Audio Quality Measurement Primer

Introduction

This audio quality measurement Application Note is geared towards describing the methods for measurement of key audio specifications. Each analysis technique described here in, can be applied to audio ADCs, DACs, Analog Mixers, or any signal within the audio band. Audio quality is typically gauged by three main components. They are Signal to Noise Ratio, Total Harmonic Distortion, and Frequency Response. Audio quality, in all cases, is subjective and has been debated for decades. The inclusion of a weighting function, to account for the ear's non-ideal perception of loudness at different frequencies, is one such example of the subjective quality of audio. In the early 1950s, transistors had just been invented. Audio amplifiers built with these devices offered better signal to noise, harmonic distortion, and frequency response than their vacuum tube predecessors. Circuits designed and built with vacuum tubes were still preferred by audio enthusiasts. The transistor amplifiers had lower power output, and when over-driven produced odd order harmonics. To the human ear, odd order harmonics have long been identified as sounding more unpleasant than even order harmonics. The vacuum tube amplifiers could be over-driven to produce more pleasing even-order harmonics while producing more power than the transistor amplifiers. Though the transistor amplifiers had better specifications, the tube amplifiers were superior in listening tests, and proved to have better subjective audio quality. Today's audio circuits far outstrip their predecessors in most audio applications. Their specifications are very often beyond the range of human perception, yet the purists still debate.

Noise and Distortion Measurement Methods

There are several competing measures of noise and distortion for audio systems. The most common are Signal to Noise Ratio, Signal to Noise plus Distortion Ratio, Dynamic Range, Idle Channel Noise, and Harmonic Distortion. Most of the measurement techniques presented here are based on the AES17-1991 and EIAJ CP-307 CD measurement standards. Traditionally, these measurements are preformed using a 1kHz signal tone. A-weighting (compensation for the human ear) is applied to the result from 20Hz to 20kHz.

Each of these measurement methods have benefits and weaknesses. You may ask why so many measures of performance are required. I will touch on these details as needed while describing each measurement technique and its properties. Appendix A contains a brief overview of the time domain and frequency domain analysis techniques used in these measurements. To achieve accurate results the stimulus signal must always have performance characteristics that exceed those of the device under test.

Signal to Noise Ratio

The signal to noise ratio, or SNR, of a signal is a measurement, in dB, of the power in the signal fundamental relative to the RMS sum of the energy of all in-band noise components excluding harmonics (EQ. 1). Testing an audio DAC, the Digital signal is set to digital Full Scale. Testing an audio ADC, the analog stimulus signal is usually set to within -0.5dB to -1.0dB of the device's full scale(FS) range. If this were not the case, gain and offset errors across devices would cause the signal to inadvertently be clipped. Clipping shears off the high and low peaks of the waveform producing a great deal of unrelated noise and harmonics. Because SNR measurement excludes harmonics, it is possible to get a false sense of the device's overall performance. Therefore, this measurement should always be considered along with Total Harmonic Distortion when evaluating total performance. Examples of some common audio applications, and their SNR, are listed in Table 1.

$$SNR = 20\log\left(\frac{\text{Full scale RMS signal energy}}{\text{RMS non-harmonic noise energy}}\right) \quad \text{(EQ. 1)}$$

TABLE 1. SNR EXAMPLES

AUDIO APPLICATION	SNR (DB)
Plain Old Telephone System or POTS	40
AM Radio, LP Records	50
FM Radio, Cassettes	70
CD Player	90
Human Ear Range (Instantaneous)	85
HiFi Studio Recording Equipment	120
Human Ear range (Total)	120

Total Harmonic Distortion

Total Harmonic Distortion, or THD, is a measure, in dB, of the relative level of the energy in the fundamental to the RMS sum of the energy of all in-band harmonics (EQ. 2). The signal is a single FS analog tone for ADCs or FS digital tone for DACs. Measurement of the THD in conjunction with the SNR of a device is a good indication of the in-band performance of the audio device. When reporting THD performance, the highest order harmonic used in the computation must be specified.

$$THD = 20\log\left(\frac{\text{Full scale RMS signal energy}}{\text{RMS harmonic energy}}\right) \quad \text{(EQ. 2)}$$

Signal to Noise + Distortion Ratio

Signal to Noise Plus Distortion, or SND, is a measure, in dB, of the relative level of the energy in the fundamental to the RMS sum of the energy of all in-band spectral components including harmonics and noise over the specified bandwidth. A FS analog or FS digital signal is used. SND is effectively the RMS sum of THD and SNR.

