THE IEM-CUBE

Johannes M Zmölnig, Winfried Ritsch, Alois Sontacchi

Institute of Electronic Music and Acoustics
University of Music and Applied Arts, Graz, Austria
zmoelnig@iem.at

ABSTRACT

Traditional multichannel-reproduction systems are mainly used for recreation of pantophonic soundfields. Fully periphonic reproduction has been limited by computational power to manipulate large numbers of audio-channels as well as needed speaker-layouts. Since in the last years digital hardware has become fast enough to meet the computational requirements, a medium-sized concert-hall for reproduction of periphonic electro-acoustic music, the so called *IEM-Cube*, has been installed at the the IEM. The room is equipped with a hemisphere consisting of 24 loudspeakers, that allows reproduction of three-dimensional soundfields following ambisonic principles of at least 3rd order. To make use of this, a linear 3D-mixing system on PC-basis has been developed. The system may be used as a production-tool for periphonic mixing into a set of ambisonic-channels, as a reproduction-environment for recreating a 3D-soundfield out of such set of ambisonic-encoded channels, and as a live-instrument that allows free positioning and movement of a number of virtual sources in real-time.

1. INTRODUCTION

One of the most demanding tasks of electro-acoustic music has always been composition of space. While the basic principles of periphonic sound (re)production have been known for decades, these approaches could not be used in "live processing"-environments due to the lack of computational power, which is needed to do multi-channel signal-processing, the basis of any periphonic production and reproduction, in real-time.

In the last few years, the computer industry has reached a point, where multi-channel digital sound processing can be done on the basis of commercial personal computers.

A system, that can be used for live-rendering of periphonic sources in concert-situation as well as for post-production, has been developed: *abcde*, an *A*mbisonic *B*ased *C*oding and *D*ecoding *E*nvironment.

2. SOFTWARE-DESIGN

Due to the rapid development of ever faster hardware, the structure of a PC-based system has to be scalable enough, to keep advantage of future improvements. In terms of audio-processing, more computational power can easily be used by incrementing the number of channels that are processed simultaneaously and by increasing the accuracy of the system.

However, this scalability should be transparent to the user. Therefore, the "frontend" (e.g.: the userinterface) has to be separated from the "backend" (e.g.: the audio-engine that does all the signal-manipulation needed to render a number of virtual-channels onto an array of loudspeakers).

This architecture allows us to define user-interfaces that provide access to exactly the functionality that is needed by the user, while the backend can be scaled to a compromise between the current needs and the hardware possibilities [1].

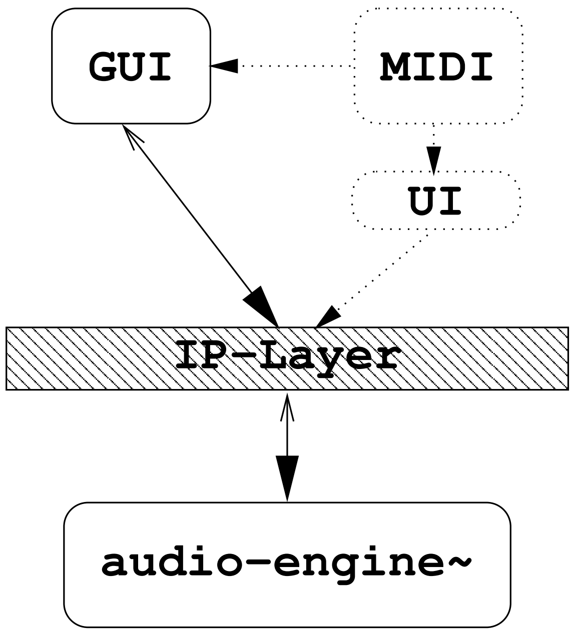

Figure 1: *Software-structure of the periphonic renderer* abcde

Frontend and backend are connected via a network-based transport layer. Execution of the audio-engine and of the user-interface can thus take place on different computers to keep the computational load in realistic ranges.

3. AMBISONICS

The main goal of the audio-engine is to render a number of virtual (mono) sources onto the speakers in such manner, that an impression of the original soundfield is given to the listeners. The input-signals are to be manipulated somehow to achieve the speaker-feeds that give this impression. However, it is often desired to record a representation of the periphonic soundfield, that enables an easy reproduction. The speaker-feeds are unfavourable, since they are bound to a special setting of speakers and normally the number of channels needed is quite high. An intermediate format that holds a maximum of information about the soundfield at a minimum number of transmission channels is therefore needed.

