ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

LINEAR PREDICTIVE CODING

Mr. Mohit Narayanbhai Raja*, Miss. Priyanka Richhpal Jangid, Dr. Sanjay M. Gulhane
[*] Electronics and Telecommunication Jawaharlal Darda Institute of Engineering and Technology Yavatmal, India
Electronics and Telecommunication Jawaharlal Darda Institute of Engineering and Technology Yavatmal, India
Electronics and Telecommunication Jawaharlal Darda Institute of Engineering and Technology Yavatmal, India

ABSTRACT
Linear predictive coding (LPC) is defined as a digital method for encoding an analog signal in which a particular value is predicted by a linear function of the past values of the signal. It was first proposed as a method for encoding human speech by the United States Department of Defense in federal standard 1015, published in 1984. Human speech is produced in the vocal tract which can be approximated as a variable diameter tube. The linear predictive coding (LPC) model is based on a mathematical approximation of the vocal tract represented by this tube of a varying diameter.
At a Particular time, t, the speech sample s(t) is represented as a linear sum of the p previous samples. The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples. Under normal circumstances, speech is sampled at 8000 samples/second with 8 bits used to represent each sample. This provides a rate of 64000 bits/second. Linear predictive coding reduces this to 2400 bits/second. At this reduced rate the speech has a distinctive synthetic sound and there is a noticeable loss of quality. However, the speech is still audible and it can still be easily understood. Since there is information loss in linear predictive coding, it is a lossy form of compression.

KEYWORDS:

INTRODUCTION
Linear Predictive Coding (LPC) is one of the methods of compression that models the process of speech production. Particularly, LPC models this process as a linear sum of earlier samples using a digital filter inputting an excitement signal. An equivalent explanation is that linear prediction filters attempt to predict future values of the input signal based on past signals. LPC "...models speech as an autoregressive process, and sends the factors of the process as opposed to sending the speech itself". Speech coding or compression is generally conducted with the use of voice coders or vocoders.

Linear predictive coding (LPC) [1] is a tool used mostly in audio signal processing and speech processing for representing speech signal in condensed form, using the information of a linear predictive model. It is one of the most effective speech analysis techniques, and one of the most beneficial methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.

Types of voice coders-
There are two types of voice coders: waveform-following coders and model-base coders.

Waveform following coders:
It will exactly reproduce the original speech signal if no quantization errors occur.

Model-based coders:
It will never exactly reproduce the original speech signal, regardless of the 4 presence of quantization errors, because they use a parametric model of speech production which involves encoding and transmitting the parameters not the signal.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

LPC vocoders and its attributes:
These are considered model-based coders which mean that LPC coding is lossy even if no quantization error occurs. All vocoders, including LPC vocoders, have four essential features: bit rate, delay, complexity, quality. Any voice coder, unconcerned of the algorithm it uses, will have to make tradeoffs among these different attributes.

Bit rate:
The first attribute of vocoders, the bit rate, is used to actuate the degree of compression that a vocoder accomplishes. Uncompressed speech is usually transmitted at 64 kb/s using 8 bits/sample and a rate of 8 kHz for sampling. Part of bit rate below 64 kb/s is considered compression. The linear predictive coder sends speech at a bit rate of 2.4 kb/s, a finest rate of compression.

Delay:
This is another important attribute for vocoders that are involved with the transmission of an encoded speech signal. Vocoders which are involved with the memory of the compressed speech, as against to transmission, are not as concern with delay. The general delay standard to transmitted speech conversations is that any delay that is more than 300 ms is considered unacceptable.

Complexity:
The third attribute of voice coders is the complexity of the algorithm used. The complexity affects the cost and the power of the vocoder. Linear predictive coding due of its high compression rate is very complex and involves executing millions of instructions per second.

Quality:
The final attribute of vocoders is quality. Quality is an instanctive attribute and it builds upon how the speech sounds to a given listener. One of the most common tests for speech quality is the absolute category rating (ACR) test. This test involves subjects being given pairs of sentences and asked to rate them as excellent, good, fair, poor, or bad. Linear predictive coders sacrifice quality in order to achieve a low bit rate and as a result often sound synthetic.

