MPEG-4 SLS –
Lossless and Near-Lossless Audio Coding Based on MPEG-4 AAC

Ralf Geiger[1], *Markus Schmidt*[1], *Jürgen Herre*[1], *Rongshan Yu*[2]

[1]Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

[2]Institute for Infocomm Research (I^2R), A^*STAR, Singapore

Address correspondence to: ralf.geiger@iis.fraunhofer.de

***Abstract*— MPEG-4 Scalable Lossless Coding (SLS) is one of the latest extensions to the MPEG-4 audio coding standard. It provides a scalable lossless extension of the well-known MPEG-4 AAC codec. The lossless enhancement is performed in a fine-grain scalable way, starting from the AAC core signal in small steps all the way up to a numerically lossless representation, thus also including near-lossless compression scenarios. This enables scalable high definition coding of audio signals.**

The paper provides a brief introduction to the technology of MPEG-4 SLS. In particular, the near-lossless coding capability is examined in detail, and its effectiveness is evaluated in a tandem coding scenario.

I. INTRODUCTION

Modern perceptual audio coding schemes, such as the well-known MPEG-4 Advanced Audio Coding (AAC) codec [1] [2], provide efficient high quality audio coding at low bitrates. Nevertheless, some applications require lossless compression (e.g. for archiving purposes) or additional coding margin for post-processing or tandem coding. The recently developed MPEG-4 Scalable Lossless Coding (SLS) technology provides an extension of MPEG-4 AAC which serves all these needs.

In this paper an overview over MPEG-4 SLS is given, and its operation in *near-lossless* and *lossless* mode is illustrated in some detail. The performance of the system in both modes is evaluated and potential application scenarios are sketched briefly.

II. AAC AND SLS - OVERVIEW

Since the MPEG-4 SLS coder was designed to operate as an enhancement to the MPEG-4 Advanced Audio Coder (AAC), its structure is closely related to that of the underlying AAC core coder. This section briefly describes the architectural features of MPEG-4 AAC as they are relevant to the SLS enhancement technology.

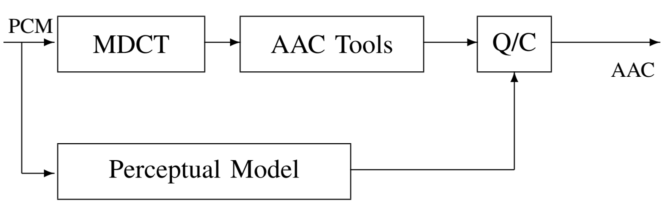

Fig. 1. Basic building blocks of AAC encoder

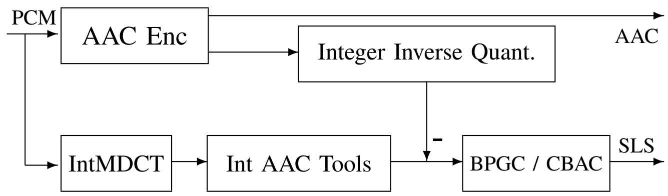

Fig. 2. Basic building blocks of combined AAC/SLS encoder

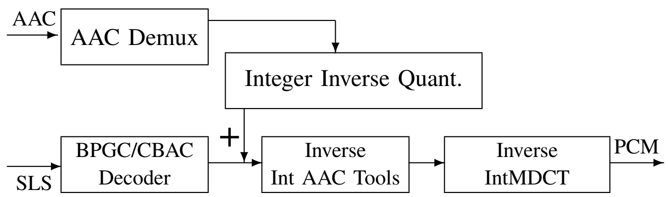

Fig. 3. SLS Decoder

The underlying AAC codec provides efficient perceptual audio coding with high quality and achieves broadcast quality at a bitrate of ca. 64 kbit/s per channel [3]. Figure 1 shows a very condensed view of the AAC encoder structure. The audio signal is handled in a block-wise spectral representation using the *Modified Discrete Cosine Transform* (MDCT) [4]. The resulting 1024 spectral values are quantized and coded considering the required accuracy, as demanded by a perceptual model. This is done to minimize the perceptibility of the introduced quantization distortion by exploiting masking effects. Several neighboring spectral values are grouped into so-called *scalefactor bands* sharing the same *scalefactor* for quantization. Prior to the *quantization/coding* tool, a number of processing tools operate on the spectral coefficients in order to improve coding performance for certain situations, most importantly:

- The *Temporal Noise Shaping* (TNS) tool [5] carries out predictive filtering across frequency in order to achieve a temporal shaping of the quantization noise according to the signal envelope and in this way optimize temporal masking.
- The *M/S Stereo Coding* tool [6] provides sum/difference coding of channel pairs, exploits inter-channel redundancy for near-monophonic signals, and protects from binaural unmasking.

