coding technologies

MPEG-4 aacPlus - Audio coding for today's digital media world

Whitepaper by:
Gerald Moser, Coding Technologies

November 2005

1. Introduction

Delivering high quality digital broadcast content to consumers belongs to the most challenging tasks of today's media infrastructure development. In digital broadcasting, one of the most critical aspects is the highly efficient use of the available transmission spectrum. Consequently, a careful choice of compression schemes for media content is essential for both, the technical and the economical feasibility of modern digital broadcasting systems. For audio, the MPEG-4 aacPlus (also known as High Efficiency AAC v2) has recently been selected within DVB as part of the overall codec toolbox. aacPlus comprises a full-featured tool set for coding of audio signals in mono, stereo and multichannel signals up to 48 channels at high quality levels over a wide range of bit rates.

aacPlus has proven in several independent tests to be the most efficient audio compression scheme available worldwide. The codec's core components are already in widespread use in a variety of systems and applications where bandwidth limitations are a crucial issue, amongst them XM Satellite Radio, the digital satellite broadcasting service in the US, as well as HD Radio, the terrestrial digital broadcasting system of iBiquity Digital in the US, and Digital Radio Mondiale, the international standard for broadcasting in the long, medium and short wave bands. In Asia, aacPlus is the mandatory audio codec for the Korean Satellite Digital Multimedia Broadcasting (S-DMB) and optional for Japan's terrestrial Integrated Services Digital Broadcasting system (ISDB). aacPlus is also a central element of the 3GPP (3rd Generation Partnership Project) and 3GPP2 specifications and applied in multiple music download services over 2.5 and 3G mobile communication networks. The paper gives an overview of the standardization, its technical components, and its compression efficiency, based on independent third party tests.

2. MPEG-4 aacPlus in international standardization

MPEG-4 aacPlus v2 is the combination of three technologies: Advanced Audio Coding (AAC), Spectral Band Replication (SBR) and Parametric Stereo (PS). All three technologies are currently being specified in ISO/IEC 14496-3 and combined in the HE-AAC v2 profile, which is referred to in ISO/IEC 14496-3:2001/Amd.4. The combination of AAC and SBR is called aacPlus v1 and is specified in ISO/IEC 14496-3:2001/Amd.1 as the HE-AAC profile.

The European Telecommunications Standards Institute (ETSI) has standardized aacPlus v2 in its Technical Specifications TS 102005 "Technical Specification for the use of video and audio coding in DVB services directly delivered over IP" and TS 101 154 "Implementation guidelines for the use of video and audio coding in

broadcasting applications based on the MPEG-2 transport stream". Based on these standardization efforts, aacPlus v2 is available for integration into all kinds of DVB services.

3. Architecture of aacPlus v2

The underlying core codec of aacPlus v2 is the well-known MPEG AAC codec. AAC is considered state-of-the-art for transparent audio quality a at typical bit rate of 128 kbps. Below this rate, the audio quality of AAC would start to degrade, which can be compensated to a maximum degree with the enhancement techniques SBR and PS. SBR is a bandwidth extension technique that enables audio codecs to deliver the same listening experience at approximately half the bit rate the core codec would require if operated on its own. Parametric Stereo increases the coding efficiency a second time by exploiting a parametric representation of the stereo image of a given input signal. Thus, aacPlus v2 is a superset rather than a substitute of the AAC core codec and extends the reach of high-quality MPEG-4 Audio to much lower bit rates. Given this superset architecture, aacPlus v2 decoders are also capable of decoding plain AAC bit streams, as well as bit streams incorporating AAC and SBR data components, i.e. aacPlus v1 bit streams. Hence, aacPlus v2 is also a superset of aacPlus v1, providing the highest level of flexibility for broadcasters as it contains all technical components necessary for audio compression over a high bit rate range (see chapter 4: "Audio quality evaluation").

