ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 2, August 2013

Low Delay SBR Filter Banks for MPEG-4 AAC-ELD

S.G. Dighe, S. S. Gundal

Abstract – MPEG's low-delay audio codecs AAC-LD and AAC-ELD are capable of providing natural audio in perceptually high quality with an algorithmic system delay1 in the range from 20 to 46.9 mS. The AAC-ELD scheme, in particular, which is basically formed out of AAC-LD and SBR, produces high-quality wideband audio at bit rates between 24 kbps and 48 kbps maintaining a relatively low algorithmic delay of 31.3 ms, where 1.3 ms are due to SBR, at 48 kHz sampling rate. These are three crucial factors for AAC-ELD to conquer the telecommunications sector.

***Keywords*–Advanced audio coding(AAC),Discrete Cosine Transform (DCT); DCT-IV; factorization; fast algorithms; filter banks; low delay audio coding; Moving Picture Expert Group(MPEG); Spectral Band Replication (SBR)**

I. INTRODUCTION

The basic MPEG-4 AAC profile is the "AAC Profile", which is commonly referred to as AAC-LC (low complexity). The most prominent application of this profile is Apple iTunes. The combination of AAC-LC with the SBR tool results in the "High Efficiency AAC profile" (HE-AAC). The SBR tool especially extends the coding efficiency for low bitrates. For a further increase of coding efficiency, HE-AAC can be combined with the PS tool to form the "High Efficiency AAC v2 Profile" (HE-AAC v2). By generating a new profile through adding new tools to an underlying profile, an encoder has all tools available to generate bit streams for the underlying profiles. The same holds true for the decoder. The "Low Delay AAC Profile" (AAC-LD) and "Enhanced Low Delay AAC Profile" (AAC-ELD) were developed for communication applications, and also for some types of contribution links. With a coding delay down to 15 ms for AAC-ELD, these codecs still offer remarkable bit rate efficiency. For highest quality and archival purposes, the MPEG-4 "High Definition AAC Profile" (HD-AAC) offers scalable, lossless audio compression. The following sections provide brief descriptions of the codecs of the AAC family that are most relevant to broadcast and cable television applications. AAC-LD, the low delay version of AAC, combines the full bandwidth and superior quality of AAC with the low coding delay that is necessary for two-way audio communication. AAC-LD features an algorithmic delay of only 20 ms, while offering CD-like audio quality at 64 Kbit/s per channel. By integrating SBR technology with the feature set of the LD codec, AAC-ELD provides full audio bandwidth at data-rates down to 24 Kbit/s per channel. Both, the AAC-LD and AAC-ELD codecs are perfectly suited for bi-directional communication applications such as VoIP, Telepresence and video conferencing.

II. MPEG-4 SPECTRAL BAND REPLICATION:-

Block diagram:-

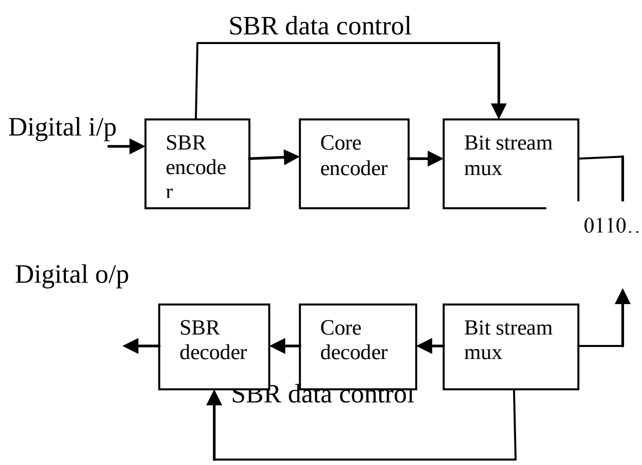

Fig 1: SBR as an add-on to a core codec

Spectral Band Replication is a hybrid1 high-quality bandwidth extension technique for speech and natural audio. It is an add-on to a conventional waveform coder, referred to as the core coder, rather than an audio coder itself. SBR can be seen as a pre-processor to the core encoder and as a postprocessor to the core decoder as depicted in Fig.2.1 In the pre processing step signal characteristics are analyzed and a moderate amount of SBR specific data –usually a small fraction of the overall data rate– is stored, which is then used in the post processing step to reconstruct the broadband signal. The core encoder codecs the low-frequency portion, alias the low band of the original audio signal up to a chosen cutoff frequency. This frequency is labeled the crossover frequency between the low-frequency band and the high-frequency band, or shorts the high band.

