Wave form Coding - Unit II

Prediction Filtering & DPCM

Prediction is a special form of estimation. The requirement is to use a finite set of present and past samples of a stationary process to predict a sample of the process in the future. It is linear if it is a linear combination of the given samples of the process. The filter designed to perform the prediction is called a predictor. The difference bet the actual value of the sample (arrived at by the process in future) to the predicted value (which is the output of the predictor) is called the prediction error. A predictor filter is designed to minimize the mean square value of the prediction error.

Consider the random samples $X_{n-1}, X_{n-2}, \ldots, X_{n-m}$ drawn from a stationary process $X(t)$. Suppose the requirement is to make a prediction of the sample X_n. Let $\hat{X}_n$ denote the random variable resulting from this prediction. We thus write.

$$\hat{X}_n = \sum_{K=1}^{M} h_{oK} \cdot X_{n-K}.$$

Where $h_{o1}, h_{o2}, \ldots, h_{oM}$ are the optimum predictor co-efficients. M is known as the order of the filter.

When an audio or video signal is sampled slightly above the Nyquist rate, adjacent samples have a good degree of correlation. This implies that by directly encoding these sample values we are permitting a good degree of redundancy. Rather symbols that are not absolutely essential to the transmission of information are generated as a result of the encoding process. By removing this redundancy before encoding, we obtain a more efficient coded signal.

The DPCM system employs a predictor, which predicts the present sample value, using a few immediate past sample values. The predicted value of the present sample is compared to the actual value and the diff bet^n the 2 samples is Pulse code modulated. The advantage lies in the fact that if the prediction is reasonably good, the difference signal will have a much smaller dynamic range than the message itself. Thus far few bits would be required to code the error sample, in comparison to the bits req for the original sample as such.

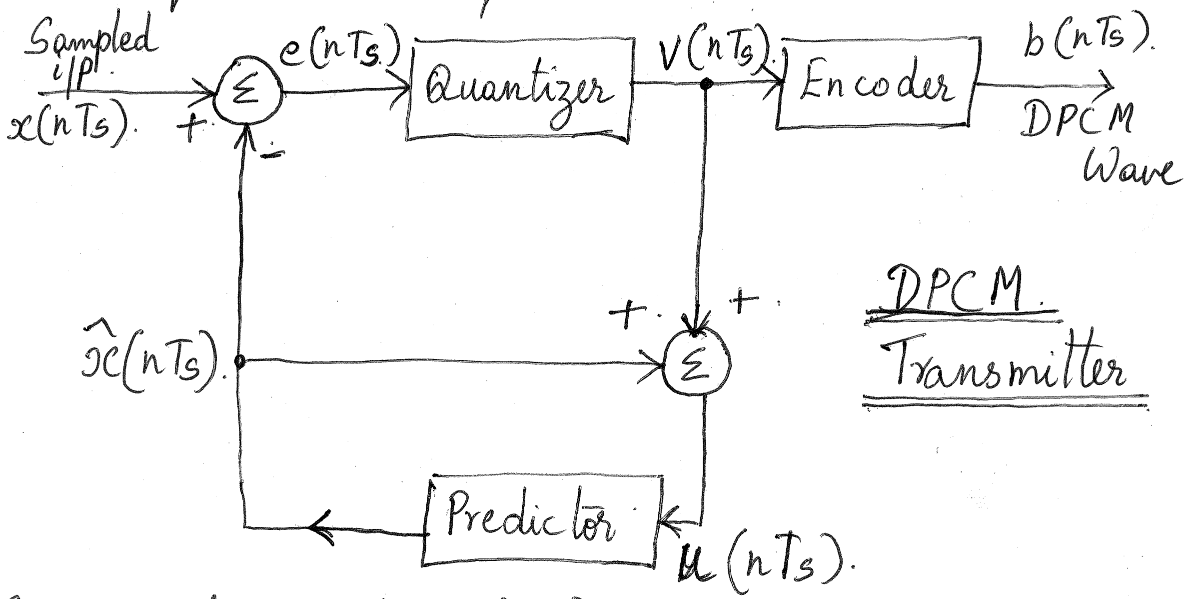
Suppose a message signal $x(t)$ is sampled at the rate $f_s = 1/T_s$ to produce a sequence of correlated samples T_s seconds apart. Let the sequence be denoted by $\{x(nT_s)\}$, where n takes on integer values. Thus the input to the quantizer is the prediction error which is given by.

$$e(nT_s) = x(nT_s) - \hat{x}(nT_s) \quad - (1)$$

Let the non linear function $Q(\cdot)$ define the i/p - op characteristic of the quantizer. The quantizer o/p may be represented as.

$$v(nT_s) = Q[e(nT_s)].$$
$$= e(nT_s) + q(nT_s). \quad - (2)$$

where $q(nT_s)$ is the quantisation noise.



DPCM Transmitter

The quantiser o/p $v(nT_s)$ is added to the predicted value $x(nT_s)$ to produce the i/p of the predictor which is

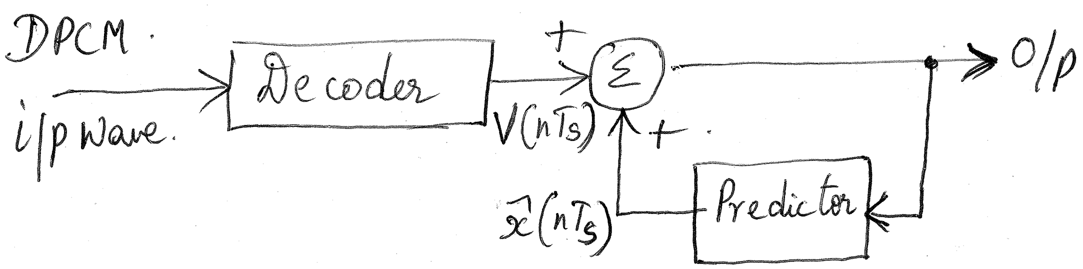
$$u(nT_s) = \hat{x}(nT_s) + v(nT_s) \quad - (3)$$ Using eq (2) in eq (3)

$$u(nT_s) = \hat{x}(nT_s) + e(nT_s) + q(nT_s).$$ Now using eq (1)

$$u(nT_s) = \hat{x}(nT_s) + [x(nT_s) - \hat{x}(nT_s)] + q(nT_s)$$
$$= x(nT_s) + q(nT_s) \quad - (4)$$

which represents a quantized version of the i/p signal $x(nT_s)$.

The receiver consists of a decoder to reconstruct the quantized error signal. The quantized version of the original i/p is reconstructed from the decoder o/p using the same predictor as used in the transmitter.



$$V(nT_s) + \hat{x}(nT_s) = e(nT_s) + q(nT_s) + \hat{x}(nT_s) \quad - (5) \text{ using Eq } (2)$$

We know $e(nT_s) = x(nT_s) - \hat{x}(nT_s)$.
$$x(nT_s) = e(nT_s) + \hat{x}(nT_s) \quad - (6)$$

Using eq (6) in eq (5) we get

$$o/p = V(nT_s) + \hat{x}(nT_s) = x(nT_s) + q(nT_s).$$

This shows that the receiver o/p is equal to $u(nT_s)$. ~~which~~ Thus it is seen that the predictors in the X'mitter & receiver operate on the same sequence of samples $u(nT_s)$

O/P signal to quantization noise (SNR).

