

Continuously Variable Slope Delta Modulation (CVSD) A Tutorial

1. Introduction

Virtually all means of wireless and wireline speech communication are, or are becoming, digital. Digital coding of speech for transmission or storage clearly has advantages over traditional analogue methods. In digital data communication or storage systems, information is transmitted or recorded as a series of binary digits -the receiver or player must only distinguish between a one or zero to exactly recover the original information. In digital voice systems the information is the human voice. Digital speech coding algorithms are judged by their ability to *quantize* (digitize) speech accurately for transmission and then perform the reverse at the decoder. In other words, the original analogue speech signal must be accurately recovered at the receiver. Sounds simple, however, without compression quantized speech can require significantly more bandwidth than analogue speech. For wireline or wireless telecom applications, the speech extends from 300Hz to 3300Hz. If an analogue signal in this band is quantized using a linear Analogue to Digital Converter (ADC) sampling just above the Nyquist rate, say 8kHz, with 256 quantization levels (8 bits) the resulting data rate is 64kbits per second. Without exotic modulation/coding, the bandwidth (BW) for the digitally encoded signal is nearly twenty times the original analogue signal. Such BW consumption is not practical, particularly for wireless applications.

Engineers working in the field of Speech Coding have been actively searching for methods to reduce the bandwidth consumed by quantized speech signals. Early algorithms attempted to take advantage of the human ear's adaptive dynamic range. The human ear has a built in ability to become more or less sensitive to audible signals—the ear can hear sound pressure levels (SPL) as low as 0dB SPL (threshold of human hearing) to 120dB SPL (on-set of pain) yet at any one time the dynamic range of human hearing is generally considered to be about 40dB. In other words, we have a difficult time hearing someone whispering at a rock concert. Some speech coding algorithms exploit this phenomena by using a greater number of progressively smaller quantization levels for low amplitude signals and fewer, more coarse quantization levels for large amplitude signals. This is known as *non-uniform quantization*. Non-uniform quantization is used in the Public Switched Telephone Network (PSTN) where it is called μ -law Pulse Code Modulation (PCM). A slightly more complex approach takes advantage of strong correlation between adjacent speech samples, quantizing the amplitude *difference* (delta) between two samples as opposed to the entire sample amplitude. This *difference* signal requires fewer quantization levels for the same signal quality and consequently, reduces the required bandwidth. Algorithms employing this technique are classified under the broad category of differential quantization or differential PCM (DPCM). Further bandwidth conservation is possible through more complex algorithms. For example, combining adaptive quantization with DPCM results in the commonly used coding algorithm, adaptive DPCM (ADPCM).

Delta Modulation (DM) and Continuously Variable Slope Delta modulation (CVSD) are differential waveform quantization techniques. Both employ two level quantizers (one bit). CVSD is basically DM with an adaptive quantizer. Applying adaptive techniques to a DM quantizer allows for continuous step size adjustment. By adjusting the quantization step size, the coder is able to represent low amplitude signals with greater accuracy (where it is needed) without sacrificing performance on large amplitude signals.

CVSD is used in tactical communications where “communication quality¹” is required yet the option for security must be available. MIL-STD-188-113 (16kb/s and 32kb/s), and Federal Standard 1023 (12kb/s CVSD) are examples of a tactical communication systems using CVSD. With the tremendous worldwide growth in wireless technology, secure communication is becoming important to everyone. In addition to point-to-point communication, CVSD is commonly used in digital voice recording/messaging and audio delay lines.

This paper attempts to describe CVSD quantization, focusing on its application to coding of speech. Before discussing the details of CVSD, the basics of uniform and non-uniform quantization (non-adaptive) will be reviewed. Next, the subject of differential quantization will be explained, showing that DM and CVSD are equivalent to one bit DPCM and ADPCM, respectively. Finally, some application suggestions for CML CVSD codecs will be presented.

2. Waveform Quantization [1, 3, 5, 9]

Waveform quantization is a process of assigning discrete levels to a sampled analogue signal. In the context of this paper the signal of interest is speech, although the concepts are applicable to signals in different frequency bands. The term “quantization” implies a relationship between the amplitude of a discrete sample and its numeric value. This relationship may be linear, non-linear, or differential.

2.1 Sampling [5, 9]

Before a signal can be quantized it must first be sampled. Sampling is the act of instantaneously capturing the level of a continuous signal at some predetermined rate. This predetermined rate is called the *sampling frequency*. As the sampling rate increases, the sampled signal begins to approximate the original continuous signal. As the sampling frequency decreases, samples move further apart in time, eventually the original signal cannot be reconstructed from the sampled version. The limit on how far apart these samples can be, without losing information, is the basis for Shannon's sampling theorem.

Shannon's sampling theorem [5] was originally stated as follows: "If a function $f(t)$ contains no frequencies higher than f cycles per second it is completely determined by giving its ordinates at a series of points spaced $(1/2f)$ seconds apart." A mathematical version of this theorem can be obtained by convolving the Fourier transform of the signal to be sampled with the Fourier transform of an infinite sequence of impulse functions

$$F_S(\omega) = \frac{1}{2\pi} [F(\omega) * S(\omega)] \quad (1)$$

or,

$$F_S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\tau)F(\tau - \omega)d\omega \quad (2)$$

Where, $F_S(\omega)$ = sampled signal spectrum

$F(\omega)$ = original signal spectrum

$S(\omega)$ = spectrum for a sequence of impulses

$\omega = 2\pi f$ (radian frequency).

The result of the convolution in equation (2) is the original signal spectrum repeated at multiples of the sampling frequency, see Figure 1. Notice that if the sampling frequency is less than twice the bandwidth of the original signal the replicas centered at multiples of the sampling frequency will overlap and distort the original. This undesirable phenomenon is *aliasing*. To avoid aliasing,

$$f_S \geq 2B. \quad (3)$$

¹ **Communication quality** is a qualitative expression widely considered synonymous with “acceptable speech communication.” It is not intended to imply “high-fidelity,” only that intelligible conversation can take place.

Where, f_s = sampling frequency in Hz

B = bandwidth of original signal in Hz.