Dynamic Range

Dynamic range, or DR, is a ratio of the greatest magnitude(undistorted) signal to the quietest(discernable) signal in a system as expressed in dB. For audio applications, DR is a ratio of the full scale RMS signal level to the RMS sum of all other spectral components over the specified bandwidth using a -60dB FS amplitude signal. DR differs from SNR in that a -60db FS signal is used. To refer the measurement to full scale, 60db is added to the resultant measurement. In this way, the device is being excited, but the harmonic components, which are generated with a full scale input, are now typically well below the noise floor. DR is often used to estimate SNR. However, if a device operates in such a way that the level of the noise floor increases with increased input level, then a measure of the Dynamic Range is no longer a good estimation of the actual SNR.

Idle Channel Noise

Idle channel noise, in dB, is a ratio of the full scale RMS signal level to the RMS sum of all other spectral components over the specified bandwidth using a zero amplitude signal. ICN is a measure of the static noise produced by the device with a zero amplitude signal. It is useful to have good ICN performance in order to avoid an irritating noise commonly referred to as hiss, random noise, or white noise. Leakage from local sources, such as power lines, will show up and degrade ICN performance as well.

Interchannel Isolation or Crosstalk

Crosstalk is a measurement, in dB, of the signal amplitude in an unused channel relative to that of a channel driven with a signal. The unused channels should be grounded, or set to an appropriate bias point. The signal going to the driven channel can be either of FS or -20db FS amplitude. Ideally, the two measurements should produce identical results. However, to avoid corruption of the signal due to non-linearities, excessive power consumption, or heat buildup from excessive signal levels, it is best to measure crosstalk with a -20db FS signal. Crosstalk is expressed in dB.

Out-of-Band Spurious Components

Out-of-Band spurious components are measured at the analog output of the device under test with no input test signal stimulus. An output high pass filter may be used to remove all in-band signals from the device output. The level of all components above the upper band-edge frequency of the device output spectrum are expressed as a dB ratio of the full scale signal level. The upper and lower band edges for the out of band range must be specified. For audible components, this range is 20kHz to 28.8kHz, while the inaudible component range is 28.8kHz to 100kHz.

Suppression of Imaging Components

This measurement is identical to the out of band spurious components measurement with the addition of a test signal. Typically, the test signal is -20dB FS. The RMS sum of all components above the upper band-edge frequency of the device under test output spectrum are measured relative to FS output level and expressed in dB FS. You can use filters to remove and isolate unwanted signals. A notch filter, at the output can be used to remove the test signal and a high pass filter may be used to isolate the out of band components.

Frequency Response

Frequency response is the variation of the system's output level relative to the input level across a frequency range (EQ. 3). The typical human hearing range is 20Hz to 20kHz. Therefore, audio systems with a flat response over this range seem to sound the most natural. As a performance benchmark, an in-band frequency flatness tolerance, in dB, is usually specified. For example, most conventional stereo audio systems specify a frequency flatness tolerance of ±3dB. In addition, the -3db points of the system at high and low frequency extremes are also specified. The -3dB point corresponds to the frequency at which the signal power has been reduced to one half of the normal 0dB signal level. A typical system frequency response plot is pictured in figure 1 with the -3dB points at 20Hz and 20kHz. It has been shown that human perception is capable of detecting signal amplitude differences as low as 0.1dB. However, under dynamic signal conditions, 1dB is more realistic. Table 2 lists some common frequency response examples. The frequency response of a system is usually measured with an input signal that is -20dB below full scale.