An output signal-to-quantization noise ratio can be defined as

$$(SNR)_o = \sigma_x^2 / \sigma_Q^2 \quad -(1)$$

where σ_x^2 is the variance of the original i/p $x(nT_s)$, & σ_Q^2 is the variance of the quantization error $q(nT_s)$.

Equation (1) may be rewritten as

$$(SNR)_o = \left(\frac{\sigma_x^2}{\sigma_E^2}\right)\left(\frac{\sigma_E^2}{\sigma_Q^2}\right) = G_p \cdot (SNR)_p .$$

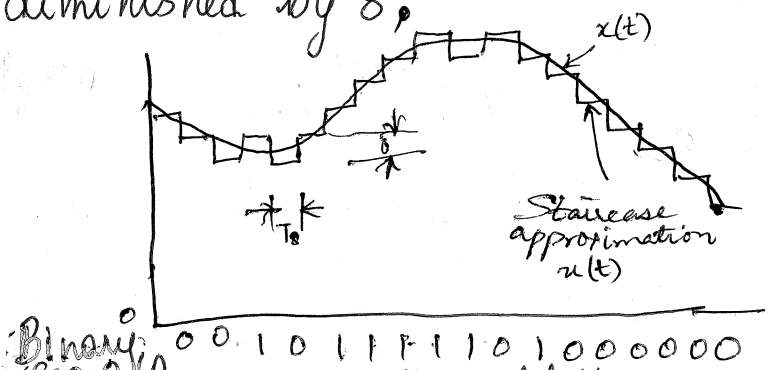
where $(SNR)_p$ is the prediction error to quantization noise ratio, defined by $(SNR)_p = \frac{\sigma_E^2}{\sigma_Q^2}$.

and G_p is the prediction gain produced by the differential quantization scheme, defined by $G_p = \frac{\sigma_x^2}{\sigma_E^2}$

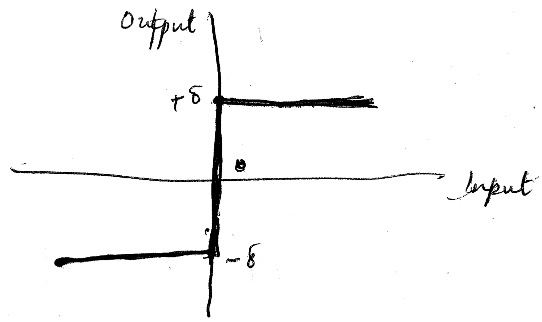
The quantity G_p, when greater than unity, represents the gain in signal-to noise ratio that is due to the differential quantization scheme. Now for a given baseband signal, the variance σ_x^2 is fixed, so G_p can be maximized by minimizing the variance of σ_E^2 of the prediction error $e(nT_s)$. So the design of the predictor filter should be made so as to minimize σ_E^2.

Delta modulation: (DM).

Delta modulation is the one-bit (or 2 level) version of DPCM. In its basic form DM provides a staircase approximation to the oversampled i/p signal. The difference betⁿ the i/p and the approximation is quantized into only 2 levels, namely $\pm\delta$ corresponding to +ve and negative differences. Thus, if the approximation falls below the signal at any sampling epoch, it is increased by δ, If the approx lies above the signal, it is diminished by δ.



Delta Modulation



I/P - O/P charac of the 2-level quantizer

δ denotes the absolute value of the two representation levels of the one-bit quantizer used in DM. The step size Δ of the quantizer is therefore related to δ by

$$\Delta(\text{stepsize}) = 2\delta. \quad (\text{refer fig 2 above})$$

Denote the input signal as $x(t)$ and the staircase approximation to it as $u(t)$. Then

$$e(nT_s) = x(nT_s) - \hat{x}(nT_s)$$

$$= x(nT_s) - u(nT_s - T_s) \quad — (1)$$

The O/P function $b(nT_s)$ is given by

$$b(nT_s) = \delta\,[\operatorname{sgn} e(nT_s)]$$

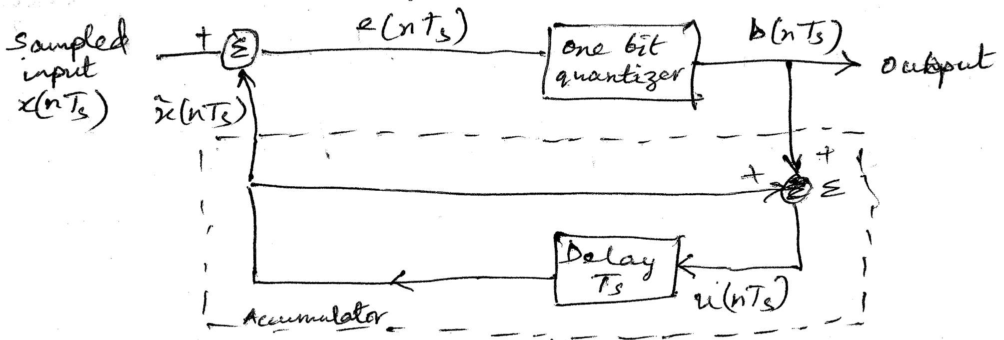
We know that sgn(z) is called the signum function and is defined by

$$sgn(z) = \begin{cases} 1 & \text{if } z > 0 \\ -1 & \text{if } z < 0 \end{cases}$$

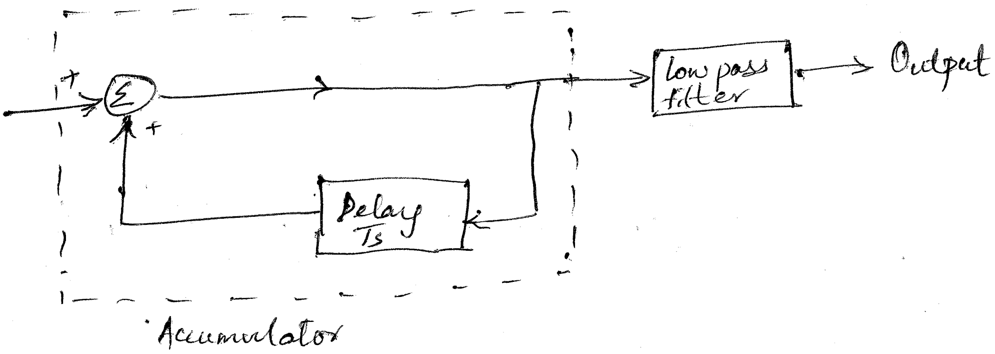
and $u(nT_s) = u(nT_s - T_s) + b(nT_s)$. (from fig ①).

where T_s is the sampling period; $e(nT_s)$ is a prediction error representing the difference betn present sample $x(nT_s)$ and the latest approximation to it, ie $\hat{x}(nT_s) = u(nT_s - T_s)$.

