Toronto, Canada
**International Symposium on Room Acoustics**
2013 June 9-11

ISRA 2013

# Parametric control of convolution based room simulators

Thibaut Carpentier (thibaut.carpentier@ircam.fr)
Tal Szpruch (tal.szpruch@ircam.fr)
Markus Noisternig (markus.noisternig@ircam.fr)
Olivier Warusfel (olivier.warusfel@ircam.fr)
UMR STMS IRCAM-CNRS-UPMC
1, place Igor Stravinsky,
75004 Paris, France

**ABSTRACT**

For several decades, reverberation processors have been based on feedback delay networks that provide an efficient way to control the distribution of early reflections and the statistical properties of room reverberation. Their algorithmic architecture also allows for a flexible control through different levels of descriptors, from low-level signal processing parameters (e.g. initial delay, echo density, reverberation decay, etc.) to high-level perceptual descriptors (e.g. source presence, envelopment, reverberance, etc.). Nowadays, most reverberation units rather opt for a convolution approach exploiting a collection of room impulse responses (RIR) measured in existing performance halls or other typical rooms. To retrieve a continuous control over important perceptual dimensions, such reverberators also provide means for tweaking the original impulse response. However the tuning is generally limited to the control of the time envelope of the RIR, whereas little attention is given to the spatial distribution. The objective of this study is to develop a signal-processing environment dedicated to the analysis and re-synthesis of directional room impulse responses (DRIRs) measured with higher-order spherical microphone arrays. The analysis step results in a space-time-frequency representation of the DRIR from which various acoustical or perceptual descriptors can be extracted. During the re-synthesis step, filtering a given time-space-frequency window of the DRIR enables the tuning of these descriptors without altering the microstructure of the original DRIR.

## 1 INTRODUCTION

### 1.1 Convolution-based reverberators

The acoustical path between an emitter and a receiver in a room can usually be considered as a linear time-invariant system, characterized entirely by its impulse response (IR). Filtering an anechoic input signal with a RIR allows for reproducing the room reverberation in a realistic way. This approach has been widely used in the last decades thanks to the increase of available processing power and the development of efficient frequency-domain filtering algorithms, such as the block-partitioned FFT convolution[1] or frequency delay lines[2].

This auralization approach guarantees for the authenticity and the naturalness of the listening experience and convolution processors propose large collections of RIRs measured in prestigious concert halls (e.g. Concertgebouw, Musikvereinssaal) or paradigmatic venues (e.g. cathedral, antique theatre). This paradigm is however restricted by essence to static situations, as for each modification of the receiver and/or source position the RIR must be updated; incorporating source directivity and changes in source orientation are treated in the same way. Obviously, strictly applying this approach would rapidly lead to an unmanageable amount of measurements. To overcome these limitations, convolution-based processors generally provide some parameters to modify the original RIR for controlling the reverberation effect. Up to date, these parameters mainly focus on the time envelope of the RIR, especially the reverberation time and the direct/reverberation ratio. The range of possible transformations provided to the sound engineers is therefore limited and does not include the control over the spatial behavior.

### 1.2 Parametric reverberators (FDN)

Feedback delay networks (FDN) processing structures have been proposed for designing digital reverberators[3,4]. They provide an adequate way to simulate the statistical properties of room reverberation, with arbitrary density and low tonal coloration in the late decay. Furthermore these structures can be efficiently implemented and allow for a continuous tuning of the time and frequency behavior of the room response. Ircam's software 'Spatialisateur' is an example of a parametric reverberator using FDN. It is controlled by a set of perceptual descriptors relying on a simplified model of the time-frequency energy distribution of the IR over four temporal segments and three frequency bands[5]. The spatial distribution can be taken into account by applying specific panning models to the different components of each of the time segments. Figure 1 gives an example of a space-time model where the direct sound (OD) arrives from a given direction, the individual early reflections (R1) are synthetized as waves coming from deterministic directions, and the late reflections (R2) and reverberation decay (R3) are spatially diffuse fields.

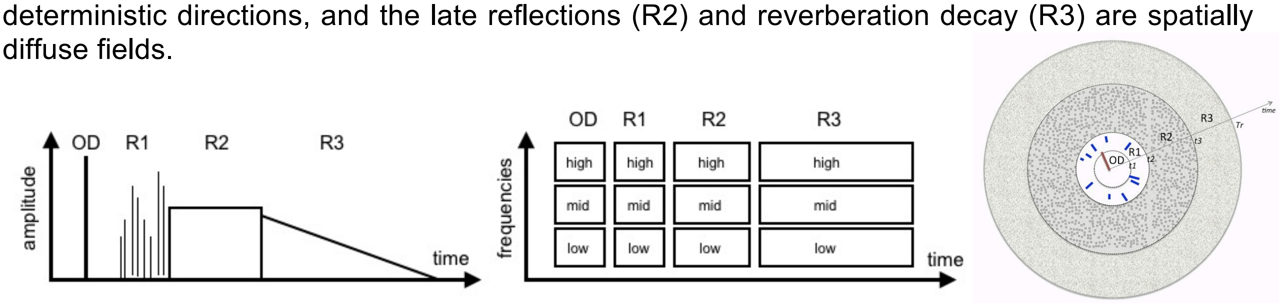


**Figure 1**: Parametric control of the IR in the FDN based Spatialisateur. Simplified IR (left), time versus frequency distribution (center) and associated spatial distribution (right).