$$\text{Frequency Response(f)} = 20\log\left(\frac{Vo(f)}{Vi(f)}\right) \qquad \text{(EQ. 3)}$$

TABLE 2. FREQUENCY RESPONSE EXAMPLES

AUDIO APPLICATION	FREQUENCY RESPONSE
Plain Old Telephone System or POTS	300Hz to 3kHz
AM Radio	100Hz to 5kHz
FM Radio	50Hz to 15kHz
Consumer Stereo System	20Hz to 20kHz
Professional Audio Equipment	5Hz to 24kHz

Weighting Functions

Our ears are capable of hearing sounds from 20Hz to 20kHz, but we don't perceive those sounds with equal loudness. For this reason, most audio band measurements use a frequency weighting function to compensate for the ear's non-uniform amplitude detection.
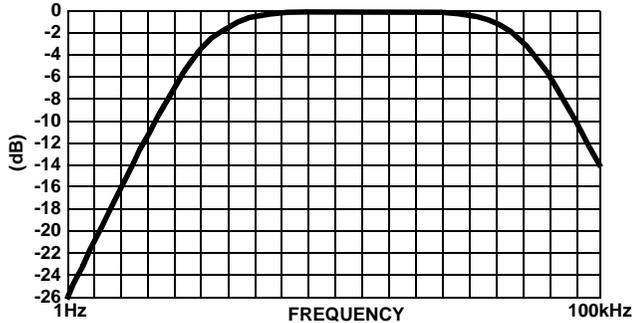

FIGURE 1. TYPICAL SYSTEM FREQUENCY RESPONSE

Weighting functions are needed to adjust the qualitative measurement of the real world to the human ear. Without this adjustment, the perceived system performance would not correlate well with the measured performance. For instance, our threshold of hearing for normal speech (200Hz - 3kHz) is much lower than the threshold of hearing at low frequencies (DC to 100Hz). Two systems, one with a great deal of noise near DC, and another with very little, would virtually sound the same. However, an SNR measurement taken with flat frequency weighting, would suggest poor performance for the first system. Three common weighting functions are used to compensate measurements for the discontinuity between perceived performance and measured performance. They are referred to as A, B, and C weighting. The respective frequency responses of these weighting functions are shown in Figure 2.

What does A-weighting mean? The first tentative standard for sound level meters(Z24.3) was sponsored by The Acoustics Society of America and published by the American Standards Association in 1936. The tentative standard shows two frequency weighting curves "A" and "B" which were modeled on the ear's response to low and high levels of sound respectively.

The most common weighting today is "A-weighting" dB(A), which is very similar to that originally defined Curve "A" in the 1936 standard. "A-weighting" is used for measuring low level signals and has become a standard for electrical and acoustic noise measurements of all signal levels. "B-weighting" is used for intermediate level sounds. "C-weighting", which is occasionally used for very loud sounds in acoustic applications, has a relatively flat response in the lower end of the spectrum. The mathematical functions that describe each filter are in Appendix B.

When an audio measurement is reported, the weighting filter must be indicated. For example, a Digital to Analog converter SNR measurement of 91dB relative to full scale using an A-weighting function would be listed as:
DAC SNR = 91dB(A) FS.
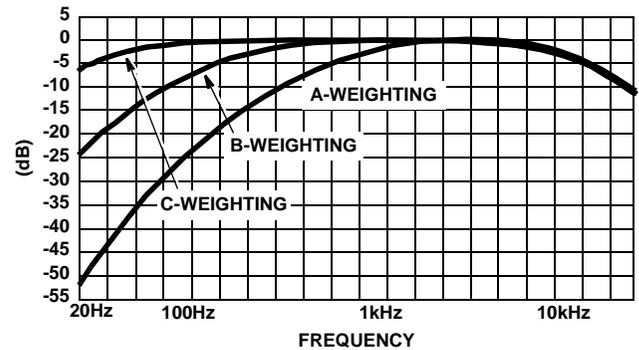
FIGURE 2. A, B, AND C WEIGHTING SYSTEMS