The scalable lossless enhancement of SLS builds on top of this architecture. The general structure of a combined

AAC/SLS encoder is shown in Figure 2.

The audio signal is represented in frequency domain using the *Integer Modified Discrete Cosine Transform* (IntMDCT) [7], [8]. This transform provides an invertible integer approximation of the MDCT and is well-suited for lossless coding in frequency domain. Other AAC coding tools, such as *Mid/Side Coding* or *Temporal Noise Shaping*, are also considered and performed on the IntMDCT spectral coefficients in an invertible integer fashion, thus maintaining the similarity between the spectral values used in the AAC coder and in the lossless enhancement.

The link between the perceptual (upper path in the figure) and the lossless (lower path) part of the coder is then achieved in the following way (see [9], [10]): The perceptually quantized spectral values are inversely quantized, rounded to integer and subtracted from the integer spectral values. The resulting residual values thus provide an integer spectral representation of the AAC quantization error.

If this error is coded and transmitted completely, a lossless reconstruction is possible in the decoder. For an efficient entropy coding of the residual values, dedicated binary arithmetic coding schemes are used, namely *Bit-Plane Golomb Coding* (BPGC) [11] or *Context-Based Arithmetic Coding* (CBAC) [12].

Figure 3 shows the general structure of the corresponding SLS decoder. The AAC values, which were subtracted in the encoder, are added to the decoded residual values. The inverse integer AAC tools are applied before the inverse IntMDCT transforms all values back into time domain.

The lossless enhancement is performed in a fine-grain scalable way, allowing for intermediate near-lossless signal representations. This feature of SLS will be discussed in more detail in the following section.

III. SLS in Near-Lossless Mode

A. Technical Background

The residual spectral values for the enhancement layer are determined based on the quantized and inversely quantized AAC values

$$quant(x) = sgn(x) \cdot \lfloor (|x| \cdot 2^{-0.25 \cdot scalefactor})^{3/4} + 0.4054 \rfloor$$

and

$$invquant(q) = sgn(q) \cdot |q|^{4/3} \cdot 2^{0.25 \cdot scalefactor}$$

by

$$residual[k] = intmdct[k] - \lfloor invquant(quant(mdct[k])) \rfloor$$

where k denotes the frequency line index, and *mdct[k]* and *intmdct[k]* represent the corresponding spectral values, possibly modified by the respective AAC tools. The residual can also be computed based on the lower bound of the AAC quantization interval, resulting in an even higher coding efficiency [13].

SLS has the capability of transmitting a coarse representation of the residual spectral values, i.e. the AAC quantization error, resulting in a near-lossless signal representation. This is achieved by dividing the residual values into bit-planes and coding the values bit by bit along the bit-planes, starting with the most significant bits (MSB) in each scalefactor band.

This approach allows to refine the initial AAC quantization successively. With each additional bit-plane the quantization error is reduced by 6 dB. Consequently, an increasing safety margin with respect to audibility is added as the bitrate is increased.

B. Evaluation of Near-Lossless Audio Quality

While it may seem sufficient for most purposes to provide perceptually transparent reproduction of audio signals by using conventional perceptual audio coders (e.g. AAC at sufficient bitrate), there are applications which demand still higher audio quality. This is especially the case for professional audio production facilities, such as archiving and broadcasting in which audio signals may undergo many cycles of encoding / decoding ("tandem coding") before being delivered to the consumer. This leads to an accumulation of introduced coding distortion and may lead to unacceptable final audio quality, unless substantial headroom towards audibility is provided by each coding step, e.g. by using coding algorithms with very high quality resp. bitrate.

