $S = (x_s, y_s, z_s)$ and the receiver's position $R = (x_r, y_r, z_r)$. The first sum shows that in a 3-dimensional field, 7 virtual sources are created in addition to the original source. The second sum shows that sound between two parallel rigid walls is infinitely reverberated. More general formulas taking into account the reflection factors of the walls are given in.[6]

2.2. Construction of the plenacoustic function

With these room impulse responses, we construct the plenacoustic function for a room. Let us consider (without loss of generality) that we study the plenacoustic function along a line in the room. We can construct a 3-dimensional graph formed by all the room impulse responses at any position along the axis. In such a graph the two base axes are space and time. Space corresponds to the listener's position, while the time corresponds to the duration of the room impulse responses. Studying the plenacoustic function along a line in the room would require to store the room impulse responses for every single point along this line. The time continuous impulse responses should also be known. This data is impossible to store, therefore we need to sample the plenacoustic function.

2.3. Plenacoustic sampling

In order to sample the plenacoustic function, we need to sample the room impulse responses at a certain temporal sampling rate depending on the desired audio bandwidth. Further, by taking an evenly spaced finite number of impulse responses, we uniformly sample the plenacoustic function in space. When the plenacoustic function is sampled, repetitions of the spectrum occur.[7]

The 2-dimensional Fourier transform (2D-FT) of the obtained plenacoustic function is shown in Fig. 1. In this figure, the two base axes represent the corresponding axes of time and space, being a temporal and a spatial frequency. Temporal and spatial pulsations are defined as $\omega_t = \frac{2\pi}{\Delta t}$ and $\omega_d = \frac{2\pi}{\Delta d}$, respectively. Δt is the sampling period of the impulse responses and Δd the sampling interval between the different positions of the measured impulse responses. The third axis is the amplitude (in dB) of the 2D-FT. We clearly see a triangular shape in this 2D-FT of the plenacoustic function. We can furthermore observe that the part outside of the triangle is of much lower amplitude than the part inside the triangle. The contents of the 2D-FT of the plenacoustic function can be schematized as in Fig. 2. As it can be seen in Fig. 2, the spatial frequency support grows for increasing temporal frequencies. Spatial sampling of the plenacoustic function at some spatial sampling frequency (e.g. ω_{d1} in Fig. 2) will lead to aliasing for all the temporal frequency above some corresponding temporal frequency (ω_{t1}). Perfect reconstruction becomes then impossible above this temporal frequency. Similarly, the sampling line at the right of the triangle in Fig. 2 shows for any temporal frequency, the minimal spatial sampling frequency to avoid aliasing. It can be shown that the equation of this line is:

$$\omega_d = \frac{2\omega_t}{c}, \tag{2}$$

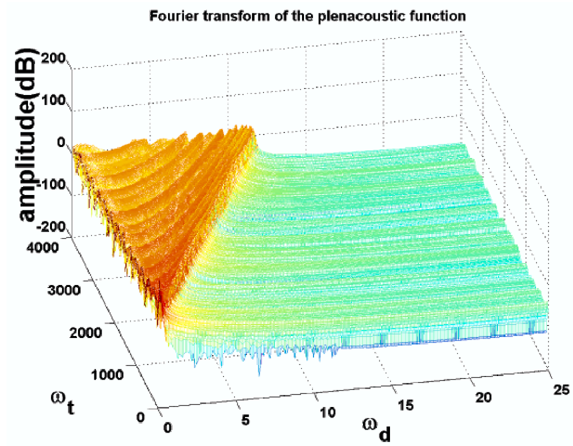


Figure 1. Two dimensional Fourier transform of the plenacoustic function.

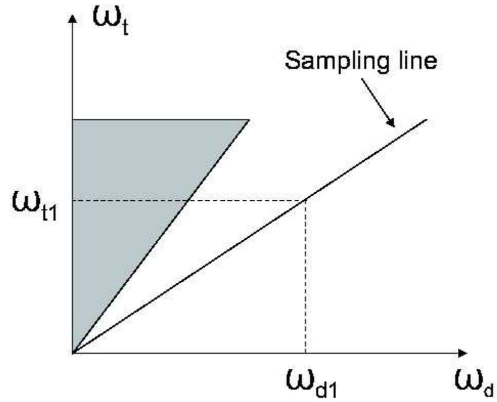


Figure 2. Schematic view of plenacoustic function in frequency domain.

with c the sound propagation speed in the air.

