

Digital Broadcast Audio Processing: Finally, the New Frontier

A look at issues involved in advancing the state of the art of DSP applications in audio signal processing.

Abstract

DSP (Digital Signal Processing) based broadcast audio processors have been, at best, a digital clone of analog counterparts. DSP is a very powerful technology. Why then, has it been so difficult to create successful-sounding digital signal processors for broadcast? This paper addresses this issue and discusses advancing the signal processing artform. Induced aliasing distortion, time delay, and sampling rate were never a concern in the analog signal processor. With DSP however, these issues can create added audible distortion, propagation delay, and/or overshoots. Algorithms that intelligently conform to the natural dynamics of an audio waveform will also be examined. Finally, with the advent of DAB and Netcasting, we'll explore processing for these new and important mediums.

To Boldly Go ...

In borrowing part of a phrase from Capt. James T. Kirk of the Starship "Enterprise," processing in the digital domain meant going where it had never been before. The introduction of DSP for transmission processors for broadcast promised to open a new frontier. Although the first wave of digital processors became available almost seven years ago, this new territory remains largely unexploited. What happened?

Initial attempts at digital processing were mere 'clones' of established analog designs. To those versed in the technical 'black-magic' of audio processing, developing a digital processor requires more than just porting over an analog design into DSP. (More on this later.) If key issues aren't dealt with, the result will add audible distortion. The following discussion is based on extensive experience developing a DSP-based processor.

In the quest for the fully digital broadcast facility, concerns involving transmission processors must be addressed. Codecs implementing an AES/EBU interface and used in digital STL systems can audibly degrade the transmission path. Depending on the coding algorithm and bit rate employed, codecs may be a source of distortion. While placement of the codec as well as the signal processor in the broadcast transmission path are critical, such discussion is beyond the scope of this paper. Suffice it to say, coded STL must be taken into account.

Also of concern are sample rate converters used in the AES/EBU interface. We will discuss the possible overshoot problems that may arise when transposing a sampling rate. Early research discovered that a processed signal that was properly peak controlled by the signal processor had the potential to cause modulation overshoots. Propagation delay can also become a problem with digital processing. This is related to the amount of time required to complete all of the processing tasks. During this period, there is an input to output delay. If this delay is too long, live monitoring off the air becomes quite difficult. For announcers especially, such delays can be very distracting. Further discussion will detail what is involved in resolving these issues.

Given the problems, is processing in the digital domain of any benefit? Yes. There are solutions along with a multitude of benefits from numeric-based processing.

Research and Design Criteria

To start, key performance criteria were established: • the digital system must operate in a transparent manner in relation to the audio signal, • the digital processor must sound at least as good as, or better than, its analog counterpart, • no additional coloration to the audio from A/D, D/A, sample rate converts, jitter, and/or the AES/EBU interface must result.

To determine the strength of the issues presented here, subjective listening and technical tests were needed. Analog and digital processing systems had to be compared, along with aural evaluations of the converter systems. Most important was evaluating analog to digital processors. Testing was meant to identify existing weaknesses, resolve such weaknesses, and help develop new algorithms.

Subjective Listening Tests

This phase of testing was conducted using the following listening conditions. The device under test was provided a clean program source from either a CD or DAT tape. Output was monitored either discrete left/right, or multiplex via a stereo monitor, amplified and heard through well known reference monitor speakers. For direct A/B processor comparison, a switching device was used that could select either multiplex A/B, or discrete A/B processors. Much the same as would be done in auditioning processing equipment by a radio station. Outputs of the units tested were peak monitored using both a modulation monitor and oscilloscope. For fair and honest evaluation, the peak levels of each system were always set at exactly the same level.

In direct comparisons of digital and analog processors, both operating with equal amounts of processing, listening tests revealed that digital systems seem to generate additional annoying distortion characteristics. In addition to the known artifacts generated in analog processors, there seems to be an added distortion signal. Where most processing artifacts can usually be described in some form of Total Harmonic Distortion, (THD) or Intermodulation Distortion, (IMD) this artifact appears as something completely different. In adjusting audio processing it is generally understood that an increase in IMD is derived from an increase in dynamic control characteristics. Items like increased release time of compressors/limiters, ratio, or amount of gain control will contribute to this. An increase in THD will usually result from an increase in hard limiting, or final clipping.