The ambisonic approach ([2], [3]) matches this criteria. The basic idea is to decompose an existing (or virtual) soundfield of plane waves into an infinitely long Fourier-Bessel-series, the so called *spherical harmonics*. Naturally, this series has to be truncated at some point, which defines the order of the approximation to the original soundfield [4]. In the periphonic case (which is considered here) the number L of transmission-channels with respect to the order M of the system is given by

$$L = (M+1)^2 \tag{1}$$

To obtain such a set of transmission channels for a plane-wave out of the direction $[\varphi, \vartheta]^{\mathrm{T}}$, directional characteristica (the "spherical harmonics" which are basically linear combinations of trigonometric terms) are applied to the source-signal. Several sets of ambisonic-channels can be superimposed. To obtain the feeds for the speakers an inverse encoding is used that ensures, that the Bessel-Fourier-series of both the analysed and the synthesized soundfield match up to the given order.

The ambisonic approach guarantees, that any number of virtual sources can be represented by a fixed number of transmission channels and that the transmission channels are completely independent of the speaker-layout used for reproduction.

Encoder (of virtual sources) and decoder (to speaker feeds) of the ambisonic representation of a soundfield are completely independent of each other, as long as there is some agreement on the actually used encoding rule.[1]

4. AUDIO-ENGINE: PRODUCTION

The task of the encoder, is to render a number of virtual sources with positional information into a full representation of the resulting soundfield.

4.1. Directional positioning

The task of directional positioning with respect to $[\varphi, \vartheta]^{\mathrm{T}}$ is fully accomplished by the ambisonic encoding.

To stay as configurable as possible, encoding rules are entirely defined as plugins that can easily be extended, modified or rewritten.

Directional movement $[\frac{\partial\varphi}{\partial t}, \frac{\partial\vartheta}{\partial t}]^{\mathrm{T}}$ can be obtained by simpling interpolating between sequential positions.

4.2. Distance and movement

The ambisonic principle is based on the assumption, that the soundfield solely consists of plane waves. Of course, this is not true for most (if not all) "real" soundfields.

Adaptions to the standard ambisonic encoding have been proposed by [5] to put this into account. For example, it is easy to ensure a blurring of very near virtual sources which occurs because near sources appear to be larger than those further away, by strengthening the omnidirectional transmission channel (the 0^{th}-order channel) with respect to the higher order channels that bear more directional information.

[1]For historic reasons there exist several encoding rules which are basically the same but differ in constant multiplication factors applied to the channels, e.g. to obtain equal levels on all channels.

Apart from this, sources that are farther away reach the ear later (due to the finite speed of sound) and lack of higher frequencies (due to the damping characteristic of the air) than sources originating from a close position.

The damping effect follows a $\frac{1}{r}$-law and can be approximated by a 1^{st}-order low-pass.

The delay imposed by the finite speed of sound does not really matter in a static and/or anechoic environment. However, if reflexions of a single sound-source are present, the run-time differences between these reflexions will be used to estimate the distance by the hearing apparatus.

4.2.1. Doppler-effect

If a sound moves towards a listener or away from the listener fast enough, the listener will notice a pitch-change of this sound. This so-called "Doppler-effect" is due to the superposition of the static propagation-speed of the sound in the air and the speed of the sound-source relative to the listener. If a sound-source moves towards a listener, the crests and troughs of the pressure waveform will be closer together than they would be if the sound-source and the listener would stay at a constant distance. The very same thing happens, if the signal is delayed (due to the finite propagation speed of sound) and the amount of this delay is changed smoothly [6].

Thus a variable, smoothly interpolating delay-line can be utilized to create the impression of fast moving sound-sources.

4.3. Spaciousness

In addition to position a virtual source in space, it is important to define the acoustic space in which the sources are positioned. The acoustic room is mainly defined via the occuring reflexions.

4.3.1. First reflexions

The first reflexions that come from the walls can be perceived individually and are used by the hearing apparatus to estimate the distance from the sound-source as well as the distance from the reflecting walls.

While these impressions are important for the perception of a room, they are computationally expensive, since each reflexion has to be calculated separately as a mirrored virtual source. In the simple case of a box-room, each virtual source will thus produce six additional mirror sources for very the first reflexions.

The computational expanse can be reduced drastically, when considering spheric rooms. If the listener is situated in the center of such room, the number of first reflexions can be reduced to one. Additionally this first reflexion comes from the same direction as the original source. It is therefore enough to apply a distance weighting (low-pass filtering and delaying) to source and mirror source separately. Afterwards the directional weighting must only be applied to a superposition of source and its reflexion.

Since spheric rooms are not very common, this simplification does not give a very good image of the room. However, it gives a good impression of the distance of the sound at an extremely low cost. If better approximations of the first reflexions are required, it is easy to define virtual sources that are mere mirror sources to induce the perception of localizable reflexions.

4.3.2. Reverberation

While the first reflexions can be perceived individually for each source, they soon become too dense and chaotic to be calculated. Therefore it is sufficient to not calculate the reverberation for each virtual source separately, but to reverberate only a mixdown of the sources – in this case, the superimposed ambisonic channels are reverberated. Since the directional information of the diffuse field is quite rough, it is sufficient to encode the reverberated soundfield only with low-order ambisonics.

