MPEG Surround—The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding*

JÜRGEN HERRE,[1] *AES Fellow,* **KRISTOFER KJÖRLING,**[2]
(hrr@iis.fraunhofer.de) (Kristofer.Kjorling@dolby.com)

JEROEN BREEBAART,[3] *AES Member,* **CHRISTOF FALLER,**[4] *AES Member,* **SASCHA DISCH,**[5]
(jeroen.breebaart@philips.com) (christof.faller@illusonic.com) (disch@tnt.uni-hannover.de)

HEIKO PURNHAGEN,[2] *AES Member,* **JEROEN KOPPENS,**[6] **JOHANNES HILPERT,**[1] **JONAS RÖDÉN,**[2]
(Heiko.Purnhagen@dolby.com) (jeroen.koppens@philips.com) (hlp@iis.fraunhofer.de) (Jonas.Roden@dolby.com)

WERNER OOMEN,[6] *AES Member,* **KARSTEN LINZMEIER,**[7] **AND KOK SENG CHONG**[8]
(werner.oomen@philips.com) (Karsten.Linzmeier@dolby.com) (KokSeng.Chong@sg.panasonic.com)

[1] *Fraunhofer Institute for Integrated Circuits, 91058 Erlangen, Germany*
[2] *Coding Technologies, 1113 30 Stockholm, Sweden; now part of Dolby Laboratories, Stockholm, Sweden*
[3] *Philips Research, 5656 AE Eindhoven, The Netherlands*
[4] *Agere Systems, Allentown, PA 18109, USA*
[5] *Fraunhofer Institute for Integrated Circuits, 91058 Erlangen, Germany; now with Laboratorium für Informationstechnologie, Hannover, Germany*
[6] *Philips Applied Technologies, Eindhoven, The Netherlands*
[7] *Fraunhofer Institute for Integrated Circuits, 91058 Erlangen, Germany; now with Dolby Laboratories, Nürnberg, Germany*
[8] *Panasonic Singapore Laboratories Pte. Ltd., Singapore*

In 2004 the ISO/MPEG Audio standardization group started a new work item on efficient and backward-compatible coding of high-quality multichannel sound using parametric coding techniques. Finalized in the fall of 2006, the resulting MPEG Surround specification allows the transmission of surround sound at bit rates that have been commonly used for coding of mono or stereo sound. The results of the standardization process are summarized by describing the underlying ideas and providing an overview of the MPEG Surround technology. The performance of the scheme is characterized by the results of recent verification tests. These tests include several operation modes as they would be used in typical application scenarios to introduce multichannel audio into existing audio services.

0 INTRODUCTION

In 2004 the ISO/MPEG Audio standardization group started a new work item on efficient and backward-compatible coding of high-quality multichannel sound using parametric coding techniques. Specifically, the technology to be developed should be based on the spatial audio coding (SAC) approach, which extends traditional approaches for coding of two or more channels in a way that provides several significant advantages, in terms of both compression efficiency and features. First it allows the transmission of multichannel audio at bit rates that so far only allowed for the transmission of monophonic audio. Second, by its underlying structure, the multichannel audio signal is transmitted in a backward-compatible way. As such the technology can be used to upgrade existing distribution infrastructures for stereo or mono audio content (radio channels, Internet streaming, music downloads, and so on) toward the delivery of multichannel audio while retaining full compatibility with existing receivers. After an intense development process, the resulting MPEG Surround specification was finalized in the second half of 2006 [1], [2].

*Presented at the 122nd Convention of the Audio Engineering Society, Vienna, Austria, 2007 May 5–8; revised 2008 September 26.

This paper summarizes the results of the standardization process by describing the underlying ideas and providing an overview of the MPEG Surround technology. Special attention is given to the results of the recent MPEG Surround verification tests that assess the technology's performance in several operation modes as they would be used in typical application scenarios introducing multichannel audio into existing audio services.

Sections 1 and 2 illustrate the basic approach and the MPEG standardization process that was executed to develop the MPEG Surround specification. The core of the MPEG Surround architecture and its further extensions are described in Sections 3 and 4. Finally system performance and applications are discussed.

1 SPATIAL AUDIO CODING BASICS

In a nutshell the general underlying concept of SAC can be outlined as follows. Rather than performing a discrete coding of the individual audio input channels, a system based on SAC captures the spatial image of a multichannel audio signal into a compact set of parameters that can be used to synthesize a high-quality multichannel representation from a transmitted downmix signal. Fig. 1 illustrates this concept. During the encoding process the spatial parameters (cues) are extracted from the multichannel input signal. These parameters typically include level/intensity differences and measures of correlation/coherence between the audio channels and can be represented in an extremely compact way. At the same time a monophonic or two-channel downmix signal of the sound material is created and transmitted to the decoder together with the spatial cue information. The downmix can be conveyed to the receiver using known audio coders for monophonic or stereophonic signals. On the decoding side the transmitted downmix signal is expanded into a high-quality multichannel output based on the spatial parameters.

Due to the reduced number of audio channels to be transmitted (such as just one channel for a monophonic downmix signal), the SAC approach provides an extremely efficient representation of multichannel audio signals. Furthermore it is backward compatible on the level of the downmix signal. A receiver device without a spatial audio decoder will simply present the downmix signal.

Conceptually this approach can be seen as an enhancement of several known techniques, such as an advanced method for joint stereo coding of multichannel signals [3], a generalization of parametric stereo [4], [5] to multichannel application, and an extension of the binaural cue coding (BCC) scheme [6], [7] toward using more than one transmitted downmix channel [8]. From a different viewing angle, the SAC approach may also be considered an extension of well-known matrix-surround schemes (Dolby Surround/Prologic,[9] Logic 7,[10] Circle Surround,[11] and so on) [9], [10] by transmission of dedicated (spatial cue) side information to guide the multichannel reconstruction process and thus achieve improved subjective audio quality [11].

Due to the combination of bit-rate efficiency and backward compatibility, SAC technology can be used to enhance a large number of existing mono or stereo services from stereophonic (or monophonic) to multichannel transmission in a compatible fashion. To this aim the existing audio transmission channel carries the downmix signal, and the spatial parameter information is conveyed in a side chain (such as the ancillary data portion of an audio bit stream). In this way multichannel capability can be achieved for existing audio distribution services for a minimal increase in bit rate, such as between 3 and 32 kbit/s.

2 MPEG SURROUND DEVELOPMENT PROCESS

In March 2004 the ISO/MPEG standardization group started a new work item on SAC by issuing a call for proposals (CfP) on SAC [12]. Four submissions were received in response to this CfP and evaluated with respect to a number of performance aspects, including the subjective quality of the decoded multichannel audio signal, the subjective quality of the downmix signals generated, the spatial parameter bit rate, and other parameters (additional functionality, computational complexity, and so on).

As a result of these extensive evaluations, MPEG decided that the basis for the subsequent standardization process, called reference model 0 (RM0), would be a system combining the submissions of Fraunhofer IIS/Agere Systems and Coding Technologies/Philips. These systems

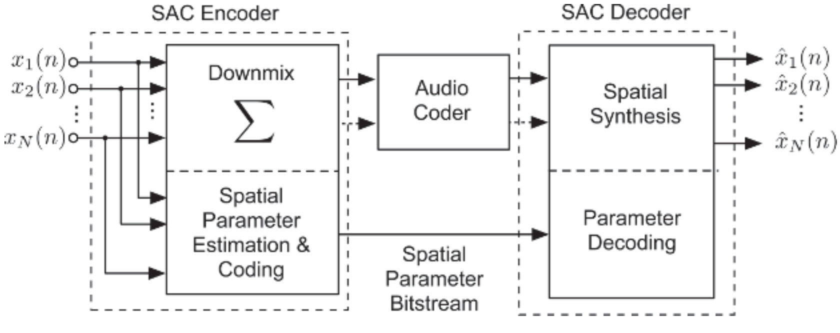

Fig. 1. Principle of spatial audio coding.

[9] Dolby, Pro Logic, and Dolby Surround are registered trademarks of Dolby Laboratories Licensing Corp.

[10] Logic 7 is a registered trademark of Harman International Industries.

[11] Circle Surround is a registered trademark of SRS Labs.

outperformed the other submissions and, at the same time, showed complementary performance in terms of other parameters (for example, per-item quality, bit rate) [13]. The merged RM0 technology (now called MPEG Surround) combines the best features of both individual submissions and was found to fully meet (and even surpass) the performance expectation [14], [15]. The successful development of RM0 set the stage for the subsequent improvement process of this technology, which was carried out collaboratively within the MPEG Audio group. After a period of active technological development, the MPEG Surround specification was frozen in the second half of 2006 and its performance confirmed by the final verification test report in January 2007.

3 BASIC CONCEPTS OF MPEG SURROUND

While a detailed description of the MPEG Surround technology is beyond the scope of this paper, this section provides a brief overview of the most salient underlying concepts. An extended description of the technology can be found in [14], [16]–[18]. Refinements of this basic framework will be described in the subsequent section.

3.1 Filter-Bank and Top-Level Structure

In the human auditory system the processing of binaural cues is performed on a nonuniform frequency scale [19], [20]. Hence in order to estimate spatial parameters from a given input signal, it is important to transform its time-domain representation to a representation that resembles this nonuniform scale by using an appropriate filter bank.

For applications including low-bit-rate audio coding, the MPEG Surround decoder is typically applied as a post-processor to a low-bit-rate (mono or stereo) decoder. In order to minimize computational complexity, it would be beneficial if the MPEG Surround system could directly make use of the spectral representation of the audio material provided by the audio decoder. In practice, however, spectral representations for the purpose of audio coding are typically obtained by means of critically sampled filter banks [for example, using a modified discrete cosine transform (MDCT) [21] and are not suitable for signal manipulation as this would interfere with the aliasing cancellation properties associated with critically sampled filter banks. The spectral band replication (SBR) algorithm [22] is an important exception in this respect. Similar to the SAC/MPEG Surround approach, the SBR algorithm is a post-processing algorithm that works on top of a conventional (band-limited) low-bit-rate audio decoder and allows the reconstruction of a full-bandwidth audio signal. It employs a complex-modulated quadrature mirror filter (QMF) bank to obtain a uniformly distributed, oversampled frequency representation of the audio signal. The MPEG Surround technology takes advantage of this QMF filter bank, which is used as part of a hybrid structure to obtain an efficient nonuniform frequency resolution [5], [23]. Furthermore, by grouping filter-bank outputs for spatial parameter analysis and synthesis, the frequency resolution for spatial parameters can be varied extensively while applying a single filter-bank configuration. More specifically, the number of parameters to cover the full frequency range (number of parameter bands) can be varied from only a few (for low-bit-rate applications) up to 28 (for high-quality processing) to closely mimic the frequency resolution of the human auditory system. A detailed description of the hybrid filter-bank in the context of MPEG Surround can be found in [14].

The top-level structure of the MPEG Surround decoder is illustrated in Fig. 2, showing a three-step process that converts the supplied downmix into the multichannel output signal. First the input signal is decomposed into frequency bands by means of a hybrid QMF analysis filter bank. Next the multichannel output signal is generated by means of the spatial synthesis process, which is controlled by the spatial parameters conveyed to the decoder. This synthesis is carried out on the subband signals obtained from the hybrid filter bank in order to apply the time and frequency-dependent spatial parameters to the corresponding time–frequency region (or "tile") of the signal. Finally the output subband signals are converted back to the time domain by means of a set of hybrid QMF synthesis filter banks.

3.2 Structure of Spatial Synthesis

MPEG Surround provides great flexibility in terms of the input, downmix, and decoder channel configurations. This flexibility is obtained by using relatively simple conceptual elements, which can be grouped to build more complex coder structures. The two most important elements are referred to as one-to-two (OTT) and two-to-three (TTT) elements. The numbers refer to the input and output channel configurations of each element at the de-

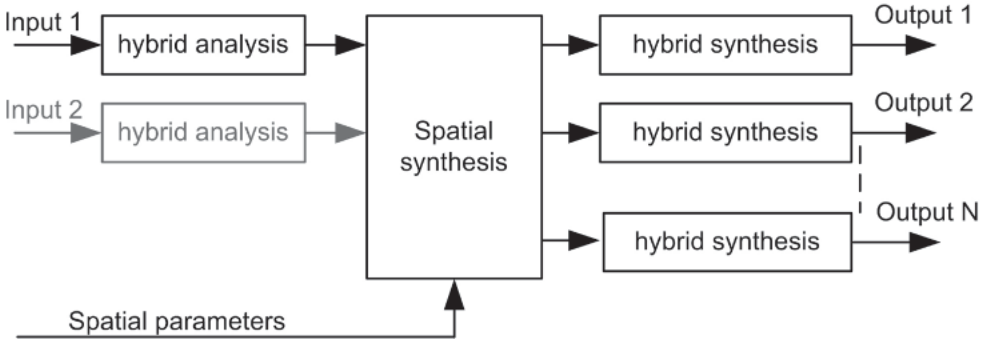

Fig. 2. High-level overview of MPEG Surround synthesis.

coder side. In other words, an OTT element describes two output channels by means of a single input channel, accompanied by spatial parameters. Similarly, the TTT element characterizes three output channels by means of a stereo input signal and parameters. Each conceptual decoder element has a corresponding encoder element that extracts spatial parameters and generates a downmix from its input signals. Using the two building blocks many encoder and decoder configurations can be built that encode a multichannel signal into a downmix and parameters, or conversely decode a downmix into a multichannel signal based on spatial parameters. An example decoder is outlined in Fig. 9, illustrating the basic idea of connecting the building blocks into an MPEG Surround decoder. The various encoding and decoding elements are described in more detail subsequently.

