Theoretical and Audible Effects of Jitter on Digital Audio Quality

4826 (P-1)

Eric Benjamin and Benjamin Gannon
Dolby Laboratories Inc.
San Francisco, CA 94103-4813, USA

Presented at the 105th Convention 1998 September 26–29 San Francisco, California

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Theoretical and Audible Effects of Jitter on Digital Audio Quality

Eric Benjamin and Benjamin Gannon
Dolby Laboratories
San Francisco, California

Abstract: Many digital audio systems now use some form of self-clocked digital interface for audio delivery. With the advent of new digital audio systems that use IEC 61937 to convey non-linearly coded audio, the total number of devices using the IEC 60958 interface has substantially increased. The digital interface may contain jitter that translates to distortion in the audio at the point of conversion back to the analog domain. Sources of digital audio, the digital interface, the mechanisms by which errors are introduced, and the effect on DACs are examined.

0. Introduction

It has been known for a very long time that jitter of the sample clock for the conversion from and to the analog domain can introduce error into the sampled or reconstructed signal. Sampling theory is predicated upon the idea of a uniform sample rate. With a changing signal, a sample reproduced at the wrong time will result in an amplitude error. The amplitude error will increase proportionally to the amount of timing error, and also depends upon the rate at which the signal undergoing sampling is changing.

A simple arithmetic calculation can approximate the amount of timing error that results in a 1-quantizer level error for an n bit system. For a sinusoidal output the maximum rate of change occurs at the zero crossing of the highest frequency full-scale signal that the system can transmit. That rate of change is $2\pi f 2^{n-1}$. If we restrict the maximum amplitude error to 1 quantization level, then the timing error can be no more than $1/(2\pi f 2^{n-1})$. For a 20 bit system with a full-scale 20 kHz sine wave this amounts to an error of 3.98 ps rms, a very small quantity indeed.

The authors, having become concerned about the possible effect of jitter upon the performance of systems which utilize the IEC 60958 Digital Audio Interface [1], embarked upon a project to quantify the amount of jitter tolerable in products that utilize this interface to transmit compressed digital audio using the IEC 61937 standard [2]. The data transmitted via this standard is sent in bursts that are packed into the space normally occupied by PCM. Previous experience indicated that these bursts could cause jitter to be introduced in the interface, and far more of these types of systems exist than separate CD players with stand-alone DACs.

The methodology of this project was to quantify the amount of jitter in bitstreams from digital audio sources, jitter added by the digital audio interface, the behavior of Digital Interface Receivers (DIRs), and the effects of clock jitter on the output of a number of digital to analog converters (DACs). These observations include objective measurements on the susceptibility of DACs to jitter, and listening tests to determine the threshold of audibility for jitter with both sine wave signals and real program material.

1.0 Theory and Simulations

An excellent discussion of the theory of the mechanism by which jitter introduces distortion can be found in [4]. Superficial analysis would indicate that the process is similar to Frequency Modulation and upper and lower sidebands would be introduced at frequencies determined by Bessel functions, symmetrically placed about the reproduced frequency. In fact, the sampling process of detecting the edges of the jitter-modulated clock reduces the effect to one similar to Amplitude Modulation. Sidebands are introduced at frequencies equal to the reproduced frequency plus or minus the jitter frequency.

Simulations of an ideal Nyquist sampler were performed using software running on a personal computer. The system is modeled by assuming that a sine wave is sampled at uniform intervals, and that the samples are then reproduced at times which are varied sinusoidally or randomly relative to the correct time. The FFT of the data from the simulation is calculated and the additional spectral components that appear are the distortion products caused by the sinusoidal jitter. An example of one such simulation for sine wave jitter appears below in Figure 1.

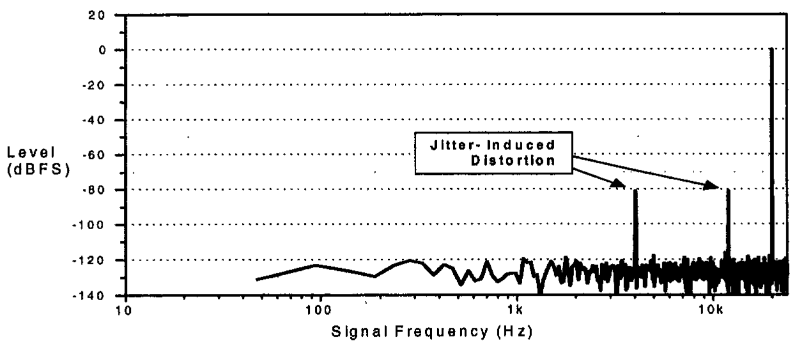

Figure 1: *Simulation of Effect of Jitter on DAC; 20 kHz at FS, 1 ns rms 16 kHz jitter*

The simulation is of a full-scale 20 kHz sine wave with 1 ns rms of 16 kHz jitter. The 16 kHz jitter causes two sidebands to appear relative to the 20 kHz sine wave that is being reproduced. The lower sideband appears at 4 kHz as expected, and the upper sideband is aliased to 12 kHz. Quantization noise appears throughout the audio spectrum for this full-scale sine wave without dither. In this simulation the jitter frequency has been chosen to create distortion products which will have maximum audibility.

Similar simulations with very high levels of jitter (>>100 ns rms) show the emergence of second order sidebands above the quantization noise floor at –120 dBFS. Since these second order sidebands do not appear except for very large amounts of jitter, and they are much smaller than the first order sidebands, their effect is negligible.

The results from numerous simulations with different frequencies and levels across the audio band, with jitter from 100 ps rms to 100 ns rms, were accumulated and are plotted below in Figure 2.

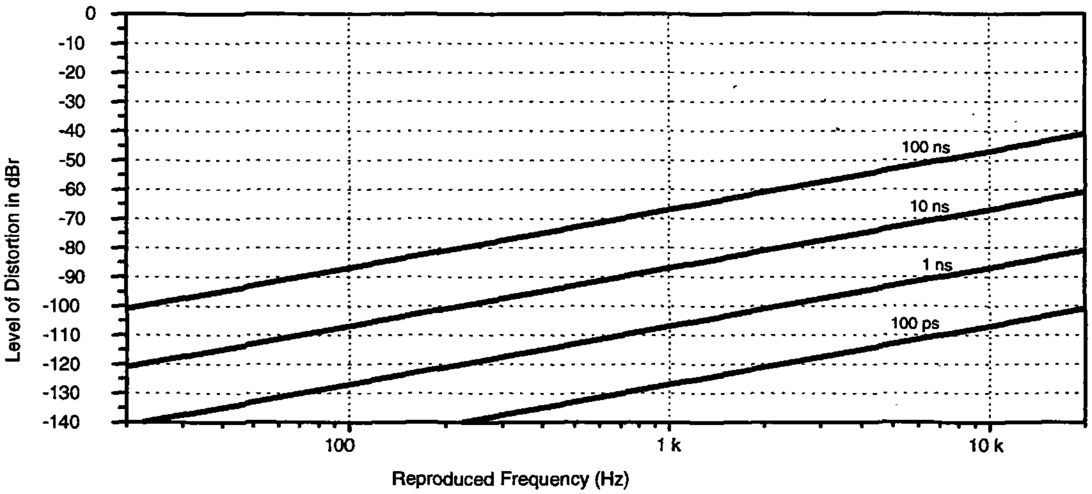

Figure 2: *Simulation of Effect of Jitter on DAC*

The primary dependency displayed in Figure 2 is that the jitter-induced distortion is directly proportional to the frequency being reproduced, and directly proportional to the level of the jitter. There is a negligible dependency on the jitter frequency. In a later section we will compare the results from these simulations to measurements performed on actual DACs. Assuming that these simulations are representative of real-world systems, the distortion varies from approximately –41 dB for 100 ns rms of sine wave jitter applied to a full-scale 20 kHz signal to –128 dB for 100 ps rms of jitter applied to a full scale 1 kHz signal. It is a reasonable assumption that distortion from the first condition is much more likely to be audible than distortion from the second one.

2.0 Jitter in the Digital Interface

Real-world jitter in the digital interface is usually a complex mixture of random white and non-white noise with many sinusoidal components. This can be analyzed by superposition. Adams [3] noted that several writers of data sheets neglect to specify whether their jitter specifications refer to rms, peak, or peak-to-peak. To this it should be added that the bandwidth and spectrum of the jitter should be specified. 1 ns rms of white phase noise jitter in a 1 MHz bandwidth is very different than 1 ns rms of sinusoidal jitter somewhere in the 0 to 20 kHz bandwidth.

Additional jitter is often inadvertently added by circuit implementation details in products. One such mechanism is the nature of decoding encoded audio in frames. A DSP which is decoding a given frame of audio processes for a period of time, usually a substantial fraction of the entire time occupied by one frame of decoded audio, and then pauses while waiting for the next frame. The current consumption is much higher during processing and decreases during the time while the DSP is idle. This change in the current drawn by the DSP produces a small modulation of the power supply voltage. Typically the clock recovered by the DIR is passed through the DSP and word clock is generated by dividing that clock by the appropriate ratio.

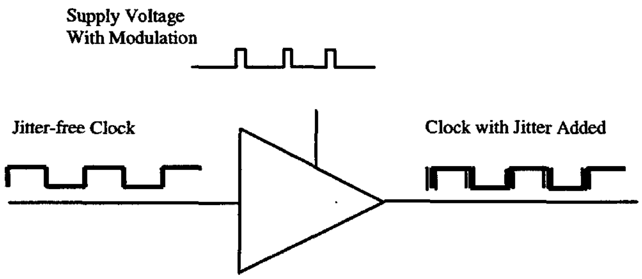

Figure 3: *Jitter Added by Supply Modulation of Gate*

Figure 3 shows that as the supply voltage is modulated, so is the transition voltage of any gates through which the clock is passed. At each of these stages the clock timing is modulated as a function of the rise time of the clock transitions and the changing transition voltage.