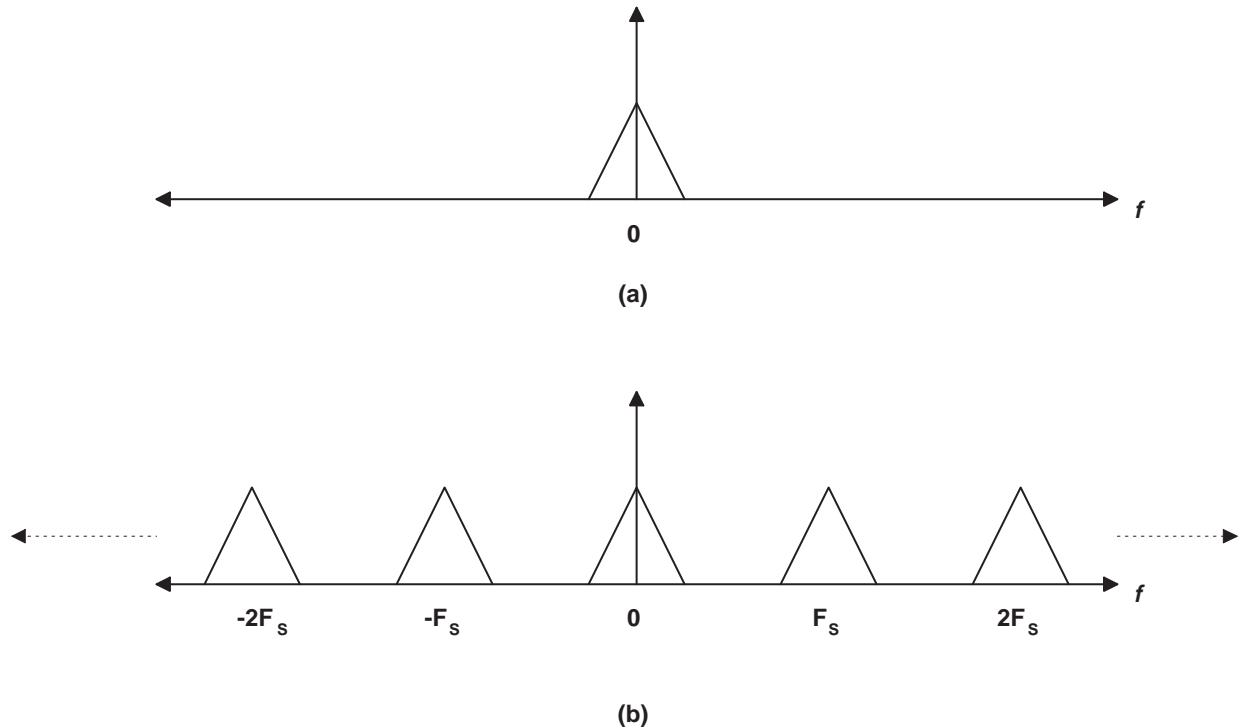


Figure 1: Fourier transform of (a) continuous time signal and (b) sampled signal.

Shannon's sampling theorem is mathematically accurate. Although, in most cases it is not practical to sample a signal at exactly twice its highest frequency. Band-limiting is necessary to avoid aliasing. For $f_s=2B$ the band-limiting filter must have a so called "brick wall" roll-off at frequency B . A filter that matches this requirement is physically unrealizable. Several factors contribute to the actual sampling frequency used. Generally there is a compromise between the complexity of the band-limiting filter versus the cost of the analogue to digital converter (ADC). As f_s becomes closer to $2B$, the band-limiting filter requires more stages to give the desired roll-off. As f_s increases, the required conversion time necessitates a faster ADC. Cost for ADCs is inversely proportional to conversion time, as conversion time decreases, cost rises.

2.2 Uniform Quantization [3, 4, 9]

Figure 2 illustrates the transfer characteristic of a seven level uniform quantizer. To represent the quantized output $q(n)$ as a binary number would require 3 bits. Once a signal has been sampled it is discrete in time. However, the amplitude remains continuous. The quantized version of a signal is obtained by applying the sampled signal to a binary encoder where the discrete voltage levels are assigned to the nearest binary numbers. The complete process, combining sampling and amplitude quantization, is known as PCM.

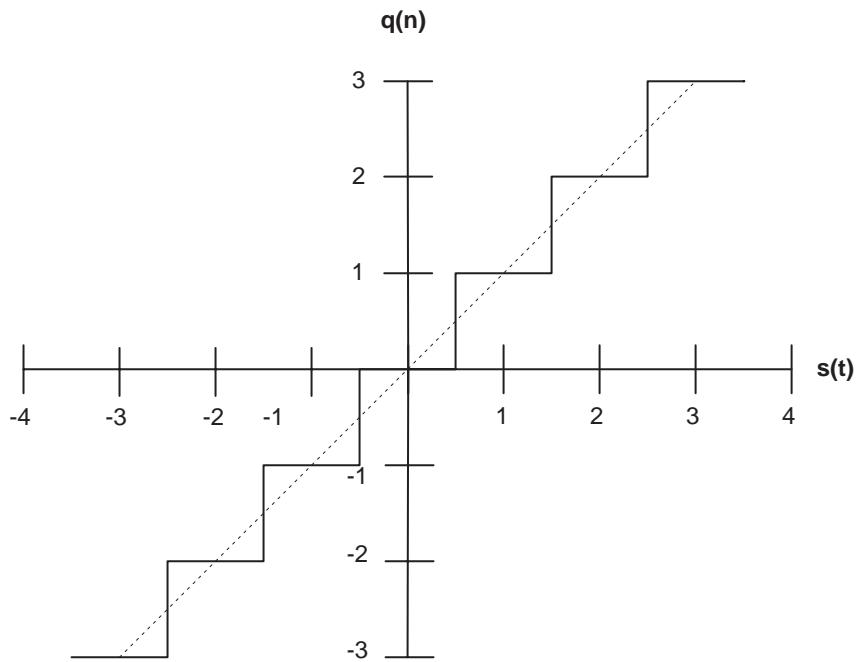


Figure 2: Uniform quantizer transfer characteristic.

The noise introduced by PCM is primarily due to the rounding to the “nearest” binary number. If a signal is quantized to eight levels (3 bits) using a quantizer transfer characteristic similar to the one shown in Figure 2, the finest resolution is the full scale magnitude divided by eight. In general,

$$\frac{A}{2^n} = \alpha. \quad (4)$$

Where, A = full scale amplitude

n = number of bits per sample

α = difference between quantization levels (i.e. finest resolution).

The uncertainty error due to rounding is based on the assumption that there is a continuous range of values within α , all of which are equally likely to have been the actual value of the original signal (uniform distribution). The error signal ε is the difference between the original signal amplitude and the quantized sample value. As a result, the mean square uncertainty error can be solved for by finding the expected value of error squared over the range of $-\alpha/2$ to $\alpha/2$

$$E(\varepsilon^2) = \frac{1}{\alpha} \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} \varepsilon^2 d\varepsilon. \quad (5)$$

Where, E = expected value operator

ε = difference between actual and quantized signal levels.

Performing the integration in equation (5) yields an expression which is equivalent to quantization noise power

$$N_{OUT}^2 = \frac{\alpha^2}{12}. \quad (6)$$

Taking the square root of both sides gives the root mean square (RMS) noise voltage

$$N_{OUT} = \frac{\alpha}{2\sqrt{3}}. \quad (7)$$

The maximum signal-to-noise ratio (SNR) for a uniformly quantized signal can be calculated by finding the ratio of the full scale quantization level to the noise voltage in equation (7). The full scale quantization level is simply the total number of quantization levels multiplied by the minimum quantization increment α

$$S_{OUT_{MAX}} = \frac{\alpha(2^n)}{2}. \quad (8)$$

Where, n = number of bits in quantizer.

