Overload in Signal Conversion

Søren H. Nielsen[1] and Thomas Lund[1]

[1] *TC Electronic A/S, Risskov, Denmark*

Correspondence should be addressed to Søren H. Nielsen (`soerenn@tcelectronic.com`)

ABSTRACT

Digital audio material is often mastered at such hot levels that frequent clipping in the digital domain occurs. While this in itself can cause listener fatigue the sound is often severely additionally distorted in the digital to analog conversion process at the listener end due to lack of headroom in the reproduction. Purely digital processing such as sample rate conversion and perceptual codecs may exhibit the same problem. This paper describes measurements to quantify the problem, and new tests performed on commercial recordings to demonstrate the phenomena. Furthermore, methods to avoid the distortion are discussed and demonstrated.

1 INTRODUCTION

The final signal processing stage in production of audio material is called mastering. While the art of mastering 20 years ago involved getting the best results from a consumer LP reproduction, it seems now to have become the process in which every tune is normalized to digital full scale. Because there are no spectral restraints from the LP emphasis/de-emphasis process, and no level restriction besides from an ancient rule of how many consecutive samples are allowed at full scale, mastering has actually turned into a loudness war. The mastering engineer is its field marshall, with a number of secret weapons to squeeze the last 0.1 dB of loudness out of any song or movie.

The loudness war, as we shall see, has now become so furious that equipment downstream of mastering is not able to pass or reproduce the audio material without adding significant distortion to it. It appears to be a contradiction that the professional audio industry promotes better standards such as 96 kHz sampling, 24 bit resolution, DSD, exotic dither schemes etc. and the hi-fi world is becoming conscious about jitter-rejection etc. – while the current main delivery format, 16 bit/44.1 kHz, is becoming more and more abused. Most pop releases now get so much distortion added during reproduction that 16 bit resolution is completely overkill.

The distortion problems arise because of more or less conscious use of clipping in the digital domain during production and mastering. Digital processors may have familiar names like "Limiter" or "Compressor", but there is no guarantee they will have much in common with their analog counterparts, and not even that they obey the fundamental sampling criteria. Clipping can also be hidden inside computer processors, DAWs or digital mixers, making later recovery difficult or in many cases impossible.

Figure 1 shows a block diagram of a music or post production studio with a computer workstation or a digital mixer at the heart of it. Headroom critical areas exist around all digital or analog filters, plug-in or external processing, summing busses, limiters, sample rate converters and digital to analog converters. Sufficient indication of internal clip in computers or processing plug-ins even on single sources is mandatory, if contamination of the final mix is to be prevented. It is also worth considering if studio speaker monitoring should include 0 dBFS+ artifacts or not, and how signal should be displayed

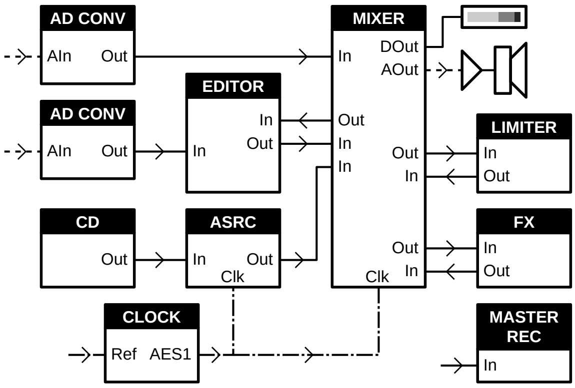

Figure 1: *Typical digital studio set-up.*

on level meters or histograms.

While frequent clipping in the digital domain can cause listening fatigue, things are made worse during mastering when clipped material is normalized to digital full scale. Such material will later be subject to additional distortion during data-reduction, sample rate conversion and digital to analog conversion.

At first glance it seems to be a paradox that overload at these stages can happen but it is not. The phenomenon is not limited to conversion into the analog domain. Also purely digital processing such as sample rate conversion or data rate reduction coding is critical.

Both narrow-band and broadband (e.g. clipped) signals can have a theoretical peak value larger than the actual representation in the digital format. The simplest example is a high frequency sine wave. Depending on frequency and phase the peak value of the digital sine wave representation may be up to 3 dB lower than the theoretical, or analog, peak value. If this headroom is used for increasing the digital gain, the peak value of the reconstructed analog signal will be higher than the DC value representing digital full scale, 0 dBFS.

Unfortunately, this is not just a theoretical problem. In practice, all digital converters are equipped with interpolation filters which by their very nature cause overshoot.

Earlier work by the authors [1, 2] have indicated this problem and presented measurements on the analog outputs of CD players showing clipping due to inter-sample peaks.

It may be argued that a single occational overshoot is inaudible, but this short overshoot does not necessarily stay short. Due to recursive filters, analog domain latch-up, protection systems or other prolonging effects in downstream processing or reproduction equipment an originally short overhoot may become much more audible.

The problem of inaccurate peak level estimation at certain frequencies has been recognised also by the IEC, as stated in Annex A.3 (Amplitude-frequency response) in [3]: "(...) However, a response irregularity may occur at specific frequencies related to the sample frequency. The requirement therefore does not apply at these exceptional frequencies."

