CHAPTER – 2
SPEECH CODING AND LINEAR PREDICTION ANALYSIS

2.1 OVERVIEW OF SPEECH CODING METHODS

2.1.1 Introduction

The main objective of compressing a digital signal is to represent information associated with the signal as economically as possible whilst retaining parameters sufficient to reconstruct the original signal. Reduction of data storage space or digital transmission rate should be balanced with the maximisation of synthesised signal quality, which is to preserve its intelligibility and naturalness for speech signals, whilst eliminating redundant signal information.

Numerous methods of speech coding have been developed to achieve the goals stated above. The objective of the research is to improve the efficiency in compression and signal quality. There has been a wide range of research and numerous publications regarding the performance of digital speech coding in real-life application where undesirable noise is introduced to the system. However, this research is focussed on the improvements for LP analysis with MLT-SPIHT Algorithm and Masking methods. Thus the compression methods discussed here are the methods related to LPC design. Here the Warped technique has also been combined with this method for getting better compression ratio.

Linear predictive coding (LPC) is one of the most powerful speech coding techniques. LPC exploits the redundancies of a speech signal by modelling the speech signal as a linear filter, excited by signal called the excitation signal. The excitation signal is also called residual signal. Speech coders process a particular group of samples called a frame or a segment. The speech encoder find the filter coefficients and the excitation signal for each
frame. The filter coefficients are derived in such a way that the energy at the output of the filter for that frame is minimized. This filter is called an LP analysis filter. The speech signal is first filtered through the LP analysis filter. The resulting signal is called residual signal for that particular frame. Actually for the decoder, the inverse of the LP analysis filter acts as the LP synthesis filter, while the residual signal acts as the excitation signal for LP synthesis filter. The process is shown in Fig. 2.1.

![LP analysis and synthesis diagram](image)

**Fig. 2.1: LP analysis and synthesis**

This method has become the leading technique for estimating the basic speech parameters, such as pitch, formant, spectrum, and vocal tract area function, to describe a signal or the system that generated the signal. Linear prediction is also beneficial for representing speech for low bit rate transmission or storage. It provides extremely accurate estimates of the speech parameters and is highly efficient in computational matters. Linear prediction is widely used in speech processing to model the speech source. A $p^{th}$ order all-pole filter is found, such that when excited by impulse, its output has a power spectrum that matches the spectrum of a given speech signal. Normally it is required that the autocorrelation of the output signal must be equivalent to the autocorrelation of the original speech signal [12]. If the modelled system is truly a $p^{th}$ order all-pole system excited by one impulse, the model derived during the linear prediction process will be the same as the original system. Voiced speech is represented by a periodic signal exciting the system and interference between successive periods can cause a derivation from the desired response.
The fundamental idea behind linear prediction is that a system sample can be approximated as a linear sum of the squared differences between the actual speech samples and the linearly predicted ones, a unique, distinct set of predictor coefficients can be determined. The predictor coefficients are the weighting coefficients used in the linear combinations. Linear prediction provides a robust, reliable, and accurate method for estimating the parameters that characterize the linear, time-varying system of speech. The common, set of linear prediction analysis techniques is usually referred to as linear predictive coding (LPC). In the usual applications of LPC, for the analysis of speech, the same parameters are used in both the voiced and unvoiced cases.

LPC-10 is an early LPC design that employees the use of fixed excitation signals (Section 2.1.2). The input signal may also be driven by a string of impulses, which is provided by an excitation generator. This LPC method is commonly referred to as the Multipulse Linear Predictive Coding (Section 2.3), which led to the development of the Code Excited Linear Prediction (CELP) coder.

2.1.2 LPC-10

This method was developed based on the channel vocoder method\(^4\). The vocal tract filter of the input signal is modelled by a single linear filter as oppose to the use of a bank of filters in the channel vocoder. Synthesised speech can be modelled from the input signal using either random noise or periodic pulse generator (as shown in Fig. 2.2).

