ON BEER AND AUDIO CODING

Why something called AAC is cooler than a fine pilsner, and how it got to be that way

By Steve Church, Telos Systems.

Munching on bratwurst and sipping a thick German beer, I first heard about something supposedly new and exciting called NBC. Now this was weird for two reasons: 1) NBC has been around since Mr. Sarnoff created it many decades ago, and 2) I was sitting in a biergarten near Nürnberg and talking with guys from Fraunhofer, the outfit that invented MP3 – not people particularly likely to be familiar with American television networks. Turned out this NBC stood for *Non-Backward Compatible* and referred to next-

Will users notice anything? Yes, they said, "just wait and see, this new stuff will be really something…"

generation MPEG stuff – 'nother thing altogether. And what it meant more specifically was that the clever Fraunhofer engineers had been turned loose to make the best audio codec possible.

This was 1995, two years after the intro of the Zephyr, which was becoming quite a hit among broadcasters who appreciated its fruitful marriage of MPEG Layer 3 and ISDN. 'MP3' was first getting noticed on the Internet at this time, too. So, naturally, this was potentially interesting news. What could be better, I asked, than Layer 3? Already, it seemed to me, we had what we needed. Layer 3 was a perfect partner to ISDN, offering plenty good fidelity on widely available Telco lines. Will users notice anything? Yes, they said, "just wait and see, this new stuff will be really something…" Already, they told me, they had lined up cooperation with Sony, Dolby, and AT&T – so it was pretty clear they were onto something.

Before, with MP3, they had been constrained by a number of things. Part of the filter bank had to be the same as MPEG Layer 2. The bit stream had to be more-or-less compatible with older formats. DSP power had been expensive. But now it was clear that the price of processing was coming down swiftly according to Moore's Law, and that it would soon be possible to do much more sophisticated calculations in real time than was feasible early in the decade. And more was being learned about audio coding every day

as people working with the technology experimented, learned, and progressed to more sophistication.

Before the stein was downed, I agreed that we should work together to get this new coding method into our next-generation Zephyrs. It took some time, but the payoff has finally arrived with the new Zephyr Xstream family – here, now. Not only does the new Zephyr have the 'NBC' codec – now called MPEG-4 AAC (*for Advanced Audio Coding*), but the very interesting and useful offshoot, AAC-LD. The LD stands for Low Delay – and it lives up to the promise, enabling smooth interaction like never before possible.

High fidelity from and to most of the world over cheap and generally available ISDN lines has been a dream realized. Maybe you remember what is used to be like. Remotes used to mean special "broadcast loops" that were installed only from one fixed point to another, having months long lead times and high cost. Two Telco technicians were usually occupied for hours manually equalizing the circuits. Long distance remotes were a near impossibility: The only vendor was AT&T, and only for very expensive circuits that connected only to a single fixed point, had crazy lead times, and marginal quality. Because of the cost, links with bandwidth reaching above 5 kHz were rare. For a short time, rented satellite uplinks mounted to trucks were being driven around the country in order to get around the Ma Bell confines. While this was better, it was still expensive, had long lead times, and was burdened by the sometimes difficult requirement to find a place to situate the truck for an unobscured shot to the bird. (I remember a *Rockline* remote out of Cleveland where we couldn't get a shot from the studio and had to get a phone loop to the transmitter site *fast* – or lose the John Mellencamp guest appearance that had been promoted heavily. We got the line going about 45 minutes before show time. Talk about a nailbiter! One of the inspirations for the creation of the Zephyr…).

Around the time the first CDs were being shipped, proposals for what has become modern audio coding were greeted with suspicion and disbelief. There was widespread agreement that it would *simply not be possible* to satisfy 'golden ears' while deleting 80% or more of the digital audio data. Furthermore, MPEG – *Moving Pictures Experts Group*, true to its name, was focused almost exclusively on video compression projects. But the audio coding pioneers were persistent and an audio group was formed within MPEG. Since 1988, they have been working on the standardization of high quality audio coding. Today almost all agree not only that audio bitrate reduction is effective and useful, but that the MPEG process has been successful at picking the best technology and encouraging compatibility across a wide variety of equipment.

Researchers who have decided to work within MPEG are dedicated to creating standard, widely usable, top-quality audio and video encoders and decoders, preempting what may become an unmanageable tangle of formats. It seems to be working. Despite persistent attempts to lock users into proprietary schemes, by far the most popular high fidelity audio codecs are developed and offered as standard under the MPEG umbrella.

The main reason MPEG has been effective in finding the best technology is that the process is open and competitive. A committee of industry representatives and researchers meet to determine goals for target bit rate, quality levels, application areas, testing procedures, etc. Interested developers that have something to contribute are invited to submit their best work. A careful double-blind listening test series is then conducted to determine which of the entrant's technologies delivers the highest performance.

The subjective listening evaluations are carried out at various volunteer organizations around the world that have access to both experienced and inexperienced test subjects. Broadcasters are common participants, with many of the important test series conducted at the BBC in England, the CBC and CRC (Communications Research Centre) in Canada, and NHK in Japan.

In 1992, under MPEG-1, this process resulted in the selection of three related audio coding methods, each targeted to different bit rates and applications. These are the famous Layers: 1, 2 and 3. As the layer number goes up, so does performance and implementation complexity. Layer 1 is not much used. Layer 2 is widely used for DAB in Europe, audio for video, and broadcast delivery systems. Layer 3 is widely used in broadcast codecs and has gone on to significant Internet and consumer electronics fame under the moniker derived from the file extension, MP3.

(Forgive, please, a moment's lapse of modesty. Telos was the first to license and use MP3 commercially and some have credited (or blamed) us with getting the whole MP3 thing rolling through our promotion and our posting of an early PC-based player on the Zephyr website. At one time, we were getting more than 10,000 downloads per day!)