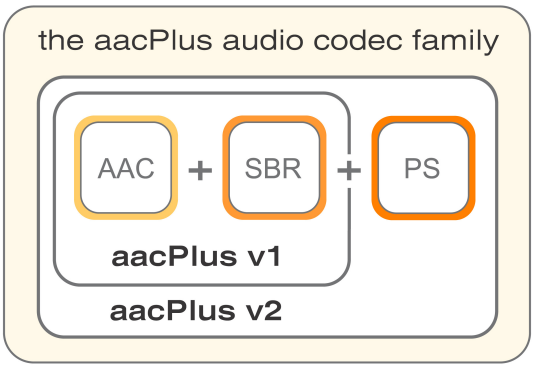

Fig. 1: The aacPlus family of audio codecs

Another important feature of the aacPlus architecture is the extremely flexible transport of metadata. These metadata can be embedded as ancillary data in a way that only compatible decoders take notice of their existence. Other than these simply ignore the metadata. A high flexibility is provided in terms of type, amount and usage of the data. Metadata play an important role in digital broadcasting, e.g. as content description data such as name of an artist or song, or as system related data such as control information for a given decoder.

3.1 Functionality of MPEG AAC

Research on perceptual audio codecs started about 20 years ago. Earlier research on the human auditory system had revealed that hearing is mainly based on a short-term spectral analysis of the audio signal. The so-called masking effect was observed: the human auditory system is not able to perceive distortions that are masked by a stronger signal in the spectral neighborhood. Thus, when looking at the short-term spectrum, a so-called masking threshold can be calculated for this spectrum. Distortions below this threshold are in the ideal case inaudible.

The goal is to calculate the masking threshold based on a psychoacoustic model and to process the audio signal in a way that only audible information resides in the signal. Ideally, the distortion introduced is exactly below the masking threshold and thus remains inaudible. Fig. 2 illustrates the quantization noise produced by an ideal perceptual coding process.

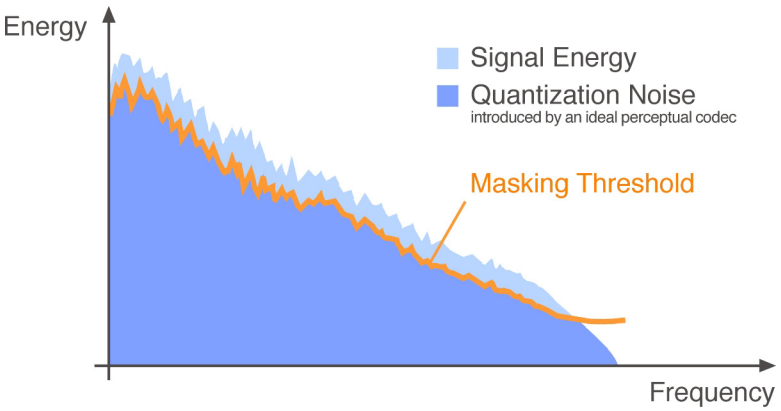

Fig. 2: Inaudible quantization noise produced by an ideal perceptual coding process

If the compression rate is further increased, the distortion introduced by the codec violates the masking threshold and produces audible artifacts.

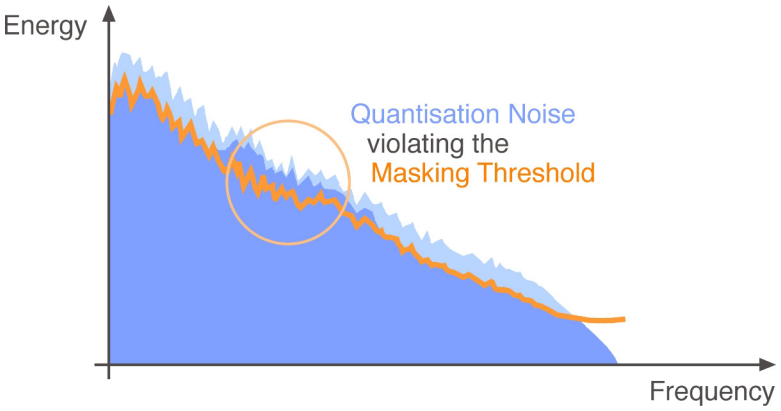

Fig. 3: Waveform coding beyond its limits: audible artifacts above the masking threshold

The main method of overcoming this problem in traditional perceptual waveform codecs is to limit the audio bandwidth. As a consequence, more information is available for the remainder of the spectrum, resulting in a clean but dull sounding signal. Another method, called intensity stereo, can only be used for stereo signals. In intensity stereo, only one channel and some panning information is transmitted, instead of a left and a right channel. However, this is only of limited use in increasing the compression efficiency, as in many cases the stereo image of the audio signal gets destroyed.