The 2.4 Kbit US Government Standard LPC-10 is the most widely used standard for this method, where an 8 kHz speech signal is divided into frames of 180 samples (frame length of 22.5 ms). This method has been documented to suffer in noisy environments [18], whilst suffering from poor sound quality due to the use of only two excitation signals.

\(^4\) This method is a conventional analysis-by-synthesis method of speech compression developed in the late 1930’s [Dudley, 1939].
Fig. 2.2: Basic speech synthesis model of the LPC-10 method

2.2 LP ANALYSIS

2.2.1 Background Theory

LP analysis of speech is historically one of the most important and currently the most popular, speech analysis technique for low bit-rate speech coding. It was initially developed in the late 1960's [Atal and Schroeder, 1967] and further studied in the early 1970's with multiple publications ([19] [20] [21]) researching this theory. The theory and algorithms surrounding this matter have matured to the point where they are now an integral part of many real-world adaptive systems.

The generation of each phoneme during speech production is dependent on two factors: source excitation and the vocal tract shape. Modelling these two factors, assumed to be independent of each other, is crucial in modelling the speech production system. Ideally the vocal tract filter is modelled by a discrete time glottal excitation signal. LP analysis in speech coding is aimed at modelling the vocal tract model.

Referring to the name linear prediction itself, this method is based on the theorem that every predicted value is a result of a linear combination of its
past values. The most popular vocal tract model for LP analysis is the autoregressive (AR) model. This analysis is a parametric method designed for discrete-time linear stochastic processes. It is based on a source-filter AR model where the vocal tract filter is modelled to be an all-pole linear filter.

**Fig. 2.3: Speech processing model in LP analysis**

Application of LP analysis on speech signals can be used to estimate the basic speech parameters, such as pitch, formants, spectra and vocal tract area functions, which in turn can be applied for compression purposes. Speech signals can be modelled as shown in Fig. 2.3 as an output of a linear time-varying system excited by pulses that are quasi-periodic (voiced) or random noise (unvoiced).

The distinct advantage of this method is its ability to estimate the important speech parameters with a reasonable degree of accuracy and efficient computational speed[5]. LP parameters, also referred to as LP coefficients, are determined from a finite sequence of samples by minimising the MSE between the original and its predicted signal. An example of a standard open-loop LP model can be seen in Fig. 2.4.
A mathematical representation of the open-loop LP model is given as below:

\[
y(n) = x(n) + \sum_{k=1}^{N} a_k x(n-k)
\]  

(2.1)

where \( N \) is the LP order and \( a_k \) is the LP parameters.

The resultant error \( \{e(n)\} \) between the actual signal \( \{x(n)\} \) and the linearly predicted signal \( \{y(n)\} \) is what is quantised for transmission purposes in signal compression schemes. The filter involved in the determination of \( e(n) \) computes its predicted results from a FIR filter. As the LP model becomes more accurate, the reduction in the quantisation error provides far less distortion in providing the synthesised signal. The difference between the true signal and the synthesised signal is called the LP residual or prediction error.

A more adaptive closed-loop model takes into account the quantised error signal into the all-pole filter equation

\[
\tilde{e}(n) = \tilde{x}(n) + \sum_{k=1}^{N} a_k \tilde{x}(n-k)
\]  

(2.2)

For a closed-loop model, \( \tilde{e}(n) \) is the quantised version of \( e(n) \). This quantised error signal is used at the receiver end as an excitation signal to the LP synthesis filter to compute the synthesised signal \( \tilde{x}(n) \).
Another approach used in LP analysis is the ARMA (autoregressive moving average) model. It is the most accurate approach for modelling the vocal tract [22]. It uses a linear combination of its past outputs in addition to a combination of its present and past inputs. This pole-zero model can only be derived by solving a set of non-linear equations. Obviously it is computationally more efficient to solve only one set of linear equations in, an all-pole model. This is the reason why the all-pole (AR) model is the most commonly used model in LP analysis. The zeros, which arise in unvoiced and nasal sounds, are thus approximately modelled by poles.