MPEG-2 opened the door for new work, and some minor enhancements were added to both

> In 1992, under MPEG-1, this process resulted in the selection of three related audio coding methods, each targeted to different bit rates and applications.

Layers 2 and 3. In 1997, the first in the AAC family was added to the MPEG-2 standard.

MPEG-4 audio, finalized in late 1999, adds some enhancements to AAC and adds the new AAC-LD codec.

MPEG-7 work is underway now. MPEGs-3, 5, and 6 have been skipped for rather strange reasons.[1]

There has been a lot of confusion regarding the naming of MPEG codecs. For one, the word *layer* probably made sense to the developers because the codecs under MPEG-1 & 2 are layered in the sense that the higher-numbered codecs build upon the previous ones. But to users, the naming is certainly a little strange – perhaps *levels* would have been better. And then there is the confusion resulting from the conflation of *MPEG-2* with *Layer 2*. All of the layers are subsets of MPEG-1, 2, and 4, so the full correct name for MP3 is *MPEG-2 Layer 3 audio,* for example. Already, some people are referring to AAC as MP4. Guess the logic here is that it is the next step up for Internet audio from MP3, and it *is* part of MPEG-4…

Perceptual coding: The miracle of acoustic masking

All of the MPEG codecs rely upon the celebrated acoustic masking principle – an amazing property of the human ear/brain aural perception system. When audio is present at a particular frequency, you cannot hear audio at nearby frequencies that are sufficiently low in volume. The inaudible components are masked owing to properties of the human ear that occur at a very low 'hardware' level – researchers say the information is dropped immediately within the ear and is not passed to the brain. This appears to be a kind of natural rate reduction that helps to keep the brain from being overloaded with unnecessary information. There is a similar effect working in the time domain, with signals coming soon after the removal of another being also inaudible.
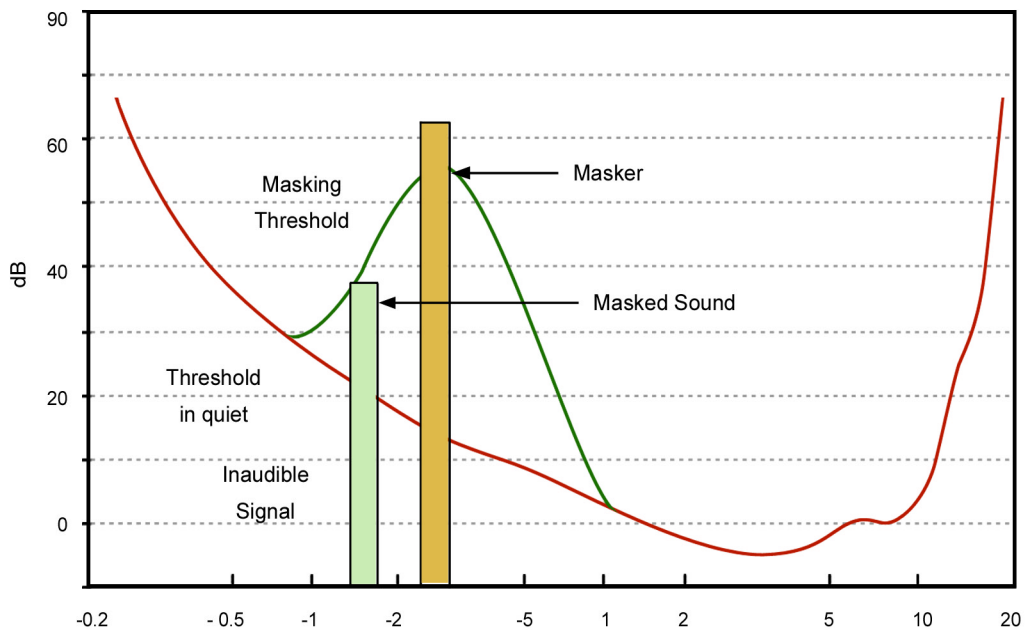
The acoustic masking principle in the frequency domain. Here the masker at 300 Hz causes the 150 Hz signal to be inaudible.
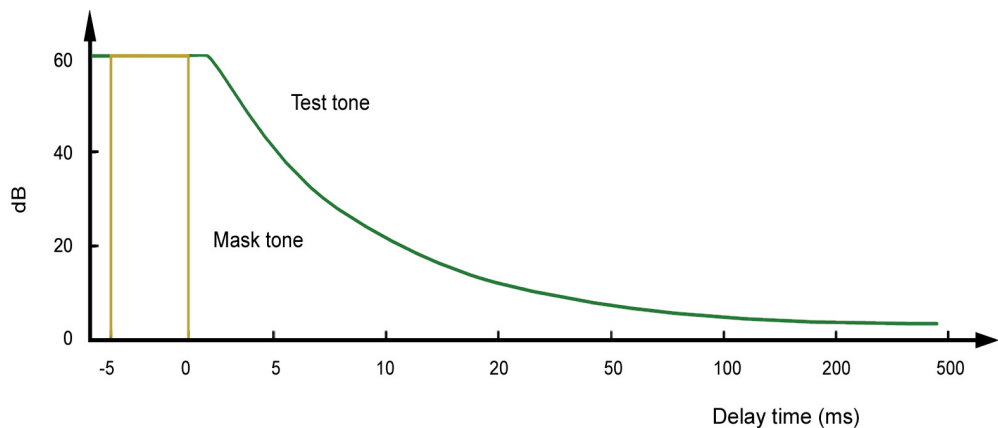
The masking principle in the time domain. For some time after the removal of the masking signal, other signals falling below the curve are inaudible.
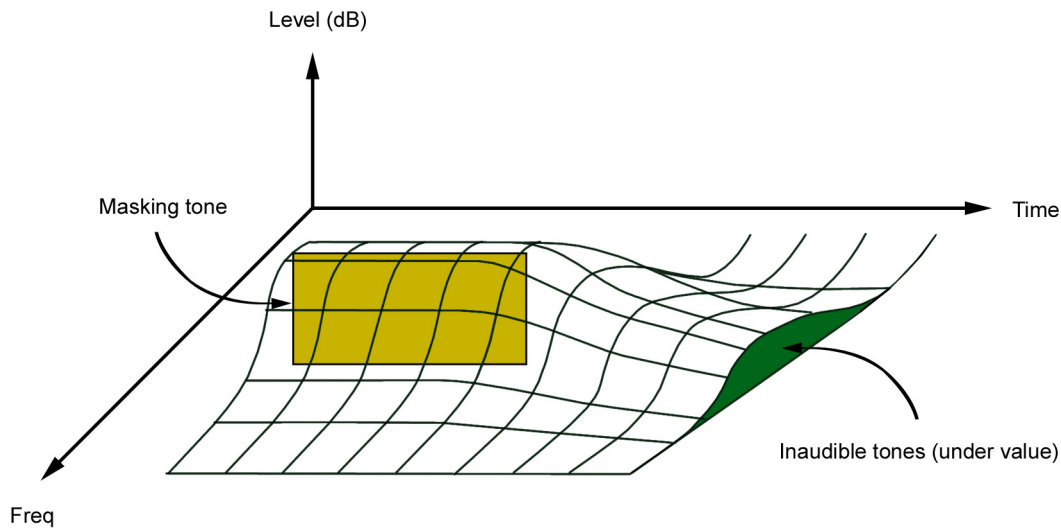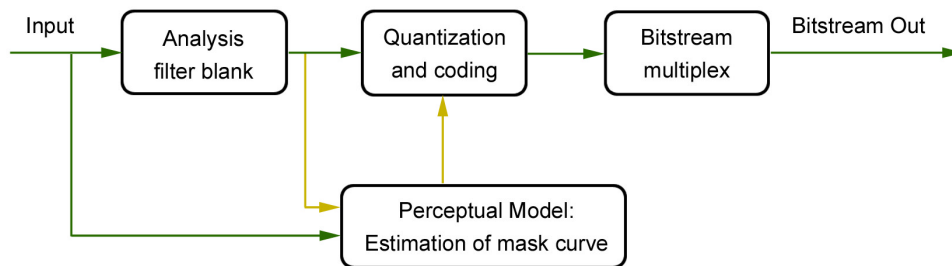
The combined result of frequency and time masking. Signals under the curve are inaudible.

