Audio Engineering Society
Conference Paper
Presented at the 19th International Conference
2001 June 21–24 Schloss Elmau, Germany

Multichannel Signal Processing Tools - differences to multiple single channel processing

Søren H. Nielsen

TC Electronic A/S

DK-8240 Risskov, Denmark

ABSTRACT
Changing sound recording and production from stereo to multichannel creates some new demands on the signal processing tools. Alone from a handling point of view there are advantages by integrating processing of multiple channels into one unit instead of using several in parallel. But many processing tasks simply need multichannel integration in order to function appropriately. Some of these tasks and proposed tools are presented, e.g. reverb, dynamics processing and format conversion.

INTRODUCTION

In the transition from stereo to multichannel many methods and much equipment can be re-used. For static processing such as equalisation multiple mono or stereo processors may be used, and maybe even be coupled together on the user interface, by using MIDI or a similar connection, if this is wanted.

The mixing consoles are often usable also for multichannel production even if originally laid out for stereo. Some tasks, however, cannot be performed appropriately by using mono or stereo tools. Dedicated multichannel tools are needed.

Even though multichannel is relatively new in the consumer world it has been common practice for many years in the cinema. Many guidelines and standards already exist in this area published by international and regional organisations and societies. Some companies have developed methods and products which to some extent serve as de-facto standards.

SOURCE PLACEMENT - ROOM SIMULATION

One of the most basic tasks is placement of sources in multichannel space. Whether this is done by a simple power panner or some advanced room simulation technique a close connection between the channels is needed.

In the rare case where a single multichannel microphone array can deliver a satisfying multichannel signal the placement issue disappears from a processing point of view.

The methods used for source placement depend on the playback loudspeaker layout. Currently, mainly two playback situations exist: Cinema and home. In the cinema the geometry is typically rectangular and the distances are large [1] whereas in the home a relatively small circular setup like ITU-R BS.775 [2] is aimed at. Due to the large distances involved in cinema reproduction the normal stereo methods using pairwise level differences and small time delays do not work well. In the ITU, or home, setup a more precise source placement can be achieved. It has proven successful to use well defined patterns of discrete reflections to support placement of sources in space [3], [4].

When using discrete reflections to place sound sources in space each of the input signals causes a large number of reflections to come from various directions. By its very nature this requires a close coupling between the channels and hence integrated multichannel processing.

It is often preferable to have uncorrelated reverb tail signals to each of the loudspeaker feeds. This can be done easily using standard stereo og mono reverbs with slightly diferent settings for each output channel. If the reverb tail output channels are not uncorrelated phasing problems will occur when the multichannel signal is mixed down to fewer channels. And this downmix will in most cases happen if the listener only listens to stereo or even mono. Only when a separate stereo mix is made to complement the multichannel mix the possible correlation between reverb channels becomes less important.

The topic of multichannel reverb is discussed further in another paper for this 19th AES Conference written by Knud Bank Christensen [5].

DYNAMICS PROCESSING

A rather late stage in the production of multichannel signals is dynamics processing of a multichannel stem or even of the final mix. In a similar fashion as for stereo the gain changes introduced in the individual channels must be coupled. If not, sound image shift is the result. There is no single best way of coupling the channels in the multichannel case.

Some material is front-back oriented, like movies, where it may be best to treat the left and right front channels as a pair, the surround channels as a pair - with stronger compression in order to render the surround channels audible also at lower listening levels -, and to treat also the center dialog channel separately. And finally, the low frequency effects channel also should be treated separately. This calls for four compressor side chains.

In material of ambient character, where channel allocation is more free than in a typical movie production, all five main channels may be treated equally, thus only needing one side chain as the LFE channel is typically not used in these cases.

Furthermore, a ducking function may be convenient, allowing the center dialog or a commentary channel to damp the others channels slightly when it is active.

A multichannel, multiband dynamics processing tool has been developed, see fig. 1.