It has also been determined that the human perception system is more sensitive to spectral poles than zeros, which is another reason why the AR model is popular for use in speech coding [23]. This leads to the most obvious deficiency of this model where it assumes speech spectra to be a perfect all-pole model without any zero present. Nevertheless, the information gathered by the all-pole filters has been deemed quite successful in predicting a sample as a weighted sum of past samples.

2.2.2 Limitations of the LP Analysis

Most signals in real life show a non-stationary behaviour. The main assumption involved in LP analysis is that for a narrow finite time frame the signal behaviour is stationary. This is still known to be the main disadvantage of the LP analysis method.

The methodology of the LP analysis involves describing large records of data by a uniform set of parameters containing its significant process information. Hence, in speech coding, a set of information computed from LP analysis (LP parameters) may represent a considerably larger set of speech data. However computing a set of LP parameters that accurately model the vocal tract, filter would be compromised by its order of analysis (LP order). Clearly a higher LP order would constitute to a more accurate LP analysis.
Another limitation of LP analysis is introduced by data windowing. The choice of windowing function will always present a trade-off between time and frequency resolution. Despite these restrictions, spectral estimation via LP analysis remains very popular because it still provides a very good frequency and time resolution, and its ease of application for signal compression.

2.2.3 Spectral Analysis

The human auditory system is known to perform spectral analysis upon speech signal, hence the original motivation in analysing speech in its frequency domain. The vocal tract produces signals that are more precise and consistent in frequency domain than in time domain.

Spectral analysis applied for use in speech coding examines the behaviour of speech mainly in its frequency domain by determining relative magnitudes of the different harmonics of the speech signal. The spectral envelope estimate of a power spectrum calculated using LP analysis for use in LPG is referred to as the LPG spectrum. The LPG spectrum offers a concise representation of important signal properties, which largely simplifies the control of synthesis models.

In processing a digital signal in its frequency domain, the quality of a synthetic signal relies heavily on how well its LPG spectrum is estimated. Fulfilling the spectral envelope properties is the main requirement in developing a spectrum estimation method. These properties are listed below:

- **Robust** - Spectral envelopes should maintain its general shape and characteristics when introduced to varying environments.

- **Envelope shape fitting** - Envelope should match the shape associated with the spectral peaks as close fitting as possible, following the link between the sinusoidal peaks,
Smoothness - General shape of the signal magnitude distribution over frequency should be easily achieved with no sudden fluctuation in the envelope.

Acclimatisation - Spectral envelope should accurately follow the sudden variations between two consecutive short-time spectral segments.

The importance in performing an accurate analysis of the power spectral can be seen via the rigorous research devoted to this topic in general [24] with regards to its robustness and most importantly in the study of the AR model [25][26] and discrete Fourier transform [27].

In order to attain specific characteristics of the power spectrum, each frame of a speech signal can be parameterised. The source-filter model of speech production is generally applied as a theoretical model in the speech processing analysis. It is used to model the physiology of the human speech system. By modelling an all-pole filter on the resonances of the speech spectrum, filter coefficients can be obtained. These filter coefficients are what is used to estimate the power spectrum of a speech frame. These coefficients can be further quantised via a conversion to spectral parameters for applications in signal compression. This analysis is achieved via the Fourier analysis.

2.2.4 Fourier Analysis

As developed by Jean Baptiste Joseph Fourier in 1807, the underlying theory behind the development of the Fourier analysis is that a set of sinusoidal waves at separate frequency locations can be used to represent a discrete time length of a speech signal. The reverse process has been shown to be true in a mathematical sense.

It is very common for information to be encoded in the sinusoidal waveforms that form a signal. Specifically for spectral analysis, the shape of the waveform in its time domain is not important, as the key spectral
information consist of its amplitude, phase and frequency of the sinusoidal representations. Any waveform that is assumed to be periodic can be analysed as a combination of the above mentioned harmonically related exponents. The Fourier transform of speech signals provide both spectral magnitude and phase with respect to its frequency. The importance of the Fourier transform in LP analysis is in the design of the power spectrum, whilst the phase spectrum is relatively unimportant perceptually [28].

