

SO WHAT'S THE BIG DEAL ABOUT AACPLUS™ ?

SBR Technology And How It Makes Aacplus™ A Major Advance

Rolf Taylor
Telos Systems
Cleveland, Ohio

In the broadcasting world, aacPlus™, otherwise known as AAC + SBR (Spectral Band Replication), has been in the news for a few years now. Not only is it being used as non-proprietary coding algorithm for transmission over POTS by a major codec manufacturer, but it is also the codec chosen for use by XM Radio for their DARS system, and the new Digital Radio Mondiale (DRM) international standard for short/medium wave transmission. The mp3PRO file format (which is MP3 + SBR) has become popular as an audio sharing file format, for many of the same reasons and is now included in many consumer-oriented software programs, such as MusicMatch Jukebox. Expect to hear more about this technology over the next months.

To best understand what makes SBR such a major advance in coding efficiency requires that one understand how previous bit rate reduction techniques work. And to understand how bit rate reduction works, it helps to recall how audio is digitized.

AUDIO CODING BASICS

Digitization

When audio is digitized, the amount of data produced is a function of the sampling rate and the sample word-size. For example, most pro-audio systems use a sampling rate of 48 thousand samples per second (usually represented as 48 kHz). This means that each second the incoming audio is sampled (or measured) 48 thousand times. To look at this another way, every 21 micro-seconds a sample is taken. Naturally, doubling the sample rate will double the amount of resulting data. The Nyquist Theorem states that the highest reproducible audio frequency will be at one half the sampling rate - thus a 48 kHz system can (in theory) reproduce audio frequencies up to 24 kHz.

The other factor that determines the raw data rate produced by digitization is the precision of measurement of the samples. This is sometimes called "resolution", "bit depth", or "word size". Typical resolutions are 16 or 24 bits per sample. Increased resolution causes a reduced noise floor, at the cost of increased raw data. 16 bits/sample yields a signal to noise ratio in the 90 dB range, suitable for distribution, but not always for production. Hence, 24-bit resolution is common in pro-audio applications.

The amount of data generated by digitizing audio is generally specified in "bits per second" or, since the numbers tend to be rather big, as "kilo bits per second" (kbps). This number is the "bit rate". The bit rate can be calculated as follows:

Sample rate x resolution x the number of audio channels. For example, the data rate for red book CD audio is:

44100 samples per second x 16 bits per sample x 2 channels = 1,411,200 bits per second. This is generally described by the round number "1.4 mega bits per second" (Mbps).

Computers measure memory in 8-bit bytes, so we can determine the actual storage requirements by dividing the bit rate by 8. 1,411,200 bits per second / 8 bits per byte = 176,400 bytes per second.

BIT RATE REDUCTION OVERVIEW

With today's reasonably priced hard drives, it is quite common to see audio stored on computers in "linear" or "uncompressed" form. However, when audio is to be transported, the cost of the transmission channel can be a significant factor. In our CD example above, it would require nearly a full T1 data circuit to transport a single stereo stream. Therefore, some form of "data reduction" can be desirable.

The most basic form of bit savings is by carefully evaluating the factors discussed above. While 24-bit resolution is desirable when recording live tracks for later production, it may very well be possible to use fewer bits later in the production chain. Reducing the sample rate (and therefore the audio frequency response) can also be a way to reduce the bit rate with only minor decreases in fidelity.

Resolution is rarely reduced below 16 bits. However, it is quite common to run across audio material with a sampling rate of 44.1 kHz, 32 kHz or even 24 kHz. For example, a sampling rate of 24 kHz will yield an audio frequency response of better than 10 kHz with a 50% savings in bit rate versus the 48 kHz sampling rate (384 kbps versus 768 kbps/channel). This still yields a very good user experience, and saves on transmission cost.

Even when advanced bit rate reduction rates are used, the sampling rate and resolution are worth taking into consideration - the more data you begin with, the harder the job will be to get it to a given target bit rate.