The input consists of the five main channels, an auxillary channel (Xt) and the LFE channel. The output consists of the five main channels and the LFE. The main signal flow is as follows: The five main inputs channels and Xt are split into three frequency bands and an optional look-ahead delay is applied. Each band in each channel then passes through a gain cell before the three bands are combined again into a broadband signal. In order to catch large compressor overshoots in a gentle way a soft clipper may be applied. The last processing stage is a fast broadband limiter.

The gain control signals for each channel and band are generated by three multiband side chains. Compressor and expander functionalities are handled by these side chains. Each of the three multiband side chain inputs can be coupled to a combination of input channels. The combination can be switched between maximum, sum and off.

The gain of each output channel can be controlled by any one of the three side chain outputs.

This combination possibility is the main difference to a collection of single or dual channel processing units. Many compressors have external side chain inputs allowing for some coupling but not as easily and flexibly as in a dedicated multichannel processor.

Due to the small bandwidth of the LFE channel it is handled with a single band processing structure.

In addition to the signal processing itself comprehensive metering is essential in dynamics processing. As indicated in fig. 1 a quite large number of meters are sensing the levels and gain reductions applied. In the practical implementation these are shown simultaneausly on one graphics page [6, p. 41-43].

THE LOW FREQUENCY CHANNEL

The low frequency effects/enhancement (LFE) channel has been introduced together with the current digital multichannel consumer distribution formats. A primary motivation for it is headroom. The LFE channel is meant to carry only very low frequency contents, below about 120 Hz. Due to the properties of hearing the sound pressure levels needed at these low frequencies in order to achieve a certain perceived loudness can be quite high. And explosions, earthquakes etc. are supposed to be loud. So instead of increasing the headroom of the five main channels in order to make room for these loud low frequency effects a separate channel was introduced. The nominal playback gain of the LFE channel is about 10 dB higher than for the main channels [7], so this enables a higher sound pressure level for the LFE channel given the same dynamic range of the transmission channel or storage media. At playback the LFE signal will typically be fed to a subwoofer, possibly together with some low frequency content from the five main channels.

In cinema practice, however, a sixth channel also exists, but with the purpose of feeding the subwoofer directly. Also this channel has its dynamics range shifted upwards by 10 dB.

The partly conflicting use of the sixth channel can lead to confusion and to practical problems of how to manage this channel. There are a few typical cases:

- Extraction of a low frequency signal from the five main channels.
- Insertion of the low frequency signal into the five main channels.

