

A SPATIAL AUDIO INTERFACE FOR DESKTOP APPLICATIONS

MICHAEL STRAUSS, ALOIS SONTACCHI, MARKUS NOISTERNIG, AND ROBERT HÖLDRICH

Institute of Electronic Music and Acoustics, University of Music and dramatic Arts, Graz, Austria
Sontacchi@iem.at

The aim of the proposed system is to create an immersive audio environment for desktop applications without using headphones. It allows the spatialisation of 3D sound fields around an almost free mobile listener. The sound field is reproduced by a several loudspeakers positioned along the desktop edges without using head tracking. Loudspeaker driving signals are derived by combining the Holographic Approach (*WFS, Wave Field Synthesis*) with different panning laws. For optimization a simulation environment in MATLAB has been implemented and beside numerical results, the quality of the synthesized wave field can be evaluated graphically.

INTRODUCTION

Beyond computer games the recent progress on advanced systems producing virtual reality has become more important in different fields of applications. Virtual reality has become part of communication (e.g. videoconferencing), entertainment, navigation systems and assistance of handicapped people. It is essential that these simulations include a realistic recreation of the intended auditory scene. Olson states four necessary conditions in [1] to achieve realism in a sound reproducing system:

- The frequency range must include all audible components.
- The dynamic range must be large enough to prevent distortion and noise.
- The spatial sound pattern of the original sound should be preserved in the reproduced sound.
- The reverberation characteristics of the original sound should be approximated in the reproduced sound.

The quality of the spatial sound reproduction has been improved by developing different 3D spatialisation strategies. These strategies depend on whether the sound field is reproduced at the ears of the listener (called transaural systems) or within a defined area around the listeners (holophone systems) [2].

Transaural systems can be realised by either using headphones or loudspeakers. In the case of using headphones the spatialization in virtual acoustic environments requires the filtering of the sound streams with head related transfer functions (HRTFs). The HRTFs capture both, the frequency and time domain aspects of the listening cues to a sound position. In the case of using loudspeakers the drawback of unwanted

channel crosstalk has to be reduced (crosstalk cancellation). This technique also called the stereo-dipol technique was introduced by Atal and Schroeder back in the 1960s. In Gardner [3] a system combining a stereo-dipole with a head tracker is presented to overcome the inflexibility of the listener. Therefore additional difficulties like HRTF interpolation must be faced.

Holophone systems can also be realised by either using headphones or loudspeakers. The case of using headphones will be described elsewhere [4]. In general using loudspeakers involves a large number of transducers, no matter whether the Wave Field Synthesis (WFS) [5] approach or the Ambisonic approach [6] is used. In the following an audio interface for desktop applications using loudspeakers is presented. The approach is based on the WFS approach to calculate the loudspeaker feeds. Several loudspeakers positioned along the desktop edges should generate the targeted sound field. Caused by the insufficient loudspeaker layout an additional panning law is required. Therefore the vector based amplitude panning technique (VBAP), which was introduced by Pulkki [7], will be considered. Since the loudspeakers are placed in front of the listener only sources in front of the listener can be reproduced faithfully. However the position of sources can vary in azimuth, elevation and distance. Sources behind and in front (see above) of the desktop can be realized. The size of the reproduction area is restricted to small extend around the listeners head. Consequently the listeners head can move free inside this region without being tracked.

1 THE AIM

The region of visual perception is depicted in figure 1. The proposed application should provide the user of screen applications to immerse into an virtual audio environment without using headphones and obtaining the possibility to zoom into icons (acoustic lense), too. Therefore this auditory interface should provide the ability to simulate an acoustic environment such that performance by the listener is indistinguishable from their performance in the real world.



Figure 1. Region of visual perception.

The paper is organized in the following sections. Chapter 2 briefly introduces the wave field synthesis approach and the calculation of the loudspeaker feeds is derived. Chapter 3 captures the soft- and hardware implementation. An objective measure to assess the simulation results for different panning laws, and also to optimize the overall system is presented in section 4. Simulation results are described and depicted in Chapter 5. Finally the paper is concluded.