In order to further test how critical the overshoots are in practice several measurements were carried out using both artificial test signals and commercially available recordings as test signals. First of all, the severity of the problem in existing equipment, both digital to analog and digital to digital converters, was determined. Next, various strategies for avoiding the problem were tested.

2 THEORETICAL CONSIDERATIONS

As trivial as it may sound, the basic cause of overload in signal conversion is lack of waveform preservation in the conversion process. There are several possible causes for this waveform change:

- The very nature of sampled signal representation
- Change of phase
- Change of bandwidth
- Non-linear distortion such as clipping

Often, a combination of these factors cause the trouble.

2.1 Phase change

Even the simplest of waveforms, the sine wave, can be sensitive to phase changes in a way which causes overload by a conversion process. This is very easily demonstrated: Consider a cosine wave at $f_S/4$. At a starting phase of $0°$, the digital representation of the

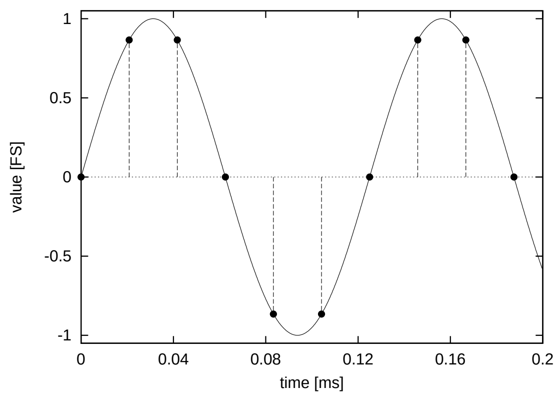

Figure 2: *Analog and a possible digital representation of 8 kHz ($f_s/6$) sampled at 48 kHz.*

waveform includes the maximum theoretical magnitude of the waveform. With a starting phase of $45°$, however, the digital representation contains only the values $\pm\frac{\sqrt{2}}{2}$times the theoretical maximum of the waveform, corresponding to -3 dB. At other frequencies related in a simple way to the sampling frequency, a similar difference between theoretical and practical peak value occurs, see for example Figure 2.

If the digital samples were derived from an analog source without overload and not processed in the digital domain, there would be no potential overload problem at the conversion to analog, or to another sample rate.

Typically some digital processing takes place, however, for example a gain and/or phase change. In the digital domain the samples may still be perfectly valid, but when converted the peak value may rise to a value larger than the DC value representing digital full scale.

The phase of the individual components in an audio signal is closely related to the crest factor, i.e. the ratio between the peak and the RMS value. Some signals, such as speech, constains many narrow peaks, corresponding to a high crest factor, whereas others, like square waves, have have a low crest factor – down to 1 for a symmetrical square wave.

In the world of broadcast audio processing a so-called phase rotator is often used in order to decrease the crest factor of the speech signals in order to squeeze more avarage energy (RMS level) into a medium with a fixed maximum peak value [4, 5]. This is done by changing the phase over a broad frequency range containing the problematic signal, thus causing peaks to be spread out in time. Depending on the input signal the phase rotation may also have the opposite effect, however, by increasing the peak value of signals with rather flat tops.

Encoding 4-2-4 matrixed surround [6] sound typically also involves a large change of phase over most of the audio frequency range in order to maintain a constant phase *difference* between the front and surround channels. The possible audibility of these phase-rotator types of phase change will not be discussed here.

2.2 Bandwidth change

Removing part of the spectral content of a broadband signal is very likely to cause a peak level *increase* – despite the reduced energy in the resulting signal.

A very simple example of this is a low frequency limitation, i.e. a high pass filter, which is present at most analog interfaces in order to avoid DC offset. Filtering a low frequency square wave with just a first order high pass filter will result in a peak level increase of up to 6 dB. Sometimes, such filters occur also within the digital domain. The effect of low frequency, high pass filtering a square wave is illustrated in Figure 3. It may be argued that some of the peak level increase is caused by the phase change rather than the gain change but that doesn't change the fact that the peak level rises when applying a high pass filter to this type of signal.

In order to illustrate the mechanisms of high frequency limitation, a relatively simple broadband signal, a square wave with a period time of 2π, corresponding to an angular frequency of $\omega = 1$, will be used. The Fourier series description consists of an infinite harmonic series of cosine waves with simple frequency and magnitude ratios and where the sign of the harmonics is alternating, see e.g. [7]:

$$square(\omega) = cos(\omega t) - \frac{1}{3}cos(3\omega t) + \frac{1}{5}cos(5\omega t) \; ...$$

The effect of constructing a square wave from the individual cosine components is illus-