Encoding and Decoding:
The common algorithm for linear predictive coding involves an analysis or encoding part and a synthesis or decoding part.

Encoding:
In the encoding, LPC takes the speech signal in frames of speech and determines the input signal and the coefficients of the filter that will be capable of reproducing the current block of speech. This information is quantized and transmitted.

Decoding:
In the decoding, LPC rebuilds the filter based on the coefficients received. The filter can be thought of as a tube which, when given an input signal, at times to output speech. Added information about the original speech signal is used by the decoder to determine the input or excitation signal that is sent to the filter for synthesis.

LITERATURE REVIEW

Historical Perspective of Linear Predictive Coding:
The history of audio and music compression begin in the 1930s with research into pulse-code modulation (PCM) and PCM coding. Compression of digital audio was in progress in the 1960s by telephone companies who were concerned with the cost of transmission bandwidth.

Origin of LPC:
Linear Predictive Coding's origins begin in the 1970s with first LPC algorithms development. Adaptive Differential Pulse Code Modulation (ADPCM), another method of speech coding, was also first conceived in the 1970s. In 1984, the United States division of Defense produced federal standard 1015 which outlined the particulars of LPC. Extensions of LPC such as Code Excited Linear Predictive (CELP) algorithms and Vector Selectable Excited Linear Predictive (VSELP) algorithms were developed in the mid 1980s and used commercially for audio music coding in the later part of that decade. The 1990s have seen improvements in these earlier algorithms and an increase in compression ratios at given audio quality levels.

Development of first vocoder:
The history of speech coding makes no mention of LPC until the 1970s. However, the history of speech synthesis shows that the beginnings of Linear Predictive Coding occurred 40 years earlier in the late 1930s. The primary vocoder was described by Homer Dudley in 1939 at Bell Laboratories. Vocoder war developed by Dudley, called the Parallel Band pass Vocoder or channel vocoder, for speech analysis and re synthesis. LPC is a descendent of this channel vocoder. The analysis/synthesis system used by Dudley is the scheme of compression that is used in many types of speech compression such as LPC.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Parametric mode:
Analysis/Synthesis schemes are based on the development of a parametric model during the analysis of the original signal which is later used for the synthesis of the source output. The transmitter or sender analyses the original signal and acquires parameters for the model which are sent to the receiver. The recipient then uses the model and the parameters it receives to synthesize an approximation of the original signal. This method of sending the model parameters to the receiver was the earliest form of lossy speech compression.

HUMAN SPEECH PRODUCTION

The process of speech production in humans can be summarized as air being pushed from the lungs, throughout the vocal tract, and out through the mouth to generate speech. In this type of description the lungs can be thought of as the source of the sound and the vocal tract can be thought of as a filter that produces the various types of sounds that make up speech. There are two categories, voiced and unvoiced sounds, that are considered by the linear predictive coder when analyzing and synthesizing speech signals.

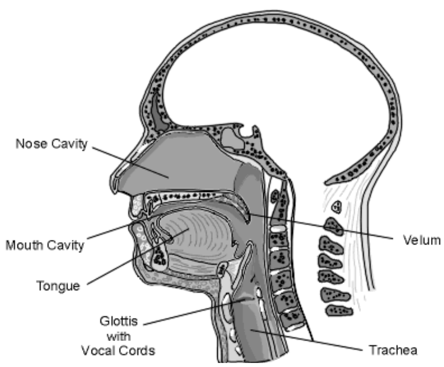

Figure 1: Path of Human Speech Production

Voiced sound:
Voiced sounds are usually vowels and often have high average energy levels and very distinct resonant or formant frequencies. Voiced sounds are produced by air from the lungs being forced over the vocal cords. As a result vibration of vocal track takes place in a somewhat periodically pattern that produces a series of air pulses called glottal pulses. The velocity at which the vocal cords vibrate is what determines the pitch of the sound produced.

Unvoiced sound:
Unvoiced sounds are usually consonants and generally have less energy and higher frequencies then voiced sounds. The construction of unvoiced sound involves air being forced through the vocal tract in a turbulent flow. Through this process the vocal cords do not vibrate, they stay open until the sound is produced. Pitch is an unimportant characteristic of unvoiced speech since there is no vibration of the vocal cords and no glottal pulses.