3.2.1 OTT Elements

At the encoder side the OTT encoder element, also referred to as reverse-one-to-two (R-OTT) module, extracts two types of spatial parameters and creates a downmix (and a residual) signal. The OTT encoder element is virtually identical to a parametric stereo coder [5], [23] and is based on similar principles as binaural cue coding (BCC, [6], [7]). The following spatial parameters are extracted for each parameter band:

- *Channel level difference (CLD)* This is the level difference between the two input channels. Nonuniform quantization on a logarithmic scale is applied to the CLD parameters, where the quantization has a high accuracy close to 0 dB and a coarser resolution when there is a large difference in level between the input channels (as is in line with psychoacoustics).
- *Interchannel coherence/cross correlation (ICC)* It represents the coherence or cross correlation between the two input channels. A nonuniform quantization is applied to the ICC parameters.

The residual signal represents the error associated with representing the two signals by their downmix and associated parameters and, in principle, enables full multichannel waveform reconstruction at the decoder side (see the section on residual coding).

At the decoder side the OTT element recreates two channels from the single downmix channel and the spatial parameters. This is visualized in Fig. 3. The OTT module takes the single input signal and creates a decorrelated version of the same by means of a decorrelator D. These two signals are mixed together based on the CLD parameter and the ICC parameter. The CLD parameter controls the energy distribution of the input signal between the two output signals, and the ICC parameter controls the amount of decorrelated signal mixed into the two output signals. If a residual signal is present, the decorrelated signal is replaced by this residual.

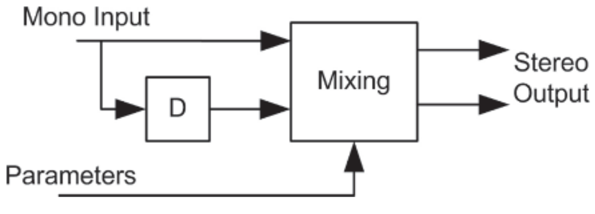

Fig. 3. Basic principle of OTT module.

3.2.2 TTT Elements

The TTT encoder element (R-TTT) generates a stereo downmix signal from three input channels, accompanied by spatial parameters. More specifically, the stereo downmix l_0, r_0 is a linear combination of the three input signals *l, c,* and *r,* in which the center input signal *c* is represented as phantom center in the stereo downmix. To provide three-channel reconstruction at the decoder side, two CPCs are transmitted. An additional ICC-coded parameter provides compensation for the prediction loss at the decoder side based on statistical properties rather than on waveform reconstruction principles. Similar to the OTT element, the TTT encoder element also provides a residual signal that may be transmitted to enable full waveform reconstruction at the decoder side. A detailed description of the TTT element and its various operation modes is provided in [24].

Conversely the TTT decoder module recreates three channels based on the two downmix channels and the corresponding parameters. The ICC codec parameter quantifying the prediction loss can be used to compensate for the prediction loss either by means of gain adjustment of the output signal, or by means of adding a decorrelated signal, or by means of a residual signal, corresponding to the prediction loss.

3.2.3 Tree-Structured Parameterization

Since OTT and TTT elements can be combined in a tree-structured manner to build more complex coder structures, many different channel configurations can be supported in a flexible way. The following two example configurations describe the encoding of 5.1 surround sound into a downmix and its decoding back to 5.1 multichannel for a mono and a stereo downmix, respectively.

Fig. 4 shows how to combine several R-OTT modules into a multichannel encoder that encodes a multichannel into a mono downmix and the corresponding parameters. For every R-OTT module CLD and ICC parameters are derived as well as a possible residual signal. (The residual may be omitted if the lowest bit rate overhead possible is desired.) The signals L, Ls, C, LFE, R, and Rs denote the left front, left surround, center, low-frequency effects, right front, and right surround channels, respectively.

In Fig. 5 the parameterization corresponding to the tree structure of Fig. 4 is visualized. The tree-structured parameterization combines the channels into larger groups of channels, and for every such combination of groups it derives parameters describing how to separate the groups given the parameters and the combined group. Hence the CLD and ICC parameters with the subscript 2 describe how to separate the Ls and Rs channels from each other given the combination of the two and the corresponding parameters. Similarly, the CLD and ICC parameters with the subscript 0 describe how to separate a combination of the surround channel (Ls and Rs) from a combination of all the front channels (C, LFS, L, and R) given the combination of the two groups and the corresponding parameters.

Among the many conceivable configurations of MPEG Surround, the encoding of 5.1 surround sound into two-channel stereo is particularly attractive in view of its backward compatibility with existing stereo consumer devices. Fig. 6 shows a block diagram of an encoder for such a typical system consisting of one R-TTT and three R-OTT encoder elements, and Fig. 7 visualizes the parameterization for the underlying tree structure.

Similarly, the OTT and TTT elements can be used to build up a multitude of different encoder/decoder structures, supporting arbitrary downmixing/upmixing configurations such as 7.1–5.1–7.1 (that is, 7.1 input channels downmixed to 5.1 downmix channels and subsequently upmixed again to 7.1 output channels).

On the decoder side the parameterization of the multichannel signal can be used to visualize a conceptual upmix of the downmix signal based on the spatial parameters. (The next subsection will clarify how the actual signal flow in the decoder differs from the conceptual upmix outlined in the following.) Fig. 8 visualizes a conceptual mono-to-5.1 decoder using OTT modules. The subscripts of the OTT modules and the parameters match those of the corresponding encoder illustration (Fig. 4) and parameterization visualization (Fig. 5). Similarly, Fig. 9 displays a conceptual decoder counterpart to the encoder visualized in Fig. 6.

All OTT/TTT analysis and synthesis operations are carried out on hybrid QMF filter-bank spectral values without

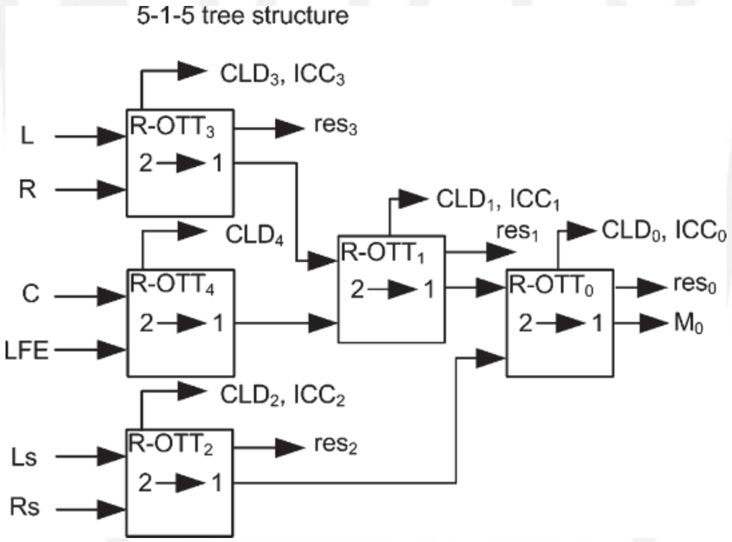

Fig. 4. OTT tree forming a 5.1-to-mono encoder.

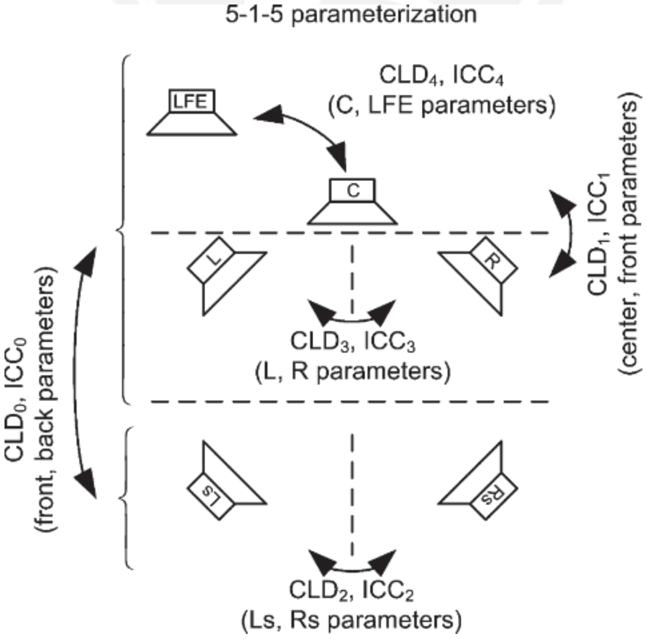

Fig. 5. Multichannel parametirization for a mono-to-5.1 MPEG Surround signal.

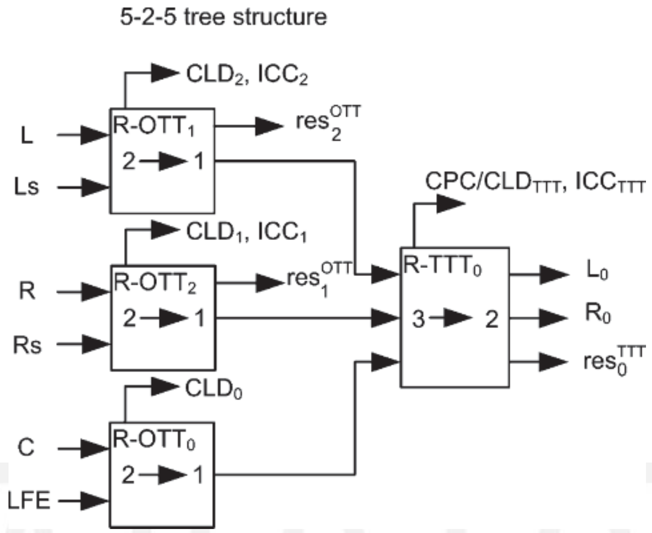

Fig. 6. OTT/TTT tree forming a 5.1-to-stereo encoder.

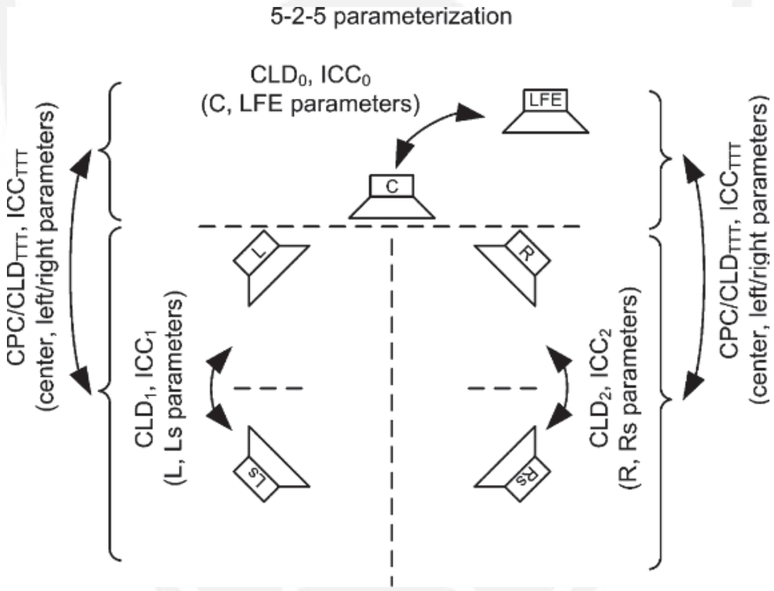

Fig. 7. Multichannel parameterization for a stereo-to-5.1 MPEG Surround signal.

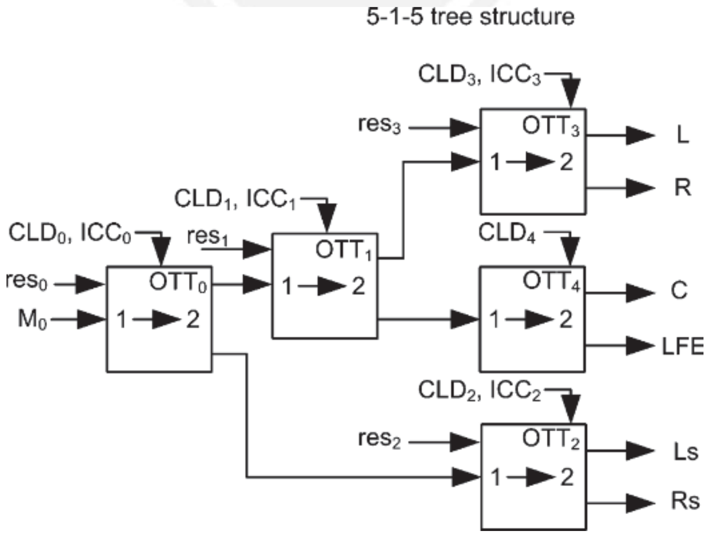

Fig. 8. Conceptual mono-to-5.1 decoder.

further algorithmic delay. Thus cascading of analysis and synthesis modules in a treelike fashion does not create additional delay in the encoder or decoder.

3.2.4 Flat-Structured MPEG Surround Decoder Signal Processing

Compared to the conceptual views presented previously the actual upmix to the multichannel output does not take place in a tree-structured fashion in an MPEG Surround decoder. Instead the decoder signal processing happens in a "flattened" way, that is, the treelike parameterization visualized by a tree-structured upmix process comprising several subsequent stages is transformed into a single-stage operation. This achieves increased computational efficiency and minimizes possible degradations due to multiple decorrelation operations, that is, sending the output from one OTT module comprising a decorrelated signal component into another OTT module, deriving a new decorrelated signal component.

