Sound Scene Creation and Manipulation using Wave Field Synthesis

Etienne.corteel@ircam.fr
Terence.caulkins@ircam.fr

I. Introduction: An overview of Wave Field Synthesis

Wave Field Synthesis vs. Conventional techniques

Wave Field Synthesis (WFS) is a sound reproduction technique using loudspeaker arrays that postpones the limits set by conventional techniques (stereo, 5.1 ...). These techniques rely on stereophonic principles allowing the creation of an *acoustical illusion* (as opposed to an optical illusion) over a very small area in the center of the loudspeaker setup, generally referred to as "sweet spot". For example, in a stereo setup, if one slightly modifies the arrival time and/or intensity of the driving signal fed to one of the two loudspeakers, a correctly positioned listener will get the impression that a *virtual* source is situated somewhere between the two *physical* sources (i.e. the loudspeakers). However, if the listener moves closer to one of the two loudspeakers, the illusion collapses and the virtual source falls back onto the nearest loudspeaker.

Huyghens' Principle

WFS, on the other hand, aims at reproducing the true *physical attributes* of a given sound field *over an extended area* of the listening room. It is based on Huyghens' principle (1678) which states that the propagation of a wave through a medium can be formulated by adding the contributions of all of the secondary sources positioned along a wave front. To illustrate this, let us consider a simple example. A rock (or *primary source*) thrown in the middle of a pond generates a wave front that propagates along the surface. Huyghens' principle indicates that an identical wave front can be generated by simultaneously dropping an infinite number of rocks (*secondary sources*) along any position defined by the passage of the primary wave front. This synthesized wave front will be perfectly accurate outside of the zone delimited by the secondary source distribution. The secondary sources therefore act as a "relay", and can reproduce the original primary wave front in absence of a primary source!

Origins of Wave Field Synthesis

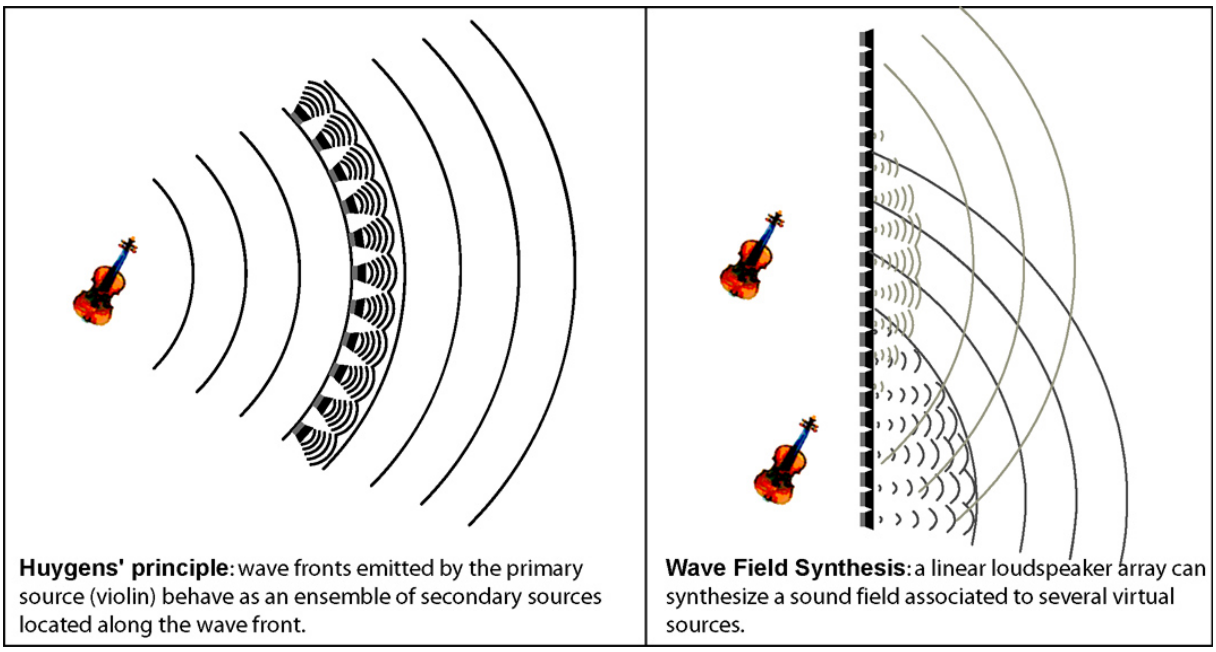
Wave Field Synthesis (WFS) [1][2][3][4] is based on a series of simplifications of the previous principle. The first work to have been published on the subject dates back to 1988 and is attributed to Professor A.J. Berkhout of the acoustics and seismology team of the Technological University of Delft (T.U.D.) in Holland. This research was continued throughout the 90's by the T.U.D. as well as by the Research and Development department of France Telecom Lannion.

Content-Coding

WFS relies on an object-based description of a sound scene. To obtain an object-based description, one must decompose the sound scene into a finite number of sources interacting with an acoustical environment. The coding of such a sound scene includes the description of the acoustic properties of the room and the sound sources (including their positions and radiation characteristics). Separate from this spatial sound scene description is the coding of the sound stream itself (encoding of the sound produced by each source).

WFS reproduction

Work conducted on the subject of Wave Field Synthesis has allowed for a very simple formulation of the reproduction of omni-directional virtual sources using a linear loudspeaker array. The driving signals for the loudspeakers composing the array appear as delayed and attenuated versions of a unique filtered signal. The maximum spacing between two adjacent loudspeakers is approximately 15 to 20 cm. This allows for optimal localization over the entire span of the listening area.



Huygens' principle: wave fronts emitted by the primary source (violin) behave as an ensemble of secondary sources located along the wave front.

Wave Field Synthesis: a linear loudspeaker array can synthesize a sound field associated to several virtual sources.

