Reverb and Room Simulation in the Multichannel Era

Knud Bank Christensen
TC Electronic A/S
Sindalsvej 34
DK-8240 Risskov, Denmark

This paper presents a family of algorithms that enable music-, film- and game producers to create engaging virtual acoustic environments around the end user and control the positioning of sound sources therein. The problems of addressing multiple formats are discussed, as is image precision vs. real-time motion capability. Progress in lateral source positioning capability and centre channel integration is reported.

THE MULTICHANNEL CHALLENGE

The introduction of multichannel audio for consumer use presents new possibilities and challenges for developers of audio production tools such as the ubiquitous Reverb. With the end user surrounded by discretely controlable loudspeakers, the audio producer should be able to do better than just adding “ambience” or present the listener with a frontal sound “stage” with added “surround effect”. The production tool should enable him or her to create an artificial acoustic environment around the listener and convincingly unfold the acoustic event under production inside that environment.

Requirements and possibilities

When we started conceptualizing such a Reverb or Room Simulation system, a number of requirements and questions eventually came to mind:

1. The algorithm(s) should be able to simulate both natural and highly unnatural acoustic environments.
2. It should facilitate credible and predictable rendering of room geometry and source- and listening positions therein, something that simple power-panning cannot achieve.
3. It should make the best possible use of at least 3, maybe 4 target formats: Discrete 5.1 in ITU-775 setup, discrete 5.1 in cinema setup, stereo and possibly headphones.
4. The produced discrete multichannel material should be down-mixable to 4:2:4 matrixed surround, stereo and even mono without notable artifacts.
5. The system should deliver good simulation within a fairly large “sweet spot” without producing notable artifacts at positions outside that sweet spot.
6. Continuous, real-time control of source positions should be possible.
7. Music producers have often refused to use the Center Channel in spite of its potential for stabilizing the stereo image, simply because it tends to sonically “stick out” of the scene. We would like to find out why, and do something to integrate it and make it useful in music production.
8. The ITU-775 setup is sometimes referred to as “3/2 format” indicating a division between a 3-speaker frontal sound stage and a 2-speaker rear “surround”. We would like to try mending this division, making one continuous sound stage albeit with direction–dependent resolution.

A SOURCE ORIENTED ROOM SIMULATION ALGORITHM

Our first multichannel algorithm aimed at high-resolution simulations of a limited number of sources in an environment with a predetermined choice of source- and listening positions. It was built from five types of elements (fig. 1):

1. An **Early Pattern Generator** (EPG) for each source input creating up to more than 50 early reflections per source and mapping them to an internal multidirectional representation (up to more than 25 directions)
2. A vectorial adder superposing the multidirectional early pattern signals (including direct sound) from the sources.
3. A **Direction Rendering Unit** (DRU) rendering the multidirectional sum of early reflection patterns via the chosen speaker setup
4. An independent **Reverb** Late Response Generator for each loudspeaker output and
5. A **Reverb Feed Matrix** controlling the level, delay and color of each source’s contribution to each reverb generator.

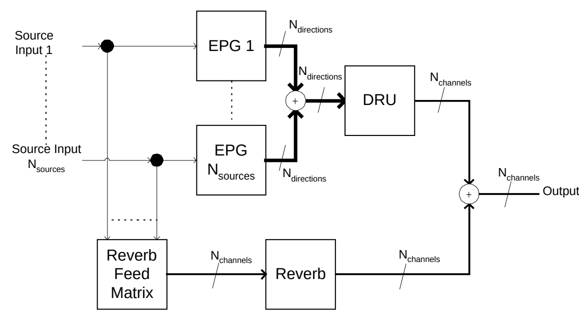

Fig. 1 Overall block diagram of Room Simulator (gain controls not shown)

The Early Pattern Generator (EPG)

The EPG produces a pattern of early reflections and maps them and the direct sound (the 0^{th}-order reflection) to an intermediate multidirectional format. The reflections are mapped into a quantized 5-dimensional parameter space according to

- Level
- Time of arrival
- Order
- Spectral color
- Direction of arrival

This is done by a matrix of delaylines as shown in fig. 2.