In order to understand physically the meaning of the triangle, consider in Fig. 3 a certain temporal frequency of v Hz emitted from a source with an angle β with respect to the line of the microphones. The period related to this frequency is $1/v$ s and its wavelength is c/v meters (m). If we are far enough from the source, we can consider that the apparent wavelength on the line of microphones is $\lambda_2 = \frac{c}{v cos\beta}$ s. The apparent wavelength is thus larger and the apparent frequency smaller. In the case where $\beta = 0$, we have the apparent frequency being equivalent to the real frequency. Similarly to the Shannon theorem in time domain, we can say that in order to perfectly reconstruct the sine of c/v m we need to sample the sine at a period of maximum $c/2v$ m in space. The spatial sampling frequency corresponding to this maximal period is $u_s = \frac{1}{c/2v} = \frac{2v}{c}$. The maximal spatial frequency that we can obtain is

$$u = u_s/2 = v/c\,. \tag{3}$$

We can verify this in Fig. 1 where the plenacoustic function has been sampled every 2 cm along a line in the room and where for a spatial frequency of 12 m^{-1} the corresponding temporal frequency lies around 4 kHz.

3. MATHEMATICAL DERIVATION OF THE PLENACOUSTIC FUNCTION

We have studied mathematically the 2D-FT of the plenacoustic function in different cases. An analytical formula is first derived for the free field case, then a general formula is developed in the case of a general room.

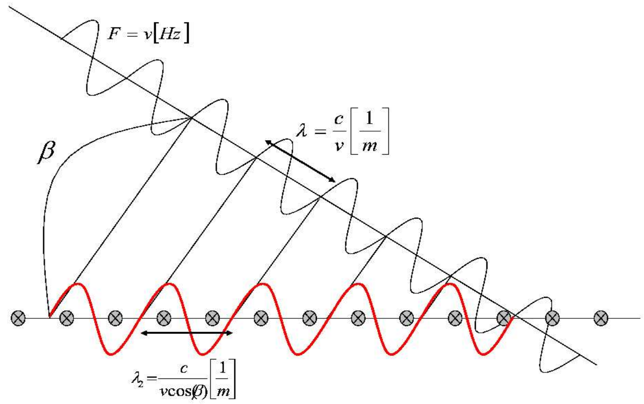


Figure 3. The apparent frequency on the line of the microphone is smaller than the real frequency of the emitted signal.

3.1. Free field case

We study the evolution of the room impulse response along an axis (e.g. the x-axis). We fix the source's position and move the microphones (x_r). The only variables are thus x_r and the time t. All the other variables are fixed. In the free field case, we have a plenacoustic function given by the following formula:

$$f(x_r, t) = \frac{\delta(t - a/c)}{4\pi a} \text{ with } a = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2} \tag{4}$$

Taking the 2D-FT of this function, we obtain:
for $|u| < |v|/c$

$$F(u, v) = -\frac{g}{4}\left[N_0(s) + j\,sign(v)J_0(s)\right], \tag{5}$$

for $|u| > |v|/c$

$$F(u, v) = \frac{g}{2\pi}K_0\left(2\pi d\sqrt{u^2 - (\frac{v}{c})^2}\right), \tag{6}$$

where $d = \sqrt{(y_s - y_r)^2 + (z_s - z_r)^2}$, $g = e^{-j2\pi u x_s}$ and $s = 2\pi d\sqrt{(\frac{v}{c})^2 - u^2}$. The part of the expression that represents the inside of the triangle is a sum of Bessel and Neuman (Bessel function of second kind) functions. For $u > v/c$, which corresponds to the part outside of the triangle, the plenacoustic function in frequency domain is a sum of modified Bessel functions of the second kind. The order of the function is zero and this is a typically decreasing function. We are thus able to determine how the function is decreasing when, for the same temporal frequency, we increase the spatial frequency.

