

CT-aacPlus — a state-of-the-art Audio coding scheme

Martin Dietz and Stefan Meltzer
Coding Technologies, Germany

CT-aacPlus is a combination of **Spectral Band Replication (SBR)** technology – a bandwidth-extension tool developed by Coding Technologies (CT) in Germany – with the MPEG **Advanced Audio Coding (AAC)** technology which, to date, has been one of the most efficient traditional perceptual audio-coding schemes.

CT-aacPlus is able to deliver high-quality audio signals at bit-rates down to 24 kbit/s for mono and 48 kbit/s for stereo signals. The forthcoming Digital Radio Mondiale (DRM) broadcasting system, among others, will use CT-aacPlus for its audio-coding scheme. CT-aacPlus will enable DRM to deliver an audio quality, in the frequency range below 30 MHz, that is equivalent to – or even better than – that offered by today's analogue FM services.

This article describes the principles of traditional audio coders – and their limitations when used for low bit-rate applications. The second part describes the basic idea of SBR technology and demonstrates the improvements achieved through the combination of SBR technology with traditional audio coders such as AAC and MP3.

Advanced Audio Coding (AAC) has so far been one of the most efficient traditional perceptual audio-coding algorithms. In combination with the bandwidth-extension technology, Spectral Band Replication (SBR), the coding efficiency of AAC can be even further improved by at least 30%, thus providing the same audio quality at a 30% lower bit-rate. The combination of AAC and SBR – referred to as **CT-aacPlus** – will be used by the Digital Radio Mondiale transmission system [1] in the frequency bands below 30 MHz and will provide near-FM sound quality at bit-rates of around 20 kbit/s per audio channel.

SBR technology is the latest development in audio-coding research and can be combined with nearly every traditional audio-coding scheme. An average improvement in bit-rate efficiency of at least 30% can be achieved. The combination of coding schemes can be done in a compatible way, thus allowing the upgrade of existing systems based on traditional audio coders. Appropriate transition scenarios will allow a nearly seamless introduction of the new technology which, in the end, will allow us to harness the full advantages of the increased bit-rate efficiency offered by SBR technology.

Traditional perceptual audio coding

Research on perceptual audio coders started about 20 years ago. Research on the human auditory system 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 neighbourhood. 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. Research then started on how to calculate the masking threshold (“psycho-acoustic model”) and on how to process the audio signal in such a way that only audible information resides in the signal. Ideally, an audio codec applies compression such that the distortion introduced is exactly below the masking threshold. *Fig. 1* illustrates the quantization noise that an ideal perceptual coder would produce.

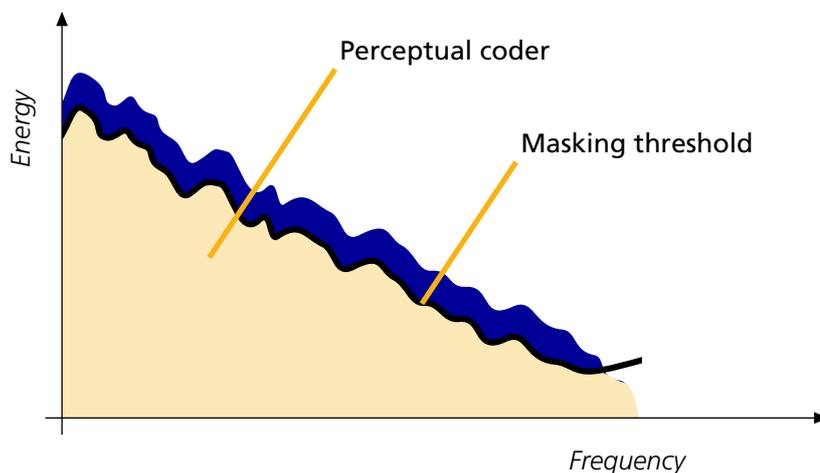


Figure 1
Ideal perceptual audio coding

This research led to today’s well-known traditional perceptual audio codecs, based on waveform codecs; for example, MPEG Layer 2, Dolby AC-3, MP3, Sony Atrac, Lucent PAC and MPEG AAC. All these codecs are based on the same principle, as shown in *Fig. 2*.