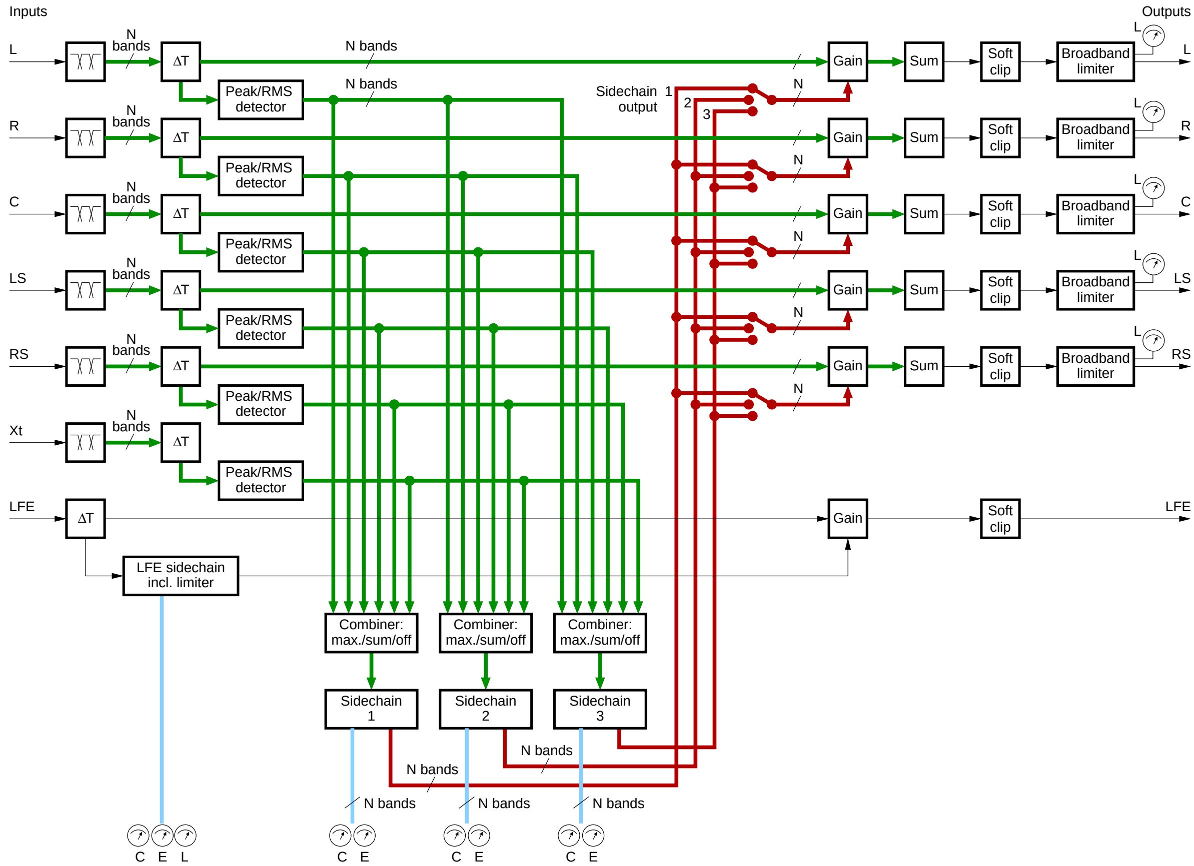

Fig. 1: Block diagram of a multichannel multiband dynamics processor. The number of bands, N, is 3.

The extraction may have the purpose of giving a starting point for an LFE channel given an already existing five channel mix. It could also be to generate a subwoofer channel for cinema use.

The insertion function could be used when changing between cinema and home format, where contents of the the cinema subwoofer signal should not be lost, as it would often be if directly used an an LFE signal. It should be remembered that the LFE channel is optional and that it dissapears in downmix matrixes [9], [10].

In both of these cases appropriate filters and mixing coefficients are needed. A configuration is shown in fig. 2. Relatively gentle second order as well as more steep fourth order filters may be selected.

Most consoles do not feature the crossover-like filter types needed here although they can easily handle the gain and summing (matrixing) functions. And external crossover filters lack the summing function. So this relatively simple function of managing a low frequency channel has been implemented as a processing block.

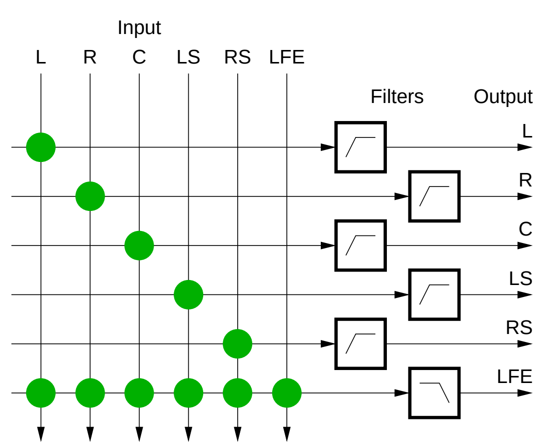

Fig. 2: Extraction of bass signal from the five main channels. The circles represent variable gain summing nodes.

The standards for low bitrate coding, e.g. [9], [11], are not particularly clear about the actual bandwidth of the LFE channel but there is a general agreement to use only up to 120 Hz. In a low bit rate coding systems the low frequency channel may be sampled at a much lower rate then the main channels. In MPEG-2 Layer II it is 500 Hz ($f_s/96$) [9, sec. 0.2.2.1] for a main rate of 48 kHz.