'That Digital Sound'

Further listening reveals that this artifact seems to produce a 'harsh' or 'metallic' effect on the 'presence' and 'brilliance' ranges of the audio spectrum. Almost as if some additional synthetic component is being added into the signal. With moderate amounts of processing, cymbals sound like 'nails on a blackboard', 'S' sounding material begins to sound like paper tearing, and high frequencies lose definition and detail! All of which has been characterized by engineers as 'that digital sound.'

For example, in the song 'Big Love' by Fleetwood Mac, at precisely eleven seconds into the song, there is a moderate crash of the cymbals. The digital processor repeatedly distorted these cymbals as compared to the analog system. Even a reduction in the amount of processing by the digital system only reduced this distortion effect, never eliminated it. To try and illustrate, this effect sounded like some form of 'smashing glass' instead of the crisp, detailed crash of the cymbals.

By contrast, the analog system did not have this problem. Only with the processing increased, would the cymbals distort, but it would happen in the known 'spitting' sound of THD generated from the increase in hard limiting. Even this form of distortion was 'easier on the ears' when directly compared to the

digital distortion. The reader may want to use this song as a test source. In the listening tests done while researching this topic, there were multitudes of programs that would cause this effect. What is it? What causes it? Is it a question of primitive digital design, or is it a technological weakness? Considering the numerous attempts at digital processing, these questions had to be answered before any development could proceed!

This problem was perceived during tests performed on numerous digital processors in comparisons with numerous analog processors. In each case, 'that digital sound' was observed.

Technical Tests and Research Data

As a researcher, engineer, and designer working with signal processing, I have a dual view of technical testing and research data. Where the data gathered from research and conducted tests can be used to prove or disprove a theory, it offers a subjective limit as to what may or may not 'sound' acceptable under dynamic processing conditions. Therefore testing and research may be used as a tool to aid in assisting 'possible' benefits or drawbacks to problems, issues, and answers. In the end, tests were developed that would break down ideas, and algorithms into the smallest common denominator so that dynamic performance could be monitored or judged.

The first phase of testing was an effort to discover the cause of that problem identified as 'that digital sound'. Since most processors are dynamic in nature, there are not any specific, static tests that will provide common results. Most systems can be evaluated with weighted noise, timed pulse bursts, or IMD signals. Frequency response can be measured, but usually with levels that are below the processing threshold to avoid generating leveling and phase errors due to the gain control action.

The Culprit

For this presentation, a digital processor operating with a 32kHz sampling rate was evaluated. A system like that used in conventional FM broadcasting, and consisting of AGC/dynamic limiting, emphasis, and the final limiting low pass filtering function was employed. In performing an audio sweep, an interesting item was noticed upon spectral analysis of the frequency range. A large level of aliasing components were observed above 4kHz! These aliasing signals would fill the entire spectrum starting at 15kHz, and work their way down the complete spectrum. This was beginning to get interesting! Figure-1 is an example of a 5kHz tone that has been clipped in the digital domain, and now produces aliasing products.

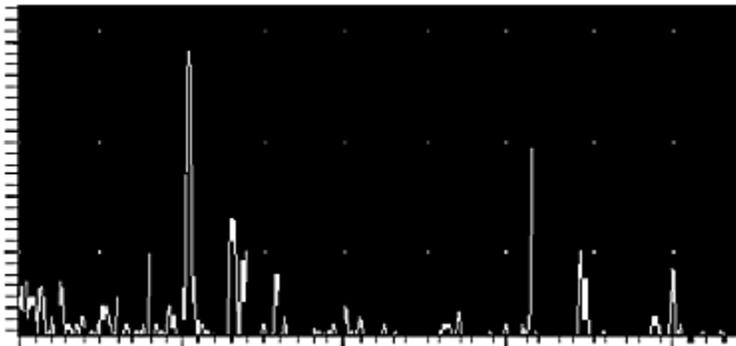


Figure-1

As can be seen in the diagram, there are aliasing products sitting near the fundamental of 5kHz. As frequency increases, more products develop in quantity, and size. At 12kHz the aliasing products are almost as significant in value as the fundamental. Now try to consider what this picture will look like if music program were used in place of a single audio tone.