In LP analysis, a short-time Fourier transform is used to represent the time-varying properties of a waveform in the frequency domain. The Fourier transform is defined as follows,

$$X_k(f) = \sum_{n=-\infty}^{\infty} w(k-n)x(n)e^{-j2\pi fn} \quad (2.3)$$

The windows function w(k-n) is a real window sequence used to isolate the portion of the input sequence that will be analysed at a particular time index k.

An ideal window function would acquire a frequency response with a narrow main lobe, which increases resolution, and no side lobes, which dictates the frequency leakage. The rectangular window function separates the signal into finite-sized frames without introducing any weighting. This introduces oscillation at the points of discontinuity, known as the Gibbs phenomenon.

Many window functions have been generated to improve upon the basic rectangular window design, such as hamming, hanning, bartlett, blackman, Kaiser, etc., each having different specification with regards to its frequency response. In this research work, LP analysis has been performed on frames weighted with the hamming window. This window, w(n) given in equation (2.4) has been chosen as it provides a good balance between its mainlobe width and sidelobe attenuation.
\[ w(n) = \begin{cases} 0.54 - 0.46 \cos \left( \frac{2\pi n}{N} \right), & 0 \leq n \leq N - 1 \\ 0; & \text{otherwise} \end{cases} \] (2.4)

The hamming window is also deemed to be adequate in determining the accuracy for approximating the transfer function of the vocal tract. This is a crucial aspect when calculating reflection coefficients for quantisation purposes.

In speech coding, a frame length of 20 to 30 ms is commonly chosen for LP analysis. Speech samples are assumed to be stationary for that period of time. This introduces the use of Discrete Fourier Transform (DFT) for applications in discrete systems.

The DFT is a widely used analogy for time-to-frequency transformation and is the central algorithm in most spectrum analysis systems. It is defined in mathematical terms as follows,

\[ X(f) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi fn} \] (2.5)

Large discrete frame lengths give poor time resolution but good frequency resolution. The DFT bandwidth is chosen to be 50 Hz. Hence an 16 kHz discrete signal constitutes to a frame length (N) of 160 samples.

DFT coefficients computed from a finite duration of samples are values of the z-transform at periodically spaced locations around the unity circle. It constitutes a unique representation of that particular sequence. Although the DFT and Inverse- DFT (IDFT) relations are developed based on its periodic sequences, it does have the ability to represent a finite duration of samples.

Application of DFT in real-world systems however is still computationally costly, especially for a long stream of discrete signals. This brought forth the development of the Fast Fourier Transform (FFT). It is a form of Fourier transform that is aimed purely to reduce the computation
time required for DFT. The most widely used FFT algorithm is the radix-2 decimation-in-time and decimation-in-frequency method [8].

Referring to Fig. 1.3, the frequency components in the power spectrum computed via the Fourier analysis can be observed. The lowest frequency component of the frequency domain is known as the fundamental frequency, while the others are known as harmonic frequencies. Usually the harmonics are represented to be harmonic with a number corresponding to the multiple of the fundamental frequency. In an ideal periodic signal, the harmonics would be located exactly at fundamental frequencies apart from each other. In a musical sense, the harmonics numbered by powers of two represents the octave levels.

Non-periodic waveforms may also be analysed by Fourier means which results in a complex integral. Disregarding phase, a spectral analysis can still be generated where the frequency components are not exact multiples of the fundamental frequencies. This would create difficulties during LP analysis, as the further away the spectral analysis resembles a harmonic model, the harder it is to perceive pitch in the signal.

When complex waveforms are introduced (for example, speech signals affected by noise), the Fourier analysis normally gives a statistical answer. It is possible to locate a particular frequency component over a large time frame; however allocating constant amplitude would remain a problem. The probability curve is same as it is normally obtained for the spectral analysis. A narrow-band noise will appear to be similar to a pitched tone, but for the complete bandwidth the distinction of the pitch tends to disappear as shown in Fig. 1.3.

