White Paper:

An Overview of the Coherent Acoustics Coding System

Mike Smyth
June 1999

Introduction

Coherent Acoustics is a digital audio compression algorithm designed for both professional and consumer applications. The algorithm is highly flexible, operating over multiple channels of audio, at sampling rates up to 192 kHz, with up to 24-bit resolution. This article outlines the overall design objectives of the algorithm, and describes the principal coding strategies employed to meet these goals. Specific functional components of the encoding and decoding algorithms are outlined in more detail, and the feature set of the coding system is summarized. Finally, the objective performance of the algorithm is illustrated using some simple test signals.

Global design objectives

The DTS Coherent Acoustics audio compression algorithm was designed with the primary objective of significantly improving the quality of audio reproduction in the home, beyond that of conventional compact discs. Consumers would benefit from more accurate sound recordings that utilized a wider range of audio frequencies, played back through more loudspeakers. The goal was to provide reproduction technology to the consumer that was as good as that found in professional music studios.

Secondarily, it was intended that the technology be used in a wide range of applications, in both the professional and consumer arenas, and that the consumer decoder be computationally simple and yet resistant to obsolescence. This required that the algorithm host a wide range of ancillary features suitable for home and studio use, and operate flexibly within a coding structure based around a complex intelligent encoder and a simple passive decoder.

Improving audio reproduction in the home

The key to delivering dramatic improvements in reproduction quality is the greater audio recording efficiency realizable through modern digital audio data-reduction techniques. In Coherent Acoustics this gain in efficiency has been used directly to improve the precision of the recorded audio. The importance of the concept warrants a more detailed explanation, which is given below. Other coding systems have used compression in a more traditional way, by simply attempting to minimize the data rate of the coded audio. While this approach can lower the cost of storing or transmitting the digital audio, it does not seek to improve quality. This is in contrast to Coherent Acoustics, which uses compression techniques primarily to maximize the quality of the audio delivered to the consumer.

Linear PCM coding

Historically, digital audio players, such as CD and DAT players, have used linear PCM coding for storing music signals. In these applications, the linear PCM data is recorded or stored in a simple, rigid format. The sampling rate and sample size are both fixed, and the data is read from the digital media at a fixed bit rate (table 1). This simple coding strategy has been very successful, and CD-audio has become the current consumer audio quality benchmark. Nevertheless, linear PCM coding is sub-optimal when used as a playback format for music. First, it is not optimized for the signal characteristics exhibited by typical audio signals; and second, it is poorly matched to the particular requirements of the human hearing system. While these inefficiencies do not necessarily limit the audio quality of linear PCM based systems, they do cause the amount of data required to represent future high quality audio formats, such as DVD-audio, to be excessive. The result has been that the rigid linear PCM format has become an impediment to continued improvements in audio fidelity, in both storage and transmission applications.

	Sampling rate	Sample size	Audio data bit rate [kbit/s/channel]
CD-audio	44.1 kHz	16 bits	705.6
Digital audio tape (DAT)	48.0 kHz	16 bits	768.0
DVD-audio (example)	96.0 kHz	24 bits	2304.0

Table 1. Audio data bit rates for consumer audio players

Linear PCM is ideal for coding full-scale, spectrally flat signals, such as white noise. However, since most music is not full-scale, and is spectrally tilted from low to high frequencies, it can be shown that linear PCM is an ‘objectively’ inefficient method for coding music. In other words, the ability of linear PCM accurately to represent the full-amplitude components of an audio signal over the full frequency range is not fully exploited for most musical passages, which normally have only small amplitude components at high frequencies. This leads to 'objective' inefficiencies, which can be measured by comparing the spectrum of the linear PCM coded signal to the full-scale, spectrally flat, ideal signal (figure 1).

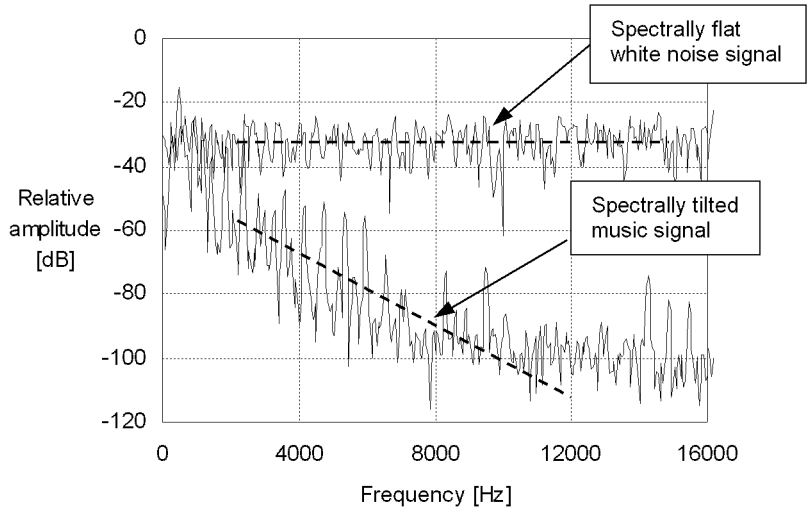


Figure 1. Spectral comparison between white noise and music signals

Similarly, the deviation from flat of the spectral sensitivity curve of human hearing (figure 2) creates a coding mismatch between the requirements for music perception and the specifications of linear PCM. The human ear is increasingly insensitive at high frequencies, and thus the coding by linear PCM of the small-amplitude high-frequency components of an audio signal is irrelevant, since these components cannot be heard. This mismatch leads to ‘perceptual’ inefficiencies in the linear PCM signal.

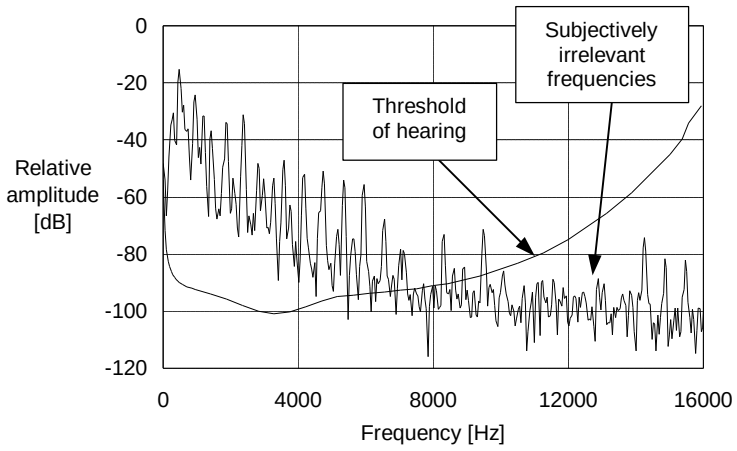


Figure 2. Spectral comparison between the threshold of hearing and an audio signal

Coherent Acoustics coding

These objective and perceptual inefficiencies, inherent in linear PCM encoded audio, can be reduced by using a more flexible coding scheme which takes explicit account of the underlying characteristics of audio signals and human hearing.

In practice, the original linear PCM coded audio is simply re-coded using a more sophisticated coding technique. This results in compressed audio data that requires fewer data bits to represent the original linear PCM audio signal. The reduction in data represents the removal of objective and perceptual inefficiencies from the linear PCM audio data, and the quality of the encoded signal is limited to that of the original linear PCM signal.

Traditionally, 16-bit CD-audio signals have been used for the original linear PCM signals, and consequently this approach reproduces 'CD-quality' audio but using fewer coding bits. An alternative approach, adopted by DTS, seeks to maximize the coded audio quality within particular data-rate constraints, and does not impose a limit on the quality of the original linear PCM data. In effect, this new approach does not exploit the objective and subjective redundancies in the signal in order to reduce storage costs, but primarily to allow more accurate audio signals than linear PCM, at similar bit rates, to be recorded and reproduced (figure 3).

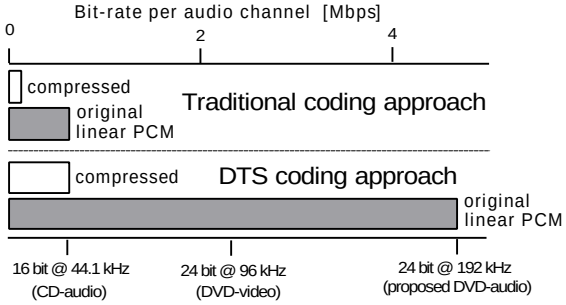


Figure 3. Using compression to improve audio quality

By substituting linear PCM with a more efficient coding methodology, DTS has been able to significantly improve the quality of digital audio reproduced in the home. For example, by replacing the linear PCM audio data with Coherent Acoustics, it is possible to provide a level of audio fidelity, from CD-audio media, in excess of that available in current professional music studios (table 2).

	Audio data-rate	Sample resolution	Sampling rate
CD-audio	705.6 kbit/s/channel	16 bits	44.1 kHz
Coherent Acoustics	705.6 kbit/s/channel	24 bits	192 kHz

Table 2. Replacing linear PCM CD-audio with Coherent Acoustics compressed audio

In practice, rather than operating at a single fixed bit rate, the coding algorithm operates over a continuum of bit rates, enabling the coded audio to vary from toll quality at very low bit rates, to mastering quality at higher bit rates (table 3). This flexibility is another key design feature of Coherent Acoustics.

Audio data-rate [kbit/s/channel]	Sampling rate	Sample resolution	Quality
8 to 32	Up to 24 kHz	16 bit	Telephony
32 to 96	Up to 48 kHz	20 bit	CD-audio
96 to 256	Up to 96 kHz	24 bit	Studio
256 to 512	Up to 192 kHz	24 bit	> Studio
Variable	Up to 192 kHz	24 bit	> Studio

Table 3. Variation of Coherent Acoustics audio data-rate with audio quality

Broadly applicable coding scheme

DTS Coherent Acoustics has also been designed to be broadly applicable in both the professional and consumer markets and, to this end, boasts a wide set of features to support all current and proposed formats, whilst retaining the flexibility to service format changes that may arise in the future.