As a result, it is not necessary to use precious bits to encode these masked frequencies. In perceptual coders, a filter bank divides the audio into multiple bands. When audio in a particular band

Perceptual coder basic block diagram.

falls below the masking threshold, few or no bits are devoted to encoding that signal, resulting in a conservation of bits that can then be used where they are needed.

While various codecs use different techniques in the details, the principle is the same for all, and the implementation follows a common plan. There are four major subsections, which work together to generate the coded bitstream:
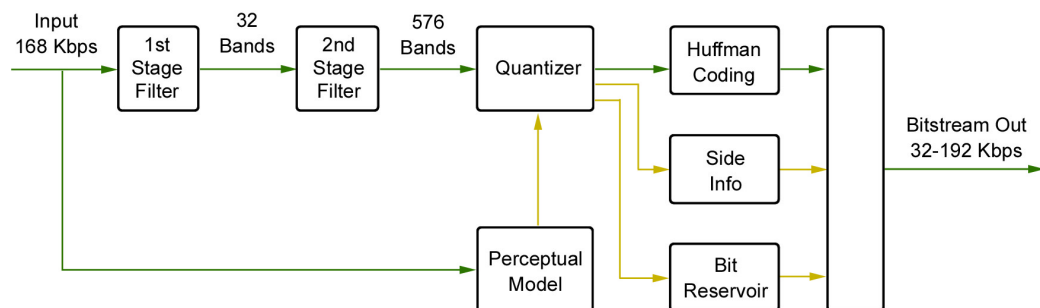
- The *analysis filter bank* divides the audio into spectral components. At minimum, sufficient frequency resolution must be used in order to exceed the width of the ear's critical bands, which have widths of 100 Hz below 500 Hz and roughly 20% of the center frequency at

higher frequencies. Finer resolution can help a coder make better decisions.

- The *estimation of masked threshold* section is where the human ear/brain system is modeled. This determines the masking curve, under which noise must fall.

- The audio is reduced to a lower bit rate in the *quantization and coding section*. On the one hand, the quantization must be sufficiently course in order not to exceed the target bit rate. On the other hand, the error must be shaped to be under the limits set by the masking curve.

- The quantized values are joined in the *bitstream multiplex*, along with any side information.

MPEG-2 Layer 3 (MP3)

MPEG-2 Layer 3 is probably the most popular audio codec in the world. So let's have a closer

Layer 3 uses a 2-stage filter and adds Huffman coding to the basic perceptual coder principle.

look at the process of the MP3 coding.

The main enhancement to the basic encoder principle is the Huffman coding section. This process causes values that appear more frequently at the quantizer output to be encoded with shorter words, while values that appear only rarely are coded with longer words. This is similar to the common PC zip-style compression and results in an increase in coding efficiency with no degradation since it is a completely lossless process.