As a result the spatial synthesis process, as shown in the previous tree-structured decoder visualizations, can be described by two matrices and decorrelators, as illustrated in Fig. 10. These matrices map a lower number of input channels to a higher number of output channels, that is, if the input signal is a mono downmix, the input signal multiplied by the premix matrix is a scalar, and if the downmix input is a stereo signal, the input is a vector of two elements. Similarly the output is a vector containing one element for each output channel. The processing takes place in the subband domain of a filter bank, and hence the matrix operations are carried out for every sample in every subband.

The matrix elements are derived from the transmitted spatial parameters given the specific tree structure used for the parameterization. The decorrelators correspond to the decorrelators as part of the OTT and TTT modules.

The input signals are first processed by a premix matrix in order to ensure that the inputs to the different decorrelators are of the same level as if the upmix were carried out in a tree-structured fashion. The postmix matrix mixes the decorrelated signals with input signals similarly to the OTT modules, albeit for all OTT modules in a single step.

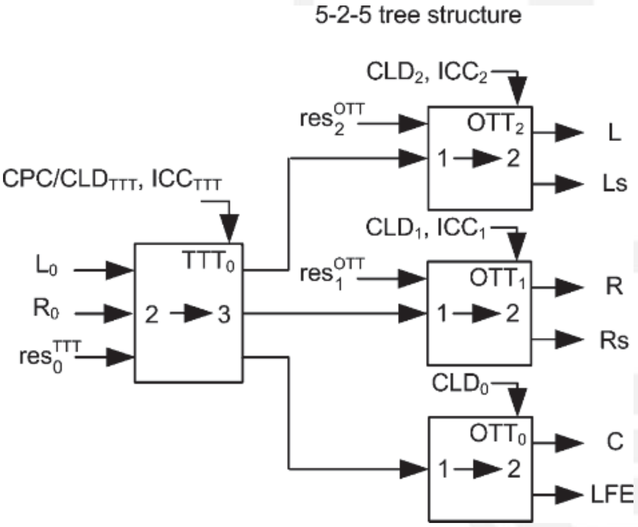

Fig. 9. Conceptual stereo-to-5.1 decoder.

3.3 Decorrelation

The spatial synthesis stage of the MPEG Surround decoder consists of matrixing and decorrelation units. The decorrelation units are required to synthesize output signals with a variable degree of correlation between each other (as dictated by the transmitted ICC parameters) by a weighted summation of original signal and decorrelator output [25]. Each decorrelation unit generates an output signal from an input signal according to the following properties:

- The coherence between input and output signals is sufficiently close to zero. In this context, coherence is specified as the maximum of the normalized cross-correlation function operating on band-pass signals (with bandwidths sufficiently close to those estimated from the human hearing system).
- Both the spectral and the temporal envelopes of the output signal are close to those of the incoming signal.

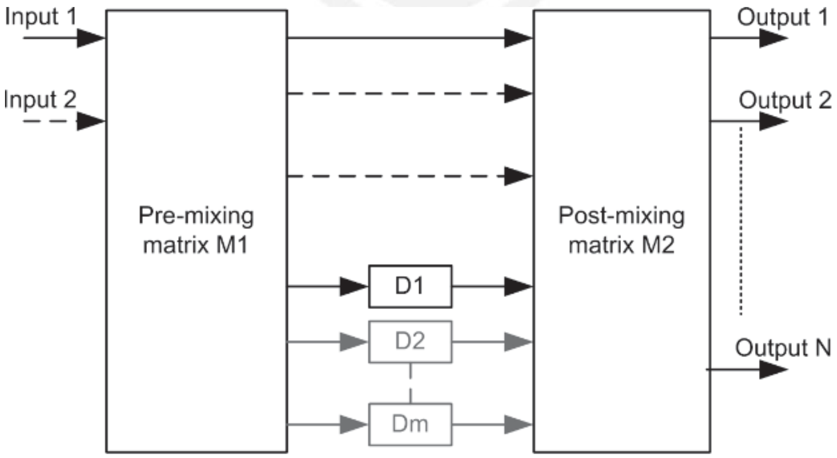

Fig. 10. Generalized structure of spatial synthesis process, comprising two mixing matrices (M_1, M_2) and a set of decorrelators, (D_1, D_2, . . . , D_m).

- The outputs of all decorrelators are mutually incoherent according to the same constraints as for their input/output relation.

The decorrelator units are implemented by means of lattice all-pass filters operating in the QMF domain, in combination with spectral and temporal enhancement tools. More information on QMF-domain decorrelators can be found in [25], [14], and a brief description of the enhancement by means of temporal envelope shaping tools is given subsequently.

3.4 Rate/Distortion Scalability

In order to make MPEG Surround usable in as many applications as possible, it covers a broad range of operation points in terms of both side information rate and multichannel audio quality. Naturally there is a tradeoff between a very sparse parametric description of the signal's spatial properties and the desire for the highest possible sound quality. This is where different applications exhibit different requirements and thus have their individual optimal "operating points." For example, in the context of multichannel audio broadcasting with a compressed audio data rate of approximately 192 kbit/s, emphasis may be given on achieving very high subjective multichannel quality, and spending up to 32 kbit/s of spatial cue side information is feasible. Conversely an Internet streaming application with a total available rate of 48 kbit/s including spatial side information (using, for example, MPEG-4 HE-AAC) will call for a very low side information rate in order to achieve the best possible overall quality.

In order to provide highest flexibility and cover all conceivable application areas, the MPEG Surround technology was equipped with a number of provisions for rate/distortion scalability. This approach permits to flexibly select the operating point for the tradeoff between side information rate and multichannel audio quality without any change in its generic structure. This concept is illustrated in Fig. 11 and relies on several dimensions of scalability that are discussed briefly in the following.

Several important dimensions of scalability originate from the capability of sending spatial parameters at different granularity and resolution:

- *Parameter frequency resolution* One degree of freedom results from scaling the frequency resolution of spatial audio processing. While a high number of frequency bands ensures optimum separation between sound events occupying adjacent frequency ranges, it also leads to a higher side information rate. Conversely, reducing the number of frequency bands saves on spatial overhead and may still provide good quality for most types of audio signals. Currently the MPEG Surround syntax covers a flexible parameter band resolution from 28 bands down to a single band.
- *Parameter time resolution* Another degree of freedom is available in the temporal resolution of the spatial parameters, that is, the parameter update rate. The MPEG Surround syntax covers a large range of update rates and also allows to adapt the temporal grid dynamically to the signal structure.
- *Parameter quantization resolution* As a third possibility, different resolutions for transmitted parameters can be used. Choosing a coarser parameter representation naturally saves in spatial overhead at the expense of losing some detail in the spatial description. Using low-resolution parameter descriptions is accommodated by dedicated tools, such as the adaptive parameter smoothing mechanism (as outlined in Section 4.6.4).
- *Parameter choice* Finally there is a choice as to how extensive the transmitted parameterization describes the original multichannel signal. As an example, the number of ICC values transmitted to characterize the wideness of the spatial image may be as low as a single value per parameter frequency band.

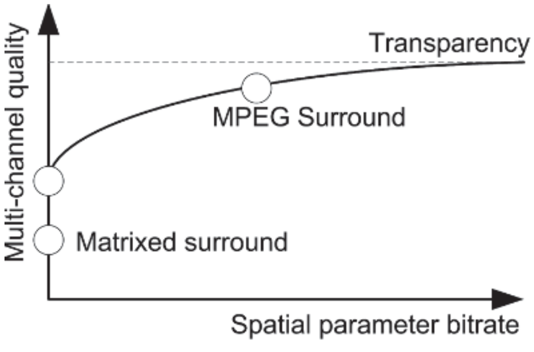

Fig. 11. Rate/distortion scalability.

Furthermore, in order not to be limited in audio quality by the parametric model used to describe the multichannel signal, a residual coding element is available that enables the MPEG Surround system to offer quality at the level of discrete multichannel coding algorithms. The residual coding tool is outlined in Section 4.5.

Together these scaling dimensions enable the operation at a wide range of rate/distortion tradeoffs from side information rates of below 3 kbit/s to 32 kbit/s and above. In addition MPEG Surround also supports a matrix-surround mode called Enhanced Matrix Mode, which will be described in more detail in Section 4.2.

3.5 Low-Power Processing

The MPEG Surround decoder can be implemented in a high-quality (HQ) version and a low-power (LP) version. The LP version is realized by simplifying the most computationally intensive modules of the HQ version, namely, the QMF filter banks and the decorrelators. Both versions operate on the same data streams, but the LP version consumes considerably less computational power.

The HQ version employs complex-valued QMF analysis and synthesis filter banks to perform time/frequency transforms. The LP version halves the complexity by using real-valued QMF filter banks. Since the real-valued QMF filter banks are critically sampled, real-valued hybrid data at parameter band borders are susceptible to aliasing when they are independently modified by spatial parameters of large difference. The aliasing problem is especially pronounced if it occurs in the low-frequency portion of signal spectrum, or if the signal at a parameter band border has a tonal characteristic.

To alleviate the aliasing effect, the low-frequency portion of the QMF data is converted to complex values before spatial synthesis, and converted back to real values before QMF synthesis. This is achieved by connecting a “real-to-complex” converter to the low-frequency output from the QMF analysis filter bank, and a “complex-to-real” converter to the low-frequency input to the QMF synthesis filter bank. These new partially complex filter banks suppress aliasing in the low-frequency spectrum significantly. This is displayed in Fig. 12, where the real-valued QMF analysis is followed by a real-to-complex converter for the lower subbands, and a delay for the higher subbands. The real-to-complex converter creates a complex-valued signal from the real-valued signal by means of filtering, in this way introducing a delay. Hence the higher subbands not subdued to the real-to-complex processing need to be delayed as well in order to maintain synchronization between the low-frequency and high-frequency parts of the signal. A similar procedure is done prior to the real-valued QMF synthesis.

For the high-frequency portion of QMF data the tonality of signals at parameter band borders is estimated. If the signal is found to be potentially tonal at a parameter band border, the spatial parameters of the adjacent bands are replaced by the average of both. This processing for the real-valued QMF subbands is more or less identical to that used in the low-power version of SBR in high-efficiency AAC [26].

While the HQ version employs high-quality lattice IIR decorrelators (optionally including fractional delays) to achieve perceptual separation among the upmix signals, the LP version employs a mixture of the real-valued version of the decorrelators described and the low-complexity decorrelators used in parametric stereo [25]. For certain decoder configurations where the downmix signal is mono, a simpler decorrelator structure is used, where the pre- and postmatrixing operations are merged to facilitate the reuse of some decorrelated signals in the spatial synthesis process.

4 ADDITIONAL SYSTEM FEATURES

4.1 Matrix Surround Compatibility

Besides a mono or conventional stereo downmix, the MPEG Surround encoder is also capable of generating a matrix-surround (MTX) compatible stereo downmix signal. This feature ensures backward-compatible 5.1 audio playback on decoders that can only decode the stereo core bit stream (that is, without the ability to interpret the spatial side information) but are equipped with a matrix-surround decoder. Moreover, this feature also enables the so-called enhanced matrix MPEG Surround mode (that is, a mode without transmission of spatial parameters as side information), which is discussed further in the next subsection. Special care was taken to ensure that the perceptual quality of the parameter-based multichannel reconstruction is not affected by this matrix-surround feature.

The matrix-surround capability is achieved by using a parameter-controlled postprocessing unit that acts on the stereo downmix at the encoder side. A block diagram of an MPEG Surround encoder with this extension is shown in Fig. 13.

The MTX-enabling postprocessing unit operates in the QMF domain on the output of the downmix synthesis block (that is, working on the signals L_{out} and R_{out}) and is controlled by the encoded spatial parameters. Special care is taken to ensure that the inverse of the postprocessing matrix exists and can be uniquely determined from the spatial parameters. Finally the matrix-surround compatible

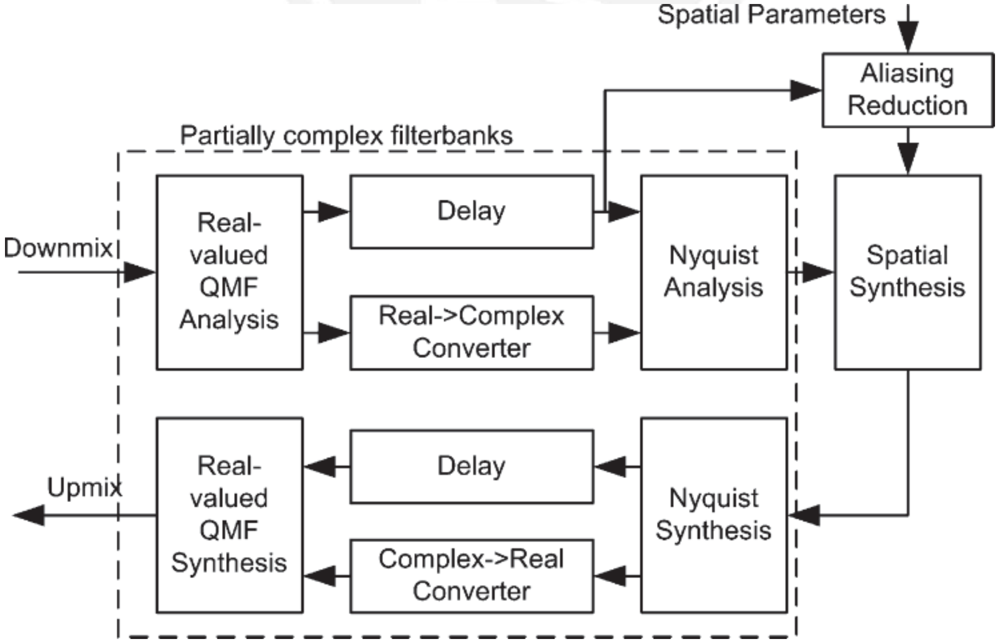

Fig. 12. Low-power MPEG Surround.

downmix (L_{MTX}, R_{MTX}) is converted to the time domain using QMF synthesis filter banks. In the MPEG Surround decoder the process is reversed, that is, a complementary preprocessing step is applied to the downmix signal before entering into the upmix process.