Observing the function in (4), we remark that the time support of the function is given by:

$$t = \frac{\sqrt{(x_r - x_s)^2 + d^2}}{c}. \tag{7}$$

We see in Fig. 4(a) that the slope of the function gets constant for high values of the microphone position. The derivative of (7) is given by

$$dt = \frac{1}{c}\frac{(x_r - x_s)}{\sqrt{(x_r - x_s)^2 + d^2}}dx_r.$$

We see in Fig. 4(b) that the derivative tends asymptotically to the values $\frac{1}{c}$ and $-\frac{1}{c}$. When $(x_r - x_s)^2 >> d^2$, we get a maximal constant slope of $\frac{1}{c}$. In this case, the 2D-FT of the plenacoustic function has the very specific shape shown in Fig. 5(a). This can simply be explained by the fact that the 2D-FT of the function $z(x,t) = f(x)\delta(\frac{x}{c} - t)$ is $Z(\omega_d, \omega_t) = F(\omega_d + \frac{\omega_t}{c})$. The shape of the frequency spectra are depending on the maximal and minimal values of the slope of (7) given in (8). In Fig. 5(c), we can observe that when the loudspeaker is symetrically located relative to the microphones, a symmetric spectrum will be obtained. For different setups of microphones and loudspeaker (see Fig. 5), the shape of the plenacoustic function can change but the support is always essentially given by (3). There is small energy present outside of the triangular shape. We will study in Section 4 how it influences the quality of the reconstruction.

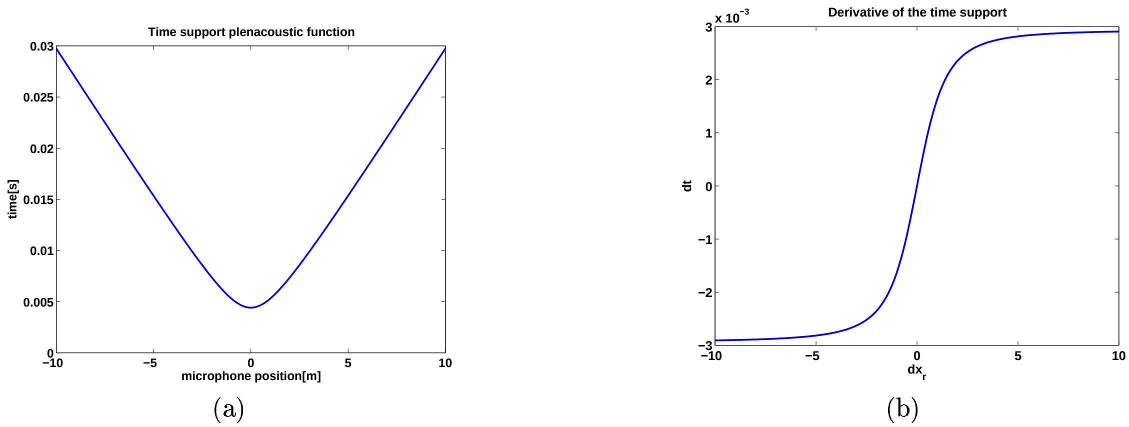


Figure 4. (a) Time support of the plenacoustic function. (b) Derivative of the time support.

3.2. General rectangular room

In the general case of a room of size (L_x, L_y, L_z), we can use (5) and (6) obtained for one source in free field to find the full spectrum of the plenacoustic function. Therefore, we need to sum up the effect of all the virtual sources of the model as shown in (1). The total spectrum is

$$F_{tot}(u,v) = \sum_{p=0}^{7} \sum_{l,m,n=-\infty}^{\infty} F_{p,l,m,n}(u,v),$$

with p, l, m, n being simply the different parameters determining the different virtual sources appearing in the model of.[6] As we are in a room, we will filter the plenacoustic function in time and space with an appropriate filter. In our case, the line of microphone along the x-axis will be filtered to take into account the size L_x of the room in this direction. Such a filter could be

$$h(x,t) = \begin{cases} 1 \text{ when } \begin{cases} 0 \le x \le L_x[m] \\ 0 \le t \le 2\,[s] \end{cases} \\ 0 \text{ otherwise.} \end{cases}$$

Let us call $G_c(u,v)$, the continuous spectrum obtained after filtering, we have thus

$$G_c(u,v) = F_{tot}(u,v) \star H(u,v),$$

with $H(u,v)$ being the 2D-FT of $h(x,t)$. In practice, the spectrum will be finally sampled. The expression for the sampled spectrum is:

$$G_s(u,v) = \frac{1}{\Delta d \Delta t} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} G_c(u - \frac{k_1}{\Delta d}, v - \frac{k_2}{\Delta t}).$$

A diagram of the situation is shown in Fig. 6.

