The IEM CUBE

Contributors: Johannes Zmölnig
Winfried Ritsch
Alois Sontacchi
Robert Höldrich robert.hoeldrich@kug.ac.at

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 for multi-channel signal processing, the basis of any periphonic production and reproduction, in real-time.
In the last few years, the computer power 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 a concert situation as well as for post-production, has been developed: *abcde*, an *A*mbisonic *B*ased *C*oding and *D*ecoding *E*nvironment.
Its high-performance application is the ***IEM CUBE.***

Figure 1: panorama view of *IEM CUBE*

2. Software-Design

Due to the very fast 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 either 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) is 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 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.
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. Audio-Engine

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.
Furthermore, it is often required to record a representation of a periphonic soundfield, that can be reconstructed easily. Direct speaker-feeds are unfavourable, since they are bound to a special setting of speakers and normally the number of channels needed is quite high.
Additionally, such a transmission-format should be backward and forward compatible. This means, that, on the one hand, extensions of the system should be able to handle "old" recordings, while on the other hand "old" systems (without extensions, p.e. due to lack of computational power) should be able to play extended recordings without problems too. (Remember the transition from mono to stereo and the so-called "MS-stereophony". Old mono-systems could reproduce a mono-mixdown out of a stereo-transmission without any problems.)
The ambisonic approach (Gerzon, 1973) matches this criteria for the periphonic case. To obtain a set of transmission channels for a plane-wave out of the direction $[\varphi,\vartheta]^T$, directional characteristica (the "spherical harmonics") are applied to the source-signal. With larger number of transmission-channels better approximations of the analysis-soundfield can be reconstructed. However, while the number of transmission channels increases with respect to the order (and thus the accuracy) of the system in the pantophonic (horizontal) case, in the periphonic case it increases with respect to the square of the order.
In basic ambisonics the accuracy of the reproduced sound-field is constant from all directions. However, the human hearing apparatus is more accurate in the horizontal plane than in the vertical one. Thus it is sensible to transmit more information on the horizontal plane. This can be achieved with mixed-order ambisonics, that uses relatively low orders for full-periphonic channels and higher orders for horizontal-only channels. Thus the number of transmission channels can be reduced. For instance, a 1st+3rd-system needs only eight transmission channels, while a fully periphonic 3rd-order system would need sixteen channels.
To achieve the loudspeaker-feeds, an inverse encoding is used. This guarantees, that analysed and synthesized soundfield match up to a certain order, that is defined by the "size" of the system.
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.

4. Production of periphonic soundfields

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

The task of directional positioning with respect to $[\varphi,\vartheta]^T$ is fully accomplished by the ambisonic encoding. Directional movements can be obtained by simpling interpolating between sequential positions.
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 (e.g. Sontacchi/Höldrich 2002). For example, it is easy to ensure a blurring of very near virtual sources - near real sources appear to have larger width than those further away - by strengthen the

omnidirectional transmission channel (the 0th-order channel) with respect to the higher order channels that bear more directional information.
Apart from this, damping effects and delays have to be applied to enable correct distance encoding. The damping effect follows a 1/r-law and can be approximated by a 1st-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.

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

4.1 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.

First reflexions The first reflexions that come from the walls can be perceived individually and are used by the hearing apparatus to get the distance of the sound source as well as the distance of 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 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 on 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.

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. Reproduction of periphonic soundfields

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.
The decoding process depends on the actual loudspeaker-layout. This layout is passed to the decoder via a configuration-file. Therefore, 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.
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 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.