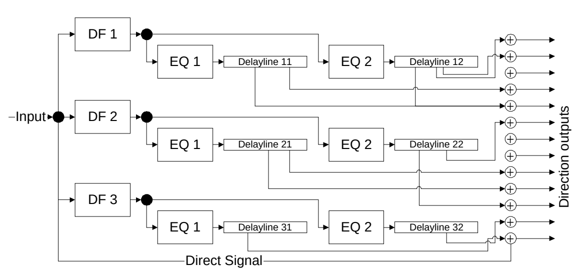

Fig. 2 Exaple of EPG (level adjustments not shown)

The rows of the matrix are fed through different Diffusion Filters producing increasing levels of diffusion, the columns represent different spectral color groups (EQ settings). When a reflection of order O, level L, color C, arrival time T and direction of arrival D is to be added to the pattern, we first choose a row based on the reflection order O and a column whose color matches C the best. Then we take out a T-delayed signal from this delayline, multiply it by the level L and add it to the output node that comes closest to representing direction D. The direct signal is not subjected to diffusion or coloring. The level of the direct signal and the group of early reflections can be set independently.

The initial setting of an EPG for a particular pair of source- and receiver positions within a particular room is derived automatically from simulation output from the ray-tracing program Odeon [1], which can import architectural CAD files for easy generation of new rooms. Besides simulating real rooms, this also enables our Virtual Room Designers with their backgrounds in classical music, rock music or film/post-production to build “ideal” or “special effects” rooms from scratch. However the sound of these automatically generated reflection patterns is often flawed one way or the other, and as theory would suggest, the ray-tracing method works best above the Schröder frequency where modal effects can largely be ignored. Hence, from this imperfect starting point, the designers put a lot of work into tuning the reflections by ear to obtain clear and consistent source positioning and timbre within each virtual room. The necessity of hand tuning makes this algorithm unfit for simulating real-time motion. However, in our view, the clarity, precision and naturalness of the sources-in-a-room simulation that can be obtained by this method justify the tuning work and the lack of motion capability.

The Direction Rendering Unit (DRU)

The job of the DRU is to render the superposition of all the early reflection patterns via the chosen loudspeaker topology. The reflection patterns come in the multidirectional internal format with up to more than 25 individual directional components. The directions are horisontal with more resolution in the front than elsewhere. Initially a number of rendering techniques was considered, and from a theoretical point of view Ambisonics [2] or HRTF-based transaural rendering [3] looked interesting. However, since these techniques rely more heavily on the precise interaction of multiple acoustic transfer functions (e.g. for interaural crosstalk cancellation) than simple panning, they seem more likely to create notable artifacts at off-center listening positions and be less tolerant to subsequent down-mixing. For these reasons and for the reason of simplicity, we decided to start out with a simple signed gain matrix (fig. 3).

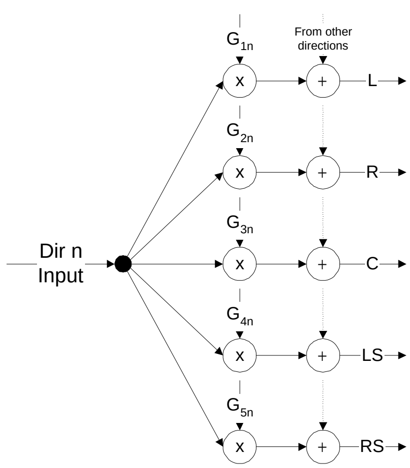

Fig. 3 Single direction slice of DRU (gain matrix) including common output summing nodes

As usual, it was tuned by ear while keeping the use of anti-phase components to a minimum. (Actually the DRU was needed for working on the EPGs, so it was tuned first). Though such a simple panner is not able to produce good localization of a single source, especially at lateral positions, it turned out to work just fine when applied repeatedly to each of a large number of multidirectional early reflections. For a non-uniform target like the cinema setup with quite different speaker types involved, additional filtering of the DRU outputs sometimes help sharpen the image and improve timbre consistency. In the ITU-775 setup, this combination of EPGs and this simple DRU seems able to produce quite credible and stable rendering of sources in a room at most horizontal positions around the listener, **even straight to the side!** By now we haven’t done a proper psychoacoustical experiment to quantify this experience, but preliminary results are reported in [4].