Another interesting and novel idea in Layer 3 is the bit reservoir buffering. Often, there are some critical parts in a piece of music that cannot be encoded at a given data rate without audible noise. These sequences require a higher data rate to avoid artifacts. On the other hand, some signals are easy to code. If a frame is easy, then the unused bits are put into a reservoir buffer. When a frame comes along that needs more than the average amount of bits, the reservoir is tapped

for extra capacity. The bit reservoir buffer also offers an effective solution for the inclusion of such ancillary data as text or control signaling. The data is held in a separate buffer and gated onto the output bitstream using some of the bits allocated for the reservoir buffer when they are not required for audio.

There is a Joint Stereo mode that takes advantage of the redundancy in stereo program material. The encoder switches from discrete L/ R to a matrixed L+R/ L-R mode dynamically, depending upon the program material.

The two-stage filterbank was used in order to have some measure of compatibility with Layer 2, which has only the first section.

Compared to Layer 2, Layer 3 is both more complicated and more powerful. Most of the improvement comes from having more frequency resolution – 576 bands vs. Layer 2's 32. The Huffman coding section requires memory to be dedicated to a look-up table, but otherwise imposes little penalty.

MPEG-2 & 4 AAC

The idea that led to AAC was not only to start fresh, but also to combine the best work from the world's leading audio coding laboratories. Fraunhofer, Dolby, Sony, and AT&T were the primary collaborators that offered components for AAC. The hoped for result was ITU (International Tele-

communications Union) "indistinguishable quality" at 64 kbps per mono channel. That is, quality indistinguishable from the original, with no audio test item falling below the "perceptible, but not annoying" threshold in controlled listening tests.

The MPEG test items include the most difficult audio known to codec researchers, so this was a daunting challenge. The thinking was that if a codec could pass this test, it would surely be transparent for normal program material like voice and pop music, which are much more easy to encode.

AAC designers chose to use a new modular approach for the project, with components being plugged-in to a general framework in order to match specific application requirements and the always present performance/complexity tradeoffs.

- An improved filter bank with a frequency resolution of 2048 spectral components, nearly four times than for Layer 3.

- Temporal Noise Shaping, a new and powerful element that minimizes the effect of temporal spread. This benefits voice signals, in particular.

- A Prediction module guides the quantizer to very effective coding when there is a noticeable signal pattern, like high tonality.
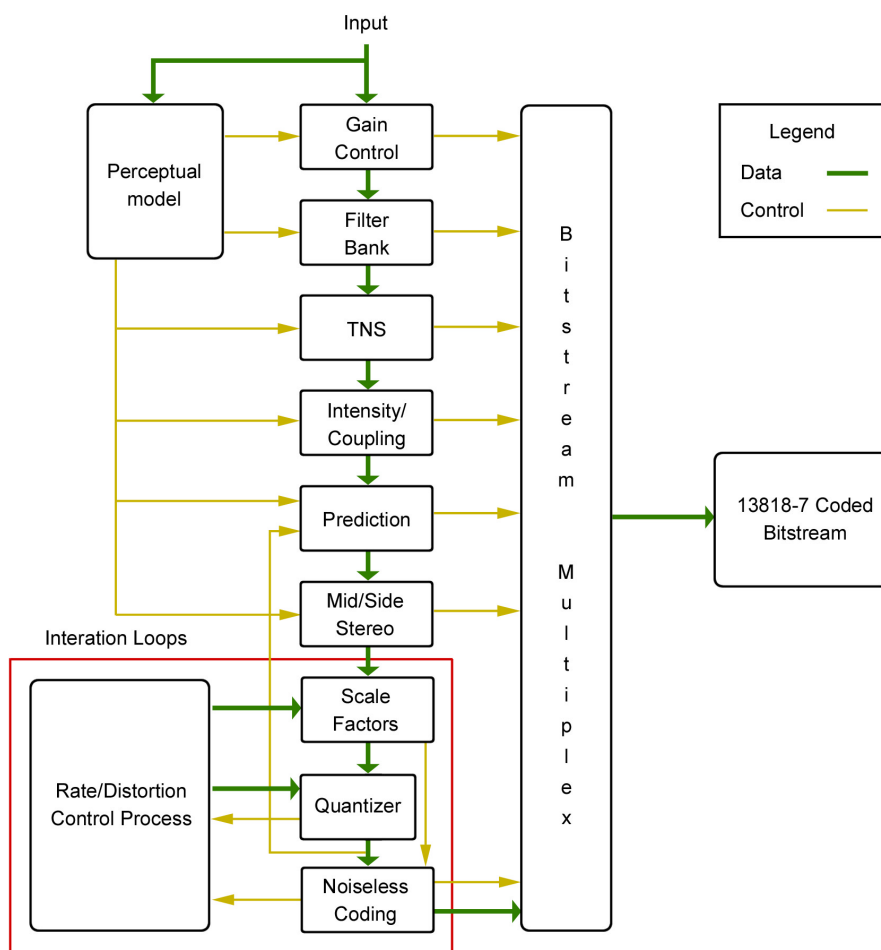
AAC is a lot more sophisticated than the previous MPEG layers 2 & 3, providing significantly more coding power. Because it uses a modular approach, new ideas can be developed and plugged in to the basic structure.

This had the additional advantage that it was possible to combine the various components from different developers, taking the best pieces from each.

AAC was built on a similar structure to Layer 3, and thus retains most of its features. But compared to the previous MPEG layers, AAC benefits from some important new additions to the coding toolkit:

- Perceptual Noise Shaping allows a finer control of quantization resolution, so bits can be used more efficiently.

The result of all this is that the researchers succeeded: AAC provides performance superior to any known codec at bitrates greater than 64 kbps and excellent performance relative to the alternatives at bitrates reaching as low as 16 kbps.

