AURALIZATION OF ROOM ACOUSTICS BY WAVE FIELD SYNTHESIS BASED ON ARRAY MEASUREMENTS OF IMPULSE RESPONSES

Diemer de Vries and Edo M. Hulsebos

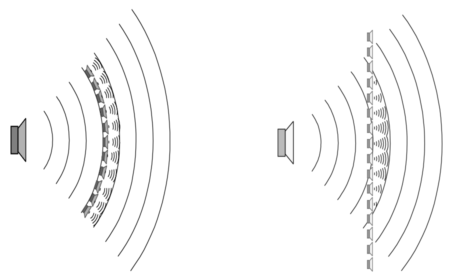
Faculty of Applied Sciences, Department of Imaging Science and Technology, TU Delft
phone: +31 15 2785220, fax: +31 15 2625403, email: diemer@akst.tn.tudelft.nl
web: www.ak.tn.tudelft.nl

ABSTRACT

Wave Field Synthesis (WFS) enables correct spatial reproduction of acoustic events for larger groups of people. Such events often include acoustic information of the recording space which is then 'auralized' after convolution with the sound sources. The acoustic information is acquired by measuring impulse responses. Spatial information is only obtained when the responses are measured along arrays (e.g., linear or circular) of microphone positions. In the presentation it will be explained how the measured datasets are parameterized and processed to be reproduced in the most efficient way: significant data reduction is possible due to the limited ability of the human hearing system to discriminate temporal and spatial details.

1. WAVE FIELD SYNTHESIS

The concept of Wave Field Synthesis (WFS) was introduced by Berkhout in 1988 [1]. The concept is based on classical wave theory going back to Huygens and further developed by Kirchhoff and Rayleigh. An extensive treatment of the theoretical background can be found in [2]. Here, a more intuitive explanation of the WFS concept is given. In 1690, Huygens [3] formulated his theorem in which he states that a wave front generated by some primary source – he had a light source in mind, but his statement also holds for a sound source – can be interpreted as a continuum of small secondary sources that together build up the next wave front, as illustrated in figure 1a.



a *b*

Figure 1: Two versions of Wave Field Synthesis (WFS).

This old representation of physics directly implies a very useful audio application. If we place an approximated continuum – i.e., an array – of small loudspeakers on the wave front and drive those individually with signals proportional to the local strength (in terms of sound pressure or particle velocity) of the primary field, then we have the ideal form of sound *amplification*: an exact copy in time and space is added up to the primary wave field. Of course, this is not very practical since for each source the wave fronts will be at different places, so – where to place the available loudspeakers? A step further to practice is the placement of a simple-shaped – say, flat - loudspeaker array on a fixed place. Knowing the positions of the sources, the primary field strength at the array position can be determined relatively easy, the loudspeaker driving signals can be adapted accordingly and the ideal amplification process still holds, as shown in figure 1b. Another limitation is that for reproduction of a 3D sound field, the theory prescribes 2D loudspeaker arrays (planes, cylinders, spheres) which are very unpractical from viewpoints of visibility and computational power. Using linear arrays, only in the plane through source and array spatial control can take place. When this is taken to be the plane roughly through sources and listeners' ears, it could satisfactorily work in practice when the operators are appropriately adapted as described in [4].

A next step, and this is what WFS is often used for, is that we leave out the primary source and just *simulate* its wave field. This gives wide possibilities in the field of sound *reproduction*. Voices and instruments which have been recorded with close-by spot microphones are virtually positioned on their true place by WFS. This way, all listeners in front of the array hear the ensemble in its natural spatial configuration – a 'super stereo' situation. Alternatively, the sound sources can be positioned in any preferred configuration to create a situation of acoustic virtual reality. Using focussing techniques, sources can even be positioned in front of the array.

Taking into account that the reflection of a primary sound wave can be represented by a mirror image source, it is seen that the acoustics of any room can be added to dry recorded sources by convolving their signals with a distribution of virtual mirror image sources created by WFS, using a 'surround' configuration of loudspeaker arrays. Creating individual images sources, however, is not the most practical way for this so-called *auralization* of room acoustics. A more appropriate approach, based on impulse response measurements will be described in the following.