Home audio is constantly evolving, with new sources and formats on offer, all promising a more involving sonic experience. These new sources, whether internet, DVD or satellite based, inevitably operate over a range of data rates, and these in turn provide different levels of quality to the consumer. The DTS coding system has been designed to accommodate these different rates, provide the necessary range of audio qualities demanded, and move seamlessly between them.

Audio systems must also handle a wider range of reproduction formats, from standard 2-channel stereo to 5.1 channels and beyond. More importantly, the direct one-to-one relationship between the audio data channels and reproduction channels (i.e., loudspeakers) has been superseded, and is increasingly under the control of the user. To service existing formats and allow new formats to emerge in the future, DTS is able to code multiple discrete audio channels, and provide additional up-matrixing (deriving new channels) and down-matrixing (folding channels together) functions.

Simple, future-proof decoder

To create a simple but future-proof decoder, the DTS coding system relies on two design features. The first puts all the 'intelligence' of the coding system in the encoding stage, creating a passive consumer decoder that is therefore relatively simple. The decoder merely follows instructions within the coded audio bit stream, generated by the encoder. This ensures that the encoding algorithm can be continually modified and improved, and that these improvements automatically benefit every consumer decoder. The second feature relates to the syntax of the data stream specification, which has been designed to provide room for additional audio data that may be needed in the future. This additional data could be used for improvements in audio quality or changes in the audio format.

In summary, the Coherent Acoustics coding system is flexible in operation for today's range of formats, but retains the necessary 'headroom' to allow improvements in audio quality to be readily implemented in the future.

Principal coding processes

By substituting linear PCM with a more efficient coding methodology, DTS is able significantly to improve the quality of digital audio reproduced in the home.

Coherent Acoustics is a perceptually optimized, differential sub-band audio coder, which uses a variety of techniques to compress the audio data. These are outlined individually below. Figures 4 and 5 show the main functional blocks involved in encoding and decoding a single audio channel. In keeping with the overall design philosophy, the complexity of the coding system is asymmetrically weighted towards the encoder. In theory, the design allows the encoding algorithm to be of unlimited complexity and, in particular, to evolve over time to include totally new methods of audio analysis. The decoder is simple in comparison to the encoder, since the decoding algorithm is controlled by instructions embedded in the encoded bit stream, and does not calculate parameters that determine the quality of the decoded audio.

Overview of encoder

At the first stage of the encoding process, a polyphase filter bank splits each single channel, full-band 24-bit linear PCM source signal into a number of sub-bands. Filtering into sub-bands provides a framework for both the exploitation of the short-term spectral tilt of audio signals, and for removing perceptual redundancies. Polyphase filters combine the advantages of excellent linearity, a high theoretical coding gain and excellent stop-band attenuation, with a low computational complexity. Each sub-band signal still contains linear PCM audio data, but has a restricted bandwidth. The bandwidth and number of the sub-bands created is dependent on the bandwidth of the source signal, but in general the audio spectrum is divided into 32 uniform sub-bands.

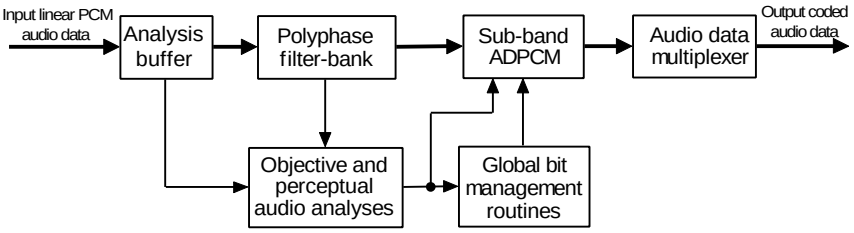


Figure 4. Functional block diagram of Coherent Acoustics encoder

Within each sub-band, differential coding occurs (sub-band ADPCM), which removes objective redundancies from the signal, such as short-term periodicity. In parallel, psychoacoustic and transient analyses are performed on the original linear PCM signal to determine perceptually irrelevant information. Depending on the bit rate, the results are used to modify the main differential coding routine operating on each signal. The combination of differential coding with psychoacoustically modeled noise-masking thresholds is highly efficient, thereby lowering the bit rate at which subjective transparency is achieved. As the bit rate rises, the dependency on psychoacoustic modeling is steadily reduced, ensuring that signal fidelity increases proportionately with bit rate.

The global bit-management routine is responsible for allocating, or distributing, the coding bits over all the coded sub-bands across all the audio channels. Adaptation occurs over time and frequency to optimize the audio quality. The bit-allocation routines translate the coding data rate into audio quality, and hence are of fundamental importance in the design of any coding system. By isolating these routines strictly to the encoding stage, the degree of complexity of the calculations involved is effectively unlimited, and can act to the benefit of all decoders. Again, as the bit rate rises, the flexibility of the bit allocation routines is reduced to ensure transparency over time and across all channels.

The final stage of the encoder is the data multiplexer, or packer, which receives encoded audio data from each ADPCM process. The multiplexer packs the audio data from all the sub-bands from all the audio channels, together with any additional optional and side-information, into the specified data syntax of the coded audio bit stream. Synchronization information is added at this stage to allow the decoder to reliably recognize this bit stream.

Overview of decoder

After synchronization, the first stage of the decoder unpacks the coded audio bit stream, detects and, if necessary, corrects data errors in the bit stream, and finally de-multiplexes the data into the individual sub-bands of each audio channel.

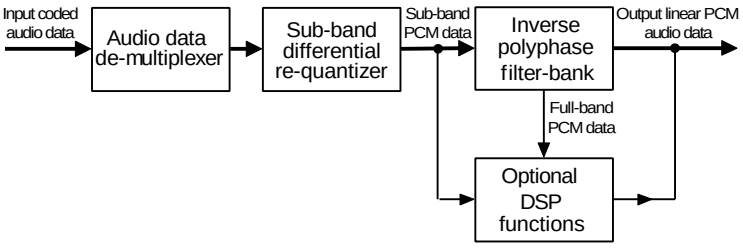


Figure 5. Functional block diagram of Coherent Acoustics decoder

The second stage inverse-quantizes the sub-band differential signals to sub-band PCM signals, following the instructions in the side information transmitted for every sub-band. The inverse-quantized sub-band signals are then inverse filtered to reconstruct the full-band time domain PCM signals for each channel. There are no routines in the decoder to adjust the audio quality.

The decoder also includes a DSP function block, which can be programmed by the user. This allows computations to be performed on either the sub-band or full-band PCM signals, on individual channels or globally across all channels. Examples of these functions include up- and down-matrixing, dynamic range control, and inter-channel differential time-delays.

Coding Strategy Analysis

The coding strategy for Coherent Acoustics revolves around two main processes: sub-band filtering and adaptive differential quantization (ADPCM). Filtering splits the uncompressed, full-band signal into narrower frequency sub-bands that can be manipulated independently of each other. It is the re-quantization process, occurring within the ADPCM, that effects the actual data compression.

Framing and filtering the input PCM signal

Coherent Acoustics operates on frames (time windows) of 24-bit linear PCM audio samples. Each frame of audio samples is filtered, then differentially coded to produce a frame of compressed output data. The choice of frame size involves a compromise between efficiency and performance. Large frames allow more compression during steady-state signal conditions, but can cause audible coding artifacts if the audio signal fluctuates rapidly in amplitude. Small frames code these transient signals better, but are less efficient in terms of overall compression.

The size of the PCM analysis frame defines the number of contiguous input samples over which the encoding process operates to produce one output frame. For Coherent Acoustics, five alternate frame sizes are permissible depending on the sampling frequency and the bit rate of the application. Figure 6 illustrates a frame size of 1024 input PCM samples, filtered into 32 sub-bands each containing 32 PCM samples. The five alternate frame sizes are 256, 512, 1024, 2048 and 4096 samples long, and the maximum PCM frame-size for sampling rate and bit rate is given in table 4. Generally, the larger windows are reserved for low-bit-rate applications, where coding efficiency must be maximized in order to retain quality. At higher bit rates, coding efficiency is less critical, and shorter windows are more applicable.

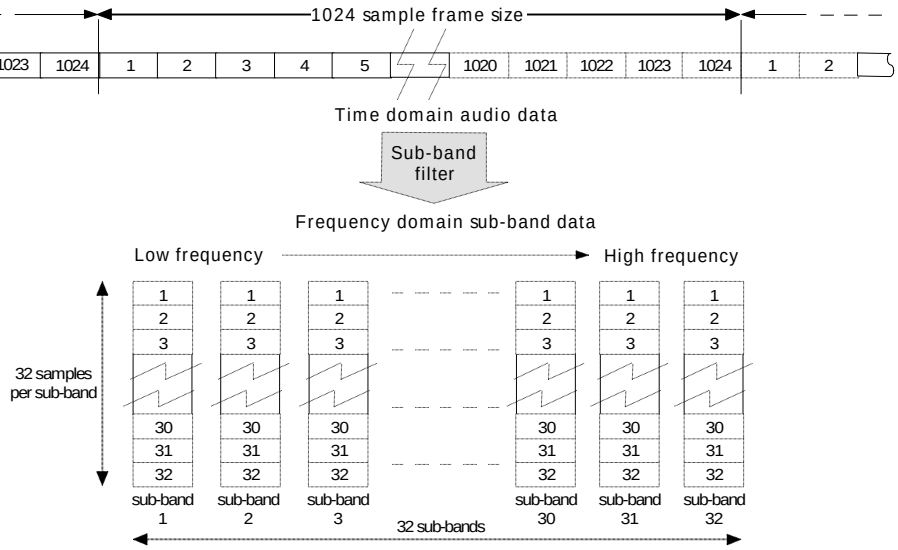


Figure 6. Framing and sub-band filtering the input linear PCM audio data

The frame size limits in table 4 were imposed in order to maintain a maximum ceiling on the decoder input buffer size. This is essentially a self-imposed constraint to reduce the complexity and cost of decoding hardware. For example, the maximum decoder input buffer size for sampling rates of 48 kHz, 96 kHz, or 192 kHz cannot exceed 5.3 kbyte, irrespective of the number of audio channels present in the bit stream.