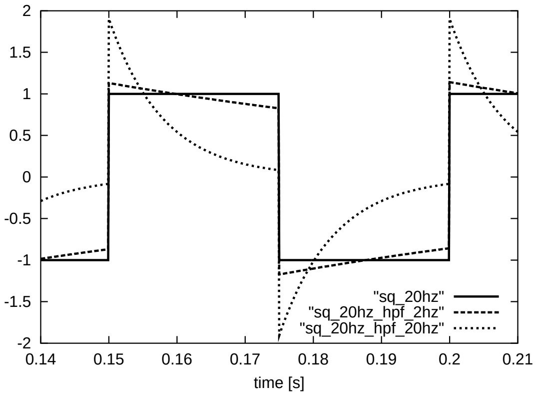

Figure 3: *First order high pass filtering a 20 Hz square wave at 2 and 20 Hz.*

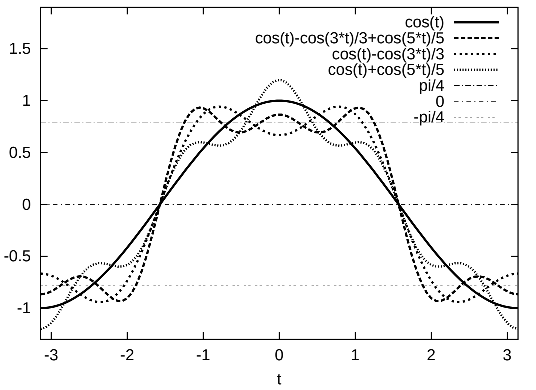

Figure 4: *Construction of a square wave of harmonic components. The effect of omitting a lower order harmonic is also illustrated.*

trated in Figure 4. For each higher harmonic added the peak magnitude decreases slightly. Continuing this Fourier series infinitely yields a square wave with a peak (and RMS) value of $\frac{\pi}{4}$.

The effect of omitting the third harmonic of a (1, 3, 5) harmonic series is also illustrated. As seen, omitting this low order harmonic (corresponding to using an ideal bandstop filter) increases the peak value substantially.

For a given filter it is possible to construct a test signal which will excite the filter in a worst-case fashion with respect to achieving a maximum output peak level, i.e. to create an impulse-like output. As such a test signal would have to be adapted to each specific filter it will not be used here.

A more common extreme signal is a pseudo-random sequence, alternating between plus and minus full scale at (pseudo-)random intervals. The MLS sequences often used for acoustical measurements are of this type.

Some audio systems change bandwidth dynamically. In particular, perceptual coding systems will change bandwidth depending on the allocation of bits to different frequency regions. This makes it difficult to predict the exact change of waveform.

2.3 Clipping

Constructing a square wave in the digital domain from a harmonic series will be limited by the Nyquist frequency, and thus the square wave will be rounded and oscillating as shown in Figure 4. An alternative way of generating square waves in the digital domain is by clipping. This may be done more or less intentionally but it is known to happen in many commercial recordings. The resulting signal is not necessarily a pure square wave but it typically has a flat top and sharp edges achieved by saturation logic.

A digital square wave with its steep slopes, sharp edges and flat top does not fulfill the sampling theorem. Such a waveshape in principle requires an infinite bandwidth, and the practical digital systems in use at present are not even of *high* bandwidth. The result is therefore aliasing – a perceptually unpleasant artefact.

The effects of truncation of the Fourier series in order to represent a square wave in a finite bandwidth system is sometimes described as *Gibb's phenomenon*, see e.g. p. 239-241 in [8]. The phenomenon plays an important role in windowed FIR filter design where the resulting ripple (oscillations) in the passband and stopband frequency response is a design parameter.

The analog output of a digital audio system normally contains a lowpass filter in order to remove the images of the baseband spectrum occuring periodically around each integer multiple of f_s. Thus, it is the output filter which creates the rounded and oscillating shape of the reconstructed square wave. A sample rate converter contains a similar filter and thus suffers the same problem.

3 CD OVERLOAD FREQUENCY

As demonstrated in the previous section, inter-sample peaks are easy to generate using artificial signals. To determine if hot signals challenging the headroom of a downstream signal path are becoming more or less frequent, we investigated a number of commercial pop and rock CDs. The CDs have not been randomly chosen. Besides from being pop or rock songs, they reflect the CD collections, and therefore to some extent the musical preferences, of one of the authors and his children.

The measurements made were:

- Slow average "loudness" reading utilizing a TC Electronic P2 dynamics processor.
- Maximum repeated digital peak level as encoded on the CD.
- Estimated typical number of occurrences of level between 0 dBFS and +1 dBFS per 10 seconds.
- Estimated typical number of occurrences with level exceeding +1 dBFS per 10 seconds.

Table 1 shows the list of tracks and the measurement results. Some further explanation to the table: The Sum is a mastering rating indicating track's susceptibility to reproduction distortion:

0: Low level. Not full use of CD dynamic range, no problems besides from reduced S/N ratio during reproduction.

1: Well aligned level. Fully used CD dynamic range, no problems to be expected during reproduction.

2: Level to the hot side. Probably no reproduction distortion.

3: Hot level. Distortion through codecs, SRC and DA converters to be expected.

4: Very hot level. Obvious distortion through codecs, SRC and DA converters to be expected.

Even if comparisons are restricted to the international "power pop" tracks of the table, we believe it signifies a trend towards more and more level maximization and digital domain clipping. They are becoming the kind of average signals all pro and consumer equipment must be prepared to process and pass. It should be noted that new tracks targeted for a mature audience, like Santana's Smooth, also have been subjected to extreme level optimization. Remastered albums are normalized to 0 dBFS, rather than their original CD level, but in these cases don't exhibit much additional clipping. This might be an indication that clipping is more frequently used in production than in mastering, or that an analog signal source like a 1/2" master tape can prevent some of the level problems from getting out of control.

