Waveform Audio Coding Overview

Introduction

Communications form a large part of modern life. Massive amounts of data are carried via telephone networks, TV channels, and private and military data networks. The provision of data-carrying capacity is expensive and using it as efficiently as possible is paramount.

This section discusses the different methods used to encode speech for efficient transmission over physical media. This topic is divided according to the compression method: waveform or parametric. Waveform coders compress the signal by encoding the redundancies of the waveform samples, such as periodicity and slowly-varying intensity. Vocoders compress the signal by assuming a speech production model and coding a set of parameters necessary to reproduce speech. In essence, vocoders emphasize the perceptual quality of speech while waveform coders recreate the signal sample by sample. Parametric coding is discussed in the next section. Unfortunately in order to keep these sections to a reasonable length we shall not be able to look at any forms of error management. Further reading of more in-depth text books, for example Jayant and Noll [1984], will provide information on this topic.

Analog Waveform Coding

In our new digital world of DSPs we still come across some old-fashioned analog waveform encoding techniques. There are three types of analog pulse modulation as shown in Figure 1, amplitude, duration (width), and position. In pulse amplitude modulation (PAM) the amplitude of a train of pulses is modulated in accordance with the message signal. In pulse width modulation (PWM) the widths of the pulses are modulated. Finally, in pulse position modulation (PPM) the position relative to a mean is modulated. In general, the zero amplitude message signal is represented by a non-zero value of pulse to ensure there are no missing pulses. This maintains a constant pulse rate allowing simpler demodulator design.

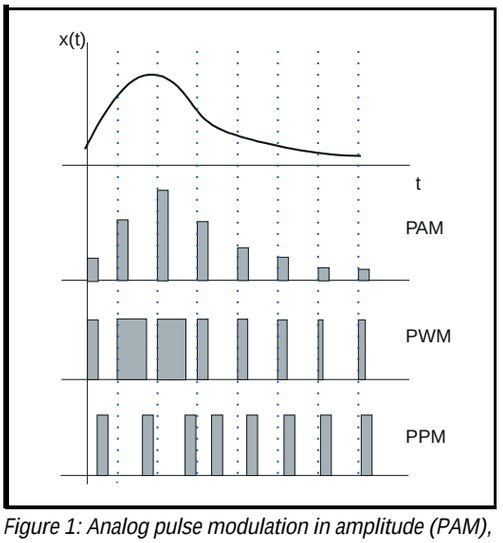

Figure 1: Analog pulse modulation in amplitude (PAM), width (PWM), and position (PPM)

Digital Waveform Coding

Pulse Coded Modulation (PCM)

In digital waveform coding we modulate a train of pulses in accordance with a digital incoming message signal. The most common form of digital waveform coding is known as Pulse Coded Modulation where the output from the coder is a serial train of equally-sized pulses chosen to represent the message signal as efficiently as possible.

PCM is actually a general description for the serial data stream produced by digitizing an analog signal, Figure 2. There are

many variations of the basic PCM scheme which are enhanced for use in specific applications. For example, μ-law and A-law PCM are generally used for telephone systems.

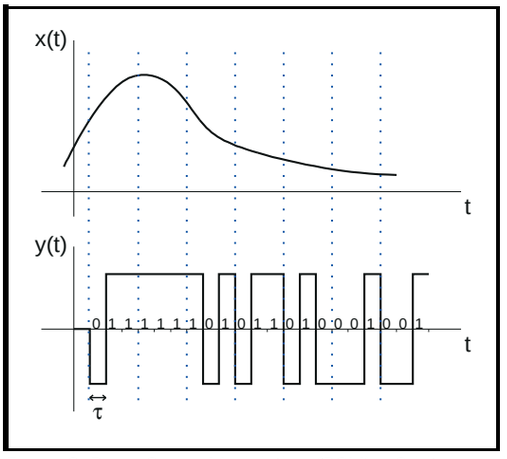

Figure 2: Pulse Coded Modulation (PCM)—3-bit example

Once we have used PCM to encode our signal, we are still left with a baseband signal with significant d.c. and low-frequency components. This means it is unsuitable for transmission over any appreciable distance. Therefore, a PCM output is usually modulated onto a higher frequency carrier wave using some form of digital continuous wave modulation; for example, amplitude shift keying (ASK), phase shift keying (PSK), quadrature amplitude modulation (QAM), etc. (see Carlson [1981], Clarke [1983]).

Generally, in any pulse-coding scheme we must try to make the channel capacity as efficient as possible and also try to make the transmitted signal robust to error. Let's now look at some of the different forms of PCM and examine how they help us increase the amount of information we can transmit.