The ITU-R recommendation BS.1548-1 [14] defines requirements for audio coding systems for digital broadcasting, assuming a codec chain consisting of so-called *contribution*, *distribtion*, and *emission* codecs. According to this recommendation, and based on ITU-R BS.1116 [15], audio codecs for contribution and distribution should fulfill the following requirements:

"The quality of sound reproduced after a reference contribution/distribution cascade [...] should be subjectively indistinguishable from the source for most types of audio programme material. Using the triple stimuli double blind with hidden reference test, described in Recommendation ITU-R BS.1116 [...], this requires mean scores generally higher than 4.5 in the impairment 5-grade scale, for listeners at the reference listening position. The worst rated item should not be graded lower than 4."

Furthermore it is noted that: *"The contribution/distribution cascade, when placed in tandem with the emission codec, should not cause a significant reduction in quality compared to the basic audio quality of the emission codec."*

In accordance with these recommendations, tests were run on signals en/decoded with AAC/SLS. The PEAQ measurement [16] provides methods for objective measurements of perceived audio quality. It focuses on applications which are normally assessed in the subjective domain by applying an ITU-R BS.1116 test. The most essential results can be seen in Figures 4, 5 and 6.

The graphs show the *Objective Difference Grade* (ODG) values which have been computed by a PEAQ system. The evaluation procedure consisted of multiple cycles of tandem coding/decoding with up to 16 cycles. The standard set of critical MPEG-4 audio items for perceptual audio coding evaluations was used. An ODG value of 0, -1, -2, -3, -4 means 'indistinguishable', 'perceptible but not

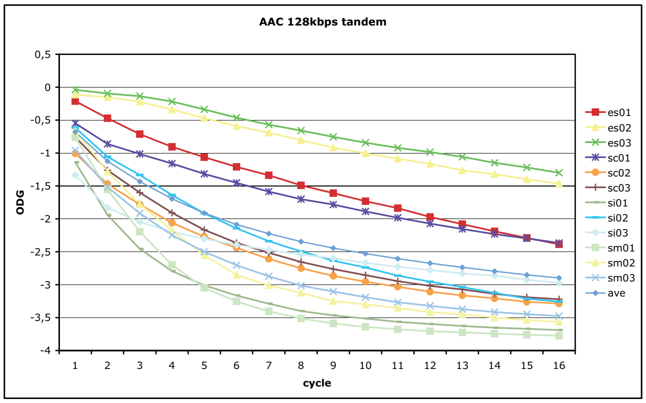

Fig. 4. Test results: AAC in tandem coding

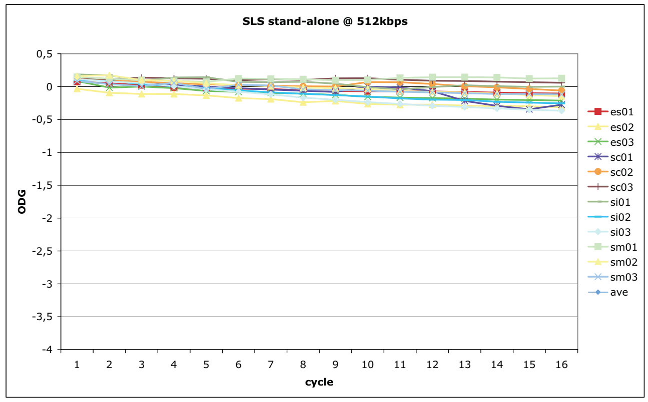

Fig. 6. Test results: SLS stand-alone in tandem coding

annoying', 'slightly annoying', 'annoying', 'very annoying', respectively.

Figure 4 shows the achieved ODG values as a function of tandem cycles for a traditional AAC coder running at a bitrate of 128 kbps/stereo. As expected, it can be observed that the audio quality significantly degrades with increasing number of tandem cycles, depending on the test item. Clearly, tandem coding is not a recommended practice for such coders.

Figure 5 shows the corresponding tandem coding results for the AAC+SLS combination running at 512 kbps/stereo (AAC @ 128 kbps + SLS enhancement @ 384 kbps). It can be observed that the audio quality remains at a very high level, even after a total of 16 tandem cycles. This illustrates the high robustness of such a representation against tandem coding. According to this measurement, the aforementioned audio quality requirement on BS.1548-1 is fulfilled.

Furthermore, when placed in tandem with AAC, the resulting audio quality is not significantly degraded by the SLS tandem cascade. More details can be found in [17].