There are several advantages to the scheme described. First the matrix-surround compatibility comes without any additional spatial information. (The only information that has to be transmitted to the decoder is whether the MTX processing is enabled or disabled.) Second the ability to invert the matrix-surround compatibility processing guarantees that there is no negative effect on the multichannel reconstruction quality. Third the decoder is also capable of generating a "regular" stereo downmix from a provided matrix-surround compatible downmix. Last but not least, this feature enables an operation mode where MPEG Surround can operate without the use of spatial side information (see later).

4.2 Enhanced Matrix Mode

In some application scenarios the transmission of spatial side information is undesirable, or even impossible. For example, a specific core coder may not provide the possibility of transmitting an additional parameter stream. Also in analog systems, the transmission of additional digital data can be cumbersome. Thus in order to broaden the application range of MPEG Surround even further, the specification also provides an operation mode that does not rely on any explicit transmission of spatial parameters. This mode is referred to as enhanced matrix mode and uses similar principles as matrix-surround systems to convey multichannel audio.

The MPEG Surround encoder is used to generate a matrix-surround compatible stereo signal (as described previously in the section on matrix-surround compatibility). Alternatively the stereo signal may be generated using a conventional matrix-surround encoder. The MPEG Surround decoder is then operated without externally provided side information. Instead the parameters required for spatial synthesis are derived from an analysis stage working on the received downmix. In particular these parameters comprise channel level difference (CLD) and interchannel cross correlation (ICC) cues estimated between the left and right matrix-surround compatible downmix signals. Subsequently these downmix parameters are mapped to spatial parameters according to the MPEG Surround format, which can then be used to synthesize multichannel output by an MPEG Surround spatial synthesis stage. Fig. 14 illustrates this concept. The MPEG Surround decoder analyzes the downmix and maps the downmix parameters to the parameters needed for the spatial synthesis.

The enhanced matrix mode thus extends the MPEG Surround operating range in terms of bit rate versus quality. This is illustrated in Fig. 11. Without transmission of spatial parameters, the enhanced matrix mode outperforms conventional matrix-surround systems in terms of quality (see also Section 5).

4.3 Artistic Downmix Handling

Contemporary consumer media of multichannel audio (DVD-video/audio, SA-CD and so on) in practice deliver both dedicated multichannel and stereo audio mixes that are separately stored on the media. Both mixes are created by a sound engineer who expresses his/her artistic creativity by "manually" mixing the recorded sound sources using different mixing parameters and audio effects. This implies that a stereo downmix, such as the one produced

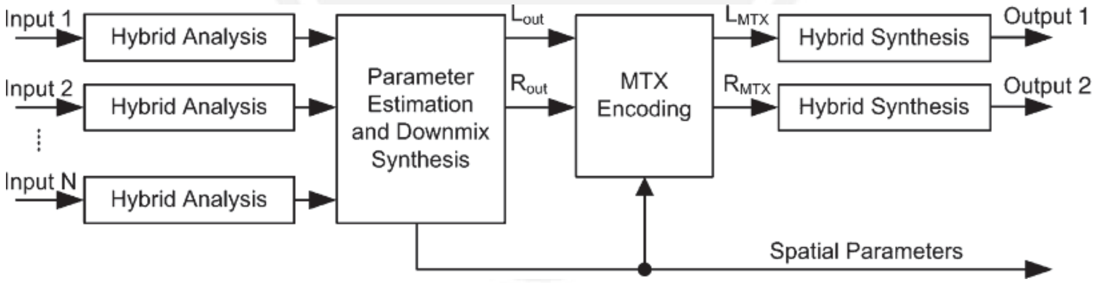

Fig. 13. MPEG Surround encoder with postprocessing for matrix-surround (MTX) compatible downmix.

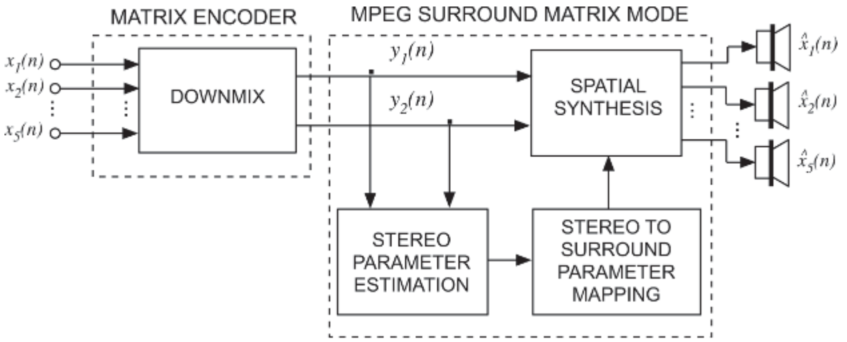

Fig. 14. MPEG Surround decoding without side information.

by the MPEG Surround coder (henceforth referred to as spatial downmix), may be quite different from the sound engineer's stereo downmix (henceforth referred to as artistic downmix).

In the case of a multichannel audio broadcast using the MPEG Surround coder, there is a choice as to which downmix to transmit to the receiver. Transmitting the spatial downmix implies that all listeners not in the possession of a multichannel decoder would listen to a stereo signal that does not necessarily reflect the artistic choices of a sound engineer. In contrast to matrix-surround systems, however, MPEG Surround allows the artistic downmix to be transmitted and thus guarantees optimum sound quality for stereo listeners. In order to minimize potential impairments of the reproduced multichannel sound resulting from using an artistic downmix signal, several provisions have been introduced into MPEG Surround, which are described subsequently.

A first layer of decoder parameters transforms the artistic downmix such that some of the statistical properties of the transformed artistic downmix match those of the MPEG Surround downmix. In addition a second enhancement layer transforms the artistic downmix such that a waveform match with the spatial downmix is achieved.

A match of the statistical properties is obtained by computing frequency-dependent gain parameters for each downmix channel at the encoder side. These parameters match the energy of the artistic downmix channels to the energy of the corresponding spatial downmix channels. These so-called artistic downmix gains (ADGs) employ the same time–frequency grid as the spatial parameters.

In order to obtain a (complete or band-limited) waveform reconstruction of the MPEG Surround downmix in the decoder, the encoder computes enhancement signals. These signals are very similar to the residual signals in the spatial synthesis in the sense that they complement the parametric reconstruction by the ADGs, obtaining a waveform match. Therefore these enhancement signals, or artistic downmix residuals, are coded as residual signals. (See Section 4.5 for a more in-depth description of residual coding.) The support for artistic downmix residual is not part of the baseline MPEG Surround profile as outlined in Section 4.7.3.

4.4 MPEG Surround Binaural Rendering

One of the most recent extensions of MPEG Surround is the capability to render a three-dimensional/binaural stereo output. Using this mode, consumers can experience a three-dimensional virtual multichannel loudspeaker setup when listening over headphones. Especially for mobile devices (such as mobile DVB-H receivers) this extension is of significant interest.

Two distinct use cases are supported. The first use case is referred to as binaural decoding. In this case a conventional MPEG Surround downmix/spatial parameter bit stream is decoded using a so-called binaural decoding mode. This mode generates a stereo signal that evokes a (virtual) multichannel audio experience when played over legacy stereo headphones.

In the second use case, referred to as three-dimensional, the binaural rendering process is applied at the encoder side. As a result, legacy stereo devices will automatically render a virtual multichannel setup over headphones. If the same (three-dimensional) bit stream is decoded by an MPEG Surround decoder attached to a multichannel loudspeaker system, the transmitted three-dimensional downmix can be converted to (standard) multichannel signals optimized for loudspeakers.

Within MPEG Surround both use cases are covered using a new technique for binaural synthesis. Conventional binaural synthesis algorithms typically use head-related transfer functions (HRTFs). These transfer functions describe the acoustic pathway from a sound source position to both ear drums. The synthesis process comprises convolution of each virtual sound source with a pair of HRTFs (such as $2N$ convolutions, with N being the number of sound sources). In the context of MPEG surround, this method has several disadvantages:

- Individual (virtual) loudspeaker signals are required for HRTF convolution. Within MPEG surround this means that multichannel decoding is required as intermediate step.
- It is virtually impossible to undo or invert the encoder-side HRTF processing at the decoder (which is needed in the second use case for loudspeaker playback).
- Convolution is most efficiently applied in the FFT domain while MPEG Surround operates in the QMF domain.

To circumvent these potential problems, MPEG Surround binaural synthesis is based on new technology that operates in the QMF domain without (intermediate) multichannel decoding. The incorporation of this technology in the two different use cases is outlined in the following sections. A more detailed description can be found in [17].

4.4.1 Binaural Decoding in MPEG Surround

The binaural decoding scheme is outlined in Fig. 15. The MPEG Surround bit stream is decomposed into a downmix bit stream and spatial parameters. The downmix decoder produces conventional mono or stereo signals that are subsequently converted to the hybrid QMF domain by means of the MPEG Surround QMF analysis filter bank. A binaural synthesis stage generates the (hybrid QMF-domain) binaural output by means of a two-channels-in, two-channels-out matrix operation. Hence no intermediate multichannel upmix is required. The matrix elements result from a combination of the transmitted spatial parameters and HRTF data. The hybrid QMF synthesis filter bank generates the time-domain binaural output signal.

In case of a mono downmix, the 2 × 2 binaural synthesis matrix has as inputs the mono downmix signal and the same signal processed by a decorrelator. In case of a stereo downmix, the left and right downmix channels form the input of the 2 × 2 synthesis matrix.

The parameter combiner that generates binaural synthesis parameters can operate in two modes. The first mode is a high-quality mode, in which HRTFs of arbitrary length can be modeled very accurately. The resulting 2 × 2 synthesis matrix for this mode can have multiple taps in the time (slot) direction. The second mode is a low-complexity mode. In this mode the 2 × 2 synthesis matrix has a single tap in the time direction and is real-valued for approximately 90% of the signal bandwidth. It is especially suitable for low-complexity operation and/or short (anechoic) HRTFs. An additional advantage of the low-complexity mode is the fact that the 2 × 2 synthesis matrix can be inverted, which is an interesting property for the second use case, as outlined subsequently.

4.4.2 MPEG Surround and Three-Dimensional Stereo

In this use case the three-dimensional processing is applied in the encoder, resulting in a three-dimensional stereo downmix that can be played over headphones on legacy stereo devices. A binaural synthesis module is applied as a postprocess after spatial encoding in the hybrid QMF domain, in a similar fashion as the matrix-surround compatibility mode (see Section 2.5). The three-dimensional encoder scheme is outlined in Fig. 16. Its postprocess comprises the same invertible 2 × 2 synthesis matrix as used in the low-complexity binaural decoder, which is controlled by a combination of HRTF data and extracted spatial parameters. The HRTF data can be transmitted as part of the MPEG Surround bit stream using a very efficient parameterized representation.

The corresponding decoder for loudspeaker playback is shown in Fig. 17. A three-dimensional/binaural inversion stage operates as preprocess before spatial decoding in the hybrid QMF domain, ensuring maximum quality for multichannel reconstruction.

4.5 Residual Coding

While a precise parametric model of the spatial sound image is a sound basis for achieving a high multichannel audio quality at low bit rates, it is also known that parametric coding schemes alone are usually not able to scale up all the way in quality to a “transparent” representation of sound, as this can only be achieved by using a fully discrete multichannel coding technique, requiring a much higher bit rate.

In order to bridge this gap between the audio quality of a parametric description and transparent audio quality, the MPEG Surround coder supports a hybrid coding technique. The nonparametric part of this hybrid coding scheme is referred to as residual coding.

As described before, a multichannel signal is downmixed and spatial cues are extracted. During the process of

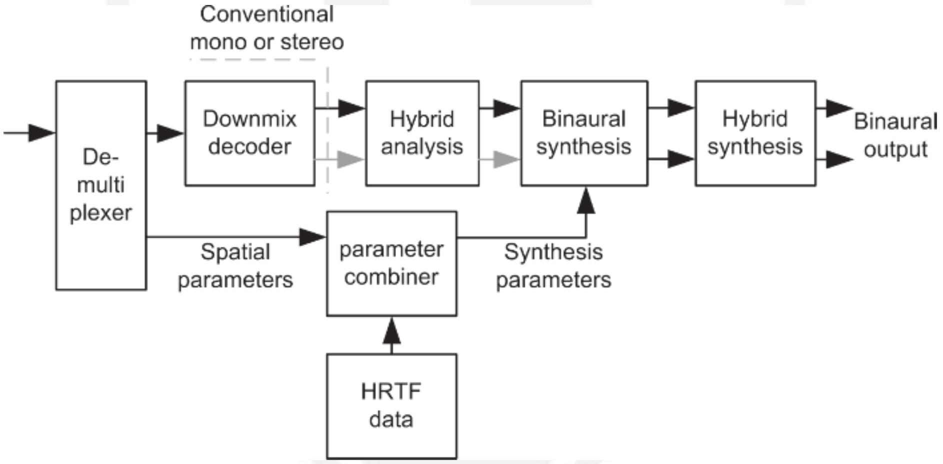

Fig. 15. Binaural decoder schematic.