Delta Modulation (DM)

Delta pulse coded modulation allows us

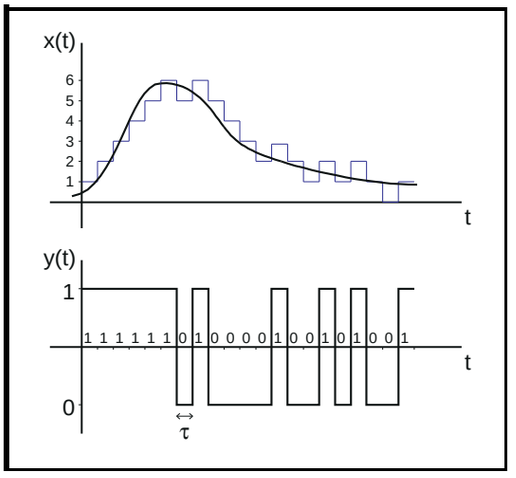

Figure 3: Delta Modulation (DM)

to remove some of the redundant bits in a PCM data stream and is by nature less sensitive to channel errors. The implementation of delta modulation is simple, as we can see in Figure 3.

If the input signal is larger than the accumulated digital value at a sampling point, then the digital value is incremented by one. Equally, if the input is less than the accumulated digital value at a sampling point, the digital value is reduced by one. The output word for each sampling period is just a 1 or a 0 depending whether we increased or decreased our running total. This implies that to transmit the information on the same signal we can use $\frac{1}{n}$ times the frequency for DM that we require for PCM (where n is the number of PCM bits). The block diagram of a delta modulator and demodulator is shown in Figure 4.

Before being input to the delta modulator, the original analog waveform is passed through a sample and hold circuit to produce a discrete waveform as shown in Figure 5. At t=0, assuming no previous inputs, x(n) is quantized to 1 or 0. The feedback ele-

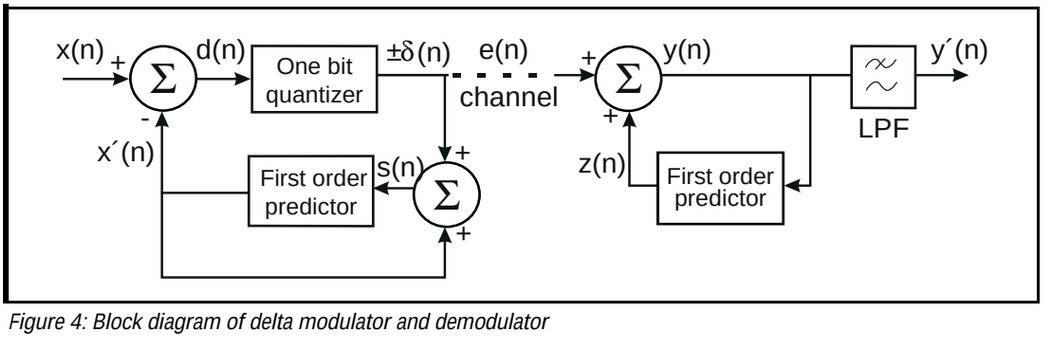

Figure 4: Block diagram of delta modulator and demodulator

ment produces a prediction of the expected next input value. If we assume that it operates as follows:

$x(n) = s(n)$

Then we can work out that the output will be as shown in Figure 6 [assuming $\delta(n) = \pm 1$]. The same predictor is used in the demodulator to reconstruct the signal.

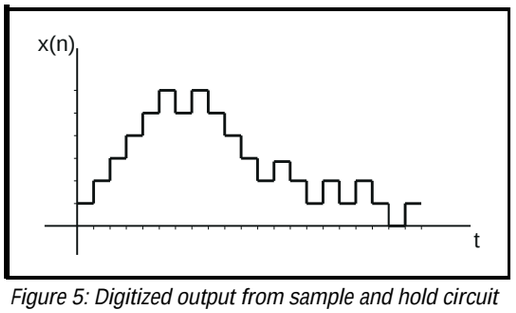

Figure 5: Digitized output from sample and hold circuit

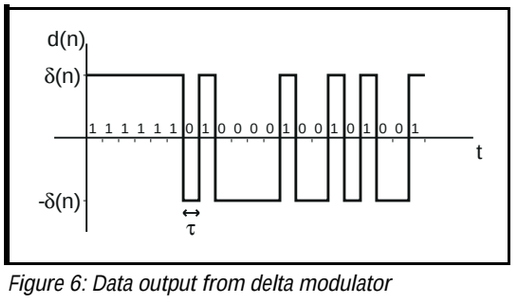

Figure 6: Data output from delta modulator

One of the main advantages of DM can be seen from Figure 6 to be the reduction in the number of transitions between 1 and 0 compared with the PCM example of Figure 2. Our PCM case had the same bit rate (τ), but we need three bits to represent each sample. This means that the effective sample rate is only one third of the bit rate. In the case of DM, we have one bit per sample and so for the same output bit rate we can have three times the number of samples and also less transitions.