2.1 Jitter From Sources

The authors have measured the jitter spectrum of approximately 50 source products (DVD players, LD demodulators, and HDTV receivers) and found a wide variation in performance.

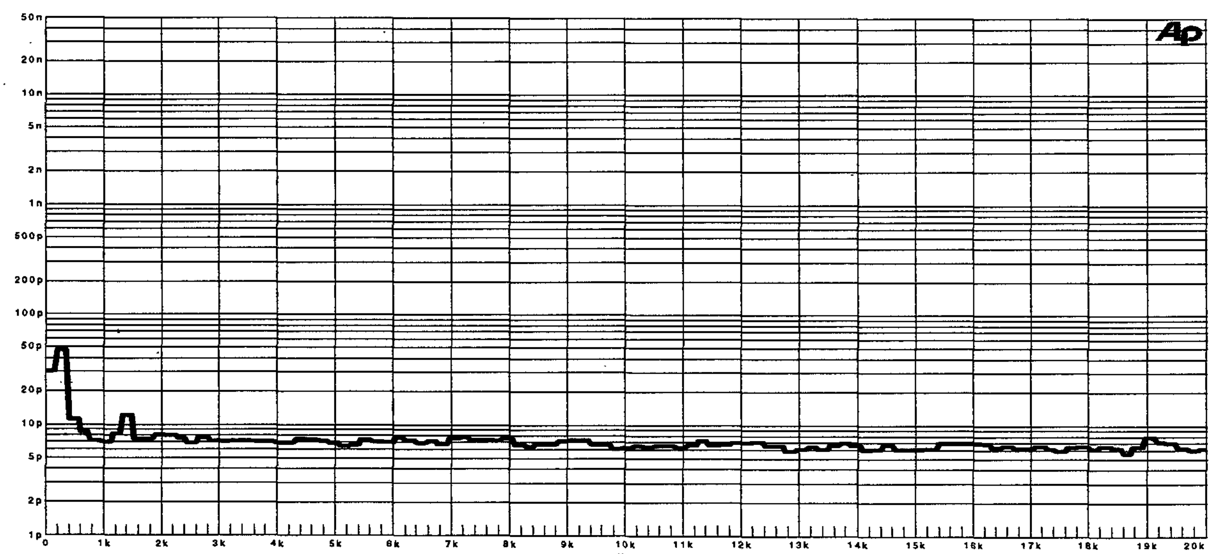

Figure 4: *Jitter Spectrum from a good DVD Player*

Figure 4 shows the spectrum of jitter measured from a laptop PC DVD ROM with optical output. Excellent performance is achieved despite the use of an optical interface. The jitter spectrum is basically white phase noise jitter with a spectrum level of 7 ps rms and small additional sinusoidal components at 250 Hz and 1.3 kHz.

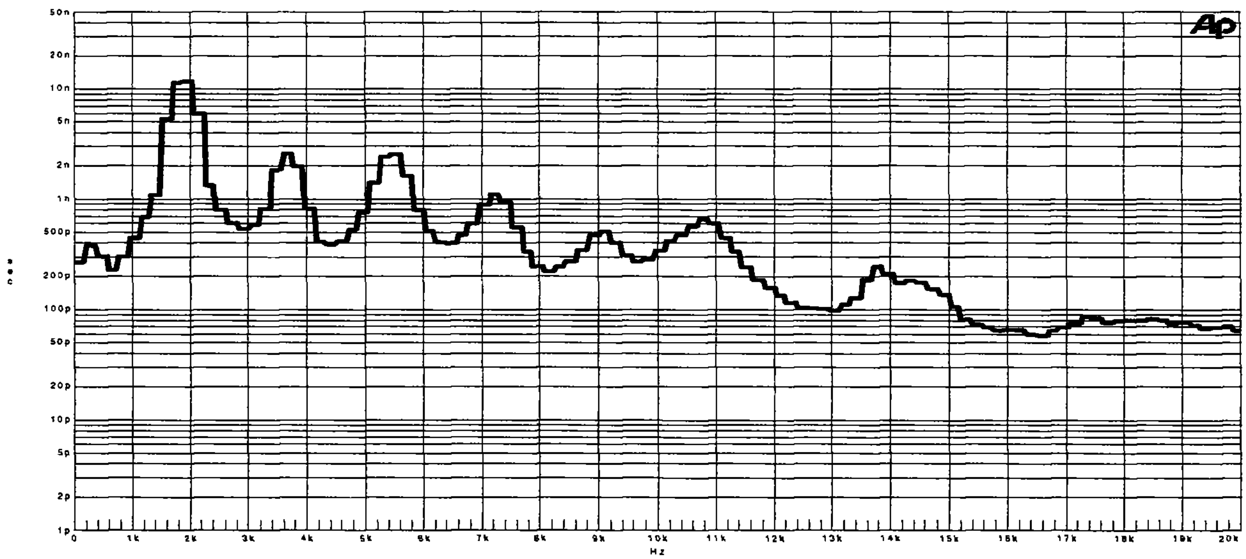

Figure 5: *Jitter Spectrum of Poor HDTV Receiver*

The graph above shows the spectrum of jitter measured from an HDTV Receiver with coaxial IEC 60958 output. The jitter spectrum shows numerous tonal components with a maximum of 12 ns rms of jitter at 1.8 kHz. The jitter performance is the worst measured in this study.

The jitter spectra from all of the measured produces were feature averaged. This was accomplished by estimating the overall noise floor of each measurement, and measuring the two largest peaks in each spectrum. The level and frequency of each of the measurements were averaged in order to give a composite result representative of a typical measurement. The typical spectrum has a noise spectrum level of 13 ps rms, and a peak of 142 ps rms at 250 Hz, and another peak of 112 ps rms at 3.7 kHz. The 250 Hz peak is present in every measurement and is caused by the status bits which are transmitted with a repetition rate of 192 frames. The other peaks are typically caused by circuit implementation details peculiar to each product design, and are at a variety of frequencies and levels, as can be seen in the two representative graphs above.

2.2 Jitter From Digital Sources

As has been demonstrated in [4], the signal carried in the digital audio interface can cause jitter to be generated. Since the interface always has some filtering affect on the digital signal that it carries, the waveform, and therefore the timing, is changed according to the pattern of transitions in the bitstream. A long cable is such a filter and will distort the waveform. This phenomenon is referred to as line-induced jitter.

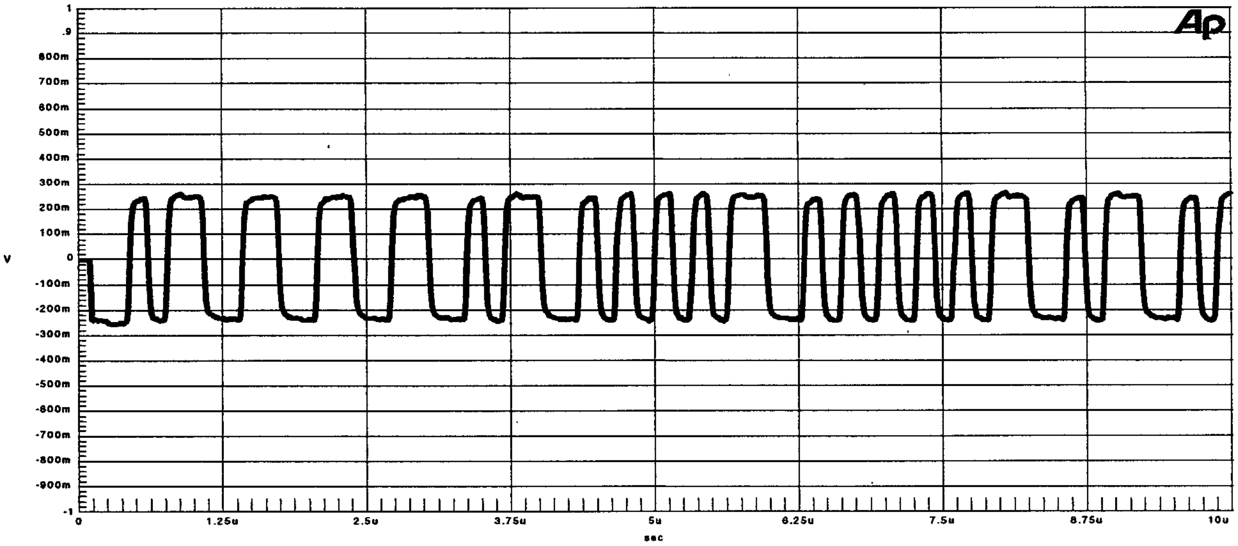

Figure 6: *Waveform from 30 m Cable Conveying IEC60958 Bitstream*

As can be seen in Figure 6 above, the zero crossing time of the waveform is dependant upon the length of the pulses preceding it. Since the clock generated by the receiver is dependant upon the detection of the transition at the time of the zero-crossings, timing jitter is generated. If the IEC 60958 bitstream is conveying linear PCM, then the pattern of the bitstream will be governed to some extent by the information contained within it. This effect is more significant for low-level signals.

