

Research on Speech Codec with Compression using Legendre Polynomials

Arun Ananthanarayanan, S.Ananthi, K. Padmanabhan, Bhagwan Shree Ram



Abstract- *Speech Codecs have been developed in several forms over the past 30 years and are used in Cellular digital telephony. Based on the throat excitation pulses and mouth cavity filter model, all the methods aim to encode the speech samples without much loss of clarity. The complexity of the methods using codebook searching, filtering and matrix inversion, all need heavy computations, leading to timing constraints. After describing a tutorial elucidating the essentials of the present methods, a different direct method of encoding, suggested by the spherical harmonic oscillator functions of Legendre, is introduced. Its performance and simplicity are such that the same may find its applications soon.*

Keywords: *Speech Processing, Legendre Polynomial, Audio codecs, CELP, LPC, Speech Compression*

I. INTRODUCTION

Since DSP processors have become available at economical costs, the efforts of senior workers in the field of speech coding have now been deployed successfully for communication systems. Even then, the goal of toll-quality low-rate coding continues to provide a challenge in research. Man has been trying since ages to create machines that can talk and to record speech and music. Early attempts at understanding speech production consisted of building mechanical models to mimic the human vocal apparatus. Our forefather mathematicians have developed all that is required for the communication technological advances of their future generation. Among them, Fourier, Galois, Hilbert, Hamilton, Lagrange, Legendre are some. With the power of computation growing up, more and more of their mathematical and hitherto considered bookish equations have been drawn into light. Processing of signals using their equations through some algorithms have not only provided means of compression, transmission and reception of signals but also provided methods of developing signals for various novel applications. Speech is one such signal which is generated by a very complex anatomy of the throat, mouth and nasal cavity. Encoding speech signals, compression of speech are handled by digital signal processing. Modeling of speech signals eliciting from the vocal chords has been attempted by many and employed to develop the presently prevalent standards of audio codecs, such as the CELP and its variants. The coding of speech through digital signal processing has seen various facets during the past thirty years.

Manuscript published on 30 July 2019.

* Correspondence Author (s)

Dr. Arun Ananthanarayanan, Department of ECE, MVJ College of Engineering.

Dr. S.Ananthi, Department of ECE MVJ College of Engineering.

Dr. K.Padmanabhan, AC College of Engineering Chennai.

Dr. Bhagwan Shree Ram, Power System Domain, SEEE, LPU Phagwara, Jalandhar, Punjab, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

From the LPC through the several CELP codecs, many standards have been evolved and employed in GSM and several communication systems [1]. The various forms of the Code excited linear prediction techniques discovered involve quite an amount of processing of the samples. Use of an extensive dictionary of excitation pulses and picking the one for a particular slot of speech in every 40 samples or so using optimization involves extensive computational time. Thus, even the simplest of CELP coders needed a highly powerful digital signal processing chip. Even with so much computational overhead, the resulting speech is having errors. Thirty years of all such research work has given us the digital communication that we find in the cellphones and media products. All the code excitation models assume that the vocal chord vibrations are just impulses and their time periodicity called 'pitch', varies as one speaks. The vibrations of a cantilever are known to be sine waves with a damped exponential decay. The differential equation for a sinusoidal oscillator is the well known simple harmonic second order linear equation.

$$\frac{d^2 y}{dt^2} + 2\zeta\omega \frac{dy}{dt} + \omega^2 y(t) = f(t) \quad \dots(1)$$

One of the above mathematicians, Legendre, had described oscillations known as spherical harmonics based on his differential equation:

$$(1-t^2)y'' - 2ty' + n(n+1)y = 0 \quad \dots(2)$$

He had mentioned that the partly spherical vocal cavity oscillations are generated by similar nonlinear oscillations. The number n in the Legendre equation provides various degrees of vibrations. Rabiner [2] and his followers [3-5] have thought that the vibrations of the vocal chords are impulses and the harmonics generated in the air are sinusoidal. To some extent, their supposition is that there are vibrating reeds, and a sinusoidal damped oscillation is the resulting source signal. It is necessary to dwell on the developments made with the simple linear vibrator mode, before describing the Legendre functions for the speech signals.

A. The throat-mouth-model based coding

The model of the human sound production having been estimated since Dudley, all research went in line with it, with the assumption that no better compression of speech can be made than with a model. It was conceived that when a model of speech is found, its representation in some parameters would be enough to fully represent the speech in a concise form. No other method would be able to do it better. Since DSP processors became available at economical costs in the late eighties of the past century, they came to be used in speech processing and digital telephony.

Linear predictive coding for speech is considered as the earliest work on model based codecs.

