Audio Compression Put To The Test

Luis A. Rovira

Scientific-Atlanta, Inc.

ABSTRACT

Transmission of digital audio almost always requires some method of compression to reduce the required bandwidth and to supply larger numbers of services. Many audio compression systems exist, based on very different principles. The tests normally associated with measuring audio performance seldom challenge the best audio compression techniques. In this paper, one audio compression system, SuperSound, is studied. Tests are performed that are intended to evaluate its performance when dealing with complex signals. These tests provide a more realistic measure of actual audio performance than simple test tones. The "compression noise" that is measured quantifies the degree to which the original signal is altered by the compression/decompression process.

INTRODUCTION

Simple test tones consisting of one or two discrete frequencies have been used for many years to characterize audio systems. Tests such as harmonic distortion, intermodulation distortion and signal to noise ratio were sufficient to characterize linear audio systems. Linear systems use no processing other than gain and frequency equalization, and the testing was thus uncomplicated.

The inadequacy of such tests first became widely apparent with the common use of noise reduction techniques. Signal to noise ratio (SNR) measurement on a noisy system using noise reduction provides a good example. In a SNR measurement, a reference tone is first applied to the system. Often, when the test signal is present, a large amount of noise can be seen to accompany the tone. Yet when the reference tone is removed, the system reduces the gain, and with it the noise we are trying to measure. This familiar "noise pumping" that such systems cause defies simple SNR tests. Though often audible and sometimes objectionable, in most cases the excess noise is adequately masked by the presence of audio. It is the function of the noise reduction system to reduce the noise when the signal is absent, and thus improve the apparent SNR. It is usually preferable to the constant hiss that is the alternative.

But today the standards have been raised. Compact disks (CD) and uncompressed digital audio tape (DAT) have accustomed the audio consumer to noise floors and reproduction quality that are limited primarily by the studio recording equipment. Now, into this setting comes a host of new products. These products promise inexpensive recording of digital audio on cheap tapes and recordable disks. They enable broadcast transmission over terrestrial airwaves, from satellites and over CATV cable. They all promise "compact disk quality audio" and they all use some form of audio compression.

Audio compression is usually necessary to reduce the transmission bandwidth or storage requirements of the signal. We all know that "you don't get something for nothing". We feel that there must be some performance cost to audio compression. But in fact, it will be shown below that there is a large amount of wasted dynamic range at higher audio frequencies. To the extent that compression is achieved solely by taking advantage of such waste or redundancy in the signal, it is possible to perfectly reproduce the original signal and "pay no price". Computer data compression systems that achieve this goal are referred to a "lossless" compression systems. Unfortunately, this alone seldom results in sufficient data reduction for audio systems.

The compression system under study in this paper, SuperSound [1], is a system that does not attempt large amounts of data compression. It attempts to gain most of its compression by exploiting the waste and redundancy in the audio signal. As will be seen below, it comes relatively close to achieving lossless compression when processing real music signals.

Compression systems differ widely. They differ both in the amount of compression achieved, as well as in the price paid in audio quality. Compression systems can be designed that perform very well when subjected to simple test tones, yet can generate high "noise pumping", distortions and other less familiar artifacts when subjected to full loading. A method is needed to quantify signal degradations that occur in the presence of the signal. Ideally, a test could be devised that would yield a single number or "figure of merit". It could be used to characterize the extent to which a compression system achieves its compression at the cost of imperfectly reproducing the original signal at its output. Test signals should be used that are both representative of actual music, and that also fully load the system.

COMPRESSION NOISE

Fortunately, digital audio techniques have not only raised the standards of expected sound quality. They have also provided more sophisticated test methods than the simple test tones. Using digital techniques, it is now possible to take a digitized signal, compress it, then decompress it, and compare each sample of the processed signal with the corresponding sample of the original signal. Differences between the processed signal and its original represent an error due to the compression system. This error is the "noise" or degradation that the compression system generates. It is the cost of compression.

