A New Perspective on Decimation and Interpolation Filters

Steve Green
Technical Marketing Manager
Mixed-Signal Products
Cirrus Logic, Inc.

1. Introduction and Historical Perspective

Early digital audio conversion systems required very high-order analog filters to provide anti-alias filtering in the A/D conversion and reconstruction filtering in the D/A conversion processes. A characteristic of these high-order analog filters was a substantial amount of phase shift within the audio band. Not only were these filters difficult to manufacture, the excessive phase shift was considered to be one of the primary sources of the poor sound quality associated with early digital audio systems. The analog anti-image filters in base-band D/A conversion systems were replaced by architectures that utilized digital interpolation filters to perform the more difficult portion of the filtering which were then followed by low order analog filters to remove the remaining images. A similar process was implemented for A/D conversion where the analog anti-alias filter was replaced by a combination of a low-order analog filter, over-sampling the input signal and using digital decimation to reduce the sample rate to the appropriate base-band rate. Partially as a result of the perceived audibility problems associated with the phase response of high-order analog filters, these digital interpolation and decimation filters were implemented with linear phase Finite Impulse Response (FIR) filters. In regards to in-band phase response, the pendulum swung from one extreme to the other and linear phase FIR filters for the conversion processes became “holy ground” and continue to be the industry standard 25 years later.

At the 75th convention of the AES in 1984, Dr. R. Lagadec presented a paper, “Dispersive Models for A-D and D-A Conversion Systems” [1], which discussed the potential degradation in audio quality that results from the “pre-echoes” that exist in the impulse and transient response of FIR filters. This premise, though of interest, apparently received little attention at the time. Now fast-forward to the mid-nineties, 96 kHz sample rates have become a reality, 192 kHz is on the horizon and SACD has been introduced. There is general agreement that the evolution to higher sample rates offers an audible improvement over either 44.1 or 48 kHz and, as with most human endeavors, we search to understand why. The obvious and intuitive explanation is that the improvements result from the extended frequency range these formats provide. However, it is generally accepted that the upper frequency limit of human hearing is in the neighborhood of 20 kHz and both 44.1 and 48 kHz sample rates have frequency response that extends beyond 20 kHz. Apparently there must be something else and there has been considerable discussion, several AES papers and magazine articles [2,3,4,5,6,7,8], since the mid 90's related to the audible improvement of the higher sample rates. Much of this work has been a continuation of the work presented by Dr. Lagadec in 1984. (It is also very interesting to note that Sony recently claimed on their website that it is the removal of the digital decimation and interpolation filters, not the extended frequency range, that produce the audible improvements offered by SACD over conventional PCM.)