The binary quantity $b(nT_s)$ is the algebraic sign of the error $e(nT_s)$ except for the scaling factor δ: $b(nT_s)$ is the one bit word X'mitted by the DM system



DM Transmitter



DM Receiver

The Delta modulation transmitter consists of a summer, a two level quantizer and an accumulator. The accumulator is initially set to zero. The accumulator o/p is given by.

$$u(nT_s) = \delta \sum_{i=1}^{n} sgn(e(iT_s)) = \sum_{i=1}^{n} b(iT_s)$$

At each sampling instant, the accumulator increments the approximation to the the input signal by $\pm \delta$. The out-of band quantization noise in the high freq staircase waveform $u(t)$ is rejected by passing it through a low pass filter with a BW equal to the original signal.

DM offers two unique features (1) a one-bit codeword for the o/p. (2) Simplicity of design for both the x'mitter & receiver.

Disadvantages of DM

(1) Severe slope overload distortion occurs if the message has steep gradients.

(2) The i/p to the single bit quantizer is the diff betn the present sample & an estimate of the previous sample ie $e(nT_s)$. Thus X'mission channel noise can cause accumulation of errors in the receiver.

Slope overload distortion & Granular noise

Let $q(nT_s)$ ~~deno~~ denote the quantizing error. We may then write the ~~quantiser o/p~~ term

$$u(nT_s) = x(nT_s) + q(nT_s). \quad — \quad (1)$$

where $u(nT_s)$ is the quantized version of $x(nT_s)$

Using eq (1) in eq of DM which states

$$e(nT_s) = x(nT_s) - u(nT_s - T_s)$$

$$\cancel{e(nT_s) = x(nT_s) - x(nT_s - T_s) - q(nT_s -}$$

$$e(nT_s) = x(nT_s) - x(nT_s - T_s) - q(nT_s - T_s).$$

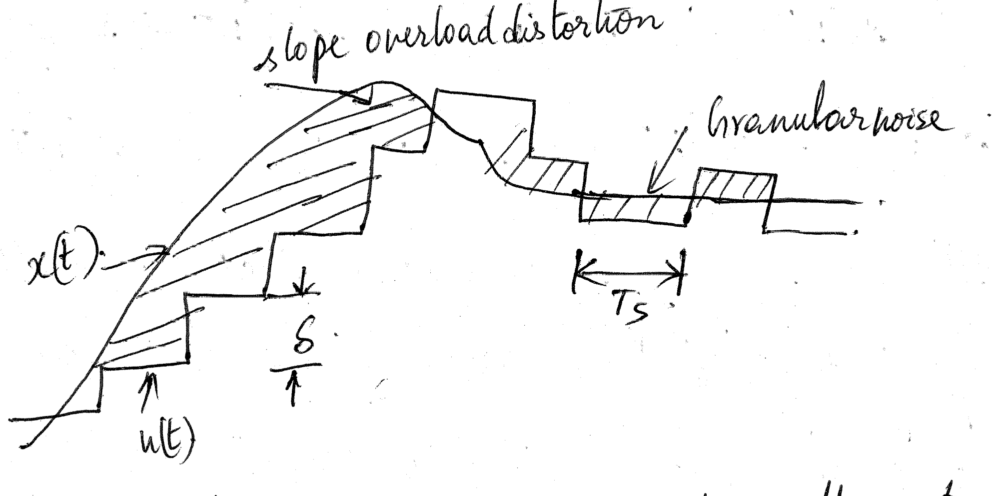
Although in fig is is shown that the delta modulator is perfectly tracking the message signal $x(t)$ it may not always do so. The average rate of increase or decrease of $x_q(t)$ is given by (Δ / T_s). $\therefore \left(\frac{\Delta}{T_s}\right) < \frac{dx(t)}{dt}$

If $\frac{\Delta}{T_s}$ is smaller than the maximum rate of change of $x(t)$ of the ie $\frac{d\,x(t)}{dt}$, then the linear Delta modulator will not be able to track the $x(t)$ properly ie the staircase waveform will be very much different from the message signal. This inability of the LDM to correctly track the message signal $x(t)$.

when $x(t)$ has steep changes is referred to as 'slope overload. There for to avoid this an LDM system should insure that

$$\frac{\delta}{T_s} > \left|\frac{dx(t)}{dt}\right|_{max}$$

where Δ is the fixed step size and $T_s = \frac{1}{f_s}$ is the sampling interval. Therefore with f_s fixed, the step size has to be large to avoid slope over load.



In contrast, granular noise occurs when the step size Δ is too large relative to the local slope characteristics of the i/p w/f $x(t)$, there by causing the staircase approximation $u(t)$ to hunt around a relatively flat sample/segment of the i/p waveform. This granular noise is simalar to the quantization noise of PCM.

Thus we find 2 conflicting req. — a large step size to avoid slope overload noise and a small step size to reduce the granular noise. Hence we would use an adaptive system in which the step size automatically varies with the rate of change of $x(t)$.

Output signal-to-noise ratio for sinusoidal modulation in DM.

Let the sinusoidal message signal be represented by $x(t) = A_0 \sin \omega_0 t$.

The max rate of change of this signal is

$$\max\left|\frac{dx(t)}{dt}\right| = \left|A_0 \omega_0 \cos \omega_0 t\right|_{max} = A_0 \omega_0 = A_0 2\pi f_0.$$

WKT
$$\left|\frac{\delta}{T_s}\right| \geq \left|\frac{dx(t)}{dt}\right| \geq 2\pi A_0 f_0.$$

We may impose the following cond^n on the amplitude of the sinusoidal modulation.

$$A_{0(max)} = \frac{\delta}{2\pi f_0 T_s} \quad \text{or} \quad A_0 \leq \frac{\delta}{2\pi f_0 T_s}.$$

Hence the max permissible value of the output signal power equals.

$$P_{max} = \left(\frac{A_0}{\sqrt{2}}\right)^2 = \frac{\delta^2}{(2\pi f_0 T_s)^2} \cdot \frac{1}{2}.$$

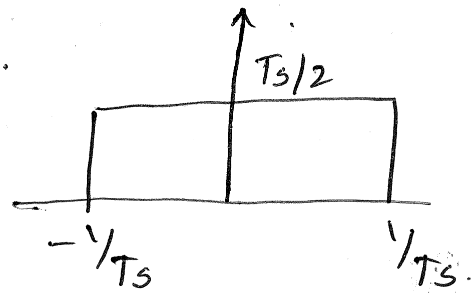
$$\therefore P_{max} = \frac{\delta^2}{8\pi^2 f_0^2 T_s^2}$$

When there is no slope over load, the max quantization error is $\pm\delta$. Assuming uniform distribution of error.

W.K.T. The variance of the quantisation noise σ_q^2 is given by $\sigma_q^2 = \frac{\Delta^2}{12} = \frac{(2\delta)^2}{12} = \frac{\delta^2}{3}$.

The receiver contains a LPF whose BW is equal to W; $W \leq f_s$.

Hence assuming that the quantization error is distributed over a freq interval from $-1/T_s$ to $\frac{1}{T_s}$.



∴ The probability density function of the error is $\frac{1}{b-a} = \frac{1}{\frac{1}{T_s} - (-1/T_s)}$

$= T_s/2$.