### 1.3 Parametric control of convolution based reverberators

The motivation of this study is to apply this parametric space-time-frequency approach to measured RIRs. Firstly, it aims at a faithful reproduction of the spatial properties of directional room impulse responses (DRIRs) recorded, for instance, with a higher-order spherical microphone array (SMA). Secondly, the proposed model allows for a parametric space-time-frequency modulation of the reverberation effect, i.e. suppressing or enhancing a certain region of this multidimensional distribution. In the context of sound mixing, this method provides sound designers with an enhanced control of measured RIRs that could be tweaked according to various acoustical or perceptual descriptors (e.g. lateral energy fraction (LEF), envelopment). In the context of architectural acoustics and room acoustics perception, it allows for investigating the perceptual influence of selected reflections impinging from a limited region of space in a given time window.

This paper discusses the different steps that are necessary for the parametric control of convolution-based reverberation. It starts with the measurement of the room response using a SMA. A method for jointly denoising the signals of the different sensors is proposed, which preserves the diffuse field properties of the late reverberation decay captured by the SMA. Then, a processing framework for the analysis, synthesis, and treatment of DRIRs is presented; e.g. it allows for filtering the space-time-frequency distribution of the DRIR. The method is illustrated using DRIR data that have been measured with an EigenMike® (i.e. 32 pressure sensors located on the surface of a rigid sphere with a diameter of 8.4cm). Measurements were conducted in the variable acoustics performance hall of Ircam (*Espace de Projection, ESPRO*) for several positions of the emitter and for various acoustical configurations of the room.

## 2 DIRECTIONAL ROOM IMPULSE RESPONSES (DRIR)

DRIRs are typically captured with microphone arrays and preserve the spatial properties of the RIR. In recent years, higher-order SMAs have received increasing interest in room acoustic analysis as they offer several advantages over other array configurations. They are well suited to the analysis in the spherical harmonics domain, which eases the development of efficient algorithms for spatial filtering and beamforming.

### 2.1 Denoising of DRIR

Measured IRs cannot be directly used for auralization as they are typically corrupted by measurement noise. Convolving an anechoic signal with a raw RIR, can lead to undesired artefacts such as a non-decaying reverb, clearly audible after the actual room reverberation.

A restoration procedure for suppressing the noise of measured RIRs was presented by Jot [6]. Essentially, the method consists in substituting the undesired non-decaying noise with a white Gaussian noise filtered by a time-frequency envelope. The energy decay relief of each IR is computed and analyzed in order to estimate the reverberation slope as well as the initial power spectrum and time limit when noise becomes predominant. From these estimations, the time-frequency envelope is modeled so as to smoothly prolong the energy decay relief, keeping the characteristics of the reverberant decay unaltered (Figure 2).

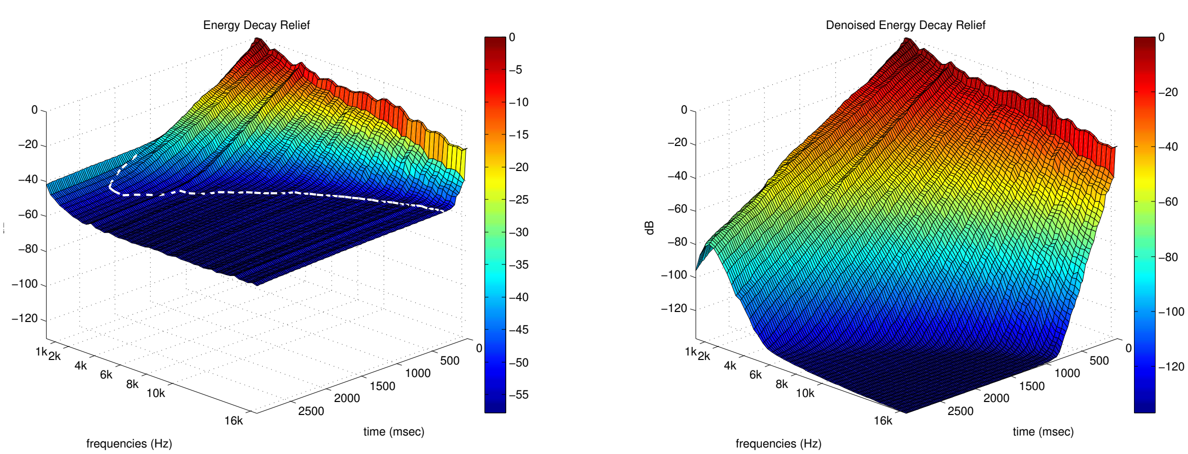


**Figure 2**: Energy decay relief (microphone #1): before (left) and after (right) denoising.

The dashed white curve in Figure 2 represents the estimated time limit before the noise predominates. The remaining noise floor on the right figure corresponds to the quantization noise (in this example, 24 bits result in -144 dB quantization noise).