4 MEASUREMENTS

4.1 General procedures

The main purpose of the measurements is to compare the tested systems under stressed and under relaxed conditions to establish to which degree the performance is altered when applying critical audio material. As the basic performance level varies greatly among audio systems, a reference measurement under relaxed (uncritical) conditions is necessary.

Three types of measurements were made: Spectrum analysis with THD+N, time domain measurements, and measurements of the error signal.

When using simple signals such a sine waves, spectrum analysis is a strong tool to identify distortions, harmonic as well as aliasing. Playing sine waves with different theoretical peak values through an audio system will reveal its basic overload behaviour.

Similarly, a time domain measurement using an oscilloscope and signals with known waveforms will show the basic overload behaviour by simple inspection.

For complex signals these methods are less suitable, however, so an alternative technique was developed. This technique aims at producing the error signal, i.e. the difference between the distorted and the ideal output

	Track	*Notes*	*Artist*	*Year*	*Slow avg. [dB]*	*Max. dig. [dB FS]*	*Hot spots 1*	*Hot spots 2*	*Sum*
1	Lose Yourself		Eminem	2002	-7.5	0.0	>25 s.	>25	4
2	Time of My Life		Macy Gray	2002	-8	0.0	16	8	3
3	Happy		Ashanti	2002	-9	0.0	18	10	3
4	La Fiesta De Amadito		Amadito Valdez	2002	-10	0.0	2	0	1
5	Don't Stop	PP	Anastacia	2001	-6	0.0	>25 s.	15	4
6	Played Alive		Safri Duo	2001	-7	0.0	>25	16	3
7	The Call	PP	Backstreet Boys	2000	-5	0.0	>25 s.	18	4
8	Livin' la Vida Loca		Ricky Martin	1999	-6.5	0.0	12	5	3
9	Need to Know		Marc Anthony	1999	-7	0.0	19	10	3
10	Razor Tongue		DJ Mendez	1999	-6	0.0	17 s.	9	4
11	I Got a Girl		Lou Bega	1999	-6.5	0.0	>25 s.	3	4
12	Let's Get Loud	PP	Jennifer Lopez	1999	-6	0.0	>25 s.	10	4
13	Smooth		Santana	1999	-7	0.0	20 s.	15	4
14	Oye Como Va	RM	Santana	1970 1999	-12	0.0	0	0	1
15	Avalon	RM	Roxy Music	1982 1999	-9	0.0	5	0	2
16	Believe		Cher	1998	-5.5	0.0	10	4	2
17	Miami		Will Smith	1997	-11	0.0	17	9	3
18	That Don't Impress...		Shania Twain	1998	-9	0.0	3	0	2
19	Vissa Har Det		Bo Kaspers Ork.	1998	-11	0.0	1	0	1
20	True Colors		Phil Collins	1998	-12	0.0	1	0	1
21	Block Rockin' Beats		Chemical Bros.	1997	-6	0.0	8	5	2
22	El Cuarte de Tula		Buena Vista SC	1997	-12	-0.2	0	0	1
23	Dimples		John Lee Hooker	1997	-11	0.0	0	0	1
24	Bla Bla Bla		Oestkyst Hustlers	1996	-9	0.0	3	0	2
25	Bob Yu Did Yu Job		Jimmy Cliff	1996	-12	0.0	6	1	2
26	Where It's At		Beck	1996	-10	0.0	1	0	1
27	Wannabe	PP	Spice Girls	1996	-8	0.0	5	0	2
28	The Only Thing		Bryan Adams	1996	-9	0.0	2	0	2
29	We'll be Together		Sting	1994	-12	-0.2	1	0	1
30	Off the Ground		Paul McCartney	1993	-12	0.0	1	0	1
31	I've Been to Memphis		Lyle Lovett	1992	-16	-0.9	0	0	1
32	Good Stuff	PP	B52's	1992	-12	0.0	5	0	2
33	Gloria's Eyes		B. Springsteen	1992	-11	0.0	0	0	1
34	Mysterious Ways		U2	1991	-11	-0.1	0	0	1
35	Something to Talk...		Bonnie Raitt	1991	-14	-0.9	0	0	1
36	Black or White	PP	Michael Jackson	1991	-11	-0.2	0	0	1
37	The End of the...		Don Henley	1989	-13	-2.2	0	0	1
38	Dirty Blvd		Lou Reed	1988	-14	-0.2	0	0	1
39	Nick of Time		Bonnie Raitt	1989	-17	-2.1	0	0	1
40	Living in America		James Brown	1986	-16	-2.6	0	0	1
41	Graceland		Paul Simon	1986	-16	-3.4	0	0	0
42	Two Tribes	PP	Frankie Goes...	1984	-6.5	-0.7	1	0	1
43	She Took Off My...		David Lindley	1981	-13	-1.9	0	0	1
44	Little Sister		Ry Cooder	1979	-22	-8.7	0	0	0

Table 1: *Statistics on average and peak levels, and on inter-sample peaks of various commercial recordings.*
The following abbreviations are used:
PP: Power pop, *RM:* Remastered.
Slov avg.: Slow average weigthed "loudness" reading utilizing the TC P2 dynamics processor.
Max. dig.: Max. repeated digital peak level as encoded on the CD.
Hot spots 1: Estimated typical number of occurrences of level between 0 dBFS and +1 dBFS per 10 seconds The label "s." is used when the track exhibits extended periods of 0 dBFS+ signal as opposed to easily identifiable 0 dBFS+ peaks.
Hot spots 2: Estimated typical number of occurrences with level exceeding +1 dBFS per 10 seconds.
Sum: Mastering rating indicating track's susceptibility to reproduction distortion, see text.

signal. The error signal will only be present when the signal itself is strong, so part of the errors may be masked by the signal. Despite this, the authors consider it useful and enlightening to measure and to listen to the error signal, especially with musical source material.