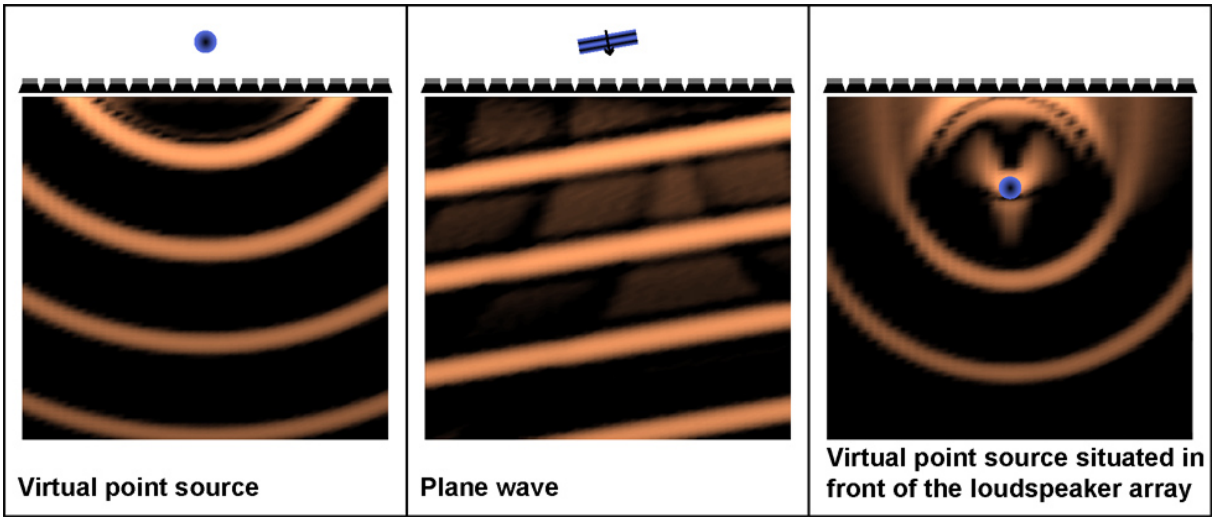
Elementary sources in Wave Field Synthesis

One can distinguish three separate types of virtual sources that are synthesizable using WFS systems:

- **Virtual point sources situated behind the loudspeaker array.** This type of source is perceived by any listener situated inside of the sound installation as emitting sound from a fixed position. The position remains stable for a single listener moving around inside of the installation.
- **Plane Waves.** These sound sources are produced by placing a virtual point source at a seemingly “infinite” distance behind the loudspeakers (i.e. at a very large distance in comparison to the size of the listening room). Such sources have no acoustical equivalent in the “real world”. However, the sun is a good illustration of the plane wave phenomenon in the visual domain. When travelling inside a car or train, one can

entertain the impression that the sun is “following” the train while the landscape streams along at high speeds. The sensation of being “followed” by an object that retains the same angular direction while one moves around inside of the listening area accurately describes the effect of a plane wave.

- **Virtual point sources situated in front of the loudspeaker array.** An extension of the WFS principle allows the synthesis of sources within the listening area at positions where no physical sources are actually present. These “sound holograms” are created when a wave front created by the loudspeaker array converges onto a fixed position inside of the listening room. The wave front is then naturally re-emitted from the target position to the rest of the listening area. The sound field is therefore inaccurate between the loudspeaker array and the target position but perfectly valid beyond it.

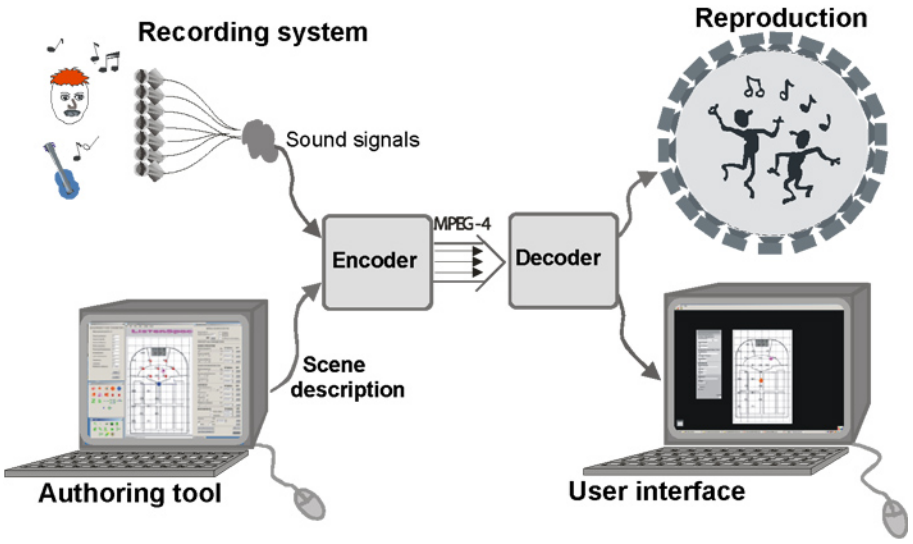


II. The CARROUSO Project

CARROUSO [5] (**C**reating, **A**ssessing and **R**endering in **R**eal **T**ime of high quality a**U**dio-visual environments in MPEG-4 c**O**ntext) is a project that was financed by the European community (IST 1999-20993) and took place between January 2001 and June 2003. This project brought together partners from the industrial and academic domains, including Delft University (inventors of WFS), France Telecom R&D, IRCAM, and the Fraunhofer Institute IIS AEMT that took care of coordinating the project.

The CARROUSO project aimed at developing techniques for recording, describing, transmitting and rendering real or virtual sound scenes. In order to achieve the prescribed objectives, the partners relied on two innovative technologies:

- MPEG-4, which is a format based on Content-Coding
- Wave Field Synthesis



1. MPEG-4

MPEG-4 [6] is a format defined by the MPEG consortium (**M**oving **P**icture **E**xpert **G**roup) for coding interactive multimedia presentations. This format is a descendant of MPEG-1 and 2 that are mostly aimed at compressing audio/video material. However, modern day multimedia applications require more flexibility when it comes to describing and reproducing audio/video content, and must also supply the end-user with the possibility to interact with this content.

This observation led to the initial work on the MPEG-4 format in 1995. A first version of the standard was defined in 1999 followed by a second one in 2000. The guiding concept for the MPEG-4 format is the concept of content-coding using a decomposition of the multimedia presentation into a set of elementary objects to which are attributed certain properties and behaviours. For spatialized sound scenes, content-coding consists in separating the sound signals associated to the different sources from the scene description.

The sound signals associated to each of the sources must ideally be “dry”, i.e. recorded with no room effect whatsoever. Different algorithms allow the reduction of the volume of data to fit the available bandwidth and desired quality during the transmission. The sound scene itself is described using a specific language named BIFS (**BI**nary **F**ormat for **S**cenes)**.** Each sound source is attributed radiation characteristics, a position, and an orientation. These sound sources are then placed in an acoustical environment that can be parameterized following two methods:

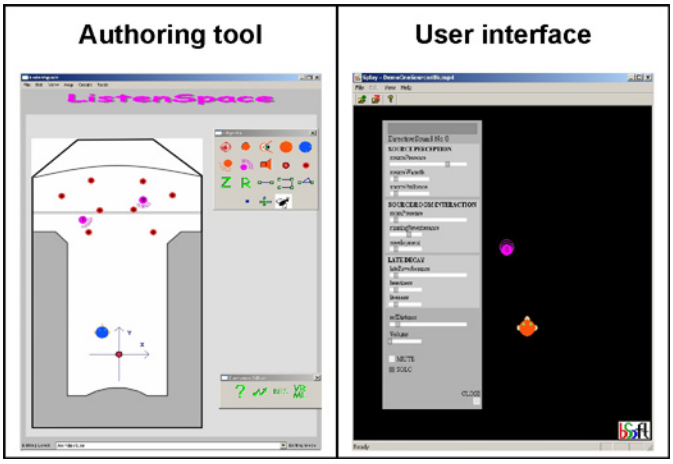
- By a geometrical description of the environment coupled with the acoustical properties of its walls. The room effect associated to a given position in the room can then be deduced from a simulation at the rendering stage [12].
- By a perceptual description of the spatial sensation undergone in the listening environment. A series of psycho-acoustical tests conducted at IRCAM suggested the use of 9 perceptual parameters for describing spatial sensations [13]. These parameters were then adopted for use in the MPEG-4 format.

