Fraunhofer IIS

TECHNICAL PAPER

EXTENDED HE-AAC – BRIDGING THE GAP BETWEEN SPEECH AND AUDIO CODING

One codec taking the place of two; one unified system bridging a troublesome gap. The fifth generation MPEG audio codec Extended High Efficiency AAC (xHE-AAC) is the first to deliver high quality for speech, audio and mixed signals at very low bit-rates.

This paper will take a brief look at the inner workings of xHE-AAC, in the process answering two key questions: what are the working principles of audio and speech codecs, and how was their structure taken into account when developing xHE-AAC?

Fraunhofer Institute for
Integrated Circuits IIS

Director
Prof. Dr.-Ing. Albert Heuberger
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Matthias Rose
Phone +49 9131 776-6175
matthias.rose@iis.fraunhofer.de

Contact USA
Fraunhofer USA, Inc.
Digital Media Technologies*
Phone +1 408 573 9900
codecs@dmt.fraunhofer.org

Contact China
Toni Fiedler
Phone +86 138 1165 4675
china@iis.fraunhofer.de

* Fraunhofer USA Digital Media Technologies, a division of Fraunhofer USA, Inc., promotes and supports the products of Fraunhofer IIS in the U. S.

UNRIVALLED AUDIO QUALITY FOR ALL SIGNAL TYPES

Prior to the development of xHE-AAC, there was the speech coding scheme on the one hand and the audio coding scheme on the other. Both are eminently capable of delivering good quality when applied to their individual operating fields, but reveal weaknesses when deployed elsewhere. Consequently, speech codecs are perfect for speech but show deficiencies when it comes to music at low bit-rates. Meanwhile, audio codecs perform very well for music signals, but cave in when it comes to speech at very low bit-rates.
This predicament provides audio developers with a challenge as there are plenty of services and applications that move through the domains of both speech and music. They work with various signal types and therefore require a suitably flexible codec. Streaming and broadcast are two prominent examples. Until now, providers of low bit-rate services have been forced to prioritize either sound or speech and the matching codec, resulting in poor audio quality for the other signal world. Alternatively, they have attempted to test and integrate two codecs at considerable expense in terms of both time and money.
This undesirable situation has been further amplified by significant changes on the receiver end. When originally developed, mobile phones were intended only for straightforward voice and texting applications. But during the past few years, major technological developments have allowed users to stream videos and music onto their mobile devices, listen to their favorite digital radio stations, and download apps to deliver audio books, magazines and other content.

xHE-AAC is the perfect solution for all of these requirements. Surpassing its goal of performing at least as good as the best speech codec, the 3GPP Adaptive Multi-Rate Wideband Plus (AMR-WB+), and the best audio codec, MPEG-4 High Efficiency Advanced Audio Coding (HE-AAC), it unites the strongest features of these two technologies, delivering unrivalled audio quality starting at bit-rates as low as 6 kbit/s per channel for any type of signal.

BEHIND THE SCENES: CODEC STRUCTURES

Nowadays, audio and speech codecs rely on the following three main elements:

- a core-coder for the representation of low and intermediate frequency signal components;
- a parametric bandwidth extension which uses various parameters to generate higher frequency content from low frequency portions;
- a parametric stereo coder to create additional stereo signals on the basis of a mono downmix and spatial parameters.

At low bit-rates, the parametric tools mentioned above provide increased coding efficiency while maintaining good audio quality. The tools can be selectively deactivated when the codec operates at higher bit-rates and the core-coder is able to independently handle a wider bandwidth as well as the discrete coding of several channels.