Addressing multiple target formats

In principle, all the target formats (ITU-775 5-channel, cinema 5-channel, stereo and headphones) could be addressed by making a separate DRU for each. However, this soon turned out to be suboptimal: In cinema 5-channel format, the rear channels are fed to large groups of loudspeakers placed all along the side and rear walls. This format does not allow active use of the rear channels for direction rendering, especially because parts of the audience are located much closer (by many meters) to the nearest surround speaker than to the front speakers. And in stereo, the rear and extreme side directions cannot be rendered without the use of artifact-prone transaural techniques, which may not work even at sweet spot position due to room reflections or individual HRTF deviations. So even though we liked the modularity of our initial algorithm, we chose to yield to reality and make separate EPG versions for the 3 main target setups: ITU-775 multichannel, cinema multichannel and stereo. The headphone target can be addressed by adding a newly developed binaural mapper to the output of the ITU-775 version of the algorithm.

The Reverb Unit

Only very dry rooms can be simulated with the early reflections described above. A diffuse reverb “tail” must be added to complete the simulation. This is accomplished by a separate feedback network of delaylines and filters for each output channel. This reverb algorithm (whose detailed structure must remain a trade secret) is characterized by maximum tuneability at the cost of computational efficiency: Each applies more than 20 delaylines and 40 filters. The reverb algorithm can be tuned to any desired decay time (within reason) as a function of frequency. Furthermore different, completely uncorrelated versions can be created with similar or differing degrees of “smoothness/liveliness/coarseness”. Earlier designs have often used modulation to average out unwanted spectral or temporal coarseness. With meticulous tuning by ear, the new algorithm becomes so smooth that modulation is no longer needed, leaving the reverb as a linear time-invariant system and thus ensuring perfect pitch-correctness. Still modulation may be added for effect. The use of completely separate and uncorrelated Reverbs for each output ensures that subsequent down-mixing to 4:2:4 matrix, stereo or mono does not cause unwanted side-effects such as cancellation of anti-phase components or misplacement of reverb signals by matrix decoders. Our reverb tuning experts have created about 20 uncorrelated tails, which can be combined into a large number of different 5- or 2-channel sets. These in combination with use of the tails’ diffusion controls make it possible to create a subjectively efficient though – from a Theoretical Acoustics point of view – conceptually self-contradicting “Diffuse Field Size” control, largely independent of decaytime and color.

The Reverb Feed Matrix

The Reverb Feed Matrix (fig. 4) connects each source input to the Reverb Unit.

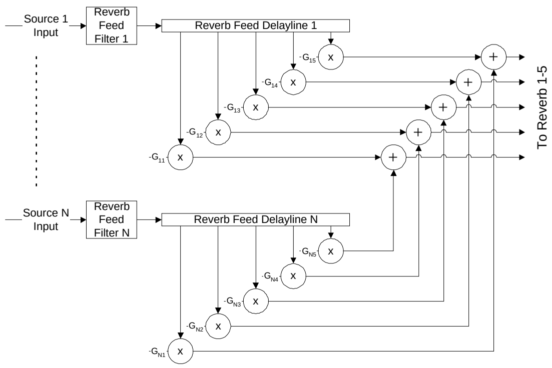

Fig. 4 Reverb feed matrix, only 2 sources shown.

The Reverb Feed Filters control the color of each source's diffuse-field contribution. The gains can be used to position the spatial "centre of gravity" of the reverb and the tapped delayline helps control its spatial/temporal properties during build-up. This leads to yet another efficient though self-contradicting control: The "Diffuse Field Position", one for each source input.