At this stage, research on classical perceptual audio coding had reached its limits, as thitherto known methods did not seem to provide more potential to further increase coding efficiency. Hence, a shift in paradigms was needed, represented by the idea that different elements of an audio signal, such as spectral components or the stereo image deserve different tools to be coded more efficiently. This idea led to the development of the enhancement tools Spectral Band Replication and Parametric Stereo.

3.2 Spectral Band Replication

In traditional audio coding, a significant amount of information is spent to code high frequencies, although the psychoacoustic importance of the last one or two octaves is relatively low. This triggered the basic idea behind SBR: based on the cognition of a strong correlation between the high and the low frequency range of an audio signal (further referred to as “high band” and “low band” respectively), a good approximation of the original input signal high band can be achieved by a transposition from the low band.

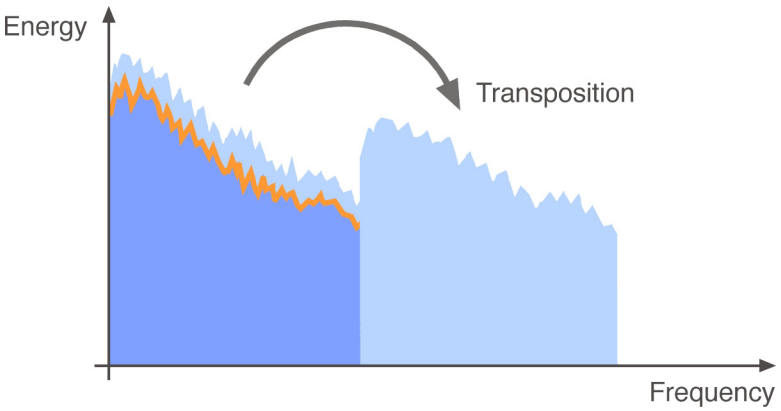

Fig. 4: Creation of high frequencies by transposition

Besides the pure transposition the reconstruction of the high band is conducted 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. This guiding information is referred to as SBR data. Also, efficient packaging of the SBR data is important to achieve a low data rate overhead.

At encoder side the original input signal is analyzed, the high band spectral envelope and its characteristics in relation to the low band are encoded and the resulting SBR data is multiplexed with the core coder bit stream. At decoder side first the SBR data is de-multiplexed, then the core decoder is being run separately and the SBR decoder operates on its output signal, using the decoded SBR data to guide the spectral band replication process. A full bandwidth output signal is obtained. Non-SBR decoders would still be able to decode the backward compatible part of the core decoder, however resulting in a band-limited output signal only.

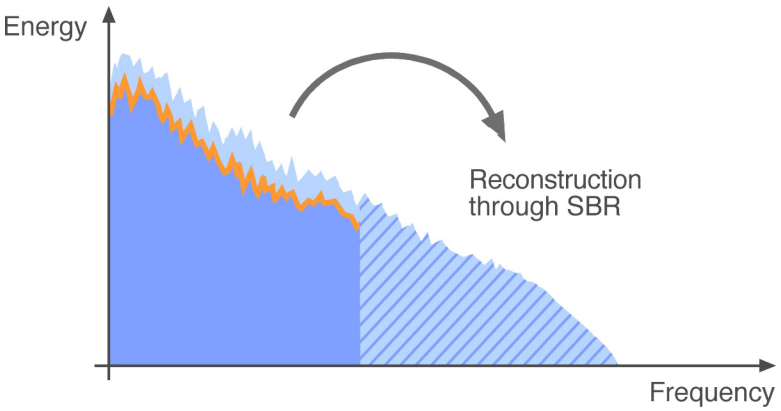

Fig. 5: Envelope Adjustment of the high band

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 guiding information in a way that all of the following criteria are met:

- Good spectral resolution is required
- Sufficient time resolution on transients is needed to avoid pre-echoes
- Cases with non-highly correlated low band and high band need to be taken care of since here transposition and envelope adjustment alone could sound artificial
- A low overhead data rate is required in order to achieve a significant coding gain

When combining AAC with SBR, the resulting codec is being named aacPlus v1. It has proven to fulfill all the criteria above and has been standardized within MPEG-4 in 2003.

3.3 Parametric Stereo (PS)

Whereas SBR exploits the possibilities of a parameterized representation of the high band, the basic idea behind PS is to parameterize the stereo image of an audio signal such as panorama, ambience, or time/phase differences of the stereo channels to enhance the coding efficiency of the codec.

In the encoder, only a monaural downmix of the original stereo signal is coded after extraction of the parametric stereo data. Just like SBR data, these parameters are then embedded as PS side information in the ancillary part of the bit stream.

In the decoder, the monaural signal is decoded first. After that, the stereo signal is reconstructed based on the stereo parameters embedded by the encoder. Fig. 6 shows the basic principle of the parametric stereo coding process.

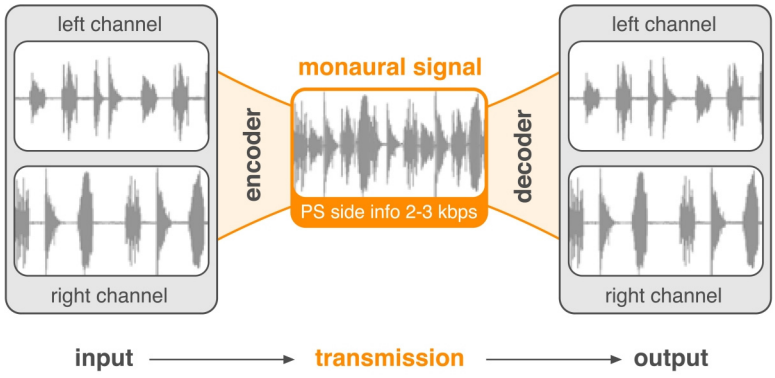

Fig. 6: Basic principle of the parametric stereo coding process

Three types of parameters can be employed in a parametric stereo system to describe the stereo image.

- Inter-channel Intensity Difference (IID), describing the intensity difference between the channels.
- Inter-channel Cross-Correlation (ICC), describing the cross correlation or coherence between the channels. The coherence is measured as the maximum of the cross-correlation as a function of time or phase.
- Inter-channel Phase Difference (IPD), describing the phase difference between the channels. This can be augmented by an additional Overall Phase Difference (OPD) parameter, describing how the phase difference is distributed between the channels. The Inter-channel Time Difference (ITD) can be considered as an alternative to IPD.

3.4 Functionality of the aacPlus v2 codec

The described technologies AAC, SBR and PS are the building blocks of the aacPlus v2 codec family. The AAC codec is used to encode the low band, SBR encodes the high band, and PS encodes the stereo image in a parameterized form. In a typical aacPlus encoder implementation the audio input signal at an input sample rate of fs is fed into a 64 band Quadrature Mirror Filter bank and transformed into the QMF domain.

If the Parametric Stereo Tool is used (i.e. for stereo encoding at bit rates below ~36 kbps), the PS encoder is extracting parametric stereo information based on the QMF samples. Furthermore, a stereo-to-mono downmix is applied. With a 32-band QMF Synthesis the mono QMF representation is then transformed back into the time domain at half the sample rate of the audio signal, fs/2. This signal is then fed into the AAC encoder.

If the Parametric Stereo Tool is not used, the audio signal is fed into a 2:1 resampler, again the downsampled audio signal is fed into the AAC encoder. The SBR encoder is also working in the QMF domain, it extracts the spectral envelope and additional helper information to guide the replication process in the decoder. All encoded data is then multiplexed into a single bit stream for transmission or storage.