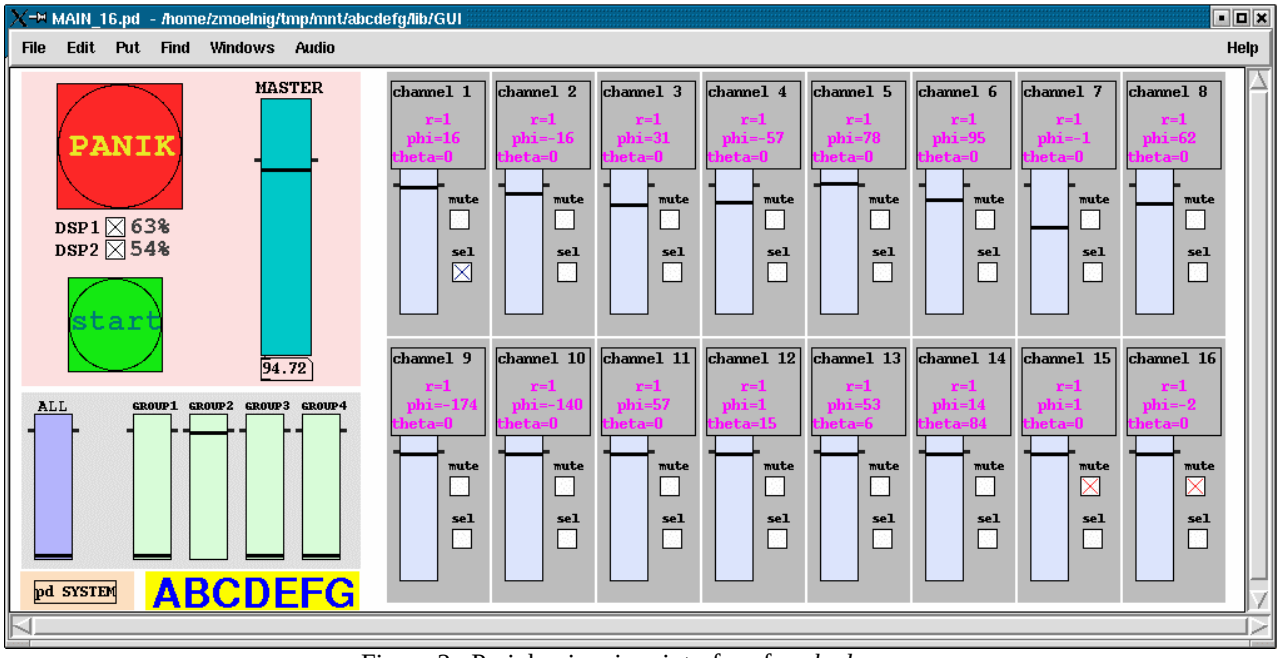

Figure 2: Periphonic mixer interface for *abcde*

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 Sequencing software

While the audio-engine can only be controlled via ethernet, the mixer-interface also allows MIDI connections. Thus it is possible to control positioning and movement of virtual sound sources by any MIDI-sequencer.

6.3 More generic controllers

Since the audio-engine is controlled via ethernet using as transmission protocols both FUDI and the well known OpenSoundControl (OSC), any controller (even internet-based) that can talk via these protocols can be used.

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. 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.

8. Conclusion

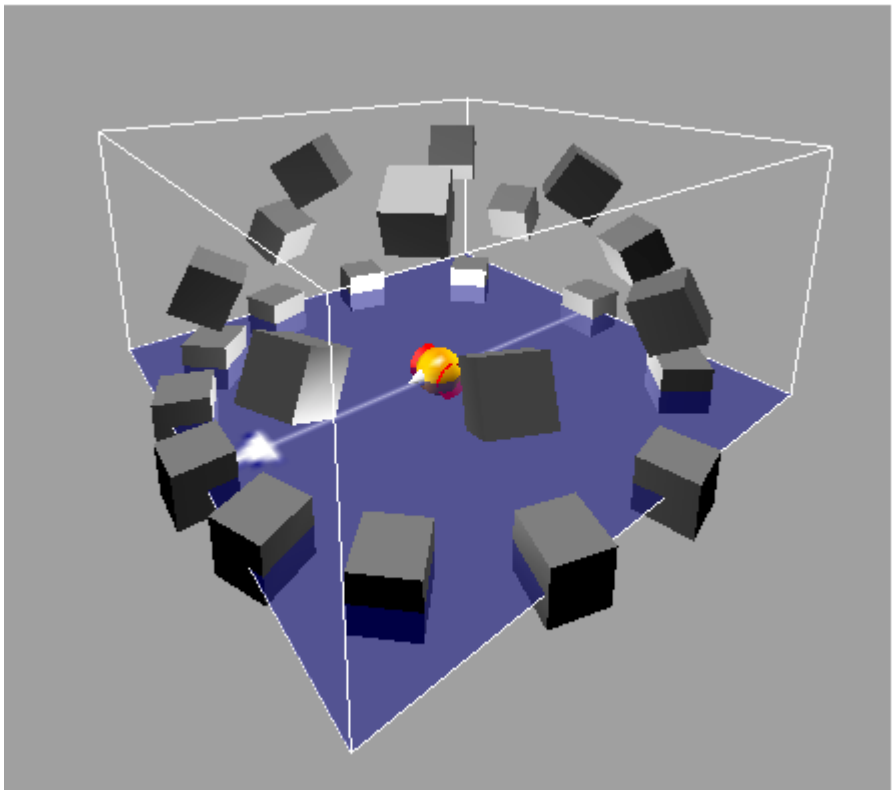

Figure 3: speaker-layout in the *IEM CUBE*

A high-end realisation of the system has been implemented at the *IEM CUBE*, a medium-sized concert-hall that enables reproduction of ambisonic soundfields of at least 3*rd* order.
24 loudspeakers have been aligned in three 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.3)
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.
On state-of-the-art PCs, it is possible to render more than 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.