Compression noise is similar to the quantization noise that occurs in analog to digital conversion. As with quantization noise, compression noise can also be treated and measured in the same ways that we now study analog noise. Like other audio noise, the human ear is more sensitive to compression noise at some frequencies than others. Thus it is appropriate to use weighting when integrating the noise over the audio spectrum in noise measurement. The same noise weighting curves, such as "A" weighting [2], or CCIR/ARM [3] are appropriate.

In measuring compression noise using digital techniques, the test signals can be simple tones, as in analog measurements. More importantly, the test signal can be any audio waveform, from actual music to broadband noise. The system degradations can be very accurately measured while the system is fully loaded, without removing the test signal.

TEST SIGNALS

Though it is interesting to test a system loaded with actual music using the method described above, most music is quite variable in time. In actual measurements of the compression noise, this can result in readings that fluctuate with the instantaneous power in the music waveform. Much of the time, such a test would not present a sufficient challenge to the system. Also, the single-number "figure of merit" we seek would be a function of time and type of music. It would be highly subject to interpretation by the tester. These problems were overcome in two ways. The first was the use of a broadband noise signal with a spectrum that is closely equivalent to that of music. The second was the use of a "composite music signal".

In arriving at a useful test signal, first the typical spectrum of music was studied. The line output of a compact disk player was connected through an audio attenuator to the input of an HP 3588A spectrum analyzer. The top reference line was set to the level of a full-scale sinusoid. The full spectrum was swept every 614 milliseconds. A resolution bandwidth of 150 Hz was used. Peak hold was used to hold the highest level encountered at any given frequency. Entire disks [4][5][6][7] were played into the analyzer and the spectra were plotted as shown in Figures 1 through 4. Though the type of music varied widely, the spectra are quite similar. Note that in each case the spectrum rolls off with increased frequency. As mentioned earlier, much of the dynamic range available from 16 bit PCM is wasted at higher audio frequencies. The music simply does not demand it. This is good news for speaker manufacturers in that the small voice coils of tweeters never need to handle the same power levels encountered by bass drivers. Certainly it is possible with electronic music synthesizers to generate high levels at high frequencies, but in reality this is quite unpleasant. It occurs about as often as blown tweeters!

Figure 5 shows the spectrum of USASI (United States of America Standards Institute) noise. This noise spectrum consists of white noise filtered to peak at approximately 200 Hz, with a 6 dB per octave falloff below 100 Hz and above 320 Hz. it is available on the Audio Precision (AP) System One audio test equipment. It is also available on the National Association of Broadcasters (NAB) Broadcast and Audio System Test CD. It is used by audio broadcast manufacturers to simulate unprocessed audio program material. Note the similarities between the spectra of actual

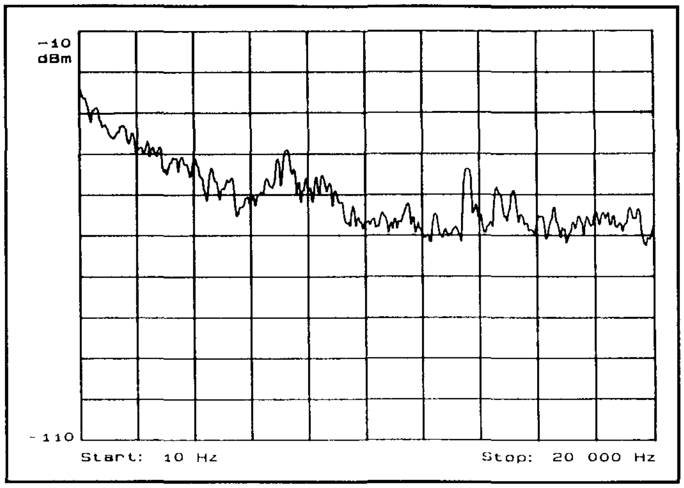

Figure 1 Spectrum of country music disk (full disk represented)