The speech processing problem is one of compression. For example, if speech is digitized at an 8 KHz rate with a 12-bit ADC, a bit rate of 96 bits per millisecond is got. If such digitized speech has to be stored or transmitted, the amount of memory or the bandwidth of the channel is quite high and is practically not possible. Reducing this data rate to a feasible level, such as 2400 bit/s needs a Compression ratio of about 40:1. But then, the reconstruction of the speech after this must be at least intelligible.

The large number of bits required for accurate reproduction of the speech waveform makes many of coder systems complex and expensive, so, more-efficient encoding of speech signals is desirable. For example, limited radio bandwidth is a major constraint in the design of public mobile telephone systems, and the speech data rate directly influences the bandwidth requirement. In military tactical communications, a system with a lower speech data rate can use less transmitter power to make detection more difficult, or it can allow higher signal-to-noise ratios to improve performance in a hostile jamming environment. Also, computer storage of speech, such as in voice mail or voice response systems, becomes cheaper if the number of bits required for speech storage can be reduced. These are just some of the applications that can benefit from the development of algorithms that significantly reduce the speech data rate.

With the technique of LPC coding, bit rates up to 2400 bits per second have been obtained and the reconstructed speech also was fairly realistic.

The name LPC stands for Linear Predictive Coding. Before explaining what is linear and what is predicted, it is necessary to consider the model of the vocal system as proposed by the early workers. A rough schematic of vocal tract is shown in the fig.1.1 below. Its real anatomy is extremely complex. The model based compression represents the speech not as raw data but as parameters based on the vocal system. This model starts with the effect of lungs which produces the breath. The breath passes through the vocal chords and produces vibrations. The filtering effects of the positioning within the mouth and nose region are the models of a filter, which produces the syllables.

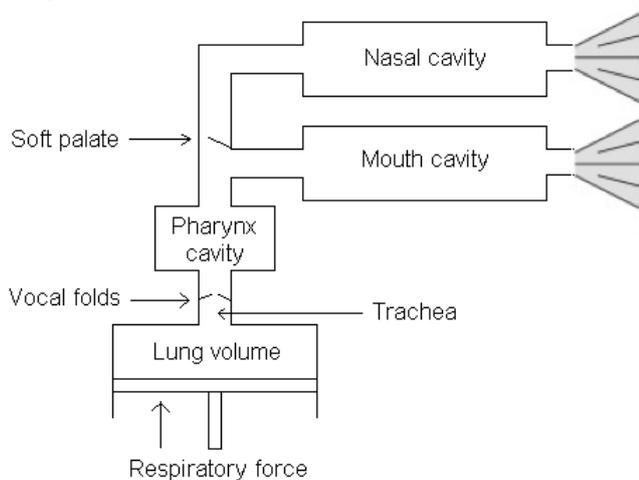


Fig. 1 A model representation of vocal sound system.

An excitation is used to reproduce the equivalent of breath through the vocal chords. This excitation is time varying and its frequency (for pitch) varies as per the speaker's mode. The throat model is used to shape or filter this excitation pulse. The filtered waveform is that which somewhat resembles the speech.

B. SHORT TUTORIAL ON LPC, CELP AND IMPULSE EXCITATION BASED ENCODING

Speech is generally modelled as a sequence of stationary signal segments, each having unique statistics. Segments are encoded using a two-step procedure:

- (1) find a model describing the speech segment,
- (2) encode the segment assuming it to be generated by the model.

Given the speech signal, or a short time segment of it, the method aims to find an excitation (breath, its pulse rate, amplitude) and the filter coefficients of the cavities. Neither of the two is constant in time. So, it is required to process sequence of short segments of speech signals and find the parameters for every segment. The number of data to get the above parameters will be having less value than the speech segment's samples. This results in the compression. The modeling of perception is an important aspect of efficient coding. The different codecs aim to model perception and reduce the distortion of the compressed speech.

The term linear prediction (LP) first proposed and continuously followed by all the workers in speech coding is by itself very difficult to substantiate its name. The question arises in a fresher to the topic as to what is linear among the speech sample values and what is the prediction that is being undertaken. The evaluation of the LP coefficients is the fundamental one in each and every one of the methods thus far developed. In this, a set of linear equations are solved involving correlation coefficients and hence the term "linear". The 'prediction' is probably the term meaning that from these coefficients (which compressed the sound), it is possible to reconstruct (or predict) the original sound to some extent. The main assumption made in this evaluation is that the vocal cavity filters are describable by a mere IIR filter structure with as many coefficients as one would use (ten in LPC10, for example).[6]. Even in this filter, there are no zeros but only poles. By looking at the *spectrum* of the speech segment of about 100 samples known as a 'frame', one could notice that there are peak points in it. These they call as 'formant frequencies'. So it has been assumed (*Assumption no.1*) that if one puts single frequencies at those places of the peaks, the same is equivalent to poles at those frequencies. The assumption of a poles-only filter for our human vocal cavities has been considered satisfactory.