2.2.5 LPC Spectrum

Quality of a decoded synthetic signal via LP analysis depends heavily on how well the spectral envelope is estimated. The accuracy of the estimation defines how sufficient the signal properties are captured. The power spectrum,
also known as Power Spectral Density (PSD), is plainly a mathematical representation of amount of power as a function of frequency. In a mathematical sense, the PSD (symbolically defined as $P_{xx}(f)$) is defined.

$$P_{xx}(f) = \sum_{k=-\infty}^{\infty} R_{xx}(k)e^{-j2\pi f k} \quad (2.6)$$

where $R_{xx}(k)$ is the autocorrelation function of an input signal. As can be seen, PSD is defined as a Fourier transform of the autocorrelation sequence in its time series. In LP analysis, the PSD is normally computed using the periodogram method of spectrum estimation.

### 2.2.6 Periodogram Spectrum

Periodogram method of spectrum estimation, or also known as the sample spectrum method, is the classic way of estimating the power spectrum. It was originally designed to observe the hidden periodicities in the data.

This method is described using a direct computation of the squared constant multiplier of the Fourier transform of its time series. The periodogram spectra $Per_{xx}(f)$ is based on a direct approach through a Fourier transform on a frame of data, generally performed through the Fast Fourier Transform (FFT).

$$Per_{xx}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n)e^{-j2\pi fn} \right|^2 = \frac{1}{N} \left| X(f) \right|^2 \quad (2.7)$$

Periodogram spectrum is a direct realisation of the PSD, hence as $N \to \infty$, then $Per_{xx}(f)$ should ideally resemble $P_{xx}(f)$. However this is not the case, where the periodogram is not a true representation of the PSD. Spectral leakage is encountered on the spectrum, which is due to the limitation of short-time analysis on finite-length data. However, as explained by Kay and Marple [29], a statistically consistent spectrum using this method can only be reached by separating the data sequence into smoothed segments. Leakage effects that occur due to data windowing can be minimised through the selection of non-uniform weighted windows.
2.2.7 Conventional LP Analysis Methods

There are a number of methods that can be used in LP analysis to obtain the LP parameters. These methods are the lattice, covariance and, the most commonly used auto correlation methods.

**Lattice Method**

This method incorporates the application of a forward and backward predictor. Although this method guarantees a stable filter, a considerable amount of storage space is needed to fulfil the computation process. A popular implementation of this method for LP analysis by Burg, also referred to as the Burg method, has a disadvantage in its line splitting tendencies and the dependency of the peak locations on phase. There is also the Recursive Maximum Likelihood Estimation (RMLE) method, which is similar to the Burg method in a way that it maximises its likelihood functions as opposed to minimising its prediction error [4].

**Covariance Method**

Covariance method is an algorithm specified for frame-to-frame basis estimation, with sets of data limited to 0≤n≤N-1 interval. This method windows the residual error signal rather than the actual signal in a way to minimise its error. For an LP order of p, the signal is assumed to be known for the set of values -p≤n≤N-1. No values outside this interval are needed for computation. This results in a covariance matrix solution that would be symmetric, however it should be noted that it is not a Toeplitz matrix.

Although this method guarantees stability in most cases, the inversion of the covariance matrix is computationally expensive. This is the reason behind the selection of the autocorrelation method for LP analysis in this dissertation.

**Autocorrelation Method**

Autocorrelation method of LP analysis, or also referred to as the Wiener-Khintchine theorem, is the most popular method of short-term LP
analysis. This method provides the most computationally efficient manner in determining the LP parameters with guaranteed stability. It takes advantage of the Toeplitz property possessed by the autocorrelation matrix.

The autocorrelation function of a signal is the inverse Fourier transform of its power spectrum. This function represents the correlation between adjacent signal samples. It measures the similarities between the current signal \( x(n) \) with its past values as a function of time. The autocorrelation function \( [R_{xx}(k)] \) is defined as

\[
R_{xx}(k) = \varepsilon[x(n)x(n+k)]
\]  

(2.8)

where \( \varepsilon \) is the expectation operator.