Since many good-sounding, easy-to-use reverberators exist as standalone devices, it possible, to plug such an external device into the encoder.

5. AUDIO-ENGINE: REPRODUCTION

The output of the encoding unit is a complete representation of the periphonic soundfield in an ambisonic format. The decoding unit recreates the soundfield in an inverse process to the ambisonic encoding. Therefore, decoder and encoder share the plugin that defines the set of encoding rules to use.

The decoding process depends on the actual loudspeaker-layout. This layout is passed to the decoder via a configuration-file. Thus, the decoding unit can be used to decode an ambisonic soundfield to virtually any loudspeaker-layout.

It is possible to rotate the whole soundfield to adjust the "front" direction to the actual orientation of the audience.

Basic ambisonic decoding works only very well in a small sweet spot. This is highly desirable in a production environment, when the mixing engineer is seated within this small area to obtain the best possible result.

However, if a periphonic soundfield is to be recreated for a large audience (as it is in concert situations), most of the listeners will be outside this sweet spot. It is therefore necessary to enlargen the sweet area of good reproduction at the expense of the excellent reproduction quality in the original sweet spot. The largest sweet area can be obtained by the so-called *in-phase*-Decoding ([7], [8]).

To optimize the trade between large reproduction area and quality, it is possible to crossfade between the two decoding algorithms.

6. FRONTENDS

Normally the user need not be aware of (and to worry about) the used encoding-rule, the loudspeaker setting and other interna of the audio-engine. Instead, it should be possible to intuitively position a number of virtual sources.

6.1. A simple Interface

The freedom of being able to position each virtual source separately might often turn out to be a pain. Especially when the system is used to play back common settings via an arbitrary loudspeaker array. This includes "standard situations", like having to simply play back a historic stereo- or quadrophonic recording. It is surely unhandy, if everytime a 5.1 recording is played back, one has to adjust all six channels by hand.

A simple interface is able to store such settings and reproduce them on demand by simply clicking a button.

Thus it is possible to easily play back any kind of decoded (this is, the speaker-feeds are directly available) multi-channel recordings.

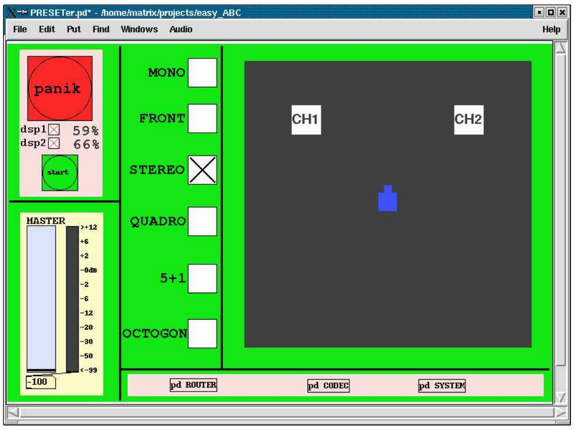

Figure 2: a simple interface for using the periphonic renderer *abcde*

6.2. A fully periphonic Mixer

However, the strength of a periphonic mixing-system is not the play-back of historic multi-channel recordings, but the availability of free positionining and movement of virtual sources. Therefore a mixer-interface that allows full control of these parameters is needed. The adjustable parameters are $[r, \varphi, \vartheta]$, since spheric coordinates are "native" to ambisonic systems.

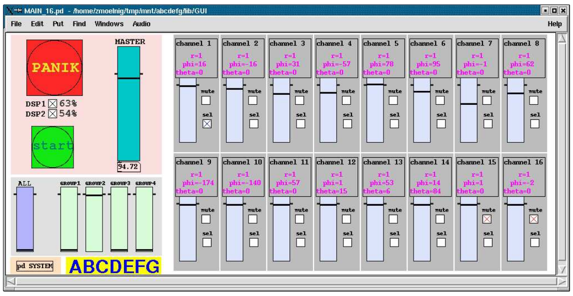

Figure 3: a periphonic mixer interface for *abcde*

Because this mixer application has no built-in sequencer, only static (non-moving) sources can be rendered. However, since the audio-engine interpolates between consecutive positions, a movement of sound-sources will be perceived.

For a smooth movement it is necessary to provide a constant stream of positions for each source. This stream can either be provided by turning a controll-knob on the user-interface (which is limited, since it is hard to move more than one source manually) or by an external sequencer.

6.2.1. Interfaces to other software

Two different interfaces for external sequencers are defined. These interfaces communicate with the (graphical) user interface rather than directly with the underlying audio-engine, to keep the used hardware totally transparent.