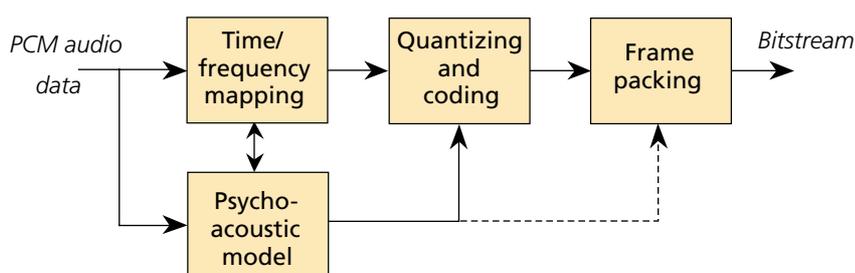


Figure 2
Traditional perceptual waveform encoder

The audio signal is transformed into the frequency domain by means of a filter bank or transform, on a block-by-block basis. The resulting short-time spectrum is quantized in such a way that the masking threshold calculated by the psycho-acoustic model is not violated. The quantized spectrum gets coded and packed into a bit-stream. The decoder performs the reverse signal-processing steps, but does not generally contain a psycho-acoustic model.

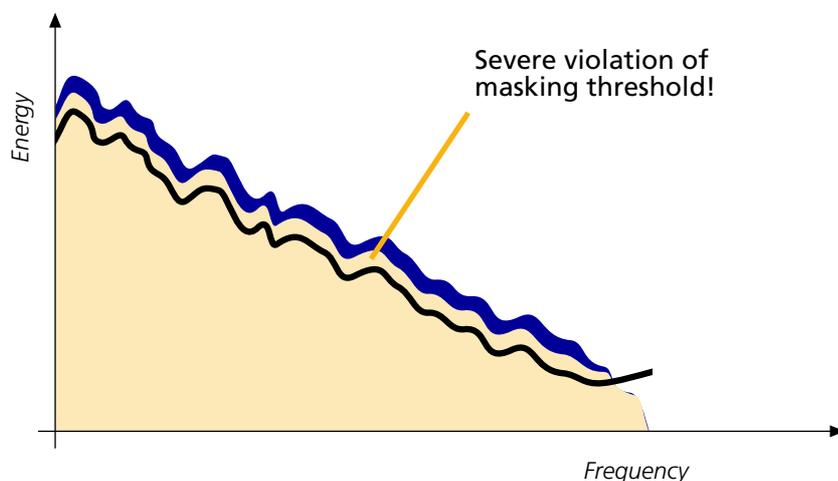


Figure 3
Waveform coding beyond its limits

Although the established perceptual waveform codecs already achieve significant compression, the efficiency is still not high enough to fulfil the needs of systems based on analogue/digital telephone lines or wireless systems and broadcasting systems. *Fig. 3* illustrates what happens if the compression rate is further increased in such a codec: the distortion introduced by the codec violates the masking threshold and produces audible artefacts.

The main method of overcoming this problem in traditional perceptual waveform codecs is to limit the audio bandwidth. As a conse-

Abbreviations

AAC	(MPEG-2/4) Advanced Audio Coding	HVXC	(MPEG) Harmonic Vector Excitation Coding
AES	Audio Engineering Society	MPEG	Moving Picture Experts Group
AM	Amplitude Modulation	MUSHRA	(EBU) MUlti Stimulus test with Hidden Reference and Anchors
CT-aacPlus	(Coding Technologies) advanced audio coding Plus	QAM	Quadrature Amplitude Modulation
DRM	Digital Radio Mondiale	RF	Radio-Frequency
DVB	Digital Video Broadcasting	SBR	Spectral Band Replication
FM	Frequency Modulation	WMA	(Microsoft) Windows Media

quence, more information is available for the remainder of the spectrum, resulting in a clean but hollow-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.