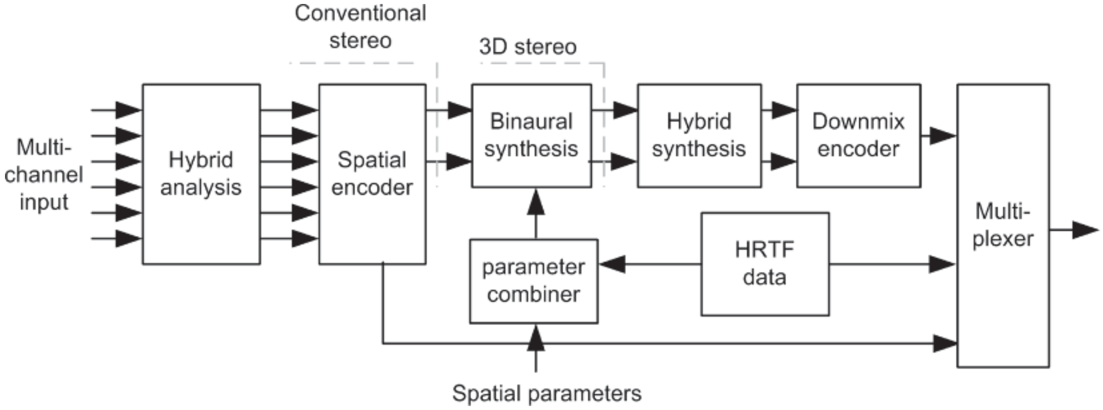

Fig. 16. Three-dimensional encoder schematic.

downmixing, residual signals are calculated, which represent the error signal. These signals can be discarded, as their perceptual relevance is rather low. In the hybrid approach these residual signals, or a band-limited version thereof, are encoded and transmitted to (partially) replace the decorrelated signals in the decoder, as illustrated in Fig. 18. This provides a waveform match between the original and the decoded multichannel audio signals for the transmitted bandwidth.

The (band-limited) residual signals are encoded by an encoder conforming to the MPEG-2 AAC low-complexity profile [27]. The resulting AAC frames are embedded in the spatial bit stream as individual channel stream elements, which is illustrated in Fig. 19. Transients in the residual signals are handled by utilizing the AAC block switching mechanism and temporal noise shaping (TNS) [28]. For arbitrary downmix residuals (Section 4.3), channel pair elements as defined for the MPEG-2 AAC low-complexity profile [27] are used as well.

A tradeoff between bit-rate increase and multichannel audio quality can be made by selecting appropriate residual bandwidths and corresponding bit rates. After encoding, the MPEG Surround bit stream is scalable in the sense that the residual signal related data can be stripped from the bit

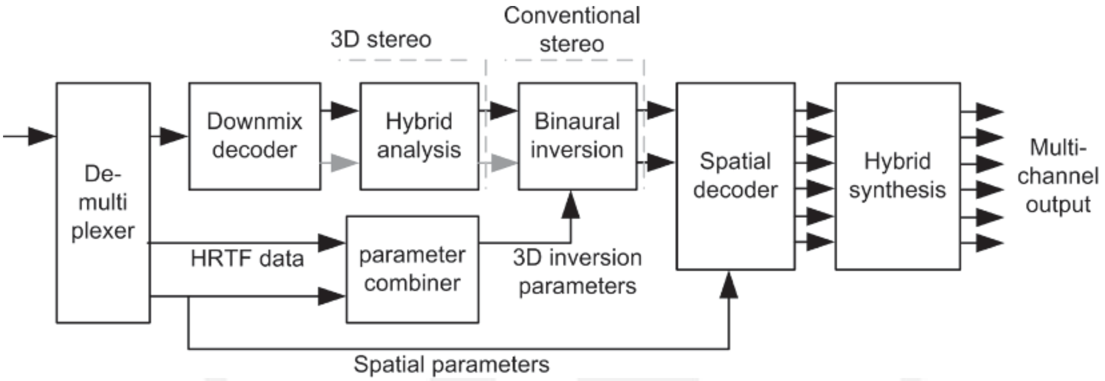

Fig. 17. Three-dimensional decoder for loudspeaker playback.

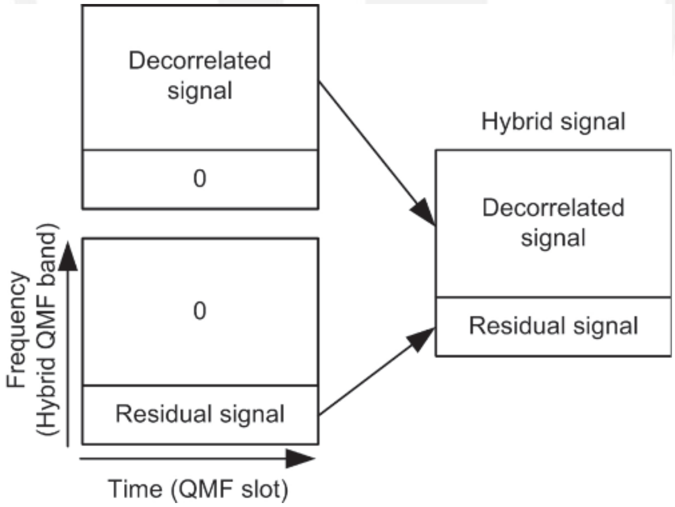

Fig. 18. Complementary decorrelated and residual signals combined into a hybrid signal.

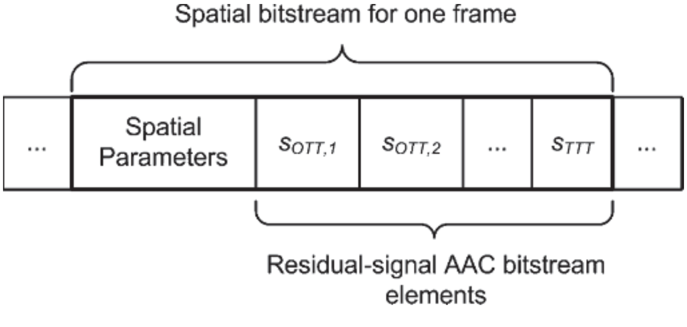
Spatial bitstream for one frame
...
Spatial Parameters
$s_{OTT,1}$
$s_{OTT,2}$
...
s_{TTT}
...
Residual-signal AAC bitstream elements

Fig. 19. Embedding of residual-signal bit-stream elements for each OTT and TTT element in spatial audio bit stream.

stream, thus lowering the bit rate, such that an MPEG Surround decoder reverts back to fully parametric operation (that is, using decorrelator outputs for the entire frequency range). Listening test results have shown that a significant quality gain is obtainable by utilizing residual signals.

4.6 Other Tools

4.6.1 Quantization and Lossless Coding

MPEG Surround provides quantization mechanisms tailored individually to each type of parameter in combination with a set of sophisticated entropy coding techniques, thus striving to minimize the amount of spatial side information while at the same time offering the best possible multichannel audio quality.

Depending on the parameter type, specific quantization schemes are used. Whereas, for example, the inter-channel correlation (ICC) parameters are quantized with as few as eight quantization steps, the channel-prediction coefficients (CPC) used in the TTT box call for a much higher precision of up to 51 quantization steps. Both uniform quantization (such as for CPCs) and nonuniform quantization [such as for channel level differences (CLDs)] are used for some parameter types. Some of the quantization mechanisms can be switched in their resolution to accommodate further reduction of the side information in the case of very low-bit-rate application scenarios. For instance, it is possible to switch to a coarse CPC quantization mode, where the coder only makes use of every second quantization step. (See also subsection on rate/distortion scalability.)

For further saving of side information, entropy coding is applied to the majority of the quantized parameters, generally as a combination of differential coding and Huffman coding. (Only the guided envelope shaping tool makes use of a combination of run length coding and Huffman coding.) Differential coding is conducted relative to neighbor parameter values in either frequency or time directions. Differential coding over time can be conducted relative to the value in the predecessor frame or between values within the same frame. A special case of differential coding is the so-called pilot-based coding, where the differences of a whole set of parameters are calculated relative to one separately transmitted constant value called the pilot value.

The differentially encoded coefficients are subsequently entropy coded using one- or two-dimensional Huffman codebooks. In case of a two-dimensional Huffman table, the pair of values which is represented by one Huffman code word belong to parameters neighboring in either the frequency or time direction. In order to keep the codebooks as small as possible, symmetries within the statistical distribution of the differentially encoded parameters are exploited by applying the same Huffman code words to groups of equally likely parameter tuples. Each of these entropy coding schemes can be combined with any of the differential encoding schemes mentioned, and there are separate Huffman codebooks that were trained for every parameter type and coding scheme.

Finally, for the rare case of outlier signal statistics where none of these entropy coding schemes results in a sufficiently low bit consumption, a grouped PCM coding scheme is available as a fall-back strategy, which consumes a fixed average number of bits per transmitted coefficient (which is only determined by the number of quantization steps). Thus an upper limit of the side information bit rate per MPEG Surround frame can be guaranteed.

4.6.2 Subband Domain Temporal Processing (STP)

In order to synthesize decorrelation between output channels a certain amount of diffuse sound is generated by the spatial decoder's decorrelator units and mixed with the "direct" (nondecorrelated) sound. In general the diffuse-signal temporal envelope does not match the direct signal envelope, resulting in a temporal smearing of transients.

For low-bit-rate applications the subband domain temporal processing (STP) tool can be activated to render temporal shaping since only a binary shaping decision needs to be coded for each upmix channel. STP mitigates the aforementioned smeared-transient effect by shaping the envelope of the diffuse signal portion of each upmix channel to approximately match the temporal shape of the transmitted downmix signal.

A schematic of the tool is shown in Fig. 20. If STP is activated, the direct and the diffuse signal contributions to the final upmix signal are synthesized separately. To compute the shaping factors for the diffuse signal portion, the temporal energy envelope of the downmixed direct portion of the upmix signal and the temporal energy envelope of the diffuse portion of each upmix channel are estimated. The shaping factors are computed as the ratio between the two energy envelopes. The shaping factors are subjected to some postprocessing to achieve a compromise between restoring a crisp transient effect and avoiding potential distortions before being applied to the high-frequency part of the diffuse signals. Finally each shaped diffuse signal is combined with its corresponding direct signal to reconstruct the particular upmix channel.

4.6.3 Guided Envelope Shaping (GES)

While the STP tool is suitable for enhancing the subjective quality of, for example, applause-like signals, it is still subject to some limitations:

- The spatial redistribution of single, pronounced transient events in the sound stage is limited by the temporal resolution of the spatial upmix, which may span several attacks at different spatial locations.
- The temporal shaping of diffuse sound may lead to characteristic distortions when applied in a rigorous way.

The guided envelope shaping (GES) tool provides enhanced temporal and spatial quality for such signals while avoiding distortion problems. An overview of the GES tool placement in the processing chain is provided in Fig. 21.

Additional side information is transmitted by the encoder to describe the broad-band fine-grain temporal envelope structure of the individual channels, and thus allow for sufficient temporal/spatial shaping of the upmix channel signals at the decoder side. In the decoder the basic upmix is performed separately for the direct and diffuse signal parts. The associated GES processing only alters the direct part of the upmix signal in a channel, thus promoting the perception of transient direction (precedence effect) and avoiding additional distortion.

Since the diffuse signal also contributes to the overall energy balance of the upmixed signal, GES accounts for this by calculating a modified broad-band scaling factor from the transmitted information that is applied solely to the direct signal part. The factor is chosen such that the overall energy in a given time interval is approximately the same as if the original factor had been applied to both the direct and the diffuse parts of the signal. Finally direct and diffuse signal parts are mixed and passed on to the synthesis filter-bank stage.

Using GES the best subjective audio quality for applause-like signals is obtained if a coarse spectral resolution of the spatial cues is chosen. In this case use of the GES tool does not necessarily increase the average spatial side information bit rate, since spectral resolution is traded advantageously for temporal resolution.

4.6.4 Adaptive Parameter Smoothing

For low-bit-rate scenarios it is desirable to employ a coarse quantization for the spatial parameters in order to reduce the required bit rate as much as possible. This may result in artifacts for certain kinds of signals. Especially in the case of stationary and tonal signals, modulation artifacts may be introduced by frequent toggling of the parameters between adjacent quantizer steps. For slowly moving point sources the coarse quantization results in a step-by-step panning rather than a continuous movement of the source and is thus usually perceived as an artifact. The adaptive parameter smoothing tool, which is applied on the decoder side, is designed to address these artifacts by temporally smoothing the dequantized parameters for signal portions with the described characteristics. The adaptive smoothing process is controlled from the encoder by transmitting additional side information.

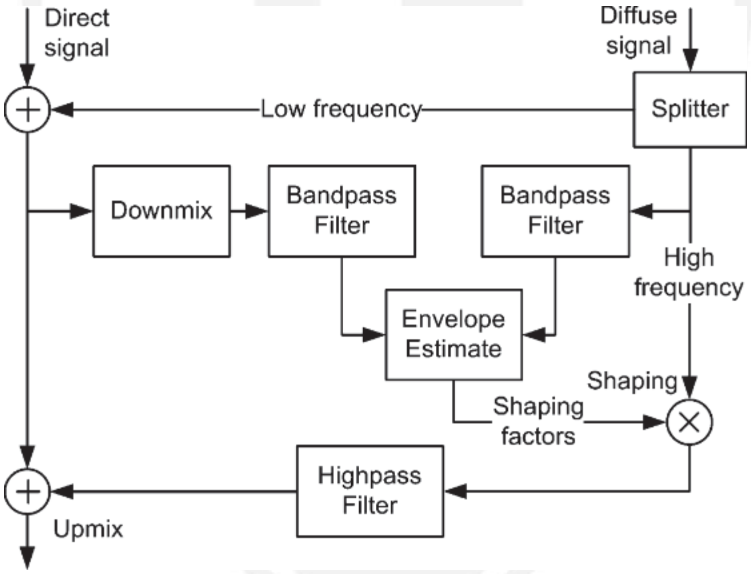

Fig. 20. Subband temporal processing (STP).