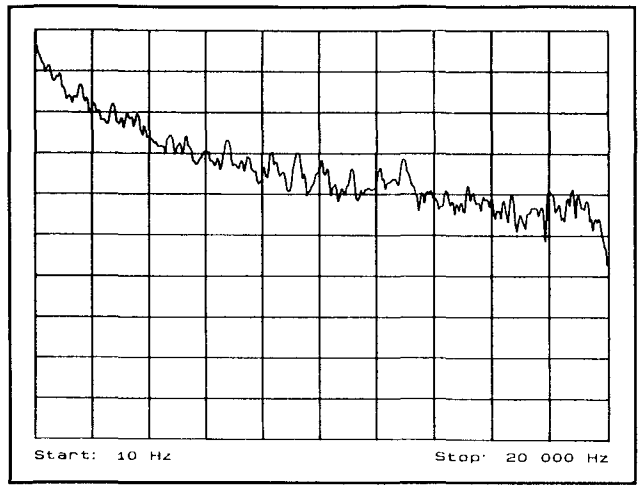

Figure 2 Spectrum of classical music disk (full disk)

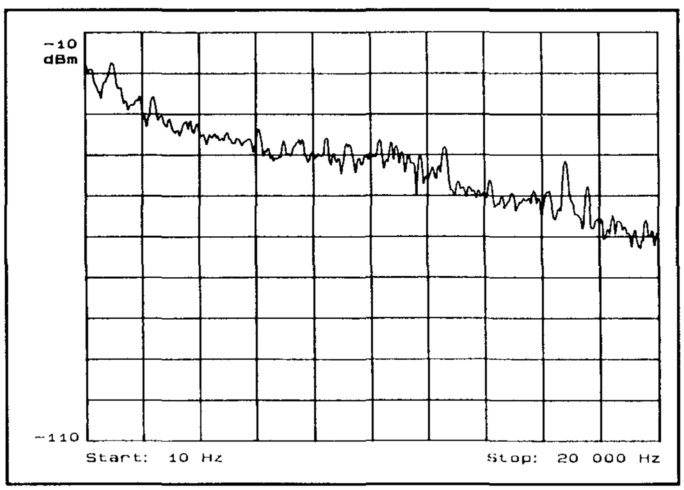

Figure 3 Spectrum of rock/Latin music disk (full disk)

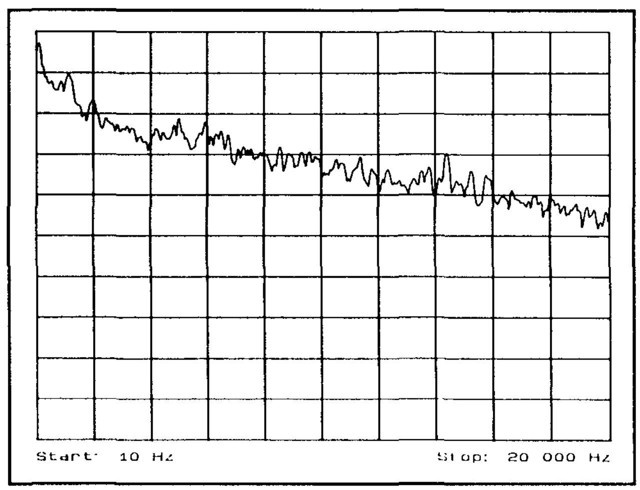

Figure 4 Spectrum of electric jazz music disk (full disk)

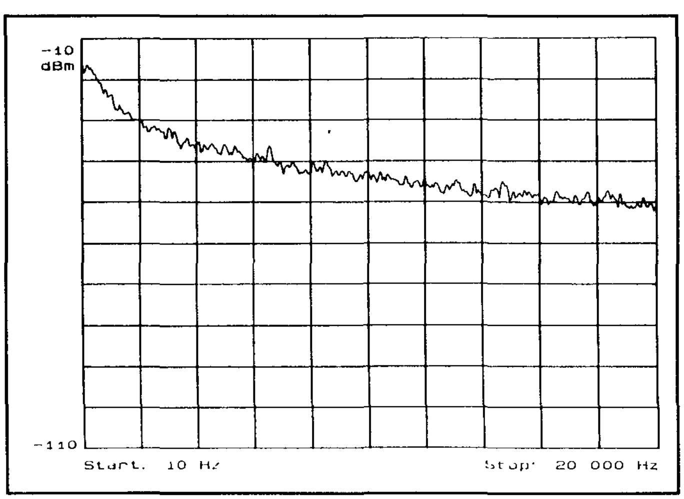

Figure 5 Spectrum of USASI noise (10 seconds)