During the CARROUSO project, alternative descriptions of room effect were also proposed:

- A “physical” description based on the measure of impulse responses at positions potentially to be occupied by sound sources in the recording room (onstage in a concert hall, for example...). These impulse responses are measured using a circular array of cardioid microphones. The signals recorded by the elements of the array are then recombined to form hyper-directive microphones, allowing the extraction of impulse responses following a set of elementary directions [14].
- An alternative perceptual description based on perceptual parameters such as apparent room size or source distance. Primary reflections are then used during the reproduction stage to stimulate the proposed parameters [15].

The principal advantage of MPEG-4 is that it supplies a sound scene description that is totally independent of the sound reproduction technique. This description can thus be decoded using a dual loudspeaker stereophonic setup, a 5.1 installation, a pair of headphones for binaural rendering or a WFS sound installation comprising dozens to hundreds of loudspeakers. The sound scene description will remain valid for all of these situations.

The Carrouso project also fostered the development of an authoring tool allowing the creation of sound scenes in MPEG-4 format using a geometrical description of the sound scene. This authoring tool is an extension of the ListenSpace interface developed for the European project Listen (IST-1999-20646) to which IRCAM contributed. A set of parameters is associated to each sound source and can be exported in BIFS format and subsequently transmitted to an MPEG-4 encoder. An interesting feature of the authoring tool is that it automatically generates a 2D user interface compatible with the MPEG-4 format that is contained within the content sent to the user-end. This can allow the user to access certain parameters of the sound scene (or all of them, according to the author’s wishes) and modify them in real time. He can then for example modify certain source positions, or vary the room effect parameters.



2. Recording

Conventional Techniques (stereo, surround...)

For conventional dual channel stereophonic recordings, a sound engineer will generally use one principal microphone and a set of spot microphones used for the close miking of certain sound sources. The principal microphone is placed around the critical distance so as to convey a global stereophonic image of the sound scene, as well as the acoustical signature of the

listening room. Close miking is used to increase the presence of certain instruments, improve localization and properly adjust the sound scene depth. The spot microphones used for close miking are placed in the vicinity of an instrument and record practically only its direct sound. A group of microphones can also eventually be placed in the far field to improve the perception of the acoustics of the recording hall.

"Surround" recordings use similar techniques [7]. The front channels are generally used to reproduce the frontal sound scene, taking advantage of the centre channel to increase the stability of sound source localization. The rear channels generally serve the purpose of increasing the sensation of being enveloped and immersed in the sound field. They are sometimes used as "special effect" channels for cinema applications. The principal microphone comprises at least three capsules that are combined to obtain the driving signals of the three frontal channels, with respect to the ITU-R 775-1 standard. A set of additional microphones are used to capture the diffuse sound field as well as general ambiance. They are distributed along the two rear channels and eventually along the front channels as well.

Recent developments

The mixing process can also be (extremely) simplified by the strict use of a principal microphone to capture the entire sound scene. To do so, more recent recording techniques can be employed. For instance, the sound engineer can use a Soundfield microphone to produce an Ambisonic recording. He can also use an artificial or dummy-head to make binaural type recordings, meant to be reproduced on headphones or dual-loudspeaker systems (transaural reproduction).
More complex microphone setups, exhibiting dozens to hundreds of microphone capsules, are beginning to appear. Delft University has developed a variation of the microphone array used for the Carrouso project which displays 288 cardioid capsules disposed around a circle, from which one can extract a precise image of the sound field arriving from 24 separate directions in the horizontal plane [8]. Evolutions of the Soundfield microphone are also being proposed by France Telecom [9] and Trinnov Audio[1] [9], which allow for better spatial precision during a recording.

MPEG-4 recording

An MPEG-4 recording theoretically implies a close miking of sound sources within the sound scene so as to record as little information as possible concerning the recording room. Each sound source is then attributed a parameterized or measured room effect according to the room in which the recording was made.

An opposite approach is to capture the sound field at one position of the recording room (using one of the recently developed techniques previously described) so as to extract the sound field components originating from a set of different directions. These directions may then be used to specify a set of distant virtual sources that convey the information pertaining to the recording room (no room effect needs to be added in this case).

Both of these approaches are similar in the sense that they lead to automatic mixing techniques that could eventually be carried out without the intervention of a sound engineer. Nevertheless, this situation is analogue to that of a stereophonic recording using one principal

[1] http://www.trinnov.com/

microphone or a set of spot microphones *automatically* distributed around the recording hall. Most of the recording sessions carried out during the Carrouso project combined "conventional" recording techniques with more recent ones described above [11]. The sound engineer played a crucial role in the entire process as he had to efficiently balance the different elements that he disposed of in order to obtain a result respecting the aesthetic frame set for the recording session.

In a way, one can consider the MPEG-4 format as an abstraction of the mixing stage. Instead of conserving the result of a mix upon a certain number of channels meant to be fed directly to the loudspeakers, one conserves the sound signal of each track after basic processing (volume, compression, equalization...) and then associates a given position and room effect to each of these signals.