The LPC "all pole" filter, when used to filter the sound emanating as pulses or white noise from the throat model, is considered to give rise to the speech of this frame.

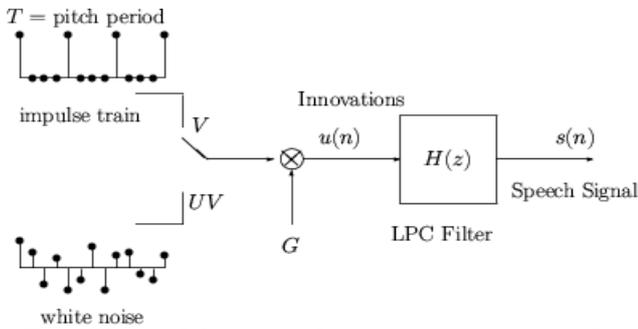


Fig.2. The model for speech using an excitation and a filter. V and UV refer to voiced/unvoiced speech. G is Gain (datacompression.com).

The transfer function between the signal output to input for such a poles only filter is:

$$H(z) = \frac{s(z)}{u(z)} = \frac{1}{1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{10}z^{-10}} \quad \dots(3)$$

This above transfer function between output and input can be expanded in a time series. The poles are at $-(a_1, a_2, \dots)$. z^{-1} represents the past sample, with a delay of one sample time. Expanding the above, since $H = s/u$, where u is input to the filter and s is the output of the filter, we get the difference equation

$$s(n) + a_1s(n-1) + a_2s(n-2) + a_3s(n-3) + \dots + a_{10}s(n-10) = u(n) \quad \dots(4)$$

which is written in sigma notation as:

$$s(n) + \sum_{i=1}^{i=10} a_i s(n-i) = u(n) \quad \dots(5)$$

So, the LPC model can be represented in vector form, including the nature of excitation of throat output in the three parameters, namely, Gain, voiced or unvoiced, pitch time, as:

$$A = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T) \quad \dots(6)$$

A changes every 20msec as per the data frame chosen at a time. The model is therefore equivalent to the expanded

$$A = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T) \text{ vect or.}$$

If we plot a speech segment in Fourier spectrum and also superpose the LPC filter response on it, we get an idea of what this filter does, as shown in fig 2 below. The filter picks the envelope of the spectrum. The peaks are the formant frequencies. There will be three major lobes in any speech segment. Each of these segments can independently be sounded and the speech will be intelligible, if not exact with the original. Later it is shown how the three segments can be singly generated by a spherical harmonic oscillation.

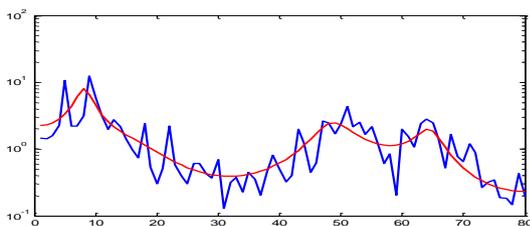


Fig.3. Fourier spectrum of a segment of speech (160 samples) and the superposed LPC filter response.

In this difference equation (4), the n^{th} value of signal is written in term of the excitation signal $u(n)$ and the past signal values $s(n-1), s(n-2)$ etc. This is misinterpreted as predicting the n^{th} value from past values. Actually, we have to know all the n values of s , in order to find the a 's.

$$S = (s(0), s(1), \dots, s(9))$$

The ten LPC parameters $(a_1, a_2, \dots, a_{10})$ are calculated so as to minimize the energy of the input signal [7]. It is assumed that to produce the sound, the energy spent will be a minimal one.

The energy is represented as being the sum of squares of the amplitudes of u .

$$E = \sum_{n=0}^{159} u^2(n).$$

To find the minimum of the function, the derivative of E with respect to a_i is equated to zero. When the square of the above equation (5) is found, products terms of $s(n)$ with $s(n-1), s(n-2), \dots$ etc. arise. They are the so called correlation terms between samples. Differentiation for minimizing E gives equations of terms of auto and cross correlations, R_0, R_1 , etc. A set of ten equations involving these have to be solved to find the filter coefficients a_0, a_1, a_2, \dots

The speech which is obtained as samples through the ADC may have 8K samples per second and the frame size chosen may be 100 samples at a time. Each frame gives rise to its own filter coefficients. To calculate them, an inversion of a matrix of 10×10 is to be done. Usually, a simple inversion method called Levinson Durbin iteration is used because the matrix is arranged with its rows in a shifted pattern. (Toeplitz matrix).

Getting these filter coefficients is not all. The excitation u_n must be created and the filter used to filter it.