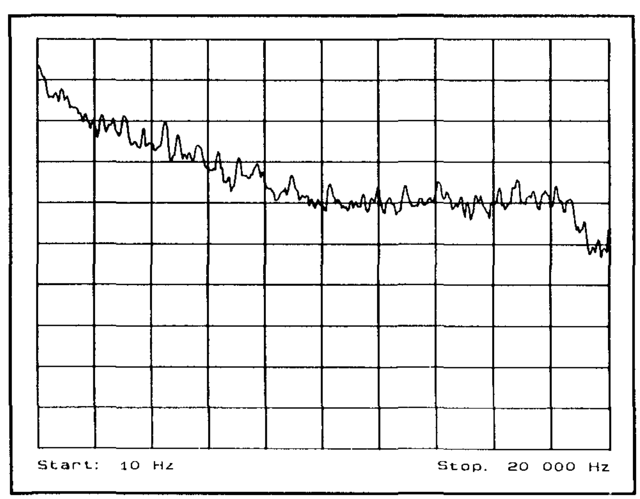

Figure 6 Spectrum of composite music mix (10 seconds)

music in Figures 1-4, and the USASI noise of Figure 5. The USASI noise allows loading the system with a realistic test signal, comparable to the peak music readings through the music spectrum, while providing a fairly constant power yielding noise measurements that fluctuate little. A time domain plot of USASI noise is shown in Figure 7.

Lest there be any concerns about the suitability of the noise waveform for audio testing, a second test waveform was generated. This time a "composite music signal" was created by mixing selected ten second sections [8][9][10][11] of country music, classical music, rock/Latin music, and electric jazz music from the sources used in Figures 1-4. The spectrum of this mix is shown in Figure 6. The ten second sections were chosen to get the largest number of instruments playing at the same time. The mix contained a full symphony orchestra, several synthesizers, two kinds of electric drums, "fuzzed" electric guitar, Latin percussion and brass, one female voice and two male voices all simultaneously. Well over 100 acoustic and electric instruments were playing at once. All were digitally summed without any scaling back of the levels. None of the highest peaks occurred at the same instants, and thus no clipping occurred. Peaks reached 92 % of full scale. The average level, however, did increase markedly and was sufficiently constant to yield steady readings. Though somewhat unrealistic, this signal provided a good "worst case torture" test. A time domain plot of this mix is shown in Figure 8.

TESTS

USASI noise from an AP System One was recorded onto an Apple MacIntosh IIcx computer hard disk using Sound Designer II and Sound Tools software and hardware. This sound file was then SuperSound compressed and decompressed. Figure 9 shows a portion of the time domain USASI waveform before SuperSound processing. Figure 10 shows the same after Supersound compression and decompression. The horizontal axis is calibrated by sample number. The vertical axis is calibrated with decimal quantization value, where 16 bits corresponds to a total of 65,360 values, or +/- 32,768 values. Display resolution is much poorer than any actual differences between the waveforms and thus it is not possible to compare the waveforms by inspection. Each sample of the original waveform was sign-inverted and added to each corresponding sample of the processed waveform. The resulting difference is the compression noise. It was stored in a sound file and is shown in Figure 11. Note that the noise is not visible on the same vertical scales as used in Figures 9 and 10. Thus the vertical scale in Figure 11 was expanded by a factor of 65.

All of the above processing was performed in the digital domain. The sound data was manipulated on the computer, and thus the difference calculations were mathematically exact. The samples that were compared represent the same sampling instants in time with zero time delay between them and zero amplitude calibration errors. This is only possible with fully digital processing, and on compression systems that accept digital input and produce digital output.

To quantify the compression noise, the file was passed digitally via the standard Sony Philips Digital Interface Format (SPDIF) [12] from the MacIntosh computer to the AP System One test equipment. Using fully digital signal processing, the noise was integrated over the entire audio spectrum and A-weighted. This resulted in a measurement of -87 dBFS (dB with respect to full scale).