Fixed-step-size DM is often referred to as linear DM (LDM). There are two major causes of error in LDM slope overload and granularity. If the step size, $\delta(n)$, is too small, then we get slope overload as shown in Figure 7, and if $\delta(n)$ is too large, we suffer from granularity as shown in Figure 8. It is usually necessary to sample the signal at a faster rate for DM than for PCM to avoid these two problems and much of the perceived advantage is lost.

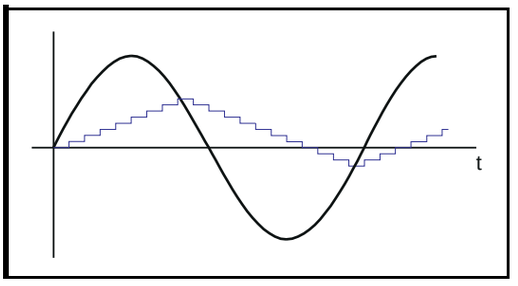

Figure 7: Delta modulation with slope overload due to too small a step size

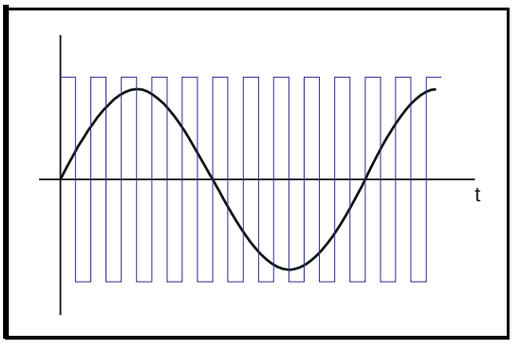

Figure 8: Delta modulation with granularity due to too great a step size

Consequently, a great deal of care is necessary when choosing a value for δ. If we know the characteristics of the incoming waveform, this will not be a problem, but if the waveform is less predictable, we have to consider another approach. An example of a waveform where linear DM is often inappropriate is human speech. There will be some very large variations in amplitude between silent sounds ("ph" or "sh") and loud sounds ("t" or "a"). This may give us a problem with particular words and certainly across the full range of speech. In such cases, adaptive predictors are used which alter the step size depending on a number of previous samples.

Differential PCM (DPCM)

Differential pulse code modulation (DPCM) is an extension of delta modulation which again utilizes the redundancy in analog signals, especially speech and image signals. In differential coding, the difference between a discrete input signal and a predicted value is quantized to one of "p" values, the complexity of the system being

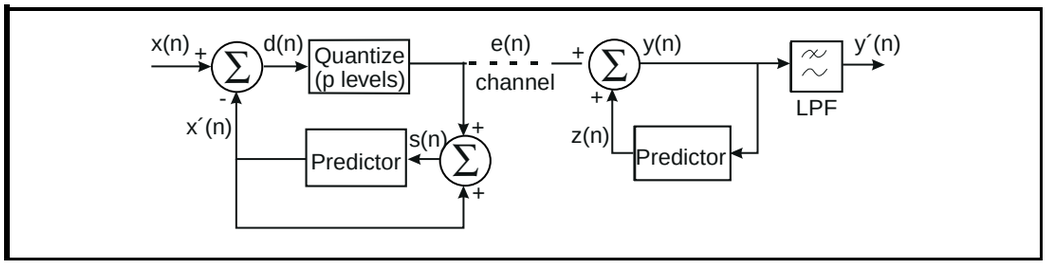

Figure 9: Differential PCM modulator and demodulator

directly related to the complexity of the predictor. A general system is shown in Figure 9.

The DPCM predictor can generally be expressed by the following equation:

$$x'(n) = \sum_{R=1}^{N} h_R S(n-R)$$

Where h_R are the set of predictor coefficients.

Note that if $h_1 = 1$ and all other coefficients are 0, we have our delta modulator.

It is also interesting to note that the general predictor equation is actually the same as the equations used for FIR filters. This is why in many text books you will see the DPCM system drawn with a digital filter in the feedback loop rather than a "predictor." In addition, since the predictor is simply a filter, we know that a DSP will provide an ideal solution.

The advantage of using DPCM over PCM is the gain in signal-to-noise ratio (SNR), i.e., the improved quality of the output waveform. As DPCM is only quantizing the difference between signals and not the absolute value, it will incur far less quantization error despite only using the same number of bits as the PCM system. DPCM is therefore much more accurate for the same system resources.

The relative merits of different order predictors are discussed in depth in Jayant and Noll [1984]. In general, we can say that the gain in SNR saturates at approximately N=2 for DPCM speech and at around N=3 for intra-frame image processing (Habibi [1971]).

Adaptive DPCM (ADPCM)

The quality of the coding algorithm depends upon our knowledge of the signal statistics. If we have a clearly defined input signal, we can design time-invariant (i.e., fixed) predictors. In most cases, although the long-term statistics are well understood, the signal departs significantly from these for shorter periods of time. An adaptive coding scheme will give an advantage where this is true.