	Sampling rate [kHz]				
Bit rate [kbit/s]	8 / 11.025/ 12	16 / 22.05 / 24	32 / 44.1 / 48	64 / 88.2 / 96	128 / 176.4 / 192
0 – 512	Max 1024 samples	Max 2048 samples	Max 4096 samples	-	-
512 – 1024	-	Max 1024 samples	Max 2048 samples	-	-
1024 – 2048	-	-	Max 1024 samples	Max 2048 samples	-
2048 – 4096	-	-	-	Max 1024 samples	Max 2048 samples

Table 4. Maximum sample frame sizes in Coherent Acoustics

Filtering the input audio signal

After framing, the linear PCM audio data is filtered into sub-bands. This is the first main computation of the Coherent Acoustics algorithm, and is of importance in analyzing objective redundancy in the audio signal. Filtering de-correlates the time-domain audio samples, and instead groups the samples into frequency ordered sub-bands. This time-to-frequency transformation rearranges, but does not alter, the linear PCM data, and simplifies the identification of those parts of the signal that are objectively redundant.

For sampling rates up to a maximum of 48 kHz, the input signal is split directly into 32 uniform sub-bands. This core stream is encoded using 32-band sub-band ADPCM. A choice of two polyphase filter banks is provided, perfect (PR) and non-perfect reconstructing (NPR). These filters have different properties, such as sub-band coding gain and reconstruction precision (table 5). The choice of filter bank is dependent on the application, and is indicated to the decoder by a flag embedded in the encoded data stream. At low bit rates, the coding efficiency is enhanced by using a filter with a narrow transition bandwidth and a high stop-band rejection ratio. This leads to a high de-correlation gain. However, in practice, filters exhibiting these properties do not reconstruct signals perfectly, and are prone to amplitude distortion at peak input levels. Nevertheless, for low-bit-rate applications, where coding noise far exceeds any filter-induced noise, the perceived audio quality is more important than the absolute reconstruction precision, and the NPR filter is normally preferred. Conversely, at high bit rates, or for lossless applications, the audio precision is critical, and the perfect reconstruction (PR) filter must be used.

Type	Taps	Transitional bandwidth [Hz]	Stop-band rejection [dB]	Ultimate Rejection [dB]	Reconstruction resolution [dB]
NPR	512	300	110	120	90
PR	512	350	85	90	145

Table 5. Properties of the two alternative filters in Coherent Acoustics, perfect (PR) and non-perfect reconstruction (NPR)

Sub-band Adaptive Differential PCM (Sub-band ADPCM)

Adaptive differential coding, or ADPCM, is the second main computational process of Coherent Acoustics and, in conjunction with sub-band filtering, forms the second level of de-correlation by operating on the sample-to-sample correlation within each sub-band. The net result is that sub-band signals are de-correlated by transforming them to time-domain difference signals (figure 7).

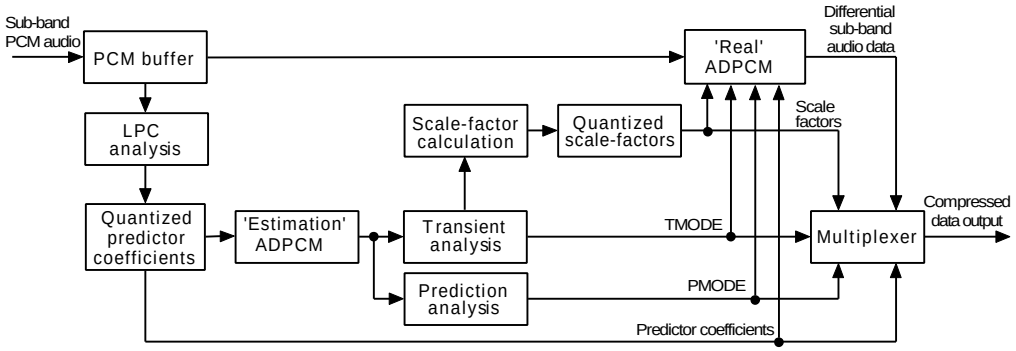
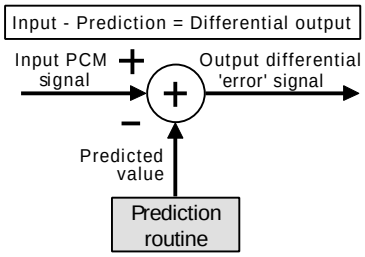


Figure 7. Encoding ADPCM routines and calculation of side-information in Coherent Acoustics

Overview of ADPCM

The encoder ADPCM process (figure 8) involves subtracting a predicted value from the input PCM signal. This leaves a residual value, which is the error or difference between the predicted and actual input values. The difference signal is then re-quantized, and sent to the decoder. At the decoder (figure 9), the predicted signal, that was removed at the encoder, can be exactly regenerated and added back to the residual signal, thereby recreating the original input signal.



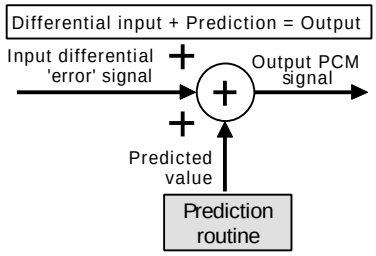


Figure 8. Encoder ADPCM process

Figure 9. Decoder ADPCM process

If the input signal is highly correlated, a good prediction can be made of the value of each sample. On subtraction of the prediction from the input, small errors are generated which are significantly smaller than the input signal, and hence can be more efficiently re-quantized than the original PCM signal. Conversely, with a random, noisy signal, only a poor prediction of the input can be made, leading to a large error signal. This error may be of comparable magnitude to the original input signal, and therefore cannot be re-quantized more efficiently than the original signal.

Since ADPCM is only effective for correlated audio signals, in Coherent Acoustics the process can be switched on or off in each of the 32 core sub-bands independently, as signal conditions dictate. An estimate of the likely coding gain from the ADPCM process is made in each sub-band and, if this gain is too low, the ADPCM may be switched off and adaptive PCM used instead (APCM).

ADPCM in Coherent Acoustics

Figure 10 illustrates the main functional features of the sub-band ADPCM routines employed at the encoder. In Coherent Acoustics, the audio data passes through two ADPCM coding loops. The first pass is an 'estimation' loop, the results of which are used to calculate all the necessary side-information that accompanies the actual audio data. With the side-information, generated from the 'estimation' coding loop,

the 'real' ADPCM coding process finally occurs, and the coded data produced is then packed with its corresponding side-information in the correct transmission syntax by the multiplexer.

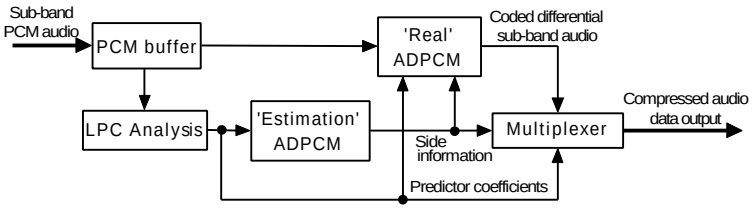


Figure 10. Encoding ADPCM routines in Coherent Acoustics

Fourth-order forward-adaptive linear predictive coding (LPC analysis) is used to generate the prediction coefficients used in the ADPCM routine in each sub-band. With forward adaptation, the prediction coefficients are generated at the encoder by analyzing current samples of audio. In order to reconstruct the predicted signal at the decoder, these predictor coefficients must be explicitly transmitted to the decoder, along with the coded differential sub-band audio data and other side information (figure11).

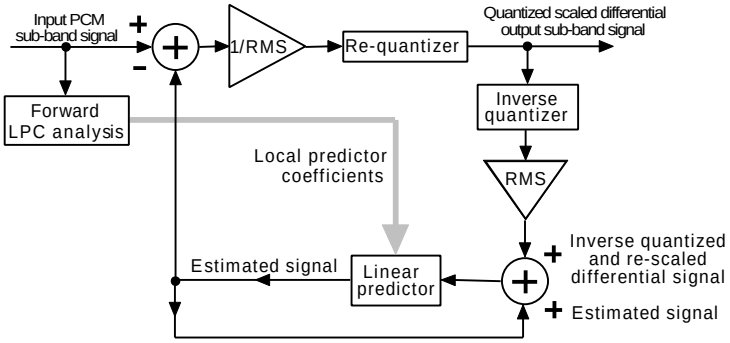


Figure 11. Generating predictor coefficients for encoder ADPCM using forward LPC analysis

The transmission of the sub-band prediction coefficients to the decoder constitutes a significant coding overhead, which can be reduced by disabling the ADPCM process in individual sub-bands. Prediction coefficients then are only transmitted for those sub-bands in which the ADPCM process is active. The decision to disable the ADPCM process is determined by estimating the prediction gain in each sub-band. If the prediction gain is too small or negative then the 'real' ADPCM process is disabled, and the prediction coefficients for that particular sub-band are not transmitted. This reduces the coding overhead. The prediction gain in each sub-band is calculated in the first pass 'estimation' ADPCM coding loop.