The term "Code Excited" has been coined to indicate that the excitation is a set of pulses in each frame. If a frame has 100 samples, there could be 4 pulses. These pulses are actually assumed to be impulses, because they place a value in certain four points anywhere in the sample space of 100. It is assumed (Assumption 2) that vocal chord is just given a hit at these sample timings and that moves the air column through the throat. The timing between the pulses, their amplitudes, signs all can be anything. Thus there are four parameters of the assumed excitation. In a frame of 100 samples, with four parameters, the permutations to pick four or more pulses are too many. Herein came the research outputs leading to various forms of CELP codecs.

The first proposal was to choose a set of random pulse positions, assuming a Gaussian distribution. The next was to place the pulses, four of them (fig.4), in a regularly spaced set of sample points. Then, another worker placed the pulses in an arithmetic progression but there starting positions could be different. Varying the signs of the pulses gave another possibility. Therefore, all the CELP algorithms concentrated only on the basic principle of single point pulses.

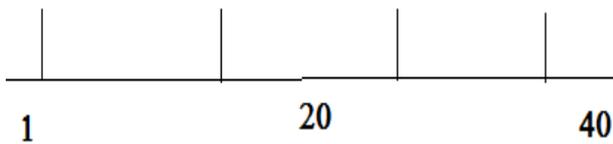


Fig.4. Multipulse LPC having about 4 pulses per frame of 40 samples. The four can be anywhere in the length of the X axis. The pulse positions, amplitudes, signs constitute a so called code vector.

Since there are innumerable possibilities of the pulses, a codebook was prepared and like a dictionary, searching through the codebook is done in every frame to fit the best pulse sequence for the frame speech signal. With such a pulse sequence after optimal choice, the filter using the all-pole filter is used to output the signal as the decoded sound. The codebook data – called code vector – along with the values are the encoded compressed data.

They have introduced two additional filters than the LPC filter. One is to filter the residual signal - the difference between the input speech signal and the synthesized speech signal – with a filter of the form

$$w(z) = \frac{A(z)}{A(z/\gamma)}$$

with γ as 0.8. This is termed as a “perceptually weighted filter” which is supposed to make the spectral valleys of the speech shallower. This is assumed (*Assumption 3*) to aid in codebook search.

Exactly as to how the dictionary of code vectors of excitation pulses is searched is not very clearly described by the codec people. Perhaps a Nelder Mead search is performed on an ordered codebook.

The compression performed leads to an output data stream, which are the compressed data values for each and every frame (of 100 samples or so). The data rate of this varies depending on the compression. The LPC10 is the simplest and crudest method which gives a compression of 26:1. So, the data rate of encoded stream is 2.4Kbps. Normal CELP codecs produce output with lesser compression values. The normal value is 8 for somewhat intelligible speech compression.

When the figures of the speech patterns as waveforms are illustrated in the reconstructed speech by the different methods and compared with the original, the sketch figures (fig.4) as the ones shown below (fig.4) are likely to be impressive in that there is but a small aberration in the noise like waveform chunks and that the methods should work well to make an understandable speech output.

However, looking inwards into the signals, for example, by comparing small segments of samples as in the fig.5, the disparity between the original sound and the reconstructed sound is considerable.

Even the highest complicated versions of CELP are only able to match the original in a very limited sense. Most of the versions of CELP methods are in use right from the days of GSM.

Table 1 GSM Audio codecs / Vocoder

Codec name	Bit rate	Compression (kbps)technology
Full rate	13	RTE-LPC
EFR	12.2	ACELP
Half rate	5.6	VSELP

AMR	12.2 -4.75	ACELP
AMR-WB	23.85-6.60	ACELP

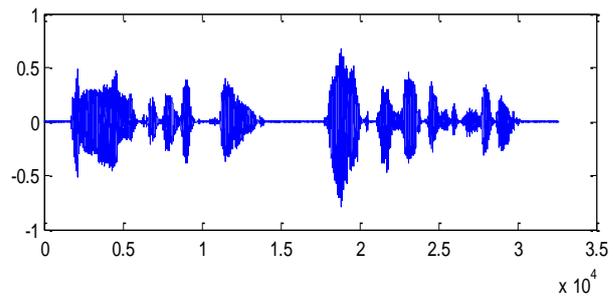
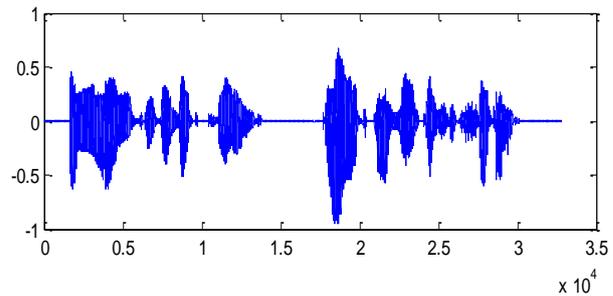


Fig.5a. The original sound file of 3 seconds and the CELP reconstructed file (data from datacompression.com [8]).