For SMA DRIR measurements denoising could be successively applied to the different sensor signals, using a new realization of the Gaussian noise for each sensor. However, the synthetic noise would thus be uncorrelated between the different sensors, regardless of their position in space, and would violate the spatial properties of a diffuse field. In order to satisfy such spatial characteristics, a synthetic diffuse field is simulated and virtually recorded by the different sensors of the SMA. It is computed as the superposition of uncorrelated plane waves with unit amplitude and random phase $\phi(k)$ impinging from $L$ directions $(\theta_l, \varphi_l)$ on the SMA. The sound pressure on a microphone array of radius $a$ can then be expressed[7] as

$$p(k,r,\theta,\varphi)_{|r=a} = \sum_{l=1}^{L}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} H_n(k,r,a)_{|r=a}\, Y_n^m(\theta,\varphi)\, Y_n^m(\theta_l,\varphi_l)^*\, e^{2i\pi\phi_l(k)} \tag{1}$$

where superscript $^*$ denotes the conjugate, $i=\sqrt{-1}$ the imaginary number, $k$ the wavenumber, $Y_n^m(\theta,\varphi)$ the spherical harmonics. $H_n(k,r,a)$ is the holographic function for a rigid sphere:

$$H_n(k,r,a) = 4\pi i^n \left( j_n(kr) - \frac{j_n{}'(ka)}{h_n{}'(ka)} h_n(kr) \right) \tag{2}$$

$h_n(ka)$, $j_n(ka)$ denote the spherical Hankel and Bessel functions of the first kind, respectively.

In order to check the consistency of the synthetic diffuse field, the coherence function $\psi(p_1, p_2)$ between two points in space is computed for various values of $L$. According to Kuttruff[8], the spatial coherence of the sound pressure signals $p_1$ and $p_2$ in an ideal diffuse field at distance $d$ calculates to

$$\psi(p_1,p_2) = \frac{\sin(kd)}{kd} \tag{3}$$

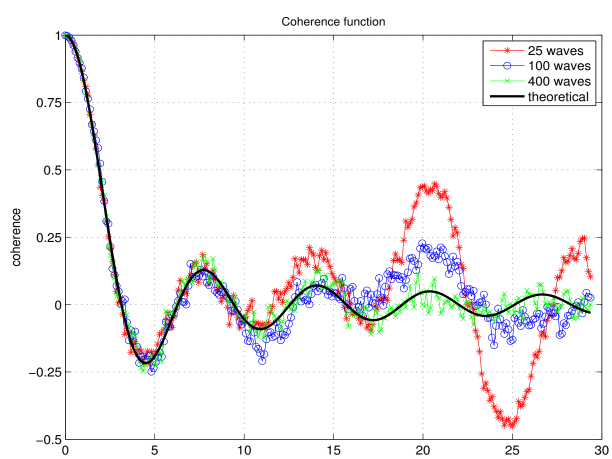


**Figure 3**: Coherence function for various values of $L$ (number of incoming plane waves).

As it can be see from Figure 3, a value of $L = 400$ (green curve) can be considered as being sufficient to generate a coherence profile that conforms to that of an ideal diffuse field (black curve). The decaying time-frequency envelope is then applied to this reference diffuse field, in order to create a synthetic late reverberation decay for denoising the DRIRs.

## 2.2 Space-time-frequency representation

Plane-wave decomposition[9] (PWD) eases the directional analysis of sound fields measured by SMAs. PWD results in the spatial amplitude density function of the plane waves composing the sound field that can be computed by

$$\Delta(k,\theta,\varphi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{1}{H_n(k,r,a)_{|r=a}} \left( \int_0^{2\pi}\int_0^{\pi} p(k,a,\theta,\varphi) Y_n^m(\theta,\varphi)^* \sin\theta \, d\theta \, d\varphi \right) Y_n^m(\theta,\varphi) \tag{4}$$