The term Adaptive DPCM, or ADPCM, is used as a general title for two different schemes—adaption of the quantizer and adaption of the predictor. Figure 10 shows an example of uniform step-size quantiza-

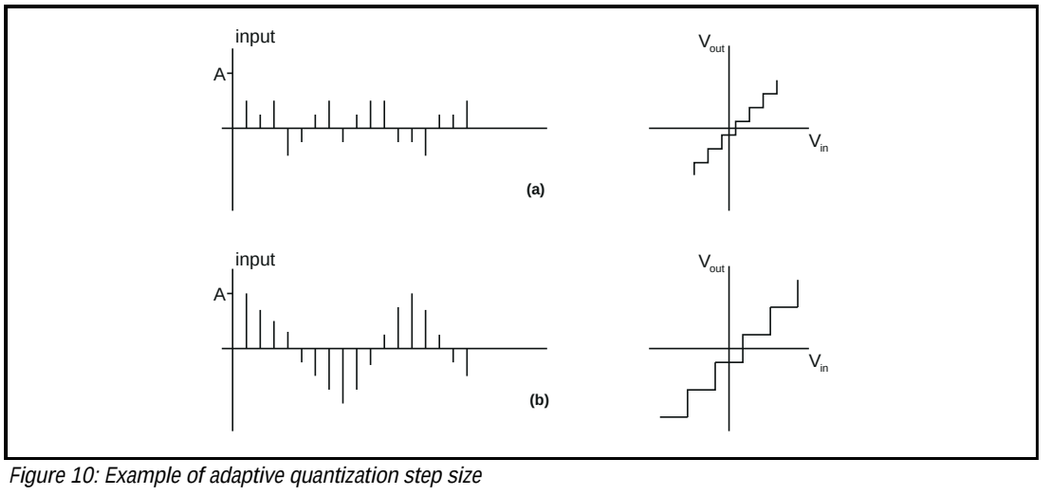

Figure 10: Example of adaptive quantization step size
a) small step size for low amplitude signal
b) large step size for high amplitude signal

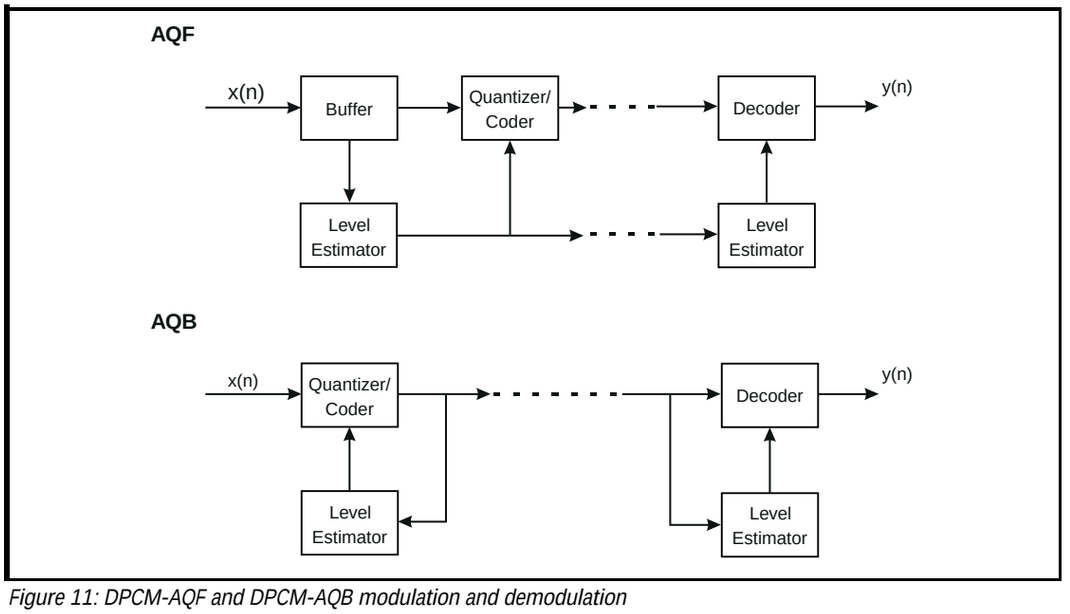

Figure 11: DPCM-AQF and DPCM-AQB modulation and demodulation

tion adaption depending on the input signal amplitude range. Step-size estimation can be performed in one of two ways, forwards or backwards estimation (DPCM-AQF and DPCM-AQB, respectively). Block diagrams for each are shown in Figure 11.

In forward estimation (AQF), the input samples are buffered and used to estimate the level of the input before coding, reducing the effects of quantization noise. AQF suffers from the need to transmit both the coded signal and the level information, for this reason AQB is more common. Another drawback which renders AQF unusable in certain applications is the delay in signal transmission due to coding. For example, if we consider speech to have a bandwidth of 4 kHz and we sample at the Nyquist rate (8k samples / sec), an AQF algorithm using 256 values for level estimation would cause a delay of 16 ms before any output was produced.