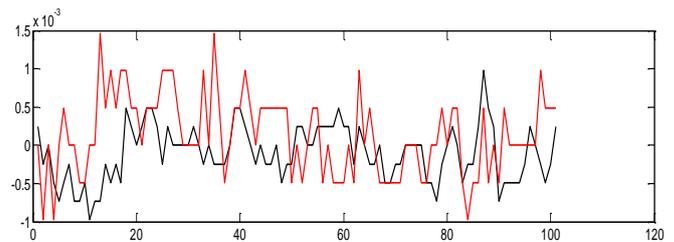


Fig.5b. For any short segment of 120 data points from the original and CELP compressed file, the plots show differences.

Therefore, all work done from the initial stages till recently have centred around this basic concept of the human throat model, the filters of the mouth and nasal cavities and so on. This concept is very interesting because it tries to connect the speech signal with its source and it provides an understanding of the human physiology behind speech production. Therefore, all work that emanated from the 1980s have used this very same concept of throat excitation and filtering. The fascination for this model, as it is linked with the human body itself, is so much that in all these years very little has been investigated to find out an alternative approach to speech signal encoding to compress it. Because of the assumption that there can be no better compression method than by characterising the model of the throat and the mouth cavities, all work thus far has been to utilise this concept of the larynx model. The model is good for understanding how speech is produced, but it is not a complete one, because, with the same model, even animal sounds could be characterised, but then, they have no intelligible speech output. Therefore, we have to accept that the human speech production needs a very sophisticated model, if such a model based compression is our aim.

With the difficulties of model based speech compression not completely solved, if we can find a compression scheme based merely on the speech signal without reference to any model, it is a worthwhile attempt. From the point of view of waveform fitting, curve fitting techniques could be used. But the speech waveforms are so very oscillating, with frequencies spread over a wide range even within a 20 ms time frame. The methods of fitting should employ waveforms of some special functions which can encompass a wide range of frequencies in them and which can be represented by just a few coefficient values. Even the Fourier Spectrum is just a representation of the sinusoidal waveforms of all frequencies and the number of Fourier coefficients forms a compressed output. But there is negligible compression with the Fourier coefficients. A scheme called “codec-2” [9] uses such a “harmonic sinusoidal speech coding”.

Herein comes what the Legendre polynomials based on spherical harmonic oscillations as suggested by Legendre himself could do. What follows is not anything original but a mere application of the same and is meant to bring out the simplicity of it for speech compression and encoding.

II. LEGENDRE POLYNOMIALS

In spherical coordinates, the solution of Laplace’s equation leads to the Legendre’s Differential equation (5).[10].

Let us now consider the Legendre differential equation.

$$(1-x^2)y''-2xy'+n(n+1)y=0 \quad \dots(7)$$

There is one solution for each value of n . The solution is a polynomial in x . There are as many as one can choose the value of n to be. Before solving this equation for different values of n , we can consider a version with time t as the independent variable.

$$(1-t^2)y''-2ty'+n(n+1)y=0 \quad \dots(8)$$

In this if we omit the factor containing the first derivative, we get

$$(1-t^2)y''+n(n+1)y=0$$

This can be rewritten as, for $n=0$, as

$$(1-t^2)y''+y=0$$

This is also rewritten as

$$y''+\frac{y}{(1-t^2)}=0$$

It is well known that the second order undamped oscillator differential equation is given by $y''+\omega^2y=0$

This is a sinusoidal oscillation of frequency ω . But the equation (7) above has a variable frequency with time t . This is but natural in a nonlinear oscillator. The frequency is at $t=0$, $\omega=1$. At $t=0.5$, it is

$$\sqrt{\frac{1}{1-t^2}} = \sqrt{\frac{1}{1-.5^2}} = 1.155$$

Thus, the nature of oscillations of the vocal chords being such that they have a nonlinear vibration, the use of Legendre polynomials for coding speech is a good suggestion.

The solution of the above differential equation (5) is given by Rodriguez as [11]

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2-1)^n \quad \dots(9)$$

The evaluation of the several polynomials for $n=1$ to 4 are given below .

$$P_0(x)=1; P_1(x)=x; P_2(x)=\frac{1}{2}(3x^2-1); P_3(x)=\frac{1}{8}(35x^3-30x^2+3)\dots etc \dots(10)$$

Since the values of x can be only between -1 and $+1$, the powers of x are all less than 1 and hence the values of P for all n do not diverge very much.

