Audio Coding Technology of ExAC

A. Ehret[1], X.D.Pan[2], M. Schug[1], H. Hoerich[1], W.M. Ren[3], X.M. Zhu[3] and F. Henn[4]

1:Coding Technologies GmbH, Nuremberg, Germany; 2:Beijing Media Works Co., Ltd, Beijing, China;
3:Beijing Eworld Tech. Co., Ltd, Beijing, China; 4:Coding Technologies AB, Stockholm, Sweden

ABSTRACT

A new low bitrate audio coding technology (further denoted as "ExAC") based on Enhanced Audio Coding (EAC) and Spectral Band Replication (SBR) is introduced. The major building blocks of the coding schemes are explained, in which EAC works as a core coder and SBR works as a powerful bandwidth extension module. The new coding technology provides a high quality audio compression scheme for a broad range of applications, including the high-density laser video diskette, HDTV and very low bitrate applications such as AM audio broadcasting and streaming.

1. INTRODUCTION

In recent years, digital audio has become an important source of information in the modern world of information systems. In linear representation, digital audio files require a lot of memory or bandwidth for transmission respectively. Many research efforts have been devoted to the problem of audio compression in the last two decades. Two different compression categories have been of particular interest: high performance and low bitrate audio coding. High performance audio coding is aimed to achieve the audio quality as high as possible at a certain bitrate. Applications requiring this type of compression include high-density laser video diskette and HDTV. Vice versa, for applications such as streaming or audio broadcasting, audio coding at lowest possible bitrate whereas maintaining a reasonable audio quality is of primary interest.

In this paper, a new audio codec named ExAC is introduced. ExAC is based on the existing audio coding technologies of EAC and SBR. In addition new tools are being applied to guarantee highest possible audio quality, even at very low bitrates and/or very low sampling rates. ExAC is being developed to fulfill the requirements of different applications, with both high performance and low bitrate audio coding characters. The codec has been proposed to the China Audio and Video Standard (AVS) Working Group and Enhanced Video Diskette (EVD) standard, and it has been selected as the audio coding standard of EVD.

The following part of the article explains the basic blocks of ExAC, and some test results are presented.

2. OVERVIEW of ExAC

The figure 1 illustrates the general structure of the ExAC Encoder, the encoder includes several components: Downsampler, EAC core encoder, SBR encoder and Multiplexer.

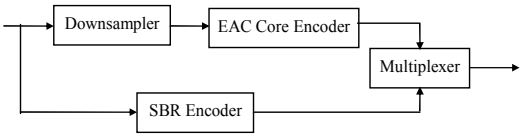

Fig1. Diagram of the ExAC Encoder

The ExAC Decoder is depicted in Fig. 2. It includes several components: deMultiplexer, EAC core decoder and SBR decoder (with implicit upsampling).

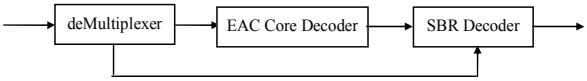

Fig2. Diagram of the ExAC Decoder

3. EAC CORE

EAC is an audio coding technology that has been developed by Beijing E-World Ltd. Co. in which a multi-resolution tiling of the T/F plane, a quantization algorithm to minimize the global perceptual distortion and entropy coding are used to compress the audio signal by utilizing the redundancy as well as the irrelevancy. The EAC codec supports mono, stereo and 5.1 surround stereo encoding and decoding modes, EAC has already been accepted as the audio codec for the EVD (Enhanded Video Diskette) system.

EAC works as the core coder of ExAC at half of the nominal sampling rate. It is composed of a 2 : 1 Downsampler, a Time to Frequency Mapping module and a Quantizer, as well as a Psychoacoustic Model. In the ExAC encoder, EAC encodes the low frequency components of the audio signal, and the result will be transmitted to the Multiplexer.