∴ Average o/p noise power = BW × Variance of error × PDF

$$= \cancel{2}W \cdot \frac{T_s}{\cancel{2}} \cdot \frac{\delta^2}{3} = W T_s \cdot \frac{\delta^2}{3}.$$

Therefore the max value of the $(SNR)_{o\,(max)} = \frac{P_{max}}{W T_s \left(\frac{\delta^2}{3}\right)}$

$$= \frac{\delta^2}{8\pi^2 f_0^2 T_s^2} \cdot \frac{3}{W T_s \delta^2} = \frac{3}{8\pi^2 W f_0^2 T_s^3} = \frac{3 f_s^3}{8\pi^2 W f_0^2}.$$

Thus $(SNR)_{o\,(max)}$ is proportional to sampling rate cube.

Adaptive Delta Modulation (ADM)

The performance of a delta modulator can be improved by making the step size of the modulator assume a time varying form. During a steep segment of the i/p signal the step size is increased. Conversely when the i/p signal is varying slowly, the step size is reduced. The resulting method is called adaptive delta modulation. In practical implementations the step size $\Delta(nT_s)$ or $2\delta(nT_s)$ is constrained to lie between minimum & maximum values.

$$\delta_{min} \leq \delta(nT_s) \leq \delta_{max}$$

The upper limit δ_{max}, controls the amount of slope overload distortion. The lower limit δ_{min}, controls the amount of idle channel noise. In the general form the adaptation rule for $\delta(nT_s)$ is expressed as.

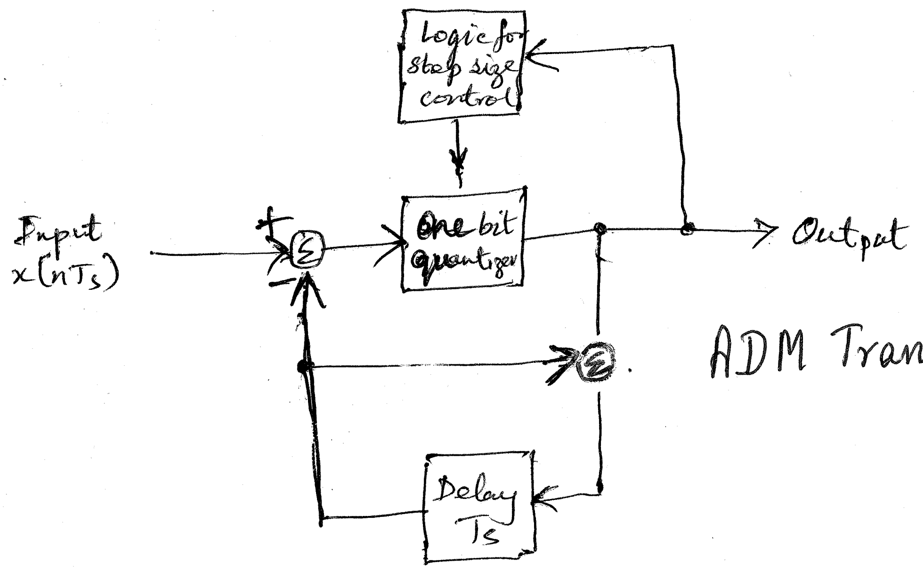
$$\delta(nT_s) = g(nT_s)\,\delta(nT_s - T_s)$$

where the time varying multiplier $g(nT_s)$ depends on the present binary output $b(nT_s)$ and the M previous values. $b(nT_s - T_s)$ - - - - - $b(nT_s - MT_s)$.

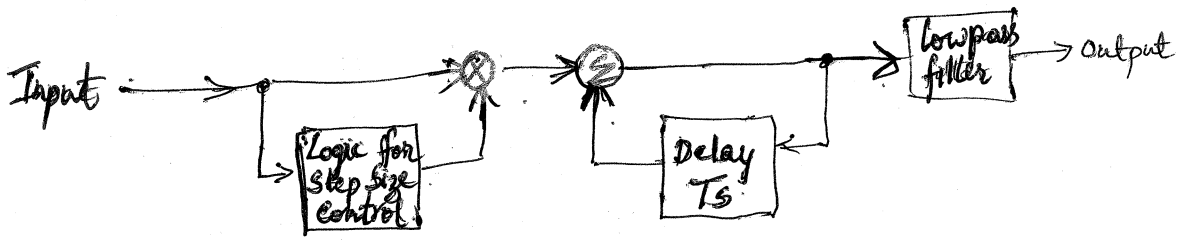
The algorithm is initiated with a starting step size $\delta_{start} = \delta_{min}$.

generally $g(nT_s) = \begin{cases} K & \text{if } b(nT_s) = b(nT_s - T_s) \\ K^{-1} & \text{if } b(nT_s) \neq b(nT_s - T_s) \end{cases}$.

This adaptation algorithm is called a constant factor ADM with one bit memory, where the one-bit memory refers to the explicit utilization of the previous single bit $b(nT_s - T_s)$. $K = 1.5$ has been found to be well-matched for speech & image inputs.



ADM Transmitter.



ADM Receiver.

Adaptive Differential PCM (ADPCM) with prediction filters

CODING SPEECH AT LOW BIT RATES:

When the channel bandwidth is at premium, there is a definite need for speech coding at low bit rates, while maintaining acceptable fidelity or quality of reproduction. A major motivation for bit rate reduction is for secure transmission. The fundamental limits on bit rate suggested by speech perception and information theory show that high quality speech coding is possible at rates considerably less than 64 kb/s. This increases the cost due to the complexity and processing delay involved in it.

The design philosophy:

1) To remove redundancies from the speech signal.

2) To assign available bits to code the non-redundant parts of the speech signal in an efficient manner.

a) ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION (ADPCM)

A digital coding scheme that uses both adaptive quantization and adaptive prediction is called adpative differential pulse-code modulation (ADPCM).

The term "adaptive quantization" refers to a quantizer that operates with a time-varying step size $\Delta(nT_s)$.

$T_s \rightarrow$ sampling period.

The adaptive quantizer is assumed to have a uniform transfer characteristic.

The step size is varied so as to match the variance σ_x^2 of the input signal $x(nT_s)$.

$$\Delta(nT_s) = \phi\,\hat{\sigma}_x(nT_s).$$

$\phi \rightarrow$ constant

$\hat{\sigma}_x(nT_s) \rightarrow$ estimate of the standard deviation $\sigma_x(nT_s)$.

