Radio World®

GUEST COMMENTARY Reprinted from September 26, 2001

On Beer and Audio Coding

Why Something Called AAC Is Cooler Than A Fine Pilsner, and How It Got That Way

Steve Church

The author is president of 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-generation MPEG stuff. And what it meant more specifically was that the clever engineers at Fraunhofer who made MP3 had been turned loose to make the best audio codec possible.

Before the stein was downed, I agreed that we should work together to get this new coding method into our next-generation gear.

It took some time, but the payoff has finally arrived. Not only is MPEG-4 AAC (for Advanced Audio Coding) here, but so is the very interesting and useful offshoot, AAC-LD.

The LD stands for Low Delay, and it lives up to the promise.

How MPEG works

The MPEG audio story begins in 1988. True to its name, MPEG, the Moving Pictures Experts Group, was focused almost exclusively on video compression projects. But persistent audio coding pioneers convinced the organization to allow the formation of an audio group.

Today almost all agree that the MPEG process has been successful at picking the best technology and encouraging compatibility across a variety of equipment.

Researchers who work within MPEG want to create standard, widely usable, top-quality codecs, preempting what may become an unmanageable tangle of formats. It seems to be effective. Despite persistent attempts to lock users into proprietary schemes, by far the most popular high-fidelity codecs are developed and offered under the MPEG umbrella.

Probably the main reason MPEG consistently has been successful at finding the best technology is that the process is open and competitive.

A committee meets to determine goals for target bitrate, quality levels, application areas and testing procedures. Interested developers who have something to contribute are invited to submit their best work. Finally, a careful double-blind listening test series is conducted to determine which of the technologies delivers the highest performance.

AAC arrives

Before becoming an MPEG standard, AAC was tested and compared carefully to other codecs in two series, the first conducted jointly at the BBC in England and NHK in Japan, the second at the CRC Signal Processing and Psychoacoustics Audio Perception Lab in Canada.

The tests conducted by CRC for MPEG were among the most extensive and thorough ever. The researchers concluded that there was a clear performance distinction among the various codecs, and that AAC was the best performer:

"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," according to G.A. Soulodre, T. Grusec, M. Lavoie and L. Thibault of the Signal Processing and Psychoacoustics/Communications Research Centre in Ontario, Canada, in a paper presented at the AES 104th Convention in 1998.

See Figure 1 (next page), which is adapted from that paper.

Therefore, AAC was selected as the general-purpose audio codec under MPEG-2 and recently was adopted by MPEG-4.

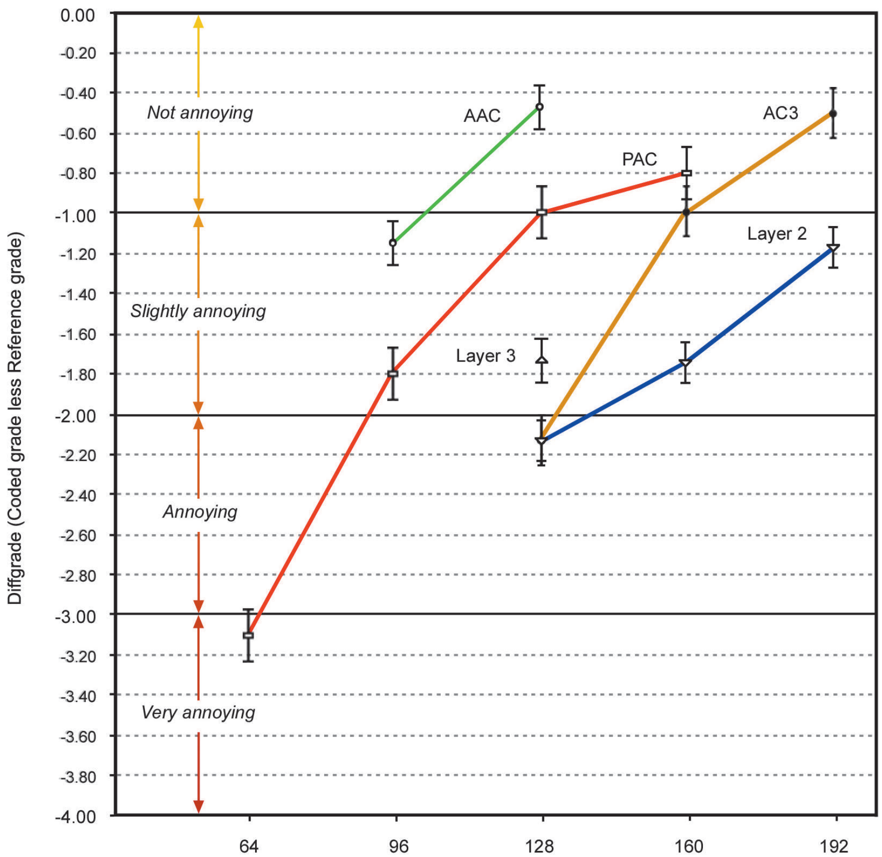

Fig. 1: This chart presents a comparison of overall quality for various stereo codecs at their recommended target bitrates. (Adapted from Soulodre, Grusec, Lavoie and Thibault.)

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

Because it is a hot subject for engineers working in the field of Internet telephony, a number of studies have been conducted to determine user reactions to delays in telephone conversations. The data apply directly to the application of professional codecs to remotes, so it may be fruitful to see what the telecom people have learned.

