

# Audio Quality and Netcasting

After a five-year hiatus, netcasting is once again making headlines in business publications. Thanks in part to streaming-enabled 3GPP devices, some analysts see netcasting as supplanting broadcasting, because netcasting offers a virtually limitless number of programming formats. Meanwhile, the local FM broadcasters in a given market, out of economic necessity, have to broadcast mass-appeal formats that don't fully satisfy anyone.

Netcasters have much to learn from broadcasters, however, about what is commonly called "broadcast-quality" audio. Without such understanding, netcasters cannot fully compete with traditional broadcast media. The topic of broadcast-quality audio can be divided into two main technology threads: audio processing and codecs.

## Audio Processing

Broadcasters have been accustomed to processing audio for AM and FM transmission with transmission audio processors like Orban's Optimod series. These processors compress dynamic range to make the signal comfortably listenable in noisy environments, and also to make the best use of the dynamic range limitations of the channel itself. In analog services (like FM radio), this dynamic range varies as a function of reception conditions, which are poorest in the fringes of the signal. Audio processing therefore also increases the potential coverage area of analog transmissions.

Digital transmissions behave differently. The technical specifications of the transmission system determine the signal-to-noise ratio. This does not change with the signal strength in wireless transmission (and is even more irrelevant in a wired environment). Internet reception anomalies are typically audio drop-outs rather than added noise.

What role does audio processing play in a system with a very low noise floor? It can still have several vital functions:

- First is dynamic range compression to accommodate the signal into typical listening environments like autos and homes. In

*autos, the acoustic dynamic range is severely limited by wind and road noise. In most apartments and multi-family dwellings, the available dynamic range is limited by the need to avoid disturbing family and neighbors with excessive sound levels. In public spaces like buses, subways, and airports, there is a wide variety of acoustic noise sources. There are relatively few environments where the full, uncompressed dynamic range of the original program material is useable or desirable.*

- Second is to ensure a consistent presentation. In radio, program material from different producers is constantly juxtaposed. Yet most successful broadcasters agree that achieving a "major market" sonic image requires an overall consistency of sound texture and

*shown that a combination of multiband compression and sophisticated peak limiting is the most effective way to do this.*

- The final function is to help improve the intelligibility of substandard program material, particularly news actualities and incoming telephone calls. Properly designed multiband compression like that used in Optimods can make startling improvements in this material without need for preprocessing in a production studio.

Preprocessing each program element before it is stored on a playout system is not as effective as preprocessing the mixed audio on the program line immediately before it is streamed. The latter technique maximizes the smoothness of transition

Experience has shown that a combination of multiband compression and sophisticated peak limiting is the most effective way to broadcast a louder and punchier audio signal.

*spectral balance from source to source. Multiband compression can achieve this. By setting a target spectral balance and automatically re-equalizing program material that does not have this balance, the multiband compression helps the radio station achieve a "big-time," highly produced sound that sounds authoritative to listeners.*

- Third is to reduce the peak-to-average ratio of the signal to increase its relative loudness by comparison to an unprocessed signal normalized to the same peak level. In netcasting, all signals share the same "pipe." Each signal has exactly the same reach. So the only thing that netcaster can do to stand out from his neighbors (and possible competitors) is to broadcast a louder and punchier audio signal. Experience has

between program elements and makes voice from announcers or presenters merge smoothly into the program flow, even if the announcer is talking over music.

It is important to understand that AM, FM, or TV audio processors that employ pre-emphasis/de-emphasis and/or clipping peak limiters are highly inappropriate for use with the perceptual audio coders used in netcasting. The pre-emphasis/de-emphasis limiting in these devices unnecessarily limits high frequency headroom. Further, their clipping limiters create high frequency components—distortion—that the perceptual audio coders would otherwise not encode.

An audio processor like Orban's Optimod-PC 1100 Professional PCI Sound Card consists of several cascaded stages. These include Input Conditioning, including defeatable highpass

filtering and defeatable phase rotation; Stereo Enhancement; Two-Band Gated Automatic Gain Control (AGC), with target-zone window gating and silence gating; Equalization, including high-frequency enhancement; Multiband Compression in either two or five bands, depending on the processing structure; and Look-Ahead Limiting.

A highpass filter removes low frequency noise that can contaminate some recordings and microphone chains. This noise can otherwise cause problems with the rest of the audio processing and with the codec, which should never waste its bit budget by encoding noise. The phase rotator makes speech more symmetrical, reducing its peak-to-average ratio by as much as six decibels without adding nonlinear distortion. Hence, phase rotation can be very useful for loudness processing of speech.