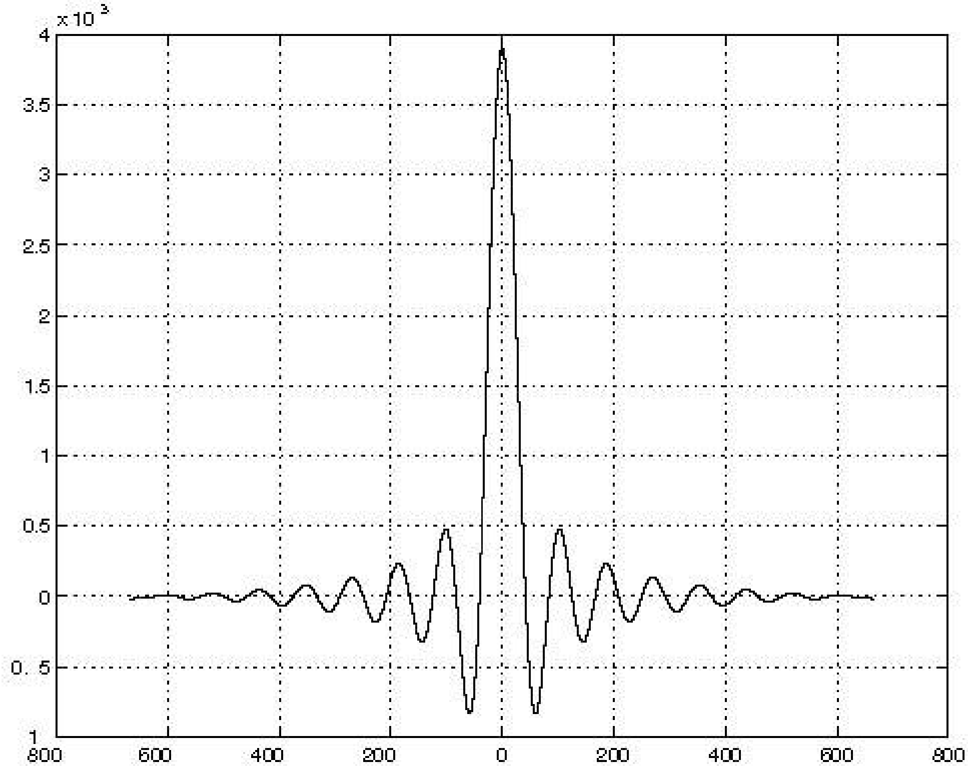

Figure 1. Impulse Response at 48 kHz Sample Rate (X-axis in µsec)

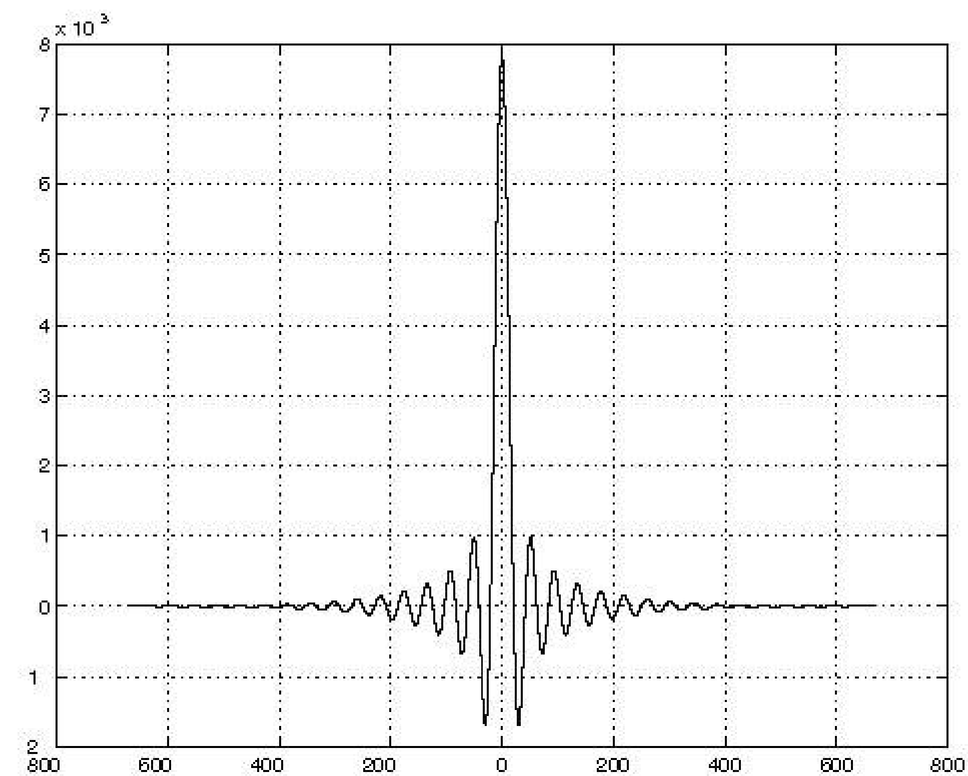

Figure 2. Impulse Response at 96 kHz Sample Rate (X-axis in µsec)