A. Encoding process:-

The SBR post processor reconstructs the high band from the decoded low band in a perceptually accurate manner, forming a wideband output signal. In the general case the core encoder operates at half the sampling frequency of SBR, resulting in a better frequency resolution of the core encoder's filter bank, which is beneficial for the exploitation of simultaneous-masking effects. This and the fact that only the low band needs to be redundancy coded boosts the coding gain of the entire system. The driving idea behind SBR is the assumption that the low- and the high band characteristics of an audio signal are strongly correlated with each other. A signal with a distinct harmonic character in the low-frequency region is assumed to basically maintain its harmonic structure in the high-frequency region. Similarly, a noisy signal is assumed to carry its noise-like character from the low frequencies over to the high frequencies. SBR also incorporates additional tools like inverse filtering, adaptive noise addition, and sinusoidal regeneration for signals that do not fit this simple model. The necessity to track partials, which is essential to sinusoidal modeling, spots SBR closely related to parametric methods. Moreover, due to the method of high band regeneration, the short-term synchronization of the high band with the low band, i.e. the temporal alignment, is close to optimal. A transient in the low band translates almost perfectly to the high band. A partial in the high band persists in time as long as the corresponding fundamental frequency is found in the low band. Therefore, following Parseval's theorem,

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|X\left(e^{j\Omega}\right)\right|^2 d\Omega,$$

The output of the filter bank is used to estimate the spectral envelope for every incoming data vector. As the core encoder operates at a different sampling rate from the SBR encoder, the whole encoding system is termed to operate in dual rate mode. Apart from the envelope estimation, the sub-band samples undergo further analysis stages: Control parameter extraction –not necessarily in that order– includes signal adaptive temporal subsample grouping, collecting guidance information for the high frequency reconstruction (HFR) in the decoder, determining the tonal-to-noise ratio in the high band, and the detection of missing harmonics which cannot be reconstructed by merely shifting the low band towards higher frequencies. The spectral envelope coefficients together with the control parameters form the SBR data stream, which is entropy coded. If a constant overall bit rate is desired, the number of bits spent on SBR encoding is to be signaled to the underlying core encoder. One can even think of subsystems which exchange information in such a way that the optimum crossover frequency between the waveform coded low band and the SBR coded high band can be found adaptively for every processed time segment. Finally, the output of the SBR encoder is serially multiplexed with the output from the core encoder into one concatenated bit stream.

B. Decoding process:-

The down sampled low band signal is provided by the embedded core decoder. The high-frequency generation and envelope adjustment is performed upon the sub-band samples of the filter bank decomposed low band data using the SBR control information. The control data further indicates the amount of adaptive filtering that is to be applied to the transposed low band signal to preserve the spectral characteristics of the original high band. The patched and filtered low band is envelope adjusted according to the transmitted time-frequency grid that was selected on the encoder side. If required, ancillary signal components are added to compensate for a weak correlation between the low- and high band characteristics of the audio material with respect to noise content and dominant harmonics. Finally, the implicitly up sampled low band and the replicated high band are combined and synthesized to form the wideband PCM output signal.

C. Delay Calculation:-

Delay reduction is obtained by zeroing out parts of the window that overlap with future input samples. For example, if the frame size is 480 samples, the length of the analysis window is 1920 samples (overlap of 4 frames), the last 120 samples of which are zeros. Similarly, the first 120 samples of the synthesis window are zeros. Thus, the delay of the analysis-synthesis filter bank chain is reduced from 480 + 480 = 960 samples to (480 – 120) + (480 – 120) = 720 samples.