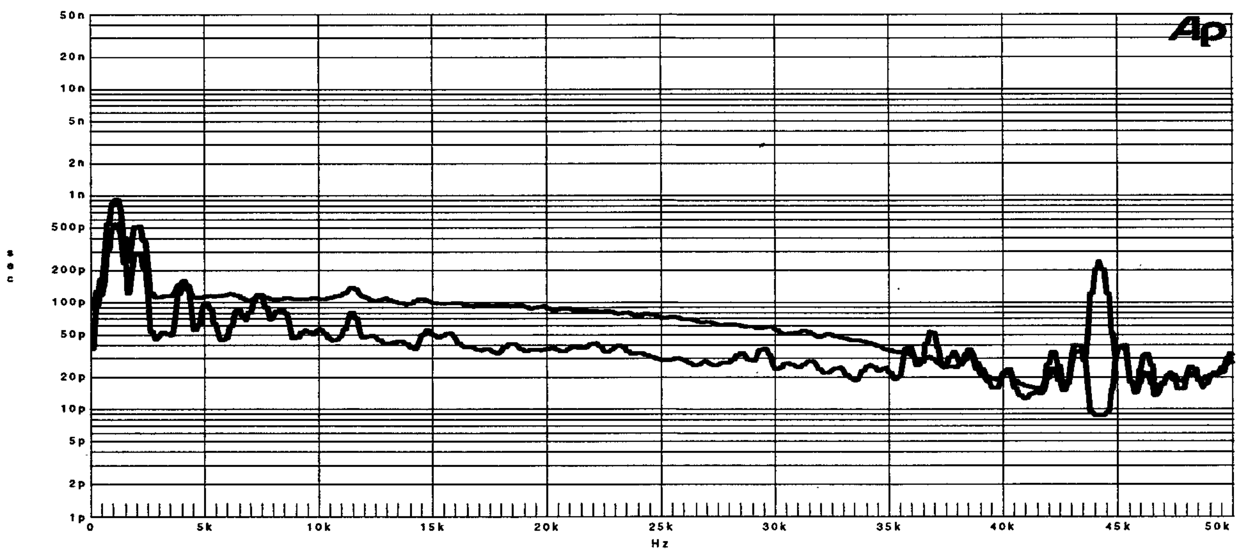

Figure 7: *Jitter Spectrum of DVD Player Reproducing –90.31 dBFS from CD*

The lower trace in Figure 7 represents the jitter generated when transmitting a –90.31 dBFS 1 kHz sine wave from the CBS CD-1 test CD from the IEC 60958 output of a DVD player. A strong component of about 550 ps rms can be seen to have been created by the 1 kHz signal transmitted by the interface. The upper trace is the spectrum of the jitter when the signal is transmitted through a 30 m. 75 Ohm coaxial cable. The overall jitter is increased and a jitter component at the sample rate is generated by transmission through the long cable.

The jitter is much greater for low-level signals than for high-level signals. Since every signal component of an audio program produces jitter signals at the same component frequency, and harmonics, the program is subjected to self-modulation. This effect could result in serious distortion of any PCM signals transmitted through an IEC 60958 interface.

In the example of Figure 7 above, the maximum jitter produced by line-induced jitter is about 900 ps. From Figure 2, the distortion produced by 1 ns of jitter is about –81 dB for a full-scale 20 kHz sine wave. In this case the signal is 90 dB lower than in Figure 2, so the distortion components will be at –171 dBFS. For higher level signals in the digital interface, the distortion caused by the jitter is greater, but the jitter produced by the signal is less. The change in line-induced jitter with level is shown in the figure below.

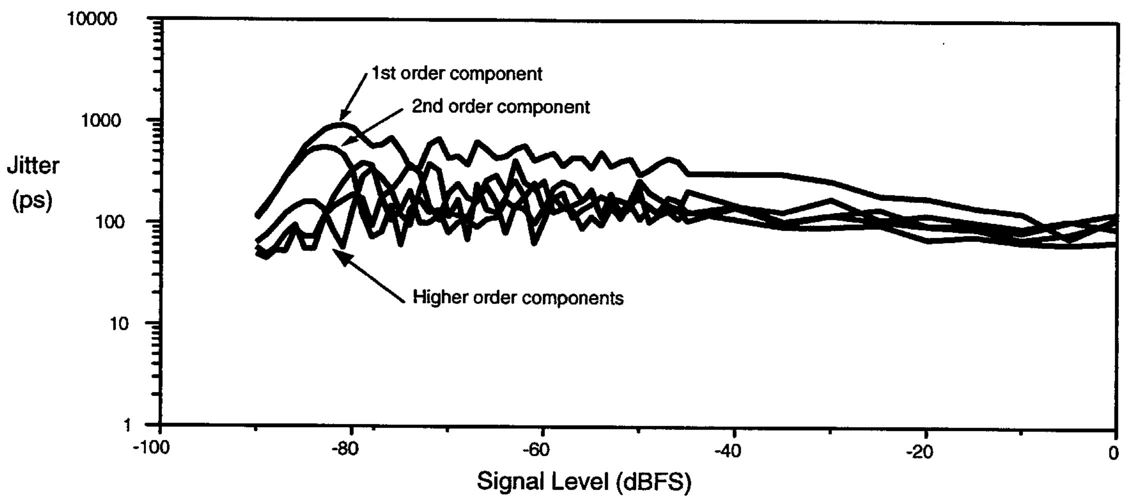

Figure 8: *Line Induced Jitter from 30 m Cable*

Figure 8 shows the first through fifth order jitter products generated by 16 bit PCM transmitted by IEC 60958 through a 30 m. 75 Ohm coaxial cable. As the signal level increases, the first order, and then higher order, jitter products appear in the jitter spectrum. They oscillate as a function of the PCM signal level, with some order components being more significant than others depending on the level. As the level increases further there is a general trend for the jitter components to decrease, and for them to converge at the same amount.

For any given frequency transmitted via the IEC 60958 digital interface, the principal jitter component will also be at same the frequency, with various harmonics showing up depending upon the level of the signal. The jitter induced by each signal component will modulate the other signal components, resulting in intermodulation distortion. Line-induced jitter results in very low levels of distortion, which are modulation products of the signal.

When the IEC 60958 interface is used to convey non-linearly coded digital audio using IEC 61937 instead of PCM, the dependence on the signal is removed. Figure 9 below shows the jitter spectrum of the IEC 60958 output of a good DVD player transmitted to the analyzer first through a 1.5 m cable, and then through a 30 m. cable.

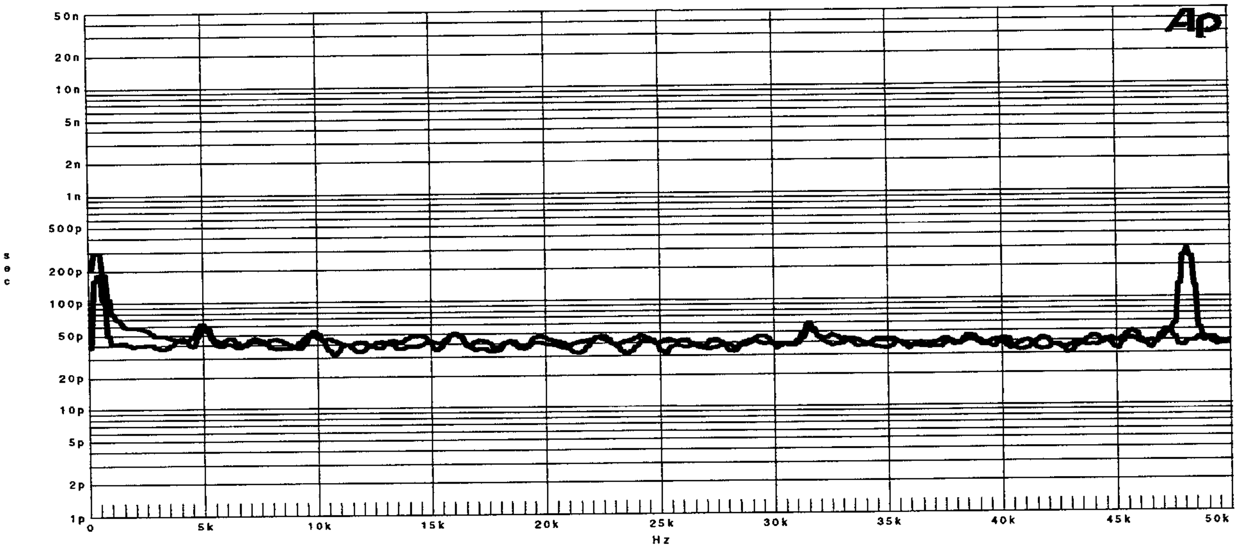

Figure 9: *Jitter in IEC 60958 Bitstream Caused by Bursts of Non-Linearly Coded Audio*

The lower trace is the spectrum of the jitter when the signal is transmitted through a 1.5 m cable and the upper trace is the spectrum when the signal is transmitted through a 30 m cable. The jitter spectrum with the short cable has two principal components. They are the usual peaks at 250 Hz and 5.0 kHz. If there is a contribution to the jitter at the burst rate of 31.25 Hz it is not visible due to the strong component at 250 Hz.

When the signal is transmitted through the long cable the jitter is increased slightly. The 250 Hz component increases. A component at 48 kHz, the frame rate of the bitstream, appears. The jitter in the range from 1 kHz to 5 kHz is increased, presumably by the harmonics of the burst rate, but this effect is small relative to the other jitter.

2.3 Receivers

Most products that are stand-alone digital audio decoders acquire their clocks for digital to analog conversion from the Digital Interface Receiver (DIR). Some small number of such products reclock using a Voltage Controlled Crystal Oscillator (VCXO). Properly designed VCXOs have very low levels of jitter and thus can recreate the clocks with low jitter. In the former case, jitter from the source product and jitter produced in the digital interface will be added to any jitter produced within the decoder product.