Adaptive quantizers offer an improvement in SNR of between 3 dB and 7 dB over non-adaptive schemes, even with 3-bit speech coding algorithms. The quality of the adaption again depends on the quality of the estimator, which could track the mean of the incoming signal or possibility its variance. Again we are faced with a trade-off of complexity against system realization. The performance of the DSP chosen directly affects the sophistication of the estimation scheme used. DPCM-AQF and AQB schemes are suitable for applications with bit rates of around 32 kbit / s.

The most common approach to speech-coding systems is to combine adaptive quantizers and adaptive predictors. With adaptive-prediction systems we can increase the gain in SNR significantly with R=10, compared to R=2 for non-adaptive. It is this form of adaptive DPCM system that we shall from now on refer to by the acronym ADPCM.

CCITT G.721

In ADPCM we have the choice of making the adaption process either feedforward or feedback, and since we are adapting both quantizer and predictor, there are four possible combinations. The CCITT standard G.721 employs feedback adaption of the predictor and feedback adaption of the quantizer (Figure 12).

The input to the encoder is the standard CCITT 64-kbit / s PCM-coded speech. The CCITT standard uses a companding algorithm, either A-law or μ-law depending on whether it is for Europe or the USA respectively. The encoder transforms the PCM digital data stream into ADPCM at 32 kbit / s and hence is usually referred to as a transcoder. The format conversion box in the encoder reconstructs a linear PCM data stream from the incoming companded version, conversely in the decoder this box performs companding on the reassembled signal.

A 4-bit adaptive quantizer is used on the difference signal. Note that as we have feedback adaption of the predictor, the encoder also contains an inverse adaptive quantizer to reconstruct the differential signal values.