The impulse response of a typical FIR filter for a 48 kHz sample rate is shown in Figure 1. Notice the ringing or “pre-echo” prior to the arrival of the impulse. This pre-echo creates what has been described as a time smear or energy dispersion of the original signal. Figure 2. shows the impulse response of the same filter at a 96 kHz sample rate. It is apparent that the pre-echo has shorter time duration and less time dispersion than at 48 kHz. It is this difference that many believe to be a primary source of the audible superiority of the higher sample rates. As a result of this, many industry experts have suggested that the additional bandwidth offered by the higher sample rates be used as a transition band to allow the use of less aggressive digital filtering, which minimizes pre-echoes and time dispersion even further. Julian Dunn [4] and Mike Story [7] have both written very nice discussions of this subject.

Unfortunately as convincing as these arguments are, there is not an industry consensus as to how the additional bandwidth of the higher sample rates should be used. Many influential manufacturers and end-users continue to believe that the extended frequency response is the key to audible improvements. It is also interesting to recognize that, despite all of the attention focused on the higher sample rates and latest delivery formats, 44.1 and 48 kHz sample rates continue to dominate the industry. Assuming the shortened pre-echoes and improved time dispersion of FIR filters are responsible for the audible improvement in the higher sample rates, how can we apply this knowledge to improve the audio quality at 44.1 and 48 kHz sample rates?

2. Audibility of signal latency or time delay

Digital audio systems introduce time delay. This delay results not only from the decimation and interpolation filters but also from transmission links and digital signal processing. In most applications, this delay is inaudible. However delay is a recognized problem in live sound applications. Consider the situation where a performer has the live sound mix fed back via headphones or loudspeakers. It becomes very difficult, if not impossible, for an artist to continue to perform if there is sufficient delay in the monitor signal chain to create the perception of an echo. This effect is documented and it is very likely that the majority of the population has experienced this phenomenon with a poor telephone connection. There is a significant amount of anecdotal information that suggests that delay of as little as 1 to 10 msec is audible and 1 msec of delay is equivalent to 48 samples at a 48 kHz sample rate.

In converter data sheets, this delay is generally referred to as “Group Delay” and specified in multiples of the sample rate. Typical specifications for interpolation and decimation filters range from 30 to 70 samples periods for each stage. Using these devices, and ignoring the additional delays from transmission or digital signal processing, the delay is in the neighborhood of 2 msec at a 48 kHz sample rate. This leaves very little, if any, time for transmission and digital signal processing before the delay encroaches into the audible range. It is apparent that the delay introduced by these devices is already marginal and adding additional taps (and additional delay) to minimize pre-echoes is not an option for live sound applications.

3. Backed into a Corner

It would appear that we've backed ourselves into a corner. The conflicting requirements of flat frequency response and minimal pre-echoes create an interesting set of challenges for designers of A/D and D/A converter ICs, especially at 44.1 and 48 kHz. Though FIR filters are considered “the industry standard” they are computationally inefficient and generally require design trade-offs when subjected to harsh economic realities. “Despite many clever schemes for increased computational efficiency, a compromise between desired response and the number of taps is not uncommon. The tradeoff is between attenuation, flat response, ripple in the pass-band (and attenuation region), transition band and more.”[9]. Not only is there an issue with the cost of FIR to properly address these issues, the filter also requires additional taps which translates to an increased time delay. Essentially, the desirable attributes of flat frequency response, absence of pre-echoes and sub-millisecond delays are mutually exclusive with FIR filters.

4. The Applicability of IIR Filters

One option available to address these issues is to transition to Infinite Impulse Response (IIR) filters. IIR filters are computationally more efficient, which allows for greater processing flexibility when compared to a FIR filter. This efficiency includes the flexibility to minimize, if not eliminate, pre-echoes, and maintain flat frequency phase response for all sample rates. Another benefit of IIR filters is vast improvements in latency characteristics. Latency specifications for IIR, with comparable frequency response to the FIR, are in the range of 5 to 10 sample periods or 75% less than FIR.

