AUDIO ENGINEERING SOCIETY
Italian Section

ANNUAL MEETING 2005

Paper: 05004

Como, 3-5 November
Politecnico di MILANO

WEIGHTED SOUNDFIELD SYNTHESIS: A PERCEPTUAL APPROACH

DAVIDE RENZI, RAFFAELE PARISI AND AURELIO UNCINI

Dept. of Scienza e Tecnica dell'Informazione e della Comunicazione – INFOCOM

University of Rome "La Sapienza"

Via Eudossiana, 18

I-00184 Roma – Italy

d.renzi@infocom.uniroma1.it, parisi@infocom.uniroma1.it, aurel@infocom.uniroma1.it

In this work an innovative sound field synthesis technique is introduced and described. Based on array processing concepts, the proposed methodology builds up the sound field by taking into account both the desired wavefront spatial distribution and psycho-acoustics features of a reference room. More specifically, psycho-acoustic effects are generated by use of a proper *ad hoc* filtering matrix. Results obtained on simulated data demonstrate that the proposed approach is able to perform accurate source positioning in the three-dimensional space by using a 2D loudspeaker array, thus resulting in a fully immersive acoustic behavior of the environment and realistic depth of the sound sources.

INTRODUCTION

In the last few decades significant progress has been made in the field of 3D-audio. Nowadays the availability of fast computers and developed DSP techniques makes it possible to reach very high performance in sound processing and rendering.

The new frontier is in the field of immersive audio and sound spazialization (auralization). The goal is the correct reconstruction of a complex acoustic scene played in a specific auditory space: the listener should experiment the same feelings that he/she would feel in a real scene, he/she should properly locate the sound source in the space, having the illusion of being in a real context.

This purpose has moved the research to extend the basic stereophonic system to the new multi-channel systems and to investigate new processing paradigms based on loudspeaker arrays (Wave Field Synthesis, Ambisonic) [1][2][3].

In this work we introduce a new point of view and develop a new concept of *acoustic field*. We deploy an artificial filtering matrix, synthesized by a few number of psychological and acoustical parameters, derived as an evolution of head-related transfer functions (HRTFs).

1 PERCEPTION OF SOURCE POSITION

In the past years a lot of efforts have been directed to study the modalities by which the listener is able to estimate the position of a source in the space [5][7][8].

In particular the listener's ear has been modelled like a filter. The corresponding frequency response has been obtained through measurements on a dummy head (i.e. KEMAR manikin) [11].

Such measurements demonstrate that a variation in a selected frequency band characterizes the elevation of the source (referred to an ideal plan passing through the ears and parallel to the ground). These frequency alterations are caused by pinnae, torso and shoulder of the listener. This means that a particular set of HRTFs are subject-dependent. Some research has been made to determine a few HRTF families, trying to reduce the subjectivity of the measurements, but experiments showed that those approximations hamper the correct perception of the source position. This can be attributed to various causes: firstly, the modalities through which the signal are synthesized and reproduced; secondarily, the integration of the information performed by the brain based on real perception and acquired experiences. Concerning the modalities used for the measurement, headphones should be employed in order to reproduce the correct signals. As a matter of fact it comes out that the listener cannot judge the correct source distance (externalisation). If a pair of loudspeakers is used instead, special care should be taken in maintaining the separation of emitted signals, trying to avoid mutual interference (cross-talk). Also in this case, even if sophisticated algorithms are adopted in order to reduce/kill the cross-talk, the risk of incorrect perception is present: sound sources are not located exactly (with the exception of simple planar configurations in which the corrected perception is allowed by the interaural level difference ILD and the interaural time difference ITD) and often the listener makes confusion between diametrically opposite positions.

These considerations lead us to closely consider the second cause of wrong perception: if the system [source]-[ear]-[brain] is assumed as the responsible of correct sound positioning, then it seems necessary to exactly define the type of information that the brain extrapolates and the modalities through which it is transformed in spatial details.

