Audio Engineering Society

Convention Paper

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Sound Source Radiation Synthesis : from Stage Performance to Domestic Rendering

O. Warusfel, N. Misdariis

Ircam, Paris, Room Acoustics Laboratory

Correspondence should be addressed to olivier.warusfel@ircam.fr .

ABSTRACT

A diffusion device based on a digitally-controlled 3D array of loudspeakers was developed in order to synthesize a given radiation pattern from the combination of a set of elementary directivities. This radiation synthesis method, designed for musical and performance purposes (real-time control, musical vocabulary associated to different directivity patterns, …), has been used in concerts and sound installations. In order to transpose the sound experience from stage to domestic contexts, the paper addresses the post-production step where the spatial sound image associated to the radiation synthesis has to be trans-coded for conventional playback setup like stereo or 5.1. The method is based on the acoustical characterization of the performance situation with the different elementary directivities which are then combined according to the musical score.

1. INTRODUCTION

Flexible control on the directivity of a sound reproduction system emerges as a significant feature in several audio domains : sound reinforcement, public address, room acoustics measurements, etc… The focus adopted in this paper originates from applications of digitally controlled radiation developed in contemporary music. On this purpose, an electro-acoustical device was designed in an instrument-like approach. It consists of a 3D loudspeaker array confined in space (typically laid out on the facets of a polyhedron), the directivity of which can be shaped through signal processing. In this application, the electro-acoustic device is not distributed in front of the stage or around the audience. It is rather used as a conventional instrument, standing on stage so that the sound it produces (sound samples, spectral or physical model of instruments,...) is merged with other instruments and undergoes the same relationship with the room effect instead of aiming at its substitution. Together with the model used for radiation synthesis, the paper describes the different steps of audio production dedicated to such a device. Special attention is paid on using a sound format, dedicated to directivity control, and as independent as possible from the technical options used to build the reproduction device. This guaranties that a same content is compatible with different versions of the hardware, including future transducer technology or processing. Going further into this direction, a post-production step was used in order to allow for a domestic rendering on conventional set-up with high fidelity of the spatial dimension of the *live* performance.

2. DIRECTIVITY SYNTHESIS

2.1. Principle

The principle considered for directivity synthesis relies, in a general way, on the following methodology : describing the target sound field and the elementary contributions of the transducers on a same spatial boundary separating the source and the reproduction domains, and then applying mathematical or signal processing optimisation in order to derive the best set of filters for approximating the equalization of the sound field on this boundary.

Actually, the model is a composite source built from the combination of N elementary sources (transducers), each producing the field $P_i(\vec{r}_o,\omega)$ on the surrounding surface S (see Figure 1). The decomposition of T on the set of $\{P_i\}$, is expressed by an equality condition defined on the boundary:

$$\sum_{i=1}^{N} a_i(\omega)\cdot P_i(\vec{r}_0,\omega) = T(\vec{r}_0,\omega) \quad ,\forall \vec{r}_0 \in S \qquad (1)$$

where $\{a_i\}$ are complex coefficients depending on frequency ω. If existing, the $\{a_i\}$ will design the filters to be applied in input of the different transducers of the device.

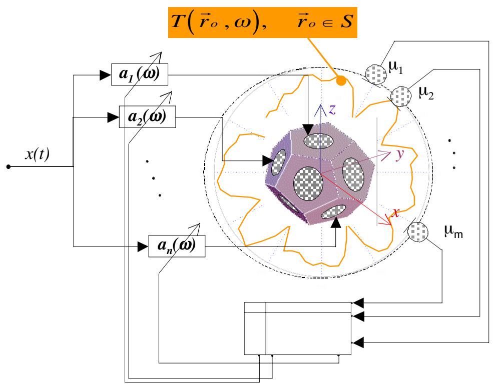

Figure 1 : *Source radiation synthesis*

In practice, the set of $\{P_i\}$ is generally not a base of the free-field radiation problem, hence the existence of $\{a_i\}$ is not ensured. The method only allows for an approximation of T and equation (1) is turned into a minimization of the least mean square reconstruction error $\left\| T-\sum a_i P_i \right\|^2$. This objective is equivalent to telling that, at each frequency, the reconstruction error is orthogonal to the vectors formed by the different P_i, as expressed in equation (2).

$$\left\langle \left(T-\sum a_i P_i\right) \middle| P_i \right\rangle = 0 \quad , \text{ for each } \omega. \qquad (2)$$

where $\langle \cdot | \cdot \rangle$ denotes the scalar product.