Fig. 7 shows the block diagram of a complete aacPlus v2 encoder.

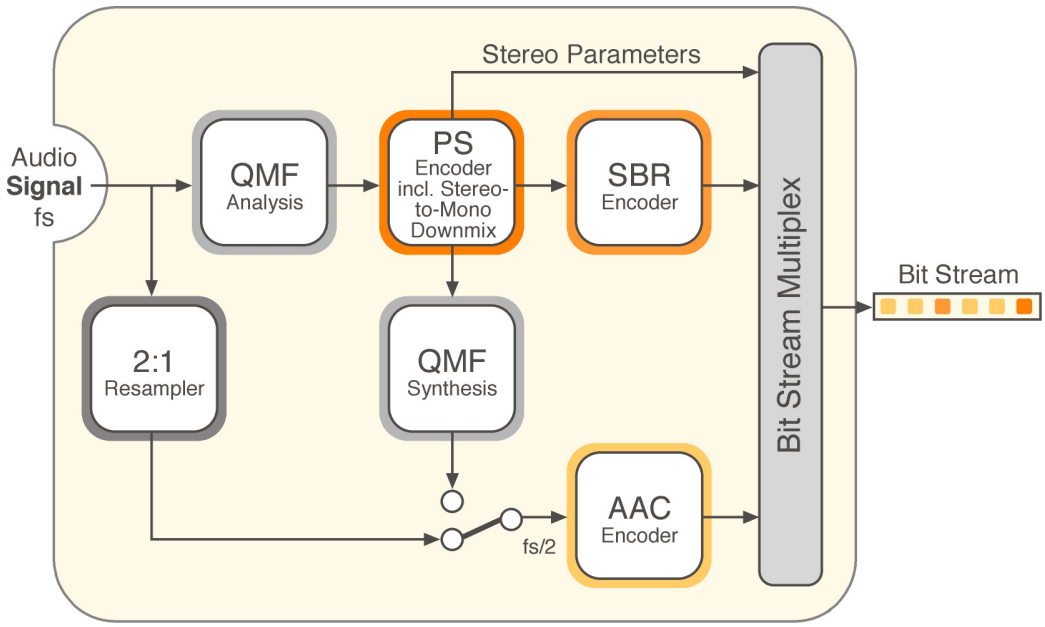

Fig. 7: Block diagram of an aacPlus encoder

In the aacPlus v2 decoder, the bit stream is first decomposed into the AAC, SBR and PS data portions. The AAC decoder outputs a time domain low band signal at a sample rate of fs/2. The signal is then transformed into the QMF domain for further processing. The SBR processing results in a reconstructed high band in the QMF domain. Low and high band are then merged into a full band QMF representation.

If the Parametric Stereo tool is used, the PS tool generates a stereo representation in the QMF domain. Finally, the signal is synthesized by a 64 band QMF synthesis filter bank. The result is a time domain output signal at the full sample rate fs.

Fig. 8 shows the block diagram of a complete aacPlus v2 decoder.