In order to derive an error signal in a single-ended system such as a CD player some assumptions about the measured system are made: 1) The measured system is largely linear in the sense that no signal or level dependent processing takes place – apart from the potential overload behaviour. 2) The measured system can reproduce at least two channels simultaneously, with only a minor and constant difference in gain and phase between channels.

With these two assumptions fulfilled, a two-channel audio system can be used to simultaneously produce the reference condition and the critical condition, as one channel is used for each. The signal level in the reference channel is kept sufficiently low that no clipping will occur. The other channel is then used for stressing the system. By mixing the two output signals together with appropriate gain and phase the result then represents the level dependent error by the conversion process. Ideally, this signal is very close to zero.

4.1.1 Test signals

The test signals were constructed in a such a way that the error caused by clipping due to limited headroom in the converters could be easily derived: The piece to test was recorded on a CD-R in one channel without changes. The other channel on the CD-R contained the same signal but at a lower level, in this case scaled by exactly 0.5 (1 bit). In this way the signal of lower level would never get clipped, whereas the loud one might be clipped.

Stationary sine waves of four different frequencies were used. One of them (997 Hz) did not have any simple relation to the sample rate and was as such well suited for general level calibration. The other three (11025, 7350 and 5512.5 Hz) were integer fractions of the Nyquist frequency and can therefore be set to have theoretical peak values exceeding the (valid) digital representation. The frequencies and the corresponding starting phases of the sine waves are shown in Table 2. A whole range of levels ranging up to the indicated ones were constructed.

Frequency		*Phase*	*Max. peak*
11025 Hz	$f_s/4$	45°	+3.01 dB FS
7350 Hz	$f_s/6$	60°	+1.25 dB FS
5512.5 Hz	$f_s/8$	67.5°	+0.69 dB FS

Table 2: *Test frequencies, starting phases and maximum theoretical peak values.*

In addition to the stationary sine waves, stationary square waves were constructed at 20, 50, 551.25, 5512.5 and 11025 Hz. The frequencies of these square waves are all at integer fractions of the Nyquist frequency and they were constructed as ideal square waves, so they consist of only two sample values – identical and with opposite signs. Due to the choice of frequencies, no intermediate values need being interpolated. It should be noted, however, that these square waves do not fulfill the sampling theorem. But by design, the aliasing products in these signals coincide with the harmonics already present in the signal, so no inhamonicity occurs. The bandwidth limitation of the output filter will truncate the infinite Fourier series of the square wave and thus generate the oscillations described earlier. The low frequency square waves are useful for checking overswings due to a possible lower cut-off frequency of the converter circuit.

A special square wave kind of signal is the maximum length sequence (MLS) – a pseudorandom binary sequence. While this signal is not the worst thinkable in terms of overload behaviour it stresses the converter significantly and it is well suited for time domain inspection.

Finally, various music material taken directly from commercial CDs were used. We found it important to verify errors sonically and *in vivo*, rather than relying only on constructed signals, like the ones described above, and FFT measurements.

Material used: TC Level Test CD with alignment material and commercially available music exhibiting various degrees of clipping. Left channel is a clone of the left channel of the original CDs. The other channel also is a clone of the left, but shifted one bit down in level. This is illustrated in the left part of Figure 5.

4.1.2 Setup

All measurements were made with a source sample rate of 44.1 kHz and a CD-R as the signal generator. It was verifyed that the digital output of the CD player did not process the audio in any way. Measurements were performed on the test signals described in the previous section.

In vivo measurements and error signal listening was performed using the setup of Figure 5. The outputs of the DUT were level calibrated using fully correlated and phase aligned tone and noise tracks at -6 dBFS and -12 dBFS on the CD. Before summing the channels, a phase reversal and enough gain to completely cancel the output was applied to the left channel. When replaying the real-world music material, non-linear discrepancies between the channels can be readily heard and measured on the output of the mixer.

Distortion and noise measurements were made with an Audio Precision System 2 Cascade analyser.

4.2 Measurements on analog outputs

In an earlier paper by the authors, measurements on typical CD players were presented [1]. As a continuation, the digital to analog converters of some studio equipment have also been measured. The source signals were hot sine waves at various frequencies. Results of the THD+N measurements are shown in Table 3.