There are a number of stereo enhancement technologies available. Orban prefers one based on its patented algorithm that increases the energy in the stereo difference signal (L-R) whenever a transient is detected in the stereo sum signal (L+R). By operating only on transients, this algorithm increases width, brightness, and punch without unnaturally increasing reverb (which is usually predominantly in the L-R channel). Gating circuitry detects “mono” material

allows the user to adjust the tonal balance in a much more detailed way. EQ has two purposes in a broadcast processor. The first is to establish a signature for a given station that brands the station, creating a “house sound” by subtly emphasizing the bass, midrange, or high frequencies. The second purpose is to compensate for the frequency contouring caused by the subsequent multiband compression and limiting. These may create an overall spectral coloration that can be corrected or augmented by carefully chosen fixed EQ before these multiband dynamics stages.

Multiband compression and limiting may occur in one or two stages, depending on the developer. If it occurs in two stages, the multiband compressor and limiter can have different crossovers and even different numbers of bands. If it occurs in one stage, the compressor and limiter functions can “talk” to each other, optimizing their interaction. Both design approaches can yield good sound, and each has its own set of tradeoffs.

Usually using anywhere between four and six bands, the multiband compressor/limiter reduces dynamic range and increases audio density to achieve competitive loudness and impact. It’s common for each band to be gated at low levels to prevent noise rush-up, and developers often have proprietary algorithms

and thus introduces sidebands that are far removed in frequency from their associated Fourier “carriers.” The “carriers” hence have little ability to psychoacoustically mask the resulting sidebands when compared with the sidebands that a look-ahead limiter introduces because the look-ahead limiter’s gain control signal has a much lower bandwidth. Therefore, compared to a hard clipper, a look-ahead limiter produces considerably less audible modulation distortion. This is particularly important when one is driving a low bitrate codec because one does not want to waste precious bits encoding this distortion.

Simple wideband look-ahead limiting can still produce audible intermodulation distortion between heavy bass and midrange material. Advanced-technology look-ahead limiters use sophisticated techniques to reduce such IM distortion without compromising loudness capability.

### Codecs

The basic principle of perceptual coding is to divide the audio into frequency bands and then to code each frequency band with the minimum number of bits that will yield no audible change in that band. Reducing the number of bits used to encode a given frequency band raises the quantization noise floor in that band. If the noise floor is raised too far, it can become audible and cause artifacts.

A second major source of artifacts in codecs is pre- and post-echo caused by ringing of the narrow bandpass filters used to divide the signal into frequency bands. This ringing worsens as the number of bands increases, so some codecs may adaptively switch the number of bands in use, depending on whether the sound has significant transient content. This ringing manifests itself as a smearing of sharp transient sounds in music, such as those produced by claves and wood blocks.

### Psychoacoustic Models

Perceptual coders exploit complex models of the human auditory system to estimate whether a given amount of added noise can be heard. They then adjust the number of bits used to code each frequency band such that the added noise is undetectable by the ear if the total “bit budget” is sufficiently high. Because the psychoacoustic model in a perceptual coder is an approximation that never exactly matches the behavior of the ear, it is desirable to leave some safety factor when choosing the number of bits to use for each frequency band. This safety factor is often called the “mask-to-noise ratio,”

*Netcasters have much to learn from broadcasters, however, about what is commonly called “broadcast-quality” audio.*

with slight channel or phase imbalances and suppresses enhancement so this built-in imbalance is not exaggerated.

AGC compensates for varying input levels. This is particularly important nowadays because many CDs are aggressively processed for loudness when they are mastered. Older CDs with the same peak levels can have average levels (which correspond approximately to loudness) more than 10dB below the average levels of current product. This means that the common technique of normalizing audio files for the same peak level causes huge problems with source-to-source loudness consistency, particularly when old and new material is juxtaposed. AGC can go a long way toward smoothing out such inconsistencies.

Equalization applies processing similar to “tone controls” to the signal but typically

for doing this while minimizing the audible side effects of the gating.

A look-ahead limiter controls the peak level at the output of the processor to prevent clipping the codec. The limiter prevents overshoots by examining a few milli-seconds of the unprocessed sound before it is limited. This way the limiter can anticipate peaks that are coming up.

One can model any peak limiter as multiplying its input signal by a gain control signal. This is a form of amplitude modulation. Amplitude modulation produces sidebands around the “carrier” signal. In the case of a peak limiter, each Fourier component of the input signal is a separate “carrier” and the peak limiting process produces modulation sidebands around each Fourier component.

Considered from this point of view, a hard clipper has a wideband gain control signal

measured in dB. For example, a mask-to-noise ratio of 12dB in a given band would mean that the quantization noise in that band could be raised by 12dB before it would be heard. (That is, there is a safety margin of two bits in that band's coding.) For the most efficient coding, the mask-to-noise ratio should be the same in all bands, ensuring that the sound elements equitably share the available bits in the transmission channel.