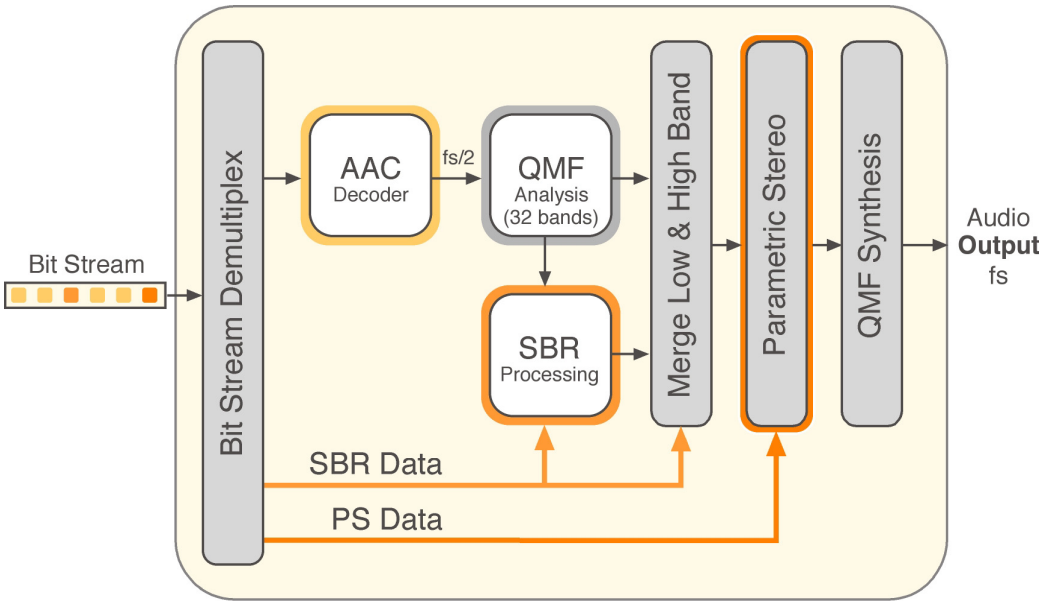

Fig. 8: Block diagram of an aacPlus v2 decoder

4. Audio quality evaluation

The audio quality of aacPlus v1 and aacPlus v2 has been evaluated in multiple double-blind listening tests conducted by independent entities such as the European Broadcasting Union (EBU), the Moving Pictures Expert Group (MPEG), the 3rd Generation Partnership Project (3GPP), and the Institut für Rundfunktechnik (IRT).

4.1 EBU subjective listening test on low-bit rate audio codecs

In 2003, the EBU conducted a comprehensive test evaluating a variety of open standard and proprietary audio codecs including aacPlus v1, AAC, Windows Media Audio, Real Audio, and others, at a bit rate of 48 kbps.[1] The test was conducted according to the MUSHRA test method (*MU*ltiple *S*timulus test with *H*idden *R*eference and *A*nchors). The results have clearly shown the superior compression efficiency of aacPlus v1. Remarkably, as the second best codec in the field scored mp3PRO, the combination of MPEG Layer-3 (MP3) and SBR.

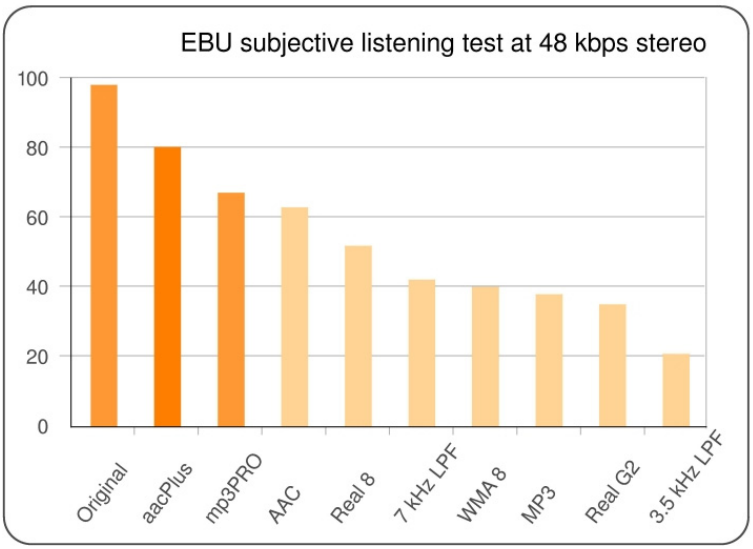

Fig 9: Results of the EBU subjective listening test

4.2 MPEG and 3GPP listening tests

Prior to standardization of aacPlus v2, MPEG has carried out a listening test to verify the efficiency improvement of aacPlus v2 incorporating Parametric Stereo over aacPlus v1. Also for this evaluation the MUSHRA test method was used. According to the scope of the listening test, the considered bit rates included 24 kbps for aacPlus v1, and 32 and 24 kbps for aacPlus v2.