A more realistic model should take into account concurrent perceptions that complete the acoustical cues (i.e. intra-cranial resonances, physical perception of sound pressure, evolutions captured by the inner ear, etc.). Clearly it cannot be admitted that the signal filtered by the ears is auto-explicative, in the sense that it is not the single alteration introduced by the pinnae that carries out the spatial information that is converted by the brain into nervous impulses to be processed. Rather, it is logical to admit that it is the brain that elaborates the content of several nervous impulses and compares them with previous acoustic experiences. This justifies the ability of the man to adapt to different situations and to traumas of the auditory apparatus.

An alternative way to reproduce the correct source position is to regenerate the actual wave front that it radiates.

2 WFS THEORY

Wave field reconstruction can be obtained by synthesizing the wave front enclosed in the auditory space. The regenerated wave intrinsically carries information on the source (position, timbre) and on the environment. In order to obtain the reconstruction of a wave front in a sufficiently wide area we can proceed by using an array of properly placed loudspeakers. Signals emitted by every single loudspeaker are mixed up, thus supplying the desired synthesized wave. The very basic idea is founded on Huygens's principle in which wave diffraction is examined. The principle finds its mathematical formulation with the Kirchhoff-Helmholtz integral (see Figure 1):

$$P(\mathbf{r},\omega) = \frac{1}{4\pi}\oint_S F(\mathbf{r}_S,\omega)\cdot \mathbf{n}dS \tag{1}$$

where $F(\mathbf{r}_S,\omega)$ is

$$F(\mathbf{r}_S,\omega) = \\ = \underbrace{P(\mathbf{r}_S,\omega)\cdot\nabla G(\mathbf{r},\mathbf{r}_S)}_{dipoles} - \underbrace{G(\mathbf{r},\mathbf{r}_S)\cdot\nabla P(\mathbf{r}_S,\omega)}_{monopoles} \tag{2}$$

and $G(\mathbf{r},\mathbf{r}_S)$ is the Green's function:

$$G(\mathbf{r},\mathbf{r}_S) = \frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|} \tag{3}$$

while $k = \omega/c$ is the ratio between angular frequency and velocity of propagation (wave number).

From Eq.(1) derives that the field in the enclosed volume V is completely described by the knowledge of the pressure $P(\mathbf{r}_S,\omega)$ and its gradient $\nabla P(\mathbf{r}_S,\omega)$ captured along the shape S that envelopes V.

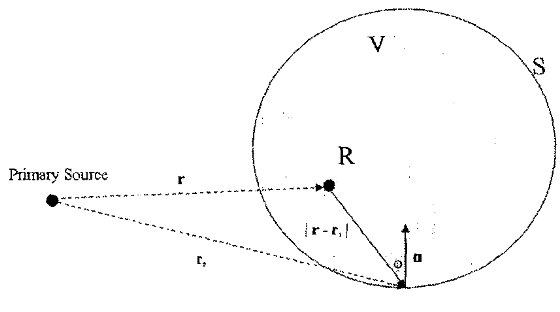
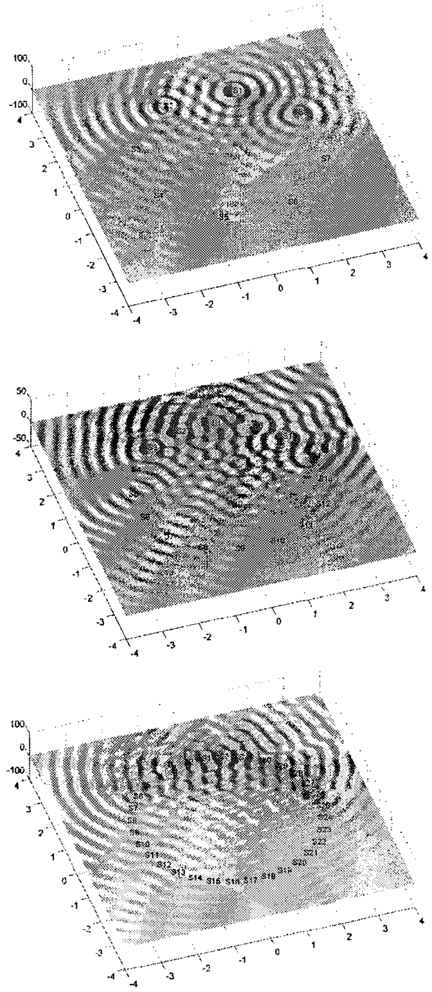