3. Mastering and Reproduction

At the reproduction stage, MPEG-4 involves separating the virtual sources that generate direct sound from their associated room effect. There is no channel oriented "master tape" created during the mixing process. The room parameters are stored separately with the audio material for each soure. Therefore, MPEG-4 allows a very flexible decoding with one mix on different reproduction systems (WFS with different number of channels, binaural reproduction methods, 5.1, ambisonics etc.). Therefore WFS can be seen as a universal reproduction system that integrates other formats: MPEG-4 allows a downward compatibility to any reproduction system.

a. Virtual sound source reproduction

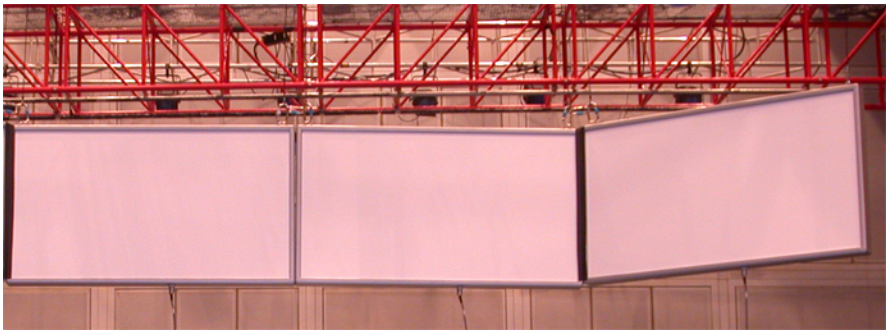
The reproduction of direct sound in the listening room is provided by the synthesis of virtual sound sources using WFS on an array of loudspeakers. During the CARROUSO project, two different loudspeaker types were employed:

- Electrodynamic(cone) loudspeakers
- MAP (**M**ulti-**A**ctuator **P**anel) loudspeakers

MAP loudspeakers

MAP (**M**ulti-**A**ctuator **P**anel) loudspeakers are derived from DML (**D**istributed **M**ode **L**oudspeaker) technology. They exhibit a vibrating plate made out of polystyrene that is excited by a set of drivers (electrodynamic devices fastened to the rear surface of the plate by their mobile coil). Each driver receives an independent signal, which allows for the creation of a multi-channel system that using a single vibrating surface. The biggest advantage of this type of setup is its low visual profile, which can allow it to be integrated into an existing environment without revealing the presence of up to hundreds of loudspeakers. Furthermore, the vibration of the surface is sufficiently faint so that it doesn't interfere with the projection of 2D images; MAP loudspeakers can therefore be used as projection screens.

The problem with these loudspeakers is that their acoustical behaviour is quite different from that of the omni-directional point sources that are theoretically needed to achieve Wave Field Synthesis. They exhibit frequency responses and radiation patterns that require specific processing. Equalization methods were therefore implemented in order to compensate for the flaws of these loudspeakers over an extended area.



Multi-channel equalization

Unlike traditional equalization schemes that aim at calculating filters for each *individual loudspeaker* without reference to the reproduction context (i.e. position and properties of the virtual sources), the multi-channel equalization scheme proposed here aims at controlling the sound field produced by the *entire array of loudspeakers* with reference to a given virtual source [16].
To do so, each loudspeaker is measured along an array of microphones. Sound field control is achieved by iteratively calculating a set of filters that will, once applied to the loudspeakers, minimize the quadratic difference between the sound field produced by the loudspeaker array and a given target for all of the microphone positions. The target is calculated for each microphone position by applying the laws of propagation between the virtual source being synthesized and the microphone array. A filter data base containing all of the virtual sources one wants to synthesize can then be put together. This process also allows the synthesis of virtual sources exhibiting complex directivities, a situation that is not covered by WFS equations.

b. Room effect synthesis

Room effect is a result of the interaction between a sound source and the different surfaces of real or virtual acoustic space. We generally distinguish:

- The first group of discrete reflections, or primary reflections, following direct sound. Their arrival time, angular direction, level and spectral content are strongly correlated with the impression of room size and source distance.
- Late reflections (after 50ms) and diffuse sound field which have very little directional dependence and can be described by their energy distribution following time and/or frequency.

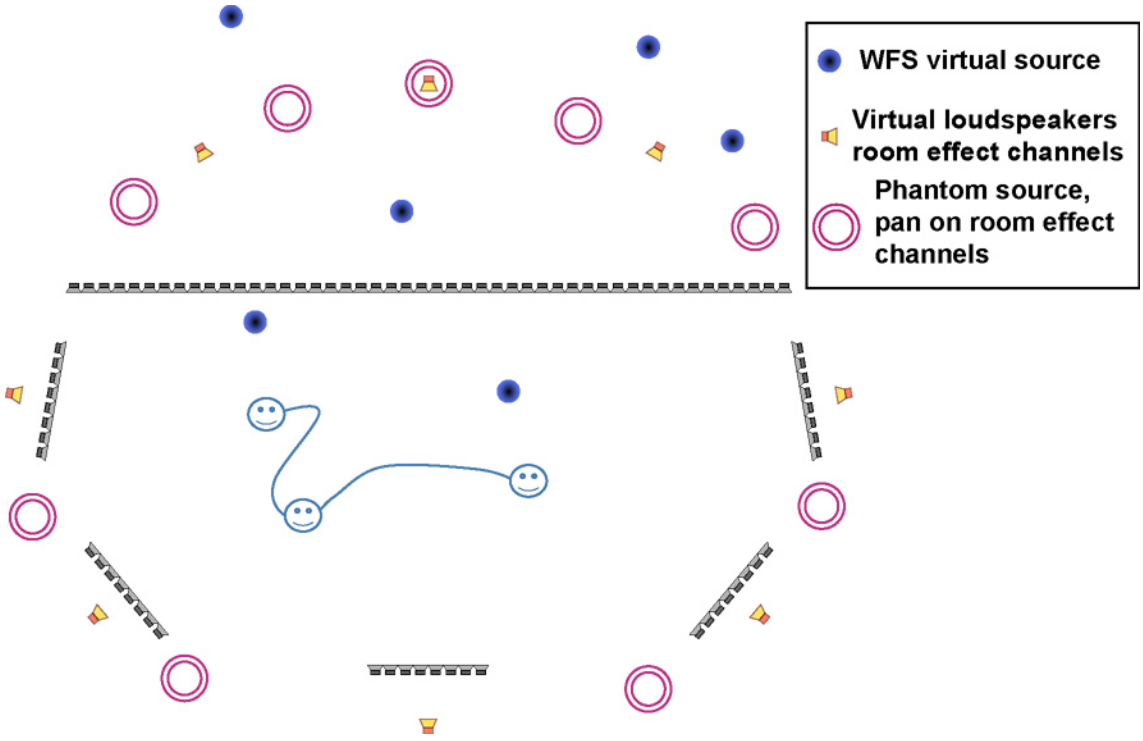
Room effect and Wave Field Synthesis

In WFS known methods to generate first reflections and a diffuse reverberation can be used, as they are used today e.g. in 5.1. But for an object oriented approach a more flexible method should be applied. Reproducing the primary reflections of a sound source in a listening room presupposes being able to synthesize each reflective component over the entire volume of the listening room. These discrete reflections can be seen as images of the virtual source seen through the “mirror” of the listening room walls. These image sources can be reproduced in their entirety if one disposes of a “complete” holophonic setup (i.e. loudspeakers covering every surface of the listening room). In a WFS setup, one can at best reproduce a source and

its reflections within the listening plane. Note that in both cases the acoustics of the reproduction room will interfere with the final result. Methods been developed to compensate for parasitical contributions of the listening room within the framework of WFS [17]. These methods are very effective in theoretical simulations; they need however to be validated with experimental results.