The same procedure was repeated with the "worst case" composite music mix. Figure 12 shows a portion of the original time domain waveform. Figure 13 shows the SuperSound compressed and decompressed waveform. Figure 14 shows the noise using the expanded vertical scale. Again, the noise was digitally integrated and A-weighted and resulted in -80 dBFS, very respectable for such torture.
This level of noise is comparable to what the best studios can produce, and is certainly inaudible when such a high level signal is present.

CONCLUSION

It is possible to design compression systems that reduce the data rate or storage requirements to very low levels. Normally, larger amounts of compression results in more compression noise. But a measurement of compression noise does not indicate how audible the degradation to the audio will be. Many systems exploit the properties of the human auditory system to very effectively mask or conceal the compression noise. Though the concealment is not perfect, it does raise an interesting question. If the concealment were so good that the noise was completely inaudible, what value would a measurement of compression noise have?

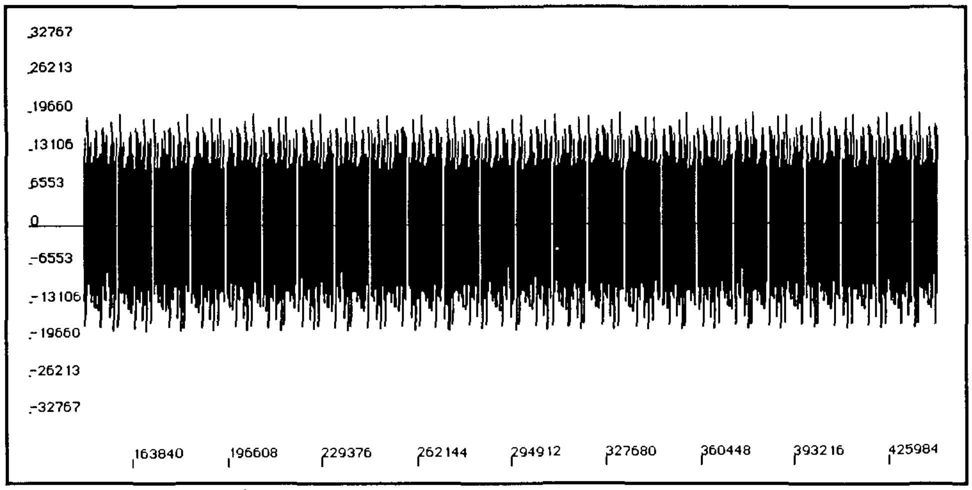

Figure 7 Time domain plot of USASI Noise
Approximately 7 seconds shown

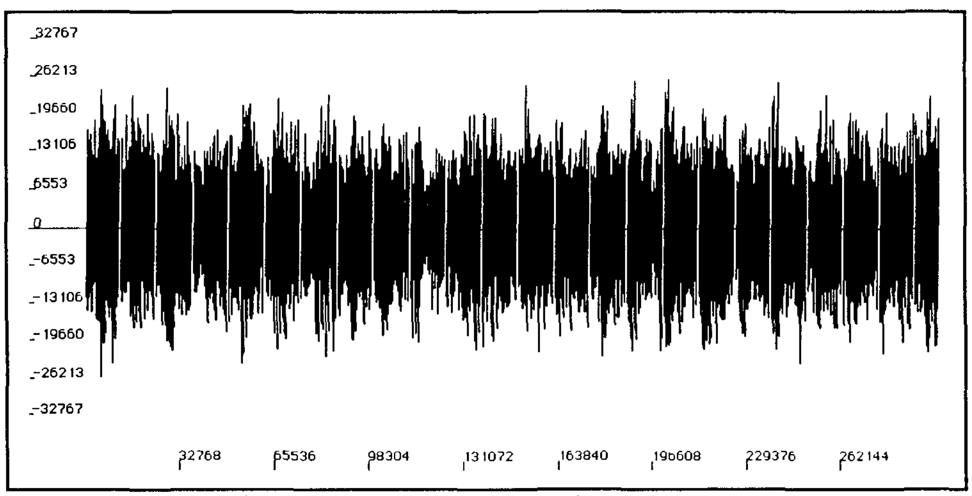

Figure 8 Time domain plot of composite music mix
Approximately 7 seconds shown

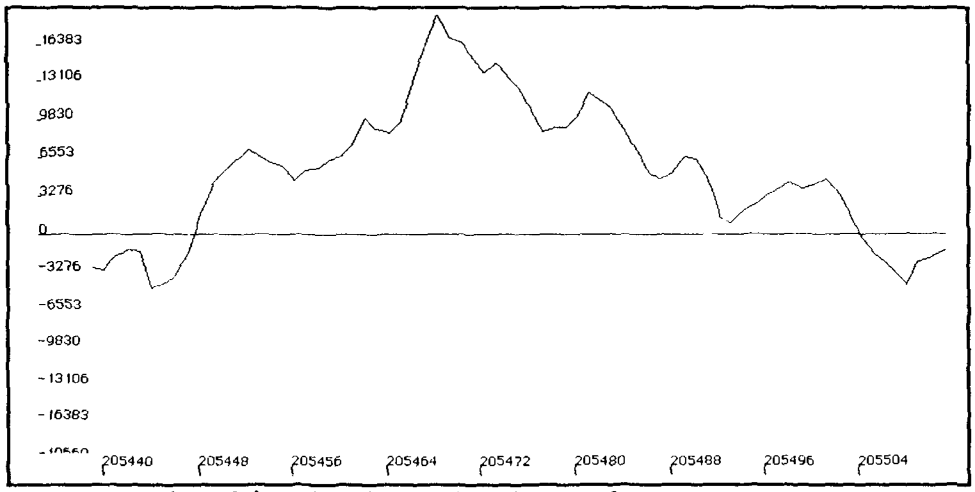

Figure 9 Section of time domain USASI noise waveform before SuperSound processing

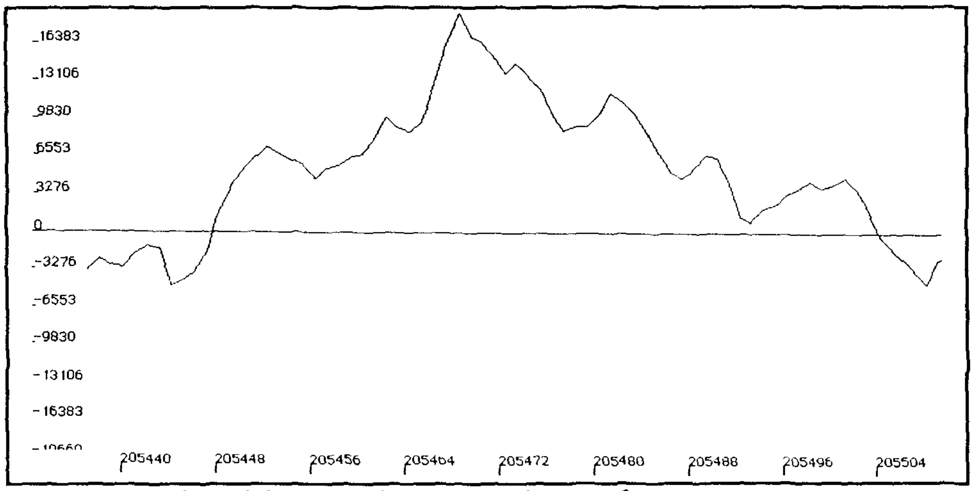

Figure 10 Section of time domain USASI noise waveform after SuperSound processing