Decibels

Why Decibels? The decibel is a logarithmic unit which is used in a number of scientific discipline to compare some quantity with some reference value. Usually the reference value is the largest or smallest likely value of the quantity. We use decibels for several reasons. Quantities of interest often exhibit such huge ranges of variation that a dB scale is more convenient than a linear scale. For example, the listening level of your home stereo can vary by seven orders of magnitude. In addition, the human ear interprets loudness on a scale much closer to a logarithmic scale than a linear scale. The 'bel', named after Alexander Graham Bell, is the logarithmic ratio of two powers, P1 and P2, and is expressed in Equation 4.

$$A(B) = \log\left(\frac{P_1}{P_2}\right) \qquad \text{(EQ. 4)}$$

The reference power is usually based on some very small natural occurrence. An example of which is sound pressure level. The reference sound pressure level in air is 0.0002 Pa(RMS). A typical loudspeaker, with 1 watt of power at 1kHz, produces 9B of sound pressure measured at a standard distance of 1 meter. The same loudspeaker is most likely limited to about 11B of pressure. Average listening levels are usually around 6B. This makes the measurement range only 5B. Due to the small size of the bel, the more convenient decibel has become standard and is expressed in Equation 5.I

$$A(dB) = 10\log\left(\frac{P_1}{P_2}\right) \qquad \text{(EQ. 5)}$$

The decibel is often used to express the relative amplitude or power of a signal. Since power is the square of amplitude, a simple translation from Amplitude to Power can be made as in Equation 6. Some common variations on the decibel are listed below.

$$10\log\left(\frac{A^2_1}{A^2_2}\right) = 20\log\left(\frac{A_1}{A_2}\right) \qquad \text{(EQ. 6)}$$

"dB FS" stands for "dB relative to full scale" and is generally used in all audio measurements unless otherwise specified.

"dBr" stands for "dB relative to some absolute level." When the absolute level is set to full scale, "dBr" is equivalent to "dB FS."

"dBV" stands for "dB relative to 1V."

Measurement Paths

In an Audio System, several measurement paths are available. It is important to measure the performance of the system as well as individual system blocks. While an individual system block may perform well, errors in offset, gain, signal path layout, and available analog headroom can cause the system to perform poorly. Figure 3 illustrates a basic system block diagram incorporating all of the available blocks. The most common measurement paths are:
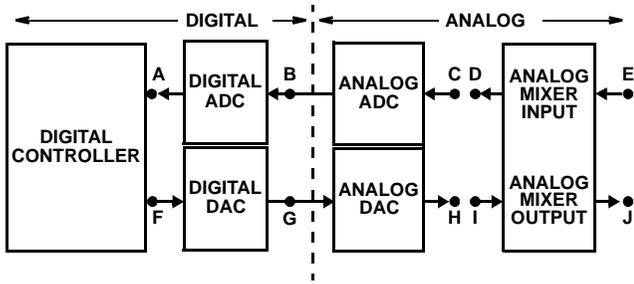

FIGURE 3. MEASUREMENT PATHS

MIX-MIX: For this measurement, Analog path D is connected to Analog path I. In this configuration, the Analog signal path can be accurately tested. Typically, this is a best-case measurement where the Analog Mixer Circuitry is superior in performance to either the ADC or DAC.

MIX-ADC: For this measurement, the path from E to A is completed. Digital Data is captured by the Digital Controller and analyzed later in software. The performance of the ADC is critical to Data Acquisition applications.

DAC-MIX: The path from F to J is completed to measure DAC performance. Digital Data must be supplied in the proper format. DAC performance is key to Multi-Media, consumer, and professional audio applications where high quality CD-Audio is regularly used.

MIX-ADC-DAC-MIX: The total system performance can be measured by completing the path from E to A and F to J. Digitally looping back the ADC data to the DAC, the Analog information sent to the Audio Codec at E can be re-captured at J, the Mixer Output, and analyzed. This is a worst case measurement exhibiting the faults of every system block in one measurement.

