Audio Engineering Society

Convention Paper 5850

Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Enhancing audio coding efficiency of MPEG Layer-2 with Spectral Band Replication for DigitalRadio (DAB) in a backwards compatible way

Alexander Gröschel, Michael Schug, Michael Beer (all [1]), and Fredrik Henn [2]

[1] *Coding Technologies GmbH, Nürnberg, Germany*

[2] *Coding Technologies AB, Stockholm, Sweden*

Correspondence should be addressed to Alexander Gröschel (gel@codingtechnologies.com)

ABSTRACT

Layer-2+SBR is an audio coding scheme which enhances significantly the coding efficiency of MPEG 1/2 Layer-2 especially for broadcasting applications like DAB (Digital Audio Broadcasting). Spectral Band Replication (SBR) is a technique to enhance the efficiency of perceptual audio codecs. High frequency parts of an audio signal are reconstructed on decoder side, so the audio encoder can focus on coding the low frequency part. Thus, a bitrate reduction can be achieved while maintaining subjective audio quality. Besides increasing the coding efficiency, the use of MPEG Layer-2 + SBR inside DAB would maintain backwards compatibility: Existing DAB receivers are capable of decoding the (bandwidth limited) Layer-2 part of the bitstream. This paper describes the functionality of SBR and Layer-2+SBR and the achievable increase in coding efficiency. Furthermore, applications and introduction scenarios are addressed.

1 INTRODUCTION

Spectral Band Replication (SBR) is a tool to enhance the efficiency of conventional waveform-oriented audio coding schemes. It allows bitrate reductions of about 30% in the medium-to-high audio quality range. SBR was combined with MPEG2 AAC (Advanced Audio Coding) first, showing promising results. Thus, SBR it has been combined with the MPEG 1/2 Layer-2 audio coding scheme also. One possible application is the use of this hybrid codec inside DigitalRadio (Digital Audio Broadcasting - DAB).

2 SBR PRINCIPLE

In traditional perceptual audio coding, quantisation noise is added to the audio signal. Assuming a sufficiently high bitrate, the inserted quantisation noise will be kept under the masking threshold and therefore be inaudible (see figure 1 a). At reduced bitrates, this masking threshold will be violated (see figure 1 b). Coding artifacts become audible.

Thus, if bitrate is restricted, usually audio bandwidth will be limited (see figure 1 c). The result will sound duller, but cleaner.

Whith SBR, exactly this is done: The lower frequency part (from 0 to typically 5..13 kHz) is coded using a waveform coder (called 'core codec'; e.g. one of MPEG audio codec family). Additionally, a reconstruction of the high frequency part happens, done by transposition of the lower frequencies (Spectral Band Replication). Thus, a significant bitrate reduction is achived while maintaining an audio quality sufficient for broadcast applications.

This principle relies on the assumption that usually there exist harmonic relations between the low and high frequency part in natural audio material. In traditional audio coding, a significant amount of bits is spent to code the higher frequency part of an audio signal. However, psychoacoustically, an exact representation of the uppermost 1..2 octaves is of less importance.

The replication is performed like this:

- On encoder side, the high frequency part of the signal is analyzed. A certain amount of side information is generated from this analysis and

