Audio Processing for Digital Television

Presented at the 1998 Broadcast Engineering Magazine Digital Television Conference, Chicago, Illinois.

By: Robert Orban
1 December 1998
revision 4

The United States DTV system specifies the Dolby AC3 lossy audio compression system as the standard for transmitting anywhere from one to six channels of digital audio. Part of the AC3 bitstream is "metadata" – data about the data. There are three important pieces of metadata in the AC3 bitstream.

- The first is Dialog Normalization, which, in essence, sets the receiver's volume control to complement the dynamic range of the program material being transmitted.
- The second is Line-Mode Dynamic Range Control, which allows the receiver to perform a wideband compression function if the listener chooses.
- The third is RF-mode Dynamic Range Control, which applies heavier processing.

The obvious question that arises is how these signals are to be generated in a real-world operational facility. And, indeed, in which situations they *should* be generated.

We should remember that the marketing landscape is littered with "features" that seemed to be a good idea at the time, but which proved to be of little or no interest to consumers. Digital technology has vastly decreased the cost of adding new features to consumer electronics, and many consumer manufacturers have responded with a blizzard of features that are confusing, hard-to-understand, or just plain useless.

For example, CDs have always offered the ability to deliver auxiliary data. According to the original CD hype, you would see the lyrics of the songs scroll by as you played them. In addition, you would see still pictures of the band members by connecting your CD player to your television set. Where are these features now? The answer, of course, is that the public did not find them compelling enough to justify the additional production expense to add them to the CD data stream, or to justify the increase in manufacturing cost necessary to add the video outputs or the LCD screens to the CD players.

Another example is the SAP channel in BTSC stereo television. Very few viewers understand it, yet a number of them manage to turn it on by accident. Then they can't understand why the sound becomes low-fidelity mono, and why everyone is suddenly speaking Spanish! Consequently, many consumer manufacturers buried the SAP control very deep in the menu structure of receivers or VCRs to prevent this confusion from occurring in the future.

Concerning the AC3 metadata, I believe that only a small minority of viewers will ever understand the concept of dynamic range control. Dolby Laboratories wisely specified that dynamic range compression would be the receiver default, because they realized that most consumers would never want full dynamic range audio.

Experience has shown that a vast majority of viewers are not interested in wide dynamic range. Instead, they want two things. First, dialog should be comfortably intelligible, and second, commercials should not be irritatingly loud by comparison to program material.

Home theater owners may want the opportunity to watch feature films while hearing a wide dynamic range signal. However, even these viewers usually consume television in a much more passive way when viewing garden-variety programs. If television is to be an acceptable part of the domestic environment, the sound cannot overwhelm household members not interested in viewing (not to mention neighbors, particularly in multi-family dwellings). For a variety of reasons, the dynamic range of sound essential to the intelligibility of the program should not exceed 15dB in a domestic listening environment. Underscoring and ambient sound effects will, of course, be lower than this.

The issue of loud commercials is particularly important – the FCC has been concerned with loud commercials ever since the mid 1960's, and has twice actively investigated the problem since then as a result of viewer complaints. It is against FCC rules to broadcast irritatingly loud commercials.

In current NTSC practice, all audio is applied to a transmission audio processor that automatically controls the average modulation and the peak-to-average ratio. This ensures that the audio will be comfortably listenable. The audio processor also has another crucial function – it smoothes out transitions between one piece of program material and the next. Several currently available transmission audio processors incorporate the CBS Loudness Controller algorithm. This uses a complex algorithm that estimates the amount of perceived loudness in a given piece of program material. If the loudness exceeds a preset threshold, the controller automatically reduces it to that threshold. The main purpose of this circuit is to control the loudness of commercials that have been processed to produce irritating loudness without such control.

Knowing how broadcasters do successful processing in the analog world, I think that the most realistic approach to handling AC3 dialog normalization is a hybrid technique. Most program material can be passed through an audio processor with a loudness controller very much like the ones currently used for analog television. This material is typically either mono or two-channel stereo. It includes commercials, live news, game shows, talk shows, soap operas; and most documentaries, sports, and pop music videos and concerts. Processors used in analog TV control their maximum loudness level very well, so a single dialog normalization value will apply to all program material whenever the processor is online. The advantage of this strategy is that the processor will guarantee that all of this material is comfortably listenable, and that commercials are not excessively loud. With the possible exception of sports, this program material does not rely on extreme dynamic range to make its point, so I do not believe that compression damages the artistic integrity of this programming. No one needs more dynamic range on the *Jerry Springer Show*, or on the local news.