After making some adjustments to the processor with the frequency sweep applied, it appeared that reduction of the final limiting would reduce the aliasing substantially. Reducing the processing on the upper frequencies totally eliminated the problem. What was this indicating? Considering that aliasing distortion is caused by signal energy that is trying to exceed the Nyquist frequency, it would appear to indicate that the final limiting seemed to be the culprit. Since the final limiting function does generate harmonic content, it is likely that energy exceeding the Nyquist frequency would result. Further spectral evaluation proved this to be true.

In a subjective test to confirm or reject the research data, music was applied to the processor. With the final limiting operating in the 'normal' range, the 'metallic' sound was evident. Upon reducing the amount of final limiting, the 'metallic' sound disappeared! This seemed like the 'proof in the pudding' for the cause of 'that digital sound'. By contrast, in the analog processor, it is the final limiting function 'where the rubber meets the road' with regard to clarity and loudness. If for some reason the digital system could not utilize a moderate amount of final limiting without clarity, then no loudness benefit would result!

Further testing revealed that indeed the final limiting function generated the majority of the aliasing distortion. However, a few other processing functions were found to be possible contributors as well. Certain cross-over designs and/or filter banks, along with system headroom are all areas that must be designed properly or processing induced aliasing will occur.

Also observed were certain processing time constants. If a timing signal was operating at a rate that exceeded the Nyquist frequency, aliasing would be generated. Given the system discussed here, any timing signal exceeding $62.5 \mu\text{s}$ ($1/16000\text{kHz}$) would create an aliasing component. With the problem found, what are the options or remedy?

Sampling Rate: Can of Worms?

When discussing anything related to aliasing, the sampling rate must be part of the equation. It is well known that increasing the sampling rate, will raise the point at which aliasing will occur. The question now becomes, how far must the sampling rate be increased to eliminate processing induced aliasing? Testing and research indicated that with a 32kHz sampled system, a multiple of at least 4 times the sampling rate, 128kHz, would be sufficient for broadcast transmission purposes.

Creating a 128kHz sampling rate can be done in one of two ways: Using a high speed A/D converter operating at the 128kHz rate, or up-sampling the 32kHz rate by a factor of four to create the new higher rate. The latter is preferable as it allows use of the industry standard A/D conversions that support popular 32kHz, 44.1kHz, and 48kHz rates. A converter at 128kHz is available but is generally more expensive and requires additional ancillary input filtering, further adding to the cost. In addition, there is no need to operate the entire system at the higher sampling rate, since that would reduce the amount of machine cycles in the DSP.

This points up another problem: When a faster sampling rate is required to remove aliasing distortion, how much DSP power is compromised to accomplish the goal? The obvious answer is 4x the power, but with all of the final filtering and overshoot control needed, is this the most efficient method to rectify the situation? Of importance here is the final low pass filter. Since it must provide ample bandstop rejection in the 16kHz Nyquist region, a filter of high magnitude is required. This will take multiple machine cycles in itself at 32kHz sampling, with 128kHz, it will be 4x the requirement. Might there be an alternative to this process that will save machine cycles, yet accomplish the same result, eliminate aliasing?

Why 32kHz Sampling Rate?

Another issue for discussion is the base sampling rate itself. Digital processors thus far have all used 32kHz as a base sampling rate, which in turn sets the Nyquist at 16kHz. Considering that conventional FM Stereo broadcasting requires 15kHz of audio bandwidth, this leaves only 1kHz of guard band spectrum before the Nyquist point. To facilitate this, a filter of very large magnitude must be employed in order to suppress all energy by at least 96dB at the Nyquist, or aliasing occurs. This can be done in DSP using a finite impulse response filter (FIR). The only drawback is that it will require many 'taps' within the filter to achieve this level of stopband rejection. The significance of the 'taps' is that for every two taps in the filter, it requires one sample to perform its duty. For a 15kHz FIR filter of this magnitude, it will need 101 taps. This in turn results in 50 required samples which equates to 1.56 milliseconds of propagation delay through the filter.