In matrix form, the autocorrelation function can be represented as follows,

\[
\begin{bmatrix}
R_{xx}(0) & R_{xx}(1) & R_{xx}(2) & \ldots & R_{xx}(p-1) \\
R_{xx}(1) & R_{xx}(0) & R_{xx}(1) & \ldots & R_{xx}(p-2) \\
R_{xx}(2) & R_{xx}(1) & R_{xx}(0) & \ldots & R_{xx}(p-3) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
R_{xx}(p-1) & R_{xx}(p-2) & R_{xx}(p-3) & \ldots & R_{xx}(0)
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\vdots \\
\alpha_p
\end{bmatrix}
= 
\begin{bmatrix}
R_{xx}(1) \\
R_{xx}(2) \\
R_{xx}(3) \\
\vdots \\
R_{xx}(p)
\end{bmatrix}
\]  

(2.9)

where \( p \) symbolizes the LP order and \( \alpha_k \) the \( k \)-th LP parameter. The series of linear equations above is commonly referred to as the Yule-Walker equation. The minimum mean square prediction error \( E \) can thus be obtained as follows,

\[
E = R_{xx}(0) - \sum_{k=1}^{p} \alpha_k R_{xx}(k)
\]  

(2.10)

Assuming that the process is stationary, the corresponding autocorrelation vector is then time-invariant. However, as the process is never entirely stationary, the Yule-Walker equation can only be true for assumed stationary processes, which is achieved by determining the process samples outside the framing window to be equal to zero (set of input values are segmented). In general, the selection of framing windows would also determine the estimation accuracy.
The autocorrelation function preserves information regarding the signal harmonics, formant amplitudes and periodicities, whilst ignoring phase. Applications of this process are commonly used for pitch detection, voiced and unvoiced speech determination and most importantly linear predictions.

2.2.8 Determination of the LP Parameters

The LP analysis produces a spectrum representative by a time-varying all pole digital filter, whose transfer function is

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

(2.11)

Where G represents the gain parameter (determined by its prediction error), p the LP order and $a_k$ the LP parameters.

It is necessary to limit the LP analysis to short-time blocks of the filtered signal sequences. The order of the LP analysis is normally dictated by the signal’s sampling frequency. For a speech signal with sampling frequency of 8 kHz, with each 1 kHz proposed one pole and two poles allocated for the beginning and end of each analysis frame, a 10th order LP is normally used ($a_1, a_2, \ldots, a_{10}$), with $a_0=1.0$ (representing the weights for the current sample as can be seen in (2.11))

The approach taken to solve the Yule-Walker equation (2.9) is the Levinson-Durbin algorithm used for the autocorrelation method. This AR method of spectral estimation has also been referred to as the maximum entropy spectral estimation method. This algorithm exploits the Toeplitz property of the autocorrelation matrix $R_x$. From (2.8), the autocorrelation function can be calculated from input signal $x$ with frame size $M$

$$R_x(k) = \frac{1}{M-k} \sum_{i=1}^{M-k} x(i)x(i+k)$$

(2.12)
Initially,
\[
a_i(1) = \frac{R_{xx}(1)}{R_{xx}(0)}
\]
and
\[
P_t = R_{xx}(0)x\left|1 - |a_i(1)|^2\right|^2
\]
Thus for the next steps, for \(m=2,3,\ldots, N\),
\[
a_m(m) = \frac{-R_{xx}(m) + \sum_{i=2}^{N} (a_{m-1}(i)xR_{xx}(m-i))}{P_{m-1}}
\]
\[
P_m = P_{m-1}x\left|1 - |a_m(m)|^2\right|^2
\]
where \(N\) is the LP order, and solving for the LP parameters for \((i=1,2,\ldots, m-1)\),
\[
a_m(i) = a_{m-1}(i) + a_m(m)x a_{m-1}(m-i)
\]
\[
a(i) = a_{N}(i)
\]
LP analysis treats problems of adaptive linear systems and explains it to sets of measurements or observations. Such observations in this case are used to analyse and estimate the power spectrum envelope of a given data set. This analysis would remove neighbouring sample correlations present in the signal, which ultimately can be used to estimate the spectral envelope.