A typical EAC encoder is illustrated as Fig.3. The Signal-type Detection analyzes the input audio signal, and the audio frames are classified and labeled as either stationary-like or transient-like. When a transient like frame is encoded, the codec should be adjusted to avoid perceptible pre-echoes. In the current EAC core, FLPVQ

(Frequency domain Linear Prediction Vector Quantization) and MR (Multi-Resolution Analysis) have a positive effect in mitigating pre-echoes. The basic idea behind FLPVQ is the linear prediction of the spectrum could further improve the time resolution efficiently for a kind of transient-like signal. On the other hand, the EAC core tunes the time-frequency resolution of the encoded signal by employing multi-resolution analysis on the frequency coefficients. The significant vectors in the time-frequency plane are quantized and coded with a Vector Quantizer to improve the coding efficiency.

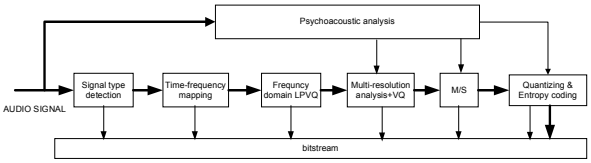

Fig3 The diagram of EAC encoder

In order to reduce redundancy of the multi-channel audio signal, M/S is implemented in which the sum and difference of highly correlated channels are coded rather than the original channels.

In the module of Quantizing and Entropy coding, the coefficients are divided into a set of scale factor bands, then a non-uniform scalar quantizer quantizes the coefficients of each band. To improve the coding gain, the Huffman coding method is implemented in this module. A bit allocation loop is introduced to distribute the budgeted bits into scale factor bands when quantizing the coefficients.

For a more elaborate description of EAC, please refer to the technical proposal to China Audio and Video Standard working group [1].

4. SBR MODULE

A. Principle

The principle of SBR is based on the fact that the high frequencies of an audio signal can be extrapolated from the low frequencies, whereas the reconstruction by means of transposition results in a coding of the high frequency portion with very low overhead.

Apart from the pure transposition (see Fig 4a) the reconstruction of the highband is further improved by transmitting guiding information such as the spectral envelope of the original input signal or additional info to compensate for potentially missing high frequency components (see Fig 4b). This guiding information is further referred to as SBR data. Of course, means must be taken to code the SBR data as efficient as possible to achieve a low overhead data rate.

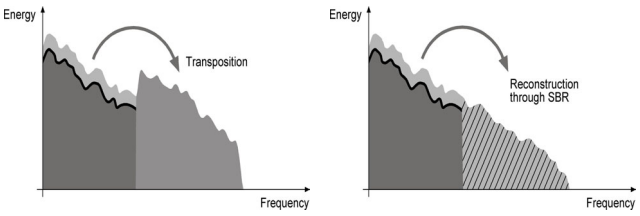

Fig 4 The principle of SBR – Transposition (a) and Reconstruction (b)

At encoder side the original input signal is analyzed at the full nominal sampling rate, the highband spectral envelope and its characteristics in relation to the lowband are encoded and the resulting SBR data is multiplexed with the EAC bitstream. At decoder side, first the SBR data is de-multiplexed, then the EAC decoder is being run separately and the SBR decoder operates on the time domain output signal at half the nominal sampling rate. The decoded SBR data is used to guide the spectral band replication process. During the transposition and reconstruction process an implicit upsampling by a factor of 2 is applied, such that a full bandwidth output signal at the nominal sampling rate is obtained.

Whereas the basic approach seems to be simple, making it work reasonably well is not. Obviously it is a non-trivial task to code the “guidance information” in a way that all of the following criteria are met:

1) good spectral resolution is required
2) sufficient time resolution on transients is needed to avoid pre-echoes
3) cases with poorly correlated lowband and highband need to be taken care for since here transposition and envelope adjustment alone could sound artificial
4) a low overhead data rate is required in order to achieve a significant coding gain

Solutions to the design criteria above as well as the major system parameters are described in more detail below

B. Sampling rate and Crossover Frequency

Generally the lowband needs to cover the frequency range from DC up to around 4 to 12 kHz, depending on the target bitrate and used sampling rate: the higher the crossover frequency between EAC and SBR, the higher the needed bitrate to fulfill the psychoacoustic masking threshold of the EAC encoder. Typical configurations are 16/32 kHz, 22.05/44.1 kHz or 24/48 kHz sampling rates, whereas 8/16 or even 48/96 kHz are also possible. The resulting audio bandwidth can be configured flexibly and may also depend on the application or audio content type. The following table shows typical examples on choices of the crossover frequency between EAC and SBR as well as the audio bandwidth at a number of bitrates, using 24/48 kHz as nominal sampling rate, mono.