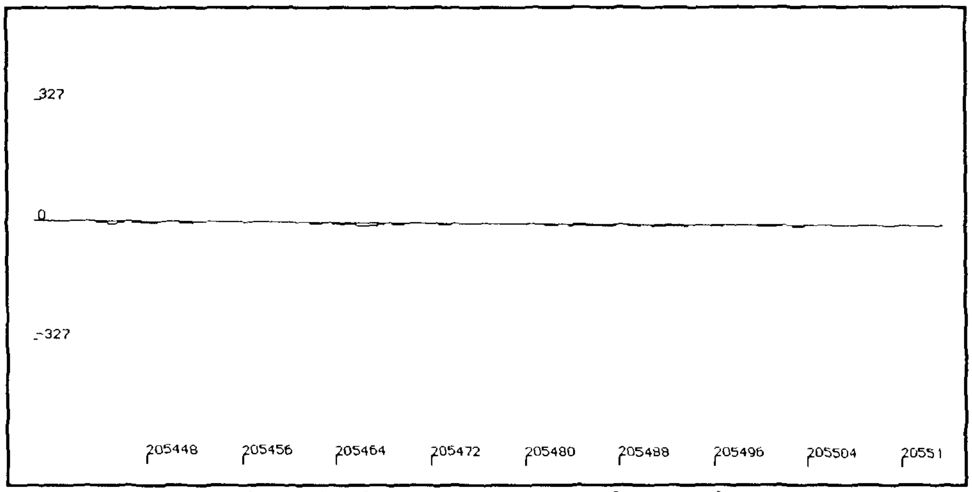

Figure 11 Compression noise for same USASI waveform section Vertical scale expanded 65x