The measurements were all made with a 16 bit source signal, so the reference measurement at 997 Hz does not show very impressive figures, even though some of the tested units have DACs with up to 24 bits and stated performance levels well above the measured figures. Using a measurement bandwidth of 80 kHz worsens matters, of course, both as a consequence of the increased bandwidth (4 times the normal, 6 dB

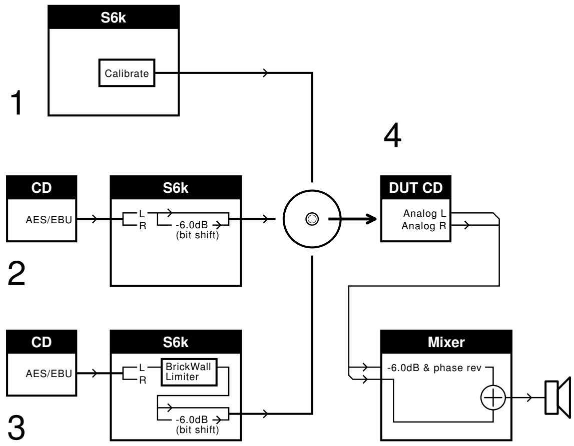

Figure 5: *Construction of signals for test CD and measurement setup for derived error signals. The device under test (DUT) could be a CD player.*

more noise power) and the possible presence of inaudible high frequency components not filtered away by the analog anti-image filter. The DAC themselves were oversampled to various degrees and contain as such a digital reconstruction filter.

A somewhat mixed picture is seen at the higher input levels. In that respect the results are similar to those of the CD players [1]. In those cases where the DACs were less sensitive to hot input signals it could be either due to a more conservative analog design, extra headroom in the DAC chip itself, or a headroom created in the digital domain before the signal is sent to the DAC.

4.3 Measurements on digital to digital conversion

The most important type of processing in this category is sample rate conversion. Other conversion types with potential problems could be time and pitch scaling.

Nowadays, sample rate conversion is almost exclusively integrated with other types of processing, such as in a hardware signal processing unit or in an audio editing software package. In the hardware case, a special purpose chip is often used instead of using the general DSP(s) also present in the unit.

The present measurements cover two sample rate conversion chips (Analog Devices AD1890 and Crystal CS8420) which are both

Freq. [Hz]	*Level [dB FS]*	Workst. ProTools	Mixer Yamaha O2R	FX proc. Lexicon PCM90	FX proc. TC Triple C	FX proc. TC S6K 3 dB HR	SACD Sony NS900V
997	0.00	-71.5	-86.0	-89.7	-82.7	-91.3	-67.0
5512.5	+0.69	-64.0	-30.1	-29.2	-81.8	-88.9	-57.1
7350	+1.25	-61.4	-24.3	-24.0	-79.2	-88.9	-61.1
11025	+3.00	-27.3	-16.2	-15.9	-23.8	-86.3	-28.0

Table 3: *THD+N in dB (20 Hz – 80 kHz) on the analog outputs of some pro-audio equipment fed with sine waves.*

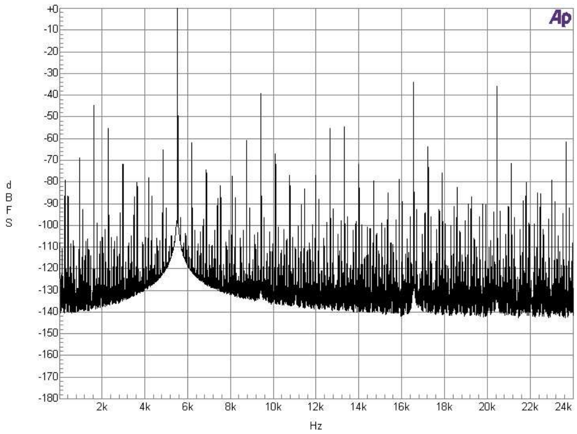

Figure 6: *Spectrum of a sample rate converter (CS8420) fed with $f_s/8$ at +0.69 dBFS. 44.1 -> 48 kHz. Notice the many aliasing products.*

in common use, e.g. in some equipment from our company. An example of a measurement is shown in Figure 6. A sample rate converter was fed with a slightly hot sine wave at a sample rate of 44.1 kHz and converted to 48 kHz. The resulting spectrum shows clearly the many harmonic and aliasing distortion products. The input frequency, if clipped at its original sample, would not generate aliasing, but as the rate is changed the integer relationship between signal and sampling frequency is no longer maintained, so the aliasing products occur all over the spectrum instead of just at the harmonics of the input frequency.

4.4 Measurements on perceptual codecs

Inside perceptually based data rate reduction coding systems the signal is filtered and quantised, often with quite narrowband filters [9, 10, 11, 12]. The filter banks or transformations used by these coding schemes are perfectly (or very near perfectly) reconstructing, provided that no quantisation takes place. Thus the perceptual codecs have the potential of being waveform-preserving and do therefore not necessarily exhibit critical behaviour when handling hot signals.

In general, it is required to have access to digital inputs and outputs of the digital-to-digital conversion systems under test, but even in the absence of such it would be possible to make *some* tests, for example to check the possible change of the time domain behaviour as a function of the data rate.

4.4.1 Coding schemes tested

In order to investigate the influence of different encoding settings, various combinations of coding algorithms, data rates and coding modes were tested, see Table 4. Especially the encoding data rate was expected to influence the effective bandwidth of the encoded signal as varying bandwidth is a relatively unobtrusive way of saving bits. Due to the varying demand for bits during a piece of audio the bandwidth may change dynamically, if the data rate available is limited and the encoder is forced to find a best compromise between quantisation noise and encoded audio bandwidth. As seen earlier, reducing the bandwidth may cause the peak value to rise.