The results of this test evidenced a clear performance gain introduced by Parametric Stereo. At 24 kbps, aacPlus v2 performed significantly better than aacPlus v1 and equal to or better than aacPlus v1 at 32 kbps. A second test conducted by 3GPP expectedly displayed similar results, also at additional bit rates.

---

[1] *"EBU subjective listening test on low bit rate audio codecs" (tech 3296), http://www.ebu.ch/CMSimages/en/tec_doc_t3296_tcm6-10497.pdf*

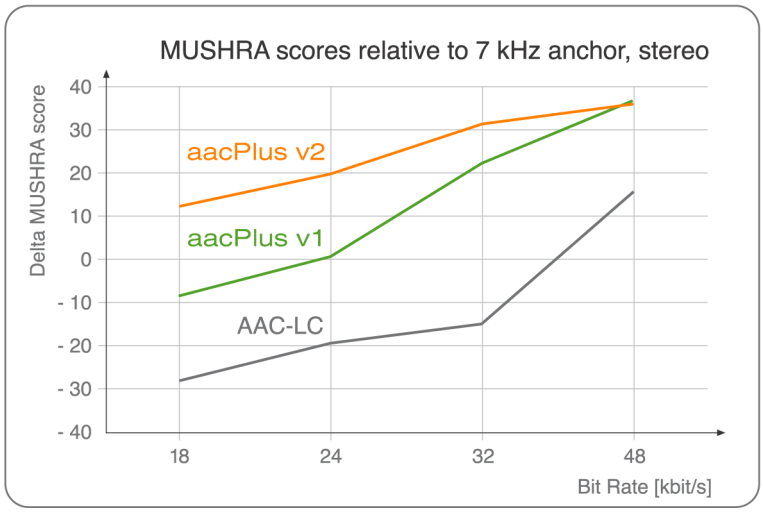

Fig. 10: Performance of aacPlus v1, aacPlus v2, and AAC

4.3 Multichannel listening test of the Institut für Rundfunktechnik IRT

In 2004, the IRT conducted a listening test comprising a number of audio codecs for multichannel applications, amongst them aacPlus, Dolby AC-3, and Windows Media. The results presented a clear advantage of aacPlus providing a significantly higher audio quality at 160 kbps compared to Dolby AC-3 operating at 384 kbps and Windows Media at 192 kbps.

Fig. 11 shows the Mean Opinion Scores (MOS) of the coded audio signals from the original. Although aacPlus operated at the lowest bit rate of all codecs, it outperformed all competitors in terms of audio quality. Calculating the average MOS of all audio items under test, it can be stated that aacPlus provides a better quality at half the bit rate compared to WMA or Dolby AC-3.

