

## Implementation of attractive Speech Quality for Mixed Excited Linear Prediction

Haresh Miyani<sup>1</sup>, Aalay Mehta<sup>2</sup>, Pratik Nai<sup>2</sup>, Harshad Patel<sup>2</sup>

Assistant professor of Electronics Engineering, Bhagwan Mahavir College of Engineering and Technology,  
Surat-395017, Gujarat, India.<sup>1</sup>

Student of Electronics Engineering, Bhagwan Mahavir College of Engineering and Technology, Surat-395017,  
Gujarat, India.<sup>2</sup>

---

**Abstract:** Nowadays the number of mobile subscribers is increasing all over the world, so the system for the communication has to be improved. Mixed Excited Linear Prediction (MELP) algorithm is developed for reducing the bandwidth of the signal as well as transmission of large data on a single channel. This results in increase channel capacity. This also results in, increasing the number of user in a channel. MELP is basically a speech coding method, relying on a Speech Encoder and Speech Decoder. The MELP speech coder reduces the redundancy of the signal and compresses it which is represented by the MELP code. Speech Decoder includes a Linear Predictive Coding (LPC) filter providing a synthesized speech at its output side in response to voice and unvoiced. MELP also reduces jitter voice. The bit rate of MELP is reducing the reserves of the code book and calculation complexity. This paper describes “the bit rates of MELP coder can be reduced to as low as 2.4kbps without apparent damage to the speech quality.”

**Keywords:** MELP, Channel capacity, Speech coder, Bandwidth, Synthetic speech.

---

### I. Introduction

As of June 2012, more than 2.4 billion people—over a third of the world's human population—have used the services of the Internet; approximately 100 times more people than were using it in 1995. Need for Mixed Excited Linear Prediction (MELP) to reduce the number of bits used to store or transmit more information. MELP is used for increasing the Bandwidth of the signal, which results in a transmission of more data over a single channel. The bandwidth of the digital communication can be increased by compressing the signal at the transmitter side and decompressing the signal at receiver side. If the data are effectively compressed, we receive an improved data on the receiver side. Speech communication is the basic communication from all of the communication. Mixed Excited Linear Prediction (MELP) coders can produce high quality speech at 2.4 kb/s. A distortion is seen when bit-rate are decreased below the 2.4 kb/s. Voice compression can be done in two steps:

1. Increasing the amount of data that can be stored in a given domain, such as space, time or frequency or contained in given message length.<sup>[1]</sup>
2. Reducing the amount of storage space required to store a given amount of data or reducing the length of message required to transfer the given amount of information.<sup>[1]</sup>

Voice compression is done by two techniques: 1) Lossy Compression—used when signals are in wave form, 2) Lossless Compression— used when storing data base records and word pressing files in the input data.

### II. Speech Coding

The term Speech Coding refers to the process of compressing and decompressing Human Speech. Speech coding is a digital technique. Transmission of Human Speech by digital technique is becoming more use in nowadays. Modern communication technologies are mobile satellite telephone, internet, landline, mailbox etc. In 1994 only 3% of American classrooms had the Internet while by 2002 92% did. A variety of Speech Coding methods are there for compressing (coding) and decompressing (decoding), to get the efficient Human Speech at the receiver side. ‘Compression’ is typically done by extracting parameters of successive sample sets; also it is referred as “frames”, of the original speech Waveform and representing the parameters as a digital signal. ‘Decompression’ is typically done by decoding the transmitted or stored digital signal.<sup>[6]</sup> In decoding the signal is the encoded versions of extracted parameters. There are three types of noise: 1) Voiced, 2) Unvoiced, 3) Jitter Voice.<sup>[1][3]</sup>

Accordingly, a need exists for a speech encoder and method for rapidly, efficiently and accurately characterizing speech signals in a fashion lending itself to compact digital representation. Further, a need exists for a speech decoder and method for providing high quality speech signals from the compact digital representations. The problem of providing high fidelity speech while conserving digital bandwidth and

minimizing both computation complexity and power requirements has been covered by the Mixed Excited Linear Prediction(MELP). The Fig.1 shows the Block Diagram of speech communication system. Speech Coding mainly divided into two parts: 1) Speech Encoder and 2) Speech Decoder.