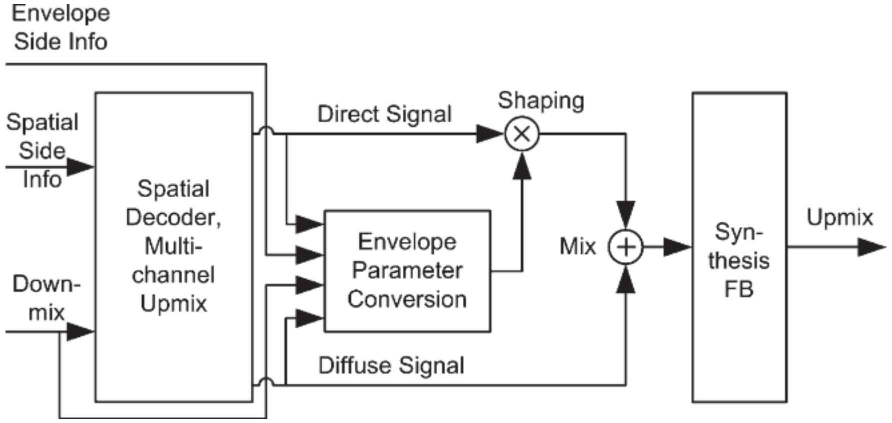

Fig. 21. Guided envelope shaping (GES).

4.6.5 Channel Configurations

In Section 4.2.3 the tree-based spatial parameterization for the standard 5.1-channel configuration was described. In addition to this, MPEG Surround supports almost arbitrary channel configurations by extending the tree structures shown in Figs. 8 and 9 by additional OTT elements. For 7.1 material (with either four surround channels or, alternatively, five front channels) dedicated 7–2–7 tree configurations are defined using a stereo downmix. Furthermore there are 7–5–7 trees defined that convey a 5.1 downmix of the original 7.1-channel configuration.

In order to enable flexible support of arbitrary channel configurations, a so-called arbitrary tree extension mechanism is available in MPEG Surround, which can extend any of the previously defined tree configurations by additional subtrees built from only OTT modules. For reasons of complexity, decorrelation is not included in these additional OTT modules.

When considering playback of MPEG Surround coded material on a device that only provides stereo output (such as a portable music player), a stereo downmix, if available, can be played back directly without any MPEG Surround decoding. In case of a 5–1–5 configuration using a mono downmix it is, however, more desirable to play back a stereo dowmix of the coded 5.1 signal than to play just the mono downmix. In order to avoid the computational complexity of a full MPEG Surround decoding to 5.1 channels, followed by a downmix from this 5.1 signal to stereo, a dedicated stereo-output decoding mode is defined for MPEG Surround coded material using a 5–1–5 configuration. In this mode spatial parameters of a virtual stereo downmix are calculated directly in the parameter domain, based on the received spatial side information. This results in just a single pair of CLD and ICC parameters for one OTT module that generates the desired stereo output from the mono downmix, similar to plain parametric stereo coding.

4.7 System Aspects

4.7.1 Carriage of Side Information

In order to achieve backward compatibility with legacy devices that are not aware of MPEG Surround, the spatial side information needs to be embedded in the transmitted downmix signal in a backward-compatible manner, such that the downmix itself can still be decoded by the devices. Depending on the technology used to code and transmit the downmix, different embedding mechanisms for MPEG Surround are defined.

In case of the downmix being coded with MPEG-1/2 Layer I/II/III perceptual audio coders, the spatial side information is embedded in the ancillary data part of the downmix bit stream. Detection of the embedded spatial side information is achieved by means of a dedicated sync word, and the consistency of the embedded data is verified using a CRC mechanism.

In case of the downmix being coded with MPEG-2/4 AAC, the extension payload data container of the fill elements available in the AAC bit-stream syntax is used to identify and convey the spatial side information. If required, the spatial side information associated with one MPEG Surround frame can be split and carried in several fill elements, and also distributed over several subsequent AAC frames. If HE-AAC (MPEG-4 high efficiency AAC) [26] is used as the downmix coder, both SBR and MPEG Surround side information can be embedded at the same time.

When MPEG Surround is used in an MPEG-4 systems environment, the spatial side information can be either embedded in the downmix bit stream (as described) or, alternatively, conveyed as a separate elementary stream (ES) that depends on the ES carrying the coded downmix itself.

Furthermore it is possible to embed the spatial side information as buried data [29] in a PCM waveform representation of the downmix. For this, a randomized version of the spatial side information is embedded in the least significant bits of the PCM downmix. This bit-stream embedding is rendered inaudible by employing subtractively dithered noise shaping controlled by the masked threshold of the downmix signal. This PCM buried data technique allows to store MPEG Surround side information, for example, on a regular audio CD, and to transmit it over digital audio interfaces such as S/P-DIF and AES/EBU in a backward-compatible manner.

While the MPEG Surround standard defines the carriage of the spatial side information for the MPEG family of perceptual audio codecs (and for PCM waveforms), MPEG Surround can be used in combination with any other downmix coding and transmission system. To this end it is merely necessary to define a way how the spatial side information is embedded in the downmix or conveyed in parallel to the downmix signal.

4.7.2 Efficient Combination with HE-AAC

When MPEG Surround is used in combination with a downmix coded with HE-AAC, it is of interest to note that both systems make use of a 64-band QMF bank of the same type. Therefore it is possible to connect the MPEG Surround decoder directly to the output of an HE-AAC downmix decoder in the QMF signal domain. This direct connection avoids the QMF synthesis at the output of the HE-AAC decoder and the QMF analysis at the input of the MPEG Surround decoder, which would be necessary if both decoders were connected using the normal time-domain representation of the downmix signal. In this way the overall computational complexity of an integrated HE-AAC and MPEG Surround decoder is reduced significantly for both high-quality and low-power processing.

4.7.3 MPEG Surround Profiles and Levels

MPEG defines decoder profiles as technology sets to be used for certain application fields and ensuring interoperability on a bit-stream basis. Currently there is a single baseline MPEG Surround profile available. The following tools are contained in this profile:

- Artistic downmix functionality
- Matrix compatibility

- Enhanced matrix mode decoding
- Temporal shaping
- Residual coding (not including artistic downmix residuals)
- Binaural decoding
- Three-dimensional audio decoding
- Low-power decoding.

This profile is further subdivided into six hierarchical levels, which come with an increasing number of output channels, range of sampling rates, and bandwidth of the residual signal when advancing to a higher level. Table 1 lists the supported tree configurations, maximum number of audio output channels, availability of residual coding, and complexity for each level.

Level 1 allows for stereo output of multichannel content with and without the use of binaural rendering techniques. It is ideally suited for mobile devices that are equipped with stereo loudspeakers or headphones and exhibits very low computational complexity. This level includes the ability to upmix from a mono-based operation to stereo loudspeaker reproduction, the 5–1–2 mode with 5–1–5 streams. Levels 2 and 3 offer in addition discrete 5.1 output for loudspeakers. Levels 4 and 5 extend to 7.1 output, while Level 6 can handle up to 32 channels and operate at up to 96-kHz sampling rate.

Independently from this level hierarchy, an MPEG Surround decoder can be implemented either as a high-quality (HQ) decoder or as a low-power (LP) decoder in order to optimally support both stationary equipment as well as battery-powered mobile devices.

The worst-case decoding complexity in terms of processor complexity units (PCUs) and RAM complexity units (RCUs) is listed in Table 2 for the different levels. Processor complexity units are specified in MOPS, and RAM complexity units are expressed in kwords (1000 words), that is, the native data representation of the computing platform (16, 24, or 32 bit). Both HQ and LP operation is shown. The MPEG Surround decoder introduces a total delay of 1281 samples for the HQ decoder and 1601 samples for the LP decoder.

Table 1. Levels of baseline profile.

Level	Tree Configuration	Maximum Output Channels	Maximum F_s	Residual Coding
1	515, 525, 727	2.0	48 kHz	N/A
2	515, 525, 727	5.1	48 kHz	N/A
3	515, 525, 727	5.1	48 kHz	Yes
4	515, 525, 727	7.1	48 kHz	Yes
5	515, 525, 757, 727	7.1	48 kHz	Yes
6	515, 525, 757, 727, plus arbitrary tree extension	32 incl. 4 LFE	96 kHz	Yes

Table 2. Processor and RAM complexity depending on decoder level and operation.

Level	PCU (HQ)	RCU (HQ)	PCU (LP)	RCU (LP)
1	12	5	6	4
2	25	15	12	11
3	25	15	12	11
4	34	21	17	15
5	34	21	17	15
6	70	38	44	32
(f_s = 48 kHz)				

For the combination of MPEG Surround decoding with an HE-AAC core decoder, the complexity savings due to sharing QMF filter banks apply as mentioned in the previous section. This leads to low total complexity numbers for a combined 525 system (in PCU/RCU assuming level 2), where for each of the two stereo channels two QMF filter-bank stages with an associated individual complexity of 1.5/0.5 (HQ) or 1.0/0.5 (LP) are omitted:

High quality	28/23 (saves 6/2)
Low power	15/17 (saves 4/2)

Here the LP operation for a combined HE-AAC + MPEG Surround decoder comes with a significant reduction in processing complexity of almost 50% when compared to HQ operation. The performance differences between both modes are discussed in the next section.

5 PERFORMANCE

In the period between August 2006 and January 2007, MPEG conducted a formal verification test for the MPEG Surround specification [30]. Such a test is not only a validation of the standardized technology but it also provides relevant test data on MPEG Surround to the interested industry. For this purpose two use cases relevant to industry were considered, a DVB-oriented use case and a music-store/portable player use case. An additional test was included to evaluate the difference between the HQ and LP decoding modes of MPEG Surround. A total of 148 subjects at eight different test sites participated in a total of five different tests. All tests were conducted conforming to the MUSHRA [31] test methodology and included a reference and a 3.5-kHz low-pass anchor.

5.1 DVB-Oriented Use Case

The test conditions for the DVB-oriented use case strive to compare MPEG Surround in DVB-like applications with other potential MPEG Audio solutions for that space. With reference to labels of the results for this multichannel test in Fig. 22, two scenarios are covered:

- The establishment of a new service, for which a stereo backward compatible part at highest quality per bit is desired. For this scenario, MPEG-4 high-efficiency AAC in combination with MPEG Surround (HE-AAC_MPS) is compared with discrete multichannel MPEG-4 HE AAC (HE-AAC_MC) at bit rates of 64 and 160 kbit/s total.
- The extension of an existing stereo service with surround sound in a backward-compatible fashion. For this scenario MPEG-1 Layer 2 (as is currently employed in DVB) is extended with MPEG Surround (Layer2_MPS)

at a total rate of 256 kbit/s. This mode is compared to Dolby Prologic II encoding and decoding using MPEG-1 Layer 2 at 256 kbit/s for coding the downmix (Layer2_DPL2).

In addition to this multichannel test a corresponding downmix test has been performed to verify the MPEG Surround downmix quality.

From the results in Fig. 22 it can be concluded that MPEG Surround provides improved coding efficiency at low bit rates and equal quality, while including backward-compatible downmix at higher bit rates. This makes it a perfect candidate for upgrading existing DVB broadcasting services to surround sound. In addition MPEG Surround is vastly superior to the matrixed system at the same total bit rate.

5.2 Music-Store/Portable Player Use Case

The test conditions for this use case strive to test the performance of MPEG Surround in music-store-like applications. With reference to the labels of the results for this multichannel test in Fig. 23, a scenario is envisioned where an online music store selling stereo material using AAC at 160 kbit/s extends the service to provide multichannel content in a backward-compatible way. This is achieved by providing in addition low-bit-rate MPEG Surround data (AAC_MPS). Hence a consumer can play the backward-compatible part on his/her legacy stereo player while enjoying a multichannel experience when playing the content at home using an MPEG Surround enabled multichannel setup.

This point of operation is compared to MPEG Surround decoding of the AAC_MPS bit streams in enhanced matrix mode (AAC_MPS_EMM, not employing the spatial side information) and Dolby Prologic encoding and decoding using AAC at 160 kbit/s for coding the downmix (AAC_DPLII).

From the results in Fig. 23 it is concluded that MPEG Surround at 192 kbit/s provides excellent multichannel quality, thus being very attractive for upgrading existing electronic music-store services. In the enhanced matrix decoding mode, where the side information is not exploited, the results are still well within the “good” quality range. Nevertheless, the MPEG Surround enhanced matrix mode shows to be superior to legacy matrix technology (Dolby Prologic II).