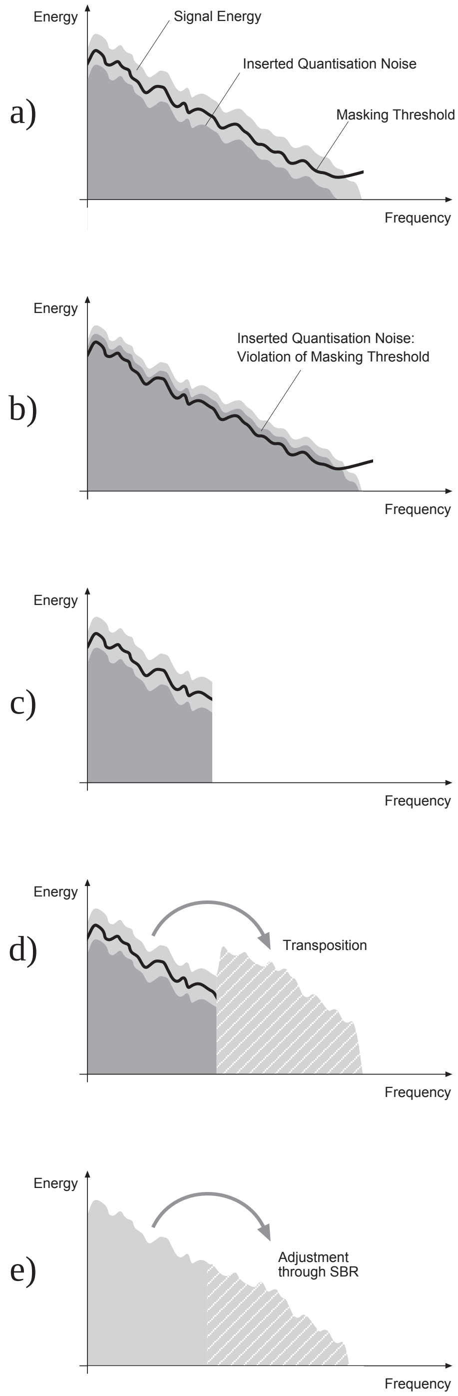

Fig. 1: SBR principle

transmitted with the core data stream, for example as ancilliary data. This side info will take typically 2..3 kBit/s per channel.

- On decoder side, the low frequency components are transposed in the high frequency part (see figure 1 d) and afterwards level adjusted using the transmitted sideinfo (figure 1 e).

A block diagram showing the hybrid coding approach is given with figure 2

The highest frequency covered by the core codec (called 'crossover frequency') can in principle be chosen arbitrary; however, its selection is determined by boundary conditions:

- The core coder may allow only distinct upper frequency borders
- A too low crossover frequency will impair the quality of the transposition result: For proper operation, SBR requires a certain minimum of spectral origin the replication can be done from.
- A very high crossover will work, but the SBR gain will go down, because most of the spectra is coded within the core codec. Assuming limited bitrate, core coder quality will go down.

Thus, a compromise has to be found on bitrate saving, reached SBR quality and remaining bandwidth of the core decoder result. The latter may be important when the application involves backwards compatible operation (i.e. there are non-SBR decoders in use).

The SBR technique can in principle be joined to any traditional audio compression scheme, for example:

- The combination of MPEG-Layer3 and SBR, called mp3PRO[7], was launched in June 2001.
- MPEG4 AAC plus SBR (aacPlus [1]) [2] is currently under standardisation inside MPEG (MPEG-4 High Efficiency AAC Profile).

3 APPLICATION TO DAB

3.1 The current DAB system

The Eureka147/Digital Audio Broadcasting System (DAB) [3] was introduced in the early 90es. One of its goals was to provide audio quality superior to FM ('radio in CD-quality') in audio broadcasting. To achieve this, it made use of MPEG Layer-2, one of the state-of-the-art coding schemes at that time. For only rudimentary error protection mechanisms are defined in MPEG, an additional CRC calculated on the scalefactor data is transmitted in DAB.

Currently, economical reasons seem to force the DAB operators to lower bitrates of single DAB programmes[1]. E.g. in the UK, 128 kBit/s for stereo transmission is widely used already, although MPEG Layer-2 is said to offer 'transparent' (indistinguishable) quality only at 192 kBit/s and above for stereo. Accordingly, there is demand for a more efficient usage of the spectra inside one DAB ensemble, even at the price of giving up 'CD quality transmission'.

One possible solution, the total exchange of MPEG-Layer 2 by a later, more efficient coding scheme, bears drawbacks:

- Old DAB receivers would not work any more
- The performance on error-prone channels is wideley examined for the Layer-2 coding scheme.

Thus, a (at least partly) backwards compatible enhancement of the Layer-2 coding scheme is desirable.

3.2 Backwards Compatibility

A Layer-2+SBR bitstream consists of a (plain) Layer-2 bitstream and additional SBR data. The SBR data is (like the plain Layer-2 data) organized in frames, so the Layer-2 and SBR data can be interleaved in the bitstream. The SBR data can be put in the stuffing area of the DAB bitstream ancilliary data field. So, it will be ignored by any conventional DAB reciever. The conventional DAB XPAD structure is not affected (see figure 3).