On the contrary, the diffuse part of room effect does not a priori require a directional component. Indeed, a sound field is considered diffuse if the signals received by a listener situated in the said field present no temporal coherence. Nonetheless, to produce a diffuse sound field one must project incoherent signals from several spatial directions. A study led by the T.U.D. shows that 8 to 10 directional channels are sufficient for the creation of a diffuse field from a perceptual point of view. The synthesized components are shown to combine with the diffuse components naturally produced by the listening room in a process of energy convolution (the room effects "add up"). Energetic deconvolution of the specified room effect with the listening room effect allows to obtain a new target room effect that compensates for the listening room influence [18]. In this context, it is impossible to withdraw diffuse energy from the room; one can only add whatever energy is lacking.
For Wave Field Synthesis installations, the synthesis of discrete reflections is generally simplified by distributing them over the available diffuse field channels. This allows a drastic drop in calculation costs. Indeed, for a sound scene containing P sources for which the Nth order reflections are to be synthesized along with 8 diffuse field channels, one needs to generate (N+1)*P + 8 virtual sources. By using the diffuse field channels to reproduce primary reflections, one need only reproduce P+8 sources.

Room effect reproduction in a WFS context is thus entirely achieved by synthesizing 8 to 10 channels reproduced as "virtual loudspeakers" around the periphery of the sound installation. These channels may also be used to reproduce the direct sound of certain sources using stereophonic panoramic techniques ("stereo pan-pot").

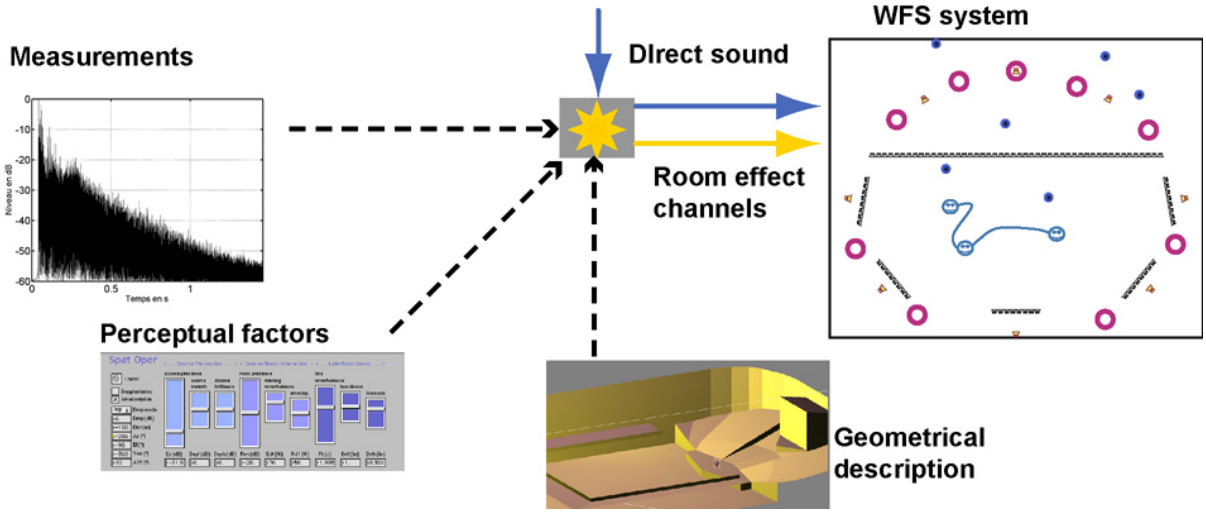


Room effect and Content-Coding

Content-coding provides a description of the target listening room. For each source, a virtual acoustics processor is responsible for "interpreting" the provided room effect description and forming 8 to 10 room effect channels.

Beforehand, several different descriptions of room effect (physical, geometric, perceptual) have been established, as well as their underlying mechanisms:

- The *physical description* provides an ensemble of impulse responses corresponding to a set of directions measured in a "target" room relatively to a given source position. The virtual acoustics processor then convolves the corresponding sound signal with the transmitted impulse responses.
- The *geometrical description* provides an architectural representation of the "target" room. Acoustical prediction algorithms are used to determine the distribution of image sources and the associated diffuse field parameters. The image sources are synthesized using delayed and filtered versions of the sound signal corresponding to the primary source and subsequently distributed over the room effect channels using a panoramic in order to synthesize the desired direction. The diffuse field is synthesized using an artificial reverberator that can produce a sufficient amount of uncorrelated signals. An alternative method consists in constructing synthetic impulse responses to be convolved with the corresponding sound signal.
- The *perceptual descriptions* involve room effect synthesis models that target the stimulation of the corresponding perceptual factors. The spatial impression specified by these models can then be synthesized using one of the methods previously delineated (reflection synthesis + artificial reverberator or synthetic impulse responses).



III. Innovative mixing techniques for Wave Field Synthesis

1. "Distance" monitoring

Wave Field Synthesis allows for the reproduction of virtual point sources. A WFS synthesized wave front will acquire a certain curvature depending upon the source's position. This curvature causes localization variations to be perceived by the listener during his movements, which allows for the creation of true spatial perspective. It therefore becomes possible to manipulate a sound source's apparent distance with a new parameter named "holophonic distance", independently of the notion of "subjective distance" that is related to the balance between direct sound and reverberation level, as well as the distribution of discrete reflections.

The notion of "subjective distance" of a given sound source was normalized for use in the MPEG-4 standard under the name "presence", defined as a quantity dependent of the amount of early energy (direct sound + early reflections) and late energy (diffuse reflections and reverberation).