LPC analysis: predicting the input signal

The size of the LPC analysis window is variable. Large analysis windows, extending over more audio samples, normally generate a more accurate prediction than a small analysis window. However, the complexity of the prediction process increases with the analysis window size, and may be limited in order to restrict the computational burden at the encoder.

In Coherent Acoustics, the optimal predictor coefficients are calculated over an analysis window of either 8, 16, or 32 sub-band PCM samples, depending on the input PCM analysis frame size (table 6).

PCM analysis buffer size	256	512	1024	2048	4096
LPC analysis window size	8	16	32	32	32

Table 6. Linear Predictive Coding (LPC) analysis window size in Coherent Acoustics

Prediction analysis: enabling and disabling the sub-band ADPCM process

Once the optimal prediction coefficients for each sub-band have been determined (assuming zero quantization error), the difference signal in each sub-band can be estimated. Using the difference signal the prediction gain can be calculated over each analysis window by comparing the variance of the difference signal to that of the sub-band signal. This estimated prediction gain may be larger than the actual gain in the 'real' coding loop. Any reduction in gain must be taken into account when deciding to enable or disable the ADPCM process in each sub-band. The first source of loss is due to the transmission of the prediction coefficients and other side-information, which constitutes an overhead in comparison to PCM. The second source arises from the use of optimal predictor coefficient values in the 'estimation' ADPCM routine, which tends to overestimate the prediction gain. The re-quantized ADPCM values that are actually transmitted to the decoder are sub-optimal, and result in a lower useable prediction gain.

If the estimated prediction gain is not sufficiently large, the ADPCM process in that sub-band is disabled by setting the prediction coefficients to zero. The use, or otherwise, of the predictors in any sub-band is indicated directly to the decoder via ‘predictor mode’ (PMODE) flags embedded in the data stream. Prediction coefficients are not transmitted for any sub-band analysis window for which the predictor mode flag is off.

In this way, the prediction process is dynamically activated in any sub-band if the accuracy of the prediction is deemed to provide a real reduction in quantization noise, at a given bit rate, over the period of that sub-band’s analysis window. For those sub-bands which do not exhibit a prediction gain, the estimated difference signal samples are overwritten by the original sub-band PCM samples for those bands.

Transient analysis and scale factor calculation

Scale factors affect the re-quantization noise within the ADPCM process, and are calculated by averaging the energy of the difference signal over the estimation analysis window. Normally, the energy of the difference signal will not fluctuate widely over the period of one analysis window, and a single scale factor value is sufficiently accurate for the all the samples in the window.

However, transient signals, which can be defined as signals which transition rapidly between small and large amplitudes, artificially boost the scale factor, and hence the quantization noise, for the entire analysis window. If the transient occurs at the end of the analysis window, the increase in noise may be audible during the quiet passage prior to the transient, leading to the phenomenon known as pre-echo. During transient conditions, therefore, it is necessary to calculate and use two scale factors. One small scale-factor value is calculated that applies to the low-level signal immediately preceding the transient, and one larger value that applies to the high-level signal during and immediately after the transient.

The detection and localization of transient signals in each sub-band is determined from the estimated difference signal, or from the original PCM samples for those sub-bands where ADPCM is disabled. On locating a transient, one of four ‘transient modes’ is assigned to the analysis window, indicating the number of scale factors required, and their position in the analysis window. A TMODE of zero indicates either the absence of a transient signal in the sub-band or the presence of a transient at the beginning of the sub-band. In either case, a single scale-factor value can be used for the entire analysis window. A TMODE of one, two or three indicates the presence of a transient at some other location, and two scale factors must be calculated. The process is illustrated in figure 12.

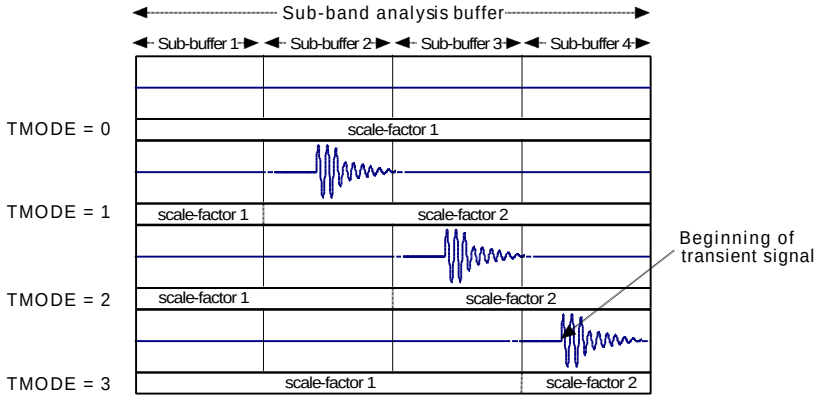


Figure 12. Transient analysis within sub-buffers and the generation of scale factors

Scale factors are calculated for each sub-buffer grouping of eight differential samples within each transient analysis window, based on either the peak amplitude or the RMS of the samples within the sub-buffers. For example, for a coding frame of 4096 audio samples per channel, there are 128 samples in each of the 32 sub-bands, and these are divided into four sub-buffers, each with 32 samples. Transient detection is conducted over four analysis windows, each with 32 differential sub-band samples. Each 32-sample analysis window is split into four sub-buffers of eight samples, and the scale factors calculated for the various combinations or groups of sub-buffers. A 2-bit TMODE flag is embedded in the bit stream for each analysis window, for every sub-band. These flags are used directly by the decoder to unpack correctly the embedded scale factors. With this method, the effective time resolution for scale factor transmission is 5.3 ms at a sampling rate of 48 kHz.

Depending on the application, scale factors are quantized logarithmically using either a 64-level table (2.2 dB per step), or a 128-level table (1.1 dB per step), allowing for the coding of sub-band samples over a 139 dB dynamic range. Scale factors for each sub-band of each audio channel are transmitted directly to the decoder and converted back to the linear domain using a simple look-up table at the decoder. The choice of quantization tables is also embedded in the bit stream for each analysis frame.

The use of dynamic transient analysis, in combination with the scale-factor generation process within each sub-band, greatly reduces the audibility of pre-echo artifacts at low bit rates, and requires only a minimal increase in the side information. Moreover, since the transient positional data is calculated exclusively in the encoder, and conveyed directly in the bit stream to the decoder, future improvements in the detection and analysis of transients will benefit all decoders.

Global bit management

The global bit management routine distributes coding bits from a common bit pool to each sub-band ADPCM process in each audio channel. The number of bits allocated to a sub-band determines the level of quantization noise, which in turn determines the overall quality of the coded signal. The maximum allowed level of quantization noise in each sub-band varies dynamically with time, frequency and audio channel, and its accurate determination is critical to the success of the coder. Depending on the application, a variety of techniques can be usefully employed to calculate the quantization noise. These vary from sophisticated calculations based on an adaptive psychoacoustic analysis of the audio signals, to more simple calculations based on constant, pre-determined thresholds of noise.

Coherent Acoustics employs a number of strategies for allocating coding bits, depending on the bit rate of the application. The basic routine uses a psychoacoustic model of the human auditory system to determine the minimum number of bits needed in each sub-band to achieve coding transparency. For high-bit-rate applications, the psychoacoustic routine can be modified to allow a higher degree of coding margin at low frequencies. Ultimately, for lossless coding, the psychoacoustic routine is ignored.

In some compression systems, the bit-allocation routines are part of the decoder algorithm, and therefore must be kept relatively simple. In Coherent Acoustics these routines only reside in the encoder, and can be of unlimited complexity. Furthermore, since the bit-allocation information is sent directly to the decoder, the global bit-allocation routines can be continuously improved and still remain compatible with all installed decoders.

Psychoacoustic analysis

Psychoacoustic analysis attempts to identify perceptually irrelevant components of an audio signal, which cannot be heard by the human ear. This may be due to the phenomena of auditory masking, whereby loud signals tend to dominate and mask out quiet signals, or due to the insensitivity of the ear to low-level audio signals at particular frequencies.

Psychoacoustic analysis operates in the frequency domain and calculates the minimum acceptable signal-to-masking ratio (SMR) for every frequency component of the signal. In Coherent Acoustics, this translates to calculating the maximum permissible level of quantization noise in each sub-band. Masking is generated by the signal itself, and is cumulative in nature. From figure 13, which shows the approximate mask generated by a single tone, it can be seen that masking occurs close to the masking signal, and that frequencies above the masker are more easily masked than frequencies below. In addition to the masking calculation, each frequency component of the signal is compared to the spectral sensitivity curve of the ear, and may be discarded if it is deemed to be below the threshold of hearing (figure 14).

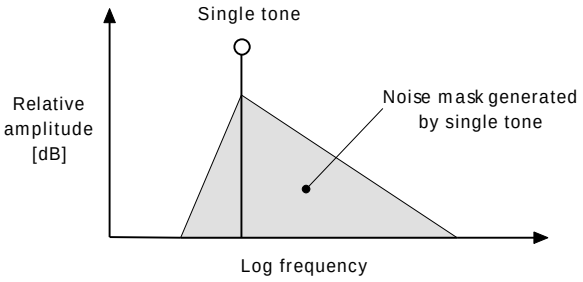


Figure 13. Psychoacoustic calculation of the approximate masking potential of a single tone

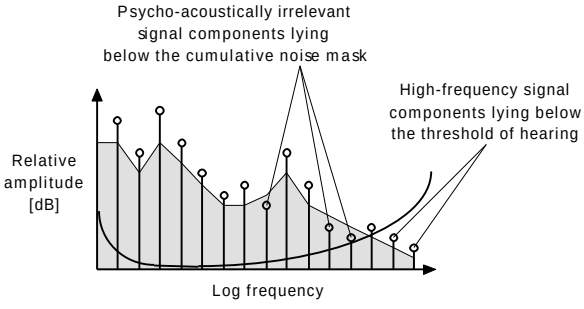


Figure 14. Determination of the irrelevant frequency components of an audio signal

Traditionally, the SMR is used to allocate the bits required for adaptive PCM encoding within each sub-band. However, when used within an ADPCM framework, the bit-allocation routine is modified to account for any prediction gain obtained in each sub-band, and appropriate care taken to ensure that sub-bands are not determined to be irrelevant erroneously.