OUTPUT FORMAT CONVERSION

For monitoring purposes a downmix matrix is needed in order to ensure that the material is also satisfactory when played back on systems with fewer (or more) than the originally intended number of channels. At least two of the distribution formats currently in use (MPEG-2 [9] and Dolby AC-3 [10], [11]) have provisions for downmix either at the encoding or the decoding side. Most multichannel decoders or amplifiers have a downmix matrix available where the user specifies the loudspeaker channels in use.

The coefficients for the optional downmix are typically set at encoding time. To make matters more complicated, and to enable an aurally satisfying result, there are several possible values for these coefficients. In MPEG-2, for instance, the center channel is mixed with -3 dB gain into L and R, whereas the surround channels can be mixed into L and R respectively with -3 or -6 dB gain. Dolby AC-3 allows a bit more flexibility in these values [10, p.4-16].

A modern mixer with presets and the appropriate routing possibilities could be used for the downmix. But it may be more convenient to do it with a dedicated processing block.

One downmix function normally not avilable in a mixer is 90°phase shifted mono mix. When converting a 4:2:4 matrix surround signal (Dolby Surround/Pro Logic [12]) to mono this special downmix is needed in order to avoid that the surround channel information (coded as L/R anti-phase) gets lost during the downmix. In broadcast mono is still highly relevant. There are many mono TV sets and portable radios out in the field.

Whenever a downmix takes place there is a risk of overload, so some place after the downmix matrix a limiter should be used to avoid overload.

Some examples

Basically when doing a downmix the LFE channel is discarded as shown in the figures. If the LFE signal is to be included (e.g. for a 5.1 movie mix) a separate bass management tool as described earlier can be used.

An example of the use of the downmix matrix for conversion from 5.1 to stereo is shown in fig. 3.

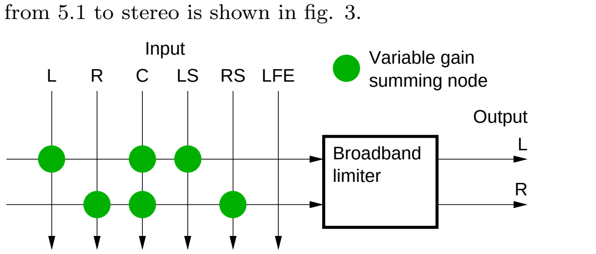

Fig. 3: Downmix matrix for 5.1 to stereo conversion.

For conversion between a full 5.1 signal and LCRS (4:2:4 matrix) the two surround channels are combined into one. In order to avoid problems with anti-phase signals in the surrounds a 90°mono circuit can be inserted, see fig. 4.

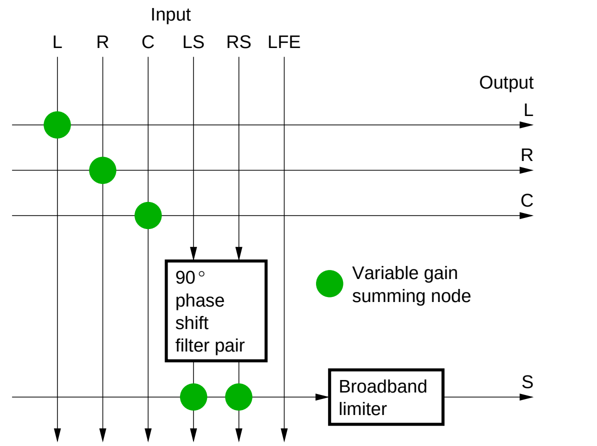

Fig. 4: Downmix matrix for 5.1 to LCRS conversion including a 90°mono filter pair.

TEST SIGNAL GENERATION

Such a simple tool as a test signal generator should not be missing in a multichannel production environment. Although confirmation of correct assignment of channels to tracks and loudspeakers may seem trivial it is nonetheless needed. Also, level calibration tones and noises are convenient to have.

For electrical calibration sine waves are quite convenient as they are stable in level. For acoustical calibration noise is more appropriate due to the many room resonances. It was chosen to implement sine waves as well as pink and white noise. Additionally, low- and highpass filtered pink noise was included to enable calibration of loudspeakers of limited bandwidth [13].