There is a recurrence formula which is useful for finding the higher values of P_n using previous two P values.

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \quad \dots(12)$$

Using this relation, from the first two values of P_0 and P_1 , all further values of P_n can be written down. The range for x is only between -1 and $+1$.

Rather than writing the equations, we can use numerical values for the x between $-1 < x < 1$ and evaluate the equation (8). This is done in Matlab easily.

Matlab has symbolic algebraic functionality. To generate the polynomials as functions of x one defines x as symbolic by the command

```
sym x
P1=x;
P2=(3*x*x-1)/2;
P3=(35*x^4-30*x*x+3)/8;
```

Then, we can evaluate the functions by substituting numerical values for x . Since the range of the Polynomials is for $-1 \leq x \leq 1$, this range is specified for x in the following statement as:

```
x = -1:0.01:1;
```

Thus, 200 points are taken in this range. For each point, the values of the Polynomials are found by evaluation in the following command.

```
p11=eval(P1);
p12=eval(P2);
p13=eval(P3);
```

We can save the polynomials for 200 points as a matrix. So, we define a matrix and make its 200 entries null and then replace with the evaluated values.

```
n=20;
p1(1:m,1:200)=0;
p1(1,:) =p11; p1(2,:) =p12; p1(3,:) =p13;
```

The number of polynomials n can be chosen as much as are required for representing the speech samples. This is given as 20 in the above statements.

So far, only $p1$ to $p3$ are now calculated and are available in $p1(1,:)$, $p1(2,:)$ and $p1(3,:)$.

The remaining polynomials 3 to 20 can be directly numerically evaluated. The recurrence relation enables this. A for loop is used to find all the m polynomials.

```
For i=2 to n
p1(i+1,1:200)=(2*i+1)*x1*p1(i,:) - n*p1(i-1,:)
end
```

Now that the 20 polynomials are found and saved in the variables p1 as a matrix, they are useful for speech data sample fitting. A plot of the waveforms for $n= 1$ to 8 is shown in figure 6 below. From a look at the waveforms, we can note that the sound waves in a small chunk (for just few samples) could be matched to one and more out of these waveforms.

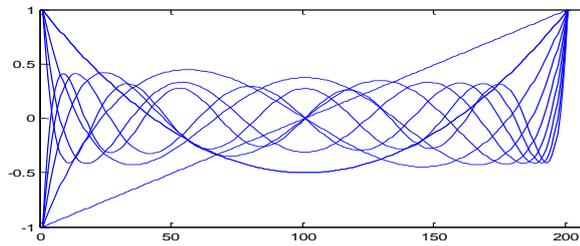


Fig.6. The Normal Legendre Polynomials for $n=1$ to 9 are shown. The range -1 to $+1$ is sampled into 200 sample points.

However, at the ends, i.e., at the points -1 and $+1$ (sample 0 and 200) the values change suddenly. This gave some peaks at the frame edges of the speech segments. So, further probe into the same Legendre functions has to be made. There are instances in literature on the use of Legendre functions for characterizing even 2-D waveforms. Associated Legendre polynomials have been used as early as 1974 for fitting such 2D data for the contours of cracks of Geological samples [12]. Applications of spherical harmonics using the functions have been useful for noise reduction from sounds.[13]. So, the use of the same is worthy of attempt. They are also based on a similar but somewhat different differential equation :

$$(1 - x^2)y'' - 2xy' + [n(n+1) - \frac{m^2}{1-m^2}]y = 0 \quad \dots(11)$$

If $m=0$, it gives the ordinary Legendre Polynomials as the solution. For $m=1$ and 2, we can generate polynomials similarly which are the associated versions of Legendre polynomials of degree 1 and 2. For $m=1$ the values are large, but for $m=2$, they are not so large. The polynomials for the first three orders (i.e., n values) are given below for $m=2$. The polynomials start with P_2 .

$$\dots P_n^m(x) = 0 \dots \text{if } m > n \dots$$

$$P_2^2(x) = 3(1-x^2); P_3^2(x) = 15x(1-x^2); P_4^2(x) = \frac{15}{2}(7x^2-1)(1-x^2) \quad \dots(12)$$

The recurrence relation for the associated Legendre polynomials is useful for finding the higher order polynomials in discrete form. The waveforms of the first few polynomials are shown in fig.7.

$$(n+1-m)P_{n+1}^m(x) - (2n+1)xP_n^m(x) + (n+m)P_{n-1}^m(x) = 0 \quad (13)$$

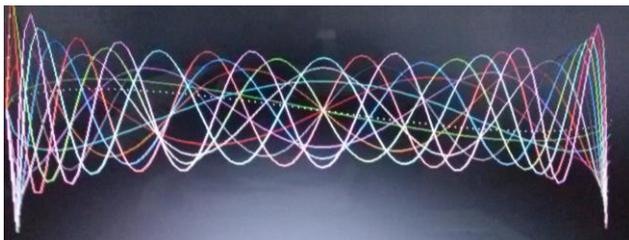


Fig.7. Showing wav deforms of Associated Legendre polynomials degree 2 for orders 2 to 10. (Photo taken from screen view of a program written in VB;6 for the purpose).