Components that influence speech production:
The categorization of sounds as voiced or unvoiced is an important consideration in the analysis and synthesis process. In reality, the vibration of the vocal cords, or less of vibration, is one of the key components in the production of different types of sound. Another component that influences speech production is the shape of the vocal tract itself. Different shapes will generate different sounds or significant frequencies. The vocal tract consists of the throat, the tongue, the nose, and the mouth. It is defined as the speech producing path through the vocal organs. This path shapes the frequencies of the vibrating air travelling from it. As a person speaks, the vocal tract is constantly changing shape at a very slow rate to produce different sounds which flow together to create words. A final component that affects the production of sound in humans is the amount of air that originates in the lungs. The air moving from the lungs can be thought of as the source for the vocal tract which acts as a filter by taking in the source and producing speech. The higher the volume of air that goes through the vocal tract, the louder the production of sound.

Source filter model:
The idea of the air from the lungs as a source and the vocal tract as a filter is called the source-filter model for production of sound. The source-filter model is the model that is used in linear predictive coding. It is based on the scheme of separating the source from the filter in the production of sound. This model is used in both the encoding and the decoding of LPC and is derived from a mathematical approximation of the vocal tract represented as a varying diameter tube. The excitation of the air travelling through the vocal tract is the source. When producing voiced sounds this air can be periodic, through vibrating vocal cords or it can be turbulent and casual when producing unvoiced sounds. The encoding process of LPC involves decisive a set of accurate parameters for modeling the vocal tract during the production of a given speech signal. Decoding involves the parameters acquired in the encoding and analysis to build a synthesized version of the original speech signal. LPC not at all transmits any estimates of speech to the receiver; it only sends the model to produce the speech and some

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

indications about what type of sound is being produced.

LPC MODEL

The particular source-filter model used in LPC is known as the Linear predictive coding model. It has two main components: analysis or encoding and synthesis or decoding. The analysis division of LPC involves examining the speech signal and breaking it down into segments or blocks.

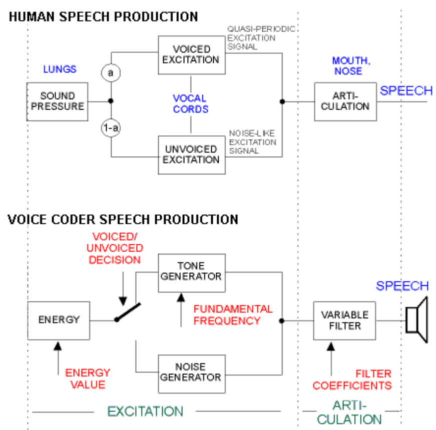

Figure 2: Human vs. Voice Coder Speech Production

LPC analysis is usually conducted by a sender who answers these questions and usually transmits these answers on a receiver. The receiver performs LPC synthesis by using the answers received to build a filter that when provided the correct input source will be able to accurately reproduce the original speech signal. Essentially, LPC synthesis tries to emulate human speech production. Figure 2 demonstrates what parts of the receiver correspond to what parts in the human structure. This diagram is for a general voice or speech coder and is not specific to linear predictive coding. All voice coders lean to model two things: excitation and articulation. Excitation is the kind of sound that is passed into the filter or vocal tract and articulation is the transformation of the excitation signal into speech.

LPC ANALYSIS/ENCODING

Input speech:

According to government standard 1014, in addition known as LPC-10, the input signal is sampled at a rate of 8000 samples per second. This input signal is then broken down up into segments or blocks which are each analyzed and transmitted to the receiver. The 8000 samples in each instant of speech 10 signals are broken into 180 sample segments. This means that every segment represents 22.5 ms of the input speech signal.