C. Stand-alone SLS Operation

The SLS codec can also operate as a stand-alone lossless codec when the AAC core codec is not used, referred to as "non-core mode". Despite the simple structure in this mode (only IntMDCT and BPGC/CBAC modules are used), this mode allows efficient lossless coding, see [17]. Furthermore, fine-grain scalability by truncated bit-plane coding is also possible in this mode. Given that the stand-alone SLS codec does not include any perceptual model to estimate masking thresholds, it is interesting to investigate the audio quality resulting from a truncation of the SLS bitstream.

A closer look into the behavior of the bit-plane coding in this mode reveals that a constant SNR per scalefactor band is achieved. With each additional bit-plane the SNR is improved by 6 dB. While this behaviour does not allow to compete with efficient perceptual codecs at low bitrates (e.g. AAC at 128 kbps stereo), this simple approach works quite well at higher bitrates in the near-lossless range.

Figure 6 shows tandem coding results for the stand-alone SLS codec operating at 512 kbps/stereo. It reaches about the same near-lossless audio quality as in the AAC-based mode presented in the previous section.

Further increasing the bitrate towards 768 kbps, most test items still require some truncation in order to guarantee this constant bitrate. Nevertheless, the corresponding PEAQ measurements indicate that both for the AAC-based mode and for the stand-alone mode no degradation of audio quality occurs in this tandem coding scenario, see [17].

This provides an interesting operating point for SLS, corresponding to a guaranteed 2:1 compression. While other stand-alone lossless codecs can also provide an average compression of 2:1 for suitable test material, their peak compression performance can be much worse, depending on the audio material to be encoded. In contrast, SLS is able to guarantee a certain compression ratio while providing lossless or near-lossless signal representation, depending on the input signal.

D. Lossless Coding Performance

The lossless coding performance of SLS is evaluated based on the MPEG-4 lossless coding test set, which includes high resolution audio material. Table I shows the average compression ratio and corresponding bitrates for this test set, both for the AAC/SLS and for the stand-alone operation modes.

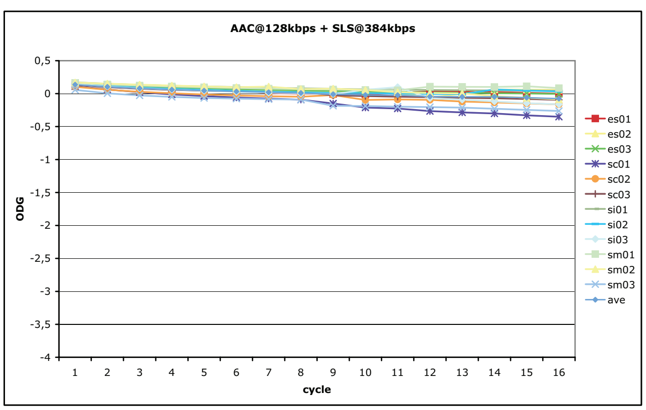

Fig. 5. Test results: AAC + SLS in tandem coding

	SLS + AAC @ 128kbps/stereo (AAC @ 48kHz sampling rate)		SLS stand-alone	
	Compression ratio	Average bitrate	Compression ratio	Average bitrate
48kHz/16bit	2.09	735	2.20	698
48kHz/24bit	1.55	1490	1.58	1454
96kHz/24bit	2.09	2201	2.13	2160
192kHz/24bit	2.60	3543	2.63	3509

TABLE I
Lossless compression results

For this test set a compression of, on average, factor 2:1 can be achieved easily at a sampling rate of 48 kHz and 16 bits word length. It can also be observed that for the AAC-based mode, where AAC is operating at 128kbps, an additional bitrate of only 30 to 40 kbps is required compared to the stand-alone mode for lossless representation. This reduces the bitrate consumption by 90 to 100 kbps compared to simulcast solutions that transmit an AAC bitstream and a stand-alone lossless bitstream simultaneously.

E. Application Scenarios

Potential applications of SLS primarily target the realm of high quality audio processing. The system may excel in:

- *Audio archiving*: Archiving applications can use SLS as a lossless storage format and its scalability to prepare the content for the customers' use.
  A very promising archiving scenario for SLS is described in [18] and [19].
- *Broadcasting*: In a broadcast environment, SLS can be used in all stages comprising storage, contribution/distribution and emission.
- *High quality audio transmission*: As mentioned above, SLS features the ability of high quality constant rate transmission.
- *Digital music distribution*: In combination with AAC, SLS can be flexibly fitted into the full life cycle of digital music distribution.