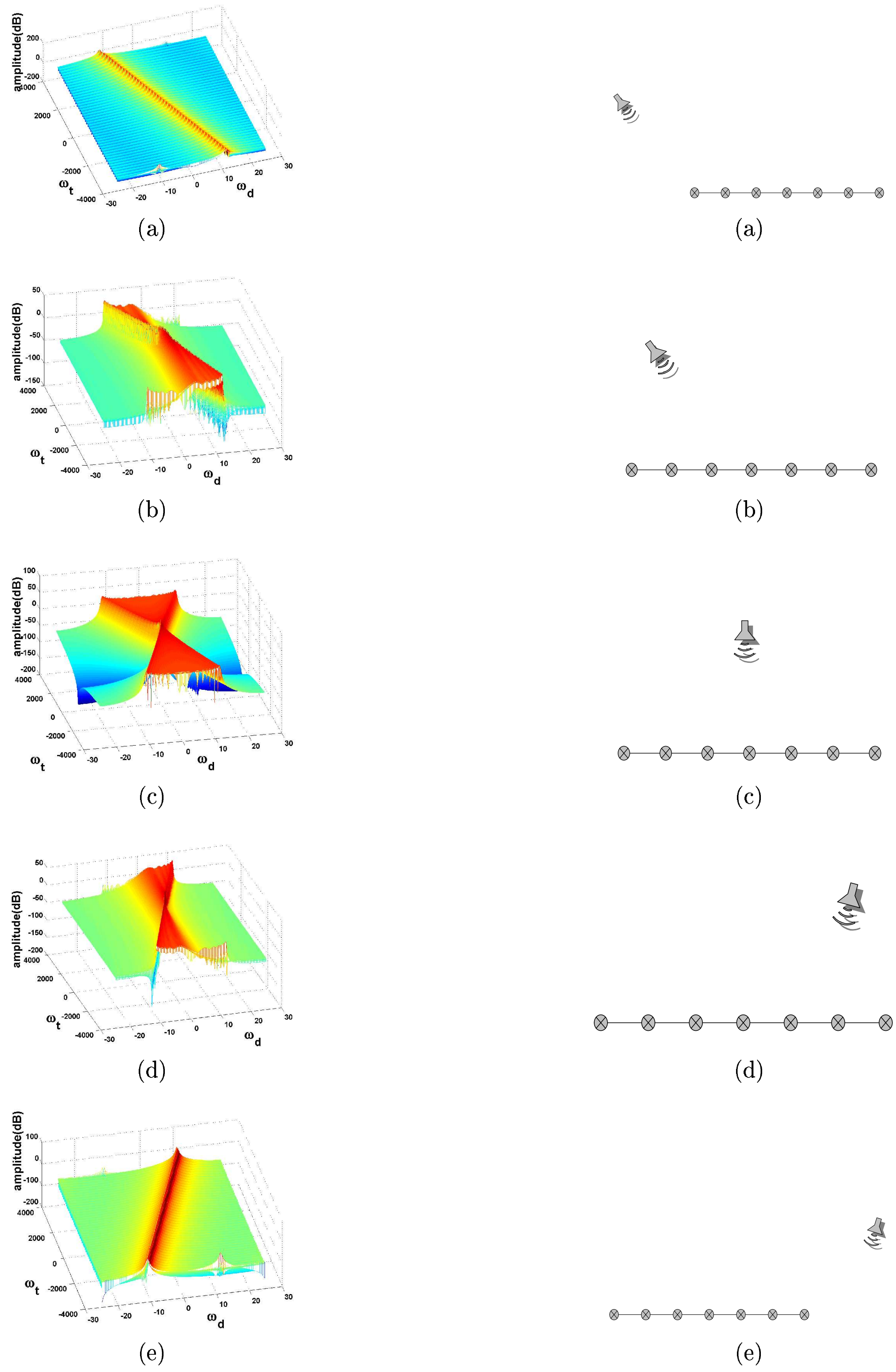


Figure 5. Different setups for the microphones and loudspeaker positions and their corresponding plenacoustic functions in frequency domain.