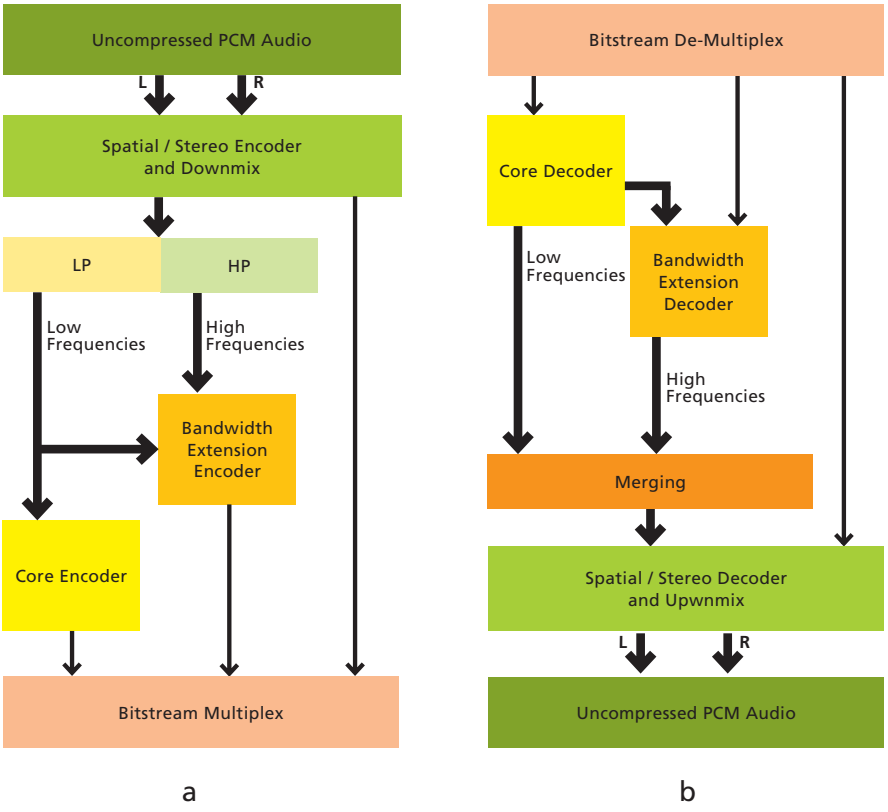

Figure 1: Depiction of a typical audio codec structure, including the core codec and parametric tools for bandwidth extension and stereo signals.
Figure 1 a): The encoder.
Figure 1 b): The corresponding decoder structure.
Bold arrows represent audio signals. Thin arrows depict side information and control data.

COMPOSITION OF AN AUDIO CODING SCHEME

As with other general transform coding schemes, the working principle of Advanced Audio Coding (AAC) is based on psychoacoustics. Aiming for 'irrelevance removal', it takes advantage of temporal and simultaneous masking effects.

The resulting audio coding scheme relies on the following three main steps:

- the conversion from time to frequency;
- the quantization of the frequency spectrum, during which process information from a psychoacoustic model helps to control the quantization error;
- the encoding of quantized spectral coefficients and corresponding side information with minimum entropy.

These steps guarantee the codec's impressive flexibility when handling different kinds of signal types at various operating points.

The audio codec HE-AAC v2 provides an even greater coding efficiency since it blends the AAC core with the tools "Spectral Band Replication" (SBR) and "Parametric Stereo" (PR):

- To enhance the operational range of speech or audio codecs, SBR regenerates higher frequency content based on similar sounding signals coming from lower and mid frequencies. Additional side information is also transmitted to reconstruct this high-frequency content later on. The SBR model profits from the human ear's tendency to sense lower frequencies much more accurately than higher frequencies.
- In combination with parameters for inter-channel level, phase and correlation, PS can generate stereo content from a mono downmix.

COMPOSITION OF A SPEECH CODING SCHEME

Speech coding models, e.g. Algebraic Code Excited Linear Prediction (ACELP), perform very well for speech signals at low bit-rates. This level of quality is possible because the time domain source-filter is closely related to the development of human speech in the human vocal tract.

The three most essential elements in modern speech coding schemes are:

- a short-term linear predictive coding scheme (LPC) for moulding the coded sound according to the shape of the human vocal tract;
- a long-term predictor (LTP) exploiting the pattern by which the vocal chords vibrate, also known as "Adaptive Codebook";
- a so-called "Innovation Codebook" which is responsible for encoding the speech signal's non-predictive portions.

Following the ACELP principle, AMR-WB employs an algebraic representation for the "Innovation Codebook" comprising a small number of individual non-zero pulses within a given time segment represented by means of an algebraic allocation formula. The speech codec's parameters - i.e. the LPC coefficients, the LTP lag and gain, and the innovative excitation - are efficiently coded in each frame. As a result, this coding model delivers high quality audio for speech signals at high as well as at low bit-rates.
To perform equally well with music signals, a transform coding mode for the excitation signal (TCX) as well as a parametric frequency and stereo extension were added to AMR-WB, resulting in AMR-WB+. Despite these enhancements, the codec cannot compete with HE-AAC v2 when it comes to music signals.