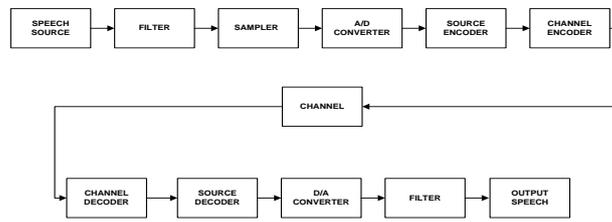


Fig.1 Block Diagram of Speech Communication System

Speech Encoder: Speech source is the input of the speech encoder, further it is divides in the Filter-removes the noise from the speech and do the framing of the signal. Sampler- it takes the most significant data from the speech on base the receiver side can regenerate the speech. AtoD converter- it is a analog to digital converter, which convert the speech into digital code. Source encoder- here the MELP main works. The MELP Encoder compresses the speech up to 2.4KB/s. and further it transmit to the channel.<sup>[2][3]</sup>

Speech Decoder: This is exactly reverses process of the encoder. It decodes the data from the cannel and give to the source decoder. This is a MELP Decoder, which regenerate the digital code send from the transmitter side. The digital to analog converter convert digital speech to the analog. And after the filtering the speech the original speech is received at the receiver.<sup>[2][3]</sup>

### III. Melp Encoder

An audio encoder is capable of capturing compression and converting audio to digital signal. Here Fig.2 shows Mixed Excited Linear Prediction (MELP) Encoder. It contains : A) Frame Segmentation, B) Pitch Detection, C) Fourier Magnitude, D) Shaping Filter, E) Gain Computation, F)LP Analysis.

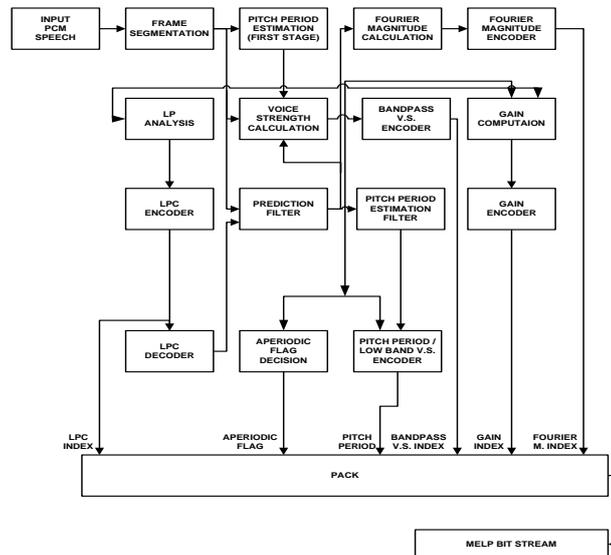


Fig.2 MELP Encoder<sup>[1]</sup>

#### A. Frame Segmentation:

Frame segmentation is also called as positions of Windows. The major frame segmentation utilized by MELP coder. MELP coder is take 180 samples in each frame for compression. This require for facilitate interpolation, because parameter of a given frame are interpolated between two different sets. It is calculated from centred at the beginning of the frame and at the end of the frame. For instance and superior Fourier magnitude utilize a 200 sample per frame is consider.<sup>[2][3]</sup>

#### B. Pitch Detection:

This parameter is a mostly plays a role to generate a synthetic speech. Pitch period is related to voicing strength. Here for detection of pitch period the fractional refinement is used throughout the encoding process.