This was confirmed in two series of tests conducted in 1997, the first jointly at the BBC in England and NHK in Japan, the second at the CRC Signal Processing and Psychoacoustics Audio Perception Lab. [3]

The tests conducted by CRC for MPEG were among the most extensive and thorough ever. Even the selection of audio test materials was a careful process, essential for unbiased codec evaluation, since each has particular strengths and weaknesses.

Consider this excerpt from the CRC paper to get some flavor of the careful work to find audio that would best reveal the limitations of the codecs:

> The selection of critical materials was made by a panel of 3 expert listeners over a period of 3 months. The first step in the process was to collect potentially critical materials (materials which were expected to audibly stress the codecs). Sources of materials included items found to be critical in previous listening tests, the CRC's compact disc collection and the personal CD collections of the members of the selection panel. Also, all of the proponents who provided codecs for the tests were invited to submit materials which they felt to be potentially critical. In order to limit the number of materials to be auditioned by the selection panel, an educated pre-selection was made regarding which materials were most likely to stress the codecs. This pre-selection was based on our knowledge of the types of materials which have proven to be critical in previous tests, as well as an understanding of the workings of the codecs and thus their potential limitations. Also, consideration was given to providing a reasonable variety of musical contexts and instrumentations.
>
> A total of 80 pre-selected audio sequences were processed through each of the 17 codecs, giving the selection panel 1360 items to audition. The selection panel listened together to all 1360 items to find at least two stressful materials for each codec. The panel agreed on a subset of 20 materials (of the 80) which were the most critical ones and which provided a balance of the types of artifacts created by the codecs. A semi-formal blind rating test was then conducted by the members of the selection panel on these 340 items (20 materials x 17 codecs). The results of the semi-formal tests were used to choose the final 8

> critical materials used in the tests. In selecting the critical materials, consideration was given to highlighting various types of coding artifacts, The two most critical materials for each codec were included in these 8 materials.

The audio samples the CRC selected using this process were, not surprisingly, quite varied:

Bass clarinet arpeggio, EBU SQAM CD

Bowed double bass, EBU SQAM CD

Dire Straits, Warner Bros. CD 7599-25264- 2 (track 6)

Harpsichord arpeggio, EBU SQAM CD

Music and rain, AT&T mix

Pitch pipe, Recording from Dolby Laboratories

Muted trumpet, Original DAT recording, University of Miami

Susan Vega with glass, AT&T mix

Then the comparative evaluations began, using double-blind procedures so that participants were not able to know what codecs were being used for which samples. 24 people participated in the tests, mostly taken from groups where it was expected that listeners with high expertise would be found. The subjects included 7 musicians of various kinds (performers, composers, students), 6 recording and broadcast engineers, 3 piano tuners, 2 audio codec developers, 3 other types of audio professionals, and 3 persons from the general public.

The results were remarkable for AAC. At 96 kbps, it give comparable quality to Layer 2 at 192 kbps and to Layer 3 at 128 kbps. The researchers concluded that there was a clear performance distinction among the various codecs:

> The results show that the codec families are clearly delineated with respect to quality. The ranking of the codec families with respect to quality is: AAC, PAC, Layer 3, AC-3, Layer 2, and ITIS (a Layer 2 implementation). The highest audio quality was obtained for the AAC codec operating at 128 kbps and the AC-3 codec operating at 192 kbps per stereo pair.
>
> The following trend is found for codecs rated at the higher end of the subjective rating scale. In comparison to AAC, an increase in bitrate of 32, 64, and 96 kbps per stereo pair is required for the PAC, AC-3, and Layer 2 codec families respectively to provide the same audio quality.

Finally, the CRC study concluded that AAC achieves the ITU "indistinguishable quality" goal – the first codec to fulfill this requirement at 128 kbps/stereo:

The AAC codec operating at 128 kbps per stereo pair was the only codec tested which met the audio quality requirement outlined in *the ITU-R Recommendation BS.1115 for perceptual audio codecs for broadcast.*