Bitrate, mono [bit/s]	EAC freq. range [Hz]	SBR freq. range [Hz]
24000	0-4000	4000-12000
32000	0-6000	6000-16000
48000	0-8000	8000-18000
64000	0-10000	10000-23500

C. Spectral and Time resolution

The most important part of the SBR data is the information for describing the spectral envelope of the original input signal. The core algorithm of SBR consists of a 64-band, complex valued polyphase filterbank (QMF). Its main design goal is, to let it be used as an equalizer without introducing annoying aliasing artifacts, providing good spectral and time resolution. A more elaborate description of the filterbank can be found in [2].
At encoder side an analysis QMF is used to obtain energy samples of the original input signal's highband, which are used as reference values for the envelope adjustment at decoder side. At a sampling frequency of 48 kHz, the theoretical maximum spectral resolution for envelope adjustment is given by

$$f_{\min} = \frac{fs/2}{numBands} = \frac{24000Hz}{64} = 375Hz$$

In order to keep the overhead low, the bitstream format of ExAC allows to group the QMF bands into scalefactor bands. By using a Bark scale oriented approach, grouping of frequency bands may result in wider scalefactor bands the higher the frequency gets, without compromising on audio quality. Furthermore, the envelope update rate in time can also be adjusted according to the audio signal characteristics.
Typically, a high spectral resolution but less time resolution is chosen for stationary like signals: the energy samples are averaged over a longer period of time but a good spectral resolution by not grouping too many bands is achieved, e.g. the energy samples are averaged over half a SBR frame, resulting in 21 ms at 48 kHz; an example SBR frequency range of 24 QMF bands would be grouped into 12 scalefactor bands. Vice versa, for percussive or transient like type of audio signals a higher time resolution is chosen but with less spectral resolution by grouping at least 2 and up to 5 bands in the higher frequency range, e.g. the energy samples are averaged only over 5 ms, the 24 QMF bands would be grouped into 7 scalefactor bands.
Further, differential coding between the QMF samples in time and frequency direction as well as Huffman coding is used. The coding scheme is held flexibly such that the trade-off between fine granularity and low overhead can be balanced according to the total target bitrate. The following table shows typical SBR data rates for a number of example configurations:

Bitrate, mono [bit/s]	SBR freq. range [num QMF bands]	SBR data rate [kbit/s]
24000	21	~1.2
32000	25	~2.0
48000	27	~2.5
64000	32	~3.5

D. Compensation Methods for poorly correlated High- and Lowband

Whereas satisfactory results for a lot of audio material can already be achieved with the transposition and spectral envelope adjustment, it tends to fail when the highband and the lowband of the original audio input signal are poorly correlated.
Examples for such cases are sounds, where the lowband contains a strong harmonic structure like with voices or tonal instruments, e.g. a saxophone or an organ, whereas the highband has a more noise-like character as with cymbals or high-hats. In such a scenario transposition plus envelope adjustment would lead to modulating the high-hat sounds with the lowband's pitch, resulting in an unnatural, annoying sound. Vice versa, the highband could contain harmonics or a single tone where a matching base tone in the lowband is missing, like it may happen with instruments like triangles, glockenspiel or synthetic sounds. In such cases narrow band noise would be transposed to bands where a harmonic structure should have been, again resulting in an unnatural, annoying sound.
With SBR there are several means to cope well with such audio material:

1) *Noise floor estimation*: the noise floor of the lowband and the highband are analyzed inthe encoder. If the transposed noise floor of the highband would not match the original input signal noise floor, additional noise may be added at the SBR decoder.
2) *Inverse filtering*: if the harmonic or noise like characteristic of the transposed highband would not match the original input signals characteristic, inverse filtering may be applied to attenuate the transposed highband harmonic structure towards the desired more noise-like character.
3) *Sine synthesis*: if a specific tone in the transposed highband is missing since no base tone in the lowband has been present which could have been transposed, it is possible to synthesize a sine tone within a QMF band at decoder side.