It must be noted that even when up-sampling, where a 'new' Nyquist frequency would now reside at a multiple of the original, the problem still remains. This is due to the downstream requirements of the AES/EBU interface. With an up-sampled signal operating internally within a host system, higher speed D/A converters can be used in the conversion process to analog without overshoots and distortion. On the other hand, the AES/EBU is a standard protocol that will only support a system sampling rate up to 48kHz. Therefore any filtering that must be done in the up-sampled domain must still adhere to the original Nyquist frequency or aliasing will result. In the case of this discussion, that frequency remains 16kHz. Still at a faster sampling rate, say 128kHz (4x the original) the number of FIR filter taps remains the same as described above.

A broad question posed to the global processing forum is why the use of 32kHz as a base sampling rate? I think, based on tests and research, that a base of 48kHz would make all of the aforementioned problems much easier to deal with. The guard band to the Nyquist is much farther out, which in turn moves out the aliasing point. This would also allow a final filter with less time restriction. Coupled with the fact that the propagation delay associated with 48kHz is much faster in itself and makes this rate more desirable.

My best guess as to why 32kHz sampling was chosen in the past, is that at that rate, there would be more machine cycles available to handle the workload. That would be the only reason to possibly support a lower sampling rate.

Alternative Anti-Aliasing Limiter Method

Since it is now apparent that aliasing, sampling rate, and machine cycles are all of importance in digital processing, what alternative is there that might allow the best performance, and yield the most efficiency? The answer is within the method used to accomplish the final limiting function. Through a

proprietary process researched and developed by the author, a mathematical analysis provided a means to accomplish the final limiting function, without aliasing, and at the base 48kHz sampling rate!

Considering that the math involved exceeds the scope of this paper, and that an entire presentation could be based on explicit digital processing design alone, in depth discussion of this analysis is best suited for another forum. Suffice it to say that this alternative method removes all of the previously discussed problems of digital processing. Should the reader desire more detail of this analysis, please contact me for more information.

Confirmation of performance of this alternative method was achieved with spectral analysis and subjective auditioning of music with this new function employed. In aural monitoring it was decided that the 'metallic' digital aliasing distortion component mentioned earlier disappeared! Now it is possible to define what 'that digital sound' is, and more importantly how to eliminate it!

Sample Rate Converters

Another innovative device used in the digital realm is the sample rate converter. This function will transform one system sampling rate to another. This becomes necessary when interfacing digital devices that use different sampling rates, and thereby adding compatibility among different systems.

This function is accomplished by scaling up, or interpolating the original sampling rate, usually by a factor of ten. Then, at the 10x rate filtering the signal with a low pass filter that is set to the Nyquist of the new desired sampling rate. Finally, the signal is scaled down, or decimated by the factor needed to achieve the new rate. While this sounds quite simple, and basically it is, there are a few issues to consider.

All transmission processors, both analog and digital apply some form of overshoot control to the output filtering section. Our concern is not the method used, rather that control is achieved. In most designs, this function is a form of integrated protection clipper working around the final low pass filter to obtain control. In each case the overshoot component can be calculated as a product of what is known as the 'Gibbs Phenomenon'[FF1]¹, which states that an overshoot will occur at one-third the cut-off frequency of any low pass filter whenever a non-linear waveform is passed through it. In the case of broadcasting, the non-linear waveform would be that of a clipped waveform. Knowing that the audio bandwidth used in FM Stereo is 15kHz, overshoot components will begin with any non-linear waveform above 5kHz. In this example, this would effect any signal above 5kHz that was clipped. Should the slope of the previously described up-sampled interpolation filter appear greater than the slope of the final filter in the audio processor, then output overshoots may result! Unfortunately, these overshoots are generated after the processing unit. To remove them would require another device.

This does not necessarily indicate that all sample rate converters will cause overshoots. But in most cases the filtering used in the sample rate converter will be of a large magnitude in the bandstop rejection area. In all probability it will be an FIR filter with at least 96 dB rejection in the stop band. Also of interest will be the direction of rate conversion. Should the host sampling rate be lower in value, than the transformed rate, chances of overshoot are small. This happens due to the up-sampled filter being set to a broader spectrum than the spectrum of the host signal. Potential problems may arise when transforming a larger sampling value to a lower rate. Then the details of the above description apply.