True lossless Coding - Entropy reduction

Many computer data files have considerable redundancy contained within them, and there are a number of redundancy reduction algorithms available to remove redundancy and therefore reduce file size. Later, the file is restored to its original form by using a complementary process. The classic examples are computer programs such as “Stuffit”, “Winzip” and “PK-Zip”.

Unfortunately, audio material has considerably less redundancy than most graphics or text files. Consequently, entropy reduction techniques offer only a small degree of bit rate reduction, typically less than 2:1.

Entropy reduction techniques do offer some value. A number of perceptual coding algorithms also employ them to enhance performance. In fact, the two techniques complement each other well.

ADPCM Coding

The nature of most audio is that the absolute difference in value between consecutive samples tends to be smaller than the average value of the samples themselves. Adaptive Delta Pulse Code Modulation operates by encoding only the difference (the “Delta”) between samples. To maximize the compression ratio a variable step size is used to encode these sample-to-sample values. A “predictor” algorithm is used to estimate the next value and then determine the step size (e.g. during highly dynamic passages a coarser step size is used). Often, a multi-band approach is used. The most common ADPCM algorithms are the international standard G.722 (offering 7 kHz frequency response) and the proprietary APT-X family of algorithms.

ADPCM methods offer a modest 4:1 compression ratio. An advantage is that coding delay is quite low, typically under 10 msec. High error resilience, and good performance with multiple passes of the same algorithm are additional features. Unlike entropy reduction schemes, ADPCM does not offer bit-for-bit transparency through the system.

The biggest drawback of ADPCM systems is the fact that the typical 4:1 compression ratio limits their usefulness. For example, 384 kbps is required for a stereo 20 kHz feed.

Perceptual Coding

Perceptual coding presented a breakthrough in the quest to achieve high fidelity at ever lower bit rates. Experimental psychologists have known for years that there are many cases where the human perceptual system ignores auditory information presented to it. Scientists call this group of phenomenon “masking”. There is considerable debate as to whether this is due to basic “flaws” in the system, or due to an evolutionary value of screening out certain “unimportant” information to make the brain's job easier at processing the rest.

In any case, this aspect of our perceptual system allows a “loophole” that can be exploited by those seeking to design bit rate reduction techniques. The perceptual coder must, in real time, determine what audio information will be inaudible by the listener's perceptual system, and remove it from the encoded bit stream. Therefore, a core component of a perceptual coder is its “model” of the human perceptual system.

In addition, various techniques to exploit the redundancy in typical stereo material can achieve improved fidelity of stereo source material at a given compression ratio (see the references for more on the subject of perceptual coding).

Remarkably, the sophisticated algorithms designed to do this can easily achieve very high fidelity at compression ratios of 10:1 - with only 10% of the original audio information intact. Without this technology, cost effective audio transmission over ISDN and audio streaming over the Internet would be impossible.

The latest perceptual coders, such as MPEG AAC (Advanced Audio Coding, also known as “MP4”), can achieve at 128 kbps (stereo) quality rated as “indistinguishable” using the ITU standard testing procedures. Moreover, “near CD” quality can be achieved at compression ratios as high as 16:1!

SPECTRAL BAND REPLICATION

Spectral Band Replication, or “SBR”, is the most recent tool available in the bit rate reduction arena. Developed by Coding Technologies (www.codingtechnologies.de), this technique works together with a perceptual coder to improve performance by 30%. Therefore, SBR technology will always be seen in the context of another coding scheme. For example, “mp3PRO” is MPEG Layer 3 with the SBR enhancement added. “aacPlus™” is Coding Technologies' trademark for their implementation of MPEG AAC (MP4) with SBR added.