Assuming this, compatibility is given the following way:

- Old DAB receivers can decode the plain Layer-2 part of a Layer-2+SBR bitstream. Depending on the crossover frequency, the result will have a more or less limited audio bandwidth.
- New (Layer-2+SBR) receivers will decode both the Layer-2 and the SBR part and therefore pro-

[1] aacPlus is a registered trademark by Coding Technologies

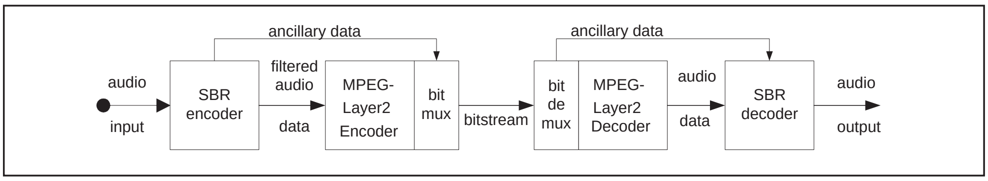

Fig. 2: SBR Encoder/Decoder Chain

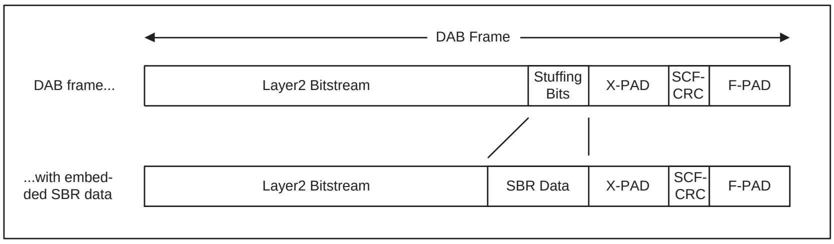

Fig. 3: Bitstream embedding of SBR data in DAB frames

duce full audio bandwidth. Besides, they can decode any plain Layer-2 bitstream.

3.3 Sampling Rates and Audio Bandwidth

DAB allows the use of 48 kHz (MPEG 1; [4]) and 24 kHz (MPEG 2; [5]) audio sampling rates (f_s). This implies two basic modes for the setup of Layer-2+SBR:

- Single Rate
  Both the plain Layer-2 bitstream and the SBR part are run at $f_s = 48\ kHz$
- Dual Rate
  The plain Layer-2 bitstream is coded at $f_s = 24\ kHz$. An SBR capable decoder will do an implicit upsampling to 48 kHz and extend the audio bandwidth beyond 12 kHz.

In general, waveform codecs are slightly more efficient when run at lower sampling frequencies (assumed that the spectral content can be represented there).

Both modes are useful for different purposes:

- In the medium to low bitrange area (below 64 kBit/s per channel), MPEG 1 (in contrary to MPEG 2) only allows very limited bitrate/channel configurations. These are (overall bitrate): 128, 112, 96 and 64 kBit/s for stereo resp. 64, 56, 48 and 32 kBit/s for mono. Thus, Single Rate operation in this bitrate range is only available for these bitrates.
- In addition, for the very low rates (48 and 32 kBit/s per channel) the audio bandwidth of MPEG 1 Layer-2 bitstreams is (determined by the standard) limited to $f_s/8$, i.e. 6 kHz. Therefore, running e.g. 96 kBit/s joint stereo Layer-2+SBR Single Rate will have full audio bandwidth on SBR-capable decoders, but only 6 kHz on Layer-2-only decoders, which may not be acceptable.
- In MPEG 2, there are a lot of bitrate/channel configurations in the medium to low area [2]. This permits a very flexible and fine-granular selection of bitrates.
- All these Dual Rate configurations allow transmission of audio data up to an audio bandwidth of 11.25 kHz only. Thus, the backwards compatible part of Layer-2+SBR Dual Rate will always have this audio bandwidth as maximum, which may be inacceptable.

[2] Allowed MPEG-2 Layer-2 overall bitrates are: 160, 144, 128, 112, 96, 80, 64, 56, 48, 40, 32, 24, 16 and 8 kBit; both stereo and mono

Besides this restrictions, the crossover frequency is reasonably set to a boundary of one of the Layer-2 polyphase filter channels, i.e. a multiple of 375 Hz. (MPEG 2) resp. 750 Hz (MPEG 1).

3.4 Stereo Modes

For efficient stereo coding, in MPEG the 'Joint Stereo (JS)' method is used. Herein, the stereo signal is approximated by one center signal waveform and separate level information for left and right channel. In Layer-2, this applies above an adjustable border (JS border) [3] This border will be put lower when using low bitrates, because the need for efficient coding raises. Joint Stereo performs good for natural stereo recordings, but can fail on artificial stereo signals, where correlation between the two channels is low.
Dependent on the crossover frequency, Joint Stereo usage is less or not necessary any more, because the upper frequency range is covered by SBR. Thus, Layer-2+SBR will provide a more stable stereo image than plain Layer-2 using JS heavily.