Digital Interface Receivers have a dual function. They detect the transitions of the signal transmitted by the interface and recover the clock and the data from that signal. Clock recovery from the bitstream of the digital interface is typically accomplished using a Phase Locked Loop (PLL). At low frequencies the PLL precisely follows the timing of the input signal and at some higher frequency it begins to attenuate the jitter of the input signal. The transfer function of the PLL loop filter determines the transfer function of the jitter rejection capability of the device. The jitter reduction capability of the DIR can vary greatly from device to device depending on the design compromises made in order to assure fast signal acquisition and good jitter attenuation. Theoretically, the clocks can be recovered with all jitter removed.

The jitter attenuation of a widely used DIR was measured by applying jitter to the interface signal input to the DIR with the recovered clocks, and then measuring the jitter, in the 70 Hz – 100 kHz bandwidth, of the clock recovered by the DIR. The transfer function of the measurement is presented below in Figure 10.

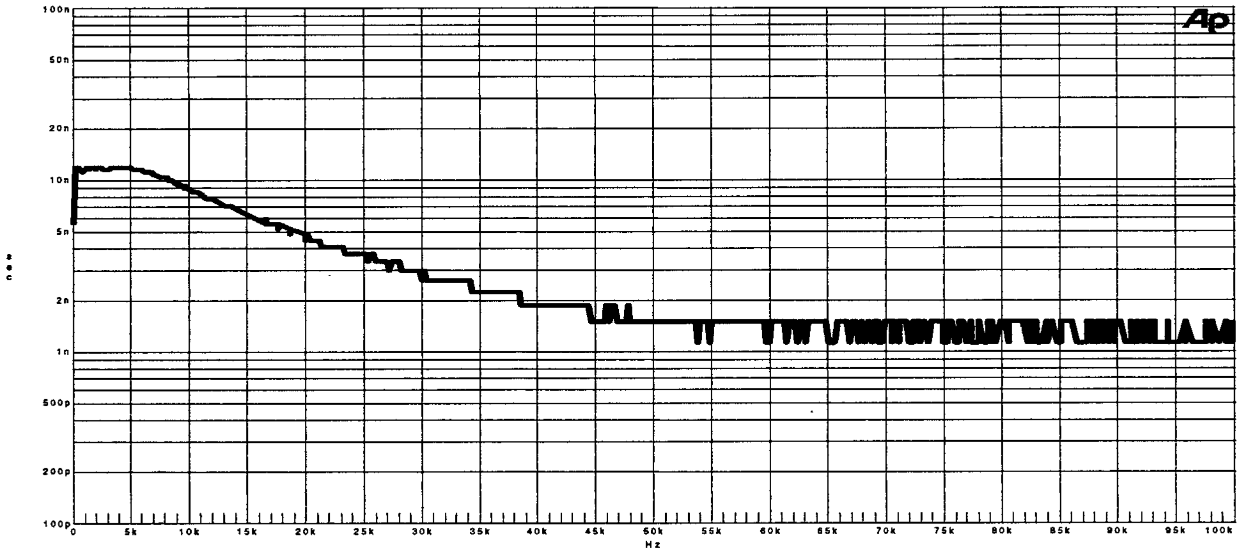

Figure 10: *Jitter Attenuation for Digital Interface Receiver with 10 ns rms Jitter vs. frequency*

The DIR in Figure 10 is specified by its manufacturer as attenuating jitter above 24 kHz. As can be seen in the graph, jitter is actually increased slightly for frequencies below about 8 kHz, and then attenuated above that frequency. The flattening of the jitter attenuation curve at a level of about 1.2 ns rms may be a limitation of the test set-up, or of the total jitter in the 100 kHz bandwidth of the measurement. From this measurement it is not possible to tell if the DIR continues to increase jitter rejection above 50 kHz. However, it should be assumed that it does so, since the measurement is wide-band and the effect of jitter outside of the audio band will influence the total quantity of jitter measured.

It can be seen that selection of a DIR with sufficient jitter attenuation could help to reduce jitter-induced distortion. A properly designed DIR similar to the one in Figure 6 will reduce the distortion caused by jitter by at least one order of magnitude for jitter frequencies above the Nyquist frequency.

3.0 Jitter Susceptibility of DACs

The jitter susceptibility of a DAC can be tested by applying jitter to the clock(s) applied to the DAC, or by applying jitter to the bitstream in the digital audio interface. Jitter in the bitstream at the level of the digital interface may be attenuated by the DIR. This will reduce the susceptibility of the DAC at the product level.

As reported by Adams [3], the intrinsic susceptibility of DACs to jitter varies according to their basic operational principles. His analysis divides DACs into three types; current division, delta-sigma with continuous-time reconstruction filter, and delta-sigma with switched capacitor reconstruction filter. The analysis shows that current division DACs and delta sigma DACs with switched capacitor filters have similar sensitivity to jitter. At the time of the writing of this paper the authors are not aware that any manufacturers of delta sigma DACs are using continuous time reconstruction filters and no such DACs were available for testing or listening. For that reason this class of DAC is not investigated in this paper.

To quantifying the sensitivity of a given DAC to jitter, it is possible to add the jitter to either the word clock of the device under test, or to add it to the signal in the digital interface. For the purposes of this study it was desired to explore the effects of the signal being transmitted by the interface, either from the source or generated within the interface. For that reason, the testing was performed by adding the jitter in the interface.

A number of DACs were evaluated for their sensitivity to distortion caused by jitter. The results for four of them are reproduced here. Each DAC was tested either on an evaluation board supplied by the manufacturer or in a product designed by one of the authors. Each DAC derived its clock from a DIR, or in the case of DAC D, the DIR was integral to the DAC. Each of the DACs had essentially flat frequency response, low distortion, and minimal modulation noise. The description of four of the DACs tested is listed below:

DAC A: two channel sigma-delta, single bit 64X modulator with fifth order noise shaper; with switched capacitor reconstruction filter followed by second order continuous time filter. Dynamic Range 92 dB.

DAC B: two channel sigma-delta single bit 128X modulator, fourth order noise shaper, with switched capacitor reconstruction filter followed by continuous time filter.
Dynamic range 105 dB. Specified as "low jitter sensitivity".

DAC C: single channel current division, 20 bits, 8X oversampling digital filter. Dynamic range 119 dB.

DAC D: six-channel sigma-delta single bit, with switched capacitor reconstruction filter followed by continuous time filter. Dynamic range 95 dB. Integral DIR.

The sensitivity of these DACs to jitter was assessed by measuring the residual distortion using a THD notch filter. During testing the input signal and jitter were varied over a range of levels and frequencies. This measurement can be accomplished using the Audio Precision System Two which has facilities for generating and analyzing jitter in the digital interface. At a given reproduced frequency, the jitter amplitude was fixed at a specified level and the jitter frequency was swept from 0 to 100 kHz. The graphs that follow represent the susceptibility of DACs A through D to jitter as a function of frequency and level, while reproducing 20 kHz at 0 dBFS. It should be noted that the residual THD+N floor of this measurement is often limited by the in-band component of the harmonic distortion of the input signal, not the jitter-induced distortion.

During this investigation, it was observed that the measurements from each DAC (Figures 11 – 14) exhibit some similar characteristics. Of interest, is the decrease in distortion at frequencies below 5 kHz is due to the THD notch filter. Low jitter frequencies cause sidebands adjacent to the reproduced frequency. Because these sidebands coincide with the THD notch filter, they are attenuated along with the reproduced frequency. For example, the notch at 20 kHz occurs because jitter at 20 kHz causes a sideband to appear at 40 kHz and 0 kHz, which is outside the pass-band of the analyzer. As the jitter frequency is swept above 35 kHz the lower sideband is swept through the notch filter and the 20 kHz low-pass filter. The upper sideband is always outside the 0 – 20 kHz pass-band. This phenomenon can be seen in Figures 11-14.

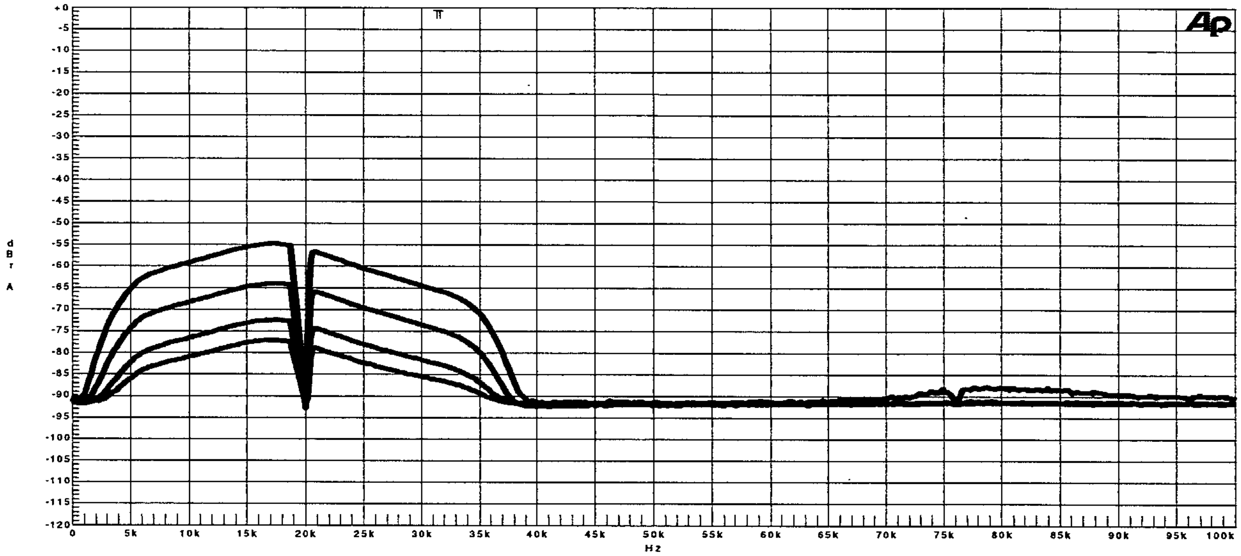

Figure 11: *Jitter Susceptibility of DAC A for 1.6, 3, 10, and 30 ns rms jitter vs. frequency*

DAC A shows a sensitivity to jitter which increases with frequency up to about 20 kHz, due to some unknown mechanism. The DIR used in this evaluation board was experimentally determined to attenuate jitter above about 8 kHz and this may explain why the distortion is less between 20 kHz and 40 kHz than it is below 20 kHz. The distortion is negligible for jitter frequencies above about 40 kHz.