For example, the basic psychoacoustic analysis might determine that a particular sub-band only requires a signal-to-mask ratio of 30dB, while the prediction gain for the same sub-band might be 35dB. Since the differential signal would then lie 5dB below the SMR, traditionally this would result in the signal being declared irrelevant, resulting in a zero bit allocation for the sub-band. However, the original psychoacoustic calculation would then be grossly inaccurate. In general therefore, sub-bands which have a negative SMR before modification receive a zero bit allocation, whilst sub-bands which have a negative SMR after modification receive at least a minimum bit allocation (figure 15).

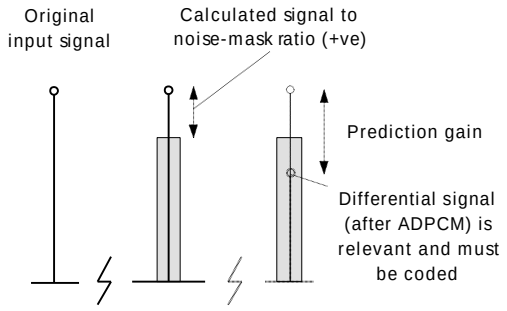


Figure 15. Modification of the bit-allocation routine for differential ADPCM signals that fall below the calculated mask threshold.

Adaptive sub-band bit-allocation

The distribution of coding bits to each sub-band is determined by calculating the maximum permissible level of residual quantization noise in each sub-band, in all coded audio channels. Once the scale factors have been generated in each sub-band, the differential samples are normalized with respect to the scale factors. The quantization noise level is then set by selecting the number of levels to be used by the differential sub-band quantizer, and the number of levels is determined, in turn, by the number of bits allocated to the sub-band. The selection of the quantization step size in each sub-band is the final step in determining the overall precision of the coded audio, and is the link between the bit rate and audio quality of the coder.

Bit-allocation strategies

In low-bit-rate applications, the acceptable, or maximum allowed, levels of quantization noise in each sub-band are determined directly from the psychoacoustically generated signal-to-masking ratios (SMRs), using either the direct or prediction-modified values. If there is an insufficient number of coding bits for this, the bandwidth of the decoded signal can be reduced.

At medium bit rates, the number of coding bits available to each sub-band may exceed the minimum number required, according to the calculated SMR values. This allows more bits to be allocated to particular sub-bands, lowering the quantization noise in those bands. This second stage bit-allocation is based on a minimum-mean-squared-error (MMSE) calculation, which tends to flatten the noise-floor across the frequency domain.

At higher bit rates, the bits allocated to each sub-band continue to accumulate, allowing a further reduction in quantization noise. Often it is desirable to have a flat noise floor across all the sub-bands, since this minimizes the noise power in the time domain. Ultimately, at a sufficiently high bit rate, the noise floor in each sub-band can be reduced until it is equal to that of the source PCM signal, at which point the coder is operating in a lossless mode.

Forward ADPCM

Up to this point, the coding process has relied on an ‘estimation ADPCM’ process to calculate the scale factors and bit-allocation indices. With this side-information and the prediction coefficients, the differential sub-band samples themselves are finally encoded within the 'real ADPCM' loop (figure 16). The resulting quantizer level codes are either sent directly to the bit-stream multiplexer, or, in low-bit-rate applications, further processed by variable-length coding to obtain an additional bit-rate reduction.

To code the differential sub-band signals with the ADPCM process, a choice of up to 28 mid-tread quantizers is available, which have a minimum of zero levels and a maximum of 16,777,216 levels.

Depending on the bit rate of the application, the bit-allocation indices are transmitted directly to the decoder, as either 4-bit or 5-bit words. Use of 4-bit words reduces the side information overheads, but limits the choice of quantizers to 16. This would be appropriate for low-bit-rate applications. The bit-allocation index word length is transmitted to the decoder via a flag in each analysis frame.

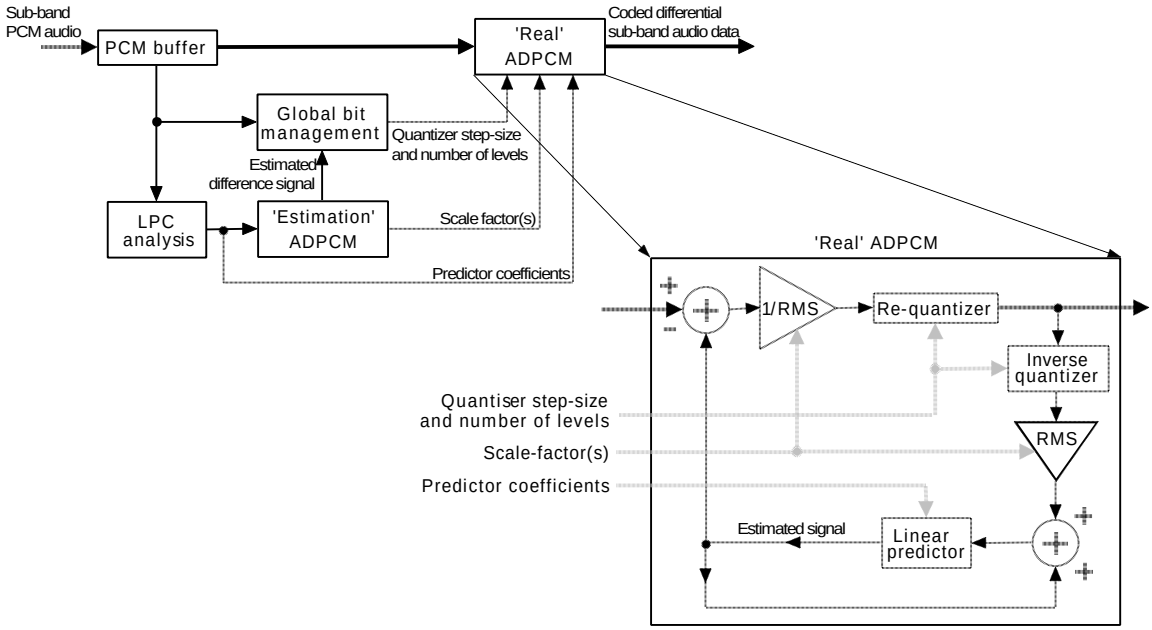


Figure 16. Global bit management within ADPCM

A single bit-allocation index is transmitted for each sub-band differential signal analysis window, and during the same period up to two scale factors can be transmitted.

Variable-length coding of differential sub-band codes

Since the statistical distribution of the differential quantizer codes is significantly non-uniform, a further improvement of up to 20% in the coding efficiency can be realized by mapping the code words to variable-length 'entropy' code-books.

Variable-length coding would normally only be appropriate in low-bit-rate applications. There are two reasons for this. First, at low bit rates the sample analysis windows are of maximum size, and the variance of the frame bit rate is at a minimum. Second, the computational complexity of unpacking variable-length codes is significantly greater than for unpacking fixed-length codes, setting an upper limit to the rate at which data containing variable-length codes can be processed. In general, for Coherent Acoustics, the unpacking computation of bit streams is relatively constant. That is, unpacking high-bit-rate streams containing fixed-length codes is computationally as complex as unpacking low-bit-rate streams containing variable-length codes.

Depending on the size of the linear quantizer, a number of statistically different entropy tables are available for mapping purposes. The codes from the table that produces the lowest bit rate are used to replace the fixed differential codes, and are then sent to the multiplexer (figure 17). Flags indicating which table has been selected are transmitted alongside the codes to facilitate proper decoding. If the bit rate of the variable-length codes is greater than that of the original fixed-length codes, they are not used and the fixed-length codes are transmitted instead.

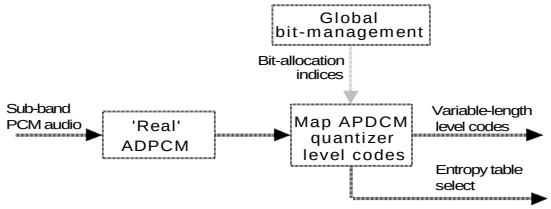


Figure 17. Variable-length coding of differential sub-band level codes

Variable-length coding of side information

The side information consists of the prediction modes, prediction coefficients, transient modes, scale factors, and bit-allocation indices.

In high-bit-rate applications, this side information can be transmitted directly to the decoder without further processing. However, in low-bit-rate applications (below 100 kbit/s/channel), the data rate of the combined side information is a significant portion of the total bit rate, and can begin to limit the quality of the decoded audio. Typically, side-information overheads of approximately 14 kb/s/channel can be expected for audio bit rates in the region from 64 kbit/s/channel to100 kbit/s/channel. This can be reduced to around 11 kbit/s/channel by re-mapping the side information using variable-length codes, in a manner similar to that employed for the differential sub-band codes. By reducing the bit rate of the side information, the bit rate allocated to coding the differential sub-band codes is increased, leading to higher audio quality at low bit rates.

As with the sub-band code words, in terms of decoding complexity, the computation necessary to unpack variable-length side-information codes at low bit rates is similar to that required to unpack the fixed-length codes at higher bit rates.

In order more efficiently to transmit the prediction coefficients to the decoder, each set of four coefficients is quantized using a 4-element, tree-search, 12-bit vector code-book prior to transmission. Each vector quantized (VQ) element consists of a 16-bit integer, creating a 32 kbyte VQ table. For example, in a 4096 PCM sample analysis frame, where the signal is decimated to 32 sub-bands each holding 128 PCM samples, the sub-band predictor coefficients are updated and transmitted to the decoder four times for each incoming frame.