The ratio of equations (7) and (8) yields the SNR

$$SNR = \frac{2^n}{\sqrt{3}}. \quad (9)$$

In dB,

$$SNR_{dB} = 4.77 + 6.02n. \quad (10)$$

Equation (10) is an objective measure of quality in systems employing uniform quantization. It should be noted: the SNR calculated in equation (10) is a theoretical maximum. In practice other factors (e.g. power supply noise) tend to reduce the final SNR. Also, the human voice is considered to have about 40dB of dynamic range, however, during most conversations it is typically about 20dB down from maximum. In other words, we generally do not shout during normal conversation. Consequently, the average signal to noise ratio for uniformly quantized speech is about 20dB less than what would be calculated using equation (10).

Another objective measure of quality which can be derived in a similar manner is dynamic range. Dynamic range pertains to the resolution of a quantization scheme. It is the ratio of the full scale amplitude to the smallest quantized amplitude change,

$$DR = \frac{1}{2} \left(\frac{\alpha 2^n}{\alpha} \right) = 2^{n-1}. \quad (11)$$

In dB,

$$DR_{dB} = 6.02(n-1). \quad (12)$$

2.3 Non-uniform Quantization [1, 7, 9]

To this point the discussion has been based strictly on uniform quantization (linear speech coding). That is, waveforms are quantized such that all signal levels have the same resolution. An alternative to linear quantization is to make α fine for low level signals and coarse for high level signals, introducing a non-uniform quantizer characteristic. This type of quantization can result in improved dynamic range for a given number of bits and effectively raise the SNR for lower level signals. The drawback is lower maximum SNR—equations 9 and 10 are not valid for non-uniform quantization. Figure 3 shows a transfer characteristic for a seven level non-uniform quantizer. Notice, when compared to the uniform quantizer transfer characteristic shown in Figure 2, there are three quantization levels for the input $s(t)$ between zero and one, where there is only one with the uniform quantizer.

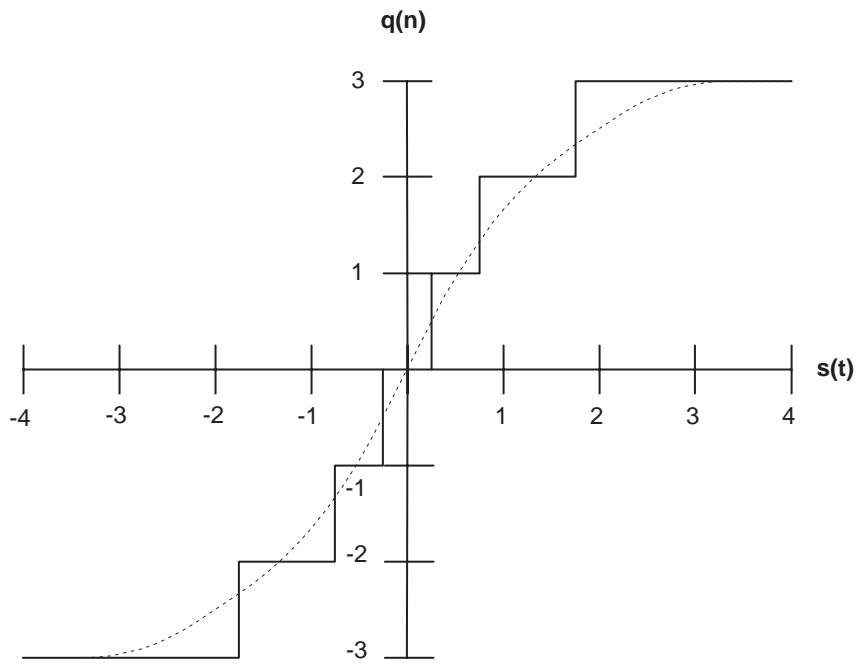


Figure 3: Non-uniform quantizer transfer characteristic.

A complete digital communication system, incorporating non-uniform speech coding, compresses the quantized signal at the transmitter then expands the signal back to linear form at the receiver. The compressed signal requires fewer bits and therefore consumes less BW. Ideally, the expander transfer function is the exact inverse of the compressor and is able to reproduce the original uncoded analogue signal. In literature pertaining to telecommunications, the words *compressor* and *expander* are generally combined in to a single word; *compandor*.

Companding is one of the most popular forms of non-uniform quantization. When compared to uniform quantization, it allows for bandwidth compression without degradation in dynamic range, at the expense of peak signal to noise ratio. For example, a signal uniformly quantized to 14 bits, using equations (10) and (12), would have a peak SNR and DR of approximately 89dB and 78dB respectively. If the same signal is non-uniformly quantized (compressed) using only 8 bits (256 levels), the minimum quantization level can be set such that DR of 78dB can be retained, although the peak SNR will be degraded. Degradation in the peak SNR is due to the coarse quantization levels used for the large amplitude signals (where fine quantization is not necessary). Low amplitude signals are quantized at finer resolution (more steps, where it is necessary). Consequently, the SNR at low signal levels is improved when compared to uniform quantization with the same number of levels. Essentially, companding strives to make SNR constant over the dynamic range of the quantizer.

2.3.1 μ -Law Companding [1, 9]

Most forms of non-uniform quantization are derived from a logarithmic transfer function. That is, the output signal is proportional to the log of the input signal. In North America and Japan, digital telecommunications networks employ μ -law companding as the standard for PCM encoding ($\mu = 255$ for North America). The general form of the transfer characteristic is:

$$y(x) = \text{sign}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} . \quad (13)$$

Where, $-1 \leq x \leq 1$

x = input signal

$y(x)$ = compressed output signal

$\text{sign}(x)$ = polarity of input signal.

2.3.2 Uniform versus Non-uniform Quantization [7]

An accepted method of objectively measuring quality in signal processing systems is SNR. In the context of digital coding this would more accurately be called: signal to quantization noise ratio (SQR). Equation (10) was derived to calculate the SNR for uniform quantization assuming maximum input signal. To find SNR for input signal amplitudes less than maximum equation (10) must be rewritten

$$SNR_{dB} = 4.77 + 6.02n + 20 \log\left(\frac{S_i}{S_{MAX}}\right). \quad (14)$$

Where, S_i = input signal level

S_{MAX} = maximum signal level.

When the quantization characteristic is not uniform (14) is no longer valid. To clarify the performance of companded quantization, the SNR must be calculated for a full range of input signal levels. For μ -law companding the signal to noise ratio is given by the following equation [9],

$$SNR_{\mu(dB)} = 10 \log \left[\frac{3(2^{2n})}{[\ln(1+\mu)]^2 \left[1 + \frac{1.732}{\mu S_i} + \frac{1}{\mu^2 S_i^2} \right]} \right]. \quad (15)$$

It should be noted that as $\mu \rightarrow 0$, equation (15) is equivalent to equation (14), thus $\mu = 0$ implies uniform quantization. Figure 4 compares SNR versus relative input signal power for 256 level non-uniform quantization ($\mu = 255$) and 256 level uniform quantization.