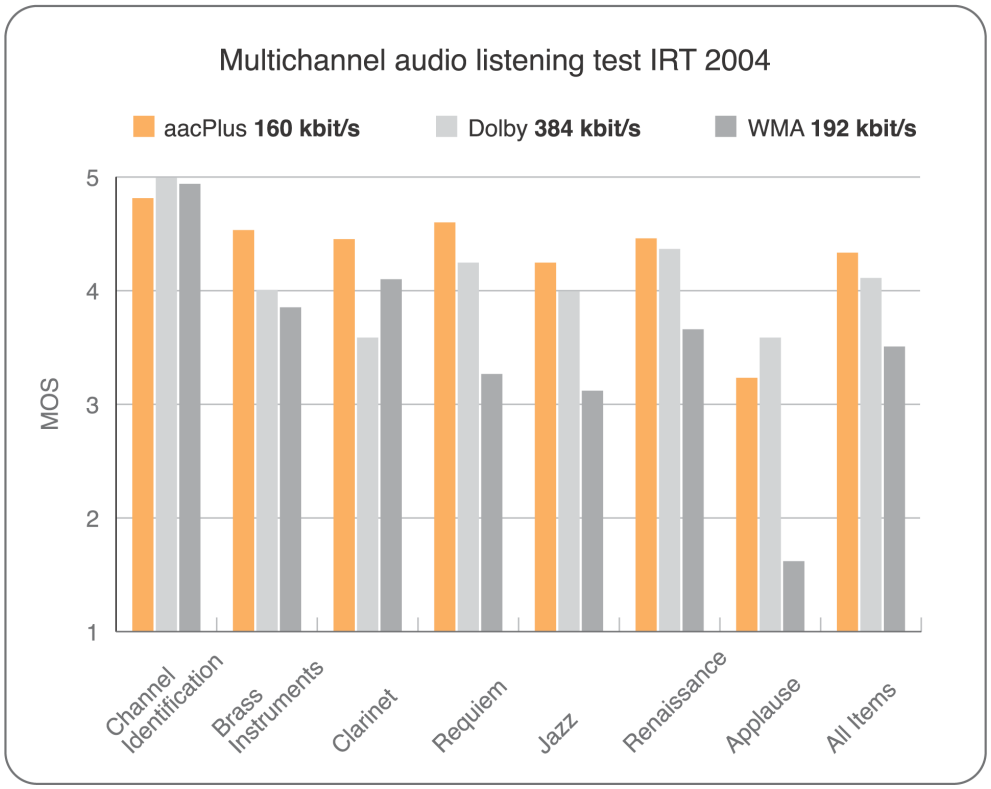

Fig. 11: Test results of the multichannel audio codec test of IRT

4.4 Interpretation of the combined results

Considering the fact that the quality of compressed audio signals scale with the bit rate, the following interpretation of the available test results can be given.

Combining AAC with SBR and PS to aacPlus v2 results in a very efficient audio codec providing high audio quality over a wide bit rate range, with only moderate gradual reduction of the perceived audio quality towards very low bit rates. Fig. 12 gives an impression of the anticipated audio quality vs. bit rate for the various codecs of the aacPlus v2 family.

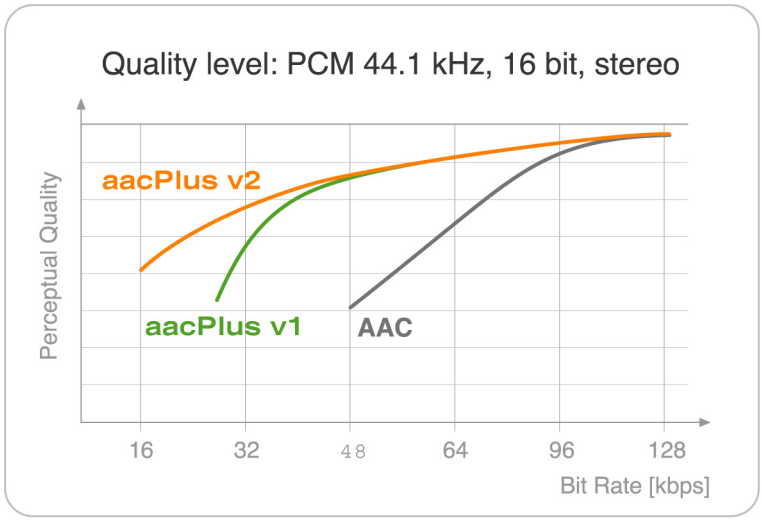

Fig. 12: Comparison of audio quality degradation of aacPlus v1, aacPlus v2 and AAC

The diagram shows only a smooth degradation in audio quality of aacPlus v2 towards low bit rates over a wide range down to 32 kbps. Even at bit rates as low as 24 kbps, aacPlus v2 still produces a quality far higher than of any other audio codec available.

For multichannel 5.1 signals, aacPlus provides a coding efficiency of a factor of two higher compared to Dolby AC-3.

5. Conclusion

This paper has described the aacPlus v2 audio codec, and how existing audio coding technologies, such AAC, can be significantly enhanced by using the novel enhancement techniques SBR and PS. Preliminary studies show that the compression efficiency of AAC can be increased by up to a factor of four.

aacPlus v2, the combination of AAC, SBR and PS, is undoubtedly the most powerful audio codec available today. It is thus the first choice for all application scenarios where bandwidth is limited or very expensive, as is in digital broadcasting or mobile applications.