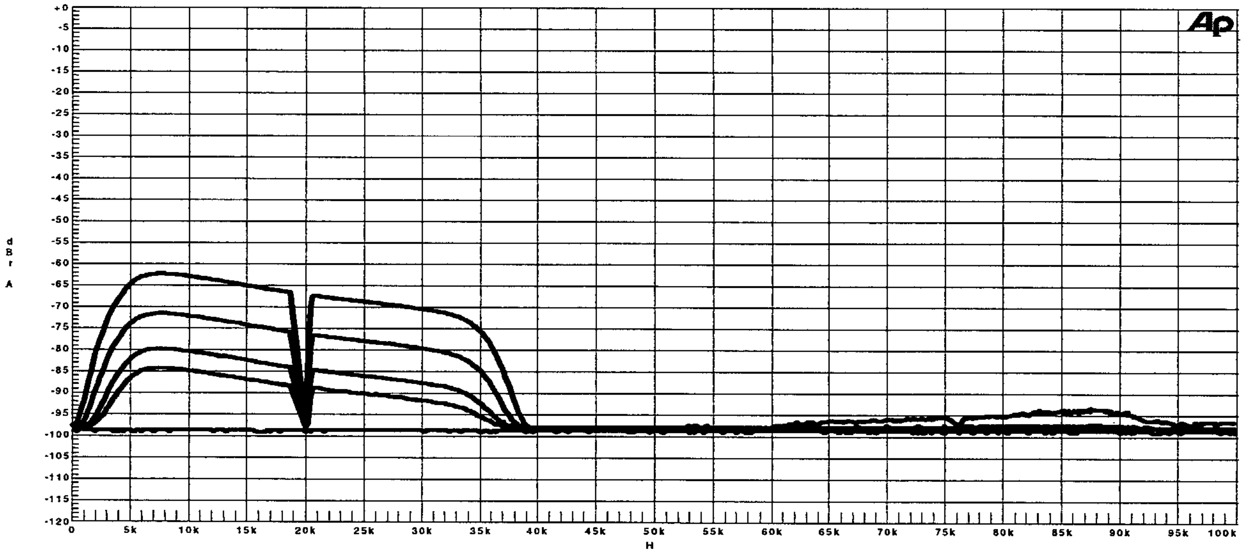

Figure 12: *Jitter Susceptibility of DAC B for no jitter, 1.6, 3, 10, and 30 ns rms jitter vs. frequency*

DAC B shows a sensitivity to jitter which decreases with frequency. The DIR used with this DAC is the same as with DAC A. The distortion is negligible for jitter frequencies above about 40 kHz.

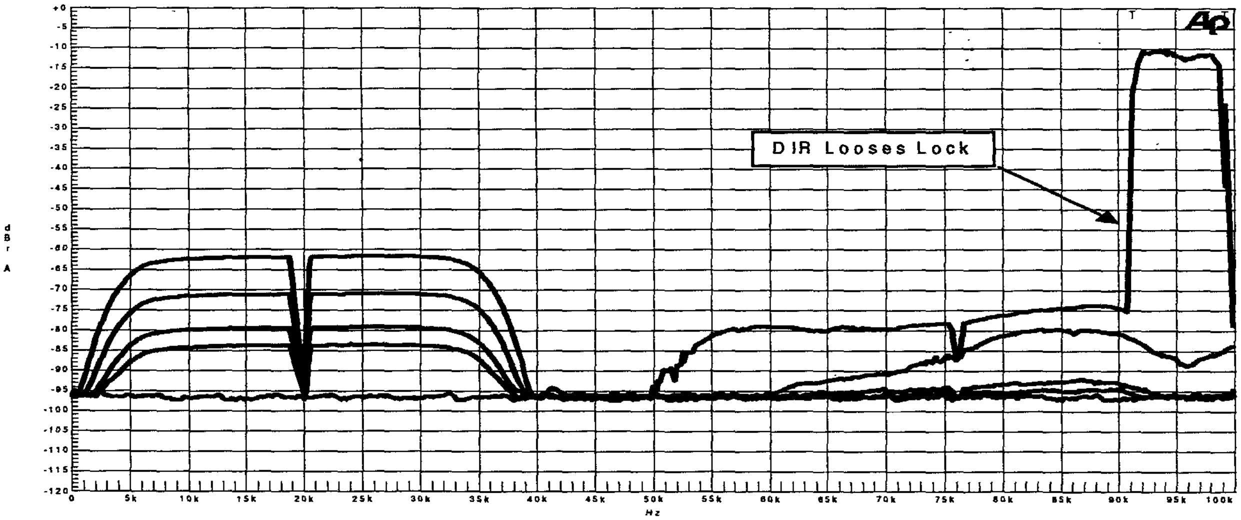

Figure 13: *Jitter Susceptibility of DAC C for no jitter, 1.6, 3, 10, and 30 ns rms jitter vs. frequency*

DAC C shows a nearly uniform sensitivity to jitter at all frequencies below the sample rate. The susceptibility is relatively high above the sampling frequency and at moderate levels of jitter the DIR loses lock.

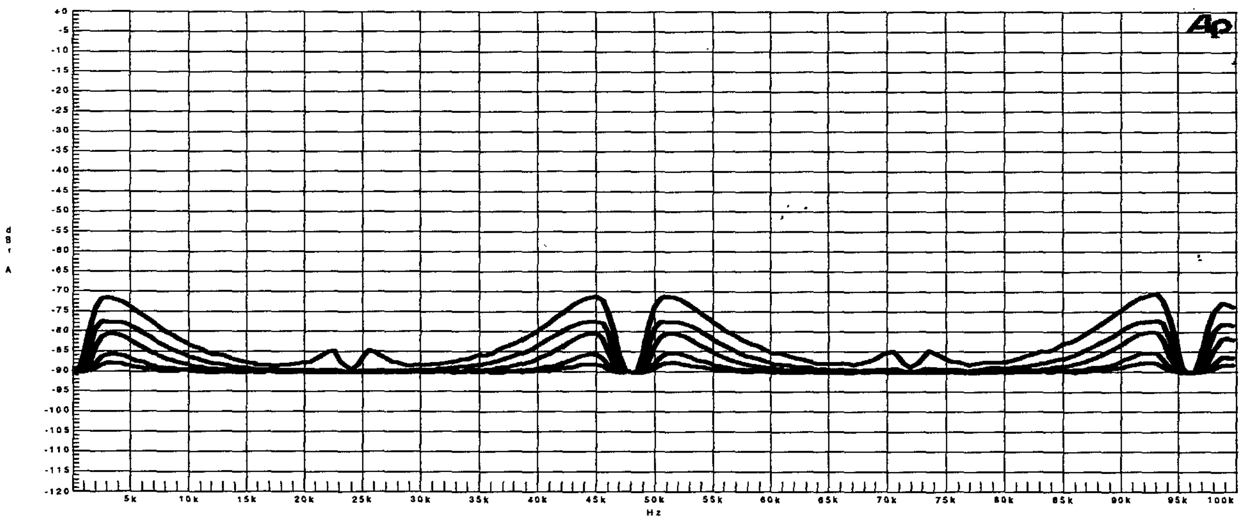

Figure 14: *Jitter Susceptibility of DAC D for 1.6, 3, 10, and 30 ns rms jitter vs. frequency*

DAC D shows sensitivity to jitter that varies dramatically with the jitter frequency. The DIR is integral to the DAC and apparently doesn't attenuate jitter at high frequencies at all.

DACs A-C are generally similar to each other in their jitter susceptibility despite the fact that they utilize very different conversion technologies. DAC C exhibited greater sensitive to jitter above 50 kHz. This is probably due to the design of the DIR used in the evaluation board, which doesn't attenuate jitter at higher frequencies. Lastly, DAC D, which has an integrated DIR, exhibited quite different results from the other DACs. The maximum distortion due to jitter was about 10-30 dB lower than the others and this DAC exhibited remarkably different distortion characteristics as a function of jitter frequency.

It is interesting to note that DACs A, B and C match the simulations fairly closely. According to the simulation, the distortion caused by sinusoidal 30 ns rms jitter in the presence of a 20 kHz tone is expected to produce –51 dBFS distortion components. DAC A, produced a maximum rms jitter distortion level of –54 dBFS, 3 dB lower than the simulations. On the other hand, DAC C produced -62 dBFS of jitter distortion. While this appears to be quite different from simulations, Adams [3] explains that 8X oversampling would theoretically cause 18 dB of jitter attenuation. DAC C actually has 7 dB greater susceptibility than this prediction. DAC B exhibited distortion at -63 dBFS, 13 dB below the simulated value, while DAC D exhibited the best jitter reduction, 30-40 dB below the expected value. This is presumably due to design considerations specific to this DAC.