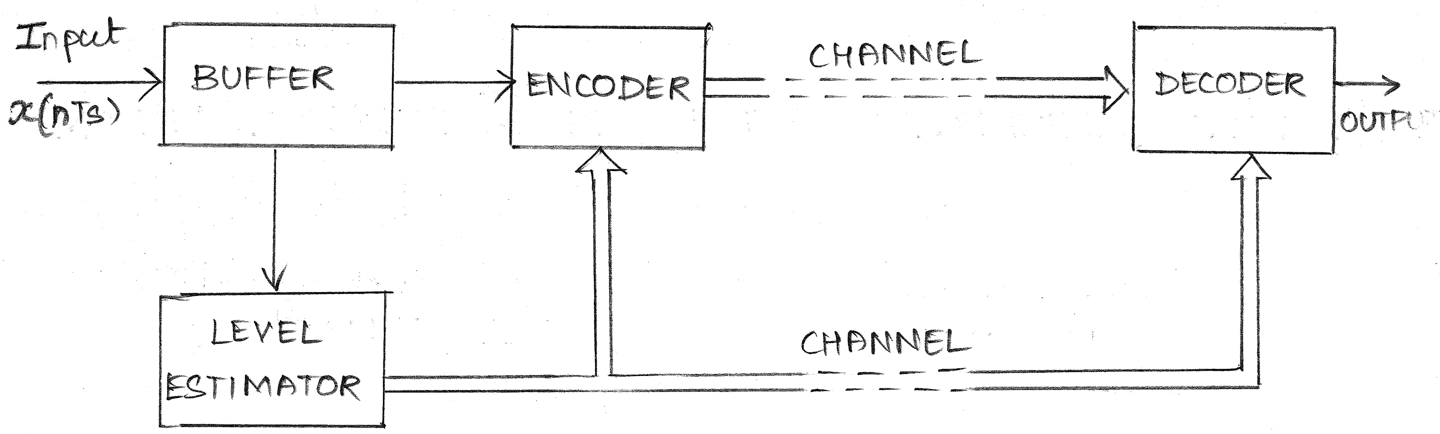
We compute the estimate $\hat{\sigma}_x(nT_s)$ in one of the two ways:

1) Unquantized samples of the input signal are used to derive forward estimates of $\sigma_x(nT_s)$.

2) Samples of the quantizer output are used to derive the backward estimates of $\sigma_x(nT_s)$.

The respective quantization schemes are referred to as adaptive quantization with forward estimation (AQF) and adaptive quantization with backward estimation (AQB)

Adaptive quantization with forward estimation (AQF):



The AQF scheme first goes through a learning period by buffering unquantized samples of the input speech signal. The samples are released after the estimate

$\hat{\sigma}_x(nT_s)$ has been obtained.

Advantages:

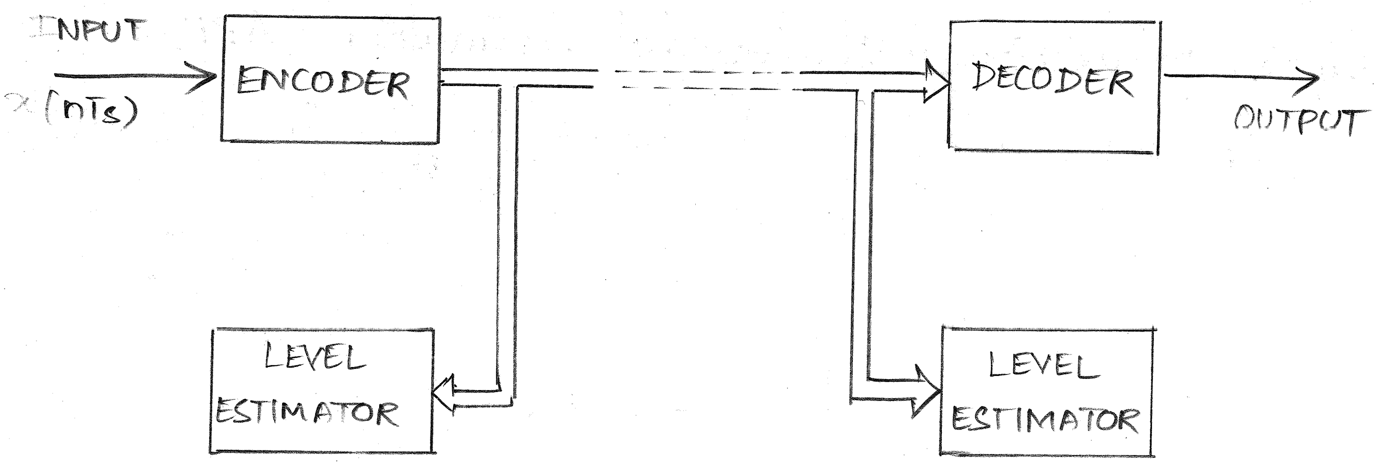
* The estimate is independent of quantizing noise.
* The step size [$\Delta(nT_s)$] obtained from AQF is more reliable.

Disadvantages:

* The use of AQF requires the explicit transmission of level information to a remote decoder, thereby burdening the system with additional side information. This introduces processing delay in the system.

Adaptive quantization with backward estimation (AQB):

The problems of level transmission, buffering and delay intrinsic to AQF are all avoided in the AQB scheme.



In Adaptive quantization with backward estimation, the recent history of the quantizer output is used to extract information for the computation of the step size $\Delta(nT_s)$.

The AQB represents a non-linear feedback system.

The system is stable in the sense that if the quantizer input $x(nT_s)$ is bounded, then the backward estimate $\hat{\sigma}_x(nT_s)$ and the corresponding step size $\Delta(nT_s)$ are also bounded.

Adaptive Prediction:

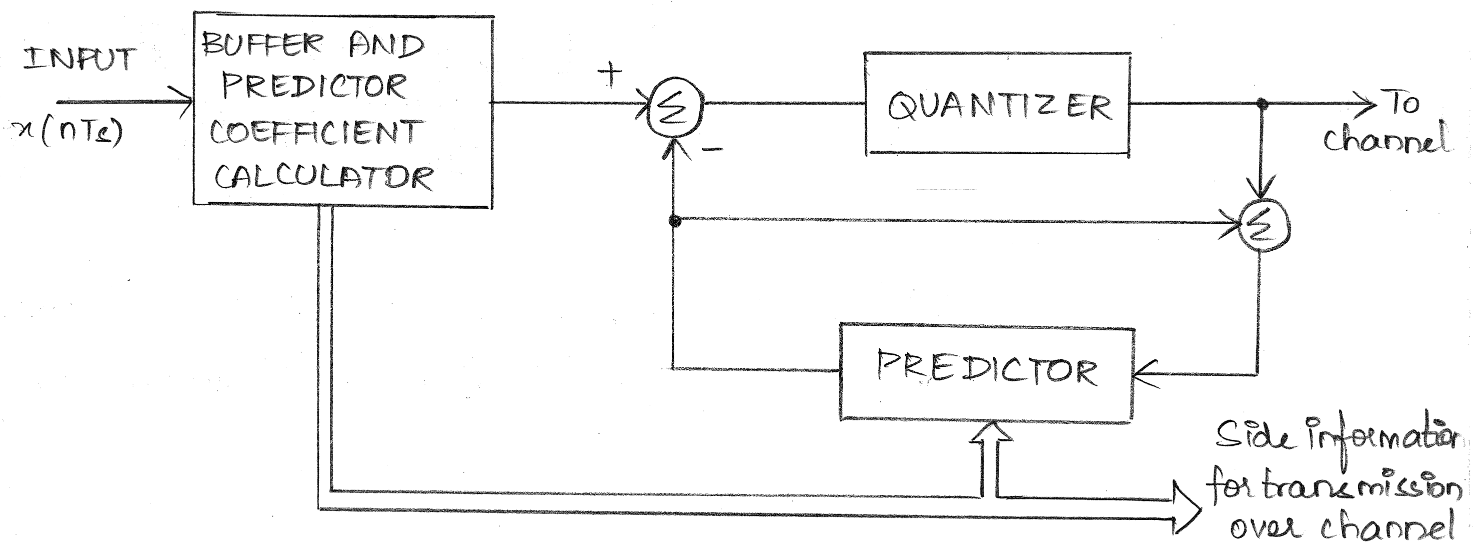
Speech signals are not stationary. The autocorrelation function and power spectral density of speech signals are time-varying functions. This implies that the design of predictors for such inputs should likewise be time-varying, (ie) adaptive.