One of the perceived drawbacks of IIR filters is their phase response is non-linear and they exhibit some degree of frequency dependent phase shift. However, this perception originates from the early days of high-order analog anti-alias and anti-image filters and ignores the advances in our understanding of the audibility of non-linear phase response or phase distortion over the past two decades. There has been a significant amount of research not only from a classical perspective but also in the research and development of head transfer functions.

5. Audibility of Phase Distortion

One of the confusing issues regarding the audibility of phase is that the discussion is generally considered to be a single topic when in reality should be discussed as two distinct situations. The audibility of phase distortion must be evaluated as follows:

1) Inter-channel phase distortion. Characterized as differences in phase response between two or more channels.
2) Intra-channel phase distortion. Characterized by non-linear phase response within a channel with the stipulation that the phase response is matched between all channels within the system (i.e. inter-channel phase distortion is equal to 0 msec)

6. Inter-Channel Phase Distortion

We use the amplitude and phase relationship between the sounds received by our ears to localize the source of the sound. Modern audio systems use this attribute to create what is known as imaging, or the perception that an instrument or vocal is coming from a location that is different than the actual speaker location. The audible effects of inter-channel phase distortion can be easily demonstrated by simply reversing the speaker connections on one channel of an otherwise properly configured stereo system. The loss of imaging is immediately noticeable even to those without a trained ear. Granted this test is rather dramatic and 180 degrees of inter-channel phase distortion is not indicative of standard operation but it does demonstrate the potential effects. As a result of this test, you would be hard pressed to find someone that would argue that 180 degrees of inter-channel phase distortion is acceptable, but where between the two extremes is the threshold of audibility? Tom Holman reports [10] that in his laboratory environment at the University of Southern California that is dominated by direct sound, a channel-to-channel time offset equal to one sample period at 48 kHz is audible. This equates to 20 µsec of inter-channel phase distortion across the entire audio band. Holman [10] also mentions, “one just noticeable difference in image shift between left and right ear inputs is 10 µsec”.

7. Intra-Channel Phase Distortion

Recall that we use the differences in signal amplitude and phase to localize or determine the source of sound and relatively small amounts of inter-channel phase distortion can be audible. But how does our hearing react when each channel in a multi-channel system is subjected to non-linear phase response but the phase response is matched between all channels? Douglas Preis [11] did an extensive survey of existing literature and Tom Holman's [10] experiences and research through his work at USC gives us an interesting insight into this phenomenon. Both report that the threshold of audibility is frequency dependent, which correlates with all other audibility thresholds. In laboratory environments when using test tones and headphones, research has shown that the human ear is sensitive to intra-channel phase differences of 0.25 msec [8] or +/-0.5 msec [9] in the mid-range with the threshold increasing at higher and lower frequencies. Preis states "the tolerances shown.... are not directly applicable to speech or music signals irradiated by loudspeakers in a reverberant environment. Most likely, the perceptual thresholds for these conditions would be at more than twice those shown". Essentially, the data suggests that for high quality music or speech reproduction in a reverberant environment intra-channel phase distortion of 1 msec is inaudible to a trained listener. Notice that this threshold is a relatively conservative statement and is still two orders of magnitude greater than that for inter-channel phase distortion!

8. Summary

In addition to the audibility of signal latency in live sound applications, there is a growing body of research that indicates that the audibility of pre-echoes is a primary source of the sonic differences between the lower sample rates, 44.1 and 48 kHz, and the higher sample rates of 96 and 192 kHz. However, minimizing pre-echoes creates an unacceptable cost structure and filter latency while minimizing filter delay creates unsuitable response in other filter response parameters. IIR filters offer the processing capabilities to minimize pre-echoes and time dispersion while maintaining acceptable cost and improved filter latency characteristics.