Variable bit-rate control by scale-factor adjustment

When variable-length codes are selected to re-map the differential sub-band audio codes, it is possible that the resultant combined bit rate across all channels momentarily exceeds the maximum bit rate of the transmission channel. In this case, an iterative approach is used to reduce the bit rate of the encoder, whereby certain high-frequency scale factors are incrementally increased, in order to force the entropy mapping process to use progressively smaller code words (figure 18). After each iteration, the ADPCM process and entropy mapping are repeated, and the total bit rate recalculated. In practice, the number of iterations necessary to reduce the bit rate is rarely more than two.

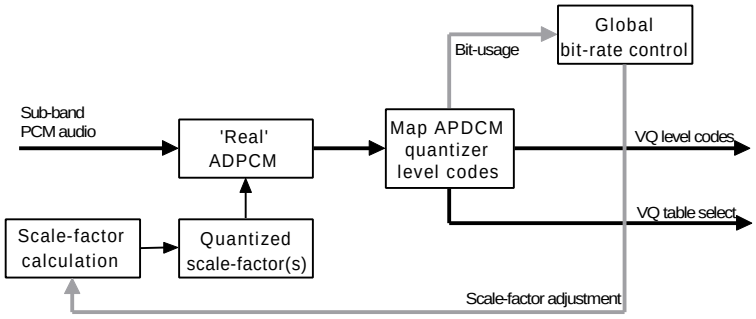


Figure 18. Encode bit-rate control by scale factor adjustment

Low frequency effects channel

A dedicated low-frequency effects audio channel (the LFE channel, commonly referred to as a 0.1 channel) is optionally available with all audio coding modes. The LFE channel is intended to enhance the listening experience by facilitating the reproduction of very low frequencies at very high signal levels, without overloading the main full-bandwidth channels. This is of particular importance in motion picture soundtracks, and is also increasingly used in multichannel music.

The LFE channel is fully independent of all other channels, and is derived by directly decimating a full-bandwidth input PCM data stream at the encoder, using either a 64X or 128X decimation digital filter. These filters exhibit bandwidths of 150 Hz and 80 Hz respectively. The decimated PCM samples are coded using an 8-bit forward-adaptive quantizer. To reconstitute the PCM channel at the decoder, the same filters are used to interpolate back up to the original PCM sample rate.

Bit-stream multiplexer and syntax

The output of the Coherent Acoustics encoder is a serial-data bit stream consisting of the coded audio data. The data is grouped into a sequence of data frames, corresponding to the filter-bank analysis frames. Within each frame the data is ordered (or packed) in a specified format by the multiplexer. Frames can be decoded independently from each other, and as such define the smallest unit of audio that can be decoded. In Coherent Acoustics the minimum decoding unit is 256 samples (or 5.3 ms at a sampling rate of 48 kHz).

A data frame is comprised of five parts (figure 19):

- a synchronization word, which defines the beginning point of each frame
- a frame header, which contains essential information about the encoder configuration
- up to 16 sub-frames, which contain the core 5.1-channel audio coding data
- optional user data, for non-essential data such as time-code information
- extension data, which contains extra audio coding data beyond that required for the 5.1 channel core.

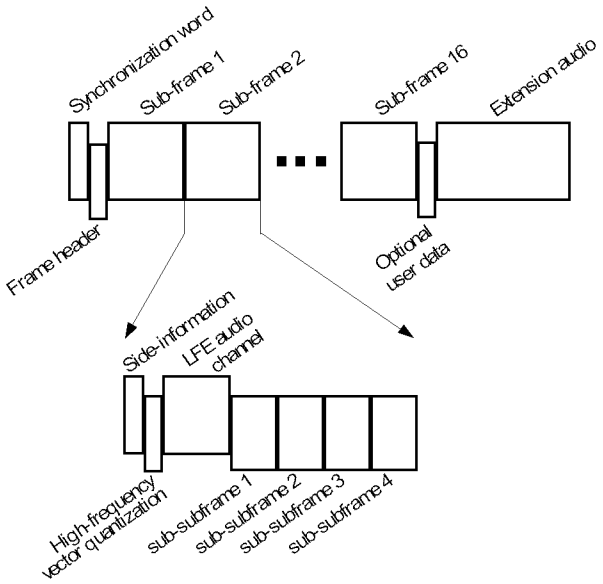


Figure 19. Structure of Coherent Acoustics data frame

An important feature of Coherent Acoustics is the ability to vary the size of an individual frame to help align the decoded audio with an external signal. These shortened termination frames allow the decoded audio to have a resolution of 32 PCM samples, enabling the audio to begin or end at specific times, such as at the beginning or end of particular video frames.

Joint frequency coding

In very low-bit-rate applications requiring two or more audio channels, it is possible to improve the overall sound fidelity by a technique know as joint frequency coding. Experimental evidence suggests that it is difficult to localize mid-to-high-frequency signals (above about 2.5 kHz), and therefore that any stereo imagery is largely dependent on the accurate reproduction of only the low-frequency components of the audio signal.

This can be exploited in multichannel applications by summing the high-frequency sub-bands of two or more of the channels into a single group of high-frequency sub-bands. Essentially this creates a monophonic high-frequency channel which, on reproduction, retains the amplitude information of the individual channels, but contains no phase information. At the decoder, the monophonic high-frequency channel is decoded and reattached to each individual low-frequency channel, thereby reconstituting each of the original full-bandwidth channels.

In Coherent Acoustics, the joining strategy is determined entirely in the encoder, and the joint frequency coding indices are transmitted directly to the decoder. These indicate which sub-bands contain summed signals and which channels have been involved in the summation process. Frequency joining is permissible in all sub-bands except in the first two, which are always coded independently. The joining strategy can be altered and improved at any time, to the benefit of all decoders.

In low-bit-rate, high-quality applications, frequency joining would typically be limited to sub-bands in the region from 10 kHz to 20 kHz. In medium-to-high-bit-rate applications the feature would be disabled altogether.

Embedded decoder functions

Down-mixing format conversion

In order that multichannel formats remain compatible with standard mono or stereo playback systems, the decoder includes fixed functions for down-mixing n channels to any fewer number, such as n-1 channels or n-2 channels. This also includes the ability to down-mix a discrete 5.1-channel soundtrack to a matrixed 2-channel L_T R_T format, compatible with any generic matrix surround decoders.

Dynamic channel mixing

The main disadvantage with using pre-determined down-mixing coefficients in the decoder is their lack of flexibility when dealing with real, dynamic signals. Essentially, there is no single, ideal set of down-mix coefficients that can be successfully applied irrespective of the actual audio signals. This is true even of a simple stereo-to-mono down-mix. For more complex format conversions, such as 5.1-channel to 2.0-channel stereo, fixed down-mix coefficients can produce bizarre results that are highly distracting. This problem is compounded since the use of 5.1 channels in motion pictures and widely varying musical material may require very different conversions to stereo. In general, therefore, for any conversion process using fixed coefficients, the resultant down-mixed signal is artistically compromised in comparison to a studio-generated down-mix using an audio mixing console. In a studio, the mixing engineer has the ability to monitor the quality of the down-mix, and can thus constantly adjust the down-mix coefficients to optimize the output.

In addition to pre-set down-mix coefficients suitable for format conversion, the decoder has to ability to utilize alternative mixing coefficients. These may originate from the bit stream itself, embedded at the encoder, or from a separate serial interface to the decoder. Channel mixing is possible either in the sub-band frequency domain or in the time domain, depending on the complexity of the decoding device. Coefficients for each channel are embedded every frame, and have an effective time resolution of around 10.6 ms, depending on the sampling rate and frame size. This facility allows program providers more flexibility in determining the optimum mix parameters for particular audio material, and alleviates many of the problems associated with using fixed down-mix coefficients for format conversion.

Dynamic range control facility

Dynamic range control seeks to adjust dynamically the output level of the decoded signal, so as to limit the volume of very loud sounds or boost the volume of very quiet sounds. This adjustment is normally used to increase the audibility of the program material to compensate for a poor or restricted listening environment, and as such is designed for specific applications or is under the control of the end user. For example, if the listening environment has a high level of ambient or background noise, such as a car interior, the listener may wish permanently to boost low level signals to make them audible above the noise. However, a more sophisticated system, designed into the player, could continually adjust the dynamic range and volume of the signal to match the changing level of noise in the car.

Whilst dynamic range control is used extensively in the recording, mixing, and mastering stages of audio program production, in all cases the external control of the dynamic range of the decoded signal is an artistic compromise, even if desirable. Furthermore, the actual control algorithm used is specific to the target playback environment, which is either a hardware design issue or under the control of the listener.

In recognition of this, Coherent Acoustics does not include a dynamic range control mechanism in the encoder or decoder algorithms. Rather, it facilitates the operation of a mechanism for controlling the dynamic range, implemented as an external post-process, by embedding dynamic range coefficient values for each audio channel in the multiplexed encoded bit stream. These values are calculated in the encoder, and are embedded in every analysis frame. At the decoder, they may be extracted by a user-defined dynamic range control algorithm, allowing the volume of the decoded audio to vary over a range of +/- 31.75 dB in discrete steps of 0.25 dB.

User data

Optional user data may also be embedded in the compressed audio bit stream at the encoder. This can be extracted from the bit stream by the decoding algorithm and presented to the end user. This data is not intrinsic to the operation of the decoder, but may be used for post-processing routines. One example is time-code data, which may be embedded in each audio frame, and used to align the decoded audio with an external video signal.