III. RESULTS

A. Algorithmic delay:-

The estimation of the algorithmic delay is conducted based on the block diagram of the SBR superstructure, which is depicted in Fig.2 In this schematic every block along the signal or data path, which contributes to the delay of the system is marked with the respective delay value counted in time samples.

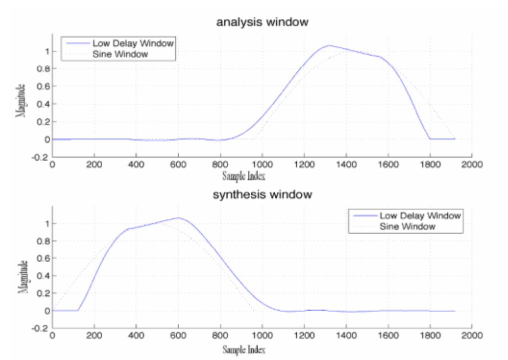

Fig 2: Analysis and synthesis windows used in MPEG-4 AAC ELD core coder filter banks.

The overall system delay is therefore given by the sum of all delay values, i. e., the delay of the down sampling filter v1, the delay of the core codec chain v2 = v21 + v22, the delay of the CPQMF bank v3 = v31 + v32, and finally the delay entailed by the high-frequency generation algorithm v4. The output from the synthesis filter bank

$$y(n) = \hat{x}(n - \nu),$$

is thus the reconstructed version ˆx(n) of the wideband input signal x(n), delayed by samples.

$$\nu = \sum_{i=1}^{4} \nu_i$$

The delay of the SBR extension is for the larger part caused by the L-length periodic version of the prototype filter function P0(n). Not including the blocking delay of the M-length transformation core, the delay of the CPQMF bank is determined by the number of samples that the prototype window overlaps towards the future.3 Hence, due to the type-I symmetry of the prototype filter p0(n), both the analysis and the synthesis jointly introduce a delay of L − M samples.

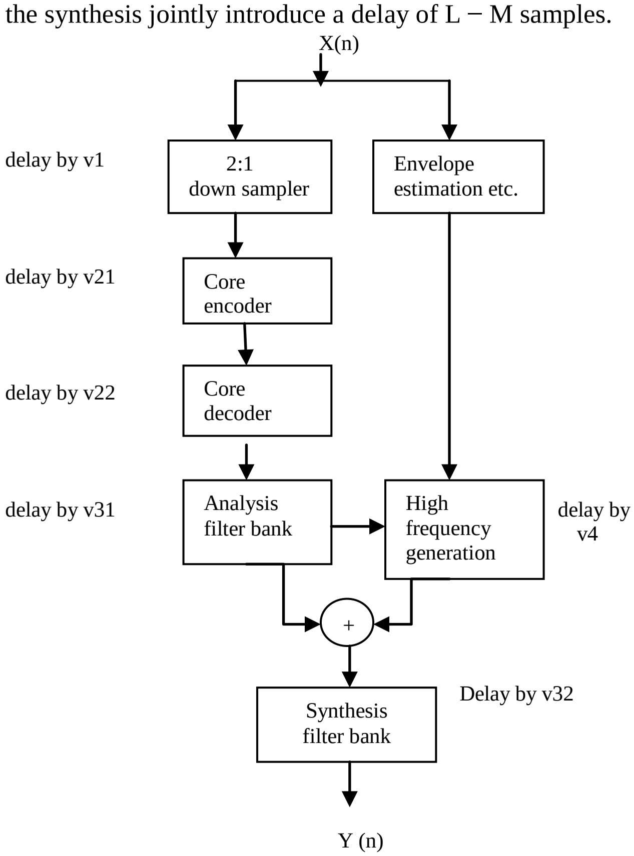

Fig 3: Algorithmic delay along the signal path.