Sampling the sound pressure field on the sphere with $M$ microphones the integral in (4) can be approximated by a discrete summation and the summation in (4) is truncated to a finite order $n \le N$. This limits the angular resolution of the PWD, which can be approximated[9] by $\Theta \approx \pi / N$. **Figure 4** depicts the space-time-frequency representation of DRIRs recorded with an EigenMike® in ESPRO. The PWD was calculated with 64-sample frames (1.5 ms) and an overlap of 75%. For sake of simplicity, data are represented only in the horizontal plane. The observed frequency is 3500Hz. In this example, the direct sound can be clearly identified arriving from azimuth 0° at time t=0. The limited spatial resolution of the PWD results in a relatively broad main lobe (that blurs the representation of the direct sound) and non-negligible “side lobes”. Further, one can distinguish several early reflections: ground floor reflection, arriving at about 2ms after the direct sound, strong lateral wall reflections (azimuth ± 50°) arriving at time t≈23ms, and so on. The sound field becomes progressively more and more diffuse after about 100ms. All these observed phenomena are very consistent with the actual geometry of the room and the source/receiver positions.

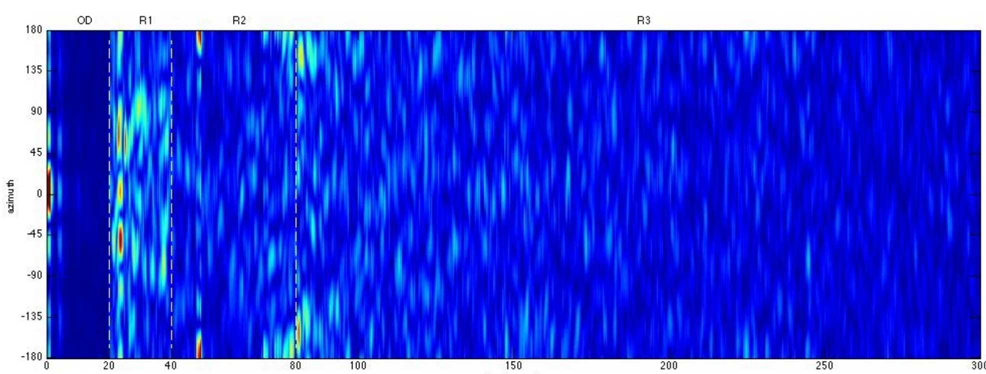


**Figure 4**: Example of a space-time-frequency visualization of DRIRs obtained by PWD analysis.

## 3 ANALYSIS/RESYNTHESIS AND SPATIO-TEMPORAL FILTERING

This section presents techniques for parametric spatio-temporal windowing of a DRIR in order to modulate its space-time-frequency distribution.

### 3.1 Beamforming in the spherical harmonics domain

Beamforming techniques enhance the sound signal from a target direction while attenuating signals from other directions therefore creating a spatial filter. The design of beamforming filters for SMA is presented in[10,11,12]. The sound pressure on the surface of the sphere is multiplied with a weighting function $w(k,\theta,\varphi)$ to produce the array output $y(k)$:

$$y(k) = \int_0^{2\pi}\int_0^{\pi} w(k,\theta,\varphi)p(k,a,\theta,\varphi)\sin\theta \, d\theta \, d\varphi \tag{5}$$

The weighting function is usually derived in the modal domain, i.e. the spherical harmonics domain for SMA processing. To do so, the spatial pressure function $p(k,a,\theta,\varphi)$ and the spatial weighting function $w(k,\theta,\varphi)$ are substituted by their spherical harmonics expansion $p_{nm}(k)$ and $w_{nm}(k)$, respectively. Taking the spherical Fourier transform of (5) and using the orthonormality of the spherical harmonics, the array output may be written as:

$$y(k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} w_{nm}(k)p_{nm}(k) \tag{6}$$

The spherical Fourier transform of the sound pressure on a rigid sphere, due to a plane wave with unit amplitude arriving from direction $(\theta_s,\varphi_s)$, can be expressed[7,10] as

$$p_{nm}(k) = H_n(k,r,a)_{|r=a} Y_n^m(\theta_s,\varphi_s)^* \tag{7}$$

where $H_n(k,r,a)_{|r=a}$ denotes the holographic function (see Equation (3)) evaluated at the sphere radius. The number *M* of available microphones typically limits the achievable order $N$ (with the constraint that $(N+1)^2 \le M$ ). To avoid aliasing artefacts we assume in the following that the coefficients beyond order $N$ are zero. The array weights for an axisymmetric beam pattern around the steering direction $(\theta_s,\varphi_s)$ write as[10,11]

$$w_{nm}(k) = \frac{b_n Y_n^m(\theta_s,\varphi_s)}{H_n(k,r,a)_{|r=a}} \tag{8}$$

where $b_n$ controls the directivity function of the array. For instance $b_n = 1, \forall n \le N$ is a widely used weighting function and is typically referred to as a plane-wave decomposition beamformer. The derivation of the array weights usually requires further treatments to avoid two phenomena: divergence of the holographic function at low frequencies and spatial aliasing at higher frequencies. This can be avoided by applying order-$n$ dependent band-pass filtering to the weighting function $w_{nm}(k)$, keeping only the relevant range of spatial frequencies $ka$ (Figure 5).