A purely theoretical analysis would be difficult due to lack of publicly available documentation for some coding schemes, so instead measurements were made on a selection of the available perceptual codecs. The tested codecs were:

- MPEG-1 Layer II and Layer III [9].
- DTS [11].

Algorithm	*Mode*	*Datarate [kbit/s]*	*Avg. per ch. [kbit/s]*	*Max. peak re. 0.5*
MPEG-1 L II	stereo	384	192	+1.3 dB
MPEG-1 L II	stereo	224	112	+1.3 dB
MPEG-1 L III	stereo HQ	320	160	+1.7 dB
MPEG-1 L III	stereo HQ	160	80	+2.3 dB
MPEG-1 L III	int-st HQ	128	64	+5.3 dB
MPEG-1 L III	int-st fast	128	64	+3.0 dB
MPEG-1 L III	int-st HQ	96	48	+4.7 dB
MPEG-2 L III	22.05 kHz, i-st HQ	80	40	+1.7 dB
DTS	6 ch.	1234	206	+0.6 dB
Ogg Vorbis	stereo	var., Q=10	157-193	+0.3 dB
Ogg Vorbis	stereo	var., Q=5	49-64	+1.8 dB

Table 4: *Maximum peak values observed in 12 hot CD excerpts (length 14-33 s) perceptually coded with various algorithms, data rates and modes.*

- Ogg Vorbis [12].

MPEG-1 Layer II encoding was done using the publicly available ISO/MPEG reference software, ver. 3.9.

MPEG-1 Layer III encoding was done using the FhG encoder integrated in an audio editing software (TC Works Spark). Stereo as well as intensity stereo coding mode was used. Quality settings of “Fast” and “High Quality” were used.

MPEG decoding was done digital to digital using two programs: The ISO software and a freeware package called “mpg123” [13], ver. 0.59.

Being international standards, detailed specifications of the MPEG family of coding schemes are publicly available [9].

DTS encoding was done using a software encoder program (Minnetonka SurCode DTS-CD). Decoding was made with a professional realtime decoder (DTS CAD-4) using digital PCM outputs. An overall description of this coding scheme is publicly available [11]. The stereo signal to be coded was placed three times, as L/R, C/LFE and LS/RS, in this 5.1 encoder in order to load it realistically.

Ogg Vorbis is one of several not formally standardised coding schemes used on the Internet. This particular scheme is open source, and full documentation is found at the website [12]. Software encoders and decoders of version 1.0 were used for these tests. All signals were encoded in stereo mode.

It should be noted here that it is theoretically possible for clipping to occur both at the very output of a codec (in the fixed point PCM domain) as well as in the subband or transformed representation (typically block floating point) used for coding. This depends on the headroom structure of the coding scheme.

The test signals used were 12 excerpts from contemporary CDs, signals that are partially clipped and show levels well in excess of 0 dBFS on an oversampling peak meter. The length of the excerpts was 14–33 seconds. The encoded signals were dual mono with a level reduction of 6 dB in the right channel. Due to the non-linear nature of perceptual coding it cannot be guaranteed that the number of bits allocated (if any) to various frequency bands of the two channels is identical. Thus, constructing a difference signal between the decoded left and right channels may not be as informative with respect to overload behaviour as in the case of digital to analog converters and sample rate converters.

4.4.2 Results

In general, it may be noticed that the peak level indeed does rise when applying perceptual codecs on very hot audio signals. As the peak level of the decoders was limited to ±1, the channel with full scale input signal was clipped more or less frequently. The half-level channel was used to measure the size of the overshoots occuring. Results of this measurement are shown in Table 4. Only the highest peak value occuring across the 12 audio excerpts is listed here. As seen in the table quite high overshoots can occur, depending on the coding scheme and its parameters.

The size of the overshoots corresponds quite well to the encoded data rate, in that the lower data rates generate higher output peak values than the higher data rates.

Looking at the results in more detail reveals some interesting facts:

The highest overshoots in this series of measurements occurred with MPEG-1 Layer III, especially for the lower data rates. An interesting exception is the lower overshoot for MPEG-2 Layer III, at half sampling rate.

There are some noticable differences between MPEG Layer II and III in terms of overload behaviour: The reconstructed peaks/overshoots in the PCM domain are larger in Layer III than in Layer II, i.e. there is more more PCM domain clipping in Layer III. This is seen by comparing the decoded half scale channels. This difference could be due to the finer frequency resolution and corresponding steeper filter shapes. Furthermore, it is noticed that the internal headroom is larger in Layer III than in Layer II. This is seen when the PCM output is scaled down to, say, half. The difference between the full scale and the half scale channels is then smaller for Layer III than for Layer II. So internally, in the narrowband domain, Layer III is more well-behaved with this type of (very hot/clipped) input signals.

DTS: In the particular encoder software used the bitrate was fixed and quite high. Only up to 0.6 dB overshoot was observed – which indicates only little bandwidth limitation. This codec is probably safe with most DACs.

Ogg Vorbis: The overshoots are highly dependent on quality setting: Q=10 is almost overshoot-free whereas Q=5 gives overshoots up to 1.8 dB. For fixed bitrate (not shown in the table) overshoots were also common and large. The Ogg Vorbis coding system seems to be designed for files and streaming over variable bandwidth connections. This is in contrast to most other perceptual coding schemes, where making the best of a fixed bitrate has been a design goal. MPEG and other coding schemes also do work with variable bitrate, however.