Creation of Perspective

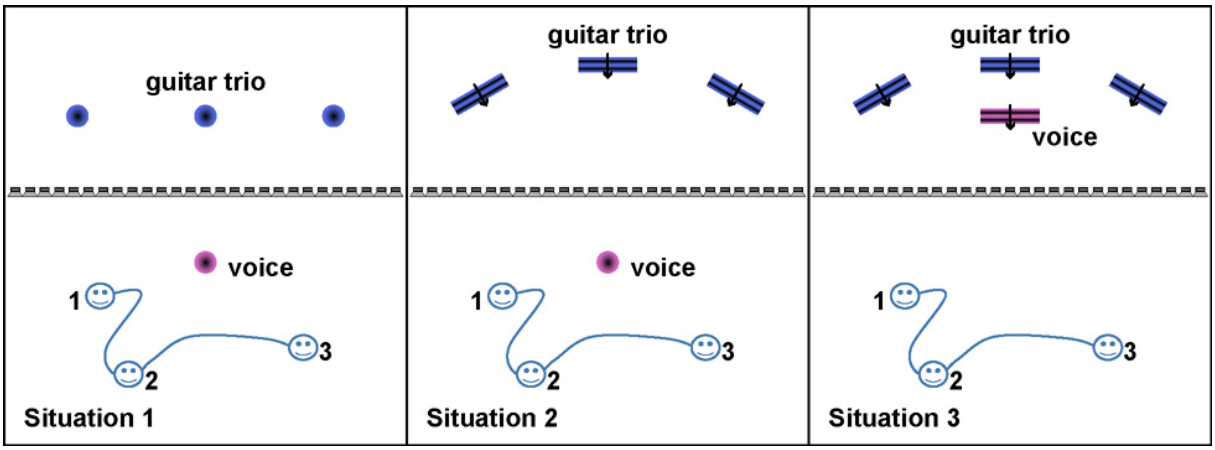
In this section we illustrate the creation of perspective in sound scenes using holophonic distance exclusively. To do so, we consider a musical ensemble composed of three guitars and a voice. In order to respect "classical" procedures, we place the guitars within one plane and the voice "in front of" this plane. Three musical situations are then constituted.

The first situation involves placing the three instruments and the voice at a short holophonic distance, i.e. close to the loudspeaker array. A spectator moving around in the sound installation can "visit" each of the sound sources by physically approaching one of them.

The second situation involves leaving the voice at a short holophonic distance and synthesizing the guitars as plane waves ("infinite" holophonic distance). The three guitars are perceived as presenting an identical angular distribution over the entire listening area, which "follows" the listener during his movements. The voice, on the contrary, remains at a fixed position, allowing the spectator to choose his point of view.

The third situation portrays all of the sources as plane waves. Identical perspective is therefore perceived over the entire listening zone. The whole scene "follows" the spectator in his movements.

It is interesting to note that all three situations are identical from the point of view of a static observer situated at the centre of the reproduction room. From this observation, we gather that holophonic distance is in fact a means for reproducing parallax effects (linked to the true position of sources in a sound scene) that arise when moving around in natural environments.



Presence vs. Holophonic Distance

At first glance, holophonic distance is not a reliable indicator of sound source distance, except at very small distances where coloration effects can be heard. Nevertheless, this parameter is closely linked to "presence" in natural listening environments. Manipulating these two parameters independently may therefore lead to conflicts in the perception of the resulting sound scene.

In order to shed light on this question, an interactive listening test was elaborated at IRCAM [19]. Variations on the three situations described above were proposed to a panel of sound engineers. The guitar trio was given a fixed holophonic distance and presence level so as to serve a reference plane in the virtual sound scene. Two variations of this reference plane were presented to the subjects in random order, one of them "close up" (short holophonic distance, strong presence) and the other "far away" (long holophonic distance, weak presence). The subjects were invited to manipulate the voice source so as to obtain a "coherent" sound scene.

The general organization of the test consisted in:

- Imposing different values of the voice's presence and letting the subject adjust its holophonic distance
- Imposing different holophonic distances of the voice and letting the subject adjust the value of its presence.

The situations are presented in random order without informing the subject of nature of the parameter he is manipulating. He disposes of a one portable MIDI fader that automatically receives the parameter retained for the experiment. He is allowed movement within the listening zone in order to correctly estimate holophonic distances for every situation.

Although the framework of this test was limited regarding the number of situations and sound samples presented to the subjects, it showed that sound engineers manipulate holophonic distance and source presence as *independent parameters.*

2. Using Wave Field Synthesis in combination with 2D video applications

When associating sound with video applications, it is advisable to distinguish two types of sounds:

- Sounds directly referenced within the image on screen (the “on”)
- Sounds that have no straightforward link with the objects on screen (the “off”, sound effects, music…)

For “on” sounds, source positions are dictated by the position of the objects on screen [20]. The manipulation of holophonic distances can therefore become a problem since the only “valid” holophonic distance is dictated by the position of the screen within the projection room. This situation is quite unfavourable for the use of WFS because it amounts to using a single loudspeaker (the closest one) to reproduce the virtual source. This is a sound reproduction situation that cannot really qualify as Wave Field Synthesis; more so the power radiated by a single loudspeaker may prove to be insufficient (WFS setups usually rely on a large number of low-power loudspeakers).

Nonetheless, two perceptual phenomena allow to maintain the validity of this type of setup:

- Precision for human auditory localization is at best 2 to 3 degrees.
- There exists a “fusion” phenomenon occurring when visual and acoustic stimuli are presented concurrently at different positions. The sound source is often seen to merge with the visual source, and the joint location is perceived as being that of the visual source. This is sometimes referred to as the “ventriloquist” phenomenon.

It is also necessary to allow for variations in the screen size that occur between two different sized projection halls. This is taken into account in visual applications by a focal adaptation of the video projector. Similarly, the 3D sound scene must be scaled to fit different projection room sizes.

For “off” sounds, the conveyed information is independent of on screen events. Hence total liberty is allowed regarding the disposition of these sources. Nevertheless, it is appropriate to confirm that the semantic link between “on” and “off” does not require certain coherence regarding the disposition of the associated virtual sources.

3. Virtual Panning Spots

Theoretically, WFS reproduction requires as many transmission channels as there are sources in the sound scene being transmitted. Each sound source is associated with a close miked signal, exhibiting coloration problems inherent to a vicinity recording. The sound sources are then reproduced as virtual point sources on the WFS system. The following observations can therefore be made:

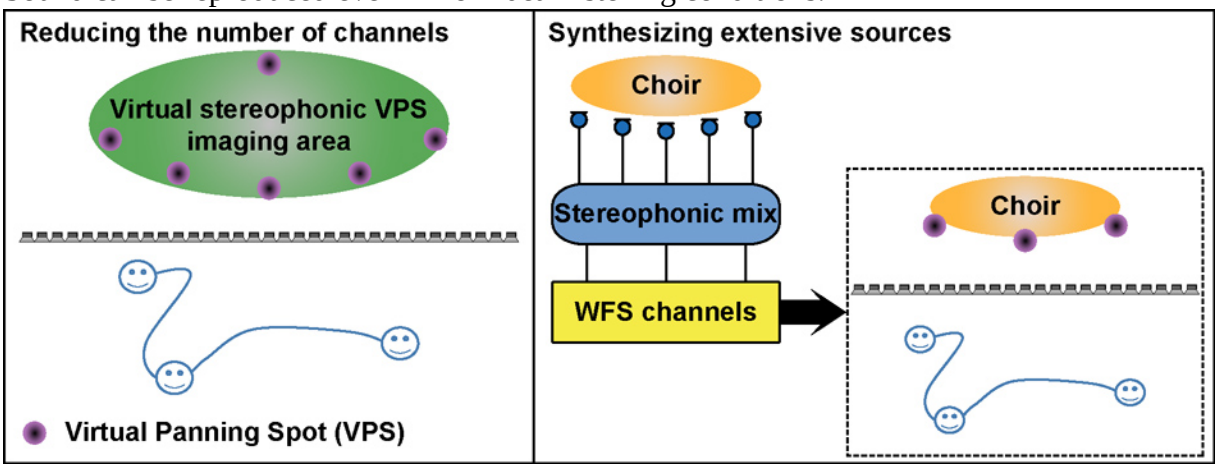
- The number of channels for an orchestra can reach large numbers. Rendering each source as a separate entity seems disproportionably precise in comparison to the spatial impressions felt during a real performance. Spectators are usually more sensitive to groups of instruments that appear to meld into unique sound masses covering spatial “zones”.
- The point sources reproduced in WFS exhibit no spatial extension. Reproducing a piano, a choir, or an organ using a single virtual source is therefore hardly conceivable.