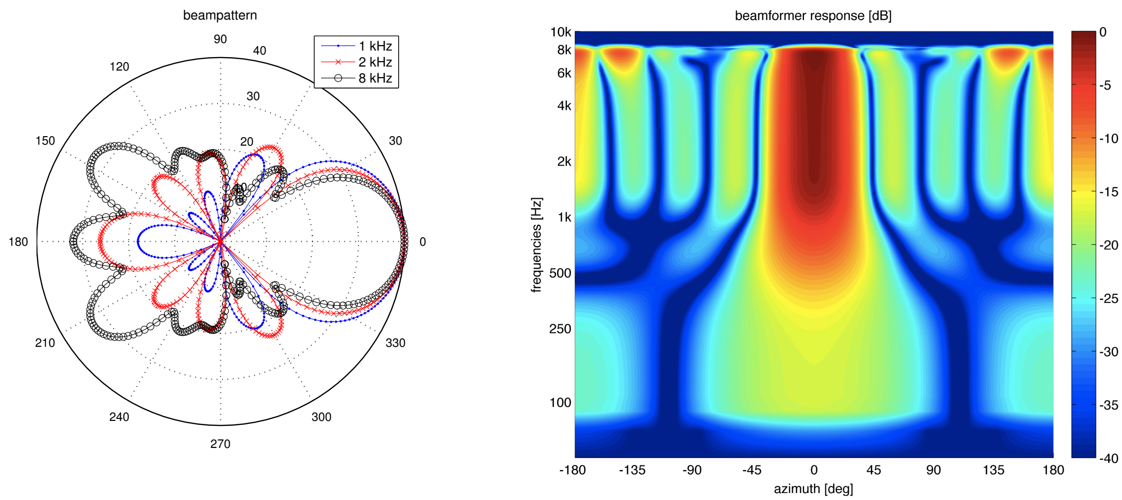


**Figure 5**: Example beam pattern for a modal beamformer. Spatial aliasing occurs above approximately 5000 Hz given a sphere radius $a = 0.042$ meters (EigenMike®) and $N = 4$.

Once the weighting function for a desired beam pattern and steering direction has been calculated in the modal domain beamforming filters (for real-time processing) can be derived in the spatial domain. They are computed by evaluating the inverse spatial Fourier transform:

$$w(k,\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} w_{nm}(k) Y_n^m(\theta,\varphi)^* \tag{9}$$

### 3.2 Analysis/Synthesis

*Analysis*

In this section an analysis/synthesis procedure for sound fields captured with a SMA is introduced. First, the measured DRIRs are processed with spherical beamforming algorithms to extract information from different room directions. The beamformer acts as a virtual microphone pointing to these directions. In order to preserve the spatial information originally captured by the SMA, the measured sound field should be sampled by an appropriate number of virtual microphones. Given the SMA order ( $N = 4$ ), the beams should be steered in at least $B = (N+1)^2 = 25$ directions evenly distributed on the 3D space. The $B$ beamformer output signals constitute a channel-based coding of the DRIRs. In the following examples a “plane-wave decomposition” beamformer has been used. This beamformer achieves the maximum directivity index and approximates the plane-wave amplitude density function. In other words, the DRIR measured by such a beamformer can be represented by a combination of plane-wave components.

Figure 6 represents the space-time PWD (in the horizontal plane only) of the original DRIRs. The DRIR used in this example corresponds to an emitter located at azimuth +80°. The space-time PWD is split in four temporal parts (OD, R1, R2, R3) according to the time segmentation presented in Section 1.2. With reference to Section 3.2, accurate analysis is limited to the operating frequency range of the beamformer. As a consequence the PWD analysis and the following figures only depict the valid frequency range (from 1000 Hz to 5000 Hz).