Figure 1: Geometrical description of the Kirchhoff-Helmholtz integral notation

Currently those measures can be extrapolated to drive the loudspeaker array by two different methodologies: *Wave Field Synthesis* (WFS) and Ambisonics.

These approaches are based on mathematical transformations applied to signals acquired by a set of microphones placed in discrete positions. This solution shows several limitations during the synthesis of the signals; ideally we must have an infinite number of measurements and a high number of speakers (Fig. 2).
For these reasons, synthesis models will not be able to catch up the faithful reconstruction of the real waves evolutions. Errors of coherence, approximation (spatial aliasing, truncation effects, ...), reproduction (loudspeakers do not have a spherical spectrum of radiation), environment (listening rooms with nonzero coefficients of reflection) makes it impossible to realize a "perfect/high fidelity regeneration" (Fig. 3).



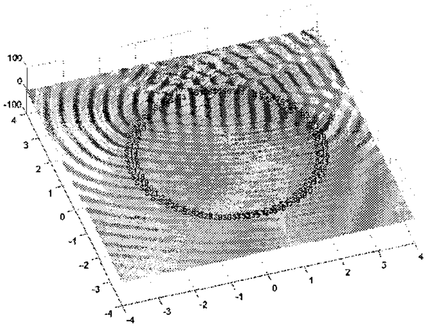


Figure 2: Reconstruction of the wave front emitted by a source positioned in the top right of the auditory space. Various sets of loudspeakers arranged in a circular array (2.5 meter ray).

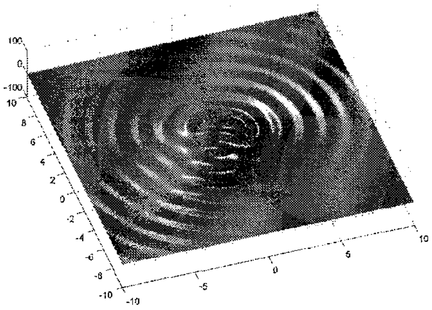


Figure 3: Example of spatial aliasing and truncation effects on the reconstructed wave front.

Theory must be adapted in order to catch up a sub-optimal solution that is based on the perception that must be recreated and not only on the physics wave that it models. Physical constraints impose a particular location for each speaker to preserve a correct WFS reconstruction (sensation) in a restricted portion of the auditory space. We studied the artificial manipulation that can be applied to the driving functions in order to obtain new perceptual sensations within synthesized wave front, thus overcoming mathematical and physical limitations.

3 PSYCHO-ACOUSTICAL FILTER

The positioning of sounds in space is obtained by comparing the time delays of arrival and the magnitude alterations in frequency bands at the listener ears (excluding the other sensorial perceptions).
The perception is the result of an integration of the signals perceived, while the displacement from the default values of the measurements can be determined as principal acoustic parameters (PAP).
In our model we use various anechoic sources positioned in the auditory space at different elevations.

We employ a planar array composed by a few number of loudspeakers. Our aim is to reproduce the elevation of the sources without using speakers located at the corresponding quota.
The source position is used to extract the parameters employed to guide the WFS engine. Each band of frequency for a defined speaker is filtered to obtain the simulated elevation of the source (psycho-acoustic effect). In the experiment we refer to the head related impulsive responses measured by the CIPIC Interface Lab.; the method used to extract the measures is described in the related documentation [11].