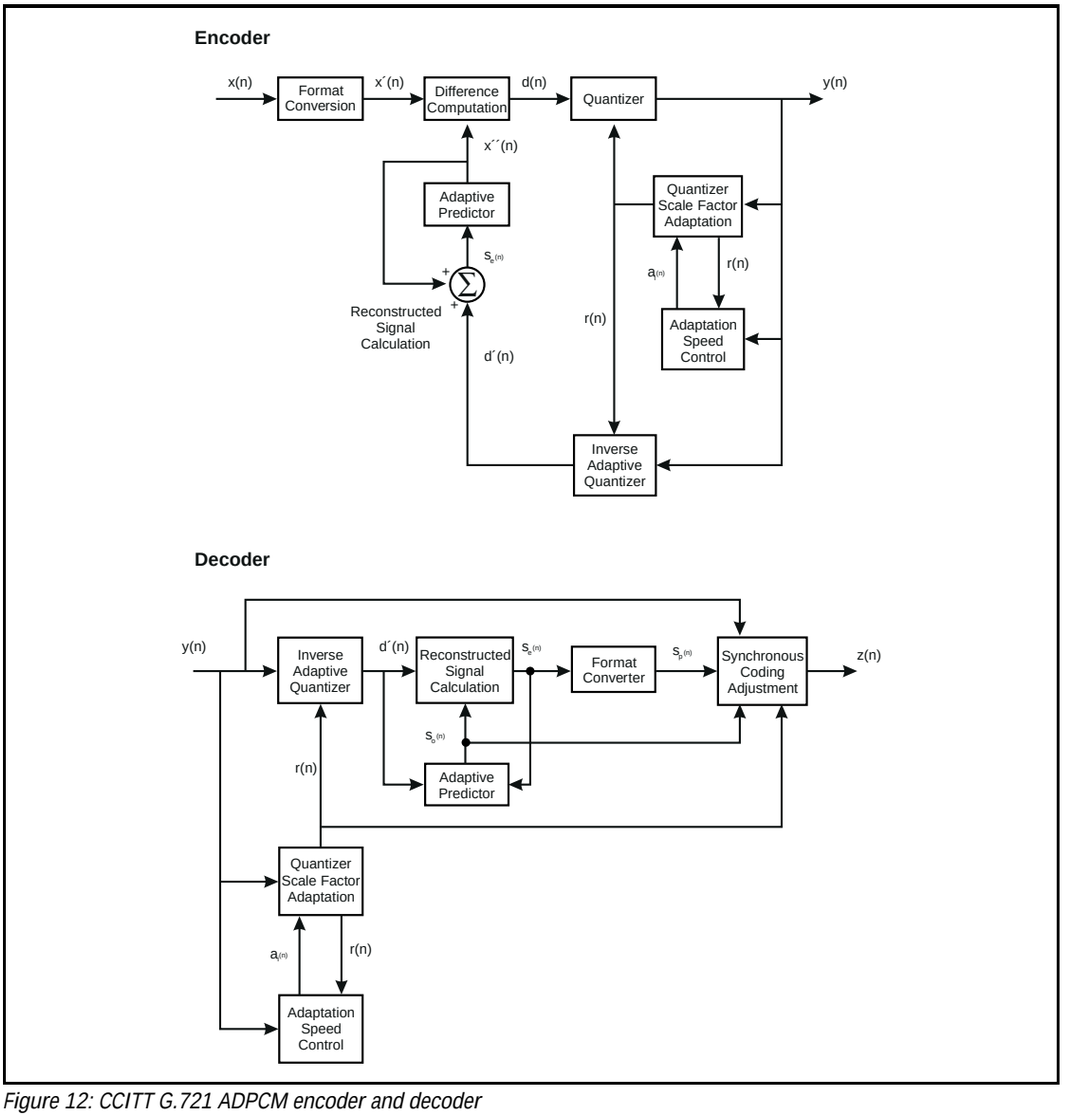

Figure 12: CCITT G.721 ADPCM encoder and decoder

The predictor then uses previous estimated input values, $s_e(n)$ to calculate the next value which is fed back and subtracted from the input to form the differential input signal d(n).

One of the advantages of using feedback predictors is that the same predictor block is used in both the encoder and decoder sections, simplifying the design. The same is true for the adaptive quantizer. Note that there is an extra block included, referred to as adaption speed control, this allows extra control of the quantizer with fast adaption for large amplitude variations (e.g., speech) and slower adaption for signals which vary more slowly (e.g., data). An example of the implementation of a half-duplex G.721 system can be found in Charbonnier et al [1985].

Adaptive Delta Modulation (ADM)

We shall not spend any time on this technique since we have covered the basics in the section on ADPCM. ADM generally uses feedback quantizer adaption to avoid having to transmit the extra level information. It has the advantage that it is simpler to implement than ADPCM.

Continuously Variable Slope Delta Modulation (CVSD)

Although ADM is simple to implement, it suffers significant degradation in speech quality if there are errors in transmission. These errors can propagate through the speech for a considerable amount of time. To recover from these errors, it is necessary to introduce some “leakage” into the predictor and quantizer.

In CVSD, the step size adaption depends on the two previous values of the encoder's output signal, y(n). The step size Δ(n) is given by:

$$\Delta(n) = \beta\Delta(n-1) + D_2$$
$$\text{if } y(n) = y(n-1) + y(n-2)$$

$$\text{or } \Delta(n) = \beta\Delta(n-1) + D_1 \text{ in all other cases}$$

$$\text{Where } 0 < \beta < 1$$
$$\text{and } D_2 < D_1 > 0$$

The values for D_1, D_2, and β, which is the "leakage" coefficient, are related to the required maximum and minimum step sizes by:

$$\Delta_{max} = \frac{D_2}{1-\beta}$$

$$\Delta_{min} = \frac{D_1}{1-\beta}$$

The values for these three coefficients are chosen carefully to ensure that the CVSD coder can follow the incoming message signal and are generally derived from long-term signal statistics. The first set of equations shows that the step size increases when we have continuous runs of either 1 s or 0 s. Looking at Figure 13 we can see that this is exactly what is required when we are suffering from slope overload. In all other cases, the step size decreases. If we choose β to be almost 1, the rate of change of Δ(n) will be slow, conversely if we choose β to be close to 0, adaption will be fast. In general, β is chosen first then D_1 and D_2 are calculated using the Δ_{max} and Δ_{min} equations.

Although CVSD produces poorer quality speech than ADM, APC, or ADPCM, it is much less sensitive to transmission errors. CVSD was quite widely used until the ADPCM schemes became popular and were then standardized by the CCITT.

Adaptive Predictive Coding (APC)

If we concentrate for now on speech coding, the next class of algorithm is adaptive predictive coding (APC). APC is slightly different from the coding schemes we have looked at so far. It considers the speech waveform to be repetitive with a period significantly greater than the average frequency content. The signal is then split into high and low frequency parts and two prediction algorithms are used. The high frequency is estimated using a "spectral" predictor and the low frequency by a "pitch" predictor.