Depending on the system under test, there can be many more measurement paths. For example, modern Sigma-Delta ADCs and DACs require complex digital filters. It is often important to test the performance of these filters to isolate noise and distortion components that arise purely from digital filtering. However, direct measurement of the filters is often done only by the manufacturer using special digital test modes that are unavailable to the general public. Lastly, external analog components in the signal path further degrade or complicate results.

Glossary

Coherent Sampling - Sampling of a periodic waveform in which there is an integer number of cycles in the data record.

Data Window - A set of coefficients by which corresponding samples in the data record are multiplied to more accurately estimate certain properties of the signal in the frequency domain.

Date Set - A sequential collection of samples acquired by a waveform recorder.

Digital Full Scale Amplitude. Amplitude of sine wave whose positive peak value reaches the positive digital full scale, leaving the negative maximum code unused. In 2's complement representation, the negative peak will be 1 LSB away from the negative maximum code.

Full Scale Signals Amplitude (Fs) - Full scale signal amplitude is the specified maximum range of the device under test.

Fundamental - The fundamental is the primary frequency component of a pure sine-wave.

Gain Drift - The change in gain value with temperature. Units in ppm/^{o}C.

Gain Error - The deviation from nominal full scale output for a full scale input. Units are in dB.

Harmonic - Harmonics are frequency components that are integral multiples of the frequency of the fundamental.

Input For Full-scale Amplitude - In systems where the output is accessible in the digital domain, the input for full-scale amplitude shall be the voltage of a sine wave that shall be applied to the input to obtain a digital signal whose positive peak value reaches the positive digital full scale.

Interchannel Gain Mismatch - For ADCs, the difference in input voltage that generates the full scale code for each channel. For DACs, the difference in output voltage for each channel with a full scale digital input. For analog signal blocks, the different in output voltage for each channel with a full scale analog input signal. Units in dB.

Maximum Input Amplitude - The maximum input amplitude shall be the maximum voltage of a sine wave that may be applied to the device under test input before introducing 1% THD+N or 0.3dB compression at the device output, whichever comes first.

Noise - Any deviation between the output signal, converted to input units, and the input signal, except deviations caused by linear time invariant system response.

Nyquist Frequency - One-half the sampling frequency of the digital system. Signals applied to the input to such a system are subject to aliasing. In systems that sample at a frequency higher than that ultimately used, the sampling frequency to be considered is the lowest sampling frequency that occurs in the signal path.

Offset Error - For ADCs, the deviation in LSBs of the output from mid-scale with the selected input grounded or properly biased. For DACs, the deviation of the output from zero with mid-scale input code. Units are in volts.

Phase Noise - Noise in the output signal associated with a standard deviation of the sample instant in time.

Relatively Prime - Describes integers whose greatest common divisor is 1.

Spurious Components - Persistent sine waves at frequencies other than the harmonic and fundamental frequencies.

Appendix A: Analysis Techniques

Time Domain

Classical time domain analysis of a sampled waveform has advantages and disadvantages. For signal to noise plus distortion ratio testing, a best-fit algorithm must be applied in order to generate an approximation to the original signal in both phase and amplitude. The approximate signal is then subtracted from the signal produced by the device. A root mean squared analysis of the residual components yields an accurate signal to noise plus distortion result. The drawback is in execution time and versatility. Harmonic distortion analysis of the signal requires subsequent runs of the best-fit algorithm. Iterative calculations of this type require a significant amount of time. In addition, there is no way to band-limit the measurement except to use analog filters in the signal path. This is generally not acceptable because it complicates the device interface board design and potentially degrades in-band performance of the device. For these reasons, a very fast algorithm called the Fast Fourier Transform is used to analyze the signal in terms of its frequency, phase, and amplitude over a finite bandwidth.