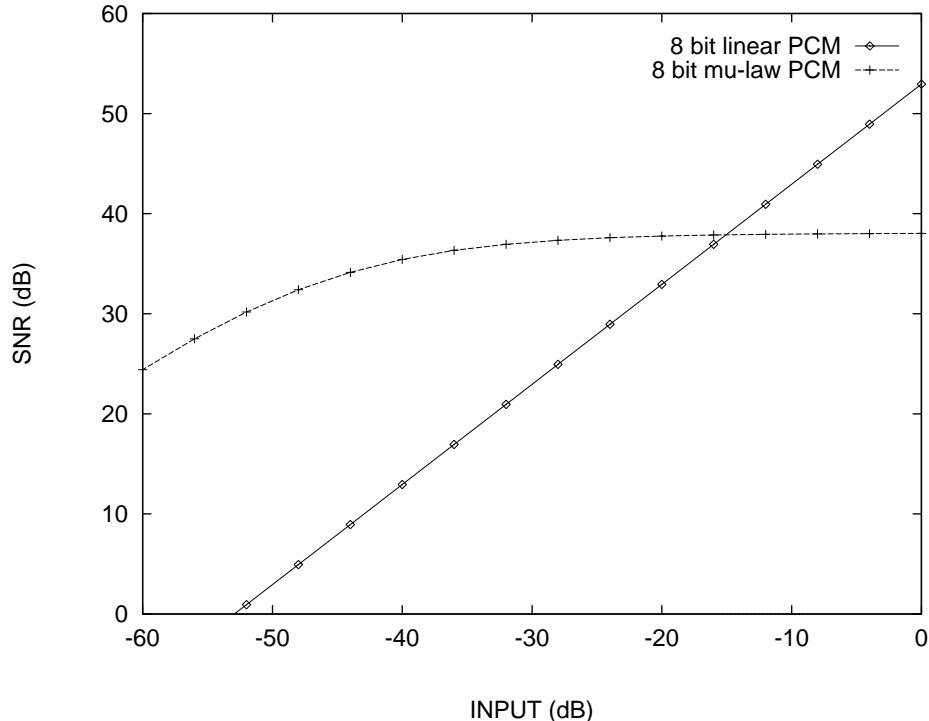


Figure 4: SNR versus input level for 8 bit μ -law and 8 bit uniform quantization.

2.4 Differential Quantization [4, 8]

Differential quantization is a coding technique where the difference between the present sample and the predicted value for the next sample is quantized. This type of quantization is advantageous because the difference signal dynamic range (variance) is less than the unprocessed input signal. Consequently, fewer quantization levels are required to retain the same Signal-to-Noise Ratio (SNR). Alternatively, the number of quantization levels can remain the same allowing the difference signal to be encoded using all the available bandwidth resulting in greater resolution and improved signal quality.

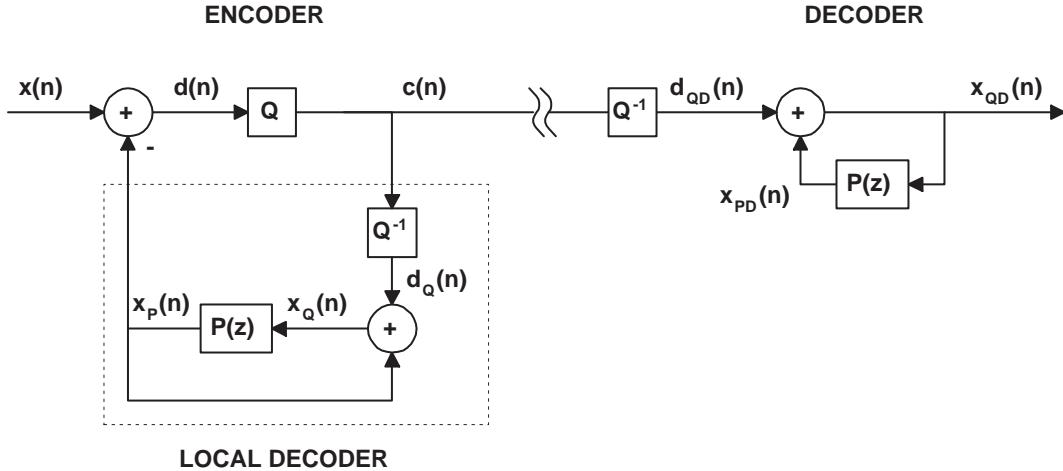


Figure 5: Differential quantization block diagram.

Figure 5 is a block diagram displaying a differential quantization system. All signals are represented in discrete time notation implying that $x(n)$ is the discrete time version of $x(t)$. Notice the decoder is in the feedback path of the encoder. Thus, the decoder is performing the inverse of the encoder. The block labeled Q converts the difference signal $d(n)$ into a binary representation suitable for transmission and the block labeled Q^{-1} does the inverse. In reality, the process of converting $d(n)$ to $c(n)$ and back to $d_{QD}(n)$ is a significant factor in the non-ideal behavior of differential quantization. Nevertheless, fundamental analysis of the system is simplified by assuming blocks Q and Q^{-1} cancel. Employing this assumption, the transfer function of the encoder in z -domain terms

$$H_{ENC}(z) = \frac{C(z)}{X(z)} = Q(z)[1 - P(z)]. \quad (16)$$

And the decoder transfer function is:

$$H_{DEC}(z) = \frac{X_{QD}(z)}{C(z)} = \frac{1}{Q(z)[1 - P(z)]}. \quad (17)$$

If $D_{QD}(z) \equiv D(z)$, then $X_{QD}(z) \equiv X(z)$ and the entire system transfer function can be written as,

$$H(z) = H_{ENC}(z)H_{DEC}(z) = 1. \quad (18)$$

Although equation (18) is based on several assumptions, it shows a differential quantization system patterned after the topology displayed in Figure 5 can be made to produce an output signal which approximates the original input signal. The quality of the approximation is what distinguishes various differential quantization schemes. Generally, high quality approximation does not come without a price.

A crucial element in differential quantization is the predictor $P(z)$. The output $x_P(n)$ of $P(z)$ is a weighted sum of past input samples. The general form is equivalent to a finite impulse response (FIR) filter

$$x_P(n) = \sum_{k=1}^P a_k x_Q(n-k). \quad (19)$$

Where, P = order of predictor

a_k = weighting factor (coefficient)

$x_P(n)$ = output of predictor

$x_Q(n) = d_Q(n) + x_P(n)$ (input to predictor).

Applying the z-transform to equation (19) yields the transfer function for the predictor

$$P(z) = \frac{X_P(z)}{X_Q(z)} = \sum_{k=0}^P a_k z^{-k}. \quad (20)$$

Equations (19) and (20) show that the predictor output is a linear combination of past inputs giving rise to the term "linear prediction". Non-linear predictors (non-linear combination of past input samples) have been studied, however, due to complexity and stability issues, their popularity is limited.

The coefficients a_k are calculated such that $P(z)$ will provide a reasonably accurate model for the behavior of human speech. In 1966 a paper was published by McDonald [6] suggesting coefficients, based on normalized autocorrelation of human speech samples, for predictors of order one to ten. Later, in 1972, Noll [4] published similar data. These two papers are generally referenced in determination of predictor coefficients. Assuming that $P(z)$ does provide a reasonably accurate model for human speech, $x_Q(n) \approx x(n)$ and $d(n) \rightarrow 0$. In other words, a good predictor should minimize the difference signal $d(n)$. This is the basis for differential quantization. In most DM algorithms the predictor order P is set to one.