At encoder side detectors for all those tools are available in order to analyze the original input signal characteristics and create the SBR data to guide the SBR decoding process.

5. TESTING RESULT

As a part of the EVD standardization process, a listening test of ExAC codec has been carried out by China National Testing and Inspection Center for Radio and TV Products in December 2004. The testing scenarios were stereo coding at 128kbit/s and 5.1 channel coding at 384kbit/s. In the test, the "double-blind triple-stimulus with hidden reference" method as described by the ITU-R BS.1116-1[3] was used, in which the subject selects one of three stimuli ("A", "B", "C") at his/her discretion. The known reference is always available as stimulus "A". The hidden reference and the object are simultaneously available but are "randomly" assigned to "B" and "C", depending on the trial.

In the test, 15 items were chosen from the available test items, where 8 items were stereo, and 7 items were 5.1 channel. The 35 involved subjects, aged from 20 to 65, are all majoring in the audio technologies or working in the audio industry for a pretty long period, most of them have been involved in listening test more than once. The listener has to evaluate the quality difference by the following scale:

Impairment	**Grade**
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very annoying	1.0

According to the testing report, at the testing configurations, the codec provide an almost transparent audio quality [4]. The listening test result can be showed as Fig 5 and Fig 6, where the values represented the so-called Subjective Difference Grade (SDG) of the quality rating. According to [5], the SDG is defined as:

$$SDG = Grade_{SignalUnderTest} - Grade_{Reference\ Signal}$$

And the SDG values should ideally range from 0 to -4, where 0 corresponds to an imperceptible impairment and -4 to an impairment judged as very annoying.

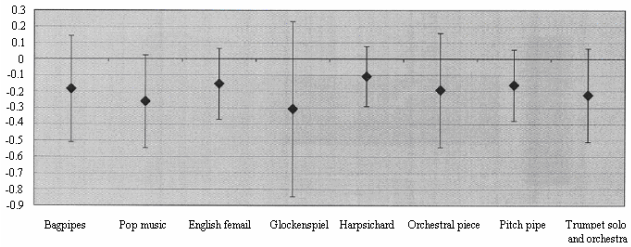

Fig 5 Test Results of 128 kbit/s, stereo

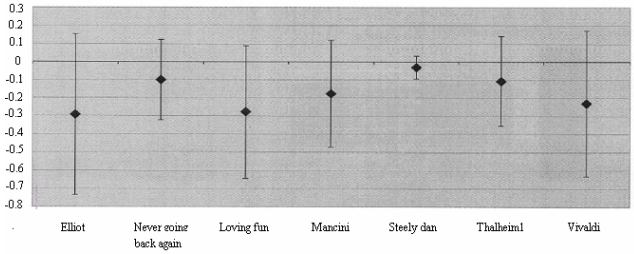

Fig 6 Test Results of 384 kbit/s, 5.1 channel

6. CONCLUSION

In this paper we presented a new audio codec based on the two existing audio coding technologies EAC and SBR. The building blocks and the basic principles of the codec are introduced. As expected, the listening tests show that the existing implementation of ExAC achieves an almost transparent audio quality. Ongoing work on the ExAC encoding scheme will even increase its performance further in the future.

7. REFERENCES

[1] Beijing E-world Technology Co., Ltd. EAC Audio Coding Technology，AVS audio proposal M1005，2002，8

[2] Per Ekstrand. Bandwidth Extension of Audio Signals by Spectral Band Replication. In *IEEE Benelux Workship on Model based Processing and Coding of Audio (MPCA-2002), Leuven Belgium, Nov 15*, 2002.

[3] ITU-R BS.1116-1, ITU Recommendation: Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems, 1994

[4] China National Testing and Inspection Center for Radio and TV Products, Listening Test Report of ExAC Audio Codec, Dec. 2003

[5] ITU-R BS.1387-1, ITU Recommendation: Method for Objective Measurements of Perceived Audio Quality, 2001