AAC at 128 kbps/stereo measured higher than any of the codecs tested. It has approximately 100% more coding power than Layer 2 and 30% more power than the former MPEG performance leader, Layer 3.
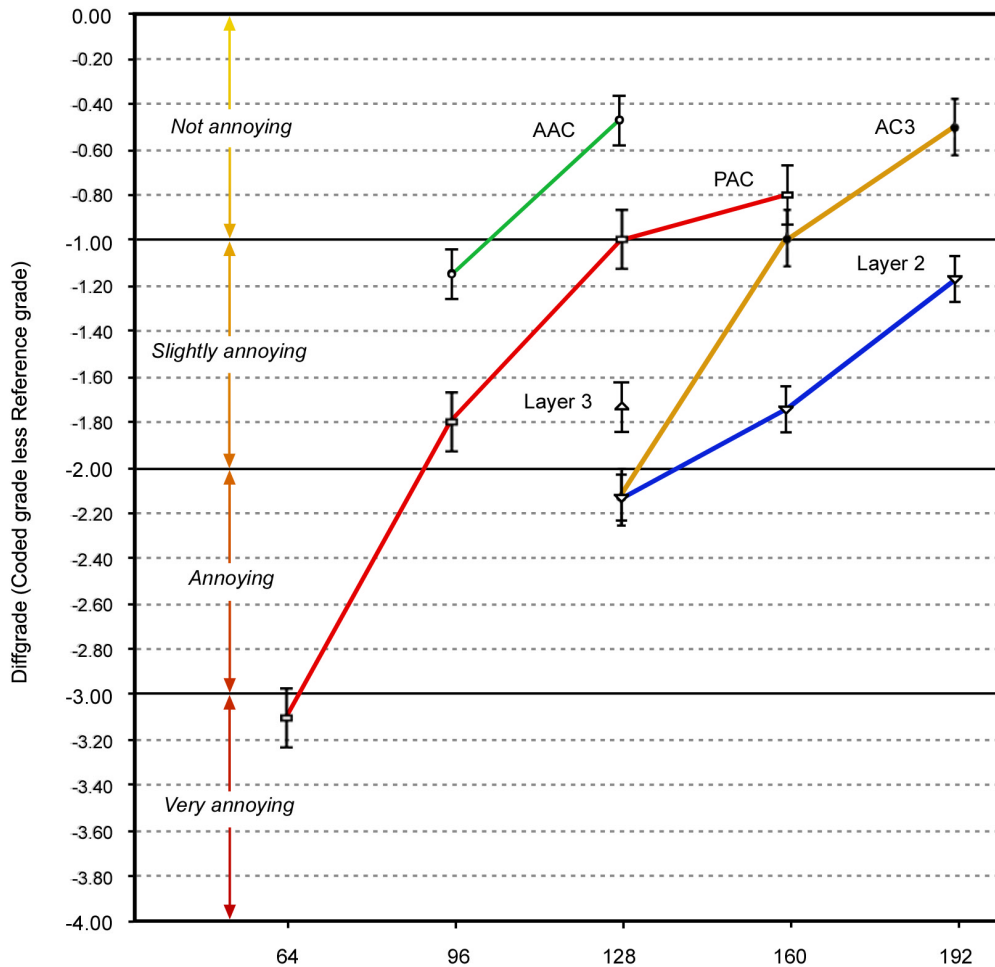
Comparison of overal quality for MPEG, PAC (Lucent) and AC3 (Dolby) stereo codecs at their recommended target bit rates from 96-192 kbps. (from the CRC test results in [3])

MPEG-4 Low Delay AAC

An important topic for many real world codec applications is delay. When announcers use codecs for a broadcast remote application, they often need to have natural two-way interaction with other program participants located back at the studio or callers via telephone lines.

Because it is a hot topic for engineers working in the field of Internet telephony, a number of studies have been conducted to determine user's reactions to delays in telephone conversations. The data apply directly to the application of pro-fessional codecs to remotes, so it is interesting to take a peak over the shoulder of the telecom guys to see what they have learned.

For broadcast remotes, we try to arrange our system so that there is no path for the field announcer's voice to return to his headphones. But sometimes echo is unavoidable. For example, this can happen when a telephone hybrid has leakage or when a studio announcer has open-air headphones turned-up loud and the audio makes its way into the studio microphone.

When there is no echo, it has been discovered that anything less than 100 ms one-way delay permits normal interactivity. Between 100 and 250 ms is considered "acceptable." ITU-T stan-

for speech signals – but cannot be used for music or mixed signals that include voice and ambient sounds.
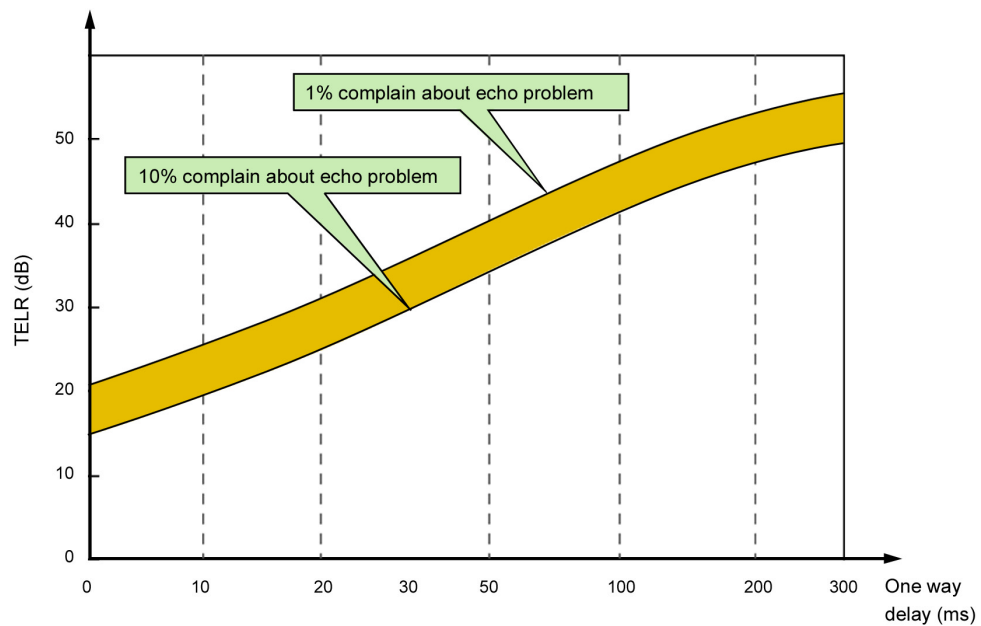
Summary of ITU-T G.13, with recommendations for designers of telephone systems that must cope with echo.

dard G.114 recommends 150 ms as the maximum for "good" interactivity.

Echo introduces a different case. As you might expect, echo is more or less annoying depending upon both the length of time it is delayed and its level. Telephone researchers have measured and quantified reactions, and ITU-T G.131 reports the findings and makes recommendations.

There are codecs using other than perceptual technologies that have lower delay, but they are not as powerful. That is, for a given bitrate, they do not achieve fidelity as good as the MPEG ones we have been examining. The common G.722 is an example. It uses ADPCM (Adaptive Delta Pulse Code Modulation), which can have delay as low as 10 ms, but with much poorer quality. So the question arises: Is it possible to have high quality and low delay in the same codec? Until recently, the answer was no. But new developments in codecs have changed the picture.