4.0 Listening Tests – What is the threshold of Audibility for Jitter?

A series of listening tests were performed to determine the threshold of audibility of distortion caused by clock jitter in the digital interface. The basic experimental protocol was to provide a low-jitter source of program material, vary the amount of clock jitter added at the digital interface, and listen to the material through a DAC with known jitter susceptibility. The amount of jitter added to the interface was adjusted downwards until the listener was no longer able to distinguish between the low-jitter source and the bitstream with added jitter.

Listening tests on both sine waves and program material were conduced using Sony MDR-V6 headphones. It was expected that the threshold of audibility would be quite different depending on whether sine waves or program material are used.

During the initial investigation, it was found that the effect of sinusoidal clock jitter on the reproduction of sine waves was indeed audible on the DAC tested. This scenario allows the use of an up-down threshold test [5], [6] as a method for measuring the threshold of audibility. In this test, the jitter level was increased or decreased linearly with time and the listener was instructed to indicate when the jitter-induced distortion became audible or inaudible. The jitter threshold level was recorded over a two-minute period. From this data the threshold of audibility can be obtained by averaging the reversal points.

Since sine waves are not representative of normal program material, a listening test to evaluate the audibility of jitter on program material was also performed. The dynamic nature of this material presents problems when trying to obtain an audible threshold because varations in the program material will cause changes in the jitter-induced distortion. It is the observation of the authors that these changes will usually cause the distortion to be inaudible. Because the audibility of jitter distortion will vary as the program material changes in frequency and level, conducting a standard up-down threshold test using normal program material could take a prohibitive number of trials before meaningful results are obtained.

In order to reduce testing time and the number of reversals necessary, a self-regulated up down threshold test was devised. This test shortens the feedback loop of the up-down threshold test by providing the subject with direct control over the amount of jitter added to the program source. This allows the subject to quickly identify and home in on the audible threshold. During this process, the subject was allowed to confirm his or her threshold using an AB comparison box. This box allowed the listener to compare the low-jitter bitstream with the jittered bitstream and further refine his or her threshold.

4.1 Test Setup

A block diagram of the test setup is shown in Figure 15. The low-jitter program source was provided by a Compact Disc player with an audio-band (20 Hz – 20 kHz) jitter spectrum level of 80 ps rms. If the threshold had tended to approach the level exhibited by the CD player, a source with lower jitter would have been procured. Because thresholds stayed well above the 1 ns rms range, this CD player was deemed acceptable.

The IEC 60958 output of the CD player was connected to a distribution amplifier with two outputs. The first output was connected to input 'A' of an AB comparator box. This input was used as the low jitter reference for comparison with the jittered bitstream. The second output from the distribution amplifier was connected a Prism JM-1 jitter modulator [7]. The Prism JM-1 has the ability to add jitter using an external generator as the jitter frequency and level source. A function generator was connected to this input and used to set the jitter frequency and level throughout these tests. The jittered output of the JM-1 was connected to input 'B' of the AB comparator box, which is used to compare the two bitstreams.

The output of the AB comparator box was connected to DAC B, which was chosen because of its widespread use in consumer electronics and professional products. The output of this DAC was connected to a headphone amplifier that drove the Sony MDR-V6 headphones.

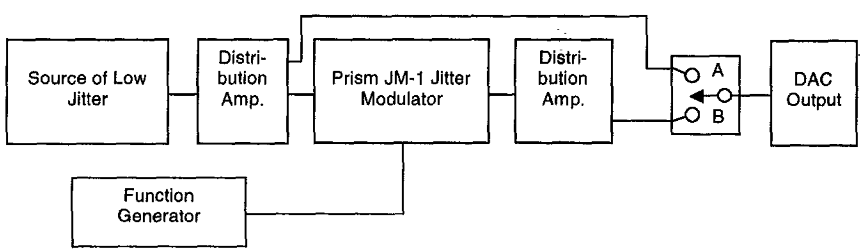

Figure 15: *Experimental Setup for Up-Down Threshold Test*

Distribution amplifiers were used to avoid problems with reflections and signal degradation when splitting the IEC 60958 output. Amplifiers with sufficient bandwidth and adjustable gain were chosen to ensure coherence between the signal levels of both bitstreams. All equipment was connected with 75 ohm coaxial cable and properly terminated to prevent reflections and the addition of jitter. After this setup was complete, measurements were made to ensure that jitter was not being added between the bitstream source and DAC.

4.2 Training

While conducting preliminary testing, it was observed that the listeners' sensitivity to distortion from jitter tended to increase during the course of testing, sometimes by a factor of two or three. It was apparent that training was taking place during the testing process. This is undesirable and would tend to artificially increase the threshold measurements. For this reason, a two part training procedure was developed that was found to greatly improve sensitivity to jitter distortion.

Using headphones and an AB comparator box, the subject was asked to compare the difference between high-jitter and low-jitter bitstream. Subjects were allowed an unlimited amount of time to familiarize themselves with the two bitstreams. Once the subject was confident that the two bitstreams were audibly different, the second stage of training began.

In the second stage of training, the subject was given control over the amount of jitter added to the system at a fixed jitter frequency, and instructed to become familiar with the change in distortion caused by adjusting the jitter level. Additionally, they were asked to indicate their threshold of audibility. This threshold was used to calibrate the JM-1 and ensure that the proper scale was used during testing. Testing began when the subject was satisfied with their ability to hear the effects of jitter on the program material.

At the end of the training session, any subjects that exhibited unusual difficulty in hearing the effects of jitter were excused from final testing. This ensured that listeners who do not have acute ability to hear jitter distortion did not inflate the measured threshold level. Training was conducted using the same program sources and jitter frequencies that were used during testing.

4.3 Program Material

Up-Down Threshold Test (Sine Wave Material)

Threshold testing was conducted at 3 frequencies: 4 kHz, 8 kHz, and 20 kHz, with jitter frequencies of 2 kHz, 5 kHz and 17 kHz respectively. The jitter frequencies chosen will create a lower sideband at 2 kHz for the 4 kHz source, and 3 kHz for the 8 and 20 kHz sine waves.

Self-Administered Threshold Evaluation

Program material was chosen during a lengthy screening process. The initial list of material was selected based on recommendations from audio industry experts, theory presented in this paper, and listening tests conducted by the authors. Given the analysis presented in this paper, program material with substantial high-frequency and minimal low and mid-frequency content was chosen. The recordings that contained substantial full-bandwidth information, excessive noise or clipping were rejected.

The remaining material was further refined by examining their spectra. The spectrum of each was rated using the following criteria:

1. Strong fundamental frequency and harmonics at or above 1 kHz
2. Minimal frequency content between 400 and 1 kHz
3. Long sustain
4. Low noise floor

The most promising recordings were selected for further investigation. Additionally, the observed spectra was used to estimate a jitter frequency that might result in audible distortion. This estimation was made by reviewing the spectra and selecting frequencies at which low-level sidebands might be audible. The jitter frequency was found by subtracting the estimated sideband frequency from the fundamental frequency of the program material.

The final selection was made during a listening test with jitter added to the bitstream. Initially, the jitter frequency was set as described above, and jitter level set as high as possible without causing the DIR to lose lock. During testing the jitter frequency and level was varied and the audible threshold at each frequency was evaluated. Eventually, the list of material was reduced to 4 recordings that were believed to be the most audible. Three of the recordings included just one note from a single instrument. The last recording was taken from a CD of synthesized music. Listed below are the titles, tracks, times and jitter frequencies used during this evaluation.

Title	CD #	Track	Time	Jitter Frequency
McGill University Master Samples Volume 3	OWCD-3-87	6	1:53 – 1:56	1700 Hz
McGill University Master Samples Volume 3	OWCD-3-87	7	2:19 – 2:23	1850 Hz
McGill University Master Samples Volume 2	OWCD-2-87	3	1:39 – 1:44	1700 Hz
Bachbusters	CD80123	14	0:06 – 0:010	1530 Hz

Table 1: *Program Material for Self-Administered Threshold Evaluation*

4.4 Test Procedures

Up-Down Threshold Test

To set up the test, the subject was given headphones and a 4 kHz full-scale sine wave was played. The subject was instructed to adjust the volume so that signal was loud but not painful. This volume setting was used throughout testing of the subject. While it would be possible to increase the volume at higher frequencies (8 kHz and 20 kHz), the correlation between the threshold levels at each frequency would be destroyed. For this reason output level changes were prevented during any single evaluation.

Testing began with the 4 kHz source and 2 kHz jitter frequency. Once testing started, the jitter level was increased until a positive response, indicating audible distortion, was indicated. This caused the trend in jitter level to reverse and slowly decrease until a negative response indicating no audible distortion was indicated. This process was repeated for two minutes and always ended on a negative response to allow for an even number of positive and negative records. The results were then analyzed and the nominal threshold for each frequency were calculated. A sample graph of the results from one up-down threshold test is shown below:

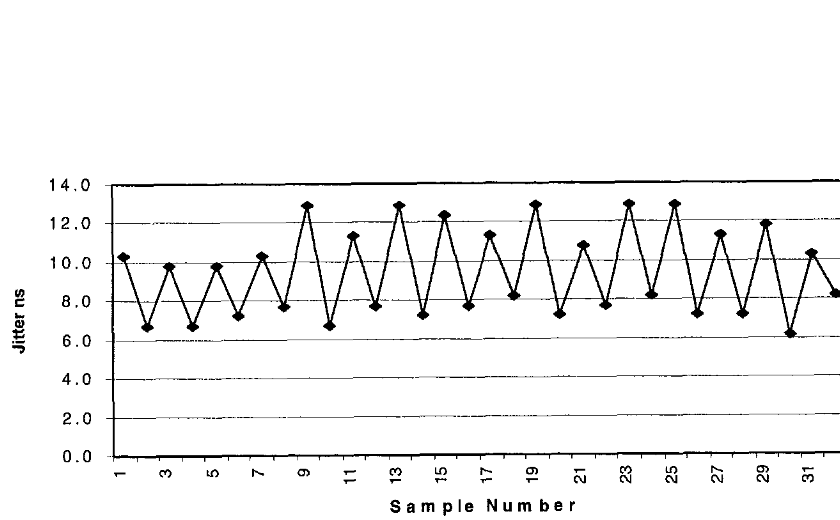

Figure 16: *Sample Up-Down Threshold Test Results*

Self-Administered Threshold Evaluation

After training was completed, the evaluator began playing the first title and the volume was adjusted so that it was loud but not painful. A variable output level was chosen because a fixed listening level might be too low for some and too loud for others. This would undoubtedly skew the threshold results. For this reason, it was decided that the over-all output level should be allowed to change for each of the 4 programs being evaluated.

