Creating Plausible Auditory Virtual Environments for Wave Field Synthesis

Renato Pellegrini
sonicEmotion AG, Bahnhofstr. 46, 8157 Dielsdorf, Switzerland, Email: renato.pellegrini@sonicEmotion.com

Introduction

Wave Field Synthesis (WFS) describes a versatile sound reproduction technology. The technology as discussed here was first introduced by Berkhout [1]. The concept of Wave Field Synthesis is based on classical wave theory going back to Huygens [2] and further developed by Kirchhoff and Rayleigh. An extensive treatment of the theoretical background can be found by Berkhout in [3]. The basic technology is well developed and documented in several AES preprints and journal papers. In this paper perceptual consequences of simplifications within the concept of WFS are discussed. This knowledge is used to define a model for a plausible auditory virtual environment. Further an implementation using a distributed, object-oriented-database network is introduced.

What we can expect from WFS in theory

In all implementations to date, WFS is restricted to a reproduction system using linear or circular loudspeaker arrays in one plane through the sources and listeners. Although the basic theory of Huygens is based on a continuum of secondary sources, practical implementations to date require a finite set of loudspeakers. Compared to binaural systems, today's WFS systems are 2-D reproduction systems since they lack control of the elevation dimension. However, since WFS can recreate the accurate wave fronts in a complete listening plane, interactive setups can be easily implemented for many listeners in which the sound scene can be explored by moving around without the use of expensive head-tracking systems, as used in binaural setups.

A basic setup therefore can create

- a number of distinct virtual sources
- an accurate acoustic perspective among different sound sources in the listening area (unlike stereophonic reproduction systems)
- Virtual point sources can be rendered to create sound sources that appear at stable positions in the listening zone.
- Plane waves can be used to create sound sources that appear from stable directions of incidences in the listening zone.
- The high number of secondary sources can be easily used to create a perceptually perfect diffuse field that is indistinguishable from a true diffuse field.

At first sight these technical elements should allow for an accurate physical model of a real environment, at least when combined with appropriate wave-field analysis techniques.

Limitations of WFS implementations

Restricting the reproduction of WFS to a plane is an important step to reduce the required processing power for a practicable implementation. Also the density of single exciting elements within the loudspeaker array is directly proportional to the required processing power.

Limitations due to spatial aliasing

However, the spacing between the loudspeakers directly defines the up-most frequency for which the phase can be appropriately controlled to create WFS-wave fronts. This so-called spatial aliasing frequency has to be above the frequencies used by human localization, to allow for at least perceptually accurate reproduction. Since we restrict ourselves to the horizontal plane through the listeners' ears, we can expect from known psychoacoustics that we can have some appropriately accurate localization in the horizontal plane for the direction of incidence by stimulating accurate ITDs and ILDs at the listeners' ears for frequencies below approximately 1700 Hz. For all frequencies above the spatial aliasing frequency we would like to at least control the spectral content of the system in the complete listening zone by using techniques as described in [4].

Limitations due to spatial windowing

Due to the windowing effect described in [5], we cannot recreate the wave fronts of an entire virtual source. This means, that the virtual source created by WFS will not feed the listening environment in the same way a real source at the same position would. Especially for sources with a high directivity index, where the main lobe is not pointing towards the listening zone, this defect becomes perceptually prominent. Possible improvements may be achieved by using additional reflections to compensate for the missing energy as described in [5]. As shown in [6, 7] the distance impression for nearby sources just on the basis of the curvature of the wave front does not serve as a sufficient cue in a WFS environment. To improve distance perception, discrete reflection patterns as described in [8, 9] can be used.

Plane wave simplifications

A real environment can be analyzed by means of measuring a set of impulse responses. This data is then used to feed a plane wave decomposition algorithm as described in [10]. Although we know from psychoacoustics that the curvature of the wave front will not result in changed perception for sources more distant than 3m in stationary situations, there is no proof that this assumption is valid when the listeners are allowed to move in a large listening zone. In contrary, referencing [11], it seems that listeners are able to judge very subtle changes in the positioning of distant sound sources, when allowed to move and explore the sound scene.