During the CARROUSO project, Gunther Theile and Helmut Wittek of the IRT in Munich (Institut für Rundfunktechnik) proposed a solution to these problems using "Virtual Panning Spots" (VPS) [21]. This technique involves reintroducing stereophonic principles into WFS.

Virtual Panning Spots are sets of sound sources reproduced on a WFS system used as virtual loudspeakers. Each microphone signal is spread over a set of VPS using classical stereophonic recording, mixing and panning techniques (intensity, delay, principal microphone …), creating a stereophonic imaging area exhibiting "phantom" sources. The information transmitted for multiple sources can therefore be greatly reduced.

Similarly, extended source reproduction can be achieved using a certain number of spot microphone signals distributed on a set of VPS. A notion of spatial extension requiring a limited number of transmission channels can thus be rendered using this method.

In the case of monitoring Multichannel Sound (e.g. 5.1) the VPS can be reproduced containing the acoustics of an ideal *reproduction* room, which means that better Surround Sound can be reproduced even in non-ideal listening conditions.



4. Wave Field Synthesis Production Chain

Different methods and problems linked to WFS reproduction were exposed in this article. Creating a sound scene in WFS involves associating spatial information to the sound signals composing the scene. Using this information, it becomes possible to synthesize virtual sources used to reproduce direct sound as well as room effect following a physical, geometrical or perceptual description.

The sound signals distributed upon the spatialized virtual panning spots correspond to the soundcard outputs of a computer equipped with a sequencer. Classical sequencer functions such as equalization or compression are therefore available. Stereophonic panning is used in order to distribute the signals over the VPS, allowing for the creation of phantom sources within the virtual stereophonic VPS imaging area. Scene description parameters are adjustable via a plug-in that gives access to source position (holophonic distance, incidence angle) as well as MPEG-4 perceptual parameters. These parameters can be entirely automated, conferring the possibility of a temporal evolution in spatialization and allowing precise synchronicity with sound events.
The author of the sound scene can choose to render direct sound using WFS virtual point sources or by panning it between the different room effect channels. This allows reproducing a larger number of sources without increasing the required processing power. More

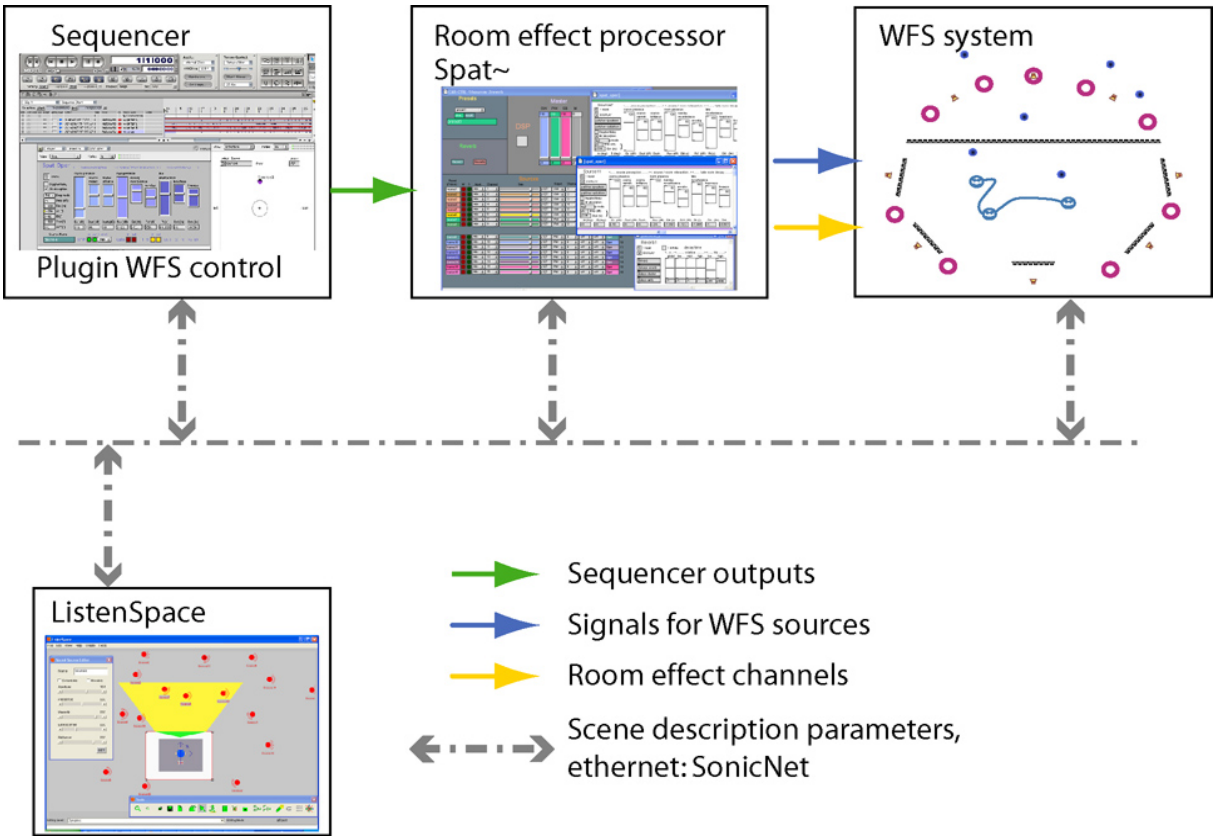
pragmatically, if the “complete” WFS setup (i.e. uninterrupted loudspeaker distribution) is restricted to the front wall of the listening room, this extends the possible positions for virtual sources to the rear and side walls.