2.3 MULTIPULSE LPC

In this method of LPC, a stream of signals is modelled as the output of an all-pole filter, driven by an excitation function. As the title of this compression scheme indicates, the excitation function consists of a pulse sequence containing a small number of pulses, defined by their location and amplitude. Atal and Remde first introduced this multipulse excitation approach of LPC [30].
A sequence of excitation pulses is computed for each frame of the signal. Increasing the number of excitation pulses would gradually improve the quality of the synthesised signal. However, a minimised number of pulses will be needed to ensure an acceptable synthesised signal quality with an optimum compression ratio. It has been shown that only a small number of pulses (4 to 1.0 pulses) for each sub-frame are enough to produce an acceptable synthesised signal. Commonly, a setting of 8 pulses per cluster of 64 samples is sufficient in generating the desired input or residual signal with minimised distortions [31].

Input signal

\[ S(n) \]

\[ Y(n) \]

\[ \text{S(n)- Input Signal; Y(n)- Output signal} \]

**Fig. 2.5: Block diagram of the Multipulse Coder**

The main focus in the design of this compression scheme is in determining the location and amplitude of the pulse. These pulses should closely represent the actual signal after being fed through a weighting filter. Excitations for the all-pole filter (or pole-zero filter, depending on its application) are created via an excitation generator that produces a sequence of pulses at certain locations and amplitudes. An LP synthesis filter is used to produce the synthetic signal waveform from the pulses.
Using an analysis-by-synthesis approach, the pulse locations and amplitudes are determined by minimising the weighted mean-squared error created by the difference between the original and the LP synthesis filtered signal. Each pulse determination process assumes that previous pulse amplitudes and locations are constant throughout the search. Although this may not be the most accurate manner in calculating the pulses, however it is deemed computationally efficient without much degradation of accuracy. For m number of pulses and a frame length of N, an exhaustive search, which involves calculating every possibility of the pulses simultaneously, would need approximately $N^m$ points of computation (depending on estimation methodology) in comparison to the chosen manner, which would only need $N \times m$ computation points.

### 2.3.1 Pulse Computation

The information content of each pulse contains two values, its amplitude ($\beta_k$) and location (denoted by its positions in the frame). Each pulse location number, referred to as $n_k$ for every $k^{th}$ pulse, can be seen in (2.19). The combination of pulses can be collectively defined as

$$u(n) = \sum_{k=0}^{m-1} \beta_k \delta(n-n_k)$$

where $m$ is the number of pulses and $\delta_n$ is the Kronecker delta. Referring to Fig. 2.5, the signal $y(n)$ is obtained by weighting the pulse $u(n)$ with an impulse response $h(n)$, such that from

$$y(n) = u(n) \times h(n)$$

we get

$$y(n) = \sum_{k=0}^{m-1} \beta_k h(n-n_k)$$
Observing from Singhal and Atal [32], the squared error \( E \) must be minimised with respect to the pulse amplitudes and locations. Optimum pulse locations are determined by calculating the minimum error for all the possible locations and its optimum amplitudes in a set sub-frame [33].

\[
E = \sum_{n=0}^{N-1} |s(n) - \beta_k h(n-n_k)|^2
\]

for \( N \) denoting length of the sub-frame. Solving for

\[
\frac{\partial E}{\partial \beta_k} = 0
\]

we get

\[
\beta_k = \frac{\sum_{n=0}^{N-1} s(n) h(n-n_k)}{\sum_{n=0}^{N-1} [h(n-n_k)]^2}
\]

Substituting \( \beta_k \) back into \( E \),

\[
E = \sum_{n=0}^{N-1} S^2(n) = \frac{\sum_{n=0}^{N-1} s(n) h(n-n_k)}{\sum_{n=0}^{N-1} [h(n-n_k)]^2}
\]

As \( s(n) \) is the original signal, the second term of the equation would then have to be maximized. This introduces the autocorrelation \( (\alpha) \) and cross-correlation \( (c) \) constants, where

\[
\alpha(n_k) = \sum_{n=0}^{N-1} h^2(n-n_k)
\]

and

\[
c(n_k) = \sum_{n=0}^{N-1} s(n) h(n-n_k)
\]

2.3.2 Pitch Prediction

In linear prediction, there is a period of underlying harmonic called the pitch period. In general, a transmitter system needs to estimate these pitch
prediction coefficients in order to obtain a better representation of the signal. This information would also need to be transmitted together with the pulse data[34].