Prime time dramatic shows, newer feature films, and classical music concerts all use dynamic range for dramatic impact, and therefore are candidates for full-blown exploitation of the AC3 metadata. Each show, film, and concert must have a dialog normalization value pre-assigned to it, ideally derived by referring to a calibrated loudness meter. The uncompressed audio is then applied to the AC3 encoder, along with Line-Mode and RF-Mode Dynamic Range Control signals to ensure that the receiver can apply compression if the viewer prefers a narrower dynamic range.

Commercials should be processed through the station's audio processor in the usual way. If the fixed dialog normalization value is correctly chosen for all material passed through the audio processor, commercials will automatically be limited in loudness to the average loudness of the dialog and will therefore be unobtrusive regardless of whether the listener is hearing compressed or uncompressed audio.

I suspect that it is impractical to pass through, without review, dialog normalization values created by program and commercial providers, because some commercial providers will inevitably try to game the system to make their commercials excessively loud. Instead, if dialog normalization is to be actively used in transmission, the broadcaster must strip its existing value from the program, and must then preview each piece of program material and replace the value with one that will ensure consistency from one piece of program material to the next. I think that very few local stations will want to devote the necessary resources to this activity. Instead, it's an obvious thing for the networks to do.

If the networks have done their job well, they will choose dialog normalization values that ensure consistency from source to source, and when the viewer changes channels. It is improbable that this can be done by automation. The best we will be able to do is to manually identify dialog or other baseline sounds, measure their loudness with a true loudness meter, and manually adjust the dialog normalization parameter so that these baseline sounds emerge with a standardized loudness. CBS's research into this area showed that no simple meter could do this accurately, including frequency-weighted meters with averaging characteristics. The errors in such measurements were so large that they were not useful in controlling the levels of commercials well enough to eliminate viewer complaints.

The CBS loudness meter divides the signal into seven octave bands and weights the gains of the bands according to the 70-phon equal-loudness curve of the ear. It then averages the output of each band with a 15-millisecond time constant. The averaged outputs of the bands are then added and the sum is applied to a 200-millisecond time constant. This is applied to the meter, which is assumed to have instantaneous response so that it clearly shows the effect of the two previous time constants.

In tests, this meter agreed with average listeners within 2dB. However, it's important to note that listeners disagreed amongst themselves by as much as 4dB when asked to assess the subjective loudness of a given piece of program material. So any loudness meter can only work for an average listener, and may show considerably greater errors when compared to any given listener.

Dynamic Range Control

I have some concerns about the wideband nature of the compression resulting from dynamic range control, particularly for the RF Dynamic Range Control signal because it applies aggressive compression. Wideband compression has been obsolete in television transmission audio processing every since the early 1980s, when a new generation of processors were introduced that superceded the old wideband Audimaxes and Volumaxes. Multiband compression prevents spectral gain intermodulation, which occurs when midrange and high frequency program material is audibly pumped up and down by bass. Because the ear is far less sensitive to bass than to midrange material, bass having low loudness but high energy will cause gain reduction that causes the loudness of the midrange to vary, seemingly inexplicably.

Experience has shown that dividing the processing into two bands above and below 200Hz, and then compressing each band independently, is sufficient to prevent audible spectral gain intermodulation when using the relatively mild compression typically applied to television audio. This option is unavailable when the Dynamic Range Control is used to determine compression, although the level detector determining the amount of DRC compression can be frequency-contoured to mimic the equal-loudness curves of the ear. This should help reduce the problem, although it introduces another.

Multiband compression is also useful in performing an “automatic equalization” function to change the frequency balance of the audio on a program-adaptive basis. In a multiband compressor, frequency bands containing excessive energy are automatically compressed more than other bands. This results in a re-equalization of the program material towards some target spectral balance. In a two-band compressor, it controls excessive bass, which can otherwise cause muddy balances. A five-band compressor, such as the one available in Orban’s current digital processor for analog audio, can perform more detailed automatic re-equalization that can be very useful for program material such as live news. We find approximately 60% of our digital Optimod-TV users are employing two-band compression, with the remaining 40% using five-band compression.