2 THEORIE

The WFS approach is based on the Huygens` Principle which is mathematically described by the Kirchhoff-Helmholtz-Integral (see eq. 1). This integral implies that the wave field of a source free volume V can be described by the knowledge of the pressure along the enclosure surface S and the gradient of the pressure normal to the surface S .

$$P(\vec{r}_R) = \frac{1}{4\pi_S} \int_S \left[P(\vec{r}_S) \cdot \nabla_S G(\vec{r}_R | \vec{r}_S) - G(\vec{r}_R | \vec{r}_S) \cdot \nabla_S P(\vec{r}_S) \right] \cdot \vec{n} \cdot dS \quad (1)$$

whereas $G(\vec{r}_R | \vec{r}_S) = \frac{e^{-jk|\vec{r}_R - \vec{r}_S|}}{|\vec{r}_R - \vec{r}_S|} + F$ is known as the

Green's function and $P(\vec{r}_S)$ is the pressure along the bordering surface. Therefore each arbitrary sound field inside a source free volume can be reproduced with distributed monopole and/or dipole sources along the surrounding surface. The Green's function can be altered by the function F which is independent of the position. A suitable choice of the function F will lead to the Rayleigh I or the Rayleigh II integral. The solution

of the Rayleigh I integral consists of monopole sources. The solution of the Rayleigh II integral consists of dipole sources. Deviding the space into two sub spaces, the space of source positions and the space of listening positions, and further restrictions will lead to a simplification of the integral in Eq.1. Therefore we get a simplified synthesis operator. These synthesis operator is used to derive the corresponding driving signals of the secondary distributed sources caused by the position and signal of the initial primary source.

2.1 Derivation of the driving functions

Starting from the holographic approach in acoustics introduced by Berkhout [5] the $2\frac{1}{2}$ -D synthesis operator [8] will be obtained. Therefore the reference sound field of the primary (line) source will be reproduced by a discrete line array of monopole sources. The driving function of each secondary monopole source is given in Eq. 2.

$$Q_m^{gen}(r, \omega) = S(\omega) \sqrt{\frac{\text{sign}(\zeta)k}{2\pi j}} \sqrt{\frac{\zeta}{\zeta-1}} \cos \varphi \frac{e^{\text{sign}(\zeta)jkr}}{\sqrt{r}} \quad (2)$$

Whereby $S(\omega)$ denotes the source spectrum. ζ is introduced due to the choice of the co-ordinates origin by the relative distance between the primary source and the secondary source over the distance between the secondary source and the listening position. The angle between the connecting line of the primary source and the secondary source with the perpendicular of the array line is denoted by φ . Finally r depicts the distance between the primary source and the secondary source.

2.2 Panning algorithm

The reproduction of sources positioned in the middle of the desktop have to be superimposed by all four line arrays, because of the insufficient loudspeaker layout. Hence each line array can be regarded to produce a virtual loudspeaker using the WFS approach. Thereby the distance of the reproduced source can be controlled (see Fig.2). Using a convincing 3D panning law each position in azimuth and elevation can be directed.

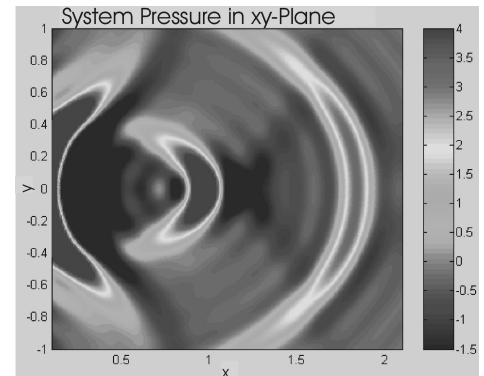


Figure 2. Reproduced "In-Front" Source.

The strategy of panning virtual loudspeakers is figured out in Fig. 3.

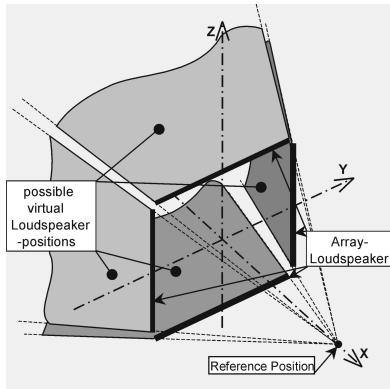


Figure 3. The strategy of panning virtual loudspeakers

3 IMPLEMENTATION

3.1 Hardware

The IEM Audio Interface consists of 16 tiny loudspeakers (100-8kHz) which are grouped to 4 line arrays (see Fig. 4). The signal generation is realized with a standard digital multichannel-soundcard which is connected to a 16 channel D/A-converter. Each channel is amplified by a calibratable channel power amplifier, which was developed at the IEM.