An additional headphone test has been conducted in order to assess the quality of the binaural decoding capability of MPEG Surround, particularly for this music-store use case. With reference to the labels of the results for this binaural test in Fig. 24, three configurations were tested. High-quality binaural decoding (AAC_MPS_BIN) and low-complexity binaural decoding combined with low-power MPEG Surround decoding (AAC_MPS_BIN_LC) at a total of 191 kbit/s for stereo head-phone listening at home. Reference signals have been created by applying

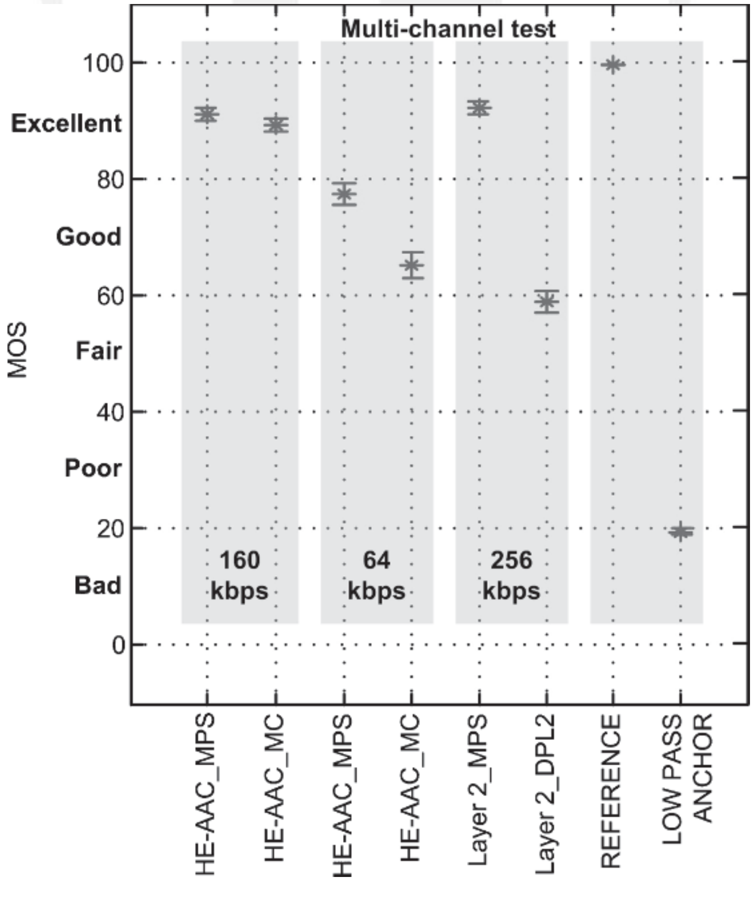

Fig. 22. Test results for DVB oriented use case.

HRTF filtering to the MPEG Surround multichannel output (AAC_MPS_BIN_Reference) and the multichannel original (reference). The low-pass anchor is the 3.5-kHz bandwidth limitation of the reference. Note that residual signals are not used in the binaural decoding mode. The results in Fig. 24 show that MPEG Surround binaural rendering offers excellent quality.

5.3 Additional Tests

An additional technology-driven test shows that the sound quality of the LP version is statistically comparable to the sound quality of the HQ version. Moreover, as shown in Table 2, the LP version requires only about half the power consumption and three quarters the RAM requirement of the HQ version. A number of further test results from earlier evaluations of other operating points within the wide range of possible MPEG Surround operating conditions can be found, for example, in [18].

6 APPLICATIONS

MPEG Surround enables a wide range of applications. Among the many conceivable applications are music download services, streaming music services/Internet radios, digital audio broadcasting, multichannel teleconferencing, and audio for games.

MPEG Surround is perfectly suited for digital audio broadcasting. Given its inherent backward compatibility and low overhead of the side information, MPEG Surround can be introduced in existing systems without rendering legacy receivers obsolete. Legacy and stereo-only receivers will simply play the stereo downmix, whereas an MPEG Surround enabled receiver will play the true multichannel signal. Test transmissions using over the air transmission of MPEG Surround and Eureka digital audio broadcasting (DAB) have been carried out on several occasions, as well as test transmissions using MPEG Surround with HDC (the U.S. terrestrial digital radio system).

Given the ability to do surround sound with low overhead and the ability to do binaural rendering, MPEG Surround is ideal for the introduction of multichannel sound on portable devices. As one very attractive example, an Internet music store can upgrade its content to surround and—using an MPEG Surround enabled portable player—the user can get a surround sound experience over headphones. When connecting the portable player to a surround sound AV receiver by means of a cradle or docking station, the content can be played over loudspeakers in true surround sound. As is always the case with MPEG Surround, legacy devices will not take notice of the MPEG Surround data and play the backward-compatible downmix.

The existing DVB-T system uses MPEG-1 Layer 2 and can only be expanded to multichannel by means of simulcast, at high bit-rate penalty, or by matrix-surround systems, resulting in low multichannel audio quality. MPEG

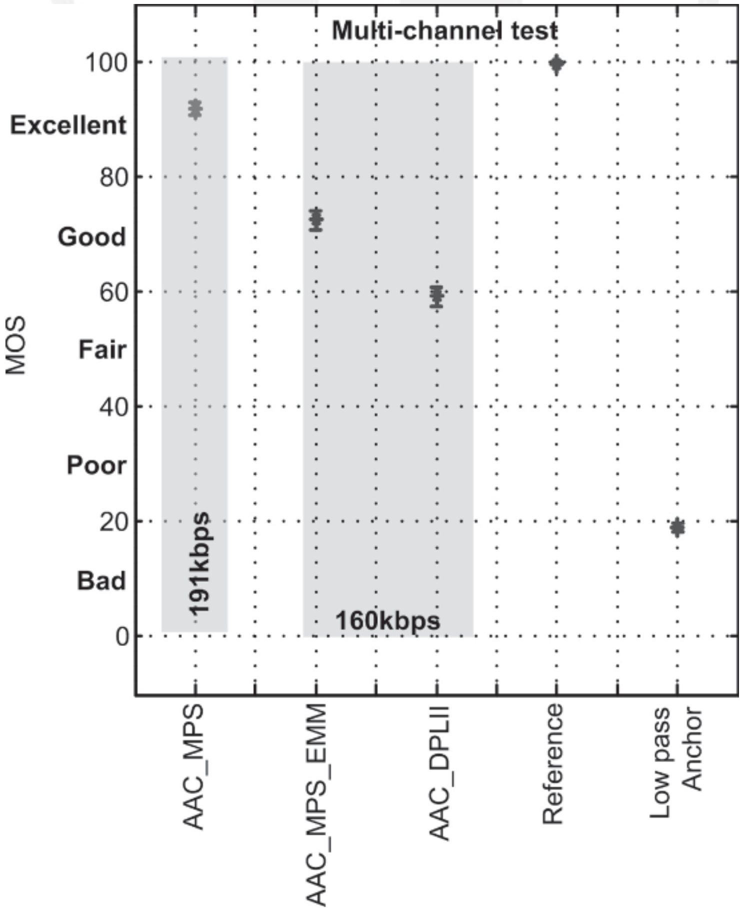

Fig. 23. Test results for music-store/portable player use case.

Surround effectively solves these problems. An MPEG Surround decoder can be introduced in the set-top box, decoding the Layer 2 downmix and the MPEG Surround data into a multichannel signal. This enables a DVB-T system to be upgraded to surround sound by simply offering new MPEG Surround enabled set-top boxes to customers who wish to have surround sound TV. Legacy receivers will decode the Layer 2 downmix as usual, not affected by the MPEG Surround data. Test transmissions have been carried out successfully using MPEG Surround over DVB-T.

7 CONCLUSIONS

After several years of intense development, the spatial audio coding (SAC) approach has proven to be extremely successful for bit-rate-efficient and backward-compatible representation of multichannel audio signals. Based on these principles, the MPEG Surround technology has been under standardization within the ISO/MPEG group for about two years. This paper describes the technical architecture and capabilities of the MPEG Surround technology and its most recent extensions.

Most importantly, MPEG Surround enables the transmission of multichannel signals at data rates close to the rates used for the representation of two-channel (or even monophonic) audio. It allows for a wide range of scalability with respect to the side information rate, which helps to cover almost any conceivable application scenario. Listening tests confirm the feasibility of this concept. Good multichannel audio quality can be achieved down to very low side information rates (such as 3 kbit/s). Conversely, using higher rates allows approaching the audio quality of a fully discrete multichannel transmission. Along with the basic coding functionality, MPEG Surround provides a plethora of useful features that further increase its attractiveness (such as support for artistic downmix, full matrix-surround compatibility, and binaural decoding) and may promote a quick adoption in the marketplace. Finally MPEG Surround enables the use of multichannel audio on portable devices because of the low overhead of the spatial data and the binaural rendering capabilities.

8 ACKNOWLEDGMENT

A great number of people made this project a reality. The following persons have contributed significantly to the development of MPEG Surround: J. Engdegård, A. Gröschel, L. Sehlström, and L. Villemoes of Coding Technologies; B. Grill, A. Hölzer, M. Multrus, H. Mundt, M. Neusinger, J. Plogsties, C. Spenger, R. Sperschneider, and L. Terentiev of Fraunhofer IIS; T. Norimatsu of Panasonic; and E. Schuijers, M. Ostrovskyy, G. Hotho, F. de Bont, F. Myburg, and M. van Loon of Philips. The authors

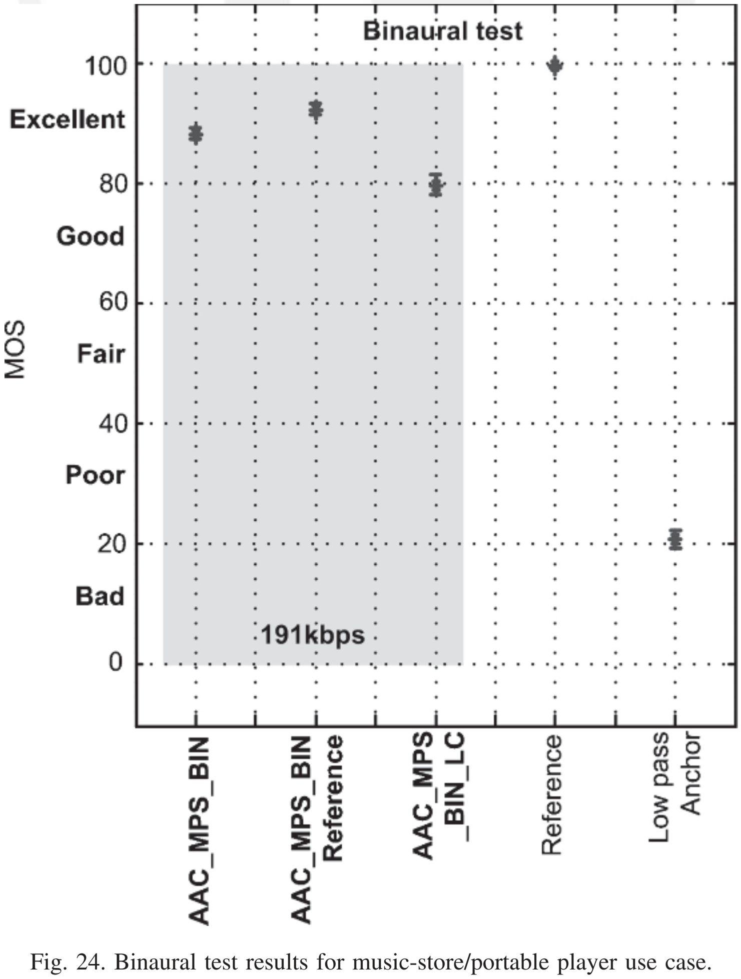

Fig. 24. Binaural test results for music-store/portable player use case.

would furthermore like to express their gratitude to the MPEG Audio Committee and its chair, Schuyler Quackenbush, for the support provided as well as to the many other contributors to the standards effort.

9 REFERENCES

[1] ISO/IEC 23003-1:2007, "Information Technology—MPEG Audio Technologies—Part 1: MPEG Surround," International Standards Organization, Geneva, Switzerland (2007).

[2] ISO/IEC 23003-1:2007/Cor.1:2008, "Information Technology—MPEG Audio Technologies—Part 1: MPEG Surround, TECHNICAL CORRIGENDUM 1," International Standards Organization, Geneva, Switzerland (2008).

[3] J. Herre, "From Joint Stereo to Spatial Audio Coding—Recent Progress and Standardization," presented at the 7th Int. Conf. on Digital Audio Effects (DAFX04) (Naples, Italy, 2004 Oct.).

[4] H. Purnhagen, "Low Complexity Parametric Stereo Coding in MPEG-4," presented at the 7th Int. Conf. on Audio Effects (DAFX-04) (Naples, Italy, 2004 Oct.).

[5] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegård, "Low-Complexity Parametric Stereo Coding," presented at the 116th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 52, p. 800 (2004 July/Aug.), convention paper 6073.

[6] C. Faller and F. Baumgarte, "Efficient Representation of Spatial Audio Using Perceptual Parameterization," presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (New Paltz, NY, 2001).

[7] C. Faller and F. Baumgarte, "Binaural Cue Coding—Part II: Schemes and Applications," *IEEE Trans. Speech Audio Process.,* vol. 11 (2003 Nov.).

[8] C. Faller, "Coding of Spatial Audio Compatible with Different Playback Formats," presented at the 117th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, p. 81 (2005 Jan./Feb.), convention paper 6187.

[9] R. Dressler, "Dolby Surround Prologic Decoder—Principles of Operation," Dolby Publi., http://www.dolby.com/assets/pdf/tech_library/209_Dolby_Surround_Pro_Logic_II_Decoder_Principles_of_Operation.pdf.

[10] D. Griesinger, "Multichannel Matrix Decoders for Two-Eared Listeners," presented at the 101st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 44, p. 1168 (1996 Dec.), preprint 4402.