THE UNIFIED SYSTEM: THE DESIGN PRINCIPLE OF xHE-AAC

For a comprehensive introduction to the xHE-AAC coding scheme, please refer to the AES Convention Paper „MPEG Unified Speech and Audio Coding – The ISO/MPEG Standard for High-Efficiency Audio Coding of all Content Types".

While generally maintaining the structure of HE-AAC v2 shown in Figure 1, xHE-AAC features updated parametric tools for even greater efficiency. These include enhanced SBR (eSBR) and improved parametric stereo coding. As illustrated in Figure 2, the core coder unites an AAC based transform coder and speech coding technology, including ACELP.

Typically, encoder structures in MPEG are not fixed and implementers are only obliged to create valid bit-streams. The same principle works for xHE-AAC, resulting in the possibility of continuously improving the encoder's performance in the years to come.
xHE-AAC shares all capabilities and tools of AAC, including features for discrete stereo or multi-channel operations. xHE-AAC decoders are compatible with AAC-LC, HE-AAC and HE-AACv2 streams.

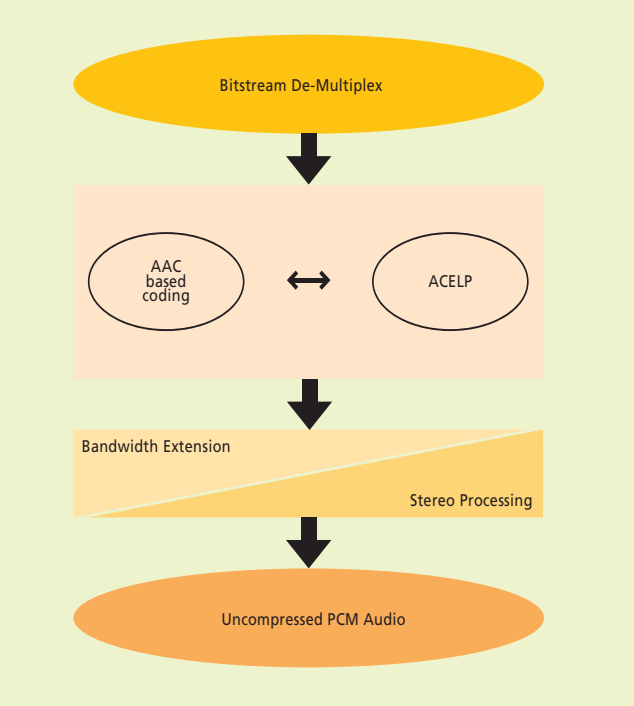

Figure 2: Working principle of the xHE-AAC core decoder

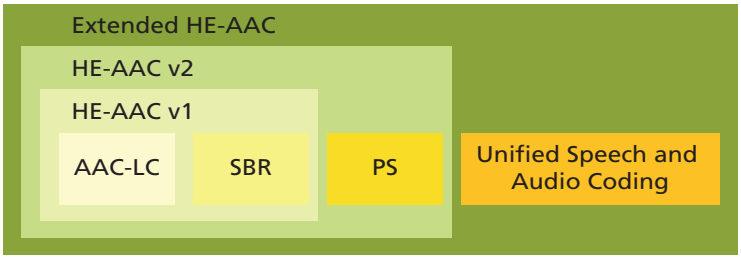

Figure 3: The new member of the AAC codec family is an upgrade of HE-AAC v2 with integrated speech coding tools and more efficient tools for general audio signal coding.

For greater flexibility and efficiency, xHE-AAC introduces the following new features:

- a more efficient context-adaptive arithmetic decoder replacing the AAC Huffman decoder;
- a Frequency Domain LPC Noise Shaping (FDNS) mechanism complementing the AAC scale factor mechanism, responsible for controlling the quantization noise shaping;
- a larger set of window lengths within the xHE-AAC MDCT for a better time-frequency decomposition.