3.1 CONFIGURATION

Consider a set of N loudspeakers positioned in a planar configuration at the same quota ($z=0$) on the same line ($y=k$); we assume to centre the reference in the point where the listener is positioned.
In the real model the source S emitting the signal $s(t)$ is located in (x_S, y_S, z_S), while listener ears are in $(x_L, y_L, z_L) \equiv (0,0,0)$. We can calculate the signal $y^r(t)$ in (x_L, y_L, z_L) using the corresponding set of HRTFs. At this step we do not include in the model the alteration introduced by the environment (reflections, frequency distortions, ...). Thus we can write the relation in the frequency domain between source signal and received signal at loudspeaker:

$$Y^r(f) = S(f) \cdot H_{(S)}(f) \tag{4}$$

where $H_{(S)}(f)$ is the transfer function in S.
If we use a loudspeaker D positioned in (x_D, y_D, z_D) to emit the signal $s(t)$ we can write:

$$Y^a(f) = S(f) \cdot H_{(D)}(f) \tag{5}$$

where $H_{(D)}(f)$ is the transfer function in D.
Now we must use a filter to model the correct elevation of the source (in general different of the position of the loudspeaker):

$$Y^{r^*}(f) = Y^a(f) \cdot G(f) \tag{6}$$

We wish to obtain $Y^r(f) = Y^{r^*}(f)$, so from Eq. (5) and Eq. (6):

$$Y^{r^*}(f) = S(f) \cdot H_D(f) \cdot G(f) \tag{7}$$

From this and Eq. (4):

$$G(f) = \frac{H_S(f)}{H_D(f)} \tag{8}$$

So we can extend our considerations to a set of loudspeakers $D_1, D_2, \ldots, D_N$ positioned in $(x_{D_i}, y_{D_i}, z_{D_i})$ with $i = 1, \ldots, N$, and a set of sources $S_1, S_2, \ldots, S_M$ positioned in $(x_{S_j}, y_{S_j}, z_{S_j})$ with $j = 1, \ldots, M$.
We can calculate a weight matrix G so that

$$\underset{N\times 1}{Y} = \underset{N\times M}{G} \cdot \underset{M\times 1}{S} \tag{9}$$

where the entry $Y(i)$ is the driving function for loudspeaker D_i; the entry $S(j)$ is the source signal $S_j(f)$; the entry $G(i,j)$ is the function defined in (8) relative to the assigned source-loudspeaker's path.
The problem at this point is to choice the right family of HRTFs to determine the matrix coefficient. As described in the previous paragraph since we cannot use an absolute measure, we must introduce a statistical study of the variation of the HRTF due to the particular ear conformation.
We collected a set of HRTF from 45 subject analyzed by the CIPIC Interface Lab.; we used only the functions referred to the central position (0 degrees azimuth) at the elevation included in the range -45° +45° (Fig. 4); we normalized our measure with the central reference measure (0° azimuth, 0° elevation) for the same listener, in order to have an independent evolution of the measure at the different elevation that was not caused by the particular subject. The measure in decibel of the variations from the base elevation can be compared for the subjects. The evolution of the averaged value can be used to populate the G matrix.
Once the matrix is populated we must introduce it in the WFS system after the generation of the loudspeaker driving signal. In this way the human ear is tricked, and the listener has the impression of a virtual source positioned in a new plan (generally different from the array plan).
We use this filter matrix to increase the convergence of the WFS model via the correct perception of the position; this goal is reached by inserting "perceptive" aids coded in the signals emitted by the loudspeakers.