To separate the input signal into five bands, a voicing strength find in every bands. First the speech is filtered by bandpass filter with passband 0 to 500Hz. <sup>[7]</sup> The normalized autocorrelation<sup>[2]</sup> is given by the,

$$r[l] = \frac{c[0, l, l]}{\sqrt{c[0, 0, l]c[l, l, l]}} \quad (1)$$

Where  $l=40, 41, \dots, 160$ .

The fractional pitch period denoted by the  $T^{(l)}=T + \eta$ .

**C. Fourier Magnitude:**

The MELP coder depends on the Fourier magnitude from the prediction error signal to generate the shape of the excitation speech or pulse which depends on the magnitude of Fourier transform. These values are quantized and transmitted on frames. The synthetic speech is can generate on the decoder side, a sequence as close as possible to the original signal. Discrete Fourier Transform (DFT) <sup>[6][3]</sup> is defined by,

$$X[k] = \sum_{n=0}^{N_0-1} x[n]e^{-j(\frac{2\pi kn}{N_0})}; k = 0, \dots, N_0 - 1 \quad (2)$$

being the analysis equation of the transform, and

$$X[n] = \frac{1}{N_0} \sum_{k=0}^{N_0-1} x[k]e^{j(\frac{2\pi kn}{N_0})}; n = 0, \dots, N_0 - 1 \quad (3)$$

$N_0=512$  is the length of the sequences.

The Fourier magnitude sequence is denoted by  $Fmag'[i]$ . It is normalized by,

$$Fmag'[i] = \alpha \cdot Fmag[i] \quad (4)$$

Where,

$$\alpha = \left(\frac{1}{10} \sum_i ((Fmag[i])^2)\right)^{-1/2} \quad (5)$$

**D. Shaping Filter:**

In MELP, shaping filter is made of five filters. This filters called as synthesis filters, which are used to synthesize the mixed excitation signal during decoding. Each synthesis filter works on particular frequency band, there are defined as 0-500 Hz, 500-1000 Hz, 1-2000 Hz, 2000-3000 Hz and 3000-4000 Hz. This all synthesis filters are connected in parallel to define the frequency responses of the shaping filters. <sup>[7]</sup>

Total response of shaping filter is given by,

$$h_p[n] = \sum_{i=1}^5 v_{s_i} h_i[n] \quad (6)$$

This two shaping filters used only during the decoding if the speech where mixed excitation synthetic speech is generated.

**E. Gain Computation:**

It is a technology to calculate the gain of the speech. Using a pitch-adaptive window length, the input speech signal gain can be measured twice per frame.

1) If  $vs_1 > 0.6$ , the window length is shortest multiple of first stage pitch period ( $T^{(l)}$ ) are longer than 120 samples. This length exceeds up to 320 samples which are divided by 2. If the value of gain is changed than the position of window is minimized.

2) If  $vs_1 \leq 0.6$ , the window length is 120 samples. This used in only for voiced or jittery voiced frames. And the length of window is default. <sup>[7][8]</sup>

In the gain computation the first 90 samples produce  $g_1$  at the first window. At the last sample are used in current frame. The gain computation can be given as follows;

$$g = 10 \log_{10}(0.01 + \frac{1}{N} \sum_n s^2[n]) \quad (7)$$

**F. LP Analysis:**

It is a simple form of the 10<sup>th</sup> order extrapolation. It uses Levinson-Durbin method (Toeplitz matrix). LP analysis is performed on the input speech signal using 200 samples. The final coefficients are bandwidth-expanded with a constant of 0.994. This are quantized and used to find the prediction error signal. <sup>[2]</sup>

**G. Bit Allocation:**

A total 54 bits are transmitted per frame at a frame length of 22.5ms. As it is result in a bit-rate of 2400 bps. Here in the MELP there is 13 bits is provided against the error protection per frame. Synchronization is alternating one/zero pattern.