The potential for instability in differential quantizers exists in the encoder. As mentioned early, equation (20) is the transfer function of an FIR filter. One of the significant characteristics of FIR filters is, by definition, they are stable (the transfer function has only zeros). However, when an FIR is placed in a feedback path, as is the case with the encoder, the zeros become poles—if one of these poles finds its way outside the unit circle the differential quantizer will be unstable.

3. Delta Modulation [1, 4, 6, 8, 10]

DM is a differential quantization scheme that uses two level quantization (i.e. one bit quantizer). By using a single bit to represent each sample, the sample rate and the bit rate are equivalent. Consequently, sample rate is directly related to signal quality (SNR). Also, BW of the input signal and band-limiting of the output are significant factors in determining signal quality.

The first DM algorithm was conceived in 1946. During the last 50 years, two algorithms; Linear DM (LDM) and CVSD, received a significant amount of attention from those interested in efficient methods to digitize speech. LDM is the most fundamental, least complex form of DM. Hence, this tutorial will attempt to provide an understanding of LDM as a foundation for understanding more complex DM algorithms. The CVSD algorithm, an accepted standard in the tactical community and finding a growing number of applications in "point-to-point" wireless applications, will also be covered.

3.1 Linear DM [1, 4, 6, 8]

The predictor $P(z)$ in LDM is first order and the quantizer \mathbf{Q} is two level. Figure 6 is a flow diagram describing the LDM encoder and decoder algorithms.

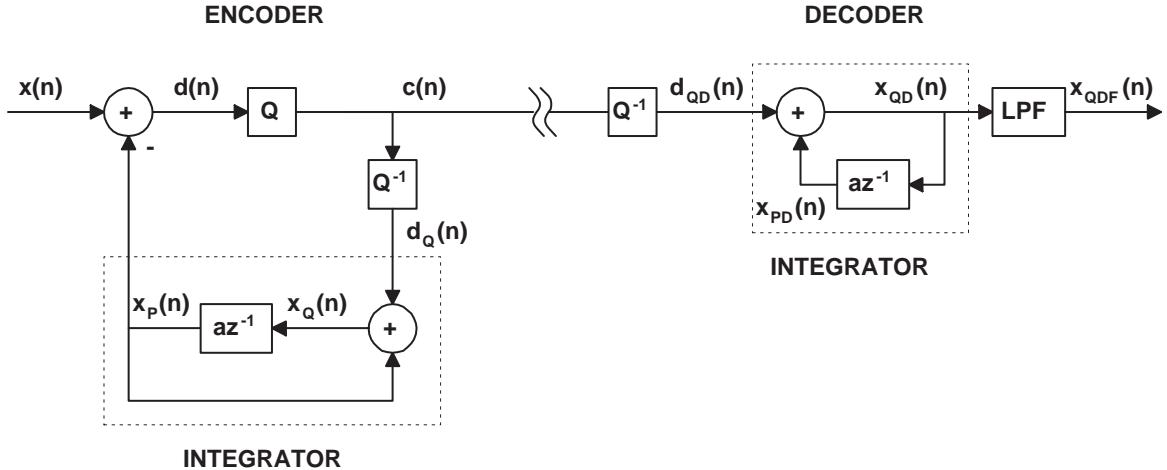


Figure 6: LDM block diagram.

The predictor is shown as a single tap FIR filter with transfer function $P(z) = az^{-1}$. In the encoder, the transfer function between $d_Q(n)$ and $x_P(n)$ can be expressed in the z -domain terms as

$$\frac{X_P(z)}{D_Q(z)} = \frac{az^{-1}}{1 - az^{-1}} \quad (21)$$

In the decoder, the transfer function between $d_{QD}(n)$ and $x_{QD}(n)$ is

$$\frac{X_{QD}(z)}{D_{QD}(z)} = \frac{1}{1 - az^{-1}} \quad (22)$$

Equations (21) and (22) represent discrete time integrators (if $a = 1$). If $a < 1$ they become damped or lossy integrators. McDonald [6] and Noll [4] both suggest $a = 1$ for optimum prediction gain. A pure integrator ($a = 1$) will cause bit errors to propagate longer than if $a < 1$. In practice, a value of $a < 1$ is preferred.

The quantizer \mathbf{Q} in figure 6 functions like a comparator. When the input $d(n)$ exceeds zero it outputs a logic one, when the input is less than zero the output is a logic zero. Hence, the output of \mathbf{Q} is a single bit indicating the sign of the magnitude of $d(n)$. The inverse quantizer \mathbf{Q}^{-1} converts logic levels to delta $d_Q(n)$ as shown in the table below.

sign of $d(n)$	$c(n)$	$d_Q(n)$
-	0	$-\Delta$
+	1	Δ

Table 1: Sign of $d(n)$ with corresponding logic levels and deltas.

The value of Δ plays an important role in the performance of LDM. If Δ is relatively small, tracking of slowly changing, low amplitude signals is quite good at the expense of poor tracking for fast, abruptly changing signals. When DM is not able to keep up with the input signal a phenomena called *slope overload* is exhibited, as shown in Figure 7.

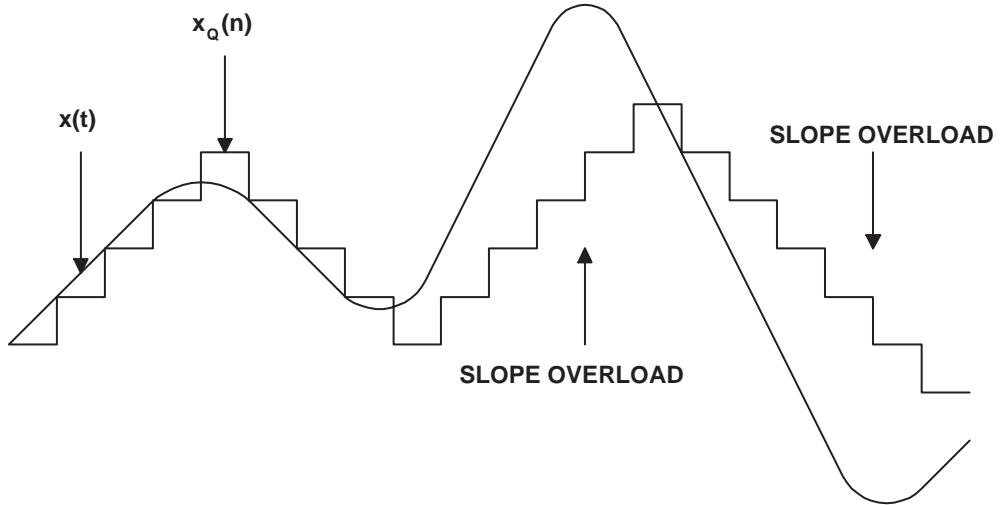


Figure 7: Slope overload in LDM.