In practice, this optimisation is driven after spatial sampling of the target directivity T and of the different P_i measured or characterized according to the microphone grid $\{\mu_i\}$. The method is conducted iteratively in the frequency domain and results in a set of filters $\{a_i\}$ for a given source radiation pattern.

2.2. Implementation optimisation

On a practical point of view, this implementation presents strong limitations with regard to audio and musical applications. On the one hand, the set of filters associated to a target directivity depends obviously on the characteristics of the loudspeaker array : number and geometrical lay out of the transducers, together with their individual electro-acoustic response. This does not allow for storing the different audio channels in order to be played back on different devices. On the other hand, musical applications often require flexibility, real-time modifications and morphing functions among a collection of presets. All these requirements imply interpolation process on predefined set of filters, which may bring signal processing artefacts. For example, some implementations do not guaranty the filter stability during the interpolation or noise-free commutation [1]. Moreover, interpolating between two set of filters does not automatically ensure the perceptual continuity nor relevance of the produced acoustical and audio effect.

A more convenient architecture consists in synthesizing permanently a subset of filters associated to elementary radiation patterns from which composite directivities will be further generated through simple weighting functions i.e. gains (Figure 2). This method based on a spatial decomposition of the radiation pattern has been already adopted in various audio applications such as Ambisonic format or multi-channel implementation of binaural rendering [4]. As stated in [1] this implementation leads to a two step process where the directivity is first "encoded" into an intermediate multi-channel format, and then decoded on the reproduction system using a set of "decoding" or said "reconstruction" filters. The Figure 2 shows the different encoding gains controlling the spatial decomposition of the targeted radiation, and the reconstruction filters here denoted $H0_i$, $H1x_i$, etc... Each reconstruction filter consists of *n* filters feeding the *n* transducers of the loudspeaker array. These filters have been derived according to the method described in §2.1 in order to approximate the associated elementary directivity.

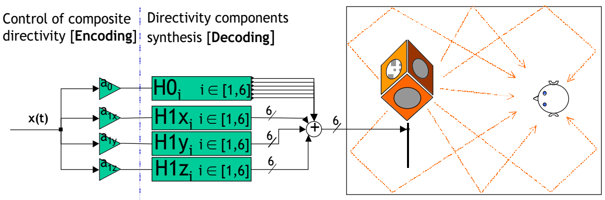

Figure 2 : Implementation of a spatial encoding of the directivity using spatial decomposition. Each reconstruction filter ($H0_i$, $H1x_i$, etc...) consists of *n* filters associated to the *n* transducers of the loudspeaker.

The choice of the spatial decomposition may depend on its efficiency (minimization of the reconstruction error for the targeted family of directivities) or its generic potential. In our case, the spherical harmonics appear the most convenient choice for they provide a structured frame with regard to important theoretical and practical concerns. Among them, an "intuitive" ordering according to the radiation pattern complexity (spatial periodicity) and simple analytical formulae used for monitoring elementary spatial transformations such as rotation or beam aperture control, etc...

The immediate advantage of such spatial decomposition implementation is the independence of the intermediate multi-channel directivity format with regard to the reproduction device. The description and the realization being de-coupled, the sound material is compatible with different hardware, including future generation using different number of drivers, different transducer technology or different signal processing implementation.

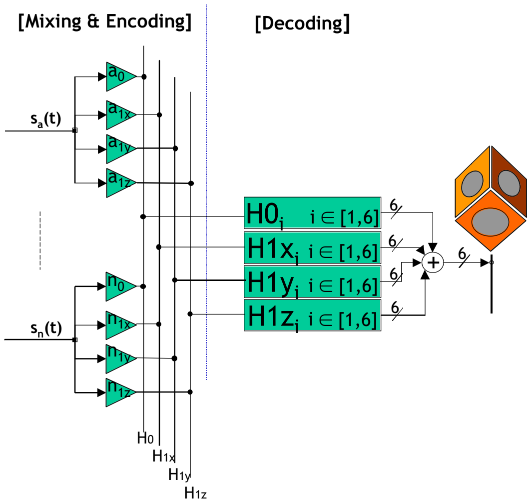

Figure 3 : Encoding simultaneous audio streams with different directivities using the multi-channel directivity format and common decoding step.