There are two schemes for performing adaptive prediction:

1) Adaptive prediction with forward estimation (APF)
2) Adaptive prediction with backward estimation (APB)

Adaptive prediction with forward estimation (APF):

In APF, unquantized samples of the input signal are used to derive estimates of the predictor coefficients.



In the APF scheme, N unquantized samples of the input speech are first buffered and then released after computation of M predictor coefficients that are optimized for the buffered segment of input samples.

The choice of M involves a comprise between an adequate prediction gain and an acceptable amount of side information.

The choice of buffer length N involves a comprise between the rate at which statistics of the input speech signal change and the rate at which information on predictor coefficients must be updated and transmitted to the receiver.

Disadvantages:

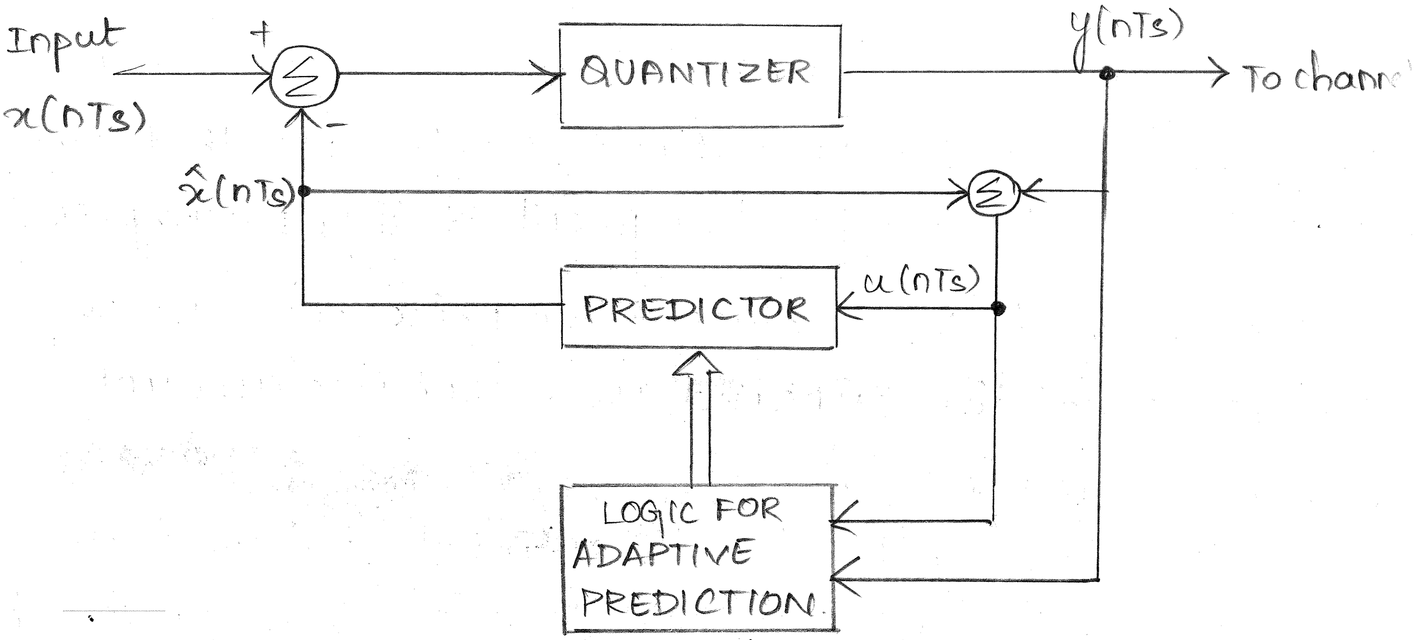
APF suffers from side information, buffering and delay.

These disadvantages are eliminated using APB scheme.

Adaptive Prediction with backward estimation (APB):

In APB, the samples of the quantizer output and the prediction error are used to derive estimates of the predictor coefficients.

The optimum predictor coefficients are estimated on the basis of quantized and transmitted data.



Let $y(nTs)$ denote the quantizer output, where T_s is the sampling period and n is the time index.

The sample value of the predictor input is given by

$$u(nTs) = \hat{x}(nTs) + y(nTs)$$

$\hat{x}(nTs) \rightarrow$ prediction of the speech input sample $x(nTs)$

Rewrite the above equation,

$$y(nTs) = u(nTs) - \hat{x}(nTs)$$

$u(nTs) \rightarrow$ represents a sample value of the predictor input

$\hat{x}(nTs) \rightarrow$ represents a sample value of the predictor output

$y(nTs) \rightarrow$ value of prediction error.

For adaptation of the predictor coefficients, we use the least-mean square (LMS) algorithm.

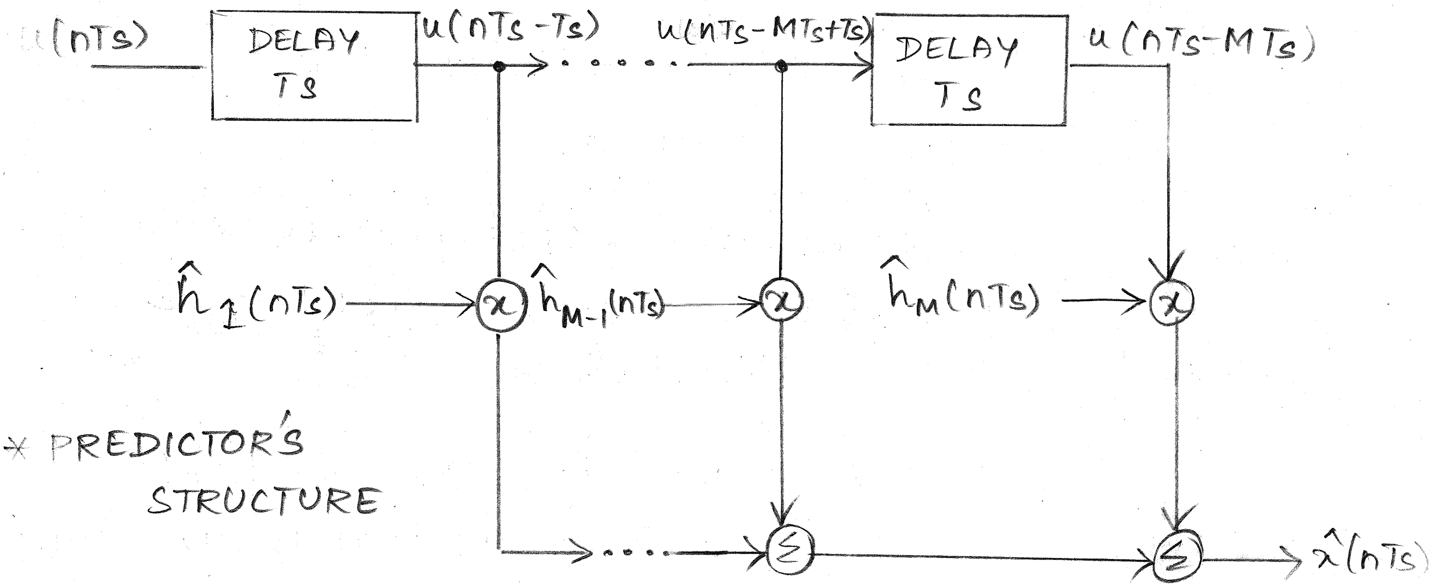
$$\hat{h}_k(nTs+Ts) = \hat{h}_k(nTs) + \mu y(nTs) u(nTs-KTs)$$