Availability of multiband compression is another argument for passing most program material through a conventional compressor with loudness control even in DTV service. Multiband compression smoothes out not only loudness variations but also variations in equalization, which can be particularly valuable with program material that has to air in a timely manner, where there is no time budgeted for careful audio post-production. Material that airs with full Dynamic Range Control implemented should be refined so that it sounds polished and consistent without further processing. A considerable amount of televised material does not meet this criterion.

Optimod-TV in the 5.1 Channel Environment

Let me discuss some of the details of implementing a 5.1 channel version of conventional Optimod-TV audio processing.

In the two-channel world, we stereo-couple the compressors by determining the gain reduction of both channels by the louder of the two channels. This works well for two-channel, but for 5.1 we will have to base the gain reduction on a power summation of the five channels, with at least two-band compression. In addition, we will have to have five loudness meters for each of the five full-range channels, and sum the loudness meters’ outputs to get an estimate of the overall loudness of the sound field. We can use this information, by feedback, to control the maximum subjective loudness of the sound field to a user-specified loudness threshold. The combination of power-summed compression and loudness control will result in consistent perceived loudness regardless of the energy distribution in the 5.1 channels.

By setting the dialog normalization parameter to provide adequate headroom, we should never have to perform peak limiting on the compressed signal. However, if we wish to perform peak limiting for some reason, we should do it with a look-ahead peak limiter that provides low modulation distortion. Such limiting does not significantly smear the spectrum of the unprocessed signal, and therefore puts minimum stress on the AC3 perceptual coder.

We are very fortunate that AC3 does not use pre-emphasis, because the 75us pre-emphasis that is used in analog television audio has long caused considerable problems. It reduces high frequency headroom by up to 17dB, and therefore requires elaborate high-frequency limiting strategies to ensure consistent loudness. Because the processing for AC3 will not require high frequency limiting or, in all probability, peak limiting of any sort, it has the potential to be cleaner and more artifact-free than current analog processing.

The Orban Optimod-DAB 6200 – A Solution for Mono, 2-Channel Stereo, and Dolby Surround Encoded Material

Orban does not yet manufacture an audio processor that can accommodate full 5.1 channel audio. However, we do offer the Optimod-DAB 6200. This is a two-channel processor that has been tuned to the requirements of digitally compressed audio services like AC3. It offers Protection, 2-Band, and 5-Band compression, followed by low-distortion look-ahead limiting. It is fully remote-controllable, either by GPI-style contact closure or by an RS-232 serial connection to a IBM-compatible PC running Orban's PC Remote Control software.

Special features for the digital television broadcaster include a facility to pad the throughput delay so that it is exactly one frame of 24, 25, or 29.97-fps video. There is also an AES3 sync input. Combined with standard sample rate converters on the digital inputs and outputs, this means that the 6200 can accept synchronous or asynchronous digital inputs at any sample rate between 32 and 48kHz, and can output a digital bitstream at 32, 44.1, or 48kHz that is in sync with the station's master sync system.

The 6200's next software revision will add a full implementation of the CBS Loudness Controller, as well as TV-specific presets similar to those in Orban's 8282 Optimod-TV.

More information on Optimod-DAB 6200 is available on Orban's web site, www.orban.com.

Conclusions

So what are my conclusions? We have a great deal of experience with conventional transmission compression and limiting, and we know that the current practice satisfies most viewers. We know this is true because well-processed audio generates very few viewer complaints. Therefore, we have to carefully consider what program material will truly benefit from the ability to be heard with unprocessed dynamic range. Any program material that will not so benefit should be processed conventionally so that we can ensure that viewers can hear the audio comfortably without being blasted by loud effects or commercials, or being forced to strain to understand dialog.

Any program that is transmitted with full dynamic range must be pre-auditioned to determine an appropriate setting for Dialog Normalization and to determine if the uncompressed dynamic range is appropriate for home theater-style viewing. We can only hope that the production houses will mix prime-time dramatic programs appropriately. Concerning feature films, full 70mm-style dynamic range is probably inappropriate for home viewing under all but the most unusual circumstances. Instead, I believe that the dynamic range used with Dolby SR optical releases is probably more appropriate for the uncompressed signal.

In addition, Hollywood seems to be making more and more mixes that have severe problems with dialog intelligibility when folded down to stereo from 5.1 or Dolby Surround. This will cause problems with many viewers who do not have full surround systems. As broadcasters, you should complain to the studios if they deliver product in which dialog cannot be understood on a typical television receiver.