From the above it is clear that perceptually based data rate reduction coding schemes are just as critical as other digital conversion with respect to handling of very hot input signals.

4.5 Observations

The measurements covered the analog outputs of studio equipment, sample rate converters, and data rate reduction codecs. Many of the commercially available recordings caused the error signal to be quite loud: Error peaks at 20-30 dB below the signal level (which was close to full scale) were common, and the error signal sounded very harsh. Digital to digital converters tended to sound worse than digital to analog converters due to clipping exclusively in the digital domain.

5 METHODS TO AVOID OVERLOAD

One way of avoiding the problem of clipping is to reduce the overall peak level of the recordings. As this also reduces the average level it will probably be considered unacceptable by many sound producing professionals, so another solution is called for.

An alternative is to to reduce the digital level in the playback environment prior to digital to analog conversion. This would reduce distortion only in future products, however, leaving hundreds of millions of playback devices with potential problems. The price for this constructional measure is a reduced signal-to-noise ratio in the D-to-A conversion circuit.

The amount of overshoot by conversion can be estimated, however, and this estimated peak value can be used to control a fast limiter which automatically reduces the gain of just those instants in the sound where overshoot would take place.

Both of these strategies have been tested and proven to deliver converted signals of significantly reduced distortion. For the fixed gain reduction case, a small headroom was found to be sufficient to avoid problems in the tested converters.

6 VERIFICATION

Using an inter-sample precision digital limiter with the threshold value set to -0.3 dBFS, that is, basically just to reduce overshoots, the distortion of the DACs disappeared almost completely. The overall loudness of the signal was not reduced, however, so for many applications this seems to be the better option.

Applying an inter-sample peak limiter before the perceptual codecs resulted in less clear results: The MPEG Layer II codec reduced the peak level both overall and the maximum peak value when the input signal had been treated with the limiter. The average size of the peaks was also reduced in the Layer III case, but certain of the limited input signals generated a few higher peaks than when the input signal had not been limited. This was surprising to the authors and shows the difficulty as estimating exactly what happens downstream in a signal chain and taking appropriate action.

7 CONCLUSIONS

In conclusion, we have demonstrated that the problem of overload in digital to analog conversion, sample rate conversion and in connection with data-reduction systems is definitely present in practice.

Measuring and listening to *in vivo* signals provide a strong evidence that headroom in the digital production chain has been exhausted, and that audio quality currently may be limited for this reason rather than a basic lack of resolution or a high enough sample rate.

We have discussed remedies that can be used to avoid the problem without having to modify the large base of playback equipment already in use and without having to change the general level of the recorded material.

Because a music or film mix may already have been polluted with processing not obeying the sampling criteria, a mastering engineer may consider "cleaning up" source material by passing digital material through the analog domain, or an asynchronous sample rate converter with the capability of attenuating the input signal. Alternatively, analog source tape in mastering is another way of preventing distortion from staying hidden until it reaches a radio processor or a consumer CD player.

Because the audio industry is currently stuck with non-intelligent level restrictions, such as counting consecutive samples at full scale, mastering engineers should start using oversampled meters and limiters, or at least normalize against e.g. -2 or -3 dBFS rather than 0 dBFS. It should also be noted that some data reduction codecs are even more sensitive to excessively hot level than a linear signalpath.

8 REFERENCES

[1] Søren H. Nielsen and Thomas Lund. 0 dBFS+ Levels in Digital Mastering. AES Convention Preprint 5251, 2000.

[2] Søren H. Nielsen and Thomas Lund. Level control in digital mastering. AES Convention Preprint 5019, 1999.

[3] IEC. Sound system equipment. Part 18: Peak programme level meters - Digital audio peak level meter. Technical Report 268-18, International Electrotechnical Commision, 1995.

[4] Frank Foti and Robert Orban. What Happens to My Recording When it's Played on the Radio? White paper, Omnia Audio, CRL/Orban, 2001. http://www.omniaaudio.com/.

[5] Orban. *Optimod-FM 8200 Digital Audio Processor Operating Manual*, 1992. http://www.orban.com/.

[6] Roger Dressler. Dolby Pro-Logic surround decoder - principles of operation. White paper, Dolby Laboratories, 1993.

[7] Murray R. Spiegel and John Liu. *Mathematical Handbook of Formulas and Tables*. Schaum's Outline Series. McGraw-Hill, New York, 2nd edition, 1999. ISBN 0-07-038203-4.

[8] Allan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1975. ISBN 0-13-214635-5.

[9] International Standard ISO/IEC 11172-3, Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s, Part 3: Audio, March 1993.

[10] ATSC. Digital audio compression standard (AC-3). Document A/52, Advanced Television Systems Committee, 20 December 1995.

[11] Mike Smyth. An overview of the Coherent Acoustics coding system. White paper, DTS, Inc., June 1999. http://www.dtsonline.com/.

[12] Ogg Vorbis. The Ogg Vorbis CODEC Project. Software, The Xiph.org Foundation, 2002. http://www.xiph.org/.

[13] MPG123. Mpeg audio decoder software, 2001. http://www.mpg123.de/.