Voice/Unvoiced Determination:

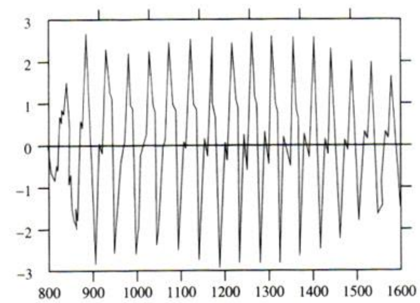

Figure 3: Voiced sound – Letter "e" in the word "test"

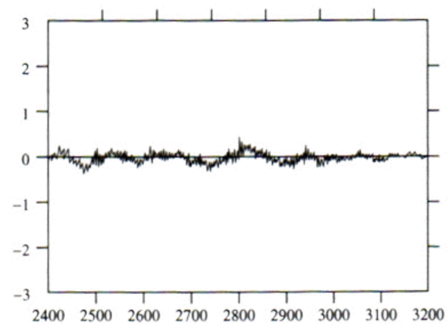

Figure 4: Unvoiced sound – Letter"s" in the word "test"

According to LPC-10 standards, before a speech segment is determined as being voiced or unvoiced it is first passed through a low-pass filter with a bandwidth of 1 kHz. Detecting if a segment is voiced or unvoiced is important because voiced sounds have a different waveform then unvoiced sounds. The differences in the two waveforms make a need for the use of two different input signals for the LPC filter in the synthesis/decoding. One input signal is for voiced sounds and the other is for unvoiced. The LPC encoder notifies the decoder if a signal segment is voiced or unvoiced by sending a single bit.

A sample of voiced speech can be seen in Figure 3 which shows the waveform for the vowel "e" in the word "test". Notice that this waveform has the characteristic large amplitude and distinct frequencies of voiced sounds. Unvoiced sounds are usually non-vowel or consonants sounds and often have very chaotic and random waveforms. Figure 4 shows that these sounds have less energy and therefore smaller amplitudes then voiced sounds. Figure 4 also shows

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

unvoiced sounds have higher frequencies then voiced sounds.

There are two steps in the procedure of determining if a speech segment is voiced or unvoiced.

First step-by amplitude:
The first step is to look at the amplitude of the signal, also called as the energy in the segment. If the amplitude levels are huge then the segment is classified as voiced and if they are small then the segment is considered unvoiced.

Second step-by frequency:
The second step is used to make the final distinction between voiced and unvoiced sounds. This step takes gain of the fact that voiced speech segments have large amplitudes, unvoiced speech segments have very high frequencies, and that the average values of these two types of speech samples is close to zero. These two facts lead to the conclusion that the unvoiced speech waveform must cross the x-axis more often than the waveform of voiced speech. This can evidently be seen to be true in the case of Figure 3 and Figure 4.

Pitch Period Estimation:
Determining if a segment is a voiced or unvoiced sound is not all of the information that is needed by the LPC decoder to accurately reproduce a speech signal. In order to generate an input signal for the LPC filter the decoder also needs another attribute of the current speech segment known as the pitch period. The period for any wave, counting speech signals, can be defined as the time necessary for one wave cycle to completely pass a fixed position.

Determination of pitch period:
It is very computationally intensive to determine the pitch period for a given segment of speech. One type of algorithm takes advantage of the fact that the autocorrelation of a period function, R(k), will have a maximum when k is equivalent to the pitch period. These algorithms generally detect a maximum value by checking the autocorrelation value against a threshold value. One difficulty with algorithms that use autocorrelation is that the validity of their results is susceptible to interference as a result of other resonances in the vocal tract. When interfering occurs the algorithm cannot guarantee accurate results. One more problem with autocorrelation algorithms occurs because voiced speech is not completely periodic. This means that the maximum will be lower than it should be for a true periodic signal. LPC-10 doesn't use an algorithm with autocorrelation, rather than it uses an algorithm called average magnitude difference function (AMDF) which is defined as

$$\mathrm{AMDF}(P) = \frac{1}{N} \sum_{i=k_p+1}^{k_p+N} |y_i + y_{i-P}|$$

Since the pitch period, P, for humans is limited; the AMDF is evaluated for a limited series of the possible pitch period values. Hence, in LPC-10 there is an assumption that the pitch period is between 2.5 and 19.5 ms.