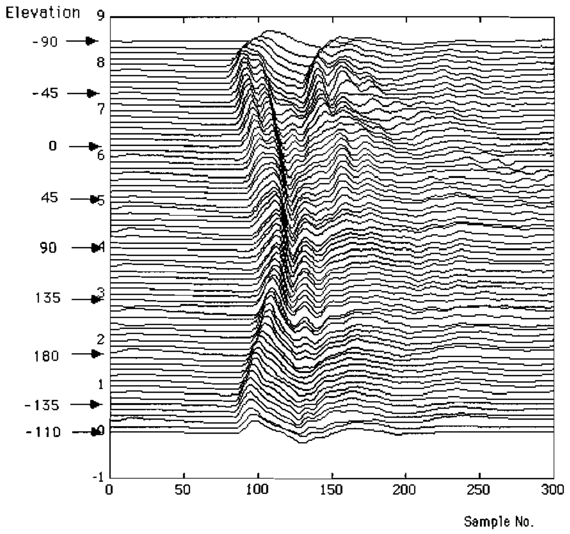


Figure 4: A possible set of HRIR in the median plane (http://www-engr.sjsu.edu/~duda/Duda.Fig4.html)

4 EXPERIMENTAL RESULT

In our experiments we used a different set of 8 loudspeakers positioned in various configurations. With planar (line) arrays is practically impossible to reproduce the correct environment feeling due to the absence of the reflections (speakers) caused by the surrounding walls. On the other hand it is possible to produce source elevation (in a very limited elevation angle) by playing a set of sounds at various distances and azimuth angles at the same time during the elevated source emission.

An absolute estimation of the sound elevation is very difficult when there are no environment background or tricking sounds.

With circular loudspeaker configurations it is possible to recreate an interesting surrounding sensation by filtering each single wave front with the relative weight, according to previous equations.

5 CONCLUSIONS

This paper has described a weighted version of the WFS theory by which it is possible to recreate psychoacoustic effects in the listener using frequency alterations. Source elevation via planar set of loudspeaker has been experimented by using HRTF differences between source and speakers.

The proposed approach will be further investigated by considering a larger set of loudspeakers and alternative frequency alteration to produce new acoustic effects.

REFERENCES

[1] A. J. Berkhout, "*A Holographic Approach to Acoustic Control*", J. Audio Eng. Soc., vol. 36, pp. 977-995, 1988.

[2] A. J. Berkhout, D. de Vries, P. Vogel, "*Acoustic Control by Wave Field Synthesis*", J. Acoustic Soc. Am., vol. 93, pp. 2764-2778, 1993.

[3] D. de Vries, E. Start, V. Valstar, "*The Wave Field Synthesis concept applied to sound reinforcement : Restrictions and solutions*", 96th AES Convention, Amsterdam, Feb. 1994.

[4] A. J. Berkhout, D. de Vires, J. J. Sonke, "*Array technology for acoustic wave field analusis in enclosures*", J. Acoustic Soc. Am., vol. 102, pp. 2757-2770, 1997.

[5] D. de Vries, "*Sound Reinforcement by Wavefield Synthesis: Adaption of the Synthesis Operator to the Loudspeaker Directivity Characteristics*", Laboratory of Seismics and Acoustics, Delft University of Technology, 1996.

[6] D. de Vries, M. Boone, "*Wave Field Synthesis and Analysis using Array Technology*", Delft University of Technology, IEEE workshop New York, 1999.

[7] E. Hulsebos, D. de Vries, P. Vogel, "*Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis*", Delft University of Technology, 110th AES Convention, Amsterdam, 2001.

[8] A. Sontacchi, R. Holdrich, "*Further Investigations on 3D Sound Fields using Distance Coding*", DAFX-01, COST G-6, 2001.

[9] M. Boone, "*Acoustic rendering with wave field synthesis*", ACM Siggraph and Eurographics campfire: Acoustic Rendering for Virtual Environments, Utah, 2001.

[10] S. Spors, H. Teutsch, R. Rabenstein, "*High-Quality Acoustic Rendering with Wave Field Synthesis*", University of Erlangen-Nuremberg, 2002.

[11] V. R. Algazi, R. O. Duda, D. M. Thompson, "*The CIPIC HRTF database*", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2001, New York, 2001.