Thus, one could fit speech data using these polynomials fairly accurately, even as much as the complex CELP methods perform.

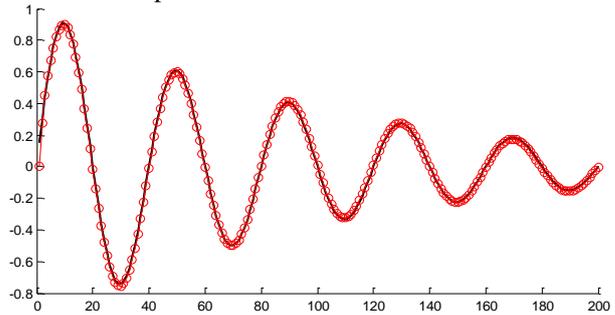


Fig.8. The Associated Legendre polynomial fit for a damped sine wave data gives exact fit with 16 coefficients.

III. FOURIER SPECTRAL COMPARISON

If we compare the Fourier Spectra of the original with the reconstructed sound, it is shown in figure.12. The higher frequency components are present in this speech which are suppressed (beyond 2KHz). The ratio of compression is 9.

The figure (Fig.12) giving the spectra of the original speech and the compressed (& reconstructed) speech shows that there is only a small disparity between the two, which is only at the frequencies above 1.5 KHz.

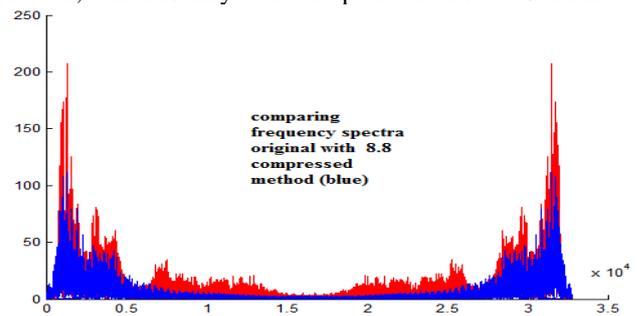


Fig.9. Showing the comparison of spectrum of the original signal and the compressed signal with Legendre polynomials. (Red is Original and Blue is compressed and reconstructed ($n=24$). (X axis shows sample numbers -- the semi axis is 4KHz).

The range of the compressed c coefficients in this total speech sample was -0.0116 to $+0.014$. This range is a rather small range in absolute values. The values are quantized into six bit binary fractions. The higher frequency components which are not present in the reconstructed sound can be improved by choosing more number of polynomials. With 32 polynomials, the difference will be negligible between the two spectra and the sound also perfect.

Let us now illustrate another important point about the Legendre Polynomial spectra. Let us casually take a look at the Fourier spectra – absolute values - plotted for the different orders of Legendre polynomials in the following figure. From $m=8,10,12,16,24$ to 28, the spectra are given in this single figure 10.

By a look at the last two spectra for 24 and 28 orders and also a look at the standard speech spectrum given in the figure 8, it is apparent how the similarity of the speech spectrum in general matches with the shape of the polynomial spectra. It is now easy to understand that speech signals are spectrally shaped similar to the Legendre polynomials (of order 24, 28 and higher). There is a low frequency end peak, followed by two peaks at the mid and high end frequencies, which are smaller. In every human speech, no matter how long it is, the total spectrum is only of a similar shape. The spectrum shown in the previous figure is actually the result of the standard test speech given by the website on data compression (datacompression.com).

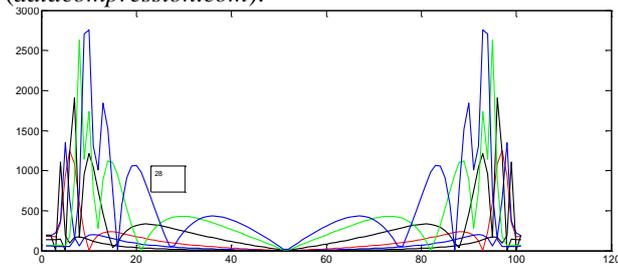


Fig.10.A set of spectra of the several Legendre polynomials starting from 8 and going up to 28.