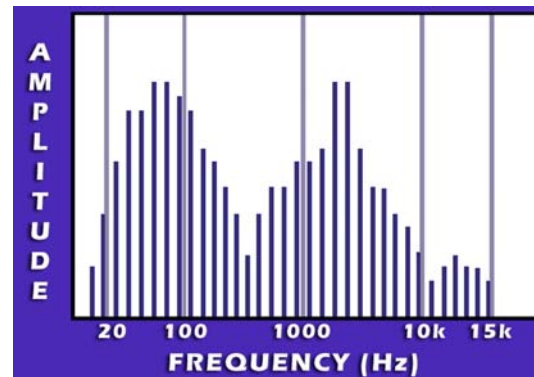


Figure 1: To transmit the above 15 kHz audio would require a sampling rate of 32 kHz. The raw data rate (mono) would be 32 k samples per second x 16 bits per sample = 512 k bits per second. If the target bit rate were 21.6 k bits per second, this would require a 24:1 compression ratio (512 / 24). Even the best codec (AAC) would sound bad at such a high ratio.

As discussed above, reducing the sampling rate by half will reduce the compression ratio by half as well. For example, changing from a 48 kHz sample rate to a 24 kHz sample rate will generate half the amount of raw data, thereby achieving a given target bit rate with half the compression ratio. The problem with this approach, of course, is that you'll have a frequency response of 12 kHz at best.

Next lets look at the bit savings incurred if we only need to transmit the audio spectrum below 7 kHz.

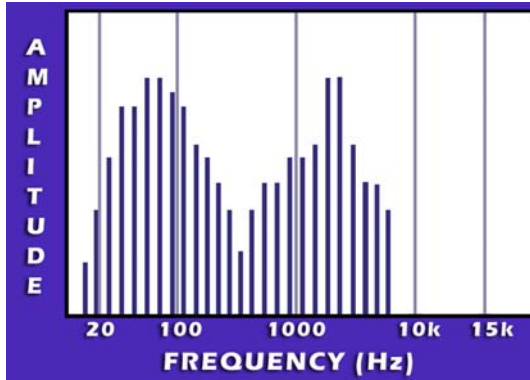


Figure 2: To transmit the 7 kHz audio shown above only requires a sample rate of 24 kHz. Since we are only transmitting the audio below 7 kHz this number is further reduced. If the target bit rate were 21.6 kbits per second, the compression ratio would be 10:1.

To transmit 10 kHz audio only requires a sample rate of 24 kHz. The raw data rate (mono) would be 24 k samples per second x 16 bits per sample = 384 k bits per second (for 12 kHz audio). Since we are only transmitting the audio below 7 KHz we multiple this number by .58 = 223 k bits per second. With a target bit rate of 21.6 k bits per second, the compression ratio would be 10:1 (223/21.6). At this reduced compression ratio, even Layer 3 could yield good results and AAC will yield excellent performance.

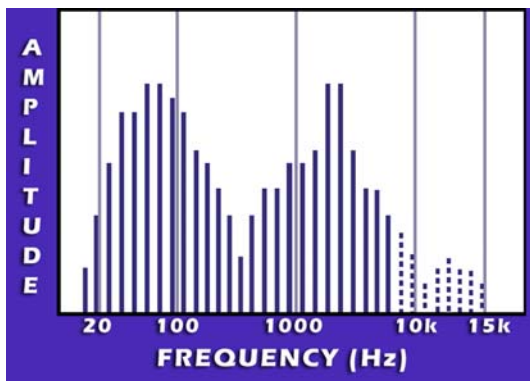


Figure 3: The combined output of the perceptual coder and the SBR process. All material above 7 kHz is generated via Spectral Band Replication

Of course, SBR restores the missing high frequency material using its replication process, and therefore full 15 KHz bandwidth can be achieved in

our examples. The final audio spectrum is a combination of the output of the perceptual coder and the SBR process as shown in Figure 3 above.

Of course the actual compression ratio of the AAC + SBR combination is over 20:1, a high ratio unobtainable by any other means.