SBR – the next step in audio coding

SBR – the Spectral Band Replication technology developed by Coding Technologies – is a novel technology that significantly increases the efficiency of audio coding. SBR is the result of the latest achievements in audio coding research, which revealed that the high frequencies of an audio signal can be represented much more efficiently than before. The main effect used is the high correlation between the low- and high-frequency content in an audio signal.

In an SBR-based coding system, waveform audio coding is only used to code the lower frequencies of an audio signal. This low frequency content is used to recreate the high frequency content at the decoding side (*Fig. 4*).

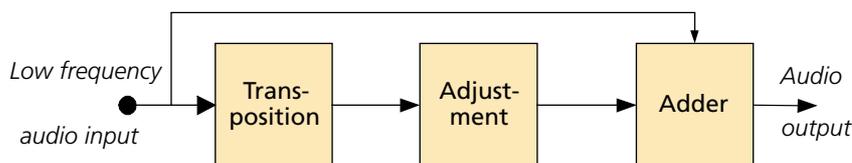


Figure 4
SBR high-frequency regeneration process in the decoder

This is done by state-of-the-art transposition methods. The recreated high-frequency content undergoes some frequency and time domain adjustment before it is combined with the low-frequency part of the audio signal.

Fig. 5 shows the integration of the AAC codec with SBR technology. At the encoding side, SBR encoding is processed prior to the AAC encoding. The input signal to the AAC encoder is band limited, as the high frequencies will be recreated by the SBR algorithm. Some additional information, needed by the SBR decoder to reconstruct the high-frequency part, is multiplexed into the coded bitstream, in the bitstream multiplexer. At the decoding side, the low-frequency output of the AAC decoder is fed into the SBR unit, which recreates the high frequencies and produces a full-bandwidth audio signal.

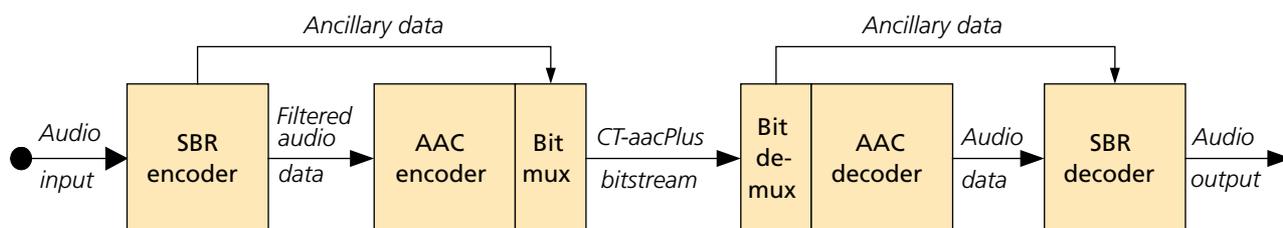


Figure 5
CT-aacPlus coding system

CT-aacPlus provides much higher compression efficiency than standalone AAC, because the AAC part in CT-aacPlus is only used to code the low-frequency content of the signal. As the SBR information needs only a very small amount of data, the AAC codec can use much more information to code the low-frequency content (see Fig. 6). Thus the masking threshold is not violated.

SBR recreates the high frequencies and combines both parts to obtain a full bandwidth audio signal (see Fig. 7). The resulting bitstream consists of two components, the AAC part and the SBR part. The bitstream format can be designed in such a way that an AAC-only decoder can decode the AAC part of the bitstream. Although the resulting audio signal is bandwidth-limited, this feature can be used to introduce SBR in existing systems that use traditional perceptual audio codecs.

SBR technology can be used together with all audio codecs which produce a clean low frequency signal. Another example of the application of SBR is the **MP3Pro** format which is the combination of MPEG Layer-3 (MP3) and SBR (see the panel below).