In the practical implementation the test signal generator has been integrated with low frequency handling, format conversion and output limitation.

PITCH CHANGE

When converting movies with 24 frames per second (FPS) to 25 FPS for European television or DVD production, the traditional solution has been to play the movie a bit too fast, 25/24 times the real speed. This simplifies the process considerably. Unfortunately, this also changes the duration of the movie, and the pitch of the sound track. The latter problem can be solved by using pitch changing tools, either in soft- or hardware.

Many pitch changing algorithms work in the time domain by dividing the signal into small blocks and stretching or compressing these according to the pitch change wanted. Stretching or compressing small blocks of signal and keeping the total duration the same means that some parts are repeated or discarded. Onset transients may therefore be shifted in time. Care must be taken that the blocks are synchronised between the channels. If not, an image shift may be caused by different delays in the individual channels.

In order to enable this synchronisation a new multichannel pitch change algorithm was devolped and implemented.

USABILITY

Having a collection of individual tools may be fine in itself but the practical aspects should not be neglected. Often sound is produced under a strong time pressure so the usability is important. A user interface which is easy to understand and use is crucial. The use of a separate remote control with a colour touch screen combined with motorised faders has proven to be successful. In a large facility with several mixing stages it may be advantageous to share processing equipment placed in a central machine room. By using a general interconnection (Ethernet) between processor and remote control it is possible to connect several processors and remote controls to the same network.

Furthermore, flexibility is nice in a field of evolving standards. Take the channel allocation on tape or disk, for example. A recommended track layout from SMPTE [7], ITU and others exists, but naturally not all material sticks to the recommendation. Therefore, the input and output routings must be flexible, see fig. 5. Also, in the practical world, both analog and digital interconnections exist side by side.

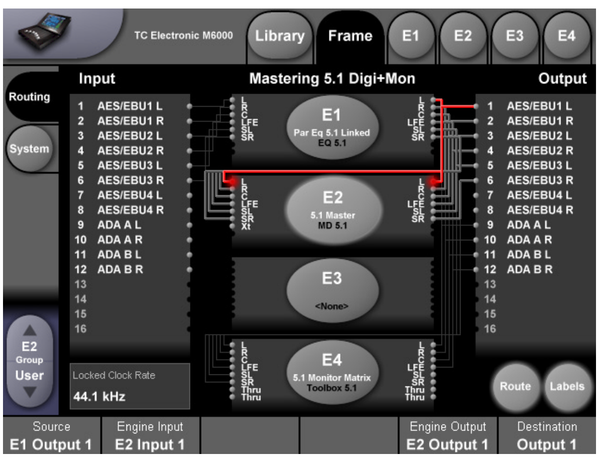

Fig. 5: Routing of inputs, outputs and processing blocks in the practical implementation [15].

Standards

It has been a design goal that the new multichannel toolbox should use and respect standards whenever possible. Standards may seem limiting and difficult to implement at first but in the long run the advantages are stronger.

The allocation of input and output channels has been mentioned already. So has the use of an industry-standard interconnect (Ethernet) for remote control.

The frequencies of various filters throughout have been selected to be in accordance with [14] wherever possible. In that way a series of numbers easy to remember is achieved, with 80 values per decade - or approximately 1/24 octave. This was chosen instead of the simple method of dividing the frequency range in 128 logarithmically spaced values in order to get a simple relationship to MIDI controller values.

Sometimes companies develop methods and products which they would like to give a status like a de-facto standard, but nevertheless wrap into licensing conditions which prevent an open public discussion of technical details. Fortunately, broadcasting authorities, international standard bodies and similar organisations do not like closed standards. A license fee may apply also on international standards, but everybody is allowed to use and discuss the standards such as the AES3 interface, or the TCP/IP data communication protocol. Even complicated low bitrate coding standards like MPEG [8], [9] can be openly discussed.

