Linear predictive coding essay sample

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 Defence 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. I will describe the necessary background needed to understand how the vocal tract produces speech. I will also explain how linear predictive coding mathematically approximates the parameters of the vocal tract to reduce a speech signal to a state that is noticeably synthetic but still understandable. I will conclude by discussing other speech encoding schemes that have been based on LPC and by discussing possible disadvantages and applications of the LPC model.

2 Introduction

There exist many different types of speech compression that make use of a variety of different techniques. However, most methods of speech compression exploit the fact that speech production occurs through slow anatomical movements and that the speech produced has a limited frequency range. The frequency of human speech production ranges from around 300 Hz to 3400 Hz. Speech compression is often referred to as speech coding which is defined as a method for reducing the amount of information needed to represent a speech signal. Most forms of speech coding are usually based on a lossy algorithm. Lossy algorithms are considered acceptable when encoding speech because the loss of quality is often undetectable to the human ear. There are many other characteristics about speech production that can be exploited by speech coding algorithms.

One fact that is often used is that period of silence take up greater than 50% of conversations. An easy way to save bandwidth and reduce the amount of information needed to represent the speech signal is to not transmit the silence. Another fact about speech production that can be taken advantage of is that mechanically there is a high correlation between adjacent samples of speech. Most forms of speech compression are achieved by modelling the process of speechproduction as a linear digital filter. The digital filter and its slow changing parameters are usually encoded to achieve compression from the speech signal. Linear Predictive Coding (LPC) is one of the methods of compression that models the process of speech production. Specifically, LPC models this process as a linear sum of earlier samples using a digital filter inputting an excitement signal. An alternate 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 parameters of the process as opposed to sending the speech itself” [4]. It was first proposed as a method for encoding human speech by the United States Department of Defence in federal standard 1015, published in 1984.

Another name for federal standard 1015 is LPC-10 which is the method of Linear predictive coding that will be described in this paper. Speech coding or compression is usually conducted with the use of voice coders or vocoders. There are two types of voice coders: waveform-following coders and model-base coders. Waveformfollowing coders will exactly reproduce the original speech signal if no quantization errors occur. Model-based coders will never exactly reproduce the original speech signal, regardless of the "presence of quantization errors, because they use a parametric model of speech production which involves encoding and transmitting the parameters not the signal. LPC vocoders are considered model-based coders which means that LPC coding is lossy even if no quantization errors occur. All vocoders, including LPC vocoders, have four main attributes: bit rate, delay, complexity, quality. Any voice coder, regardless of the algorithm it uses, will have to make trade offs between these different attributes. The first attribute of vocoders, the bit rate, is used to determine the degree of compression that a vocoder achieves. Uncompressed speech is usually transmitted at 64 kb/s using 8 bits/sample and a rate of 8 kHz for sampling. Any bit rate below 64 kb/s is considered compression. The linear predictive coder transmits speech at a bit rate of 2. 4 kb/s, an excellent rate of compression.

Delay is another important attribute for vocoders that are involved with the transmission of an encoded speech signal. Vocoders which are involved with the storage of the compressed speech, as opposed to transmission, are not as concern with delay. The general delay standard for transmitted speech conversations is that any delay that is greater than 300 ms is considered unacceptable. The third attribute of voice coders is the complexity of the algorithm used. The complexity affects both the cost and the power of the vocoder. Linear predictive coding because of its high compression rate is very complex and involves executing millions of instructions per second. LPC often requires more than one processor to run in real time. The final attribute of vocoders is quality. Quality is a subjective attribute and it depends on how the speech sounds to a given listener. One of the most common test 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. An alternate method of speech compression called adaptive differential pulse code modulation (ADPCM) only reduces the bit rate by a factor of 2 to 4, between 16 kb/s and 32kb/s , but has a much higher quality of speech than LPC. The general algorithm for linear predictive coding involves an analysis or encoding part and a synthesis or decoding part. In the encoding, LPC takes the speech signal in blocks or 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. 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, attempts to output speech. Additional information about the original speech signal is used by the decoder to determine the input or excitation signal that is sent 4 "to the filter for synthesis.

3 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 started in the 1960s by telephone companies who were concerned with the cost of transmission bandwidth. Linear Predictive Coding’s origins begin in the 1970s with the development of the first LPC algorithms. Adaptive Differential Pulse Code Modulation (ADPCM), another method of speech coding, was also first conceived in the 1970s. In 1984, the United States Department of Defence produced federal standard 1015 which outlined the details 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.

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 first vocoder was described by Homer Dudley in 1939 at Bell Laboratories. A picture of Homer Dudley and his vocoder can be seen in Figure 1. Dudley

developed his vocoder, called the Parallel Bandpass Vocoder or channel vocoder, to do speech analysis and resynthesis. LPC is a descendent of this channel vocoder. The analysis/synthesis scheme used by Dudley is the scheme of compression that is used in many types of speech compression such as LPC. The synthesis part of this scheme was first used even earlier than the 1930s by Kempelen Farkas Lovag (1734-1804). He used it to make the first machine that could speak. The machine was constructed using a bellow which forced air through a flexible tube to produce sound. 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 5 Figure 1: Homer Dudley and his vocoder "sent to the receiver. The receiver then uses the model and the parameters it receives to synthesize an approximation of the original signal. Historically, this method of sending the model parameters to the receiver was the earliest form of lossy speech compression. Other forms of lossy speech compression that involve sending estimates of the original signal weren’t developed until much later.