This process is akin to the old “high frequency enhancers” made by Aphex and BBE. However, the addition of sophisticated digital algorithms and a small amount of guidance data allows for remarkably accurate reconstruction of both harmonic and “noise-like” high frequency materials.

One decided advantage of Spectral Band Replication is that it easily offers backwards compatibility with the core coder, albeit with reduced audio bandwidth. This is accomplished by embedding the guidance data into the ancillary data mechanism of the host encoder bitstream, thereby maintaining the standard data-framing format.

AACPLUS™ (AAC + SBR)

The combination of MPEG AAC (sometimes referred to as “MP4”) and SBR is referred to by the trademark aacPlus™. The MPEG-4 standard refers to this combination as “High Efficiency AAC” (HE AAC). aacPlus™ is used in the Telos Xport POTS terminal, by XM Satellite Radio, and has been chosen by the DRM Consortium for terrestrial broadcasting at 30 MHz and below. In tests performed by the DRM consortium, aacPlus at 24 kbps out performed AAC at 32 kbps. Similar results were found by tests performed by the European Broadcast Union (EBU).

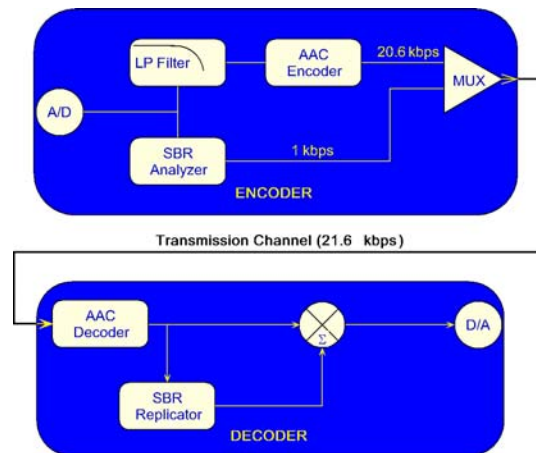


Figure 4: The aacPlus™ system. Note that while a discrete multiplexer (MUX) is shown as part of the encoder, the usual practice is to embed the guidance data into the ancillary data stream of the AAC encoder to achieve backwards compatibility.

Let's use a 15 kHz POTS codec operating at 21.6 kbps as an example to demonstrate how this works.

The AAC encoder is operated at a 24 kHz sample rate, configured for operation at a bandwidth of 7 kHz. The AAC is therefore operating at a compression ratio of 10:1, well within its normal

operating abilities, yielding high quality 7 kHz audio at the decoder. The SBR analyzer also generates a small amount of guidance data, typically about 1 kbps.

At the decoder the AAC decoder works in the usual fashion. SBR then takes the decoded 7 kHz audio, plus the guidance data, and “replicates” the audio from 7 to 15 kHz. The process is shown in Figure 4.

CONCLUSIONS

Spectral Band Replication can improve the coding efficiency of virtually any coding algorithm, reducing required bit rates by 30%. Its ability to produce clean high frequency material is particularly complementary to perceptual encoders, which tend to have high frequency artifacts present when operated at higher compression ratios. The development of this technology promises to drive a new level of applications based on its significantly improved coding efficiency.

REFERENCES

“CT-Aacplus; A State Of The Art Audio Coding Scheme”, M. Dietz & S. Melzer, Coding Technologies. EBU Technical Review, July 2002 (www.ebu.ch/trev_291-dietz.pdf).

“On Beer and Audio Coding: Why Something Called AAC is Cooler Than a Pilsner, and How It Got To Be That Way”, S. Church, Telos Systems (www.telos-systems.com/techtalk/aacpaper_2/).

“Audio & Multimedia MPEG-2 AAC”, Fraunhofer Institute IIS (www.iis.fraunhofer.de/amm/techinf/aac/).

“Facts About MPEG Compression”, Telos Systems (www.telos-systems.com/techtalk/mpeg/).

“Facts About MPEG AAC”, Telos Systems (www.telos-systems.com/techtalk/aac/).