3.5 Bitrates and Listening Test Results

For audio quality evaluation, a subjective listening test was performed comparing plain Layer-2 to Layer-2+SBR at several bitrates. This test was carried out at the Institut für Rundfunktechnik, München [4] The Layer-2+SBR under test was Dual Rate, with the crossover frequency optimized for best SBR performance (i.e. not optimized for high backward compatibility). The following configurations were under test:

Mono:

	bitrate (kBit)	f_s (kHz)	SBR
48_hm	48	24	no
48_m_SBR	48	48	yes
64_hm	64	24	no
64_m_SBR	64	48	yes
80_hm	80	24	no
96_hm	96	24	no

Stereo:

	bitrate (kBit)	f_s (kHz)	stereo mode	SBR
80_js_sbr	80	48	JS	yes
96_hjs	96	24	JS	no
96_js_sbr	96	48	JS	yes
112_js_sbr	112	48	JS	yes
128_js	128	48	JS	no
128_st_sbr	128	48	S	yes
160_js	160	48	JS	no
160_st_sbr	160	48	S	yes
192_js	192	48	JS	no

A MUSHRA [6] test was performed on a set of 6 different test items (speech, music and mixed items). 17 listeners took part in the test. The results are displayed in figure 4 (Mono) and figure 5 (Stereo). It can be concluded that

- for mono, 48 kBit/s Layer-2+SBR performs as good as 80 kBit/s plain Layer-2
- 64 kbit/s Layer-2+SBR performs as good as 96 kbit plain Layer-2
- for stereo, 96 kbit/s Layer-2+SBR almost reaches the quality of 128 kbit/s plain Layer-2 JS.
- 128 kbit/s Layer-2+SBR almost reaches the quality of 160 kbit/s plain Layer-2 Stereo
- 160 kBit/s Layer-2+SBR is rated better than 192 kBit/s plain Layer-2 JS. This probably results from the fact that Joint Stereo coding is used for the latter.

Layer-2+SBR is able to maintain audio quality while reducing bitrate significantly compared to plain Layer-2.

3.6 Error Sensitivity

Inside the Layer-2+SBR bitstream, the largest part (representing the low frequency range) still is plain Layer-2 data. Therefore, here, no change in error sensitivity is to be expected.
On the SBR data itself a 10bit Cyclic Redundancy Check is calculated; therefore the probability of recognizing disturbed SBR frames is sufficiently high.

[3] possible JS bordersin DAB are: 3, 6, 9, 12 kHz (MPEG 1); 1.5, 3, 4.5, 6 kHz (MPEG 2)

[4] This test was done in the course of a diploma thesis by Christian Kain. This thesis was not yet available to the public at current point of time.