Therefore, the justification that human speech is produced by oscillations of the kind of spherical harmonics, as described by Legendre in his famous differential equation given in the beginning is well supported. The advantages of the Legendre method could be summarized:

1. The method is simple to implement with least computational load
2. It gives good compression and good sound.
3. It can be adapted to variable bit rate easily by choice of number of polynomials.
4. Even simple embedded controllers can use it for full duplex telephony.
5. The latency is very low compared to the CELP codecs.

IV. CONCLUSION

Performance tests of codecs are based usually by listening. Listeners test the sound reproduced. In all the codecs, if we observe the wave shapes of frame signals – original and reconstructed – there is considerable disparity. Though the wave patterns differ, the sound output seems alright because of our auditory perception capacity. The power of mathematical relations given by Legendre’s differential equation is useful for speech encoding in a direct way, with simple calculations taking much less computational load than existing complex schemes prevalent in digital cellular and wireless communication.

REFERENCES

1. Jacob Benesty, M. Mohan Sondhi, Yiteng Huang (Eds.), Handbook Springer of Speech Processing Springer-Verlag Berlin Heidelberg, 2008.
2. Rabiner .R. and Schafer,.R.W., Digital Processing of Speech Signals, Prentice Hall Inc., Englewood Cliffs, N.J., 1978.
3. Atal and Schroeder B.S.: Predictive coding of speech signals and subjective error criteria, IEEE Trans. Acoust. Speech Signal Process. 3, 247–254 (1979).
4. C. Laflamme, J.-P. Adoul, H.Y. Su, S. Morisette: On reducing computational complexity of codebook search in CELP coder through the use of algebraic codes, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (1990) pp. 177–180.

5. J.H. Chen, M.S. Rauchwerk: An 8 kb/s low-delay CELP speech coder, Proc. IEEE Global Communication. Conf. (1991) pp. 1894–1898.
6. B.S. Atal: The history of linear prediction, IEEE Signal Proc. Mag. 23(2), 154–161, (2006).
7. K.Padmanabhan ,S.Ananthi, R.Vijayarajeswaran, A practical Approach to Digital Signal Processing, New Age International Publishers, New Delhi, revised edition 2002.
8. www.datacompression.com.
9. Voice and Audio Compression for Wireless Communication, Book, Lagos, Hanco, Clare Sommerville and Jasan Woodland. Wiley IEEE Press, 2007.
10. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.–Chapter 8: Applied Mathematics Series 55, Dover Publications. p. 332. ISBN 0-486-61272-4.
11. Spiegel, Math. Handbook of Formulas and Tables, Schaum Series Publications, 1990.
12. Attewell. P.B. and Sandford M.R, On the Creiterion of failure for Geological Rock Crack Distribution, Symposium on Anisotropic Rocks, 1974.
13. Techniques for Noise Robustness in Automatic Speech Recognition, edited by Tuomas Virtanen, Rita Singh, Bhiksha Raj, Wiley Inc.,2007.
14. <http://opus-codec.org>.
15. Mouser.com, DS70146A – Code for DSPIC 30 Family for CELP.

AUTHORS PROFILE



Dr. Arun Ananthanarayanan has done his B.E from Anna University,M.S in Electrical Engineering from SUNY Buffalo, USA. His research interest includes Computer Networks,Signal processing ,Communication systems He completed his PhD from University of Madras . Having served as Assistant Professor for more than 7 years in reputed private Engineering college, currently he is working as Assistant Professor at MVJ College of Engineering Bangalore.



Dr. K.Padmanabhan, did his Grad.Brit.IRE, B.E. from Guindy Engg.College and Doctorate from the Madras University and has served as Professor and Head of the Instrumentation Centre, University of Madras. After retirement, he is A.I.C.T.E. Emeritus Professor in the Anna University. He is a Fellow of IETE,IEE and Sr. Member IEEE. His areas of specialization range from Applied Electronics, Microprocessors, Telecommunications and DSP.



Dr. S.Ananthi, got her B.E. in ECE from the Anna University. She did M.Tech. at the Indian Institute of Science and later did Doctorate from University of Madras. She worked in the University as Professor and Head i/c of NSIT department till June 2018.After retirement, she is Professor, ECE in MVJCE. Her areas of specialization includes in DSP, Adv. Communication and Biomedical telemetry. She was awarded the Environmental TN state award for the year 2004.



Dr. Bhagwan Shree RAM holds a Bachelor, Master and Doctoral degrees in Electrical Engineering., and working as Professor in The School of Electronics and electrical Engineering, Lovely Professional University, Phagwara, Punjab, India. He had reviewed technical paper for the WSEAS Journal and African Journal of Agricultural Research. He has published more than 60 technical research papers in his credit. His principal research interests and experience include HRM, Entrepreneurship studies, distributed power generation, renewable energy conversion, analysis, design of electrical machines, variable-speed drives, power electronics and Entrepreneurship in renewable sources of energy systems.