It has been well understood that the human ear is highly sensitive to pitch errors [4]. This has brought forth the development of more accurate pitch detection algorithms. The technique used here employs the autocorrelation (2.26) and cross-correlation (2.27) functions. This autocorrelation function provides a suitable approach in predicting the pitch period of the signal. This function should have a maximum value at each pitch period point. A pre-determined maximum coefficient is needed to help establish the pitch coefficient. The pitch coefficient is deemed to be reached when the auto correlation value is larger than the set threshold.

2.4 PERFORMANCE OF THE LPC CODER

In measuring the quality of a synthesised signal, the mathematical representation of signal-to-noise ratio (SNR) is used.

\[
\text{SNR} = 10 \log_{10} \left[ \frac{\sum_{i=0}^{M-1} x(i)^2}{\sum_{i=0}^{M-1} (x(i) - y(i))^2} \right]
\]  

(2.28)

Where \(x_i\) and \(y_i\) are the samples of the original and synthesised signals respectively, and \(M\) is the signal length. Signal quality would then naturally improve as the RMS difference is minimised.

In speech coding SNR is a poor estimate of speech quality especially when a wide range of speech direction is introduced. SNR is not specifically designed to model the subjective attributes of a speech signal; hence determining its speech quality would be unreliable.

In (2.28), the time domain error of a sequence of speech is weighted equally. This does not necessarily correspond well with the behaviour of speech where its energy varies with time. A quick solution to this problem is
by applying the SNR calculation to speech on a frame-by-frame basis, thus
weighting each short-time speech frame independently and then averaging the
overall ratio. This Segmental-SNR can be defined,

\[
SNR_{\text{seg}} = \sum_{j=0}^{n-1} 10 \log_{10} \left[ \frac{\sum_{i=(n-M+1)}^{n} x(i)^2}{\sum_{i=(n-M+1)}^{n} (x(i) - y(i))^2} \right]
\]

(2.29)

where the \( n_j \)'s are the end points of each frame (frame length \( M \)) and \( N \) is
number of frames.

\( SNR_{\text{seg}} \) allows an objective measure of speech quality by assigning equal
weight to loud (large energy) and soft (small energy) portions of speech. Normally to avoid unnecessary large distortions in speech, silence needs to be
identified and excluded (most likely located at the beginning and end of each
sample sequence). It has also been widely understood that thresholds can be
set at its extreme ends of the scale (e.g. ratios below 0 dB and above 35 dB are
left out). This is done because the ratios outside the set threshold limits are
regarded as not offering any contribution to the overall speech quality [35].

It should be pointed out here that the SNR, and \( SNR_{\text{seg}} \) respectively, is
only a mathematical ratio in comparing the performance of a particular speech
coder. Although it is a reasonably reliable mathematical representation of
signal quality, it is still possible to have two synthesised speech samples with
one sample having a worse SNR but better sound quality than the other. In
the end, human subjects would still be needed to gather an subjective quality
measurement, especially in producing synthesised speech.

This introduces the Mean Opinion Score (MOS), which is an average
numerical opinion score for a set of untrained subjects. The MOS is the most
commonly used subjective measure for determining signal quality. It uses
human subjects to determine the quality of speech using a predetermined
setting of 1 to 5, with 5 representing excellent quality.
2.5 CONCLUSION

In this chapter, the importance of speech coding and the detail of Linear Prediction Analysis method has been discussed. The various methods of LPC like Lattice method, Auto correlation method, spectral estimation method are discussed in this chapter. The performance factors of the Speech Compression like Signal to Noise ratio, Segmental Signal to Noise ratio, Mean Opinion Score are also discussed. These performance factors have been calculated and analyzed in the following chapters.