Vocal Tract Filter:
The filter that is used by the decoder to recreate the original input signal is created based on a set of coefficients. Original signal extracts these coefficients during encoding and are transmitted to the receiver for use in decoding. Each speech segment has dissimilar filter coefficients or parameters that it uses to recreate the original sound. Not only are the parameters themselves unusual from segment to segment, but the amount of parameters differing from voiced to unvoiced segment. Voiced segments use 10 parameters to construct the filter while unvoiced sounds use only 4 parameters. A filter having n parameters is referred to as an nth order filter. In order to detect the filter coefficients that best match the current segment being analyzed the encoder attempts to minimize the mean squared error. The mean squared error is given as:

$$e_n^2 = (y_n - \sum_{i=1}^{M} a_i y_{n-i} + G\epsilon_n)^2$$

Where {yn} is the set of speech samples for the current segment and {ai} is the set of coefficients. In order to offer the most accurate coefficients, {ai} is chosen to minimize the average value of en2 for all samples in the segment. The initial step in minimizing the average mean squared error is to take the derivative. Taking the derivative gives a set of M equations. To solve for the filter coefficients E[yn-iyn-j] has to be estimate. There are 2 approaches that can be used for this estimation: autocorrelation and auto covariance. Autocorrelation is widely used.

Transmitting the Parameters
In an uncompressed form, speech is usually transmitted at 64,000 bits/second using 8bits/sample and a rate of 8 kHz for sampling. LPC reduces this rate to 2,400 bits per second by breaking the speech into

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

segments and then sending the voiced/unvoiced data, the pitch period, and the coefficients for the filter that represents the vocal tract for every segment. The input signal used by the filter on the receiver end is determined by the classification of the speech segment as voiced or unvoiced and by the pitch period of the segment. The encoder sends a single bit to inform if the current segment is voiced or unvoiced. Pitch period is quantized using a log-companded quantizes to one of 60 possible values. 6 bits are mandatory to represent the pitch period. If the segment has voiced speech than a 10th order filter is used. This means that values are needed: 10 reflection coefficients and the gain. If the segment is an unvoiced speech than a 4th order filter is used. It means that 5 values are required: 4 reflection coefficients and the gain.

1 bit	voiced/unvoiced
6 bits	pitch period (60 values)
10 bits	k1 and k2 (5 each)
10 bits	k3 and k4 (5 each)
16 bits	k5, k6, k7, k8 (4 each)
3bits	k9
2 bits	k10
5 bits	gain G
1 bit	synchronization
54 bits	TOTAL BITS PER FRAME

Figure 5: Total Bits in each speech segment

Sample rate = 8000 samples/second
Samples per segment = 180 samples/segment
Segment rate = Sample Rate/ Samples per Segment
= (8000 samples/second)/(180 samples/second)
= 44.444444.... segments/second
Segment size = 54 bits/segment
Bit rate = Segment size * Segment rate
= (54 bits/segment) * (44.44 segments/second)
= 2400 bits/second

Figure 6: Bit Rate of LPC speech segment

The total number of bits required for each segment or frame is 54 bits. This total is explained in Figure 6. Recall that the input speech is sampled at a rate of 8000 samples per second and that the 8000 samples in each second of speech signal are broken into 180 sample segments. This means that there are roughly 44.4 frames or segments/ second and therefore the bit rate is 2400 bits/second as show in Figure 6.

LPC SYNTHESIS/DECODING

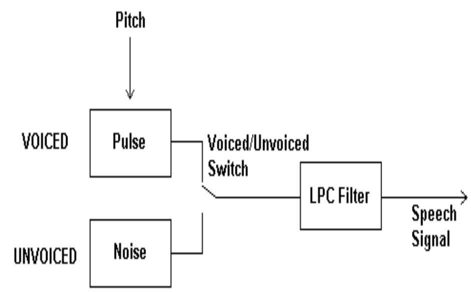

Figure 7: LPC Decoder

The process of decoding a sequence of speech segments is the reverse of the encoding process. Every segment is decoded individually and the sequence of reproduced sound segments is joined together to represent the entire input speech signal. The synthesis or decoding of a speech segment is based on the 54 bits of information that are transmitted from the encoder.