After the output level was adjusted, testing began and the subject was instructed to adjust the jitter level until their threshold of audibility was reached. During testing, the subject was free to use the AB comparator to compare the two bitstreams for audible differences. This allowed for further refinement of the threshold by providing the subject with ready access to a low-jitter reference. Once a threshold was reached the result was recorded. This test was repeated for all of the selected program material.

4.5 Listening Test Results

Up-Down Threshold Test

The results from each test were compiled, and the average threshold level of each subject was calculated along with the maximum and minimum threshold levels. The results for a three tests are presented in Figures 17-19.

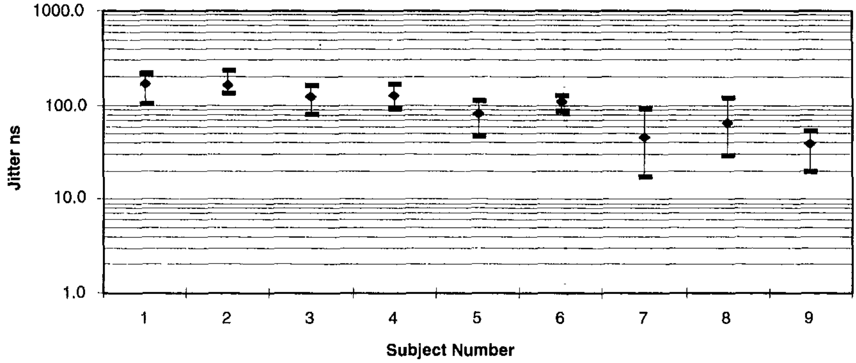

Figure 17: *Threshold of Audibility of Distortion for 4 kHz Sine Wave*

Figure 17 shows the results of the 4 kHz sine wave test. The threshold for this test is generally higher than the threshold for the 8 kHz and 20 kHz test. This is understandable because the 4 kHz tone and 2 kHz jitter will result in a 2 kHz distortion tone which will be masked at lower jitter levels.

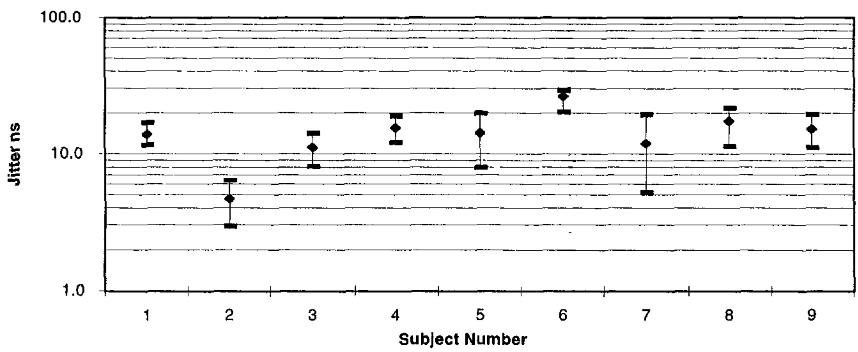

Figure 18: *Threshold of Audibility of Distortion for 8 kHz Sine Wave*

The results shown in Figure 18 are considerably lower than the 4 kHz tone. This is mostly due to the large reduction in the effect of masking because the fundamental is farther away from the distortion tone.

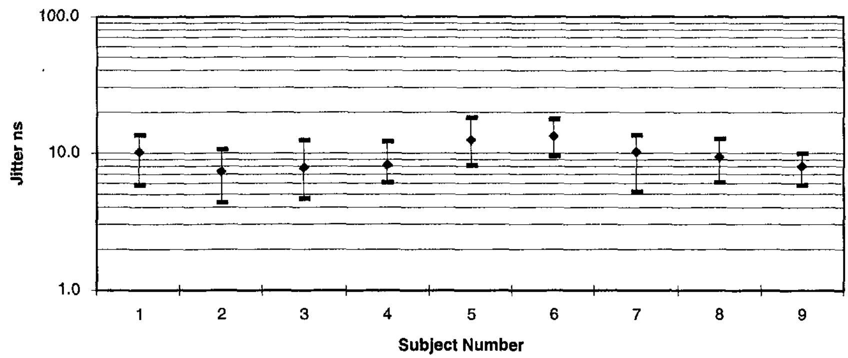

Figure 19:*Threshold of Audibility of Distortion for 20 kHz Sine Wave*

As expected, figure 19 shows a slightly lower thresholds than the 8 kHz test. The 20 kHz tone and 17 kHz jitter result in a 3 kHz distortion component. This is the same distortion frequency that was caused by the 8 kHz sine wave test. It is evident that the reduced masking and increased distortion relative to the 8 kHz test should result in increased sensitivity for the 20 kHz tone. While a 50% increase in sensitivity was observed between the two tests, the results do not entirely agree with masking theory.

Figure 2 shows a plot of the Distortion Level vs. Frequency for various levels of jitter. This figure shows that an 8 to 9 dB increase in level is expected between the distortion in the 8 kHz test relative to a 20 kHz test. This difference was not observed during testing. The authors are not aware of a reason for this discrepancy and it is the subject of future investigation.

Self-Administered Threshold Evaluation:

The results from the self administered threshold evaluation for 8 subjects are shown in table 2. The three selections from the MUMS discs yielded the most audible results. The selection from Bachbusters included several synthesized notes that resulted in significantly higher threshold levels than the MUMS material.

Subject Number	1	2	3	4	5	6	7	8
Program Material Disc and Track	Jitter in ns rms							
MUMS 3 track 6	120	90	130	50	270	250	140	60
MUMS 3 track 7	40	60	110	32.5	90	100	90	50
MUMS 2 track 3	80	120	150	26	310	130	150	20
Bachbuster track 14	NA	330	222	112	NA	256	370	130
NA: Subject could not hear the distortion caused by jitter.								

Table 2: *Self Administered Threshold Evaluation Results*

These results are not unexpected because each track from the MUMS discs contains a single that is very tonal in nature. This produces distortion components that are relatively constant and are similar in nature to the sine wave tests presented earlier. The material from Bachbusters, on the other hand, consists of about 5 different notes. The distortion components resulting from this program material will vary in level and frequency relative to the chosen jitter frequency. This means that only some of the notes will produce the most audible distortion components, while other may result in nearly inaudible distortion. This makes it more difficult for the listener to identify the jitter-induced distortion, as the distortion component may only be audible during a single note of the material. Additionally, forward and backwards masking could also help hide the distortion. For these reasons it is understandable that this track resulted in a much higher threshold than the first three tracks.

5.0 **Psychoacoustics** – What does masking predict about the audibility of jitter?

The distortion introduced by jitter in the digital audio interface is directly proportional to the frequency and amplitude of the signal being produced. This means that, although the error is approximately -81 dBFS for a full-scale 20 kHz sine wave with 1 ns rms of jitter, the distortion will be about -95 dBFS for a 4 kHz signal with the same jitter. If the reproduction level is 97 dB SPL, then the distortion of 4 kHz by 1 ns rms of jitter will be at 2 dB SPL and near the threshold of audibility, even in the absence of the signal. For a signal at a frequency lower than 4 kHz the distortion caused by 1 ns rms of jitter will always be inaudible at this reproduction level; the threshold of hearing is higher at lower frequencies and the distortion caused by jitter is less.

Signals cause auditory masking of other smaller signals that are close to them in frequency. This means that weak sounds which are close in frequency to louder sounds can't be heard. For sine wave signals, the masking effect can be visualized as families of curves which show how much these adjacent frequencies are masked, depending on the level of the masker, and the level and frequency of the signal being masked. An example of such a making curve from [8] is shown below.