where $k = 1, 2, \ldots, M$

$\mu \rightarrow$ adaptation constant.

For initial conditions, we set all of the predictor coefficients equal to zero at $n=0$. The correction term in the update consists of the product $y(nT_s)\,u(nT_s - kT_s)$ scaled by the adaptation constant μ.

When the value of μ is small, the correction term will decrease with the number of iterations n.



* PREDICTOR'S STRUCTURE

ADAPTIVE SUB-BAND CODING:

PCM and ADPCM are both time-domain coders in that the speech signal is processed in the time-domain as a single full-band signal. Here we describe a frequency-domain coder, in which the speech signal is divided into a number of sub-bands and each one is encoded separately. The coder is capable of digitizing speech at a rate of 16kbits/sec.

To accomplish this performance, it exploits the quasi-periodic nature of voiced speech and a characteristic of the hearing mechanism known as noise masking.

Periodicity of voiced speech manifests itself in the fact that people speak with a characteristic pitch frequency. This periodicity permits pitch prediction.

Reduction in the level of prediction error requires quantization.

The number of bits per sample that needs to be transmitted is thereby greatly reduced. The number of bits per sample can be reduced further by making use of the noise-masking phenomenon.

That is, the human ear does not perceive noise in a given frequency band if the noise is about 15 dB below the signal level in that band.

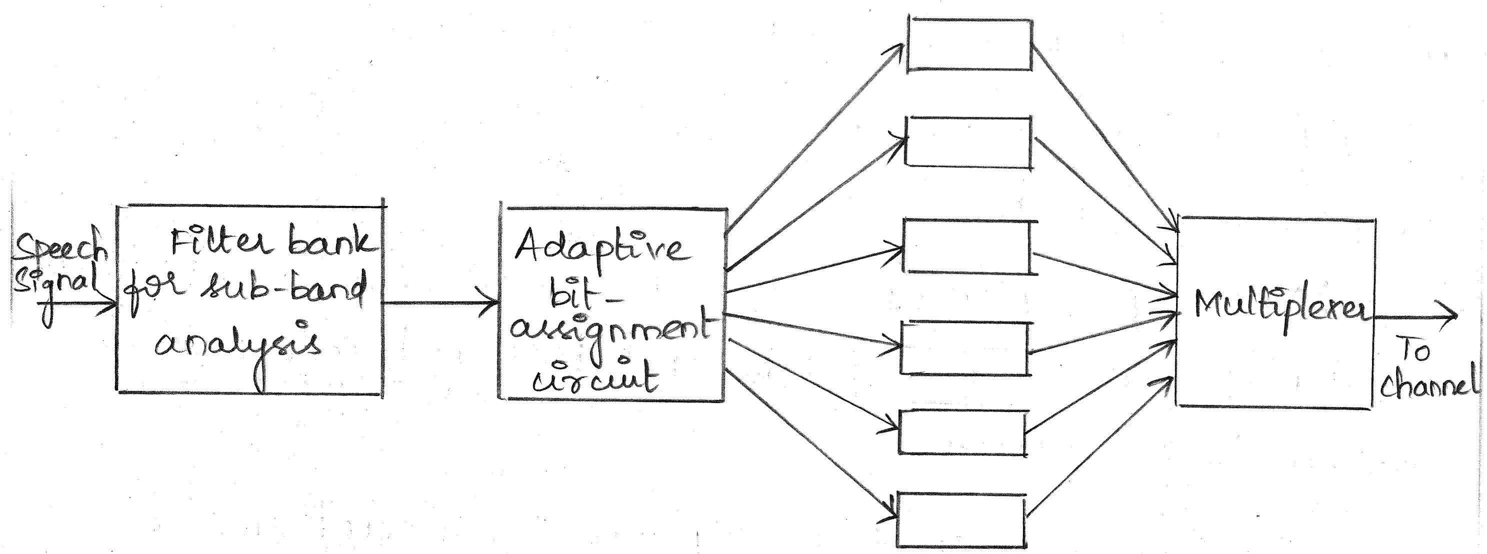
This means that a relatively large coding error can be tolerated near formants and that the coding rate can be correspondingly reduced. In speech production, the formants are the resonance frequencies of the vocal tract tube. The formants depend on the shape and dimension of the vocal tract.

In **Adaptive Sub-band coding** (ASBC), noise shaping is done by adaptive bit assignment. In particular, the number of bits used to encode each sub-band is varied dynamically and shared with other sub-bands.

Sub-bands with little or no-energy may not be encoding at all.

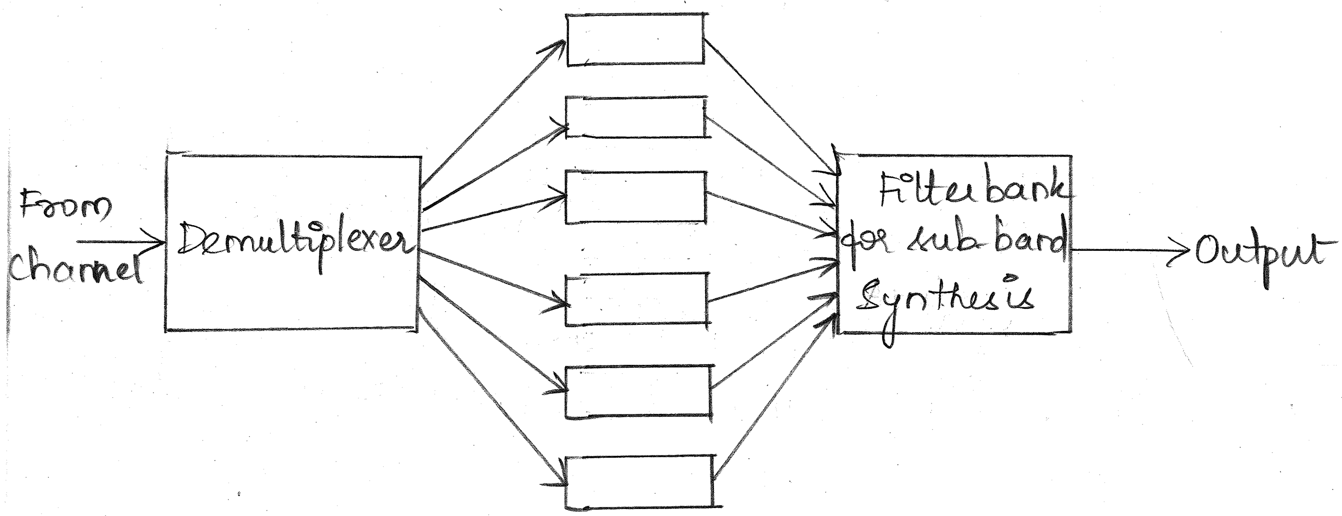
* TRANSMITTER

DPCM encoders



* RECEIVER

DPCM decoders



The speech band is divided into a number of contiguous band by a bank of bandpass filters (4 to 8).