PARAMETER	VOICED	UNVOICED
LPC	25	25
PITCH PERIOD/LOW-BAND VOICING STRENGTH	7	7
BANDPASS VOICING STRENGTH	4	****
FIRST GAIN	3	3
SECOND GAIN	5	5
APERIODIC FLAG	1	****
FOURIER MAGNITUDES	8	****
SYNCHRONIZATION	1	1
ERROR PROTECTION	****	13
<b>TOTAL</b>	<b>54</b>	<b>54</b>

Table 1 Bit Allocation in MELP

**IV. Melp Decoder**

The decoder is a device which is regenerates the original data from the packed data received at the receiver. The MELP Decoder shows in the Fig 3.

MELP decoder, where the bit stream is unpacked as directed in the decoder.

Two additional filters are added to the processing path:

- 1) The spectral enhancement filter taking the mixed excitation as input.
- 2) The pulse dispersion filter at the end of the processing chain.

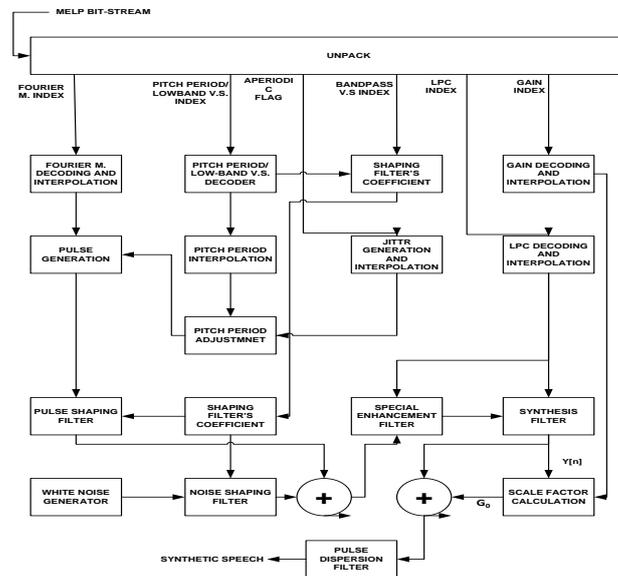


Fig.3 MELP Decoder<sup>[1]</sup>

For unvoiced frames the different values are used. In the MELP decoder, the 50 included to the pitch period, 0.25 for the jittery voice, all 0's for the voicing strengths and all 1's for the Fourier magnitude. the jitter is 0.25 if aperiodic flag is equal to one. The maximum pitch period distortion is around  $\pm 25\%$ . The interpolation factor<sup>[2][3]</sup> is given by,

$$\alpha = \frac{n_0}{180} \quad , \quad \text{Where is} \quad n_0 \in [0,179] \quad (8)$$

This gives the LPC (LSF), pitch period, jitter, Fourier magnitudes, and shaping filters coefficients. By putting all this coefficients together the final speech generated. This speech is approximately exactly same as the input of the speech. This speech is generated from the codebook of the different amplitudes so it is a synthetic

speech. (Not a natural. Its artificial or contrived). As the result the final a synthetic speech is generated at the end of receiver.

## V. Result

### Input and Output for five.wav

TABLE 2Five.wav Coefficients

Wave file	duration	MSE	Absolute	SNR
five.wav	4329	7.0410e+006	1.6810e+003	-0.8683

