SPECTRAL BAND REPLICATION AND AACPLUS CODING – AN OVERVIEW

Clark Novak
Telos Systems
Cleveland, Ohio

Introduction

Over the past decade, huge advances have been made in the area of audio coding for bit-reduced transmission. Fast, effective perceptual audio coders like MPEG Layer 3 and MPEG-2 AAC (Advanced Audio Coding) have been proven to deliver studio-quality audio with little or no perceptual loss, at bit rates as low as 64 Kbps (over digital transmission paths such as satellite and ISDN networks).

Meanwhile, breakthroughs in efficient coding able to transport satisfactory audio over low-bandwidth connections, such as analog POTS (Plain Old Telephone System) phone lines, have been less successful. While the widely-used CCITT G.722 Wideband Speech Coding and CELP standards produce acceptable results at bit rates as low as 48 Kbps when used for speech, many analog connections are not capable of data rates even that high, and severe deficiencies are apparent when music or mixed program material is transmitted. Equipment manufacturers have advanced several proprietary schemes, but none represented a substantial improvement of the compression-to-quality balance.

Thus, another approach to coding for low-bandwidth transmission was called for, and was achieved with the introduction of Spectral Band Replication (known as SBR) by Coding Technologies.

Limitations of Previous Coding Methods

When discussing audio transmission, the question invariably arises, "Why use low bit rate connections at all?" Logistically, it is still a necessity. As modern technology struggles to make use of existing analog infrastructure for transmission of digital applications, analog transmission circuits are overwhelmingly more available than digital. Additionally, low bit rate audio coding is an "enabling technology" for applications such as digital radio, Internet streaming, mobile multimedia applications and, in the case of broadcast mediums, transmission of audio from remote locations to a central broadcast facility using the Public Switched Telephone Network, whether over ISDN, POTS or cellular data paths.

Advanced Perceptual Audio Coding techniques (like MPEG Layer-3 or MPEG-2 AAC) exploit the properties of the human perceptual system by eliminating audio frequencies and tones that are "masked" by other tones to achieve transmission of audio with almost perceptible loss of quality, often reducing the size of transmitted audio data by as much as 12 times. This makes such schemes perfect for high quality low bit-rate applications, remote ISDN broadcasting, like soundtracks for CD-ROM games, solid-state sound memories, Internet audio, digital audio broadcasting systems, and other similar applications.

The use of perceptual codecs at low bit rates, though successful, can be risky. The best performing perceptual audio coder on Earth, MPEG2-AAC, achieves "transparent" audio quality (indistinguishable source from output) at 128 Kbps (stereo), or a compression of approximately 12 to 1. Below 128 Kbps, the perceived audio quality of most of these codecs begins to break down, either reducing the overall audio spectrum or generating coding artifacts while trying preserve spectrum when insufficient bandwidth is present to do so. Either one of these scenarios can result in unacceptable audio quality, as shown in the figure below.
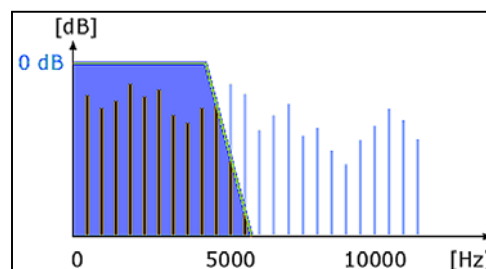

Figure 1: Effects of band-limiting a typical audio signal

Overcoming the Limitations

Conquering the restrictions of previous perceptual coders to provide a new encoding method that works well with low-bandwidth transmission paths required a new technology that could work compatibly with existing coding methods to substantially improve their capabilities.

Spectral Band Replication (SBR), a new audio coding enhancement tool developed by Coding Technologies, improves the performance of low bit rate audio and speech codecs by either increasing the audio bandwidth of existing codecs (such as Layer 3 or AAC) at a given bit rate, or improving coding efficiency at a given quality level.

SBR can increase the limited audio bandwidth that a conventional perceptual codec offers at low bit rates so that it equals or exceeds analog FM audio bandwidth (15 kHz). SBR can also improve the performance of narrow-band speech codecs, offering broadcasters speech-only channels with 12 kHz audio. Where most speech codecs are very band-limited, SBR can not only improve speech quality, but also intelligibility and comprehension.

From a technical point of view, SBR makes possible highly efficient coding of high frequencies in audio compression algorithms. When used in conjunction with SBR, the underlying coder is only responsible for transmitting the lower part of the spectrum, and so can be operated at a reduced sample rate. The SBR decoder generates the higher frequencies, which is mainly a post-process following the conventional waveform decoder.