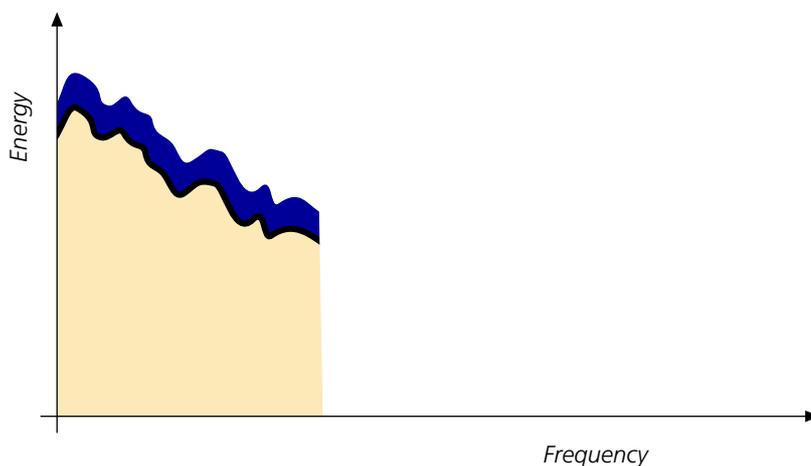


Figure 6
Low frequency part, coded by AAC

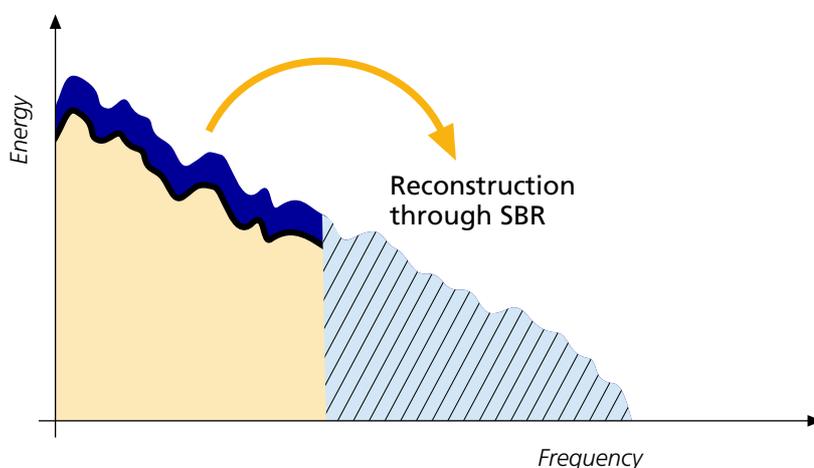


Figure 7
High frequency reconstruction through SBR

Comparison of SBR-enhanced codecs and traditional audio codecs

During the selection of the source-coding algorithm, the DRM consortium conducted several listening tests. Fig. 8 shows the results of a listening test conducted at the BBC, comparing AAC with AAC + SBR at 24 kbit/s mono. The test was conducted using the MUSHRA test method for comparison tests [2]. The MUSHRA test contains several “anchor” signals, such as bandwidth-limited (3.5 kHz and 7 kHz) versions of the original signal, as well as the original signal itself. These anchor signals increase the reliability of the test results. In the MUSHRA scale, 100 represents the original signal.

MP3Pro and CT-aacPlus

The first technology developed and marketed by Coding Technologies was the enhancement technology called Spectral Band Replication (SBR). The combination of MP3 and SBR has been named "**MP3Pro**" and was released by Coding Technologies and its licensing partner, Thomson Multimedia, in June 2001. Products using MP3Pro have been launched or announced by Thomson/RCA, Philips, MusicMatch, STMicroelectronics, Texas Instruments, Gracenote, Syntrillium, Ahead and many others.

"**CT-aacPlus**", the combination of AAC and SBR, is the most powerful audio codec available today. It has been selected as the audio coder for systems such as XM Satellite Radio in the US and the Digital Radio Mondiale broadcast system (DRM) for short-, medium- and long-wave transmissions.