Spatialization parameters are shared and accessible by the other elements of the production chain through a distributed database on the ZsonicNet network developed by sonicEmotion[2]. A large number of parameters are therefore made available at all locations within the installation using an Ethernet type connection. The network displays very low latency (~10 ms), allowing for a global refreshment of parameters in real time from any location within the network. ZsonicNet allows the control of distributed processes from any client inside the network. It enables a synchronous transfer of audio data to all clients and provides a consistent database of parameters. In practice WFS rendering on a large set-up with different rendering machines can be controlled from one ore more audio workstations. Therefore it allows the integration of WFS into currently available audio workstations. The audio and control data are transferred from the audio workstation to different WFS rendering machines inside the network. Still the network is server-less and allows a dynamic configuration with changing reproduction systems.

Thus the spatialization parameters can be made available on a ListenSpace interface installed on a portable PC tablet using a wireless Ethernet connection. The author can then modify spatialization parameters in real time while moving around in the sound installation.

The virtual acoustics processor Spat~ developed by IRCAM has been adapted for WFS rendering. The Spat~ creates a set of room effect channels and transmits the direct sound signals associated to WFS point sources to the reproduction system.
The proposed interface gives access to scene description parameters, as well as a few basic mixing operations (level settings, routing, mute, solo…).
It also contains a multi-channel sound-file player synchronized with a MIDI sequencer. The system can therefore function without the use of the audio sequencer. Spatialization parameters are then translated into fixed MIDI controller values. The MIDI files associated with the multi-channel sound-files therefore form a complete content-coding of the sound scene.

[2] http://www.sonicemotion.com


Sequencer
Plugin WFS control
Room effect processor
Spat~
WFS system
ListenSpace
Sequencer outputs
Signals for WFS sources
Room effect channels
Scene description parameters,
ethernet: SonicNet

Glossary

Holophony: Sound reproduction technique proposed by Jessel in 1973. The sound field is captured on an array of pressure (omni) and pressure gradient (figure 8) microphones placed on a closed surface. At the reproduction stage, the signals recorded on the omni microphones are fed to dipolar sources (e.g. non-baffled loudspeakers) and the signals recorded on the figure 8 microphones are fed to monopole sources (e.g. baffled loudspeakers). An imperative of this technique is that the loudspeakers be situated at exactly the *same position* as their associated microphones. This requires a large number of transmission channels and implies that all recording and reproduction setups be *identical*. The sound field reproduced within the listening zone is

WFS is sometimes referred to as “Holophony”. However, although WFS is derived from similar theoretical considerations, it allows for more flexibility at both the recording and the reproduction stages and a reduced number of transmission channels.

IRT: Institute für RundfunkTechnik

IRCAM: Institut de Recherche et de Coordination Acoustique/Musique

T.U.D.: Technological University of Delft, Holland

V.P.S.: Virtual Panning Spots

WFS: Wave Field Synthesis

Selective Bibliography

Reference articles in WFS

[1] Berkhout A. J., *A Holographic Approach to Acoustic Control*, Journal of the Audio Engineering Society, Vol. 36, No. 12, December 1988, pp 977-995.
[2] Berkhout, A.J., de Vries, D., and Vogel, P., *Acoustic control by wave field synthesis*, J. Acoust. Soc.Am., Vol. 93, pp 2764-2778

Reference theses in WFS

[3] Start, E.W., *Direct Sound Enhancement by Wave Field Synthesis*, thesis, T.U.D., 1997.
[4] R. Nicol, Restitution *sonore spatialisée sur une zone étendue: Application à la téléprésence*, Thèse, Université du Maine, Le Mans, France, 1999 (http://gyronimo.free.fr/audio3D/Guests/RozennNicol_PhD.html)

CARROUSO

[5] S. Brix et al, CARROUSO, *A European approach to 3D audio*, 110th AES convention, Amsterdam, 2001

MPEG-4

[6] Väänänen R., Warusfel O., Emerit M., *Encoding and Rendering of perceptual sound scenes in the Carrouso Project*, 22nd AES conference, Espoo, Finland, 2002.

Recording

[7] Theile G., *Multichannel Natural Music Recording Based on Psychoacoustic Principles*, downloadable at http://www.hauptmikrofon.de/theile.htm
[8] Hulsebos E., Schuurmans T., de Vries D., Boone R., *Circular microphone array for discrete multichannel audio recording*, 114th AES conference, Amsterdam, 2003.
[9] Daniel J., Moreau S., *Design refinement of high order ambisonics microphones – Experiments with a 4th order prototype*, CFA/DAGA 04, Strasbourg, 2004.
[10] Laborie A., Bruno R., Montoya S., *A new comprehensive Approach of Surround Sound Recording*, 114th AES convention, Amsterdam, 2003.
[11] Kuhn C., Pellegrini R., Leckschat D., Corteel E., An Approach to Miking and Mixing of Music Ensembles Using Wave Field Synthesis

Room Effect

[12] J.P. Jullien, *Structured model for the representation and the control of room acoustical quality*, in Proceedings of the 15th International Conference on Acoustics, 1995, pp 517-520
[13] Jot, J.-M., and Warusfel, O. 1995. "A real-time spatial sound processor for music and virtual reality applications". Proc. 1995 ICMC.
[14] Hulsebos E., de Vries D., *Parameterization and reproduction of concert hall acoustics measured with a circular microphone array*, 112th AES conference, Amsterdam, 2002.

[15] Pellegrini, R., *Perception-Based Design of Virtual Rooms for Sound Reproduction,* Audio Engineering Society, conference paper No. 245, presented at the 22th International Conference on Virtual, Synthetic and Entertainment Audio, Espoo, Finland, June 15-17, 2002

Loudspeaker equalization and Listening room compensation in WFS

[16] Corteel E., Horbach U., Pellegrini R., *Multi-channel Inverse Filtering of Distributed Mode Loudspeakers for Wave Field Synthesis,* preprint no 5611, 112th AES convention, Munich, 2002.

[17] Von Zon R., Corteel E., de Vries D., Warusfel O., *Multi-Actuator Panel (MAP) loudspeakers: how to compensate for their mutual reflections?,* 116th AES convention, Berlin, 2004.

[18] Corteel, E, Nicol, R. *Listening Room Compensation for Wave Field Synthesis. What can be done?,* 23rd AES conference, Copenhague, Danemark, 2003.

Distance Monitoring in WFS

[19] Noguès M., Corteel E., Warusfel O., *Monitoring Distance Effect with Wave Field Synthesis,* DAFX03, Londres, 2003.

Wave Field Synthesis in combination with 2D video applications

[20] Melchior F., Brix S., Sporer T., Röder T., Klehs B., *Wave Field Synthesis in combination with 2D Video Projection,* 24th AES conference, Banff, 2003.

VPS

[21] Theile G., Wittek H., Reisinger M., *Potential Wave Field Synthesis Application in the Multichannel Stereophonic World,* 24ème AES conference, Banff, Canada, 2003.