[11] J. Herre, C. Faller, S. Disch, C. Ertel, J. Hilpert, A. Hölzer, K. Linzmeier, C. Spenger, and P. Kroon, "Spatial Audio Coding: Next-Generation Efficient and Compatible Coding of Multichannel Audio," presented at the 117th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, p. 81 (2005 Jan./Feb.), convention paper 6186.

[12] ISO/IEC JTC1/SC29/WG11 (MPEG), "Call for Proposals on Spatial Audio Coding," Doc. N6455, Munich, Germany (2004).

[13] ISO/IEC JTC1/SC29/WG11 (MPEG), "Report on Spatial Audio Coding RM0 Selection Tests," Doc. N6813, Palma de Mallorca, Spain (2004).

[14] J. Herre, H. Purnhagen, J. Breebaart, C. Faller, S. Disch, K. Kjörling, E. Schuijers, J. Hilpert, and F. Myburg, "The Reference Model Architecture for MPEG Spatial Audio Coding," presented at the 118th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, pp. 693, 694 (2005 July/Aug.), convention paper 6447.

[15] ISO/IEC JTC1/SC29/WG11 (MPEG), "Report on MPEG Spatial Audio Coding RM0 Listening Tests," Doc. N7138, Busan, Korea (2005); available at http://www.chiariglione.org/mpeg/working_documents/mpeg-d/sac/RM0-listening-tests.zip.

[16] J. Breebaart, J. Herre, C. Faller, J. Rödén, F. Myburg, S. Disch, H. Purnhagen, G. Hotho, M. Neusinger, K. Kjörling, and W. Oomen, "MPEG Spatial Audio Coding/MPEG Surround: Overview and Current Status," presented at the 119th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, p. 1228 (2005 Dec.), convention paper 6599.

[17] J. Breebaart, J. Herre, L. Villemoes, Craig Jin, K. Kjörling, J. Plogsties, and J. Koppens: "Multi-Channel Goes Mobile: MPEG Surround Binaural Rendering," presented at the 29th AES Int. Conf. (Seoul, Korea, 2006).

[18] L. Villemoes, J. Herre, J. Breebaart, G. Hotho, S. Disch, H. Purnhagen, and K. Kjörling, "MPEG Surround: The Forthcoming ISO Standard for Spatial Audio Coding," presented at the 28th Int. Conf. (Piteå, Sweden, 2006).

[19] B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hear. Research,* vol. 47, pp. 103–138 (1990).

[20] J. Breebaart, S. van de Par, and A. Kohlrausch, "Binaural Processing Model Based on Contralateral Inhibition—I. Model Setup," *J. Acoust. Soc. Am.,* vol. 110, pp. 1074–1088 (2001).

[21] J. Princen, A. Johnson, and A. Bradley, "Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation," in *Proc. IEEE ICASSP* (1987), pp. 2161–2164.

[22] M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral Band Replication—A Novel Approach in Audio Coding," presented at the 112th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 50, pp. 509, 510 (2002 June), convention paper 5553.

[23] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric Coding of Stereo Audio," *EURASIP J. Appl. Signal Process.,* vol. 9, pp. 1305–1322 (2005).

[24] G. Hotho, L. F. Villemoes, and J. Breebaart, "A Backward-Compatible Multichannel Audio Codec," *IEEE Trans. Audio, Speech, Language Process.,* vol. 16, pp. 83–93 (2008 Jan.).

[25] J. Engdegård, H. Purnhagen, J. Rödén, and L. Liljeryd, "Synthetic Ambience in Parametric Stereo Coding," presented at the 116th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),*

vol. 52, pp. 800, 801 (2004 July/Aug.), convention paper 6074.

[26] M. Wolters, K. Kjörling, D. Homm, and H. Purnhagen, "A Closer Look into MPEG-4 High Efficiency AAC," presented at the 115th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 51, p. 1221 (2003 Dec.), convention paper 5871.

[27] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.,* vol. 45, pp. 789–814 (1997 Oct.).

[28] J. Herre and J. D. Johnston, "Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS)," presented at the 101st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 44, p. 1175 (1996 Dec.), preprint 4384.

[29] A. W. J. Oomen, M. E. Groenewegen, R. G. van der Waal, and R. N. J. Veldhuis, "A Variable-Bit-Rate Buried-Data Channel for Compact Disc," *J. Audio Eng. Soc.,* vol. 43, pp. 23–28 (1995 Jan./Feb.).

[30] ISO/IEC JTC1/SC29/WG11 (MPEG), "Report on MPEG Surround Verification Test," Doc. N8851, Marrakech, Morocco (2007), available at http://www.chiariglione.org/mpeg/working_documents/mpeg-d/sac/VT-report.zip.

[31] ITU-R BS.1534-1, "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)," International Telecommunications Union, Geneva, Switzerland (2001).

THE AUTHORS

J. Herre

K. Kjörling

J. Breebaart

C. Faller

S. Disch

H. Purnhagen

J. Koppens

J. Hilpert

J. Rödén

W. Oomen

K. Linzmeier

K. S. Chong

Jürgen Herre joined the Fraunhofer Institute for Integrated Circuits (IIS) in Erlangen, Germany, in 1989. Since then he has been involved in the development of perceptual coding algorithms for high-quality audio, including the well-known ISO/MPEG-Audio Layer III coder (aka MP3). In 1995 he joined Bell Laboratories for a postdoctoral term, working on the development of MPEG-2 advanced audio coding (AAC). Since the end of 1996 he has been back at Fraunhofer, working on the development of advanced multimedia technology, including MPEG-4, MPEG-7, and secure delivery of audiovisual content, currently as the chief scientist for the audio/multimedia activities at Fraunhofer IIS, Erlangen.

Dr. Herre is a fellow of the Audio Engineering Society, cochair of the AES Technical Committee on Coding of Audio Signals, and vice chair of the AES Technical Council. He also is an IEEE senior member, served as an associate editor of the *IEEE Transactions on Speech and Audio Processing,* and is an active member of the MPEG audio subgroup. Outside his professional life he is a dedicated amateur musician.

●

Kristofer Kjörling was born in Ljungby, Sweden, in 1972. He received an M.Sc. degree in electrical engineering from the Royal Institute of Technology in Stockholm, Sweden, in 1997, concluded by a thesis project on harmonic bandwidth extension (later to be developed into spectral band replication standardized in MPEG as MPEG-4 SBR).

The work was carried out at Coding Technologies AB in Stockholm, where he was working from 1997 to 2007 as a researcher, holding the position of head of research strategy. The main research work has been SBR development and MPEG Surround development. Since 2008 he has been employed by Dolby, where he holds the position of principal member, technical staff.

Mr. Kristofer is an active member of the ISO/IEC MPEG Audio standardization group.

●

Jeroen Breebaart received an M.Sc. degree in biomedical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1997 and a Ph.D. degree in auditory psychophysics from the same university in 2001.

From 2001 to 2007 he was with the Digital Signal Processing Group at Philips Research, conducting research in the areas of spatial hearing, parametric audio coding, automatic audio content analysis, and audio effects processing. Since 2007 he has been the leader of the biometrics cluster of the Information and System Security Group at Philips Research, expanding his research scope toward secure and convenient identification.

Dr. Breebaart is a member of the AES and IEEE. He contributed to the development of audio coding algorithms as recently standardized in MPEG and 3GPP such as HE-AAC, MPEG Surround, and the upcoming standard on spatial audio object coding. He also actively participates in the ISO/IEC IT security techniques standardization committee and is significantly involved in several EU-funded projects. He published more than 50 papers at international conferences and journals and coauthored the book, *Spatial Audio Processing: MPEG Surround and Other Applications.*

●

Christof Faller received an M.S. (Ing.) degree in electrical engineering from ETH Zurich, Switzerland, in 2000, and a Ph.D. degree for his work on parametric multichannel audio coding from EPFL Lausanne, Switzerland, in 2004.

From 2000 to 2004 he worked in the Speech and Acoustics Research Department at Bell Laboratories, Lucent Technologies and Agere Systems (a Lucent company), where he worked on audio coding for digital satellite radio, including parametric multichannel audio coding. He is currently a part-time postdoctoral employee at EPFL Lausanne. In 2006 he founded Illusonic LLC, an audio and acoustics research company.

Dr. Faller has won a number of awards for his contributions to spatial audio coding, MP3 Surround, and MPEG Surround. His main current research interests are spatial hearing and spatial sound capture, processing, and reproduction.

●

Sascha Disch received his Dipl.-Ing. degree in electrical engineering from the Technical University Hamburg-Harburg (TUHH) in 1999.

From 1999 to 2007 he was with the Fraunhofer Institut für Integrierte Schaltungen (Fraunhofer IIS), Erlangen, Germany, where he worked in research and development in the fields of perceptual audio coding and audio processing, including parametric coding of surround sound. Currently he is pursuing a Ph.D. degree in electrical engineering at the Laboratorium für Informationstechnologie, Leibniz University, Hannover (LUH), Germany, in cooperation with Fraunhofer IIS. His primary research interests include audio signal processing/coding and digital audio effects.

Mr. Disch is a co-editor of the MPEG Surround standard.

●

Heiko Purnhagen was born in Bremen, Germany, in 1969. He received an M.S. degree (Diplom) in electrical engineering from the University of Hannover, Germany, in 1994, concluded by a thesis project on automatic speech recognition carried out at the Norwegian Institute of Technology, Trondheim, Norway.

From 1996 to 2002 he was with the Information Technology Laboratory at the University of Hannover, where he pursued a Ph.D. degree related to his research on very low-bit-rate audio coding using parametric signal representations. In 2002 he joined Coding Technologies (now Dolby Sweden) in Stockholm, Sweden, where he is working on research, development, and standardization of low-bit-rate audio and speech coding systems. He contributed to the development of parametric techniques for efficient coding of stereo and multichannel signals, and his recent research activities include the unified coding of speech and audio signals.

Since 1996 Mr. Purnhagen has been an active member of the ISO MPEG standardization committee and editor or co-editor of several MPEG standards. He is the principal author of the MPEG-4 parametric audio coding specification known as HILN (harmonic and individual lines plus noise) and contributed to the standardization of MPEG-4 High Efficiency AAC, the MPEG-4 Parametric Stereo coding tool, and MPEG Surround. He is a member of the AES and IEEE and has published various papers on low-bit-rate audio coding and related subjects. He enjoys listening to live performances of jazz and free improvised music, and tries to capture them in his concert photography.

●

Jeroen Koppens was born in The Netherlands in 1980. He received an M.Sc. degree in electrical engineering from the Eindhoven University of Technology, The Netherlands, in 2005. He did his graduation project in the Signal Processing Group of Philips Applied Technologies, where he worked on a state-of-the-art psychoacoustic model. After graduation he joined Philips Applied Technologies and contributed to the development of the MPEG Surround standard.

●

Johannes Hilpert received a Dipl.-Ing. degree in electrical engineering from the University of Erlangen-Nürnberg, Germany in 1994.

Upon graduation he joined the Fraunhofer Institute for Integrated Circuits (IIS) in Erlangen, where he worked on perceptual audio measurement and MPEG perceptual audio codecs such as MP3 and AAC. Starting in 2000 he headed the team for real-time audio coding algorithms on digital signal processors and since 2001 he has been in charge of the Audio Coding and Multimedia Software Group. His recent research topics are parametric multichannel and multiobject audio coding.

Mr. Hilpert is a co-editor of the MPEG Surround standard.

Jonas Rödén was born in Östersund, Sweden, in 1975. He received a B.S. degree in electrical engineering from Mid Sweden University in 1998 and an M.Sc. degree from Blekinge Institute of Technology, Sweden, in 2000.

In 2000 he joined Coding Technologies (now Dolby Sweden) in Stockholm, Sweden, where he worked on research, development, and standardization of low-bit-rate audio coding tools such as SBR, parametric stereo, and MPEG Surround. Since 2006 he has been involved in product management with the main focus on MPEG Surround and with continued involvement in application standards.

Werner Oomen received an Ingenieur degree in electronics from the University of Eindhoven, The Netherlands, in 1992.

He joined Philips Research Laboratories in Eindhoven, in the Digital Signal Processing Group in 1992, leading and contributing to a diversity of audio signal processing projects. His main activities are in the field of audio source coding algorithms. Since 1999 he has been with Philips Applied Technologies, Eindhoven, in the Digital Signal Processing Group, where he leads and contributes to different topics related to digital signal processing of audio signals.

Since 1995 Mr. Oomen has been involved with standardization bodies, primarily 3GPP and MPEG, where for the latter he has actively contributed to the standardization of MPEG2-AAC, MPEG4-WB CELP, parametric (stereo) coding, lossless coding of 1-bit oversamples audio, and MPEG Surround.

Karsten Linzmeier received a Dipl.-Ing. degree in electrical engineering from the Friedrich-Alexander University in Erlangen, Germany, in 2000.

Upon graduation he worked for the Fraunhofer Institute for Integrated Circuits (IIS) in Erlangen, where he most recently was involved in the development of novel parametric multichannel audio coding algorithms such as MP3 Surround, MPEG Surround, and spatial audio object coding (SAOC). In 2007 he joined Coding Technologies, today part of Dolby Laboratories, and is working on the improvement of floating-point implementations of various audio codecs.

Kok Seng Chong received B.E. and Ph.D. degrees in electrical and computer system engineering from Monash University, Australia, in 1995 and 1998, respectively.

He is a senior staff engineer at Panasonic Singapore Laboratories, specializing in digital signal processing for audio coding, acoustics, and medical instruments.