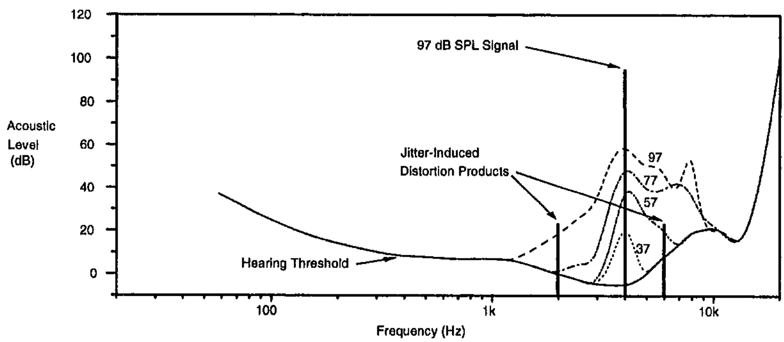

Figure 16: *Masking by 4 kHz Sine Wave*

Usually the effect of masking is more significant for frequencies above the frequency of the masker. A 4 kHz signal causes significant amounts of masking both above and below it in frequency. For example, a 4 kHz signal at 97 dB SPL causes the masked threshold at 2 kHz to be approximately 20 dB higher than the threshold of hearing in the absence of signal. The distortion due to 10 ns rms of jitter is –75 dBr for a 4 kHz sine wave. Since the threshold has been raised to approximately 20 dB SPL from the unmasked 0 dB SPL, the distortion at 22 dB SPL is 18 dB below the masked threshold of hearing. Jitter-induced distortion caused during the reproduction of sine wave signals becomes audible only when the jitter is very large, or in the case of very high frequencies where the effect of masking in the range of 1 to 4 kHz is minimal and the distortion due to jitter is greater. Virtually all audio program sources have a spectrum that emphasizes the middle frequencies more than high frequencies (or low frequencies). Audio program signals that have strong high frequency components almost always have even higher levels of mid range or high mid-range frequencies. This acts to make the effect of jitter extremely difficult to hear with normal program material.

6.0 Discussion

Numerous sources of IEC 60958 or IEC 61937 program bitstreams were measured and the spectrum of the jitter from the bitstreams analyzed. The jitter spectrum can be characterized as a white noise spectrum with various single frequency components superimposed. The noise jitter averages about 0.92 ps/$\sqrt{\text{Hz}}$. The sinusoidal jitter components varied from negligible to about 1 ns rms for the range from DC up to the sample rate, except for one source that exhibited sinusoidal components up to about 12 ns rms in amplitude.

Four DACs representative of modern implementations were tested and characterized as to their sensitivity to clock jitter as a function of the frequency and level of the signal being reproduced, and the frequency and level of the jitter. These DACs were found to be relatively similar in their sensitivity despite the differences in the technology of the implementation. DAC D showed markedly reduced sensitivity to jitter at most frequencies for reasons unknown to the authors but undoubtedly due to some specific detail of the design.

Listening tests were conducted to determine the threshold of audibility for sinusoidal jitter both for sine wave signals and for program material. Previous work has suggested various thresholds of audibility for clock jitter based on subjective tests [9], a masking model [10], and another hearing model not assuming any masking [11]. The minimum acceptable clock jitter from those studies ranged from 35 ns rms [9] for subjective tests down to 14 ps rms [11], when the hearing threshold is taken into account but masking is not.

Sine wave thresholds were measured using the UDTR technique. The threshold for sine waves is lowest for the case of 20 kHz FS reproduction where the mean threshold was found to be 10 ns rms of sinusoidal jitter at frequencies at 17 kHz. For 4 kHz sine wave full-scale reproduction, the threshold is raised due to the decrease in the amount of distortion due to jitter, and to the masking of the distortion components in the frequency range between 1 kHz and 7 kHz where the ear is most sensitive. For 4 kHz sine waves the threshold was found to be 100 ns rms.

Determination of the sensitivity to clock jitter for program material is much more problematic. The threshold is completely determined by the spectral characteristics of the program material. The vast majority of program material either has not enough high frequency signal to cause audible jitter-induced distortion, or has substantial mid-range content which masks the distortion. The authors were able to find 4 selections containing solo instruments in which the effect of jitter induced distortion was audible. Extensive listening tests using several subjects were conducted to determine the amount of jitter that caused a Just Noticeable Difference (JND) on these particular program examples. It was found that training enabled the subjects to hear the distortion caused by jitter at a lower level than at initial exposure to the phenomenon. The threshold of detection ranged from about 30 ns rms to 300 ns rms of sinusoidal jitter for critical source material.

Subsequent to the end of the listening tests, the spectrum of the audio reproduced by DAC B was analyzed in order to attempt to determine what distortion of the signal was occurring that caused the effect of jitter to be audible. The signal was placed in a recirculating buffer and 256 averages of a 4096 length FFT were calculated using the Audio Precision System Two to measure the output of the

DAC. The measurement was performed with no jitter added, 100 ns rms of jitter added, and 300 ns rms of jitter added. The resultant spectra appear below in Figure 17.

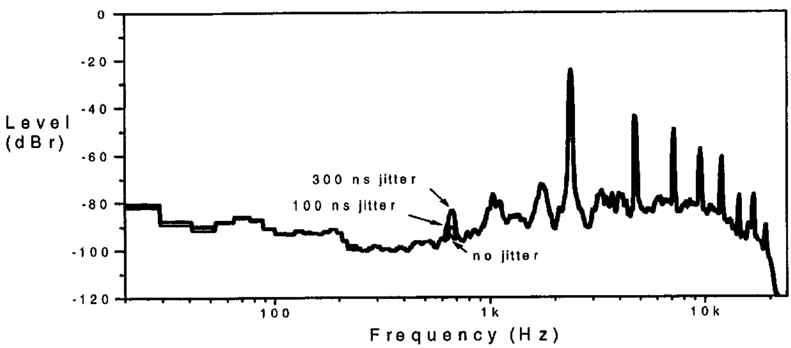

Figure 17: *Audible Distortion of Piccolo Playing D4 by Addition of Sine Wave Jitter*

The spectra show an easily visible tonal component at about 670 Hz that is proportional to the amount of jitter added. This component is the lower sideband of the two that are created by the 1700 Hz jitter applied during the listening test. There is presumably a corresponding upper sideband but it is not visible in these tests due to the noise in the spectrum of the piccolo between the fundamental and the second harmonic. Additional spectra measured with jitter of 30 ns rms and less showed no observable difference relative to the unjittered spectrum. The distortion for the 300 ns rms jitter level is approximately 60 dB below the fundamental component of the piccolo note, as predicted by the simulation and by the measurements of the DAC.

7.0 Conclusion

The effect of clock jitter in the digital interface was studied extensively. Measurements of the jitter spectrum of numerous digital audio sources, primarily DVD players, were conducted. A wide range of performance was found. The jitter spectrum of a typical source can be characterized as a white noise floor with one or many sinusoidal jitter components with a magnitude in the range of 10 ps to 10 ns rms. The effect of jitter induced in the interface was studied and found not to be a significant factor for short interconnection runs likely to found in a domestic environment. Several DACs and their DIRs were measured in order to characterize the sensitivity to distortion induced by jitter. These results were compared to each other and to results derived from simulations. Most DACs were found to be similar to each other and to the simulation in terms of susceptibility to jitter-induced distortion. That distortion is approximately $-107+20\log(F)+20\log(J)$ dBr for sine wave signals at F kHz with J ns rms of clock jitter.

Up-Down threshold and AB comparison listening tests were conducted to determine the threshold of audibility for jitter-induced distortion. The threshold of audibility for pure tones was found to be about 10 ns rms at 20 kHz and higher at lower frequencies. For nearly all program material no audible degradation was heard for any amount of jitter added below the level at which the DIR lost lock. Certain program material was found in which an audible degradation due to jitter was heard. The threshold of audibility for these programs was generally found to be in the range of 30 ns rms to 300 ns rms for sinusoidal jitter. Finally, the audible degradation was found to correspond to measurable changes in the spectrum of the program material.

The influence of jitter in causing audible distortion was found to be less than anticipated by the authors, and less than that predicted by both the technical and consumer audio press. Jitter induced by the digital audio interface was not found to be an audible problem for any of the program material auditioned.

It should not be assumed that jitter-induced distortion is a non-issue. Distortion induced by jitter is a real phenomenon and work to reduce its effects should continue. Although the threshold of audibility was found to be relatively high in the authors' experiments, the effect of all distortions in the audio chain is cumulative and it is reasonable to reduce them to the lowest practical levels. Manufacturers of DACs may find the methodology for evaluating jitter susceptibility presented in this paper useful in characterizing and presenting meaningful jitter specifications for their products.

8.0 Acknowledgments

The authors would like to express their appreciation to Edmund Chu, who designed and built the wide bandwidth AB comparator system used in the listening tests. We would also like to thank our many colleagues at Dolby Laboratories who participated in long and difficult listening tests.

9.0 References

[1] IEC 60958; "Digital Audio Interface" (1989)

[2] IEC 61937; "Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958" (1998)

[3] Robert Adams; "Jitter Analysis of Asynchronous Sample-rate Conversion", Presented at the 95th AES Convention, Preprint 3712, (1993 October)

[4] J. Dunn; "Specification and assessment in Digital Audio Equipment", Presented at the 93rd AES Convention, Preprint 3361 (1992 October)

[5] G. Von Békésy; "A New Audiometer", Acta Oto-Laryngologica, vol. 35, pp. 411-422, (1947)

[6] H. Levitt; "Transformed Up-Down Methods In Psychoacoustics", JASA, Vol. 49, No. 2, pp. 467–477 (1971)

[7] Prism Sound JM-1 Jitter Modulator Data Sheet, Prism Media Products Limited, Cambridge, UK

[8] Fielder, L. D., "Evaluation of the Audible Distortion and Noise Produced by Digital Audio Converters," J. Audio Eng. Soc., Vol. 35, pp. 517-535 (1987 July/Aug.)

[9] C. Manson, "Digital Sound Signals; Subjective Effect of Timing Jitter", BBC Research Department Engineering Report 1974/11 (1974 March)

[10] J. Dunn, "Considerations for Interfacing Digital Audio Equipment to the Standards AES-3, AES-5, AES-11", 10th International AES Conference, pp. 115–126, (1991 September)

[11] M. Hawksford, C. Dunn; "Is the AESEBU/SPDIF Digital Audio Interface Flawed", Presented at the 93rd AES Convention, Preprint 3360 (1992 October)