A MULTICHANNEL PROGRAM ORIENTED VERSION

The source-oriented room simulator described above does not fit the normal use of Reverb processing. First of all it is an insert effect, where "normal" reverbs are additive effects that leave the direct signal untouched. And secondly, the user does not always have single source signals to place in a virtual room, but instead a half-finished mix in need of added "space". To make such an additive 5-channel in, 5-channel out version of the algorithm, we used the source-oriented room simulator (without the direct signal "0^{th} order reflection") to place virtual loudspeakers in our virtual rooms and then play the multichannel input material through these virtual loudspeakers. Additionally the input can be fed directly to the output speakers (fig. 5).

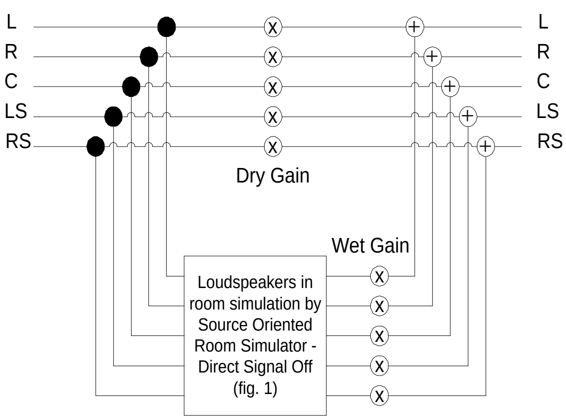

Fig. 5 Implementation of "conventional" additive multichannel reverb.

In general, but most notably in this situation, the "dry" input may not be dry at all: It could be a multichannel recording of a complete acoustic event, including its own room reflections. So the user is given means to trim and adjust the algorithm's early reflection patterns based on early/late arrival, direction or order, to make them complement rather than mask the spatial information in the original recording.

REAL-TIME SOURCE POSITIONING

As mentioned above, the sources-in-a-room simulation algorithm is depending on factory tuning by ear of the large early reflection patterns that define the room geometry and source- and receiver positions within it. Until we find a way to generate sufficiently good reflection patterns automatically and in real time, we have to produce another answer to the need for continuous real-time source positioning. Our solution to this has been to add an 8-input/5-output gain/delay panning matrix to the input of the multichannel program oriented reverb described in the previous section (fig. 6).

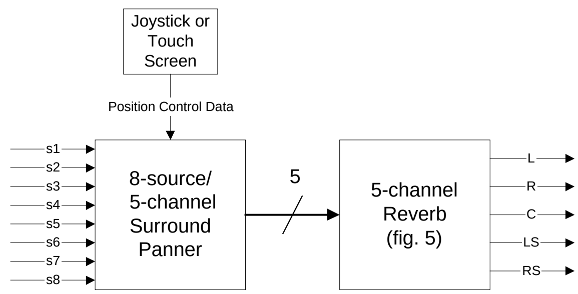

Fig. 6 Real-time multiple source positioning using additional panner.

THE CENTER CHANNEL INTEGRATION PROBLEM

For the past several decades, high quality loudspeakers for studio monitoring as well as domestic use have been aimed at the traditional stereo setup. Thus, when optimizing loudspeakers, the primary objective is timbre neutrality of the phantom center image where the solo singer or instrument is normally placed. When switching from stereo to multichannel audio, the phantom center was replaced by a physical center speaker. Immediately music producers started complaining about the new format, declaring the center channel virtually useless because – for some reason – it stuck out of the stereo image like a sore thumb, even when using three carefully matched identical front speakers. With the benefit of hindsight, this problem is easily explained (fig. 7).