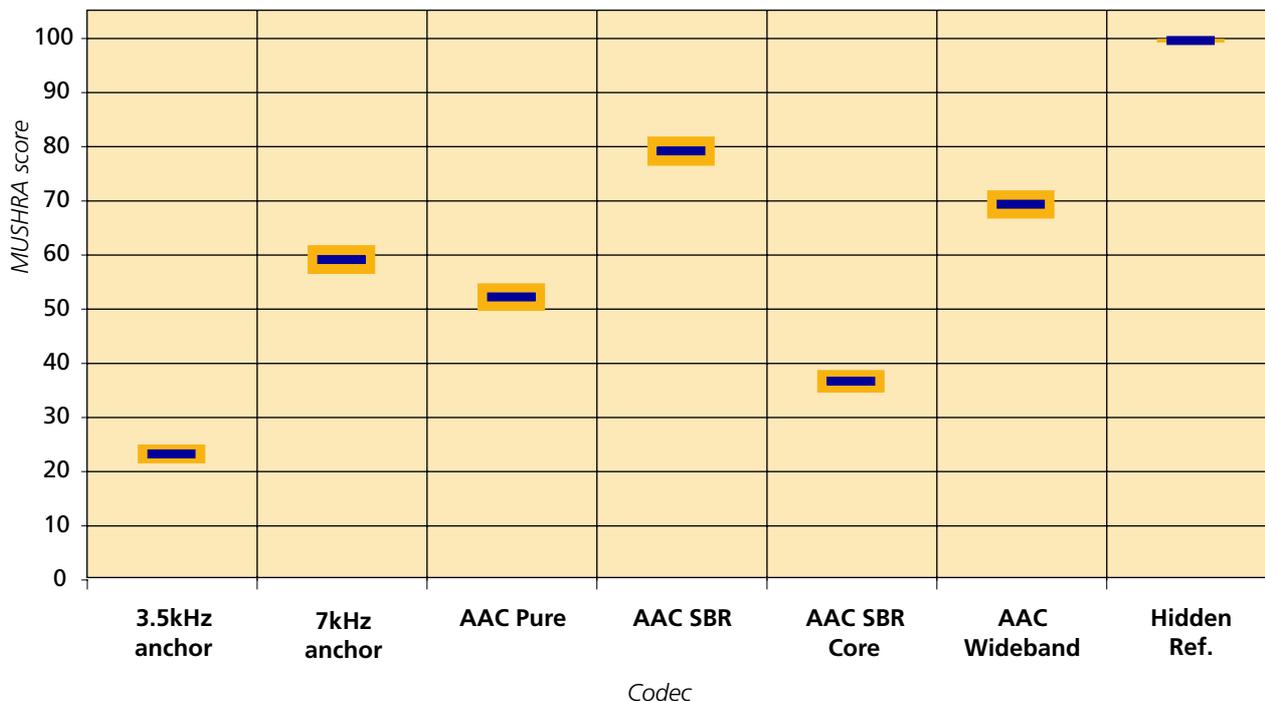


Figure 8
DRM listening test comparing AAC and CT-aacPlus at 24 kbit/s mono

The MUSHRA tests carried out at the BBC included the following signals:

- 3.5 kHz anchor and 7 kHz anchor;
- AAC Pure – i.e. MPEG-4 AAC at 24 kbit/s;
- AAC SBR – i.e. CT-aacPlus at 24 kbit/s;
- AAC SBR Core – the AAC part of the CT-aacPlus signal – at 24 kbit/s;
- AAC Wideband – i.e. MPEG-4 AAC at 32 kbit/s;
- Hidden Reference – the original signal, not bandwidth limited.

The results clearly show the higher coding efficiency of CT-aacPlus when compared to AAC. CT-aacPlus, at 24 kbit/s, yielded an improvement in coding efficiency of more than 30% when compared with AAC at 30 kbit/s.

A further comparison test, this time using several commercially-available codecs, has been performed by the EBU. The first results were shown at a recent AES conference in the UK [3]. Fig. 9 shows the result of this MUSHRA test, where MP3Pro and CT-aac-Plus were compared with MP3, AAC, WMA 8.0 and Real Audio at 48 kbit/s stereo.