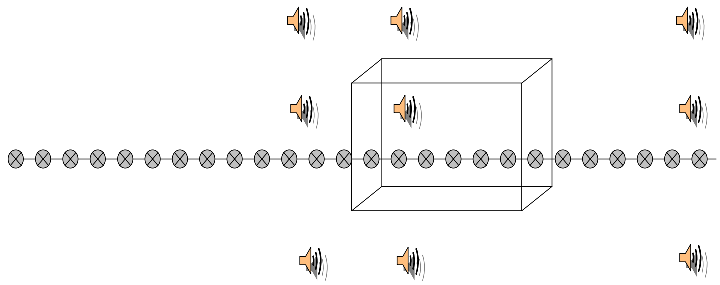


Figure 6. The effect of all the virtual sources are added to obtain an expression of the 2D-FT of the plenacoustic function. The function has been calculated for an infinite line, therefore we will need to filter and sample the plenacoustic function to consider the finite length of the room and the finite number of microphones.

4. RECONSTRUCTION

Knowing the sound field at every point of the sampling grid, we apply the usual interpolation techniques[7] in order to reconstruct the sound field at any location. First, we need to upsample our time domain signal accordingly to the desired location. We then filter the upsampled plenacoustic function with an appropriate 2-dimensional filter. The value at the location of interest is then obtained by interpolation. The interpolation filter to be used is dependent on the sampling grid.

4.1. Rectangular sampling

In the case of rectangular sampling, we just sample the plenacoustic function in time domain with a sampling grid shown in Fig. 7a. Convolution of the sampling grid with the spectrum of the plenacoustic function leads us to Fig. 7b. In this figure, we can also see the filter needed. It is just a rectangular filter. In Fig. 7b we observe that the spatial sampling frequency is $2\omega_0$. The corresponding spacing between the samples on the spatial axis is $d_1 = \frac{2\pi}{2\omega_0} = \frac{\pi}{\omega_0}$. In order to gain some processing time we need to find a better sampling.

4.2. Quincunx sampling

Better performance can be achieved using quincunx sampling. We can actually fill the space lost in frequency domain in Fig. 7b, in order to fill the whole frequency space. This is achieved in Fig. 7d. The corresponding sampling grid in time domain is shown in Fig. 7c. In this case the filter used for interpolation is a fan filter.[8, 9] The filter is shown in Fig. 7d. In the quincunx sampling the spatial sampling pulsation is now only ω_0. This corresponds to the distance between two samples on the space axis of $d_0 = \frac{2\pi}{\omega_0}$. This shows that using the quincunx sampling we need to sample the even microphones at even times while the odd microphones need to be sampled at odd times. This leads to a gain of factor 2 in the processing. Note that once the temporal and spatial sampling frequencies are fixed, the shape of the fan filter remains constant.

4.3. Reconstruction results

In this example, we have applied the rectangular sampling method. We have simulated 200 room impulse responses separated with a spatial spacing of 1 cm along a line in a room. Then, by keeping only every second room impulse responses, we obtain a downsampled version of the original plenacoustic function. Subsequently, we interpolate our data by zero-padding in frequency domain. In Fig. 8(a) and (b), we compare an interpolated with the original room impulse response. We see that the interpolated signal is not perfectly the original signal. The difference between the maximal amplitude of the signals is typically of an order of 10^{-3} of the magnitude of one of the room impulse responses. This difference is due to the fact that the spectrum of the plenacoustic function is not perfectly bandlimited but still contains small energy outside of its triangular shape. By comparing these two waveforms with respect to the impulse response obtained at the next microphone position, we remark in Fig. 8(c) that the interpolation result is still very acceptable.