The spectral predictor usually is of 4th

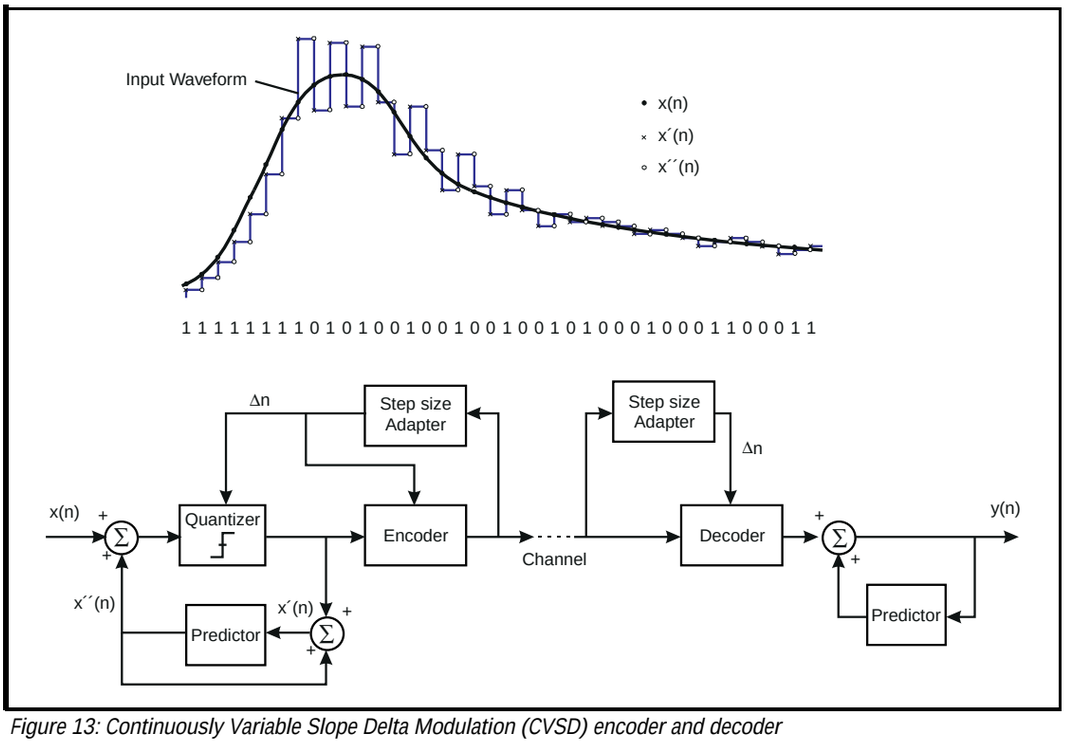

Figure 13: Continuously Variable Slope Delta Modulation (CVSD) encoder and decoder

order, deriving its coefficients from the input signal in the same way as we saw for ADPCM. The input to the pitch predictor is the output of the spectral predictor.

APC has been used for coding speech at 9.6 and 16 kbit/s. Its main drawback is the complexity of the transmitter. This has to estimate the pitch period, calculate the adapted spectral and pitch predictor coefficients, and finally produce an output. Nevertheless there are many implementations of APC as it works well in noisy environments. Further analysis and references can be found in Papamichalis [1987] and an example of the implementation of an APC circuit can be found in Lin (ed.) [1987].

Subband Coding (SBC)

Subband coding is a general term that can use any of the previous PCM-based techniques. Basically the speech waveform is firstly split into subbands using bandpass filters and then each subband is encoded using either ADM, ADPCM, APC, or any other technique (see Figure 14).

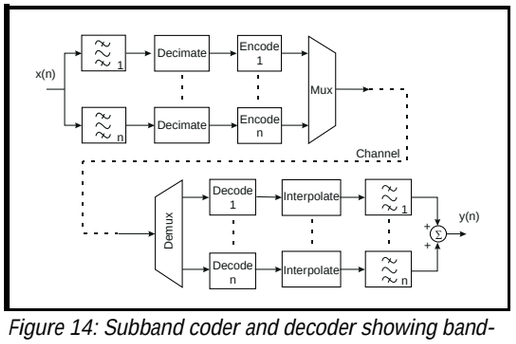

Figure 14: Subband coder and decoder showing bandpass filtering and multiplexing

In splitting the speech into subbands, any quantization noise is kept within that band and cannot interfere with any other subband. Also, the available bits can be allocated between bands according to perceptual criteria, enhancing the speech quality as perceived by the listeners, though not necessarily improving the SNR.

The encoder section includes decimation circuits which allow the control of the number of bits per subband. In SBC, decimation simply refers to throwing away unwanted samples. For example, assuming we have a fixed number of bits to represent the input signal, then, if we have two equally important subbands, half the bits will be assigned to each. When we then sample both subbands at the original sampling frequency, we obtain twice the permitted number of bits and must therefore discard every other sample in each subband. After encoding, the two channels are simply multiplexed for transmission.

At the decoder, interpolation is used to recreate the missing samples before the subbands are filtered and summed to produce the reconstructed signal.

At bit rates of between 9.6 kbit/s and 32 kbit/s, subband coding with APC compares favorably with ADPCM and ADM although the complexity of the system may be higher depending upon the number of subbands. In addition, the design of the bandpass filters must be carefully controlled to avoid interference. Special filters known as quadrature mirror filters (QMF) are generally used (Esteban and Galand [1977]). As with the other coders, an example of the implementation of subband coding can be found in Lin (ed.)[1987].