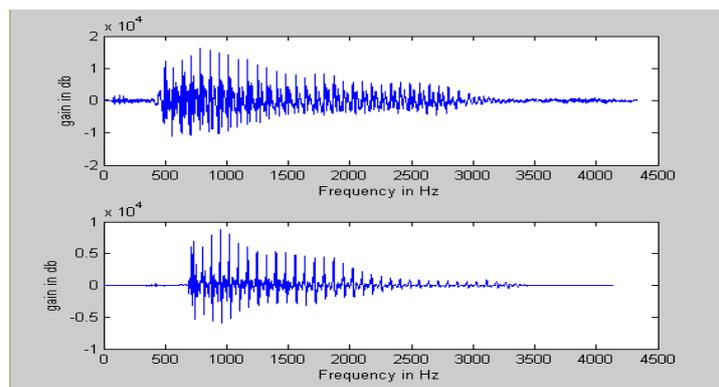


Fig. 3 Input Output Waveform of Five.wav

## VI. Conclusion

After working on this project, compressed and transmitted speech is received with highest possible quality using the least channel capacity. At the low bit rate, MELP speech codec satisfy all the requirements of being good quality speech coder. The lower the bit-rate at which the coder can deliver toll quality speech, the more speech channels can be compressed within a given bandwidth. The speech received at the bit rate of 2.4KB/s without damage our original speech. The results shows the low SNR ratio compress the speech at the same rate but gives a good quality speech, while increasing in SNR ratio gives a slight lower quality speech.

## Acknowledgement

Many have contributed to the successful preparation of our project. We would like to place on record my grateful thanks to each of them. We wish to express our gratitude to our guide Mr. Hareesh Miyani, for his demonstrative and undaunted guidance and support during our project work and giving an opportunity to represent our project. We would like to thank our parents for their kindness support to us and inspiring us. Lastly, we would thank each and every person who has helped us directly or indirectly.

## References

- [1] Wai C. Chu, "Speech coding Algorithms: Foundation and Evolution of Standardized Coders " A John Wiley & Sons, Inc.
- [2] John G Proakis & Dimitris G Manolakis : "Digital Signal Processing : Principles, Algorithms, and Applications" (Prentice Hall India, 3rd edition).
- [3] Sanjit K Mitra : "Digital Signal Processing : A computer Based Approach" (TMH, 2<sup>nd</sup> edition.)
- [4] Theodore S. Rappaport: "Wireless Communications Principles and Practice "( PEARSON Education, 3<sup>rd</sup> edition ).
- [5] Antoniou, A. "Digital Filters: Analysis, Design, and Applications"( McGraw-Hill, New York.).
- [6] Burrus, C. S. and T. W. Parks (1985). "DFT/FFT and Convolution Algorithms", John Wiley & Sons, Hoboken, NJ. (1993).
- [7] Gersho, A. and R. M. Gray. Vector Quantization and Signal Compression, 4th printing, Kluwer Academic Publishers, Norwell, MA. (1995).
- [8] United Status patent paper, "Speech coding system and method including harmonic generator having an adaptive phase off setter" (Richard Louis Zinser, Mark Lewis Grabb, Steven Robert Koch, Willium Brooksby.).
- [9] Advanced Electronic Communications Systems (Wayne Tomasi, 5<sup>th</sup> Edition.).

## AUTHORS



**Hareesh Miyani** was born in Bhavnagar, Gujarat, India in 1987. He passed his B.E. from C.K. Pithawala Engg. & Tech., Surat. He passed M.Tech from VJTI in 2012. He is qualified for GATE. His area of interest is Digital Signal Processing, Image Prosessing, MATLAB. He is Asst. Proff. At the Bhagwan Mahavir College of Engg. & Tech., Surat. He has one year training at the IIT, Bombay campus.



**Aalay Mehta** was born in Mehsana, Gujarat, India in 1993. He is Post Graduate in BE-Electronics Engineering, Bhagwan Mahavir College of Engineering and Technology (Gujarat Technical University), Surat. His area of interest is to work on Digital Signal Processing, Power Electronics, Embedded System and software (MATLAB).



**Pratik Nai** was born in Bharuch, Gujarat, India in 1992. He is Post Graduate in BE-Electronics Engineering, Bhagwan Mahavir College of Engineering and Technology (Gujarat Technical University), Surat. His area of interest is to work on Digital Signal Processing, Power Electronics, Embedded System and software (MATLAB), Digital logic Analog System.



**Harshad Patel** was born in Surat, Gujarat, India in 1992. He is Post Graduate in BE-Electronics Engineering, Bhagwan Mahavir College of Engineering and Technology (Gujarat Technical University), Surat. His area of interest is to work on Digital Signal Processing, Embedded System and software (MATLAB), Digital Communication.