Frequency Domain

Frequency domain analysis pertains to the use of such analysis algorithms as the Discrete Fourier Transform (DFT), Fast Fourier Transform (FFT), and many others. In general, the FFT is used because of its speed. However, application of the standard FFT is restricted to a factor of 2 number of samples. The harmonic estimates obtained through use of the FFT are N/2 frequency spectra uniformly spaced at $f_S(N/2)$ increments beginning at DC where f_S is the sampling frequency and N is the number of samples. This approach is elegant and attractive when the results are cast as a spectral decomposition as in Figure 4. It can be seen that the time domain waveform data set contains exactly 11 cycles. This uniqueness is called coherent sampling, and is very important for obtaining good results with the FFT.
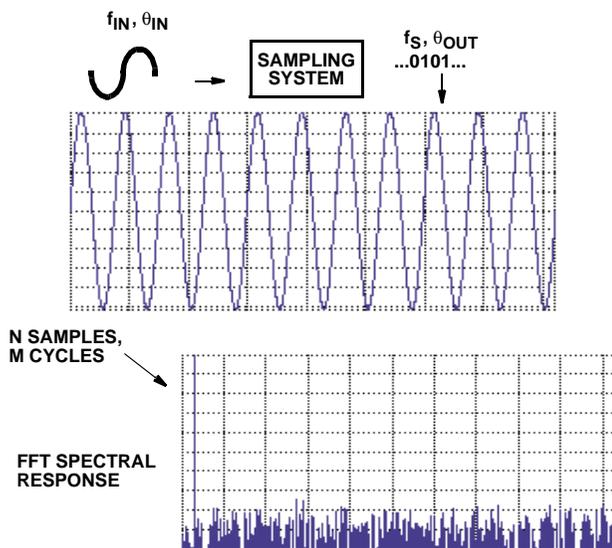

FIGURE 4. EXAMPLE OF COHERENTLY SAMPLED SYSTEM AND FFT SPECTRAL RESPONSE

Coherent Sampling

The primary application of coherent sampling is sine wave testing. Coherent Sampling of a periodic waveform occurs when an integer number of cycles exist in the data record. In other words, coherent sampling occurs when the relationship of Equation 7 exists. This greatly improves the FFT spectral purity and creates an ideal environment for evaluating the system performance. Special analysis techniques, such as windowing, are required if coherent sampling is not observed. Care must be taken to insure the spectral purity and stability of the input frequency and sampling frequency in the testing environment.

$$\frac{f_{IN}}{f_S} = \frac{M}{N} \qquad \text{(EQ. 7)}$$

Where:

f_S is the sampling frequency
f_{IN} is the input frequency
M   is the integer number of cycles in the data record
N   is the integer, factor of 2,
           number of samples in the record

The coherence relationship will work for any arbitrary M and N, but practical values provide better results. A good choice for N is a power of 2. The FFT requires the number of samples to be a power of 2 because of its inherent periodicity. If a DFT is to be used, an arbitrary positive N may be used. M should be odd or prime. Making M odd eliminates many common factors with N. When M and N are relatively prime, all common factors are eliminated. The uniqueness of M is absolutely imperative to Signal to Noise Ratio and Harmonic Distortion calculations. Common factors, between M and N, lead to different harmonics of fin having the same frequency bin in the FFT after aliasing, thus THD and SNR measurements are corrupted.

Non-Coherent Sampling

Non-Coherent sampling occurs when M in Equation 4 is not an integer. In this case, the FFT or DFT will produce a spectrum with leakage, or skirts, surrounding the fundamental. This effect is illustrated in Figure 5. Under these circumstances, a special technique called windowing is required to approximate the signal quality. This technique attempts to modify the time domain data set to more closely fit the requirements of the coherence relationship. Windowing is perhaps the easiest, most versatile, and most widely used of the techniques available for analyzing non-coherently sampled frequency related data sets.