- **MIDI:** To allow communication with almost every sequencing software a bi-directional MIDI-interface has been implemented. This allows recording and play-back of position-tracks for up to 32 virtual sources.
- **OSC/FUDI:** An unlimited number of sources can be positioned via an ethernet-connection that is based on either the Open-Sound-Control- or the (similar) FUDI-protocol. Thus, complex movement paths can be generated in real-time with Software à la pure-data (and similar Max-like dialects) and SuperCollider.

7. LOUDSPEAKER-SETUP

While the decoding-unit of the audio-engine is capable of decoding ambisonic soundfields to almost any loudspeaker-layout, best results will be achieved with regular positioning [7]. Although it is easy to get regular two-dimensional polyhedrons, this is not trivial for the three-dimensional case.

Additionally, it is often impossible to mount loudspeakers in an approximate sphere around the listener(s), simply because of architectonic reasons.

While the human hearing apparatus is capable of localization of periphonic sources (if not, this article would be void), sound-sources that are within the horizontal plane can be localized much better than sources that come from "above".

Since the number of reproduction-channels is always limited, it therefore makes sense, to mount relatively more speakers in the horizontal plane than in the third dimension.

The layout of a hemisphere that is sectioned into several horizontal rings of speakers proved to be good layout.

At the *IEM-Cube*, 24 loudspeakers have been aligned in three such rings, where the lower ring consists of twelve speakers to achieve a maximum localization in the horizontal plane. The middle ring is built of eight loudspeakers, where the remaining four speakers form the top ring. (see fig.4, for a detailed description of the layout, see [1])

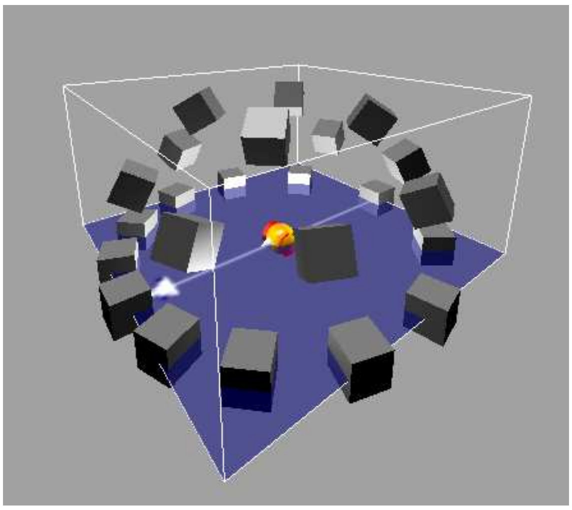

Figure 4: speaker-layout in the *IEM-Cube*

8. CONCLUSIO

A periphonic production and reproduction system for the use in realtime-environments has been shown.

While the software is scalable and configurable to run in almost any environment, it proved necessary to provide a high-end realisation of such an environment. This has been implemented at the *IEM-Cube*, a medium-sized concert-hall that enables reproduction of ambisonic soundfields of at least 3^{rd} order.

The audio-engine has been realized originally on two PentiumIII-800MHz computers, which enables the live encoding of up to 24 virtual sources that are fully positionable and smoothly moveable in real-time.

For even higher requirements, the encoding-unit can be executed on a high-performance PC, which makes it possible to render about 50 individual virtual sources into a 3^{rd}-order ambisonic soundfield in real-time.

Thus an musical instrument has been made, that enables a large auditory in typical concert situations to perceive compositions of space.

9. REFERENCES

[1] Johannes M. Zmölnig, *Entwurf und Implementierung einer Mehrkanal-Beschallungsanlage*, master's thesis, Institute of Electronic Music and Acoustics, Graz, Austria, Jan. 2002.

[2] P. Fellgett, "Ambisonics Part One: General System Description," in *Studio Sound*, vol. 17, pp. 20-40, 1975.

[3] Jeffery S. Bamford, *An Analysis of Ambisonic Sound Systems of First and Second Order*, master's thesis, University of Waterloo, Waterloo, Ontario, Canada, 1995.

[4] Robert R. Höldrich and Alois Sontacchi, "Konzepte zur Schallfeldsynthese und Schallfeldreproduktion," in *Proceedings des FA für Akustik, Jahrestagung der ÖPG*, Graz, Austria.

[5] Dave G. Malham and Richard Furse, "the Furse-Malham Set," at *http://www1.york.ac.uk/inst/mustech/3d_audio/secondor.html*, 1992

[6] Durand R. Begault, "Challenges to the Successful Implementation of 3-D Sound," *J. Acoust. Soc. Am.*, 1991.

[7] Jérôme Daniel, *Representation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, phd thesis, Université Paris 6, Paris, France, 2000.

[8] Dave G. Malham, "Experience with Large Area 3D Ambisonic Sound Systems," in *Proc. Inst. Acoust.*, York, 1992