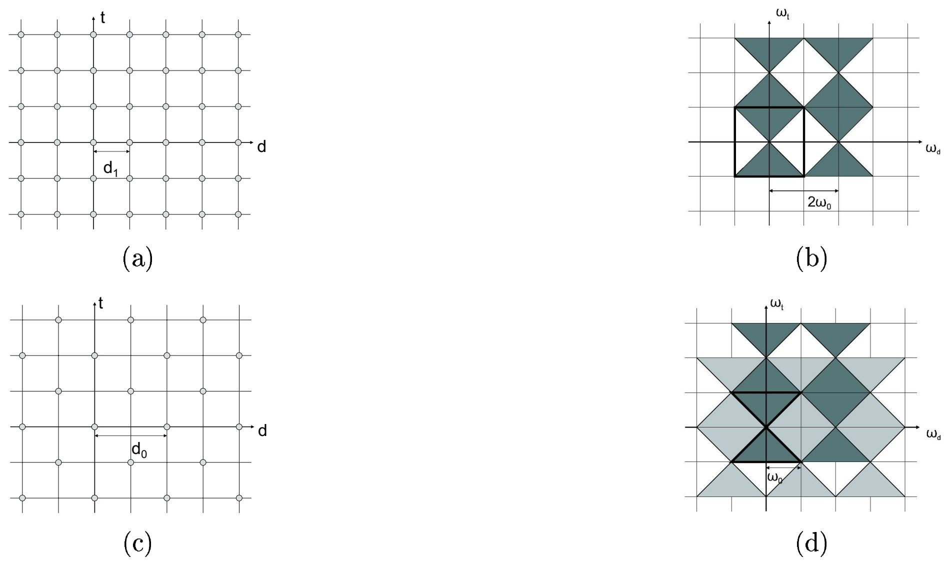


Figure 7. (a) Rectangular sampling grid. (b) Plenacoustic spectrum with its repetitions for a rectangular sampling grid. (c) Quincunx sampling grid. (d) Plenacoustic spectrum with its repetitions for a quincunx sampling grid.

5. EXPERIMENTS

In this section, we review some experimental results. More information can be found in.[3] We made some experiments in order to check if the triangular shape of the 2D-FT of the plenacoustic function was also observed in measured room impulse responses. We used two sets of measures.

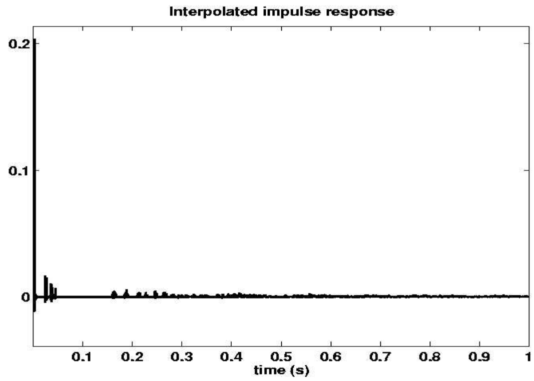
The first set of data comes from measurements made in the varechoic chamber at the Bell Laboratories, Lucent Technologies. These measures were done using a logarithmic sweep signal as excitation signal.[10] The room impulse responses are given for an array of 22 microphones where the spacial spacing between the microphones was 10 cm.

We measured the second set of impulse responses using the Maximum Length Sequences (MLS)[11] technique. The measurements were made in a sound insulated room (LCAV audio room at EPFL). 40 measures were taken with 1.8 cm spatial spacing.

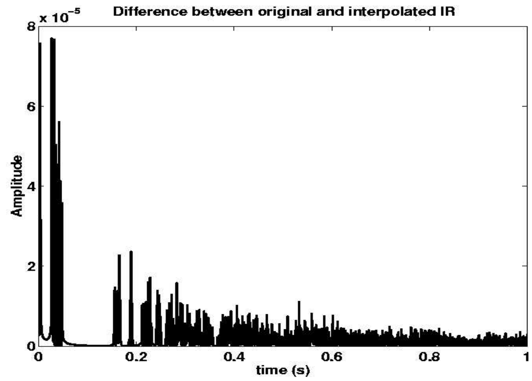
For the first data, we can reconstruct the signal up to a temporal frequency of 1.7 kHz, as can be seen using (3). For the second set of data we can reconstruct the sound field up to a temporal frequency of 9.2 kHz. The 2D-FT of the plenacoustic functions of the two sets of data are shown in Fig. 9a and 9b. In these two figures, we show the whole spatial axis until the sampling frequency (and not until Nyquist) to see clearer until where we get the triangle shape and when the aliasing begins. We can see in Fig. 9a that the aliasing begins around 1.7 kHz while in Fig. 9b the aliasing begins around 9 kHz. We get thus a good correspondence between theoretical and practical results.