In regards to phase distortion, research indicates that the most critical area for phase distortion is inter-channel phase distortion where differences as little as 10 µsec can be perceptible in the most critical of situations. According to available research, it also appears that intra-channel phase distortion of 0.5 msec is inaudible with speech or music. The criteria stated in this paper to keep phase distortions inaudible can be easily achieved for IIR filters. One of the primary advantages of digital filters is that their response is repeatable. To meet the requirements for inter-channel distortion, this characteristic is crucial and it enables decimation and interpolation filters to have inter-channel phase differences approaching zero that is well below the audible range of 10 µsec. Various techniques are also available to keep intra-channel phase distortion well below the audible threshold. Knowledge of these thresholds creates the degrees of flexibility required to address the commonly acknowledged issue with digital filter latency for live sound applications as well as maximize other filter parameters to improve audio quality.

9. Acknowledgements

I would very much like to thank Rich Cabot, Peter Craven, Tom Holman, Bruce Jackson, Stanley Lipshitz, George Massenburg, James (JJ) Johnston and Ed Meitner for taking the time to discuss these topics and share their insights. I would also like to thank fellow Cirrus Logic employees Ray Rayburn, John Melanson and Raghu Rao for their contributions.

10. References

1) Lagadec, R. “Dispersive Models for A-to-D and D-to-A Conversion Systems” AES 75th Convention, March 27 - 30, 1984. Paris, France. Preprint 2097
2) Craven, Peter G. “Controlled Pre-response Anti-Alias Filters for use at 96kHz and 192kHz” AES 114th Convention, March 22-25, 2003. Amsterdam, The Netherlands. Preprint 5822
3) Craven, Peter G. “Anti-alias Filters and System Transient Response at High Sample Rates” Journal of the Audio Engineering Society; Volume 52 Number 3. March 2004
4) Dunn, Julian. “Anti-Alias and Anti-Image Filtering: The Benefits of 96kHz Sampling Rate Formats for those who cannot hear above 20kHz” AES 104th Convention, May 16-19, 1998. Amsterdam, The Netherlands. Preprint 4734
5) Reefman, Derk; Nuijten, Peter. “Why Direct Stream Digital is the best choice as a digital audio format” AES 110th Convention, May 12-15, 2001. Amsterdam, The Netherlands. Preprint 5396
6) McGuire, Monte. “96kHz: Hip or Hype?” Recording; June, 1999
7) Story, Mike. “A Suggested Explanation For (Some Of) The Audible Differences Between High Sample Rate And Conventional Sample Rate Material” private publication available at the dCS website; http://www.dcsltd.co.uk/Papers. September 1997
8) Stuart, J. Robert. “Coding for High-Resolution Audio Systems” Journal of the Audio Engineering Society; Volume 52 Number 3. March 2004
9) Lavry, Dan. “Understanding FIR (Finite Impulse Response) Filters - An Intuitive Approach” private publication available at the Lavry Engineering website; http://www.lavryengineering.com. 1997
10) Holman, Tomlinson. “Sample Rate” & “Word Length aka Bit Depth, Resolution, and many other aliases” Surround Professional, Volume 1, Issues # 5 & # 6
11) Preis, Douglas. “Phase Distortion and Phase Equalization in Audio Signal Processing --- A Tutorial Review” AES 70th Convention, October 30 - November 2, 1981. New York, New York. Preprint 1849

11. Related and Interesting Articles

Lavry, Dan. “Understanding IIR (Infinite Impulse Response) Filters - An Intuitive Approach” private publication available at the Lavry Engineering website; http://www.lavryengineering.com. 1997

Rao, Raghunath. “Optimal Head Related Transfer Functions for Hearing and Monaural Localization in Elevation: A Signal Processing Design Perspective” IEEE Transactions on Biomedical Engineering, Vol. 43, No. 11, November 1996

Contacting Cirrus Logic Support

For all product questions and inquiries contact a Cirrus Logic Sales Representative.
To find one nearest you go to **http://www.cirrus.com/**