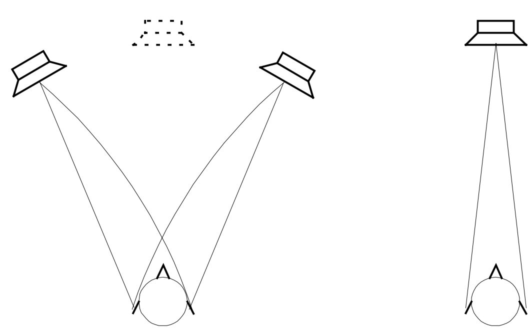

Fig. 7 Phantom vs. physical center channel

The signal reaching each ear from the phantom center channel is the sum of two signals: One from the loudspeaker located to the near side of the head, and one arriving a bit later from the loudspeaker located to the far side of the head. Thus the familiar phantom center signal at each ear is – to a crude approximation ignoring the HRTF's involved – a comb-filtered version of the signal that we're exposed to from the physical center channel. And the loudspeakers are optimized for the phantom situation, so the physical center channel got its bad reputation for **lack** of comb filtering! So, rather than waiting for everybody to design, distribute and apply special center channel loudspeakers, we offer a special optional filtering of the center channel signal for music production. And – in our own humble opinion – it helps a lot, but again we haven't attempted to quantify the effect via a psychoacoustic experiment.

META PARAMETERS – OBTAINING SIMPLICITY BY ADDING COMPLEXITY

The translation between a few dozen of intuitively useful user parameters [5] and the underlying more than one thousand physical parameters (gains, frequencies, delays, directions, etc…) is done by a rule based "Metaparameter Engine". These "Meta Rules" are themselves parameters of each of the designed virtual rooms and is – like almost everything else in this system – tuned by ear. The meta rules determine how each metaparameter affects a number of physical parameter, possibly depending on the setting of other user parameters also. Often it takes six or eight of these "metaparameter-parameters" to define the relationship between one metaparameter and one physical parameter. Thus a complete setup of the algorithm may require more than 3000 parameters.

IMPLEMENTATION

The algorithms described above have been implemented on our commercially available Multichannel Processing Platform [6] where they occupy between two and four 100 MHz Motorola DSP563xx signal processors with separate banks of high-speed SRAM, depending on sampling-rate and panning options. The use of hundreds of delaylines and filters, make these algorithms so demanding in terms of DSP instructions and memory bandwidth, that - for the time being - they can only be implemented economically on dedicated hardware (our development prototype runs on a rather expensive 4-CPU 64-bit RISC based SGI computer).

CONCLUSION

A flexible family of Room Simulation and Reverb algorithms has been presented. It enables audio producers to place sources in virtual acoustic environments with good precision and render them for stereo, cinema 5-channel format, ITU-775 format or (with additional processing) headphones. The core algorithm relies on detailed patterns of early reflections and separate, uncorrelated reverb tails for all output channels. The early reflection patterns allow source positioning in most horisontal positions via the ITU-775 setup. Perceptually they enhance localization and room definition to a level that we don't believe could ever be achieved by a simpler and much cheaper "power panning plus reverb" algorithm. The use of uncorrelated reverb tails prevents anomalies from occurring after subsequent down-mixing to other formats.

A conventional additive 5-channels in/5-channels out Reverb version of the algorithm has been made. In conjunction with an 8-input surround panner this allows real-time motion of sources in a virtual room, albeit with reduced precision compared to that of the stationary source-oriented algorithm.

A special center channel filtering has been devised, which helps to integrate the center speaker into the frontal sound stage, making it useable in music production, while retaining its image stabilizing effect.

The good source positioning precision in virtual rooms and the improved center channel integration in music productions that we have experienced, have so far not been formally quantified through systematic experiments. But the market response has been good.

REFERENCES

[1]: Odeon is a ray-tracing based room simulation program under constant development at the Technical University of Denmark (http://www.dat.dtu.dk/~odeon/). We have only used the initial part of it that calculates reflectograms.

[2]: Jérôme Daniel, Jean-Bernard Rault & Jean-Dominique Polack: "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions", AES Preprint no. 4795, 1998

[3]: K.C.K.Foo, M.J.Hawksford & M.P.Hollier: "Pair-wise loudspeaker paradigms for multi-channel audio in home theatre and vitual reality", AES Preprint no. 4796, 1998.

[4]: Thomas Lund: "Enhanced Localization in 5.1 Production", AES Preprint no. 5243, 2000.

[5]: TC Electronic System 6000 Manual, sections "VSS-5.1 Source" and "VSS-5.1 Reverb" (http://www.tcelectronic.com/files/M6_Algos_Presets_130.pdf)

[6]: TC Electronic System 6000, http://www.system6000.com, http://www.tcelectronic.com.