Increasing the value of Δ can lessen the effects of slope overload but creates a new problem; *granular noise*. With Δ to large, low amplitude signals will not be quantized at fine enough levels and they appear as *idle channel noise*, see Figure 8. The idle channel pattern is simply an alternating one-zero sequence indicating the input signal amplitude is not changing. Since an alternating one-zero bit pattern has a mean value of zero, the signal out of the decoder will integrate to zero.

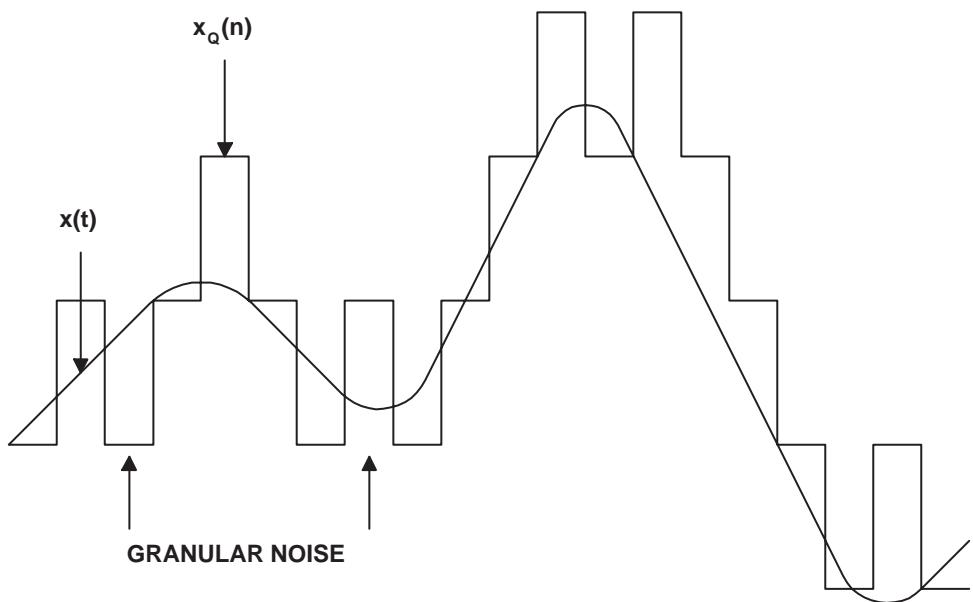


Figure 8: Granular noise in LDM.

3.2 CVSD [2, 4, 10]

Issues with granularity and slope overload can be drastically reduced by making dynamic adjustments to the quantizer step size Δ . Adaptive DM (ADM) algorithms attempt to do this by making Δ small for slowly changing signals and large for rapidly changing signals. The most publicized ADM algorithm is known as CVSD. It was first proposed by Greefkes and Riemens in 1970 [2]. In their CVSD algorithm, adaptive changes in Δ are based on the past three or four sample outputs (i.e. $c(n)$, $c(n-1)$, $c(n-2)$, $c(n-3)$). Figure 9 and Figure 10 show flow diagrams of the algorithms for the encoder and decoder respectively. Notice only the last three samples of $c(n)$ are used here.

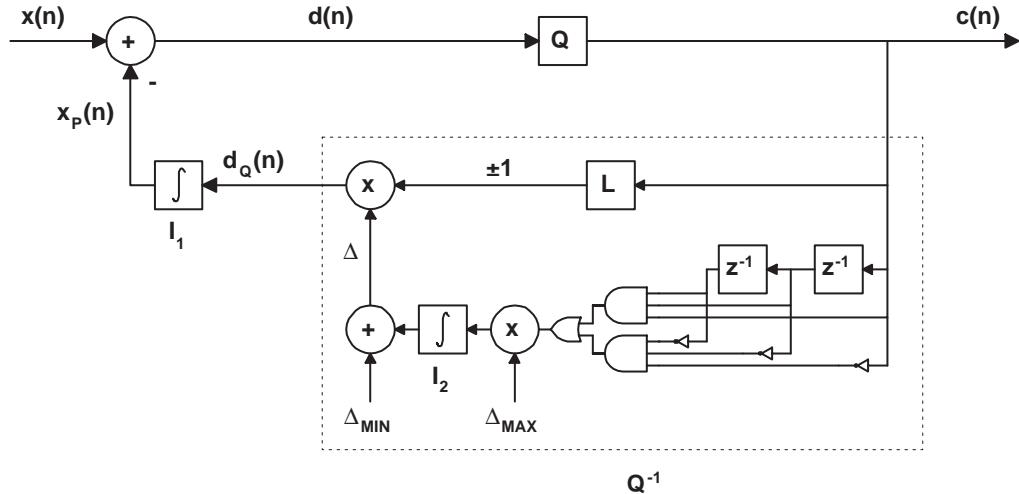


Figure 9: CVSD encoder block diagram.

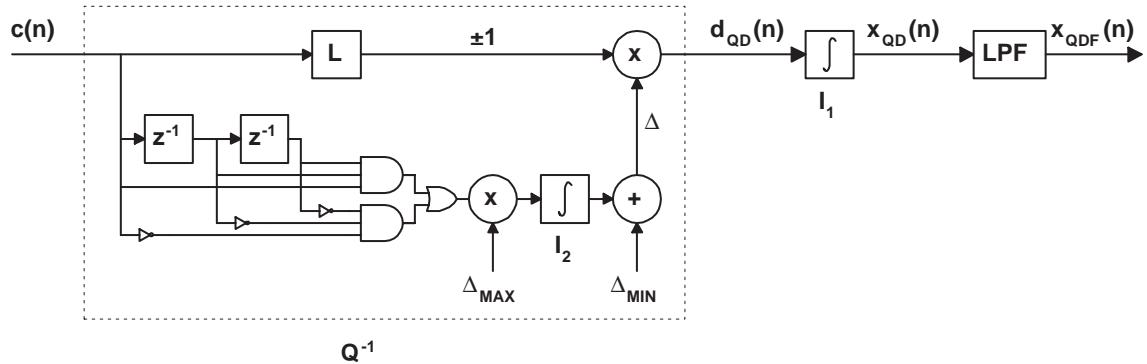


Figure 10: CVSD decoder block diagram.

The block labeled \mathbf{Q}^{-1} , previously shown in Figure 6, has been replaced by two cascaded one sample delays (z^{-1}) and logic to determine when three consecutive ones or zeros have occurred. The minimum and maximum step height is set by Δ_{MIN} and Δ_{MAX} , respectively. When strings of three consecutive zeros or ones have not occurred for a period of time long enough for the output of the integrator \mathbf{I}_2 to decay to near zero, the algorithm is equivalent to LDM (section 3.1). The time constants for the integrator \mathbf{I}_2 and integrator \mathbf{I}_1 are typically 4 ms and 1 ms respectively. In CVSD literature integrator \mathbf{I}_1 is referred to as the *principle integrator* and integrator \mathbf{I}_2 as the *syllabic integrator*. The so called *syllabic integrator* derives its name from the length of syllable. Actually a syllable is about 100 ms in duration, however, pitch changes are on the order of 10 ms. Consequently, 4 ms seems to work best for the CVSD *syllabic* (*pitch*) time constant. The block labeled \mathbf{L} performs simple level conversion (i.e. $c(n) = 1$, \mathbf{L} outputs 1; $c(n) = 0$, \mathbf{L} outputs -1). This CVSD algorithm is also known as "Digitally Controlled DM" [10].