MPEG Audio Compression

The Motion Pictures Expert Group (MPEG) audio-compression algorithm is a subband coder with adaptive quantization. Figure 15 illustrates a single audio channel of the coder. The encoder divides the incoming audio signal first by frequency into 32 subband channels, then by time into frames of approximately 8 or 24 milliseconds in length. A psychoacoustic model estimates the ear's sensitivity to quantization noise in each frame and subband. Using this information, the quantization level for each is chosen to produce the best subjective sound quality within the total available transmission band-

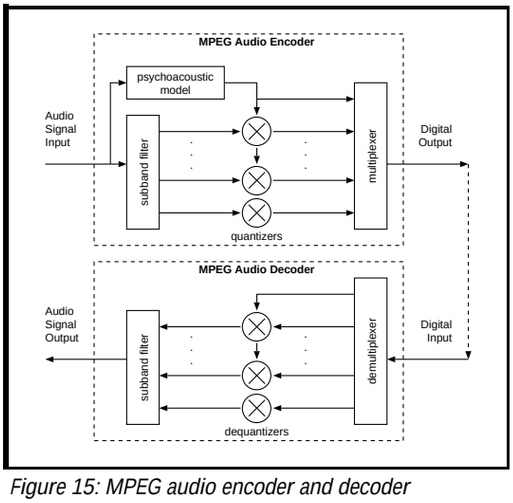

Figure 15: MPEG audio encoder and decoder

width. The quantization information is also sent along with the encoded samples of each subband to the decoder, which performs the inverse operations to reconstruct the signal.

There are three levels of complexity in the MPEG Audio standard, referred to as layers 1, 2, and 3. Each successive layer adds complexity to the encoder and decoder, in return for improved signal quality at higher compression ratios. The highest standard of signal quality is referred to as transparent quality, meaning the that any coding artifacts in the reproduced sound are imperceptible to a trained listener. This algorithm can achieve transparent quality at a compression ratio of 6:1 using layer 2 coding and 4:1 at layer 1.

For stereo audio input signals, a so-called joint-stereo encoding method is available. This method exploits similarity between the left and right channels to further compress the signal. This allows transparent coding at higher compression ratios.

In analog recording, a technique called pre-emphasis is often used. This technique attenuates the low frequencies and enhances the higher frequencies in the audio signal, allowing the perceptually more significant higher frequencies to be recorded with greater dynamic range. Two standard pre-emphasis filters are used in practice, the so-called 50.15 μs and CCITT J.17 types. MPEG audio streams can indicate that the compressed audio samples include pre-emphasis, so that analog or digital circuitry in the decoder unit can apply the appropriate de-emphasis operation.

Each MPEG Audio frame includes a synchronization marker and configuration information. The configuration information describes the settings of the parameters used by the encoder to compress the signal. These parameters include the transmission bit rate and audio sample rate, the number of audio channels, the coding layer, and the type of pre-emphasis used. Using this information, the decoder adapts automatically to the encoder configuration and follows changes in the configuration in realtime.

An optional ancillary data stream can be included in the MPEG Audio stream. This can be used to transmit application-specific information, such as textual information, about the audio stream or user system control data. Space for the ancillary data in each audio frame is taken from the area that would otherwise be used for audio sample data.

References

Carlson, A.B. [1981]. *Communication Systems,* Second Edition. McGraw Hill, New York.

CCITT [1984]. *32 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM),* CCITT Recommendation G.721.

Charbonnier, A., Maitre, X., and Petit, J. P. [1985], "A DSP Implemention of the CCITT 32 kbit/s ADPCM Algorithm," *Proceedings of IEEE International Conference on Communications,* Vol 3, pp. 1197–1201.

Clark, A.P. [1983]. *Principles of Digital Data Transmission,* Pentech Press, Devon, UK.

Esteban, D. and Galand, C. [1977]. Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes, *Proceedings of the 1977 IEEE International Conference on Acoustics, Speech, and Signal Processing,* Hartford, CT, pp. 191–195.

Habibi, A. [1971]. Comparison of Nth - Order DPCM Encoder with Linear Transformations and Block Quantization Techniques, *IEEE Transactions on Communications,* December 1971, pp. 948–956.

ISO-IEC/JTC1 SC29 (MPEG) committee DIS 11172-1, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1.5 Mbit/s."

Jayant, N.S. and Noll, P. [1984]. *Digital Coding of Waveforms,* Prentice Hall, Englewood Cliffs, NJ.

Lin, Kun-Shan [1987]. *Digital Signal Processing with the TMS320 Family, Volume 1,* Prentice Hall, Englewood Cliffs, NJ.

Papamichalis, P.E. [1987]. *Practical Approaches to Speech Coding,* Prentice Hall, Englewood Cliffs, NJ.

Contact:

Excerpted from *A Simple Approach to Digital Signal Processing*
by Craig Marven and Gillian Ewers
Texas Instruments
ISBN 0-04047-00-8