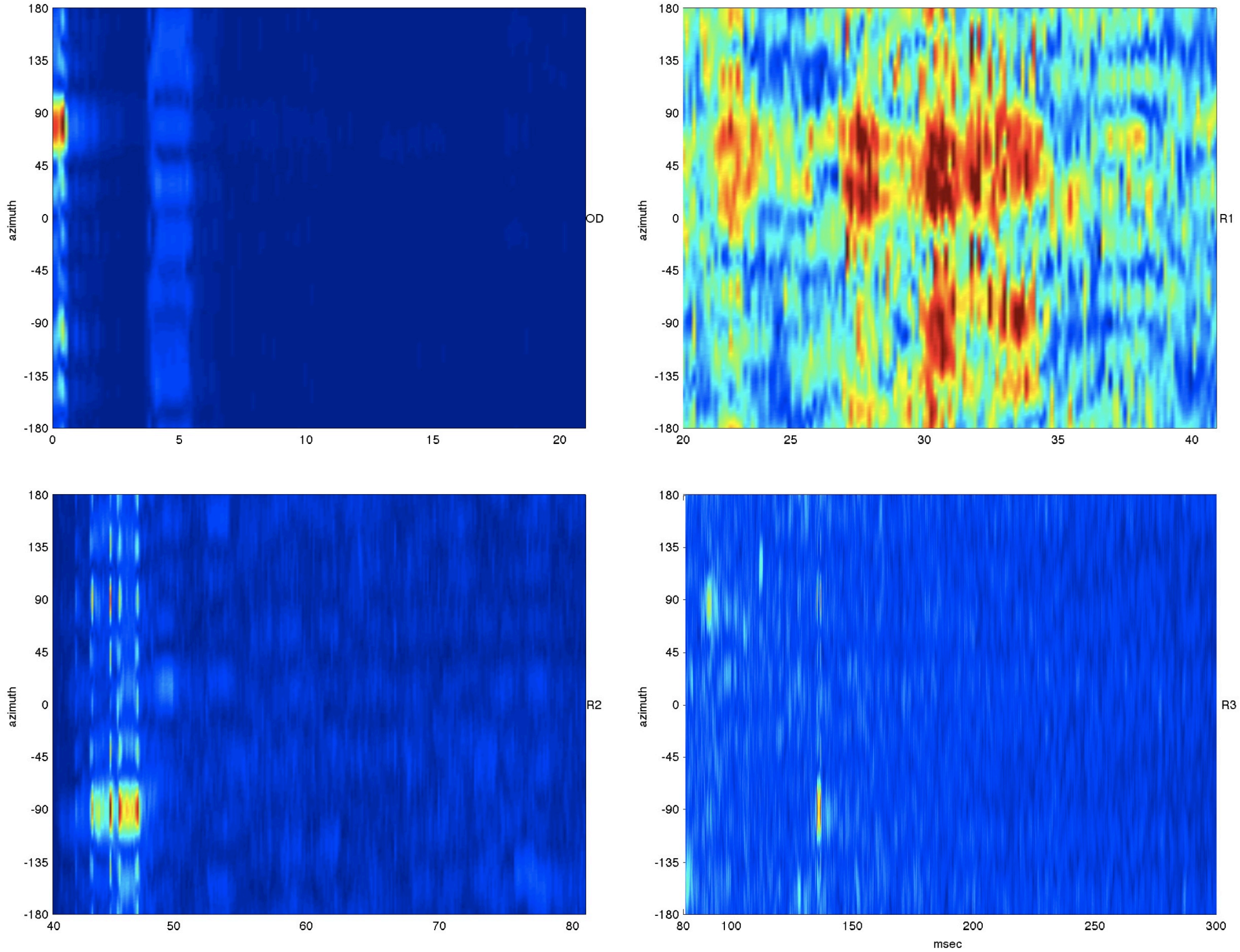


**Figure 6**: Space-time analysis of DRIRs using PWD: OD (top left), R1 (top right), R2 (bottom left) and R3 (bottom right).

*Synthesis*

Based on the analysis data, a sound field with similar spatial and energetic characteristics can be synthesized. This is achieved by creating $B$ plane waves which directions of arrival (DOA) are chosen accordingly to the steering directions. Higher-Order Ambisonic (HOA) rendering is a natural choice to reproduce these $B$ beamformer signals as it intends at creating plane waves. To achieve an accurate reconstruction of the original sound field, the spatial resolution of the reproduction system should match the order of the analysed sound scene i.e. the order of the HOA setup should be at least equal to $N$ .

Figure 7 illustrates reconstruction artefacts due to an insufficient number of steering directions. This results in angular energy fluctuations that can be clearly seen in the late decay (R3) of the reconstructed sound field. This reveals that a certain number of plane waves are required to recreate a sufficiently diffuse sound field. In addition the direction of arrival of the direct sound (OD) is shifted to the nearest reproduction direction (azimuth ≈90° instead of 80°).