Instead of transmitting the spectrum, SBR reconstructs the higher frequencies in the decoder based on an analysis of the lower frequencies transmitted in the underlying coder. To ensure an accurate reconstruction, some guidance information is transmitted in the encoded bit stream at a very low data rate.
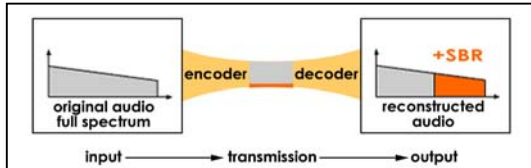

Figure 2. SBR increases the efficiency of PACs by decreasing the amount of spectrum transmitted

The reconstruction of these high frequencies is efficient for harmonic as well as for noise-like components, and allows for proper shaping in the time domain as well as in the frequency domain. As a result, SBR allows full bandwidth audio coding at very low data rates, thus significantly increasing the compression efficiency of the core coder.

In terms of performance, SBR can enhance the efficiency of perceptual audio codecs by approximately 30% in the medium to low bit rate range. The exact level of improvement also depends on the underlying codec; for instance, using SBR in conjunction with MP3 (called MP3 Pro) produces quality at 64 Kbps stereo comparable with conventional MP3 at a bit rate of better than 100 Kbps stereo. SBR offers maximum efficiency in the range of bit rates where the underlying codec itself is able to encode audio signals with an acceptable level of coding artifacts at a limited audio bandwidth.

Combining SBR with AAC

MPEG2-AAC is a perfect choice to combine with SBR, since its qualities include the ability described above.

AAC is the newest audio coding method selected by MPEG, and became an international standard in April 1997. It is a fully state-of-the-art audio compression tool kit that provides performance superior to any known approach at bit rates greater than 64 Kbps, and excellent performance relative to the alternatives at bit rates reaching as low as 16 Kbps.

AAC is the first codec system to fulfill the ITU-R/EBU requirements for indistinguishable quality at 128 Kbps/stereo. It has approximately 100% more coding power than Layer 2 and 30% more power than the former MPEG performance leader, Layer 3. AAC takes advantage of such tools as temporal noise shaping, backward adaptive linear prediction and enhanced joint stereo coding techniques in addition to the techniques used in MPEG Layer 3, resulting in superior high-fidelity audio at lower bit rates and with less delay than Layer 3 or Layer 2.

Combining AAC with SBR results in a superset of AAC technology, called by the trade name aacPlus. This blend of technologies helps to greatly increase the efficiency of AAC — by more than 30% according to independent evaluations.

As a result, aacPlus delivers digital audio broadcasting quality at or even below 48 Kbps for stereo signals, and at even lower bit rates when used with mono signals. In several independent quality evaluations, aacPlus could demonstrate it's superior performance at low bit rates. In careful double-blind listening tests conducted in DRM, MPEG or by the EBU (European Broadcasting Union), aacPlus outperformed each and every other codec it competed with.

The selection of aacPlus for the Digital Radio Mondial and XM Satellite Radio broadcasting systems demonstrates that the benefits of high coding efficiency can be combined with error robustness for transmission over difficult channels.

The SBR technology (used in combination with AAC) has been selected as the reference model for bandwidth enhancement technologies inside MPEG. This process will result in an extension to the MPEG-4 specification. MPEG-4 is expected to become a widely used standard in areas such as Internet streaming and mobile multimedia services. aacPlus (or more precisely the MPEG flavor of aacPlus) will help MPEG-4 to perform even better in a variety of bitrate-sensitive applications.

Conclusion

We have selected aacPlus as the coding method for the new Zephyr Xport for just these reasons — excellent spectrum reproduction and low delay at bit rates commonly available using analog telephone circuits. It is, without much doubt, the best low-bit rate codec yet devised.

The result is that, for the first time ever, true 15 kHz mono audio capable of high fidelity reproduction of both voice and musical materials is possible over POTS telephone circuits. MPEG AAC has been independently tested using a double-blind procedure and found to be superior to any other scheme at rates down to 16 Kbps. Spectral Band Replication takes this already excellent performance to jaw-dropping amazing. Because it is an enhancement designed specifically for very low bit rates, it is perfect for POTS codecs.

References

On Beer and Audio Coding, Steve Church, Telos Systems,
www.telos-systems.com/techtalk/aacpaper_2/default.htm

MPEG-2 AAC Explained, Fraunhofer IIS,
www.iis.fraunhofer.de/amm/techinf/aac/

The CCITT G.722 Wideband Speech Coding Standard, Esin Darici Haritaoglu,
www.umiacs.umd.edu/~desin/Speech1/node3.html

SBR Explained, Coding Technologies,
www.codingtechnologies.com/technology/sbr.htm