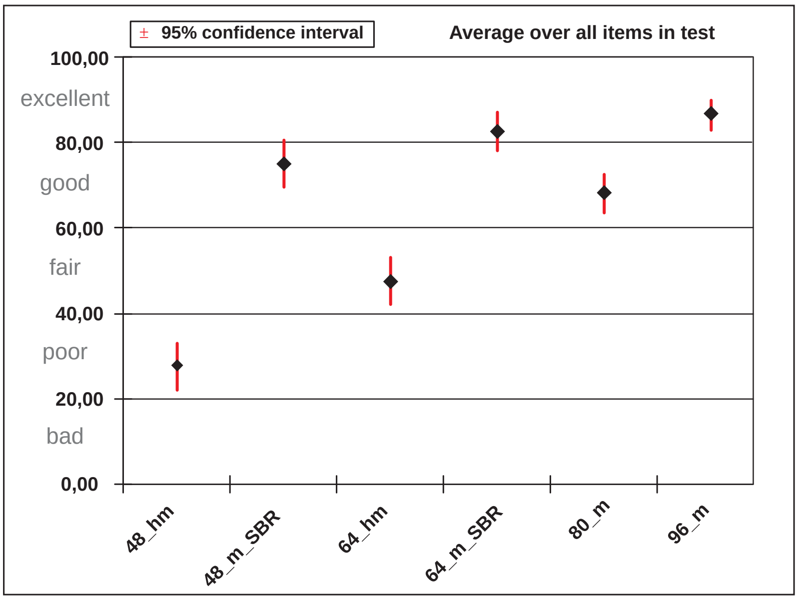

Fig. 4: Listening test mono

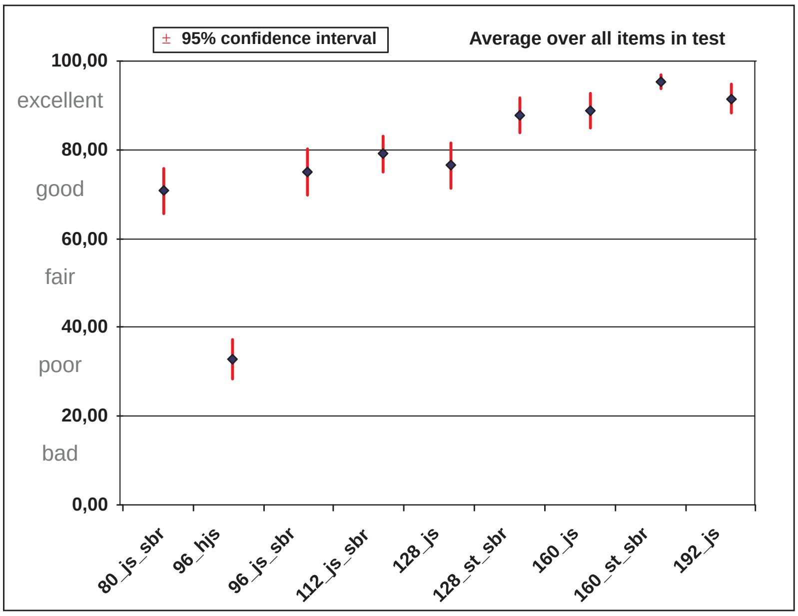

Fig. 5: listening test stereo

In that case, the high freequency part of the signal can be muted. For the low frequency part still persists, impairment of overall audio quality will be moderate.

3.7 Transition Scenario

The introduction of Layer-2+SBR in DAB can be done gradually:

- Layer-2+SBR programmes can be mixed with plain Layer-2 programmes inside one ensemble
- The crossover frequency can be chosen depending on the need to be backwards compatible (i.e. on the portion of 'old' non-SBR receivers in the market).

It may be desirable to use Layer-2+SBR Single Rate in the beginning (high core bandwidth), switch over to Dual Rate (high efficiency) at a later point in time and gradually reduce the crossover frequency further.

4 FUTURE WORK

- An implementation of a Single Rate system and corresponding listening tests have to be performed.
- The assumptions about error sensitivity of Layer-2+SBR has to be proven in real-world tests. This has started already (in cooperation with IRT), and results will be available probably in 2003.
- Work on the audio quality of the Layer-2+SBR system is still ongoing. Even though it has been proven that maintaining audio quality at reduced bitrates by means of SBR, there still is room for improvement, and further bitrate savings can be expected.

5 CONCLUSION

With Layer-2+SBR, a hybrid scheme has been presented that can increase the efficiency of the audio coding in DAB. Bitrate saving without losing audio quality allows more programmes to be broadcast in the same frequency range.

Technical background on the technology of Spectral Band Replication as well as listening test resuls have been presented. It has been shown how Layer-2+SBR can be introduced step-by-step without the need to exchange all equipment at once. In particular, the aspect of backwards compatible operation has been demonstrated.

6 REFERENCES

[1] http://www.wohnort.demon.co.uk/DAB/index.html, private DAB info site.

[2] Andreas Ehret, Martin Dietz, and Kristofer Kjörling. State-of-the-Art Audio Coding for Broadcasting and Mobile Applications. In *114th AES Convention, Amsterdam, March 22 - 25*, 2003.

[3] ETSI EN 300 401 V1.3.3, Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers, May 2001.

[4] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 11172-3 "Information technology – Coding of moving pictures and associated audio for digital storage media up to about 1,5 Mbit/s", Part 3: Audio, 1991.

[5] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 13818-3 "Information technology – Generic Coding of Moving Pictures and Associated Audio:", Part 3: Audio, 1994.

[6] ITU-R BS.1534, ITU Recommendation: Method for the subjective assessment of intermediate quality level of coding systems, June 2001.

[7] Thomas Ziegler, Andreas Ehret, Per Ekstrand, and Manfred Lutzky. Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm. In *112th AES Convention, Munich, May 10 - 13*, 2002. Preprint 5560.