### Coding Efficiency

Different sounds will vary greatly in the efficiency with which a perceptual coding system can encode them. Therefore, for a constant transmission bitrate, the mask-to-noise ratio will constantly change. Pure sounds having an extended harmonic structure (such as a pitch pipe) are particularly difficult to encode because each harmonic must be encoded, the harmonics occupy many different frequency bands, and the overall spectrum has many "holes" that are not well-masked, so that added noise can be easily heard. The output of a multiband audio processor that uses clipping is another sound that is difficult to encode, because the clipper creates added distortion spectrum that does not mask quantization noise well, yet may cause the encoder to waste bits when trying to encode the distortion.

### The AAC and aacPlus Codecs

AAC is intended for very high-quality coding with compression up to 12:1. The AAC codec is about 30% more efficient than MP3 and about twice as efficient as MP2.

The AAC codec can achieve "transparency" (that is, listeners cannot audibly distinguish the codec's output from its input in a statistically significant way) at a stereo bitrate of 128Kbps, while MP2 requires about 256Kbps for the same quality. The MP3 codec cannot achieve transparency at any bitrate, although its performance at 192Kbps and higher is still very good.

AAC stands for Advanced Audio Coding. Intended to replace Layer 3, AAC was developed by the MPEG group that includes Dolby, Fraunhofer (FhG), AT&T, Sony, and Nokia—companies that have also been involved in the development of audio codecs such as MP3 and AC3 (also known as Dolby Digital™). (AAC does not stand for Apple Audio Codec, although Apple was one of the first to implement this technology with the introduction of Apple iTunes, the most successful downloadable music source, and QuickTime 6.)

Coding Technologies' "Spectral Band Replication" (SBR) process can be added to

almost any codec. This system transmits only lower frequencies (for example, below 8kHz) via the codec. The decoder at the receiver creates higher frequencies from the lower frequencies by a process similar to that used by "psychoacoustic exciters." A low-bandwidth signal in the compressed bitstream provides "clues" to modulate these created high frequencies so that they will match the original high frequencies as closely

provide the absolute best possible sound per bit that the current state-of-the-art will allow, without the typical resonant, phasey, watery character of older-technology codecs, like the Windows Media Audio. WMA has become a de facto "lowest common denominator" codec in netcasting, mostly because Microsoft ships a player with every copy of Windows. However, all third-party bias-controlled tests known to us

*There's little reason not to consider using aacPlus to deliver quality audio to the increasingly sophisticated Internet streaming audio audience.*

as possible. Adding SBR to the basic AAC codec creates aacPlus, which offers the best subjective quality currently available at bitrates below 128kbps. At bitrates below 128kbps, full subjective transparency cannot be achieved at the current state-of-the-art, yet the sound can still be very satisfying. (In the phraseology of the ITU 1 to 5 subjective quality scale, this means that audible differences introduced by the codec are judged by expert listeners to be "detectable, but not annoying.")

Coding Technologies' aacPlus v2, the latest in MPEG-4 Audio and previously known as "Enhanced aacPlus," is aacPlus coupled with the new MPEG Parametric Stereo technique created by Coding Technologies and Philips. Where SBR enables audio codecs to deliver the same quality at half the bitrate, Parametric Stereo enhances the codec efficiency a second time for low-bitrate stereo signals. Both SBR and Parametric Stereo are backward- and forward-compatible methods to enhance the efficiency of any audio codec. As a result, aacPlus v2 delivers streaming and downloadable 5.1 multichannel audio at 128Kbps, near CD-quality stereo at 32Kbps, excellent quality stereo at 24Kbps, and great quality for mixed content down to 16 Kbps and below. MPEG standardized Coding Technologies' aacPlus as MPEG-4 HE-AAC (MPEG ISO/IEC 14496-3:2001/AMD-1: Bandwidth Extension).

With the addition of MPEG Parametric Stereo (MPEG ISO/IEC 14496-3:2001/AMD-2: Parametric coding for high quality audio), aacPlus v2 is the state-of-the-art in low-bitrate standards-based audio codecs. The Coding Technologies codecs

that compared aacPlus to WMA at the same bit rate always rated WMA lower when subjective audio quality is the performance metric. Because WMA has never been passed through formal MPEG testing procedures, it is unclear what bitrate one must use with WMA to achieve parity with any given aacPlus bitrate, if even possible.

Orban's Opticodec-PC family of Streaming and File Audio Encoders uses the genuine Coding Technologies aacPlus codec. Because Opticodec-PC supports both RTSP/RTP (Real and QuickTime) and HTTP/ICY (SHOUTcast and Icecast2) standards-based streaming servers, and because free player clients are available from Real, Winamp, and many mobile phones, there is little reason not to consider using aacPlus to deliver quality audio to the increasingly sophisticated Internet streaming audio audience while saving on bandwidth costs thanks to aacPlus's higher efficiency than MP3 and WMA.

contact info



Readers can obtain more information about Orban/CRL at:

Orban  
1525 Alvarado St.  
San Leandro, CA 94577  
510.351.3500  
www.orban.com

Greg Ogonowski, gogonowski@orban.com  
John Schaab, jschaab@orban.com