3.2 Software

The 2½-D synthesis operator was implemented in Pure Data for driving the linear loudspeaker arrays on a standard 2GHz PC. Thus “in-front” and behind positioned as well as moving sources can be reproduced with this algorithm. Pure Data (PD, by Miller Puckard) is a graphically based, open source real time computer music language and is free available at <http://www.iem.at/services>.

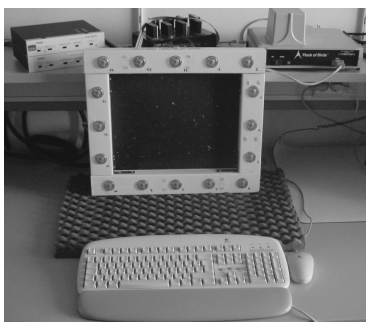


Figure 4. IEM Audio Interface

4 SIMULATION AND OPTIMIZATION

To simulate and evaluate the three dimensional sound field synthesised by an array a special MATLAB®-development tool has been established. The features of

this tool provides the simulation of 3D-wave propagation of arbitrary arrays with any weighting function (incl. WFS). Extensive possibilities for graphical evaluation can be used to examine the simulated system performance.

Furthermore an objective 3D-error function has been developed to assess the overall system performance. The relative synthesis error is defined in Eq. 3.

$$L_{rel} = 10 \log \left\{ \frac{|P_{sec}(r_M, f) - P_{prim}(r_M, f)|^2}{|P_{prim}(r_M, f)|^2} \right\} \quad (3)$$

To obtain a single value for the system performance the relative error is summed up over a sphere. To reduce the account of errors which occur far away from the ideal listening position the radial error weighting function is introduced (see Eq. 4).

The core range of the weighting function which is controlled by the parameter r_1 can be related to the head extent.

$$w(r) = 1 \quad \text{für } r \leq r_1$$

$$w(r) = w_{r_{max}} \left[\frac{(r_1 - r)}{(r_1 - r_{max})} \right] \quad \text{für } r > r_1 \quad (4)$$

The weighting function $w(r)$ is depicted in fig. 5. The weighted objective 3D-error function (single value) is given in Eq. 5.

$$\bar{L} = \frac{\sum_{n=1}^N L(r_n, f) \cdot w(r_n, f)}{\sum_{n=1}^N w(r_n, f)} \quad (5)$$

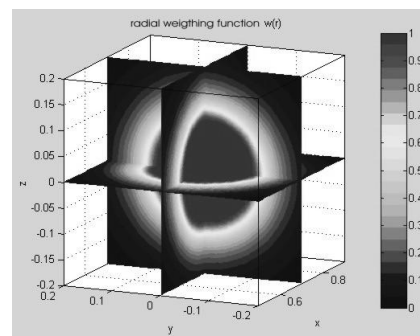


Figure 5. Radial error weighting.

5 RESULTS

In the following an example of the simulated source reproduction and the introduced relative error with the IEM Audio Interface is depicted. The parameters of the figured primary source position are given by: $x=-0.4\text{m}$, $y=-0.15\text{m}$, $z=+0.15\text{m}$. In this case the vector based

amplitude panning (VBAP, [7]) was used to weight the four line arrays according to the reproduced source position.

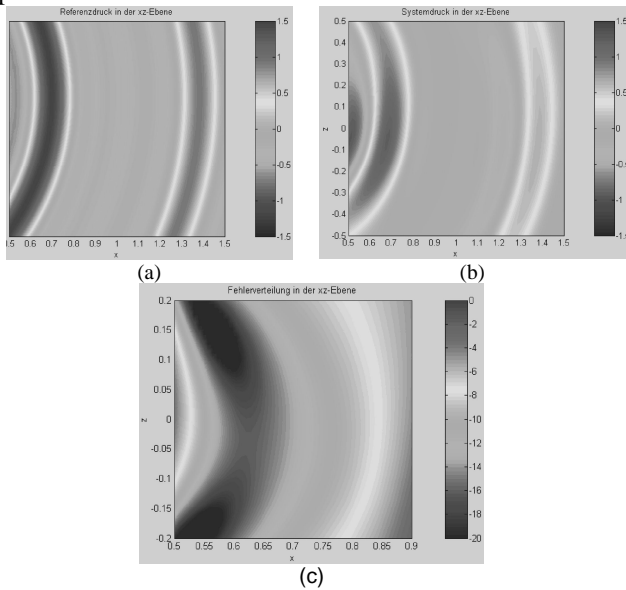


Figure 6. Results of the IEM audio interface.