4 Human Speech Production

Regardless of the language spoken, all people use relatively the same anatomy to produce sound. The output produced by each human’s anatomy is limited by the laws of physics. The process of speech production in humans can be summarized as air being pushed from the lungs, through 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. The above is a simplification of how sound is really produced.

Figure 2: Path of Human Speech Production

In order to understand how the vocal tract turns the air from the lungs into sound it is important to understand several key definitions. Phonemes are defined as a limited set of individual 6 "sounds. There are two categories of phonemes, voiced and unvoiced sounds, that are considered by the Linear predictive coder when analysing and synthesizing speech signals. Voiced sounds are usually vowels and often have high average energy levels and very distinct resonant or formant frequencies. Voiced sounds are generated by air from the lungs being forced over the vocal cords. As a result the vocal cords vibrate in a somewhat periodically pattern that produces a series of air pulses called glottal pulses.

The rate at which the vocal cords vibrate is what determines the pitch of the sound produced. These air pulse that are created by the vibrations finally pass along the rest of the vocal tract where some frequencies resonate. It is generally known that women and children have higher pitched voices than men as a result of a faster rate of vibration during the production of voiced sounds. It is therefore important to include the pitch period in the analysis and synthesis of speech if the final output is expected to accurately represent the original input signal. Unvoiced sounds are usually consonants and generally have less energy and higher frequencies then voiced sounds.

The production of unvoiced sound involves air being forced through the vocal tract in a turbulent flow. During this process the vocal cords do not vibrate, instead, they stay open until the sound is produced. Pitch is an unimportant attribute of unvoiced speech since there is no vibration of the vocal cords and no glottal pulses. The categorization of sounds as voiced or unvoiced is an important consideration in the analysis and synthesis process. In fact, the vibration of the vocal cords, or lack 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 produce different sounds or resonant 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 through 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 flowing from the lungs can be though of as the source for the vocal tract which act 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 sound. 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 sound production. The source-filter model is the model that is used in linear predictive coding. It is based on the idea of separating the source from the filter in the production of 7 "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. This air can be periodic, when producing voiced sounds through vibrating vocal cords, or it can be turbulent and random when producing unvoiced sounds. The encoding process of LPC involves determining a set of accurate parameters for modelling the vocal tract during the production of a given speech signal. Decoding involves using the parameters acquired in the encoding and analysis to build a synthesized version of the original speech signal. LPC never transmits any estimates of speech to the receiver, it only sends the model to produce the speech and some indications about what type of sound is being produced.

In A Practical Handbook for Speech Coders, Randy Goldberg and Lance Riek define the process of modelling speech production as a general concept of modelling any type of sound wave in any medium. “ Sound waves are pressure variations that propagate through air (or any other medium) by the vibrations of the air particles. Modelling these waves and their propagation through the vocal tract provides a framework for characterizing how the vocal tract shapes the frequency content of the excitation signal” [7].

5 LPC Model

The particular source-filter model used in LPC is known as the Linear predictive coding model. It has two key components: analysis or encoding and synthesis or decoding. The analysis part of LPC involves examining the speech signal and breaking it down into segments or blocks. Each segment is than examined further to find the answers to several key questions: • • • Is

the segment voiced or unvoiced? What is the pitch of the segment? What parameters are needed to build a filter that models the vocal tract for the current segment? LPC analysis is usually conducted by a sender who answers these questions and usually transmits these answers onto 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 imitate human speech production. Figure 3 demonstrates what parts of the receiver correspond to what parts in the human anatomy. This diagram is for a general voice or speech coder and is not specific to linear predictive coding. All voice coders tend to model two things: excitation and articulation. Excitation is the type of sound that is passed into the filter or vocal tract and articulation is the transformation of the excitation signal into speech.

Figure 3: Human vs. Voice Coder Speech Production

6 LPC Analysis/Encoding

Input speech According to government standard 1014, also known as LPC-10, the input signal is sampled at a rate of 8000 samples per second. This input signal is then broken up into segments or blocks which are each analysed and transmitted to the receiver. The 8000 samples in each second of speech "signal are broken into 180 sample segments. This means that each segment represents 22. 5 milliseconds of the input speech signal. Voice/Unvoiced Determination 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. Determining 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 creates a need for the use of two different input signals for the LPC filter in the synthesis or 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. Recall that voiced sounds are usually vowels and can be considered as a pulse that is similar to periodic waveforms. These sounds have high average energy levels which means that they have very large amplitudes. Voiced sounds also have distinct resonant or formant frequencies. A sample of voiced speech can be seen in Figure 4 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.