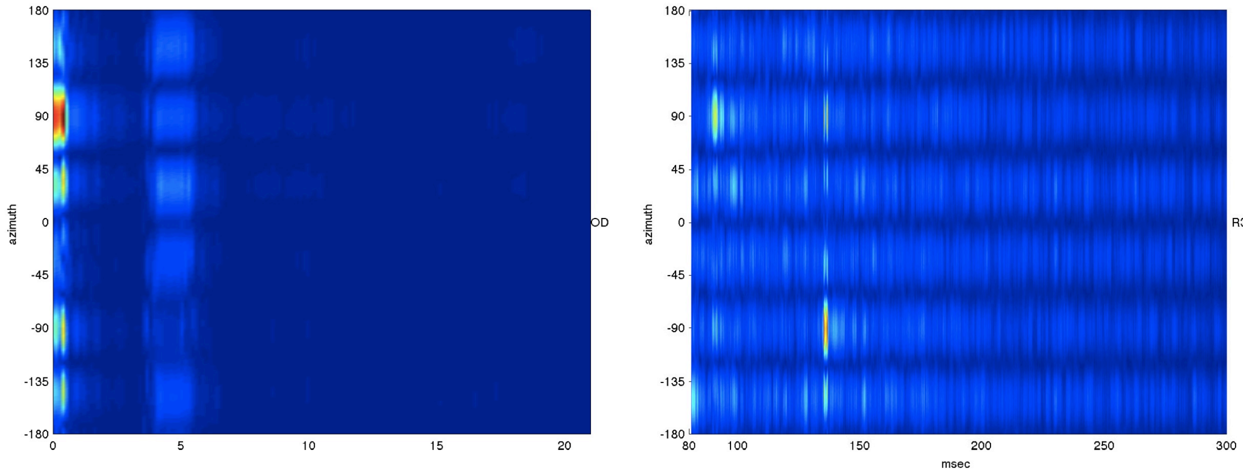


**Figure 7**: Space-time analysis of DRIR resynthesized from six beams (in the horizontal plane): OD (left), R3 (right).

Assuming a sufficient number of analysis beams a very good approximation of the original field (Figure 8, top and bottom left) can be achieved when the order of the reproduction system is equal or superior to the analysis order. Lower reproduction orders result in a spatially blurred reconstruction (Figure 8, top right).

*Spatio-temporal filtering*

The analysis/synthesis procedure introduced in the previous section allows for spatio-temporal filtering of DRIRs: amplitude gains or filtering (in the frequency domain) can be applied on one or several channels of the channel-based representation of the DRIRs. This transformation only affects the spatial area covered by filtered beams. Furthermore the transformation can be restricted to a given time frame. Figure 8 (bottom left and right) demonstrates a very basic example of such a spatio-temporal filtering: a gain envelope is applied in the 26—28 ms time frames on the two beams steering to azimuths 45 and 90°. A similar operation is repeated for the time frames 28—35 ms on the beams pointing to -45° and -90°. This procedure thus allows parametrically dimming or enhancing reflections located in any space-time-frequency region of the DRIRs.