The EBU test can be interpreted as follows: the hidden reference (Ref.) gets a grading below 100, indicating that some listeners could not distinguish between it and a very good coded version (most likely the CT-aacPlus codec) for some items. CT-aacPlus performs best – with a mean grading of around 88. Next is MP3Pro, followed by MPEG-4 AAC. Below that are

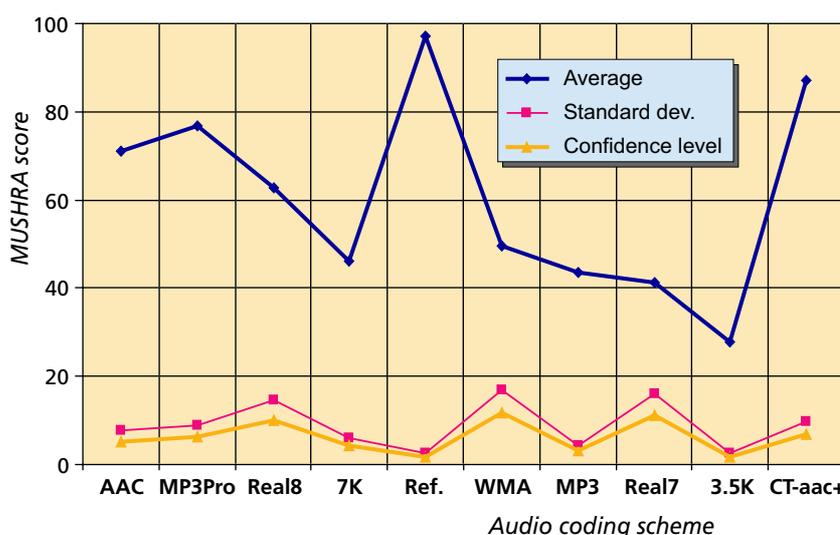


Figure 9
EBU test at 48 kbit/s stereo (IRT results only)

RealAudio8 (v8.5), Windows Media (WMA) 8.5 and MP3. The result is a verification of the outstanding performance of the SBR-enhanced codecs – MP3Pro and CT-aacPlus. It should also be noted that WMA does not perform much better than MP3 – very much in contrast to the marketing hype for the WMA coder!

Use of CT-aacPlus in DRM and other digital radio systems

Keeping the existing channel spacing in the AM bands has been one of the important requirements in the development of the DRM system [1]. Accordingly, the amount of spectrum that is available for one transmission channel is 9 or 10 kHz (depending on the region and band). Although highly-advanced modulation and channel-coding methods are used [4], the characteristics of the channel only allow spectral efficiencies of between 2 and 3 bit/Hz (below this, in very difficult cases). This results in a bit-rate available for source coding of between 18 and 30 kbit/s, approximately equivalent to analogue modem bit-rates. Therefore, selection of the most efficient audio-coding method was essential to achieve high audio quality under such conditions.

The unprecedented performance of CT-aacPlus enables the DRM system to deliver very high quality in the AM bands. In one transmission channel, high-quality mono can be achieved and, if sufficient bit-rate is available, even a limited amount of stereo imaging can be realized. The bundling together of two transmission channels – which might be possible in the future, if the regulatory situation allows it – would result in bit-rates of around 40 to 48 kbit/s, which would provide for high-quality stereo programmes using CT-aacPlus.

For special cases – such as extra-robust transmissions, analogue/digital simulcasts or multilingual transmissions – the lowest bit-rate AAC or speech coding can be used. For ultra-low bit-rate speech transmissions – most likely in parallel with a normal audio programme – the 2 kbit/s HVXC speech codec can be used. The combination of speech coders with SBR is also anticipated.

The DRM system uses error-resilient MPEG-4 AAC to increase the robustness in cases where there are transmission errors. A special hyperframing format has been developed to accommodate the properties of the channel-coding and modulation part and to provide a common framing for the different codecs used in DRM.