As depicted in Figure 3 this implementation also allows for the mixing of simultaneous audio streams with individual directivity patterns. In other words, different input signals will be associated to different set of gains resulting in the creation of specific radiation patterns attached to these signals. On the perceptual point of view, this will contribute to "audio stream segregation" similar to the well known cocktail party effect where the segregation does not rely on the location of sources but on their associated directivity. The perceptual relevance of the different elementary pattern has been studied and discussed in [7]. Moreover, this mixing is very efficient in terms of signal processing resources since most of the calculation is confined in the reconstruction filters which do not depend on the number of input streams.

Another example exploiting this architecture is given in Figure 4 where the input stream is split into different frequency bands, each one being processed separately in order to synthesize a frequency dependent directivity. A foreseen application of this principle, called *spatial additive synthesis,* is suggested in [5]. The signal output of the different vibration modes of an instrument, simulated with modal synthesis approach, are associated to their respective radiation pattern and reproduced by a 3D loudspeaker array. Instead of adding the different contributions in the signal domain and then playing back the resulting signal on a conventional loudspeaker, the synthesis is produced by adding up the different acoustic fields directly in the air medium.

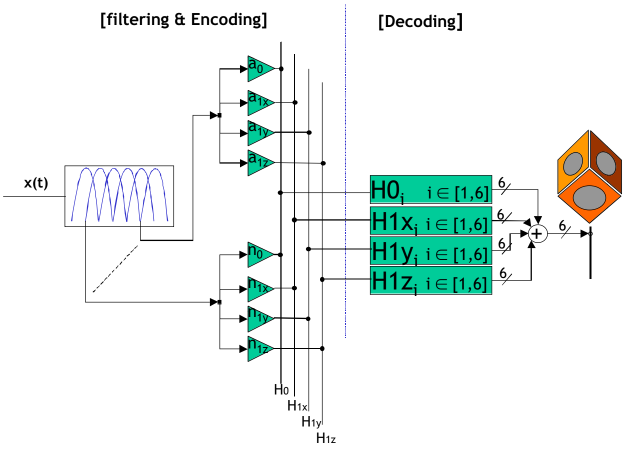

Figure 4 : Implementation of a frequency dependent directivity using the spatial decomposition format.

3. MUSICAL APPLICATIONS

3.1. Acoustic versus Electro-acoustic worlds

In most of the contemporary music productions, the sounds of live acoustic instruments coexist with electronic sounds delivered by loudspeakers hanging in front of the stage or surrounding the audience. As illustrated in [8] the radiation behaviour of acoustic instruments differ from each other, presents complex patterns and may even vary strongly for a same instrument according to the pitch, the fingering, etc... These acoustic features are transmitted and relayed by the room which results in perceptual differences from the listener standpoint. The current state of loudspeaker design does not allow reproducing these aspects of sound imaging and all electronic or recorded sounds are submitted to the radiation of a loudspeaker, i.e. a cardioid shape slowly varying with frequency. This makes difficult the balance between acoustic and electro-acoustic worlds which generally ends by amplifying the acoustic instruments even in small or medium size halls.

Another consequence is that, being typically more directive than most of the musical instruments, the loudspeakers tend to mask the response of the room. Actually, it is generally a pursued goal of the loudspeaker design in order to reduce feedback or echo problems in sound reinforcement systems.

From this point of view, the 3D loudspeaker array described in this paper tries to introduce a relevant alternative to this common approach. Built on the instrumental model (localised in space and spatially controllable) such radiation control device restores the interaction between the source and the room and could facilitate the balance between electronic and acoustic sounds.

3.2. Musical possibilities of the system

A collaboration with composers was undertaken in order to explore the validity of the proposed approach. Main aspects of the collaboration were, first, the assessment of the perceptual relevance of the elementary components (the spherical harmonics) and, second, the relevance of their monitoring.

For this purpose, the DSP algorithm described in the previous paragraphs has been implemented in a real-time software environment (Max/MSP) and user interface controllers have been attached to drive the system. For example, Figure 5 shows elementary controllers that allow for monitoring of the shape (from omni-directional to dipole) and of the direction of the radiation pattern.