Extension data

The use of extension data permits improvements to be made to the capabilities of Coherent Acoustics, and allows the new bit stream containing the extra data to remain compatible with first generation decoders. These improvements may include coding at higher precision, and at higher frequencies. The technique involves generating a core signal from the input audio, and an extension signal that is the difference between the input and the coded core signal. The extension signal therefore includes the additional audio information that is in the input signal, but which has not been coded in the core bit stream. This additional information can be encoded with a variety of methods, generating an extension bit stream that is appended to the main core bit stream. This creates a single compressed audio bit stream, containing the core and extension data, that is compatible with all decoders.

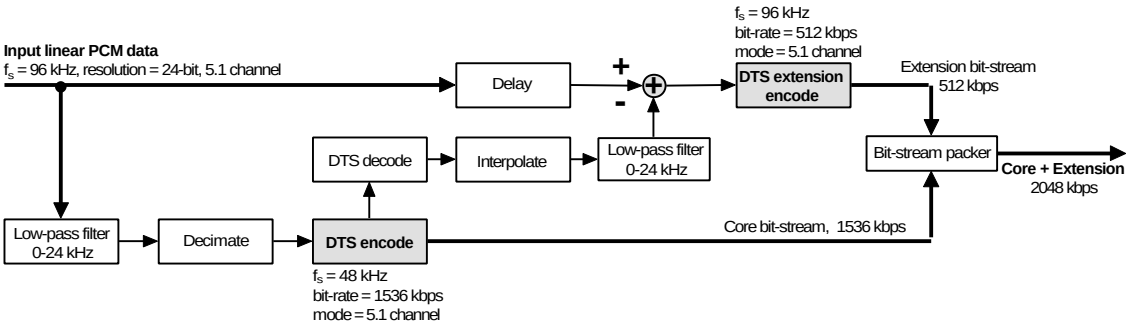


Figure 20. Encoding high sampling rate audio with extension data

Figure 20 outlines the main steps involved when encoding a 96-kHz 24-bit linear PCM signal containing 5.1 channels of audio. The 0 – 48 kHz core signal is generated by low-pass filtering and decimating the full-band input signal. This core signal is then coded normally with 32-band ADPCM, generating the compressed core

audio bit stream. Simultaneously, the encoded core signal is decoded and subtracted from the original full-band input signal. This creates a difference signal that contains the coded error components from 0 - 48 kHz and the original high frequency components from 48 – 96 kHz. This difference signal is also encoded using 32-band ADPCM, generating extension data that is packed with the core data into a single compatible bit stream. A simple decoder will ignore the extension data and decode only the core information. A more sophisticated decoder will decode both the core and extension data, and add them together, recreating the original audio signal (figure 21). An example of this technique applied to an audio signal is shown in figure 22.

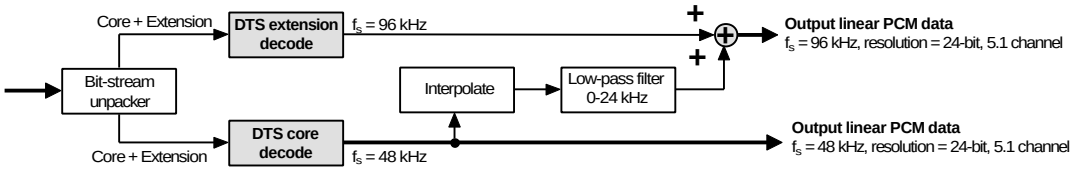


Figure 21. Decoding extension data containing high-sampling-rate audio

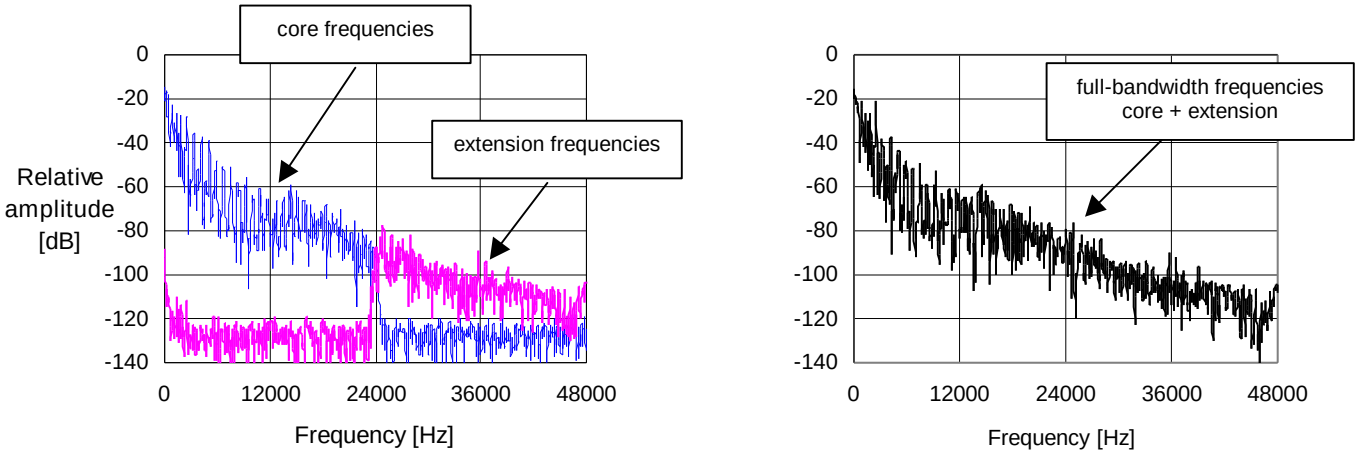


Figure 22. Encoded individual core and extension frequencies and the full-bandwidth composite output for a music signal (5.1 channel, 24 bit, 96 kHz coded at 2.048 Mbit/s)

This technique can also be used for lossless coding, where the extension data is coded with adaptive PCM, creating a scalable bit stream that is fully compatible with first-generation decoders.

Performance of Coherent Acoustics

Subjective performance

Subjective listening tests with experienced listeners still provide the only means of accurately evaluating the sonic performance of any audio reproduction system. For evaluating digital audio data reduction systems, such as Coherent Acoustics, these subjective tests must directly compare an original linear PCM audio signal and the encoded/decoded version of this signal. If done correctly, these comparative listening tests are very difficult, being both time-consuming and expensive. Nevertheless, they are essential. Evaluating the Coherent Acoustics system took place over a period of more than two years, and involved expert listeners from the music and entertainment industry, using custom-built digital audio evaluation equipment installed in professional listening rooms (figure 23). Using digital audio signals sampled at 48 kHz with a resolution of up to 24 bits, these tests consistently demonstrated the transparency of Coherent Acoustics. Irrespective of the audio material, there was no perceived difference between the original linear PCM signal and the coded Coherent Acoustics version. More recent tests have also demonstrated the subjective transparency of Coherent Acoustics at a sampling rate of 96 kHz .

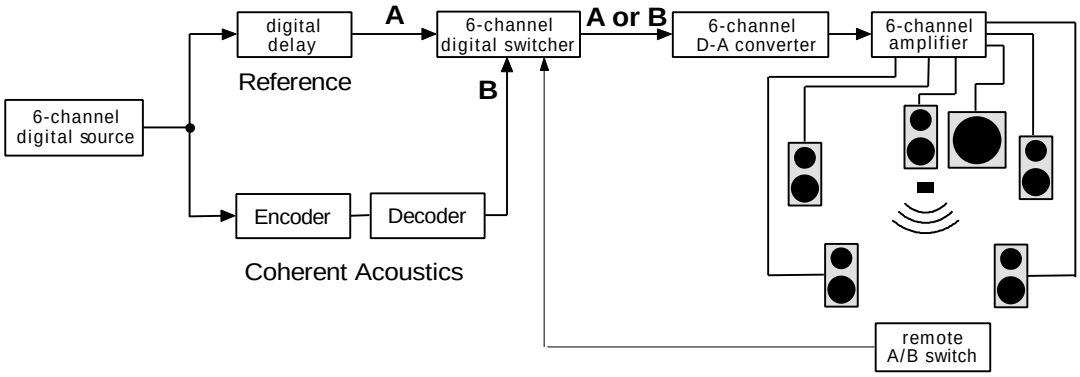


Figure 23. Subjective listening test set-up for evaluating Coherent Acoustics at sampling rates of 48 kHz and 96 kHz

Objective performance

The objective performance of an audio coding system is only relevant if the measurements are strongly indicative of the subjective performance. For digital audio compression systems, the relationship between the measured objective performance and the perceived subjective quality has proven to be very difficult to establish. Most of the standard measurements, such as harmonic distortion, that have been successfully used to determine the quality of analog and linear PCM coding systems, are irrelevant, or even misleading, when applied to audio compression systems.

Nevertheless, some simple test signals can reveal the performance envelope of an audio coding system, and thus provide useful indicators of possible limitations in the underlying coding system. These coding limitations may ultimately be exposed as audible artifacts in subjective listening tests with suitable audio signals. For example, multi-tone test signals can determine the dynamic noise floor and bandwidth of a coding system, while single or dual-tone signals can indicate the dynamic range, and its possible variation with frequency. Multi-tone test signals tend to stress coding systems, and thus give a more realistic measurement of the coded bandwidth than the traditional swept sine-wave test signal.

Figures 24 and 25 illustrate the performance of Coherent Acoustics when coding high-resolution tones sampled at 48 kHz. These indicate a wide dynamic coding range, and the ability of Coherent Acoustics accurately to reproduce 24-bit audio at any signal level over a bandwidth of 20 kHz.