6. CONCLUSION

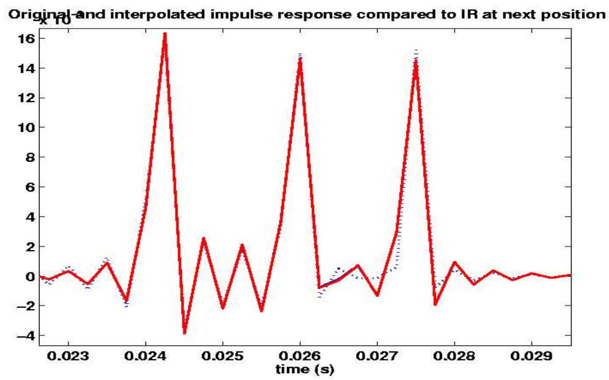
In this article, we have studied the plenacoustic function. The plenacoustic function characterizes the sound field at any point in space. We then studied the sampling of this function. We are able to predict the number of microphones to place in a room to perfectly reconstruct the sound field at any point in the room up to a certain temporal frequency. An analytical derivation of the plenacoustic function was given in free field and for any general rectangular room. An optimal sampling pattern in time and space is given in order to achieve the best performance for reconstruction of the sound field. Reconstruction of the sound field is then discussed. Finally, we presented experimental results and compared theoretical results with measurements.



(a)

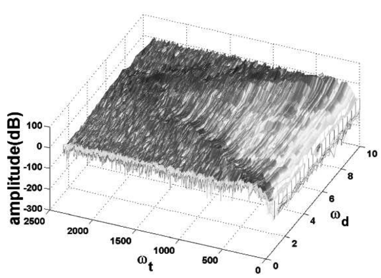


(b)

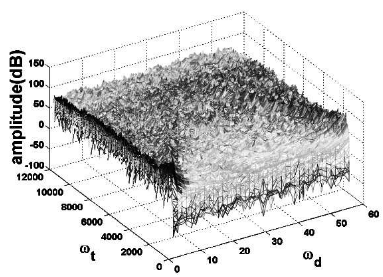


(c)

Figure 8. (a) Original room impulse response. (b) Difference between the original and the interpolated room impulse response. (c) In solid lines, the original and the interpolated room impulse response are shown. In dotted line is the original room impulse response shown for the next microphone position. We see that with respect to the next position, the original and interpolated version are hard to distinguish.



(a)



(b)

Figure 9. Measured plenacoustic functions in frequency domain in (a) the varechoic Bell Labs chamber, (b) the LCAV audio room.

ACKNOWLEDGMENTS

The work presented in this paper was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

REFERENCES

1. M. Kubovy and D. Van Valkenburg, "Auditory and visual objects," *Cognition* **80**, pp. 97–126, 2001.
2. T. Ajdler and M. Vetterli, "The plenacoustic function and its sampling," in *IEEE Benelux Workshop on Model Based Processing and audio Coding*, 2002.
3. T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *Proc. of IEEE ICASSP*, 2003.
4. E. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, pp. 3–20, MIT Press, 1991.
5. J. Chai, X. Tong, S. Chan, and H. Shum, "Plenoptic sampling," in *Proceedings of the conference on Computer graphics*, pp. 307–318, 2000.
6. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *JASA* **65**, 1979.
7. M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Signal Processing, Prentice Hall, A.V. oppenheim ed., 1995.
8. P. Vaidyanathan, *Multirate Filter Banks and Wavelets*, Prentice Hall, 1992.
9. C. Zhang and T. Chen, "Generalized plenoptic sampling," Tech. Rep. Amp 01-06, Carneggie Mellon University, September 2001.
10. A. Harma, T. Lokki, and V. Pulkki, "Drawing quality maps of the sweet spot and its surroundings in multichannel reproduction and coding," *21st Conference on Architectural Acoustics and Sound Reinforcement, AES* , pp. 317–325, June 2002.
11. J. "Borish and J. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc* **31**(7), 1983.