Processing and Coded STL Systems

A technique that is very popular is the use of the audio codec to reduce the data requirement for a digital STL system. These devices make use of 'lossy' data reduction algorithms to compress the bitrate down to a size that will fit within the existing bandwidth of the STL system. While there are a number of specific algorithms to choose from, most STL manufacturers have made use of proprietary digital formats that are derivatives of prior development. Most common usage has been done with ISO/MPEG Layer-II, ISO/MPEG Layer-III, apt-x, and Dolby AC-2.

Detailed operation of the above mentioned algorithms is not needed for this discussion, as my focus is on the actual effects that data reduction algorithms have upon signal-processed audio. Each system possesses many strengths and possible weaknesses for this application. It is not my intention to advocate any audio coding technique to be considered as a standard or preference when compared to one another.

Dealing with Pre-emphasis

All broadcast applications make use of some form of pre-emphasis boost. For FM broadcasting, North American Countries utilize a 75 μ s emphasis, whereas 50 μ s is used elsewhere the World over. Medium-wave, or AM makes use of an optional modified 75 μ s emphasis.

Transmission signal processors employ pre-emphasis within their system architecture. Since emphasized audio must also fit within the imposed modulation limits, the processor employs specialized high frequency control sections that provide both the emphasized boost and control of the high frequency energy. In this manner, efficient high levels of modulation are easily obtained since the processor is designed and set to limit any tradeoffs resulting from pre-emphasis and high frequency limiting requirements. Basically, these two sections work in concert with one another to allow pre-emphasis to be employed, and yet control the emphasized energy content.

One of the critical requirements of any codec is that the audio signal must be "flat" in spectral origin. The term "flat" is used here in the context of a signal where no additional EQ has been added to the original component. This is due to the operation of the masker signal used in the coding process. Any significant change, or imbalance of the frequency spectrum can cause the threshold curve of the coding system to possibly have a profound effect on the output of the coded audio².

At issue here are what the coded STL encoder/decoder requires and what the processing system will provide. A paradox exists because the processor is designed to output a pre-emphasized signal, and the codec is designed to accept a "flat" signal. To accomplish this, the output of the processor must be de-emphasized so that the output signal is returned to a "quasi" flat form.

The weakness that this function creates, is that the output of the coded STL must restore the emphasis to the signal. Thus adding another generation of emphasis which might add some distortion, but in all probability will add modulation overshoot to the total transmission system. To eliminate the added overshoot, another limiter must be employed. Unfortunately, tests have shown that operating a transmission processor with an emphasized output into a codec will generate audible high frequency distortion. This occurs because the spectral balance to the codec masking process is not spectrally flat, which is what the masker signal wishes to operate on.

A discussion of codecs in the transmission system exceeds the scope of this paper. For further reading, see a paper by the author presented at another industry forum³.

What about That New Frontier?

So far, the focus has been on finding the weaknesses of and remedies for prior digital implementations. The most important of which was the discussion of the non-aliasing limiter. Although an accomplishment I believe that resolving this problem only brought the technology up to date. Now, how might this technology finally move forward?

Phase Linear Dynamically Flat Cross-Over

A topic of vast interest in any multiband signal processor is the cross-over network. The goal is to achieve maximal flat response, with gain control employed, and maintain as linear as possible phase response over the entire spectrum. Easier said than done!

In the analog derivative this was virtually impossible. With gain control active, phase errors between audio bands would develop due to the difference in propagation delay of each cross-over filter. As gain levels would shift near the cross-over frequency, additional gain or even loss would occur at the final summation point. This might result in possible 6dB of gain or loss at this juncture. Some designs would offset phase at the cross-over region in an effort to minimize this problem. However, the compromise disrupted the linearity of the phase response over the whole spectrum.

The digital cross-over makes use of time aligning to each of the audio bands. In this manner, true phase linearity can be maintained, while maintaining dynamic flatness of the program signal. This in turn eliminates any added gain or loss at the final summation. This analogous to using time aligned loudspeakers.