CT-aacPlus is also the audio-coding system used by XM Satellite Radio [5]. XM Satellite Radio is one of the two subscription-based satellite radio providers in the USA. Based on two geostationary satellites and a network of terrestrial transmitters, XM Satellite Radio has provided 100 different audio channels across the entire USA since November 2001. The main market for XM Satellite Radio is car receivers, but home reception is



Martin Dietz studied electrical engineering at the Friedrich-Alexander-University in Erlangen, Germany. After obtaining his diploma degree in 1992, he joined the Fraunhofer Institute for Integrated Circuits (FhG/IIS-A) in Erlangen as a researcher in the field of audio coding. Between 1992 and 1995 he worked on the standardization of ISO/MPEG layer 3 (MP3) and the first single-chip MP3 decoder implementation, which is used in several MP3 players today. Between 1995 and 1997 he was head of the research group at IIS that was responsible for the development of ISO/MPEG AAC.

In 1998 Mr Dietz became head of the audio department at IIS, responsible for research and development related to MP3, AAC and MPEG-4 audio algorithms. Since Summer 2000 he has been CEO and President of Coding Technologies, a Swedish/German company specializing in improved audio and speech processing technologies, with a special focus on audio coding. In December 2001, he was awarded a Fellowship from the Audio Engineering Society for his outstanding work in the area of audio coding.

Stefan Meltzer studied electrical engineering at the Friedrich-Alexander University in Erlangen Germany. After he received his Dipl.-Ing. degree in 1990, he joined the Fraunhofer Institute for Integrated Circuits in Erlangen. After working in the field of IC design for several years, in 1995 he became the project leader for the development of the WorldSpace Satellite Broadcasting system at the Fraunhofer Institute. From 1998 until 2000 he led the development team at the Institute for the development of the XM Satellite Radio broadcasting system.

In 2000, Mr Meltzer joined Coding Technologies in Nuremberg, Germany. He currently occupies the position of Vice President for business development.



also anticipated. To achieve high system reliability, the system uses a large amount of diversity and redundancy. Taking into account the limited amount of bandwidth available for the service, maximally-efficient audio coding was essential to achieve the goal of 100 audio channels. As traditional audio codecs were not able to achieve this goal at a satisfactory audio quality, CT-aacPlus has been selected for use in the XM Satellite Radio system.

Conclusions

This article has described how the existing audio-coding technologies, such as MP3 and AAC, can be significantly enhanced by using the novel bandwidth-extension technology, SBR. Preliminary studies show that a 30% increase in coding efficiency can be achieved through the use of SBR.

CT-aacPlus – a combination of AAC and SBR technologies – is undoubtedly the most powerful audio codec available today. It is used in the DRM and XM Radio Satellite broadcasting systems, but the SBR approach could also be extended to other codecs used in digital broadcasting, mobile communications and internet streaming applications.

Bibliography

- [1] DRM website: <http://www.drm.org/plansite.htm>
- [2] G. Stoll and F. Kozamernik: **EBU listening tests on Internet audio codecs**
EBU Technical Review No. 283, June 2000
http://www.ebu.ch/trev_283-kozamernik.pdf
- [3] Martin Link, IRT GmbH: **Internet Audio Quality and the MUSHRA Test**
AES UK Conference 2002: Audio Delivery: the changing home experience
<http://www.aes.org/sections/uk/conference/02program.html>
- [4] Jonathan Stott: **DRM – Key Technical Features**
EBU Technical Review No. 286, March 2001.
http://www.ebu.ch/trev_286-stott.pdf
- [5] XM Satellite Radio: <http://www.xmradio.com/>
- [6] DRM Source Coding Group, 2001: **Report on Subjective Listening Tests of SBR-LC, an AAC-based Audio Bandwidth Widening Tool**
<http://www.drm.org/plansite.htm>

Acknowledgements

The authors would like to thank their colleagues at Coding Technologies in Sweden and Germany for their contributions to this article. Furthermore, the authors would like to thank their DRM colleagues for all the hard and excellent work that went into creating the DRM Standard.