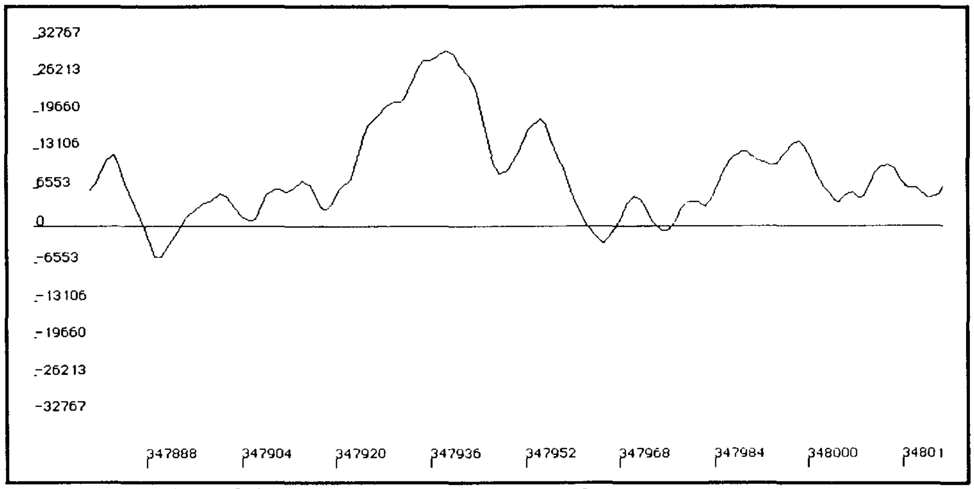

Figure 12 Section of time domain music mix waveform
before SuperSound processing

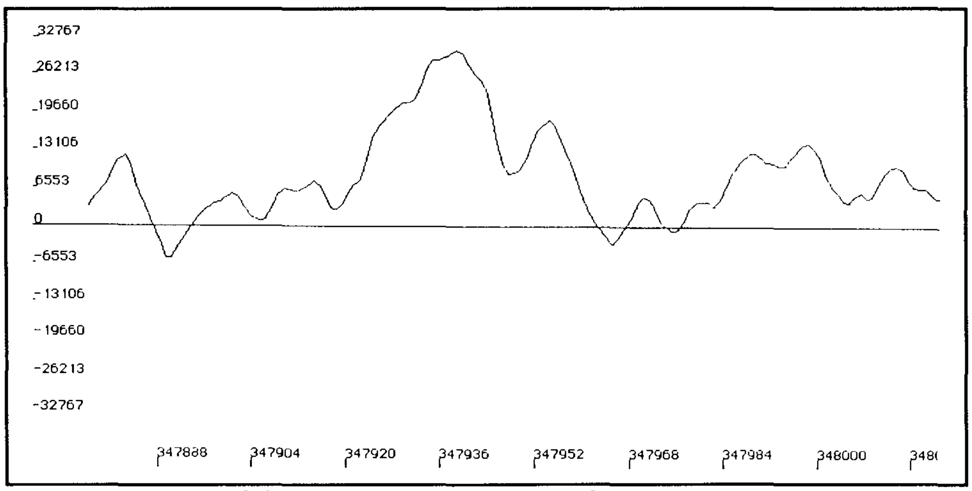

Figure 13 Section of time domain music mix waveform
after SuperSound processing

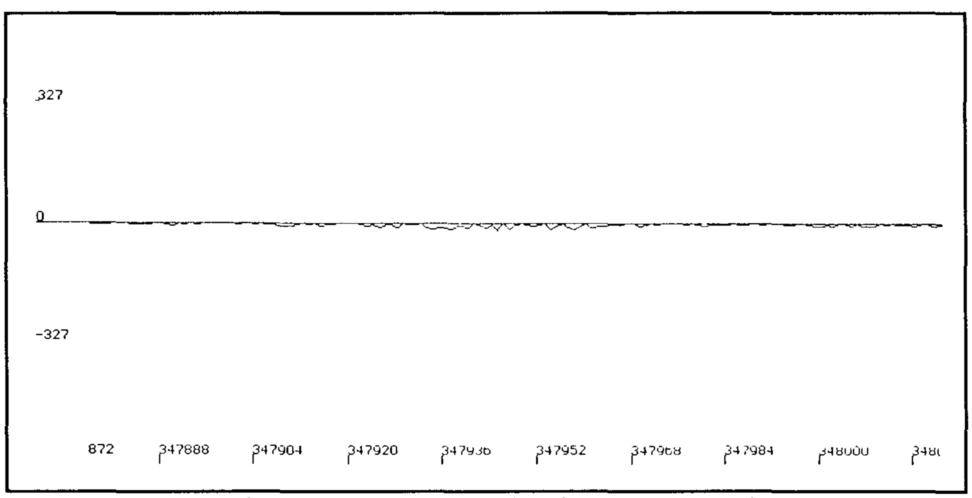

Figure 14 Compression noise for same music mix waveform section
Vertical scale expanded 65x

The problem is that no one yet knows the effects of cascaded audio compression. As mentioned earlier, many products are being conceived and introduced that use audio compression. Each of these generates compression noise. The noise and artifacts generated by a transmission system may be undetectable to the human ear, but may be very real to the compression system in a recorder. In the recorder, the noise and artifacts will interact with the second compression system in unpredictable ways. The compounded artifacts generated may be much more objectionable than either of the two systems alone.

Short of not compressing the audio, the safest approach would be to use as little compression as possible, and thereby generate the minimum amount of noise.

We normally consider two pieces of digital audio equipment compatible if they use 16 bit PCM, share sampling rates, and interface using either SPDIF (consumer equipment) or AES/EBU (professional) [13] standards. But compatibility may mean more than that. Compatibility may be determined by the specific compression system used, and the degree to which compression is achieved by altering the individual samples as opposed to exploiting redundancy and waste in the data. Compression noise provides a measure of the degree to which the original signal is altered. As such, it should provide a good tool for evaluating the compatibility between audio systems using compression.

ACKNOWLEDGEMENT

I gratefully acknowledge Jay McMullan who wrote much of the software used in this work and whose help was indispensable.

REFERENCES

[1] Quinn, S.B. Jr., Frederiksen, J.
"A Compressed Digital Audio Transmission System," Paper presented at AES 7th International Conference, May 14-17, 1989.

[2] "Sound System Equipment," IEC Standard 268-1 (1968).

[3] "CCIR/ARM: A Practical Noise-Measurement Method," Journal of the Audio Engineering Society, Vol.27, No.3, March 1979.

[4] "Pancho & Lefty," Merle Haggard and Willie Nelson, Epic EK 37958

[5] Telarc Sampler Vol 1, used 17 classical selections by various composers and orchestras, CD-8001

[6] "Primitive Love," Miami Sound Machine, Epic EK 40131

[7] "The Chick Corea Electric Band," Chick Corea et. al. GRP D-9535

[8] ref. [2] Cut: "All the Soft Places to Fall," time: 1:40-1:50

[9] ref. [3] Cut: R. Wagner, "Riezi Overture," time: 0:05-0:15

[10] ref. [4] Cut: "Conga," time: 3:39-3:49

[11] ref. [5] Cut: "Elektric City," time: 2:43-2:53

[12] "Digital Audio Interface," EIAJ Standard CP-340 (1987).

[13] Audio Engineering Society Standard AES3-1985, also ANSI S4.40-1985.