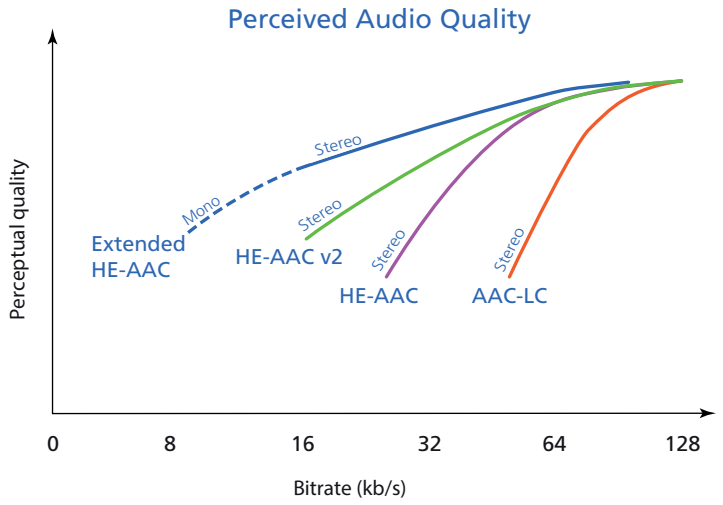

Figure 4: Qualitative illustration of perceived audio quality

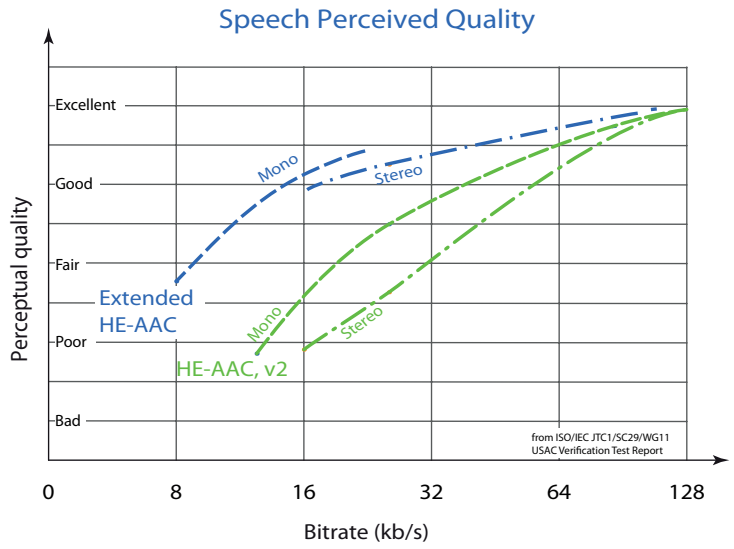

Figure 5: Speech perceived quality. From ISO/IEC JTC1/SC29/WG11 USAC Verification Test Report.

WHERE xHE-AAC CAN MAKE A CHANGE

The latest addition to the MPEG AAC codec family delivers high audio quality for a wide variety of bit-rates, making it especially valuable for applications where bit-rates vary or are set very low.
This includes the streaming of media content, especially on to mobile devices. Since unstable network conditions are no obstacle for the codec, users can enjoy their multimedia experience with fewer dropouts and shorter buffering time. The codec's bit-rate efficiency also creates new space which can either be used for new programs or tracks, or can simply be converted into cost savings – a feature digital broadcasting services can benefit from.

The first radio system to adopt xHE-AAC and benefit from its features is Digital Radio Mondiale (DRM).

ABOUT FRAUNHOFER IIS

When it comes to advanced audio technologies for the rapidly evolving media world, Fraunhofer IIS stands alone. For more than 25 years, digital audio technology has been the principle focus of the Audio and Multimedia division of Fraunhofer Institute for Integrated Circuits (IIS). From the creation of mp3 and the co-development of AAC to the future of audio entertainment for broadcast, Fraunhofer IIS brings innovations in sound to reality. Today, technologies such as Fraunhofer Cingo for virtual surround sound, Fraunhofer Symphoria for automotive 3D audio, AAC-ELD for telephone calls with CD-like audio quality, and Dialogue Enhancement that allows television viewers to adjust dialogue volume to suit their personal preferences are among the division's most compelling new developments.

Fraunhofer IIS technologies enable more than 7 billion devices worldwide. The audio codec software and application-specific customizations are licensed to more than 1,000 companies. The division's mp3 and AAC audio codecs are now ubiquitous in mobile multimedia systems.

Fraunhofer IIS is based in Erlangen, Germany and is an institute of Fraunhofer-Gesellschaft. With 23,000 employees worldwide, Fraunhofer-Gesellschaft is comprised of 67 institutes making it Europe's largest research organization.

For more information, contact Matthias Rose, matthias.rose@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/audio.