One of the main objectives in audio coding is to provide the best tradeoff between quality and bit rate. In general, this goal can only be achieved at the cost of a certain coding delay. Codecs for voice telephone applications have use ADPCM and CELP because they have much lower delay than perceptual codecs. These are optimized for voice and can have reasonably good performance

How AAC-LD Gets its Low Delay

The new AAC-LD uses new techniques, some only very recently discovered, in order to offer both low delay and high fidelity. Compared to speech coders, AAC-LD handles both speech and music with good quality. Unlike speech coders, however, audio quality scales up with bit rate, and transparent quality can be achieved.

Perceptual codec designers must manage several tradeoffs. Most important is the number of samples coded together in one frame. Long frames have high delay but are more efficient because the header and side information is transmitted less frequently. Longer frames offer the possibility to use filter banks with better frequency resolution. A fundamental principle in signal processing is that spectral splitting filters may have either good time resolution, or good frequency resolution, but not both. This makes sense when you consider that a longer time window means that the analyzer has more complete information, more full audio cycles, to work with. (Perhaps this is the DSP designer's equivalent to the economist's *TANSTAAFL* – There Ain't No Such Thing As A Free Lunch).

In the case of rapidly changing input signals (transients) long frames are not as good as short

ones because the time spread will lead to so-called *pre-echoes*. For such signals, the size of the frame should correspond to the temporal res-

The result: AAC-LD is clearly better than Layer 3 for half of the test items, and as good for the remaining half.

olution of the human ear. This can be achieved by using short frames or by changing the frame length according to the immediate characteristics of the signal.

In addition to frame length, delay in perceptual codecs is dependent upon filter bank delay, the look-ahead delay for block switching, and the time requirements of the bit reservoir buffer. The overall delay is a combination of all of these components divided by the sampling rate, and scales linearly and inversely with the sampling frequency.

Layer 3 and AAC use filter banks with high

is a broad sine-shaped curve, but AAC-LD can dynamically switch to a shape that has a lower overlap between the bands. This significantly improves performance with transients, without adding any delay.

- Problems with transients and pre-echoes are also handled by the new Temporal Noise Shaping module.

The result is delay that falls well below the 100 ms required for natural conversation.

Amazing Performance

Low delay would not be useful if the quality was not acceptable. So how does AAC-LD stack-up? Because most codec users are familiar with Layer 3, listening tests were performed to compare AAC-LD to it at the standard single-channel ISDN 64 kbps rate. The result: AAC-LD
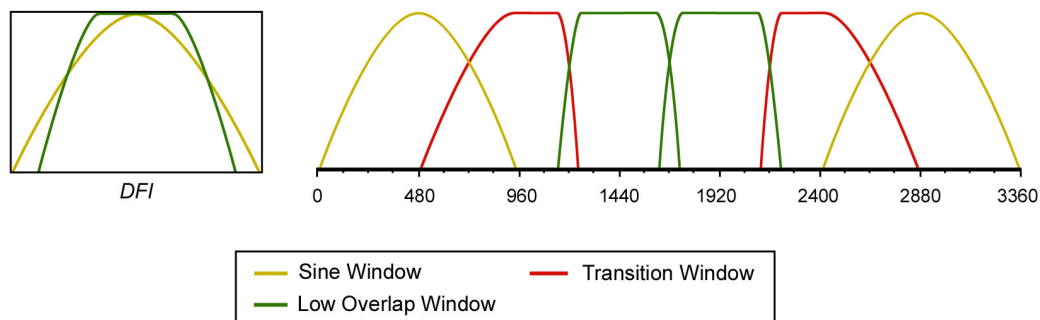
A dynamically adaptive window shape is used in the AAC-LD analysis filter bank to minimize pre-echo and other time-related effects. Normally the window shape has a broad curve, but it is switched to a sharper response when the audio material can benefit from the change.

frequency resolution. But when there are transients, a block switching process shifts to a filter bank with lower frequency resolution and better time resolution. In order to correctly decide when to make this change, a look-ahead process is required, which is a significant cause of delay.

AAC-LD is based on the core AAC work, so much is similar, but each of the contributors to the delay were addressed and modified:

- The frame length is reduced to 512 or 480 samples, with the same number of spectral components at the filter bank output.

- The "window shape" of the spectral filter is enhanced to be adaptive. Normally, the shape

is clearly better than Layer 3 for half of the test items, and as good for the remaining half.

AAC-LD's coding power is roughly the same as Layer 3, meaning that mono high fidelity 15 kHz audio may be sent via one ISDN channel. With ISDN's two channels, you can have near CD quality stereo. Since most mono remote broadcasts are speech, you can expect audio quality even better than with the familiar – and already quite good – Layer 3.
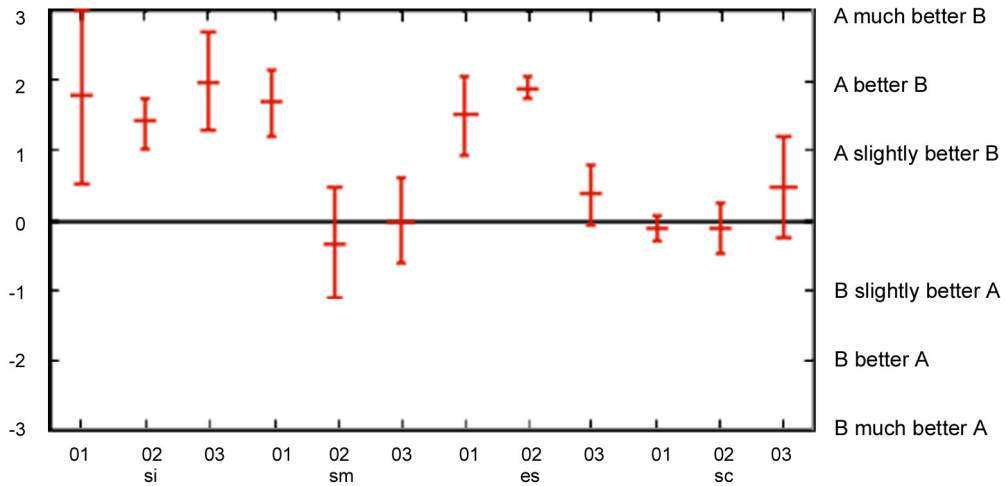
AAC-LD compared to Layer 3 at 64 kbps mono. The test items[2] es01-03 are speech and show very good quality. Pop music is the test item sc03 and performance is good here as well. The one slight trouble spot is the glockenspiel at sm02.3 [7]