2. AURALIZATION

Auralization, as indicated above, is the process of making the acoustics of a hall audible in another space than the room itself. When the hall really exists, this can be done by reproducing measured impulse responses, preferably convolved with an anechoic music or speech signal in order to hear the acoustics in a 'natural' context. This way, the acoustic quality of a space can be perceptually evaluated without being there. It also enables to virtually replay a nice concert recorded by close-miking in a hall with poor acoustics in a space with brilliant acoustics. Auralization is also used to let architects and acousticians evaluate the acoustic quality of a hall under design. Since this hall does not yet exist, the impulse responses have to be simulated using one of the software packages available for that purpose.

In commonly used auralization techniques, the (often binaural) sound field at one listener position is reproduced, to be perceived with headphones or a pair of near-field loudspeakers. This way, the spatial properties of the sound field are quite difficult to assess, since often the full acoustic image is localized within the listener's head. In the WFS approach of auralization, however, listeners can 'walk around' within the wave field generated by the loudspeaker arrays around the listening area. In order to auralize a 3D sound field this way, it is theoretically necessary to surround the listening space with a continuum of loudspeakers, such as a spherical or shoebox-shaped configuration of planar arrays. However, since most perceptual cues are determined in the horizontal ear plane of the listeners, also in this case a rectangular or circular array in this plane can do most of the job. Moreover, the floor of a regular auditorium is covered with absorptive chairs or well-dressed audience such that in the vertical direction only the first ceiling reflection is of significant importance. Investigations on the additional value of a ceiling array to correctly include this first reflection in the wave field auralized by WFS are in progress at TU Delft.

3. RECORDING AND PROCESSING OF IMPULSE RESPONSES

In the remainder of this paper, only auralization based on measured impulse responses will be discussed. In order to auralize the acoustics of a hall not at a single listener position, but at an extended area, impulse responses have to be measured along an array of microphone positions. When, during the measurements, the acoustic conditions in the hall are stationary, instead of an array of microphones one microphone can be used that moves from one position to each other. Figure 2 shows an example of a multi-trace impulse response, measured in a small lecture hall along a linear array of microphone positions over the full width of a hall, for a source position at the front center. The horizontal axis specifies the lateral position ('offset') of the microphone, vertical is the travel time axis. The picture clearly shows the complex wave structure of the sound field. Several components can be easily identified. As discussed in [5], cross arrays and circular array are alternatives, the latter one having many advantages for further processing.

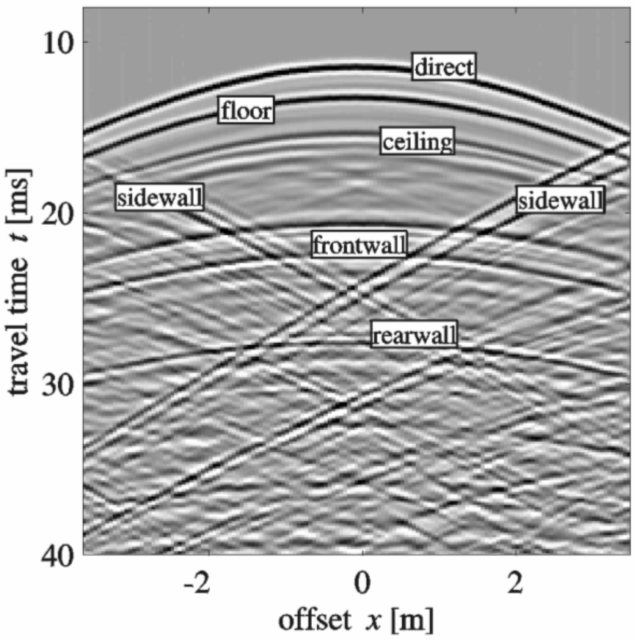


Figure 2. Multi-trace impulse response of a small lecture hall.

In order to process multi-trace impulse responses for WFS-based auralization, there are basically three possible approaches: *holophony*, *wave field extrapolation* and *wave field decomposition*. In case of *holophony*, impulse responses are being measured at microphone positions that correspond with the WFS loudspeaker positions. Here, no further processing of the measured impulse responses is required; the impulse responses can directly be used as convolution filters to drive the corresponding loudspeakers. Besides requiring the use of microphones with unrealistic directivity properties, a mayor disadvantage of holophony is that it is very inflexible; it cannot be used for a different WFS speaker configuration. In case of *wave field extrapolation* the impulse responses are not necessarily measured at positions that correspond to the WFS speaker configuration. They can be extrapolated from the microphone array positions to the WFS loudspeaker array positions using the Kirchhoff-Helmholtz integrals, as described in [5]. Successful auralization results have been achieved with this method. However, time-consuming measurements along a very large array of positions have to be done for a medium-size extrapolation area and a reasonable sound quality.