Figure 11 displays the output of the integrator \mathbf{I}_1 ($x_p(n)$ in the encoder and $x_{PD}(n)$ in the decoder). When compared to Figure 7 and Figure 8 slope overload and granular noise are reduced.

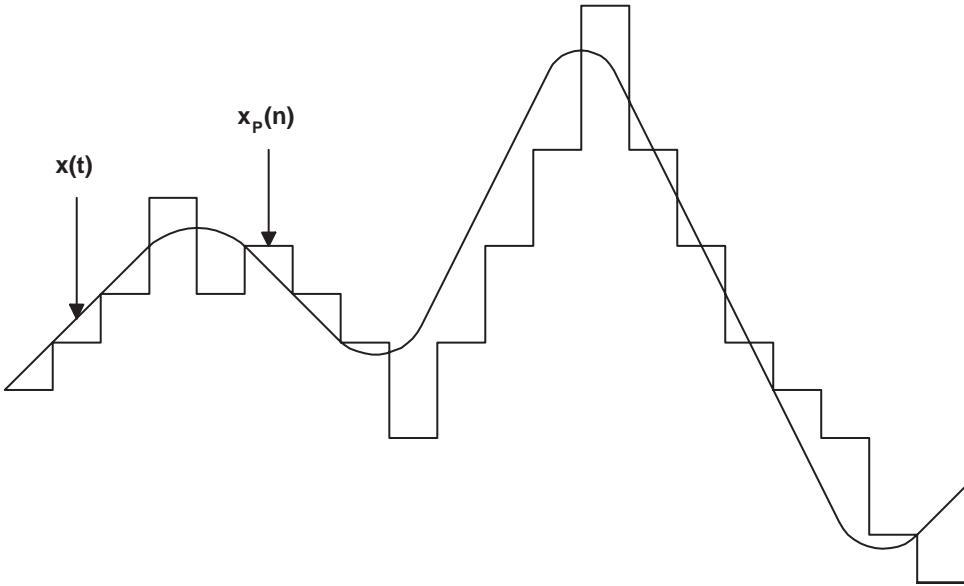


Figure 11: CVSD quantization noise.

3.3 Performance Measurement of DM [1, 4]

The maximum SNR for DM speech coders is [1]

$$SNR_{dB} = 10 \log \left(\frac{f_S^3}{f_{BW} f^2} \right) - 14.0. \quad (23)$$

Where, f_S = data rate (sample rate)

f_{BW} = signal BW (cutoff frequency of LPF)

f = signal frequency.

Most literature on DM coding agrees on the terms involving f_S , f_{BW} , and f in equation (23) but most do not agree on -14 dB. This term varies from -17.0dB to -11.7dB depending on a variety of different input signals and assumptions about environment. Nevertheless, equation (23) is considered a maximum because it assumes no slope overload and minimum granular noise, i.e. optimum step size.

The adaptive behavior of CVSD results in a SNR versus input level characteristic similar to non-uniform quantization (mentioned in section 2.3). This non-linear SNR characteristic is due to *companding*, where the quantization level (slope) is adjusted to a larger or smaller value according to past pitch changes of the input signal. The number of past samples (bits) used to make a prediction is normally three or four. Four bit *companding* has proven to be most effective for data rates greater than 32kb/s, three bit *companding* seems to work better for data rates less than 32kb/s.

Figure 12 shows a typical SNR versus input level characteristic for 32kb/s CVSD (four bit *companding* algorithm, input signal = 820Hz sinewave). Equation (23) predicts a maximum SNR of 27.6dB which is nearly met in Figure 12 for an input signal level of -15 dbm0 (0dBm0 = 489 mVrms). Above and below the -15dBm0 input level granular and slope overload noise start cause distortion. Finally, It should be noted that CVSD does not perform well with continuous tone or data input signals, particularly at data rates below 32kb/s.

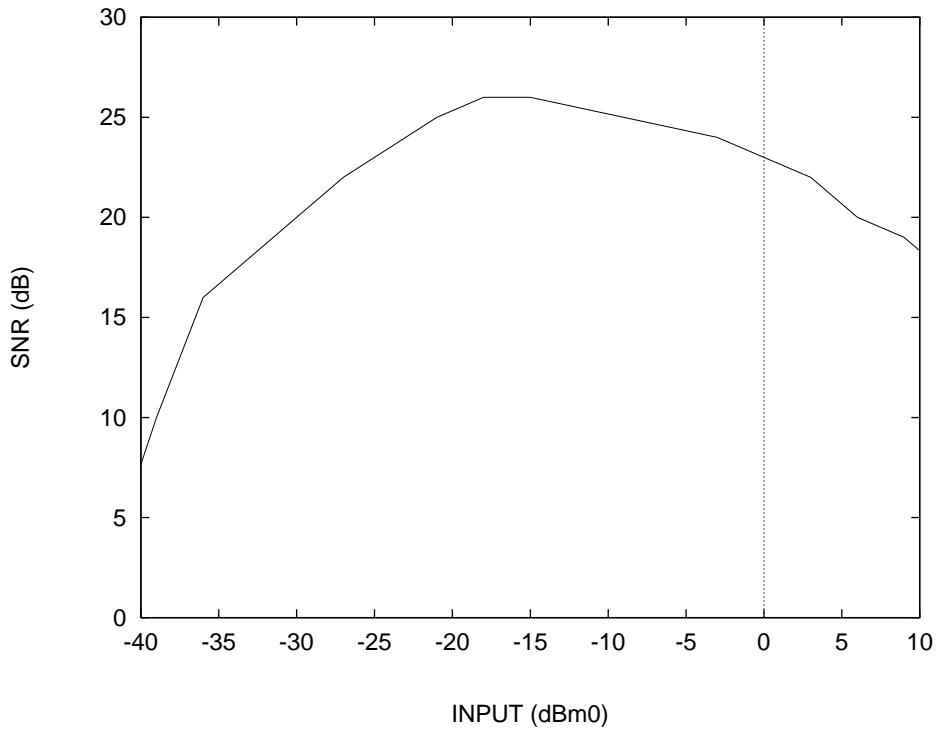


Figure 12: Measured SNR versus input level for 32kb/s CVSD (input = 820Hz sinewave).

SNR is the most used method to objectively quantify performance of speech coding algorithms. However, it does not always correspond with perceived quality, particularly for differential and adaptive algorithms using actual voice as the input. In addition, it is difficult to make reliable SNR measurements in the presence of random bit errors. In an effort to quantify the perceived quality of a speech coding algorithm, *Mean Opinion Score* (MOS) testing was developed [4]. Table 2 summarizes the five point scale used to judge quality and impairment.