So What Coding Should You Use?

The Zephyr Xstream has many coding possibilities – so which one should you use? Before AAC, the choice was usually a tradeoff between quality and delay. G.722 was lowest delay and poorest quality, Layer 2 good fidelity and medium codecs and should be used when delay has priority over fidelity. G.722 can be used when delay must be at minimum, and Layer 2 or Layer 3 for compatibility with older codecs.

The Xstream offers lower delay compared to

Coding	Bitrate	Sampling Rate	Delay
AAC-LD stereo	128 kbps	32 kHz	80 ms
AAC-LD stereo	128 kbps	48 kHz	56 ms
AAC-LD mono	64 kbps	48 kHz	50 ms
AAC stereo	128 kbps	48 kHz	172 ms
Layer 3 stereo	128 kbps	48 kHz	326 ms
Layer 2 stereo	128 kbps	48 kHz	224 ms
Layer 2 stereo	128 kbps	24 kHz (half mode)	398 ms
G.722	64 kbps	48 kHz	10 ms

Zephyr Xstream coding characteristics, including the new AAC and AAC-LD modes

delay, and Layer 3 best fidelity and most delay. Things are easier now. AAC has lower delay than Layer 2 or Layer 3 and higher quality than both, so it should be used for most applications. AAC-LD has the lowest delay of the MPEG the original Zephyr even on the legacy coding methods. This is a result of DSPs becoming more powerful and our being able to use a single one, rather than a chained set as was required before.

For stereo, these numbers also include the contribution from the ISDN channel splitter function, which adds delay because there must be a buffer to compensate for any time shift that may be encountered on the Telco line when two channels are being combined.

The Zephyr Xstream lets you independently choose the coding mode for the send and receive directions, so you can optimize for the specific requirements of the application. For example, often the return monitor can be lower fidelity than the field-to-studio direction, so you can choose a coding method that reduces round-trip delay.

That pilsner was probably the most productive drinking I've had the good fortune to experience!

Links

http://www.tnt.uni-hannover.de/project/mpeg/ audio/
Official website for MPEG audio.

http://www.cselt.it/mpeg/
Official website of the MPEG Committee.

http://www.zephyr.com
Telos Systems website, with information about the Zephyr MPEG codec family.

References

[1] ISO/IEC 13818-7 Advanced Audio Coding (AAC).

[2] ISO/IEC 14496-3:1999 Information technology – Coding of audio-visual objects – Part 3: Audio, December, 1999.

[3] G. A. Soulodre, T. Grusec, M. Lavoie, and L. Thibault, Signal Processing and Psychoacoustics/ Communications Research Centre, Ottawa, Ont.: Subjective Evaluation of State-of-the-Art 2-Channel Audio Codecs, Paper presented at the AES 104th Convention, 1998.

This paper was edited and reprinted in the Journal of the Audio Engineering Society:

G. A. Soulodre, T. Grusec, M. Lavoie, and L. Thibault. Subjective evaluation of state-of-the-art 2-channel audio codecs. *J. Audio Eng. Soc.*, 46(3):164 – 176, March 1998.

[4] ITU-R: Low Bit Rate Audio Coding, Recommendation ITU-R BS.1115.

[5] ITU-R: Methods for the Subjective Assessment Of Sound Quality – General Requirements, Recommendation ITU-R BS.1284, 1996.

[6] Karlheinz Brandenburg: MP3 and AAC Explained. Presented at the AES 17th International Conference, Florence, Italy, 1999 Sept. 2-5. Available at http://www.cselt.it/mpeg/tutorials.htm.

[7] E. Allamanche, R. Geiger, J. Herre, T. Sporer: MPEG-4 Low Delay Audio Coding Based on the AAC Codec. Paper resented at the AES 106th Convention, 1999.

[1]Many people have wondered about the strange numbering system of MPEG standards: 1, 2, 4, and, on the horizon, 7. Here is the story. Work was begun on an MPEG-3 standard for high-definition television, but it became clear that the tools needed were very similar to those in MPEG-2, so MPEG-3 was quickly abandoned, and HDTV support was included in MPEG-2. When the latest work item was started, the first question taken up was what number to use. One participant recalled that the conversation was something like, "Shall the number for the next job be 5, which follows 4, or should it be 8, attractive in its own binary way, to follow 1, 2 and 4? After some thought, MPEG members decided that their new work item was so different from what had gone before that they threw both ideas overboard and chose 7 as the lucky number."

[2] The test items this time were:

sc01	Trumpet solo & orchestra
sc02	Symphonic orchestra
sc03	Contemporary pop music
es01	English female speaker
es02	German male speaker
es03	English male speaker
sm01	Bagpipes
sm02	Glockenspiel
sm03	Plucked strings
si01	Harpsichord
si02	Castanets
si03	Pitch pipe