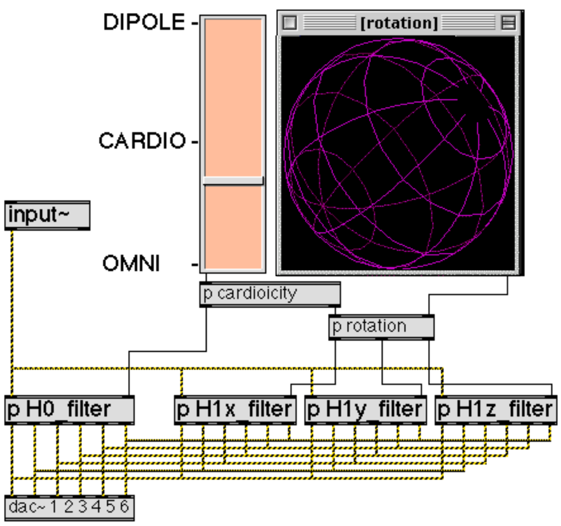

Figure 5 : Max/MSP Implementation and GUI

The slider drives the cardioicity of the pattern through the ratio between zero and first order harmonics, while the trackball controls the phase relation (relative weighting among harmonics belonging to the same order) between the three dipoles which results into driving the 3D orientation of the directivity pattern.

Moreover, specific tools were developed (temporal envelopes, random functions, constrains, ….) in order to propose different authoring situations, static or dynamic. Two of these situations are illustrated on *Figure 6* and *Figure 7*: respectively, a directivity morphing example where the radiation shape is modified in real time along the envelope of a sound (a gong hit), and a spatial counterpoint that consists in underlining the voices of a musical piece by assigning them different directivities.

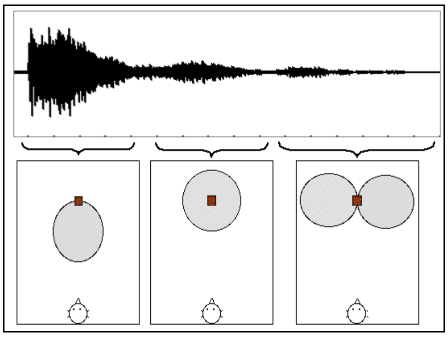

Figure 6 : Directivity morphing applied on the decay sound of a gong.

In both figures a symbolic representation of the directivity is provided. Such symbolic or textual description can be easily stored in a musical score or in a sequencer in order to be reproduced on time of the performance. Auxiliary tools used in sound design and sound production can be introduced for automation or algorithmic control such as slaving the radiation to an envelope or pitch follower.

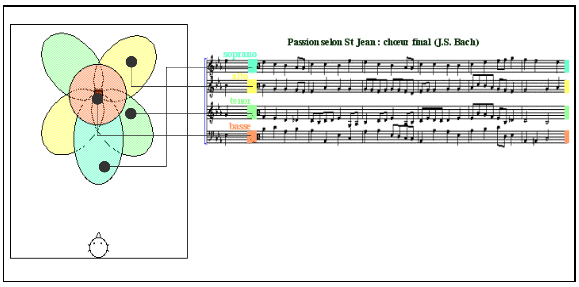

Figure 7 : Directivity counterpoint on a choral.

4. POST-PRODUCTION

4.1. Proposed method

This last section describes the method used for transmitting the performance experience to conventional reproduction systems such as domestic stereo or 5.1 set-up. As evoked in the previous sections, the effects of the radiation are perceived through the interaction between the direct sound and the room effect, i.e. through the modification of the perceptual attributes linked to room perception such as envelopment, apparent source width, source presence (auditory perception of source distance), etc...

One of the simplest methods for conveying these impressions to the listener is to use a convolution process based on a preliminary characterisation of the room with an impulse response. The room effect being dependent on the directivity of the source which is variable, it needs to collect a data base of responses allowing to reconstruct the directivity variations. Thanks to the implementation architecture described in the section 2.2 this data base can be restricted to the elementary radiation patterns used for the spatial decomposition of the directivity function.

Once getting these room responses, the auditory image of the *live* performance may be recomposed *a posteriori* by substituting the reconstruction filters with these responses and decode the sound tracks of the intermediate multi-channel format. The quality of the result mainly depends on the existence and spatial fidelity of the microphone technique used to characterise the sound field in the room.

A convenient choice for recording the impulse response is a binaural technique adapted to the reproduction on headphones or a stereo set-up via a further transaural decoding step. For multi-channel and surround set-up such as 5.1 a B-format or high order ambisonic microphone technique should be used in order to capture the spatial organisation of the sound field.