The third approach, *wave field decomposition*, decomposes the multi-trace impulse responses into plane waves. Plane wave decomposition can be regarded as making an acoustic photograph of the sound sources, including the secondary (mirror image) sources that can be represented as generating the reflections. This way, the sound field is split up into sound waves coming from different directions of incidence. A relatively small array with a size of the order of 2 meter - e.g., a circular array with a radius of 1 meter - can be used to obtain a plane wave decomposition with satisfactory angular resolution in the audio range. In that case the reproduction (extrapolation) area of the full plane wave decomposition will also be limited. However, by using only a limited number of plane waves together with a parameterization technique that will described below, a much larger reproduction area can be achieved using a small array setup as described above. The theory behind plane wave decomposition is described in [5].

4. PARAMETRIZATION AND REPRODUCTION

Full plane wave decomposition of a multi-trace impulse response recorded along a circular array of microphone positions still yields a large amount of data. Besides, proper reconstruction and reproduction of such plane wave decomposition data is limited to the inner area of the recording circle. To reduce the data size and enlarge the listening area size some simplifications are made in both temporal and spatial properties of the sound field, which are within the thresholds of human auditory perception.

For the spatial reduction the direct sound is reproduced separately as a point source. Also dominant early reflections are isolated from the impulse responses, parameterized and, if feasible in practice, reproduced as individual point sources. For the WFS reproduction of the remaining diffuse part of the plane wave decomposed impulse responses, a limited number (8 – 16) of plane waves is used. Their content is created by panning the components of the full plane wave decomposition. The comb filter effects inherent to such panning procedure are perceptually not relevant, since only the diffuse part of the impulse response is being panned, not the direct sound. If it is not feasible to reproduce the dominant early reflections as separate virtual WFS sources, they can be panned and reproduced over the limited number of plane wave channels used for the diffuse part of the impulse response.

By isolating the direct sound and dominant reflections, aperture effects from the microphone array that limit the wave field extrapolation area, are avoided, since these isolated events are reproduced as ideal virtual sources that do not suffer from such aperture effects. For the diffuse part of the plane wave decomposed impulse responses, the contents of the plane wave reproduction channels can be made mutually uncorrelated. In this way a large listening area can be obtained, much larger than the original microphone setup.

Figure 3 shows an example of an appropriate auralization setup.

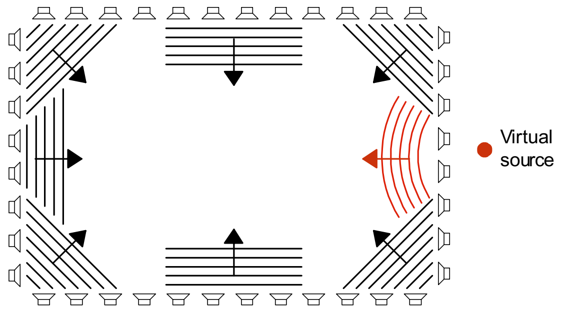


Figure 3. Direct and reflected sound reproduction for auralization.

The data size can be further reduced using temporal parameterization. The method consists of splitting the impulse response into critical frequency bands, calculating their time envelopes or RMS levels and temporally integrating them within appropriate time integration windows. What remains are positive amplitude functions that can be sampled at very low rates. In this way only the perceptually relevant large scale variations of the impulse responses are maintained while their fine structure is ignored and will be random. Critical band filters that can be used for example are the Patterson filters [6]. Suitable time windows are frequency band dependent Hanning of Gaussian windows. This temporal paramterization works quite well for the diffuse part of the sound field; it can not be applied to the direct sound and strong early reflections without significant loss of perceptual quality. The late reverberant part of the sound field can be parameterized even simpler by making an exponential fit in time for each critical band.

The spatial and temporal parameterization methods described above can be combined. This way, a multi-trace impulse response can be appropriately described (encoded) with a limited number (order 10^3 – 10^4) of parameter values. This data reduction by parameterization is also very important when the recorded data has to be transmitted (e.g., in MPEG-4 format) between the recording and the reproduction spaces.