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

Echo introduces a different case. We try not to have echo in our broadcast setups by using mix-minus arrangements, but sometimes it is unavoidable. An open-air headset might be cranked up or a phone hybrid might have leakage.

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 recorded people's reactions, and you can see the ITU-T G.131 findings in the graph shown in Fig. 2.

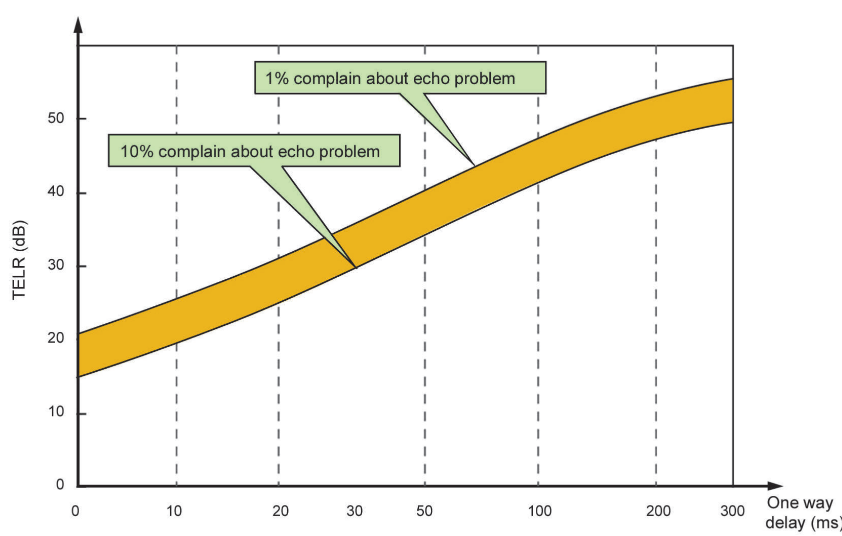

Fig. 2: Designers of telephone systems often must cope with echo. From ITU-T G.131.

The main goal with audio coding is to provide the best tradeoff between quality and bitrate. Codecs for voice telephone applications use ADPCM, the technology used by the familiar G.722, and CELP, used in mobile phones. These are optimized for low-delay and speech, but they are terrible for music.

Because they are also bad for mixed signals that include ambient sounds such as from spectators at sporting events, they are not optimum for broadcast remotes.

Is it possible to have high quality and low delay in the same codec?

Tradeoffs

Perceptual codec designers must manage several tradeoffs. The most important comes from a fundamental principle in signal processing: 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 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.)

AAC-LD uses new techniques, some freshly discovered, in order to offer both low delay and high fidelity. Layer 3 and AAC use filter banks with high-frequency resolution. But when there are transients, the encoder switches 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. In AAC-LD the shape of the spectral filter is adaptive, dynamically switching to a shape that has a lower overlap when necessary.

This and other enhancements result in delay that can reach as low as 50 ms, well within range for good conversational flow.

But how does it sound?

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, a series of tests was performed to compare AAC-LD to it at the standard single-channel ISDN 64 kbps rate. The result: AAC-LD is clearly better than Layer 3 for half of the test items, and as good for the remaining half. See Figure 3.

AAC-LD's coding power is roughly the same as Layer 3, meaning that mono 15 kHz audio may be achieved on one ISDN channel. With two channels, you can have near CD quality stereo.

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 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 perceptual 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.

Fnd more graphics and a much more in-depth version of this article at www.rwonline.com, *including links to MPEG resources on the Web.*

RW welcomes other points of view.

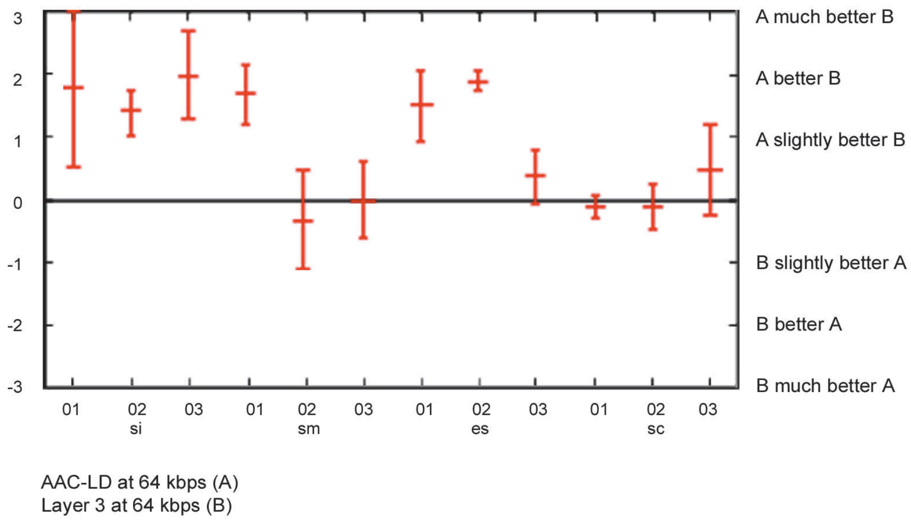

Fig. 3: AAC-LD is compared here to Layer 3 at 64 kbps mono. (Source: E. Allamanche, R. Geiger, J. Herre, T. Sporer: 'MPEG-4 Low Delay Audio Coding Based on the AAC Codec,' presented at the AES 106th Convention, 1999.)