IV. CONCLUSIONS

The MPEG-4 SLS codec represents a new approach to audio coding at high bitrates and unifies the concepts of perceptual, lossless and scalable near-lossless audio coding into one single powerful coder architecture. Evaluations using objective perceptual measurement confirmed the high degree of robustness of the format to repeated encoding/decoding (tandem coding), both for combined AAC/SLS and for stand-alone SLS operation. SLS is about to become an amendment to ISO/IEC MPEG-4 Audio [1].

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11, "Coding of Audiovisual Objects, Part 3 Audio", International Standard 14496-3

[2] J. Herre, H. Purnhagen, "General Audio Coding", in F. Pereira, T. Ebrahimi (Eds.), *The MPEG-4 Book*, Prentice Hall IMSC Multimedia Series, ISBN 0-13-061621-4, 2002

[3] ISO/IEC JTC1/SC29/WG11, "Report on the MPEG-2 AAC Stereo Verification Tests", MPEG1998/N2006, San Jose, USA, Feb. 1998

[4] J. Princen, A. Johnson, A. Bradley, "Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation", *ICASSP*, 1987, pp. 2161 - 2164

[5] J. Herre, J.D. Johnston, "Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS)", *101st AES Convention*, Preprint 4384, Los Angeles, USA, Nov. 1996

[6] J.D. Johnston, J. Herre, M. Davis, U. Gbur, "MPEG-2 NBC Audio - Stereo and Multichannel Coding Methods", *101st AES Convention*, Preprint 4383, Los Angeles, USA, Nov. 1996

[7] R. Geiger, T. Sporer, J. Koller, K. Brandenburg, "Audio Coding based on Integer Transforms", *111th AES Convention*, New York, USA, Sep. 2001

[8] R. Geiger, Y. Yokotani, G. Schuller, "Improved integer transforms for lossless audio coding", *Proc. of the Asilomar Conf. on Signals, Systems and Computers*, 2003

[9] R. Geiger, J. Herre, J. Koller, K. Brandenburg, "IntMDCT - A link between perceptual and lossless audio coding", *ICASSP*, Orlando, USA, May 2002

[10] R. Yu, R. Geiger, S. Rahardja, J. Herre, L. Xiao, H. Haibin, "MPEG-4 Scalable to Lossless Audio Coding", *117th AES Convention*, San Francisco, USA, Oct. 2004

[11] R. Yu, C.C. Ko, S. Rahardja, X. Lin, "Bit-plane Golomb code for sources with Laplacian distributions", *Proc. ICASSP*, 2003, pp. 277-280

[12] R. Yu, X. Lin, S. Rahardja, C. C. Ko, H. Huang, "Improving Coding Efficiency for MPEG-4 Audio Scalable Lossless Coding", *IEEE 2005 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005)*, Philadelphia, PA, USA, May 2005

[13] R. Yu, X. Lin, S. Rahardja, C.C. Ko, "Advanced Audio Zip - A fine granular scalable to lossless audio coder", to appear in *IEEE Trans. Speech Audio Processing*

[14] Recommendation ITU-R BS.1548-1, "User requirements for audio coding systems for digital broadcasting", International Telecommunications Union, Geneva, Switzerland, 2001-2002

[15] Recommendation ITU-R BS.1116-1, "Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems", International Telecommunications Union, Geneva, Switzerland, 1994

[16] Recommendation ITU-R BS.1387-1, "Method for Objective Measurements of Perceived Audio Quality", International Telecommunications Union, Geneva, Switzerland, 1998

[17] ISO/IEC JTC1/SC29/WG11, "Verification Report on MPEG-4 SLS", MPEG2005/N7687, Nice, France, Oct. 2005

[18] M. Militzer, M. Suchomski, K. Meyer-Wegener, "LLV1: layered lossless video format supporting multimedia servers during realtime delivery", *Proc. SPIE*, Vol. 6015, Oct. 2005, pp. 436-445

[19] M. Suchomski, M. Militzer, K. Meyer-Wegener, "RETAVIC: using meta-data for real-time video encoding in multimedia servers", *Proc. ACM SIGMM NOSSDAV*, Stevenson, Washington, USA, Jun. 2005, pp. 81-86