- (a) Reference pressure field in the xz-plane,
- (b) Pressure field reproduced by the distributed secondary sources – array loudspeakers,
- (c) weighted relative synthesis error around the ideal listening position of 0.7 meters.

Each distance is given in meters. The pressure and the weighted relative synthesis error are depicted in dB.

The weighted objective synthesis error \bar{L} is obtained by calculating Eq. 5. Choosing the weighting parameter $r_1=0.2\text{m}$ we get the weighted objective synthesis error $\bar{L} = -13\text{dB}$.

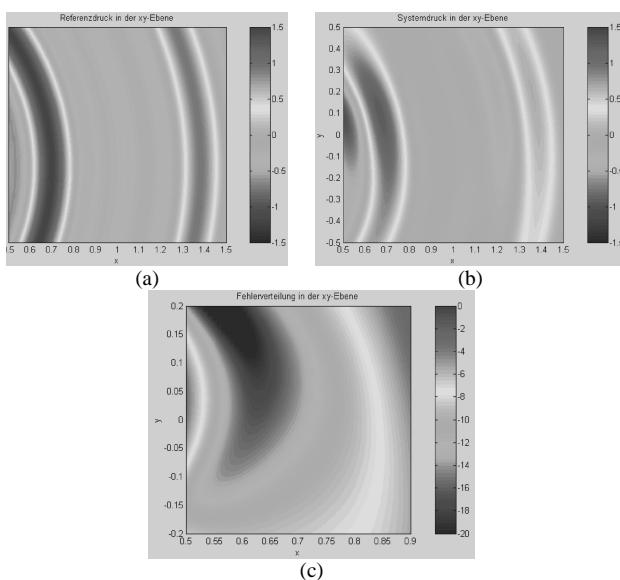


Figure 7. Results of the IEM audio interface in the xy-plane. Description see figure 6.

6 CONCLUSIONS

An audio interface for desktop applications has been presented. The proposed application should provide the user of screen applications to immerse into an virtual audio environment without using headphones and obtaining the possibility to zoom into icons (Acoustic Lens), too. The theoretical model approach is based on the WFS and can be combined with various panning techniques to overcome the insufficient number of loudspeakers. It provides the possibility to direct the reproduction position in distance, azimuth and elevation. The interface can be applied as an auditory interface for blinde people (information spatialisation) and of course to all kind of computer entertainment (e.g. games), too.

Furthermore we introduced a objective single value error to assess the overall system performance.

Beyond that listening tests have to be conducted to verify the reliability of the objective measure and to subjectively confirm which panning law fits best.

REFERENCES

- [1] H.F. Olson, “Modern Sound Reproduction,“ Van Nostrand Reinhold, New York, 1972.
- [2] M. Poletti, “The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems,“ in *J. Audio Eng. Soc.*, Vol. 44, No. 11, pp. 948-963, 1996.
- [3] W. Gardner, „3-D Audio Using Loudspeakers,“ MIT, Kulver Academic Publishers Group, 1998.
- [4] M. Noisternig, A. Sontacchi, T. Musil and R. Höldrich, “A 3D Real Time Rendering Engine for Binaural Sound Reproduction,“ presented to ICAD 2003.
- [5] A.J. Berkhout, „Holographic approach to acoustic control,“ in *J. Audio Eng. Soc.*, Vol. 36, pp.977-995, 1988.
- [6] M. A. Gerzon, “Ambisonic in multichannel broadcasting and video,“ in *J. Audio Eng. Soc.*, vol. 33, pp. 859-871, 1985
- [7] V. Pulkki, „Virtual Sound Source Positioning Using Vector Base Amplitude Panning,“ *J. Audio Eng. Soc.*, Vol. 45, No. 6, pp. 944-951, 1997.
- [8] E.W. Start, D. de Vries, A.J. Berkhout, „Wave Field Synthesis Operators for Bent Line Arrays in a 3D Space,“ *ACUSTICA – acta acoustica*, Vol. 85, 1999.