Windowing

In many cases the signal or sampling variables can not be precisely controlled. This makes it difficult to obtain exactly an integer number of cycles. A coherently sampled waveform inherently generates data that begins and ends at the same level. Windowing can force the data to begin and end at the same or nearly the same level. The technique of mathematical windowing is accomplished by multiplying the sampled waveform data set by an appropriate windowing function as depicted in Figure 6. This prevents discontinuity at the window edge. Leakage occurs when the data set does not contain an integer

number of cycles of the frequency of interest. Eliminating the discontinuity does not always eliminate the leakage, but it does help to reduce it.

A window is typically chosen such that FFT results yield the best spectral purity, reducing leakage to a minimum, near the frequency or frequencies of interest. The actual source of leakage is not the signal itself but the window used in acquisition. The amount of leakage depends on the window shape and how the signal fits into the window.

A few common windowing functions are listed in Table 1. The window shape must be carefully chosen such that it performs its intended function. In general, an Exact Blackman, 3-term, or 4-term Blackman perform well for single-tone audio measurements. Many more windowing functions are available, but are not listed here.

TABLE 3. TYPICAL WINDOWING FUNCTIONS

WINDOW NAME	WINDOW FUNCTION
Rectangular	$A(n) = 1$ For n = 0 to N
Triangular	$A(n) = 2n/N$ For n = 0 to N/2 $A(n) = 2 - 2n/N$ For n = N/2 to N
Hanning	$A(n) = 0.5(1-\cos(2\pi n/N))$ For n = 0 to N
Hamming	$A(n) = 0.08 + 0.46(1-\cos(2\pi n/N))$ For n = 0 to N
Exact Blackman	$A(n) = 0.42 + 0.50\cos(\pi(2n-N)/N) + 0.08\cos(2\pi(2n-N)/N)$ For n = 0 to N
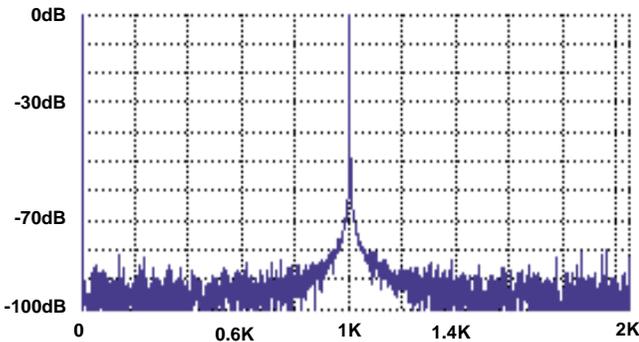
FIGURE 5. EXAMPLE OF SPECTRAL LEAKAGE
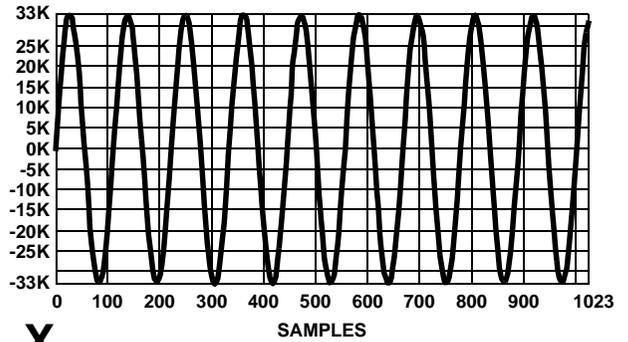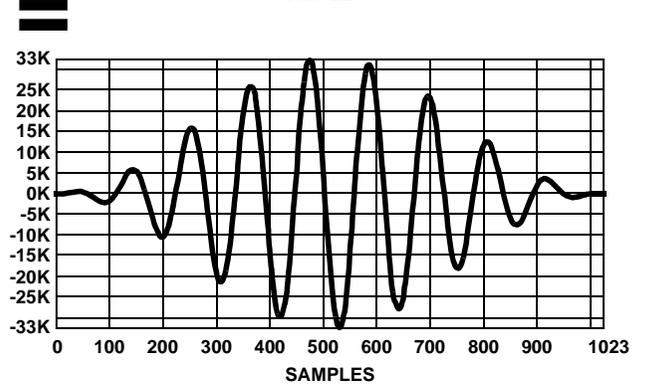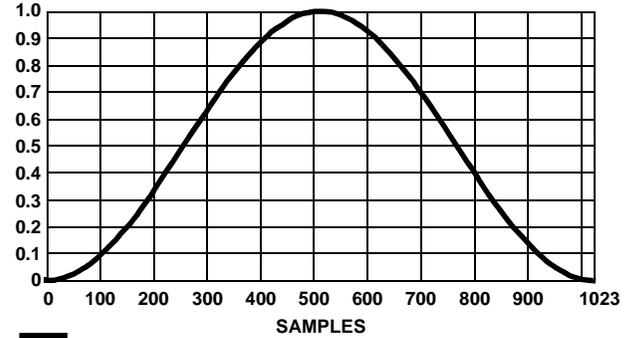
FIGURE 6. EXAMPLE OF A HANNING WINDOW