The output of each band pass filter is frequency translated.

It is then sampled at a rate slightly higher than its Nyquist rate (twice the width of the pertinent sub-band) and then digitally encoded by using an ADPCM with fixed prediction. A specific coding strategy is employed for each sub-band.

More representation levels are used for lower frequency bands where pitch and formant information have to be preserved.

If high frequency energy is dominant in the i/p speech signal, the scheme automatically assigns a larger number of representation levels to the higher frequency components of the input.

Bit assignment information is transmitted to the receiver, enabling it to decode the sub-band signals individually and modulate them back to their original locations in the frequency band.

Finally, they are summed to produce an output signal that provides a close replica of the original speech signal.

The complexity of a 16 kb/s adaptive sub-band coder is typically 100 times that of a 64 kb/s PCM coder for about the same reproduction quality.

Linear Predictive Voice coders (LPVC).

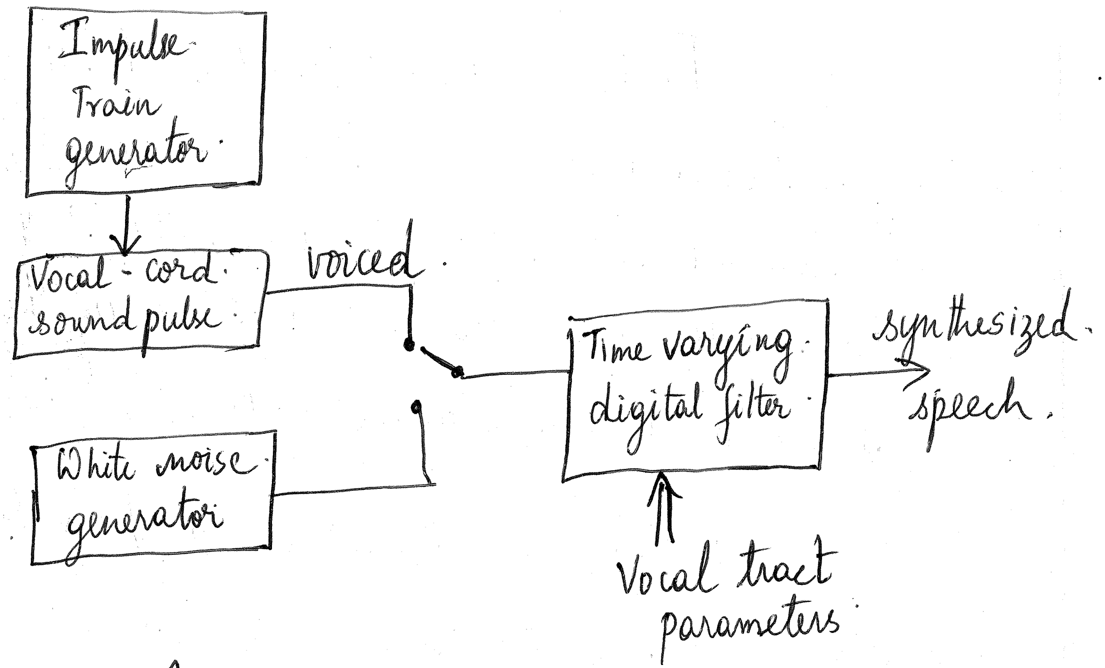
To transmit a voice signal, it is not necessary to transmit the entire speech wave. Instead only the necessary limited information from which the speech can be regenerated needs to be transmitted. This is the basic idea behind linear predictive coding (LPC).

LPC can operate at lower bit rates [1.2 to 2.4 Kbps] But the reproduced voice has a artificial quality and is termed as synthetic voice. Basically used for military applications.

Voice Model / Speech Synthesizer.

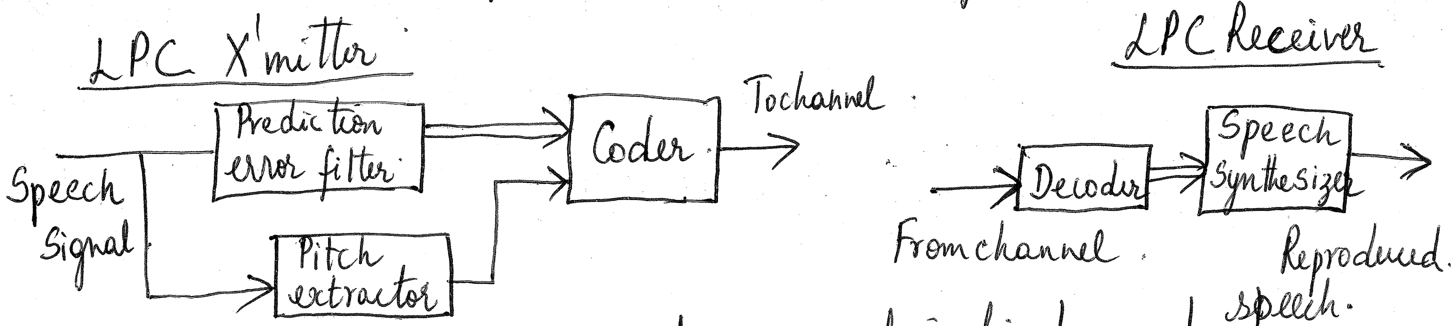
The speech consists of a sequence of voiced and unvoiced sounds. The voiced sounds are generated by the vibrations of the vocal cords. The unvoiced sounds are generated when a speaker pronounces letters such as 's', 'f', 'p' etc. where the sounds are produced by expelling air through lips & teeth.

In speech synthesis, the voiced sounds are simulated by an impulse generator which operates on the fundamental freq of vibration of the vocal cords. The unvoiced sounds are generated by a noise source. The filter represents the effects of generated sounds of the mouth, throat and nasal passages of the speaker.



LPC transmitter & Receiver :-

The input signal is analysed as blocks which have a length of 10-30ms. The analysis results in the foll parameters a) Prediction error filter co-efficients b) Voiced / unvoiced parameters c) Pitch period.



The receiver performs decoding operation first and then it synthesizes the speech signal. The O/P of the synthesis filter is an artificial quality synthezised speech signal.