PROCESSING PLATFORM

The tools described here are basically independent on the processing platform. Whether implemented in software or on a dedicated hardware platform the same functionality is needed and in principle possible. When it comes to practice, however, a hardware solution is often preferable for a variety of reasons:

- Reliability. Auditory "features" such as sample slips do not belong to a professional environment. Neither do sudden reboots.
- Ease of operation. A user interface adapted to the task at hand and with physical interaction allows a quicker and safer way of operation than the one-finger analogy in typical software interfaces.
- Remote control. Although modern computer hardware is perfectly capable of being remotely controlled the operating systems are typically not well suited for this.
- Protection of development effort, especially the software part. To have many users is nice but to have many customers is nicer.

It is no small task to develop a multichannel processing platform. The processing power needed is beyond the capability of one DSP chip so a number of these must be used together. The needed coupling between channels and processing blocks calls for high bandwidth connections between the DSP chips.

Many seemingly trivial tasks such as routing and metering take their share of the processing power and communication bandwidth.

Our present hardware DSP platform consists of four DSP chips for doing the signal processing and a fifth DSP chip for routing and metering. These are controlled by a RISC processor for coefficient calculation.

The DSP module is placed in a mainframe with Ethernet connection to a remote control unit. Modules for analog and digital I/O are placed in the same mainframe.

The remote control unit features a touch-sensitive colour LCD screen and motorised faders. By using a touch-sensitive screen the use of dedicated hardware pushbuttons is avoided enabling a more flexible and compact design.

For sound source placement the touch screen may be inconvenient, so one or more joystics can be connected via MIDI.

CONCLUSIONS

A number of tools for production of multichannel sound has been presented. Due to the coupling needed between the channels they have been designed as genuine multichannel tools with both the flexibility and ease of handling needed. Some differences to single and dual channel processing have been pointed out. The tools have been implemented in a general signal processing hardware platform at present.

REFERENCES

[1] Society of Motion Picture and Television Engineers (1994): SMPTE Engineering Guideline 18 - Design of Effective Cine Theatres. www.smpte.org

[2] International Telecommunication Union (1994): Recommendation ITU-R BS.775-1: Multichannel Stereophonic Sound System with and without Accompanying Picture, First edition. www.itu.int

[3] Knud Bank Christensen and Thomas Lund (1999): Room Simulation for Multichannel Film and Music, presented at the 107th AES Convention, New York, preprint 4993. www.aes.org

[4] Thomas Lund (2000): Enhanced Localization in 5.1 Production, presented at the 109th AES Convention, Los Angeles, preprint 5243. www.aes.org

[5] Knud Bank Christensen (2001): Reverb and Room Simulation in the Multichannel Era, presented at the AES 19th International Conference, Schloss Elmau. www.aes.org

[6] T.C. Electronic (2001): System 6000 Users Manual, Algorithms, Rev. 1.30. www.tcelectronic.com

[7] Society of Motion Picture and Television Engineers (1999): SMPTE Standard 320M-1999, Television - Channel Assignments and Levels on Multichannel Audio Media allocation. www.smpte.org

[8] International Standard ISO/IEC 11172-3 (1993): Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s, Part 3: Audio. www.iso.ch

[9] International Standard ISO/IEC 13818-3 (1994): Information Technology - Generic Coding of Moving Pictures and Associated Audio: Audio. www.iso.ch

[10] Dolby Laboratories (1998): Dolby Digital Professional Encoding Manual, Issue 01, Part No. 91535, www.dolby.com

[11] Advanced Television Systems Committee (1995): ATSC Doc. A/52, Digital Audio Compression Standard (AC-3). www.atsc.org

[12] Dressler, R. (1997): Dolby Pro Logic Surround Decoder Principles of Operation, Dolby, Inc. www.dolby.com

[13] Surround Sound Forum (1999): Multichannel Recording in 3/2 Format - Parameters for Programme Interchange and Archiving, Alignment of Reproduction Equipment. Recommended Practice SSF-02. www.tonmeister.de

[14] International Standard IEC 1260 (1995): Electroacoustics - Octave-band and fractional-octave-band filters, First edition. www.iec.ch

[15] T.C. Electronic (2001): System 6000 Users Manual, Operation, Rev. 1.30. www.tcelectronic.com