The complete parameterization procedure can now be summarized as follows: first the plane wave decomposition is calculated from the measured impulse responses. The direct sound and early reflection parameters (position, relative strength, spectra etc.) are estimated and the corresponding events are removed from the plane wave decomposition. The remaining data is distributed over the reproduction channels by using a suitable panning algorithm. The remaining channels are temporally parameterized and an exponential fit is made for the late reverb. Figure 4 shows a block diagram of the complete parameterization process.

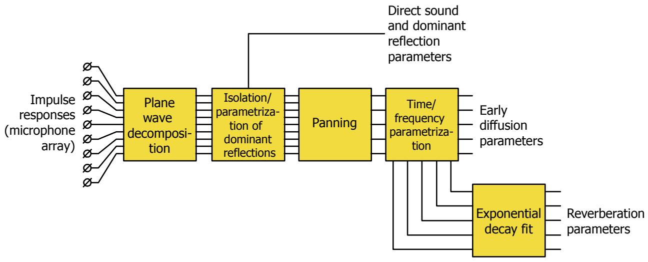


Figure 4. Block diagram of the parameterization process.

In order to be auralized, the parameterized data has to be reconstructed (decoded) to a format suitable for WFS reproduction. This is presently done in the following way, being an improved procedure as compared with an older method described in [7]. The reverberation parameters are reconstructed from the exponential fit parameters and combined with the parameters for the early diffuse part into a single set. Impulse responses are now reconstructed from the parameters by using mutually uncorrelated noise signals for different channels and properly adjusting their large scale variations. The direct sound and discrete early reflections are reconstructed from its parameters and reproduced with their proper wave shapes using WFS. The reconstructed remaining impulse responses are reproduced as plane waves on the WFS system. By using mutually uncorrelated noise signals for the different channels, comb filter effects from events in the sound field that do not match well with a selected channel such that they must be panned over at least two channels, are removed: the different fine structure of the noise signals is used as a decorrelation filter to reduce interference effects between the channels. Figure 5 shows a block diagram of the reconstruction process.

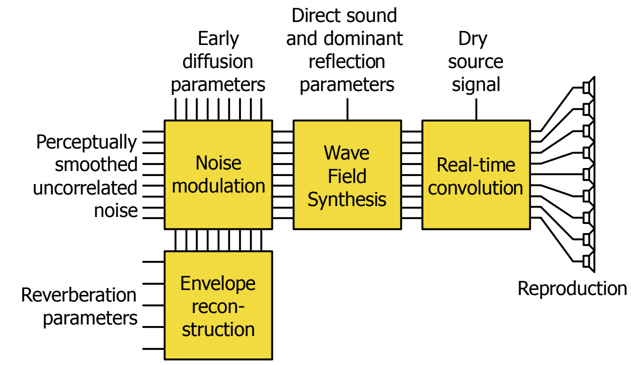


Figure 5. Block diagram of the reconstruction and reproduction process.

During the presentation, more details and results of the new method will be discussed.

CONCLUSIONS

- Wave Field Synthesis (WFS), a reproduction concept based on array technology, enables the auralization of an extended part of a hall.; listeners can 'walk around' in the auralized environment.
- Auralization by WFS is based on measured or simulated impulse responses along arrays of positions.
- Measured multi-trace impulse responses can be processed for WFS auralization by different approaches, the most advantageous being plane wave decomposition.
- Significant data reduction without loss of perceptual quality can be obtained by spatial and temporal parameterization of the multi-trace impulse responses.

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, "Applied Seismic Wave Theory", Elsevier, Amsterdam etc, 1987.

[3] C. Huygens, "Traite de la lumiere", ed. P. van der Aa, Leiden, 1690.

[4] E. W. Start, "Direct sound enhancement by Wave Field Synthesis", PhD dissertation Delft University of Technology, chp.3, ISBN 90-9010708-8, 1997.

[5] E. M. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sounds fields by wave-field synthesis", J. Aud. Eng. Soc., Vol. 50, pp. 779-790, 2002.

[6] R. D. Patterson and B. C .J. Moore, "Auditory filters and excitation patterns as representation of frequency resolution" in B. C. J. Moore, Frequency selectivity in hearing, Academic Press, London, pp. 123--174, 1986.

[7] E. M. Hulsebos and D. de Vries, "Spatial decomposition and data reduction of sound fields measured using microphone array technology", proceedings 17th ICA, Rome, 2001.