Program Dependent RMS Calculation

Within the design of many AGC sections for compression, there is some function to calculate an average level. This in turn is used within the compressor's control function to alter the gain structure. It has been found that the use of RMS detection seems to produce a natural sound to the AGC operation. In essence, the RMS function calculates the averaged root mean square value of a signal as it occurs over a period of time. Within a block diagram, this average over time is achieved by the use of a simple time constant that is nested within the square root of the squares function. Figure-2 is a block diagram of the theoretical concept behind the RMS calculation.



Figure-2

This style of RMS detector has been implemented in numerous forms recently. All of which have found their place in processing applications. The drawback is that the averaging time is normalized for a broad range of control. This is sufficient for RMS calculation of static signals, but when processing any audio signal, there might be the need to alter the averaging time, as if to create a rolling average.

It is important in the design of the RMS detector, that the lowest frequency passed through the detector does not generate any AC ripple to the control signal. This is possible if the averaging filter is set too

fast⁴. If this occurs, distortion and gain control errors will result. Therefore many RMS detectors must operate with a compromised averaging time to not generate distortion and control errors.

Using program material as content, there will be instances when certain audio frequencies would benefit from a rolling average, as compared to a compromised static value in the detector. This is easily done in DSP as the processor can calculate different averaging times, as well as insert them to create the rolling average. Also, rules can be implemented to allow this process to occur only during desired situations. Through the use of on board memory, it is easy to store a 'history' of what has transpired over time regarding a signals content. Then gain changes can be calculated based upon historic statistics, and figured into the processing algorithm.

Due to this paper's limitations, please accept that volumes can be generated about the subjective nature of controlling an audio signal. The point is that with the digital processing function, the realm of possibilities for control now becomes endless!

Intelligent Interactive Processors

This discussion of digital processing has centered on specific individualized operations. With this in mind, there are a few systematic operations that can be explored.

Part of the workload of a dynamic processor, compressor, limiter, or clipper, is to calculate a value for an audio signal and then use that as a control operation to alter the gain. The transmission system processor is usually made up of a combination of all of the aforementioned processes. Through that, it is possible to know the precise amount of RMS and peak level present. Along with having the historical values mentioned earlier. This information can be used on an interactive basis among the AGC, peak limiter, and clipper sections to systematically provide information to one-another.

An example of this would be the operation of the final clipper. Should significant clipping occur, information can be routed back to the previous AGC and limiter sections altering the gain in an effort to reduce the amount of clipping beyond a specific amount. Through this additional processing, induced distortion can be controlled. This can even be performed on a frequency dependent basis.

DAB and the Web

Past designs for transmission system processors have focused on FM, AM, and television. It is time to enlarge the focus and assess the requirements of Digital Audio Broadcast (DAB) and netcasting, transmitting audio via the World Wide Web. These new mediums have very specific needs, and in each case, signal processing can be used to augment their performance.

For DAB, processing can do more than create a radio station's signature sound and control overmodulation. It can be used to help minimize the effects of the data reduction algorithms that are required by digital transmission systems. Considering there is no need for an emphasized signal in the transmission path, this will further reduce the rigor set upon the processing system. That alone, will improve audio clarity and reduce distortion.

As for the Web, processing will play an even larger role. Knowing that the bit rate requirements for audio on the web are quite small, signal processors can be used to pre-condition the signal so that aural enhancement, and intelligibility, at the lowest bit rates, will improve. Here is a truly untapped area of potential in the world of communications.

Net Results

The goal was to review the progress of audio processing in the digital domain. In my opinion, quality digital processing was always possible, but early implementations were very primitive in nature, design, and most importantly in aural presentation. As we've seen, DSP dynamic processing has been given a 'bad name' because of its disappointing performance thus far. But, I believe through the implementations presented here, this technology may yet clear its name. Only time will tell!

REFERENCES

- [1] Baher, H. Analog & Digital Signal Processing, J. Wiley & Sons, 1990
- [2] Mendenhall, G. Pre-emphasis and Limiting Considerations for Audio Processors and Digital Studio-to-Transmitter Link, White Paper, 1995
- [3] Foti, F. Broadcast Signal Processing and Audio Coding: Are We Trying to Mix Oil with Water?, AES Pre-print #4203, 1996
- [4] Kitchin, C. and Counts, L. RMS to DC Conversion Application Guide, Analog Devices, 1986