Conclusions for the applicability of WFS

Taking the limitations described above it becomes evident that a direct copy of a real environment is hardly feasible. Interestingly, in many applications such an authentic reproduction is not required. An advanced concept for the creation of auditory virtual environments is based on plausibility as described in [12]. Taking plausibility as the basis of our quality definition, we can allow for non-physically motivated rendering to compensate for defects of the rendering system including discrete reflections to improve directive source perception or distance perception. Taking the possibilities and limitations of current WFS implementations we can summarize: WFS is a versatile reproduction technique for AVEs, however, a true physical model is bound to severe limitations driven from windowing, aliasing, and array-topology simplifications. Using an appropriate perceptual model a plausible AVE is feasible, although we have to incorporate knowledge about the exact reproduction setup. Given a high number of individual sources in a virtual acoustic scene, a network for control is required.

Mandatory network features

Since WFS requires a high amount of processing power, multi-processor networks have to be applied to implement a manageable system. A distributed approach is introduced that allows for high scalability. In contrast to classical audio processing schemes, in this application, the basic signal flow is more complex. Audio and control data have to be shared by many clients simultaneously. Peer-to-peer communication schemes, as used frequently in today's audio networks, are unfavorable. An appropriate database-distribution network can be seen as a synchronous TDM bus. However, a truly synchronous network is costly and not well suited for fault tolerance and redundancy. Synchronicity and a defined latency is key for WFS to avoid artifacts in the auditory percept. A simple example to illustrate this fact is a moving sound source. The levels and delays for all secondary sources have to be changed synchronously to create consistent wave fronts for each moment in time. Typical implementations of WFS include many different modules (renderer, coder / decoder, graphical user interface, control interfaces to standard digital audio workstations …) that interact on the same database of control and audio data. Different modules will require different operating systems, which enforces the network to run on multiple platforms. Further, nice-to-have features of the network include the possibility to consume only the relevant data for each client. Since systems can become quite large, appropriate redundancy and fault tolerance schemes have to be applied. The highest possible scalability can be achieved with a server-less design with a distributed database, where each module will implement the portion of the database that is handled by the module. A GUI, for instance, will implement the necessary data to visualize the current setup, while the renderer database includes all relevant data for rendering. If a specific module is switched off (or crashes), the portion of the database handled by that module will not be available. The loss of the functionality of that specific module is transparent to all other modules since that part of the database is (temporally) not accessible. In this way, fault tolerance mechanisms can be implemented.

Conclusions

WFS is a promising technology for AVEs. A perception-driven reproduction model is an adequate tool to compensate for physical defects and simplifications in WFS. Peer-to-peer connections for audio & control are disadvantageous. An advanced networking scheme based on a server-less, TCP/IP-based, distributed database has been introduced.

References

[1] A .J. Berkhout, “A holographic approach to acoustic control”, J. Audio Eng. Soc., Vol. 36, pp 977-995, 1988.
[2] C. Huygens, “Traite de la lumiere”, P. van der Aa, Leiden, 1690.
[3] A. J. Berkhout, “Applied Seismic Wave Theory”, Elsevier, Amsterdam etc, 1987.
[4] E. Corteel, U. Horbach, R. S. Pellegrini, “Multichannel Inverse Filtering of Multiexciter Distributed Mode Loudspeakers for Wave Field Synthesis”, presented at the 112th AES convention in Munich, Germany, 2002.
[5] T. Caulkins, E. Corteel, O. Warusfel, “Wave field synthesis interaction with the listening environment, improvements in the reproduction of virtual sources situated inside the listening room”, Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, September 8-11, 2003.
[6] H. Wittek, “Spatial perception in WFS rendered sound fields: Distance of real and virtual nearby sources”, (same session), joint convention CFA, DAGA, Strasbourg, France 2004.
[7] M. Boone, “On the Audibility of Spatial Reflection Changes in Relation to WFS”, joint convention CFA, DAGA, Strasbourg, France 2004.
[8] R. S. Pellegrini, “A Virtual Reference Listening Room as an Application of Auditory Virtual Environments”, Ph. D. Thesis, Ruhr-University Bochum, Germany, 2002.
[9] R. S. Pellegrini, U. Horbach, “Perceptual Encoding of Acoustic Environments”, ICME, Lausanne, Switzerland, 2002.
[10] U. Horbach, E. Corteel, E. Hulsebos, R. S. Pellegrini, “Real-Time Rendering of Dynamic Scenes using Wave Field Synthesis”, ICME, Lausanne, Switzerland, 2002.
[11] M. Nogues, E. Corteel, O. Warusfel, “Monitoring distance effects with wave field synthesis”, Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, September 8-11, 2003.
[12] R. S. Pellegrini, “Quality Assessment of Auditory Virtual Environments”, ICAD, Espoo, Finnland, 2001.