Determining the type of speech signal:
The speech signal is declared voiced or unvoiced based on the voiced/unvoiced determination bit. The decoder wants to know what type of signal the segment contains in order to determine what type of excitement signal will be given to the LPC filter. For voiced segments pulse is used as the excitement signal. This pulse contains of 40 samples and is locally stored by the decoder. For unvoiced segments white noise created by a pseudorandom number generator is used as the input for the filter.

Determining the pitch period:
This combination of voice/unvoiced determination and pitch period are the only things that are need to produce the excitement signal. 10 reflection coefficients are used for voiced segment filters and 4 reflection coefficients are used for unvoiced segments. These reflection coefficients are used to generate the vocal tract coefficients or parameters which are used to create the filter which are generated by the use of these reflection coefficients

Synthesis of speech signal:
The final step of decoding a segment of speech is to pass the excitement signal through the filter to produce the synthesized speech signal. Figure 7 shows a diagram of the LPC decoder.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

ADVANTAGES AND DISADVANTAGES OF LPC

Advantages of LPC:

1. The main advantage of linear predictive coding is to reduce the bitrates of the speech i.e. reduces the size of the transmitting signal.
2. The signal transmitted through LPC required less bandwidth and hence no. of users can be increased
3. This method of coding uses the encryption of data so the data is secured until the destination.

Disadvantages of LPC:

1. Due to reduce in the bitrates of the speech signal, the quality of voice signal is reduced.
2. This technique is lossy compression technique, hence data gets faded if transmitted to the long distance.

LPC APPLICATIONS

In general, the most common usage for speech compression is in standard telephone systems. In fact, a lot of the technology used in speech compression was developed by the phone companies. Below shown the bit rates used by different phone systems. Linear predictive coding has many applications in the area of secure telephony because of its low bit rate. Sheltered telephone systems require a low bit rate since speech is initially digitalized, then encrypted and transmitted. These systems have a primary objective of decreasing the bit rate as much as possible while maintaining a level of

North American Telephone Systems (uncompressed)	64kb/s
International Telephone Network (can range from 5.3-64kb/s)	32 kb/s
Digital Cellular standards	6.7-13 kbps
Regional Cellular standards	3.45-13 kb/s
Secure Telephony	0.8-16 kb/s

Speech quality that is understandable.

A second area that linear predictive coding has been used is in Text-to-Speech synthesis. In this type of synthesis the speech has to be generated from text. Since LPC synthesis involves the generation of speech based on a model of the vocal tract, it provides a ideal method for generating speech from text. Further applications of LPC and other speech compression schemes are voice mail systems, telephone answering machines, and multimedia applications.

CONCLUSION

Linear Predictive Coding is an analysis/synthesis technique to lossy speech compression that attempts to model the human production of sound instead of transmitting an estimate of the sound wave. Linear predictive coding achieves the bit rate of 2400 bits/second which makes it ideal for use in secure telephone systems. Protected telephone systems are more concerned that the content and meaning of speech, rather than the value of speech, be preserved. The trade off for LPC's low bit rate is that it does have some difficulty with certain sounds and it produces speech that sound synthetic.

Linear predictive coding encoders break up a sound signal into different segments and then send information on each segment to the decoder. The encoder send data on whether the segment is voiced or unvoiced and the pitch period for voiced segment which is used to create an excitement signal in the decoder.

REFERENCES

[1] Jeremy Bradbury "Linear predictive coding" December 5, 2000.
[2] Mrs.Savitha S Upadhya EXTC Department" Linear Predictive Vocoder" International Journal of Electronics Communication and Computer Technology (IJECCT) Volume 3 Issue 2 (March 2013)
[3] Gupta Rajani, Mehta Alok K. and Tiwari Vebhav "Vocoder (LPC) Analysis by Variation of Input Parameters and Signals" ISCA Journal of Engineering Sciences Vol. 1(1), 57-61, July (2012).
[4] Akshya K. Swain Waleed Abdulla" Estimation of LPC Parameters of Speech Signals in Noisy Environment"
[5] Richard V. Cox. Speech Coding (1999) 5-8
[6] Mark Nelson and Jean-Loup Gailly. Speech Compression, the Data Compression Book (1995) 289-319.