4.2. Practical aspects

As detailed in *Figure 8*, the measurement step consists in capturing the response of the concert hall for each elementary directivity pattern used for the spatial decomposition. The figure represents the case of a binaural measurement and can be easily replicated in the case of a B-format or higher order ambisonic measurement.

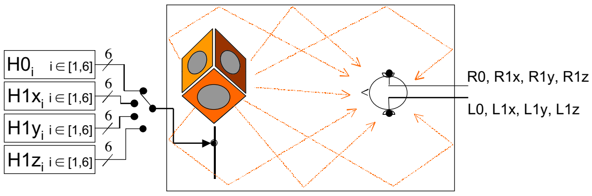

Figure 8 : Binaural measurement of the room effect associated to the different elementary directivities

The measured impulse response will usually need a denoising process based for example on a statistical time frequency model of the reverberant decay as presented in [2]. Starting from the analyse of late response decay the end of the decay that might have been truncated or corrupted by noise may be synthesised.

As described on *Figure 9*, the substitution of the decoding filters by the corresponding room impulse responses allows one to maintain the flexibility of control during the post-production step. In particular, it gives access to real-time monitoring by hand, sequencer or algorithms. An interesting feature is also that one can "listen" and "explore" the sound of the room response while varying the directivity of the source.

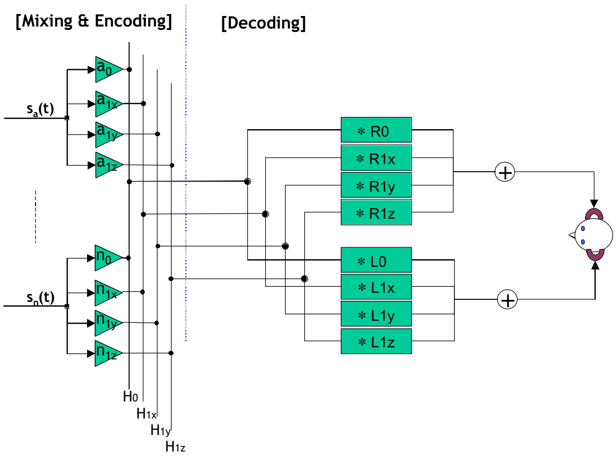

Figure 9 : Decoding the directivity on binaural simulation of the room response to the elementary directivities

5. CONCLUSIONS

An approach for controlling the directivity of a sound reproduction system, based on an encoding format upon a truncated spherical harmonics basis, has been implemented on prototypes gathering both hardware and software developments. The system has been explored for musical applications including stage performance situations. The method also covers the post-production step where the sound experience of the performance is conveyed to conventional audio formats using a data base of room impulse responses corresponding to the elementary directivities of the spatial decomposition.

6. REFERENCES

[1] Jot, J.-M., Wardle, S., Larcher, V., Approaches to binaural synthesis. Presented at the 105th Convention of the Audio Eng. Soc (1998) [preprint 4861].

[2] Jot, J.-M., Cerveau, L., and Warusfel, 0.. Analysis and synthesis of room reverberation based on a statistical time-frequency model. Presented at the 103rd Convention of the Audio Eng. Soc. (1997).

[3] Jot, J.-M., Larcher, V., and Warusfel, 0.. Digital signal processing issues in the context of binaural and transaural stereophony. Presented at the 98th Convention of the Audio Eng. Soc. (1995) [preprint 3980].

[4] Larcher, V., Jot, J-M., Guyard, J., Warusfel, O. Study and Comparison of Efficient Methods for 3D Audio Spatialization Based on Linear Decomposition of HRTF Data. Presented at the 108th Convention of the Audio Eng. Soc. (2000). [preprint 5097].

[5] Warusfel O., Corteel E., Misdariis N., Caulkins T., Reproduction of sound source directivity for future audio applications, International Congress on Acoustics, Kyoto, april 2004.

[6] Warusfel O., Misdariis N., Directivity Synthesis with a 3d Array of Loudspeakers : Application for stage performance, DAFX–01, Limerick, december 2001.

[7] Misdariis N., Warusfel O., Caussé R., Radiation Control on a Multi-Loudspeaker Device, International Symposium on Musical Acoustics, ISMA 2001, 2001.

[8] Warusfel O., Derogis P., and Caussé R., Radiation synthesis with digitally controlled loudspeaker, 103rd AES Convention, New-York, 1997.