The delay of the HFR unit is determined by the size of the look-ahead buffer, which is a quarter of the block size N.4 Presuming that the down sampling filter has a low order, and hence has a negligible delay, the total SBR delay in milliseconds amounts to

$$\tau = \frac{1}{f_s}\left(L - M + \frac{N}{4}\right),$$

Where, fs are the sampling frequency.

Sampling Rate	**SBR delay**	**Optimized SBR delay**
32 KHZ	34 m S	30 m S
44.1KHZ	24.7 m S	21.8 m S
48KHZ	22.7 m S	20 m S

Table 1: SBR delay for L = 640, M = 64, and N = 2048.

IV. APPLICATION

1. iTunes.
2. YouTube.
3. Sony PS3, Nintendo Wi-Fi.

V. CONCLUSION

Since, we had studied the fast algorithms for MPEG-4 AAC-ELD using SBR filter banks. Low-Delay Spectral Band Replication (LD-SBR) is a derivative of the standard SBR tool, which as a bandwidth extension coder complements the waveform coding low-delay core, forming a low-delay codec for high-quality bidirectional communication at low bit rates. Its algorithmic delay is reduced to a minimum, such that the overall system delay does not significantly exceed the algorithmic delay of widely used speech codecs, which is round about 20 m S.

VI. ACKNOWLEDGEMENT

This project started in june, 2012 and was completed in July, 2013. The goal of this project is to design "Low delay SBR Filter banks using MPEG-4 AAC-ELD". The function has been realized successfully. I want to give my whole sincere to my supervisor and grateful appreciation to Prof. S. S. Gundal, as my supervisor of dissertation work; she tried her best to help me. Without her help and guidance I cannot bring the theories into practice. On the other hand, I want to thank all my family members and friends for their always support and spiritual motivation.
Thank you very much!

REFERENCES

[1] Karlheinz Brandenburg, Oliver Kunz, and Akihiko Sugiyama."MPEG-4 Natural Audio Coding. Signal Processing: Image Communication", pages 423444, January 2000.

[2] Eric Allamanche, Ralf Geiger, Jurgen Herre, and Thomas Sporer. MPEG-4 Low Delay Audio Coding Based on the AAC codec. In 106th AES Convention, Munich, Germany, May 1999.

[3] Andreas Spanias, Ted Painter, and Venkatraman Atti. Audio Signal Processing and Coding. John Wiley Sons, 2007.

[4] Martin Dietz, Lars Liljeryd, Kristofer Kjorling, and Oliver Kunz. "Spectral Band Replication, a novel approach in audio coding". In 112th AES Convention, Munich, Germany, May 2002.

[5] MartinWolters, Kristofer Korling, Daniel Homm, and Heiko Purnhagen. "A closer look into MPEG-4 High Efficiency AAC". In 115th AES Convention, New York, NY, USA, October 2003.

[6] Stefan Meltzer, Reinhold Bohm, and Fredrik Henn. "SBR enhanced audio codec's for digital broadcasting such as Digital Radio Mondiale (DRM)". In 112th AES Convention, Munich, Germany, May 2002.

[7] Alexander Groschel, Michael Schug, Michael Beer, and Fredrik Henn. "Enhancing audio coding efficiency of MPEG Layer-2 with Spectral Band Replication for digital radio (DAB) in a backwards compatible way". In 114th AES Convention, Amsterdam, The Netherlands, March 2003.

[8] Andreas Ehret and Michael Schug. "AAC Plus for ring tones". In AES 12th Regional Convention, Tokyo, Japan, July 2005.

[9] Erik Larsen and Ronald M. Aarts. "Audio Bandwidth Extension". John Wiley Sons, 2004.

[10] "ISO/IEC. Information technology generic coding of moving pictures and associated audio information". Part 3: Advanced Audio Coding (AAC), AMENDMENT 9: Enhanced low delay AAC, April 2007. ISO/IEC 14496-3:2005/FPDAM 9.3