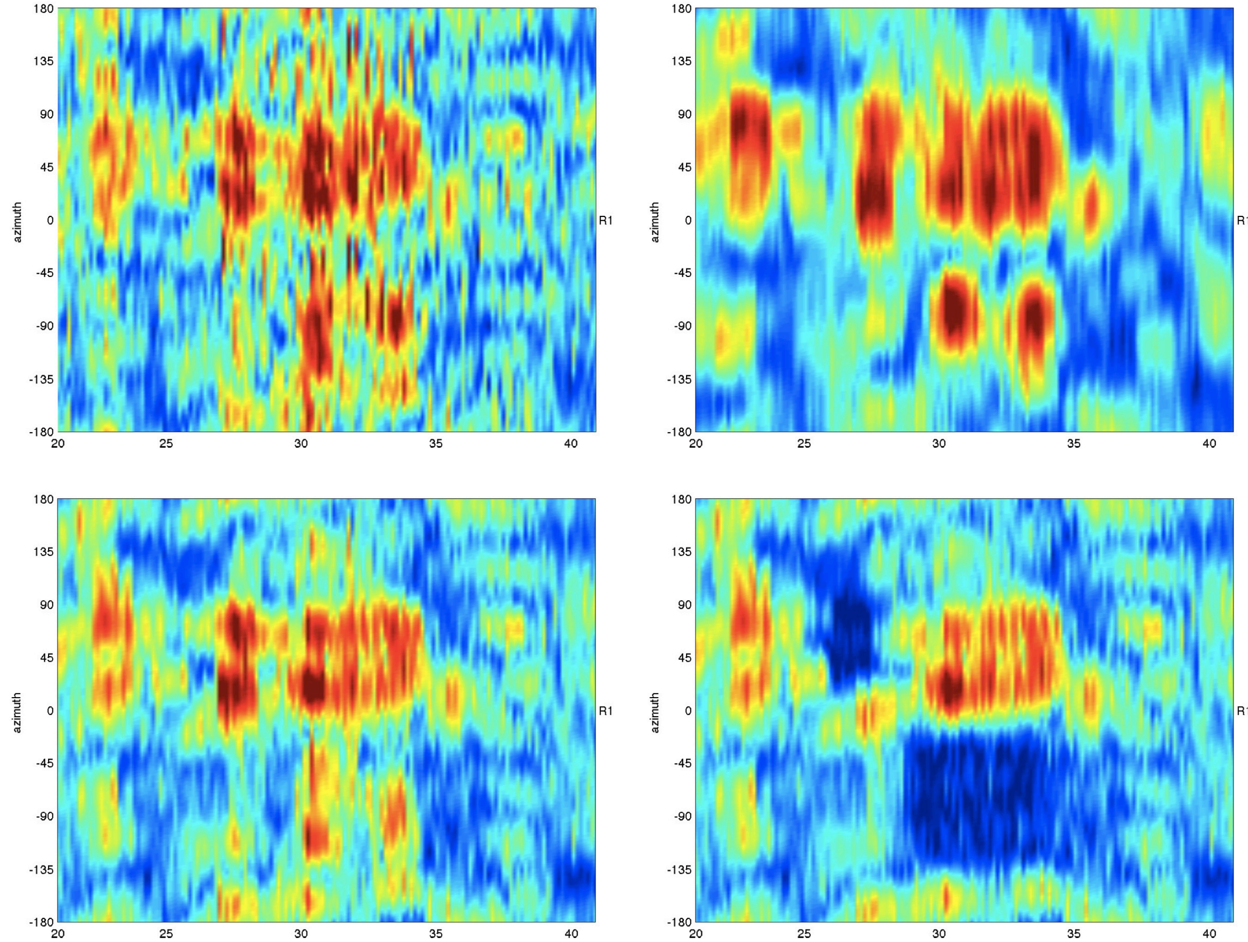


**Figure 8**: Space-time analysis of R1 segment: original (top left), resynthesized with 4$^{th}$ order reproduction setup (bottom left), resynthesized with 3$^{rd}$ order reproduction setup (top right), resynthesized with 4$^{th}$ order reproduction setup and filtered (bottom right).

## 4 CONCLUSIONS

This paper presented a framework for manipulating DRIRs recorded with a higher-order spherical microphone array in the context of convolution-based reverberators. The method is based on an analysis-synthesis procedure that samples the sound field through beamforming and provides a space-time-frequency representation of the room response. The beamforming is conducted in the spherical harmonics domain.

A technique for synthesizing a spatially diffuse noise by superposition of uncorrelated plane waves has been introduced. It can be applied to the denoising of the late decay of DRIRs recorded by a SMA.

The study outlined the conditions for accurate sound field reproduction, which can be achieved provided that the number of beams is coherent with the order of the SMA and that the order of the rendering setup is equal or superior to that of the SMA.

Space-time-frequency encoding of the sound field allows for further manipulations such as gain control or filtering of selected spatial and time windows. This method can enhance the parametric control of convolution-based reverberators and provides means for tuning the measured DRIRs along objective and perceptual descriptors related to spatial dimensions.

## ACKNOWLEDGMENTS

The authors would like to thank Orange Labs Lannion for providing the EigenMike® used for the measurements. This study has been supported by the French ANR CONTINT project "Sample Orchestrator 2" (SOR2).

## REFERENCES

1 William G. Gardner, "Efficient convolution without input-output delay", in Journal of the Audio Engineering Society, Vol.43, No.3, March 1995.

2 Guillermo Garcia, "Optimal filter partition for efficient convolution with short input/output delay", in Proceedings of the 113th AES Convention, Los Angeles, 2002.

3 John Stautner and Miller Puckette, "Designing multi-channel reverberators", in Computer Music Journal, Vol.6, No.1, Spring 1982.

4 Jean-Marc Jot and Antoine Chaigne, "Digital delay networks for designing artificial reverberators", in Proceedings of the 90th AES Convention, Paris, 1991.

5 Jean-Marc Jot, "Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces", in ACM Multimedia Systems Journal, Vol.7, No. 1, 1999.

6 Jean-Marc Jot, Laurent Cerveau and Olivier Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model", in Proceedings of the 103th AES Convention, New York, 1997.

7 Earl G. Williams, *Fourier Acoustics: Sound radiation and nearfield acoustical holography* (Academic Press, 1999).

8 Heinrich Kuttruff, "Room Acoustics – Fourth Edition," *Spon Press, New York* (2000).

9 Boaz Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution", in Journal of the Acoustical Society of America, Vol.116, No.4, October 2004.

10 Boaz Rafaely, Yotam Peled, Morag Agmon, Dima Khaykin and Ethan Fisher, "Spherical microphones array beamforming", Chap.11 in *Speech Processing in Modern Communications*, edited by Israel Cohen, Jacob Benesty and Sharon Gannot (Springer-Verlag, Berlin, 2010).

11 Jens Meyer and Gary W. Elko, "Spherical microphones array for 3D sound recording", Chap.3 in *Audio Signal Processing For Next-Generation Multimedia Communication Systems*, edited by Yiteng Huang and Jacob Benesty (Kluwer Academic Publishers, 2004).

12 Haohai Sun, Edwin Mabande, Konrad Kowalczyk and Walter Kellermann "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing", in Journal of the Acoustical Society of America, Vol.131, No.4, April 2012.