Appendix B: Weighting Functions

The s-domain transfer function for C-weighting is:

$$Hc(s) = \frac{4\pi^2 12200^2 s^2}{(s + 2\pi 20.6)^2 (s + 2\pi 12200)^2} \qquad \text{(EQ. 8)}$$

Adding an extra real-axis pole to the C-weighting transfer function gives us B-weighting:

$$Hb(s) = \frac{4\pi^2 12200^2 s^3}{(s + 2\pi 20.6)^2 (s + 2\pi 12200)^2 (s + 2\pi 158.5)} \qquad \text{(EQ. 9)}$$

Adding two real-axis poles to the C-weighting transfer function gives us A-weighting:

$$Ha(s) = \frac{4\pi^2 12200^2 s^4}{(s + 2\pi 20.6)^2 (s + 2\pi 12200)^2 (s + 2\pi 107.7)(s + 2\pi 738)} \qquad \text{(EQ. 10)}$$

where $\pi = 3.14159...$etc and s is the complex variable.

If you are only interested in the steady-state response then the weightings may be expressed in terms of frequency alone:

$$Rc(f) = \frac{12200^2 f^2}{(f^2 + 20.6^2)(f^2 + 12200^2)} \qquad \text{(EQ. 11)}$$

$$Rb(f) = \frac{12200^2 f^3}{(f^2 + 20.6^2)(f^2 + 12200^2)(f^2 + 158.5^2)^{0.5}} \qquad \text{(EQ. 12)}$$

$$Ra(f) = \frac{12200^2 f^4}{(f^2 + 20.6^2)(f^2 + 12200^2)(f^2 + 107.7^2)^{0.5}(f^2 + 738^2)^{0.5}} \qquad \text{(EQ. 13)}$$

These filters show a loss at 1kHz of 2.0dB, 0.17dB, 0.06dB for A, B and C weightings respectively, and since it is usual to normalize the response of each filter to 1kHz, this loss must be added to the modulus. In other words, the normalized responses may be expressed (in dB) as follows:

C = 0.06 + 20*log(Rc(f))

B = 0.17 + 20*log(Rb(f))

A = 2.0 + 20*log(Ra(f))

References

[1] AES17-1991, AES standard method for digital audio engineering measurement of digital audio equipment, Audio Engineering Society, New York, NY.

[2] American National Standard Specification for Sound Level Meters, ANSI S1.4-1985.

[3] Frederick J. Intersil, "On the Use of Windows for Harmonic Analysis with the Discreet Fourier Transform," Proceedings of the IEEE, Vol. 66 No. 1, 1978.

[4] Intel Corporation, "Audio Codec '97 Revision 1.03", Sep. 1996, www.intel.com

[5] PC98 Hardware Design Guide, Chapter 17 Audio Components, Microsoft Corp, www.microsoft.com/hwdev/pc97.htm

[6] EIAJ CP-307, Methods of Measurement for CD Players, Electronics Industries Association of Japan, June 1985.

[7] PC Audio Measurement Techniques, www.crystal.com

intersil