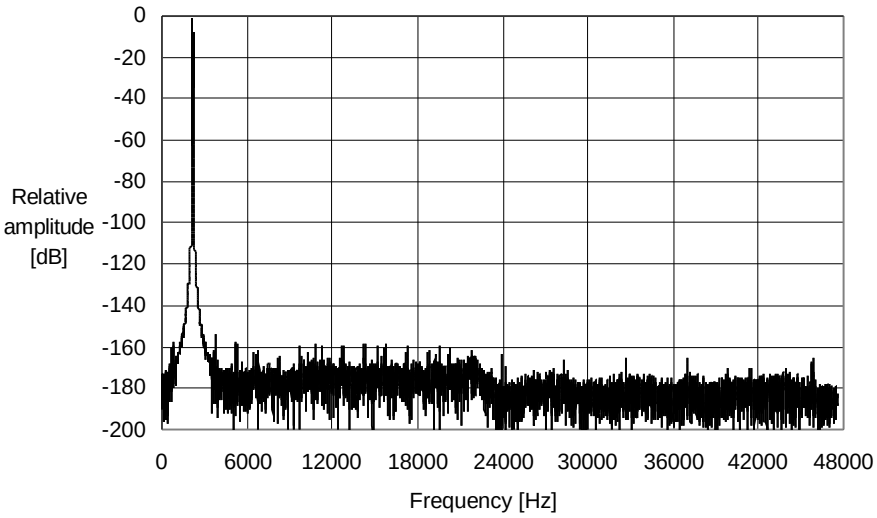


Figure 24. Encoded high resolution 1 kHz tone (24 bits, 48 kHz)

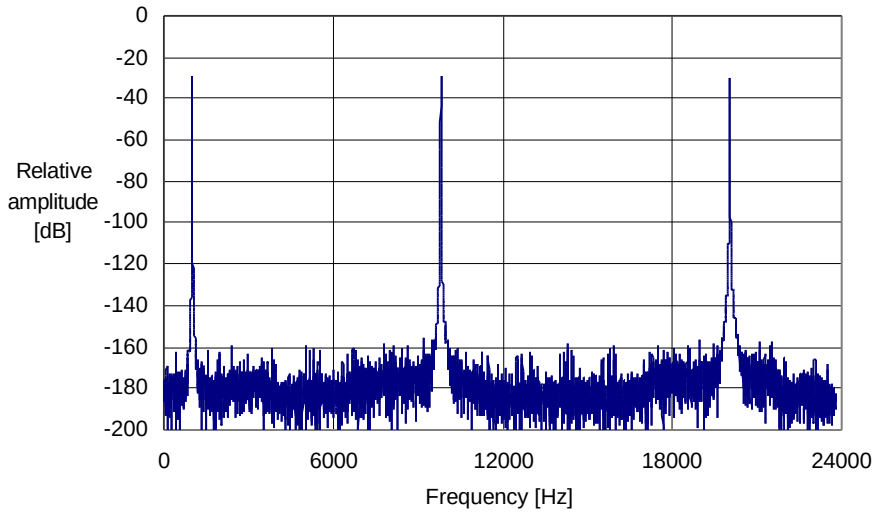


Figure 25. Encoded high resolution tones, 1, 10, 20 kHz (24 bits, 48 kHz)

At a sampling rate of 96 kHz, the core coding algorithm (0-24 kHz) continues to operate with a wide dynamic range, while the high frequency components (24-48 kHz) of the signal are coded at a reduced accuracy (figure 26). With high-resolution single-tone and multi-tone signals, figures 27, 28, and 29 illustrate the 0-48 kHz dynamic bandwidth and the coding noise floor of Coherent Acoustics.

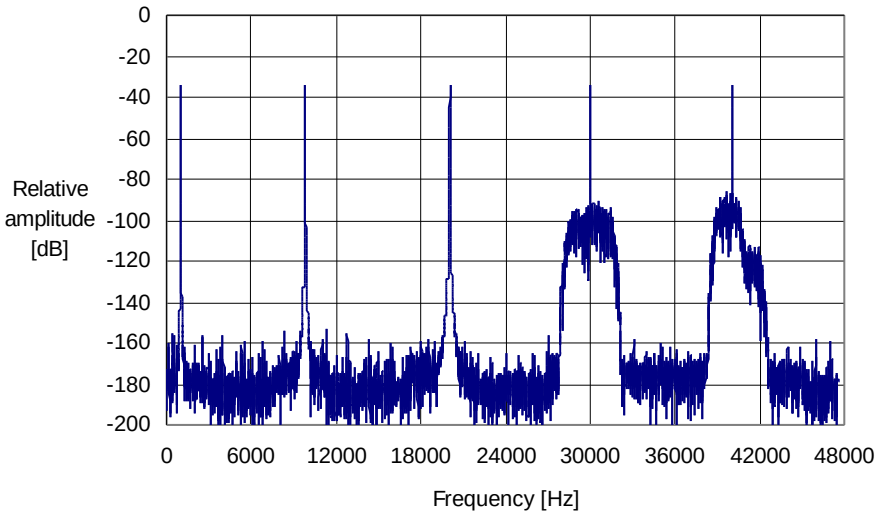


Figure 26. Encoded multi-tone signal 1, 10, 20, 30, 40 kHz (24 bits, 96 kHz)

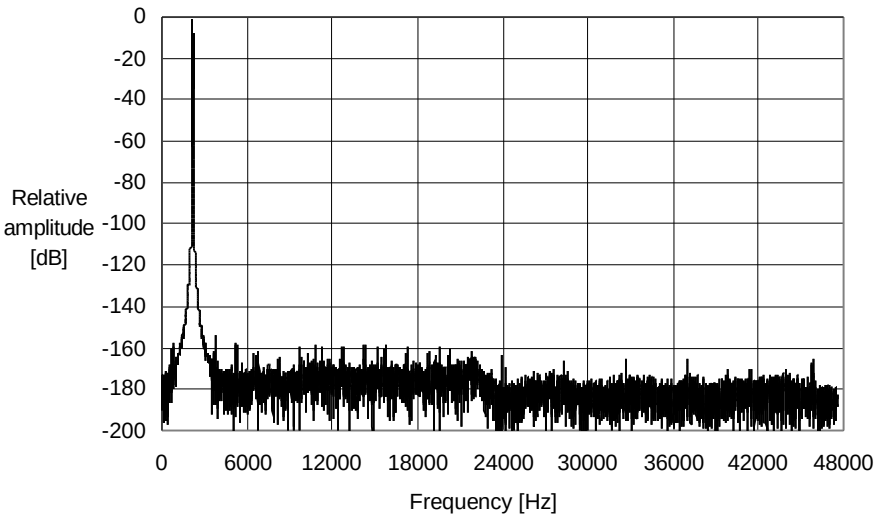


Figure 27. Encoded high resolution 2 kHz tone (24 bits, 96 kHz)

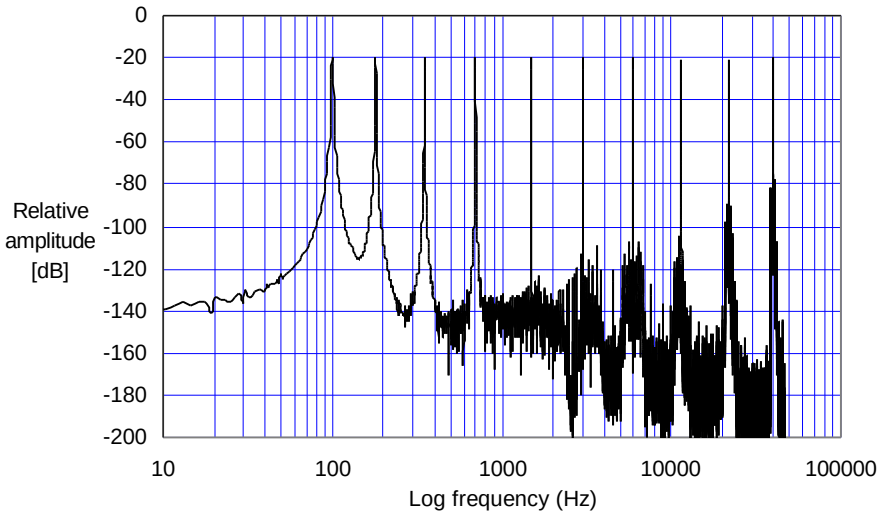


Figure 28. Encoded multi-tone signal, 100 Hz to 40 kHz (24 bits, 96 kHz)

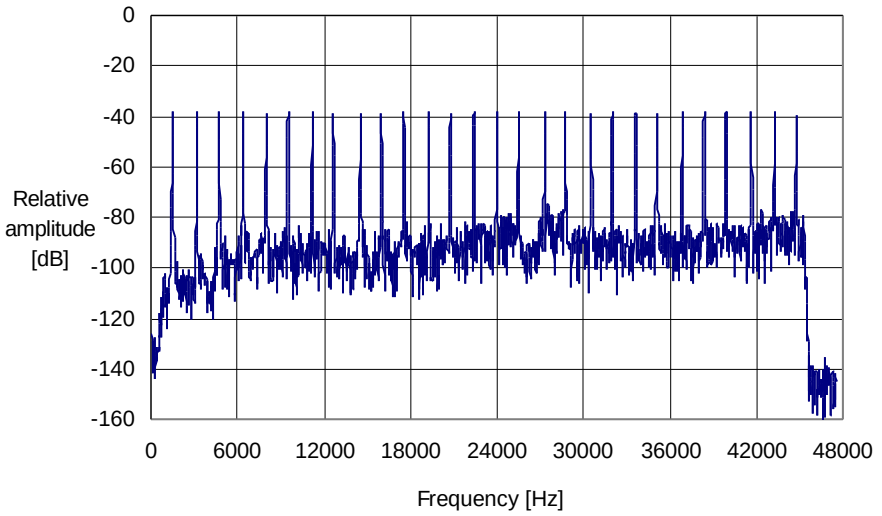


Figure 29. Encoded multi-tone signal (24 bits, 96 kHz)