Number	Quality	Impairment
5	excellent	imperceptible
4	good	perceptible but not annoying
3	fair	slightly annoying
2	poor	annoying
1	bad	very annoying

Table 2: Mean Opinion Score testing guidelines.

An MOS rating of 4 to 4.5 is considered *Toll Quality* (equivalent to commercial telephony). Whereas *Communication Quality* MOS ratings are 3 to 4 (barely perceptible distortion, but no degradation in intelligibility).

Figure 13 compares MOS ratings for μ -law PCM (the standard for *Toll Quality*), CVSD and ADPCM. Notice CVSD performs as well or better than both μ -law PCM and ADPCM in the presence of bit errors. Specifically, CVSD retains quite good MOS ratings at bit error rates exceeding 1%, and at 10% has an MOS rating of 3 (*Communication Quality*). It is this robustness to bit errors (channel noise) that makes CVSD an ideal solution for many wireless speech communication applications.

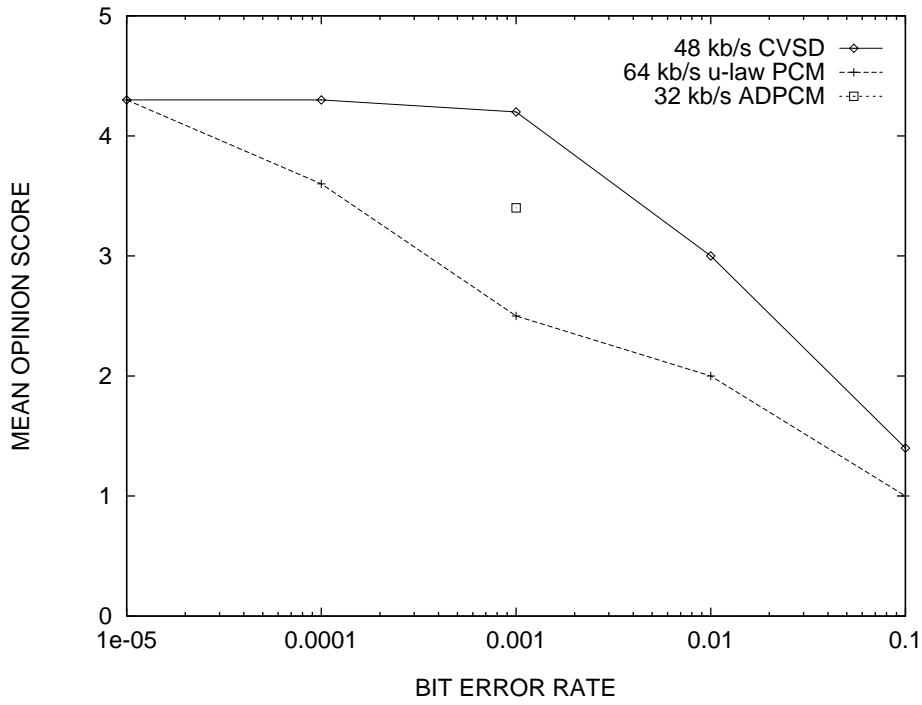


Figure 13: MOS versus bit error rate for ADPCM, μ -law PCM, and CVSD from [4].

4. Summary

CVSD has several attributes that make it well suited for digital coding of speech. One-bit words eliminate the need for complex framing schemes. Robust performance in the presence of bit errors make error detection and correction hardware unnecessary. Other speech coding schemes may require a digital signal processing engine and external analogue to digital/digital to analogue converters to convert the analogue signal in to a form that can be processed digitally—the entire CVSD codec algorithm, including input and output filters, can be integrated on a single silicon substrate. Despite this simplicity, CVSD has enough flexibility to allow digital encryption for secure applications.

Finally, CVSD can operate over a wide range of data rates—it has been successfully used from 9.6kb/s to 64kb/s. At 9.6kb/s audio quality is not particularly good, however, it is intelligible. At data rates of 24kb/s to 48 kb/s it is judged as quite acceptable. And above 48 kb/s it is comparable to *toll quality*. All of these attributes make CVSD attractive to wireless telecommunication systems (e.g. digital cordless telephones, digital Land Mobile Radio). The defence industry has been using CVSD for decades in wireline and wireless systems as specified in Mil-Std-188-113.

More recently, Federal Standard 1023 proposed CVSD for 25kHz channel radios operating above 30MHz. Figure 14 is a block diagram showing a CVSD Codec (FX609) in a digital mobile radio system.

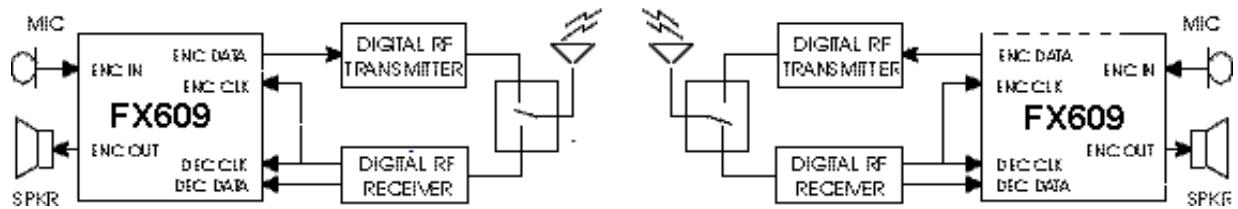


Figure 14: Digital wireless system incorporating CML's FX609 CVSD Codec.

This tutorial has attempted to shed some light on the fundamental aspects of CVSD. It has shown that CVSD is a differential adaptive quantization algorithm with one bit coding and first order prediction (one bit ADPCM). In addition, objective and subjective methods of measuring signal quality showing that CVSD performs quite well in the presence of bit errors (noisy channel) have been presented.

5. References

- [1] J. C. Bellamy, Digital Telephony, Wiley and Sons, New York, 1982.
- [2] J. A. Greefkes and K. Riemens, "Code Modulation with Digitally Controlled Companding for Speech Transmission," Philips Tech. Rev., pp. 335-353, 1970.
- [3] A. Gersho, "Principals of Quantization," IEEE Transactions on Circuits and Systems, pp. 427-436, July 1978.
- [4] N. S. Jayant and P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Prentice-Hall, Englewood Cliffs, N. J., 1984.
- [5] A. B. Jerrí, "The Shannon Sampling Theorem—Its Various Extensions and applications: A Tutorial Review," Proceedings of the IEEE, pp. 1565-1596, November 1977.
- [6] R. A. McDonald, "Signal-to-Noise and Idle Channel Performance of Differential Pulse Code Modulation Systems—Particular Applications to Voice Signals," Bell System Technical Journal, pp. 1123-1155, Sept. 1966.
- [7] P. Noll, "A Comparative Study of Various Quantization Schemes for Speech Encoding," Bell System Technical Journal, pp. 1597-1614, November 1975.
- [8] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, N. J., 1978.
- [9] M. Schwartz, Information, Transmission, Modulation, and Noise, McGraw Hill, New York, 1980.
- [10] R. Steele, Delta Modulation Systems, Pentech Press, London, England, 1975.