A Technical report on the Design and MATLAB Implementation of

Residual Excited Vocoder

By
Amogh Rajanna
4JC02EC006
ECE Department, SJCE.
amoevol@yahoo.co.in

Submitted to

DEPARTMENT OF ELECTRONICS AND COMMUNICATION
Sri Jayachamarajendra College of Engineering, MYSORE.
570 006

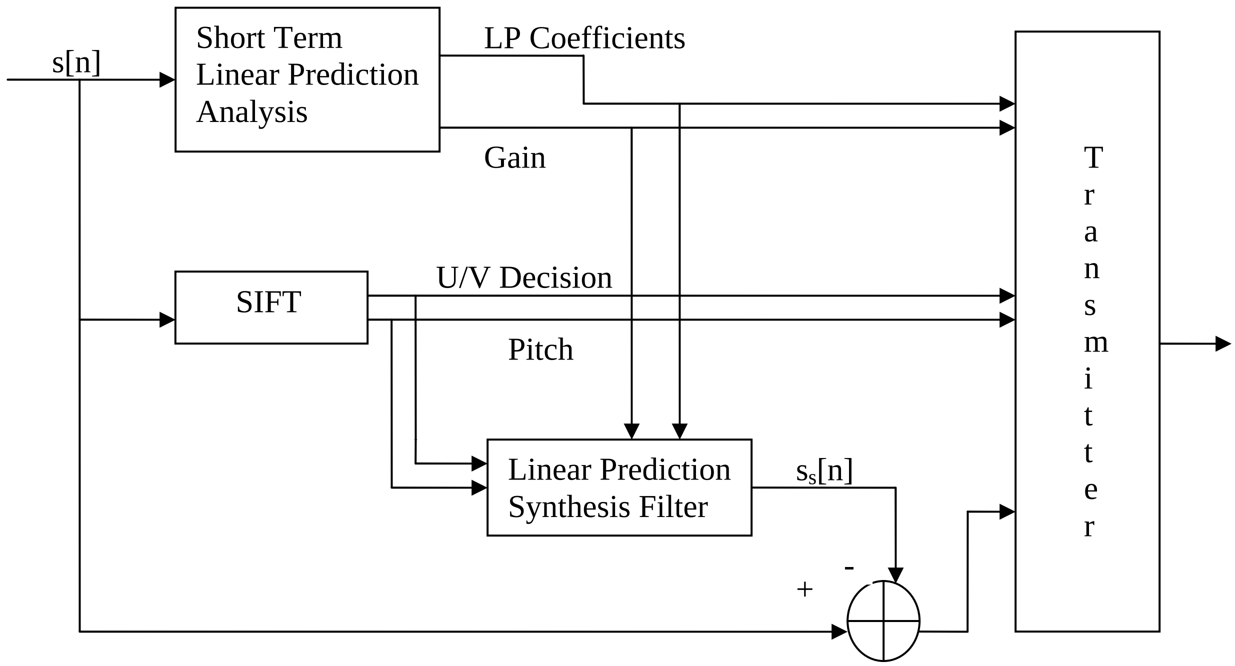


Diagram 1: Transmitter portion of the Vocoder

Short Term Linear Prediction Analysis:

The input speech signal sp[n] is segmented into frames. A speech frame consists of 320 samples.

SF[i, n] = sp[(i – 1)* 320 + n] ; 1<=n<=320

Each speech frame is windowed by a Hamming window of length 320.Each speech frame is windowed to taper the speech frame at the edges to zero.
Hamming Window w[n] = 0.54 - 0.46 cos(2πn/319);   0<=n<=319
SF[i,n] = w[n-1]SF[i,n]; 1<=n<=320

The autocorrelation analysis of the windowed speech frame leads to an autocorrelation vector given by

AC[m] = $\sum_{1}^{320} F[n]F[n+m]$; 0<= m <=p

where F[n] is the n^{th} sample in the speech frame, p is the LPC order.

LPC Coefficients are derived recursively from autocorrelation coefficients using the standard Levinson-Durbin Algorithm.

k[i] = (AC[i] - $\sum_{1}^{i-1} AL[i-1, j]\, AC[i-j]$) / E[i - 1]

AL[i , i] = k[i]
AL[i , j] = AL[i-1,j] - k[i]AL[i-1,i - j];
1<= j <= (i - 1)

E[i] = (1 - $k[i]^2$)E[i - 1]

The algorithm is repeated for the range
1<= i<= p and then, the LPC coefficients are AL[p , j] for 1<= j <=p.

Pitch Extraction SIFT algorithm:

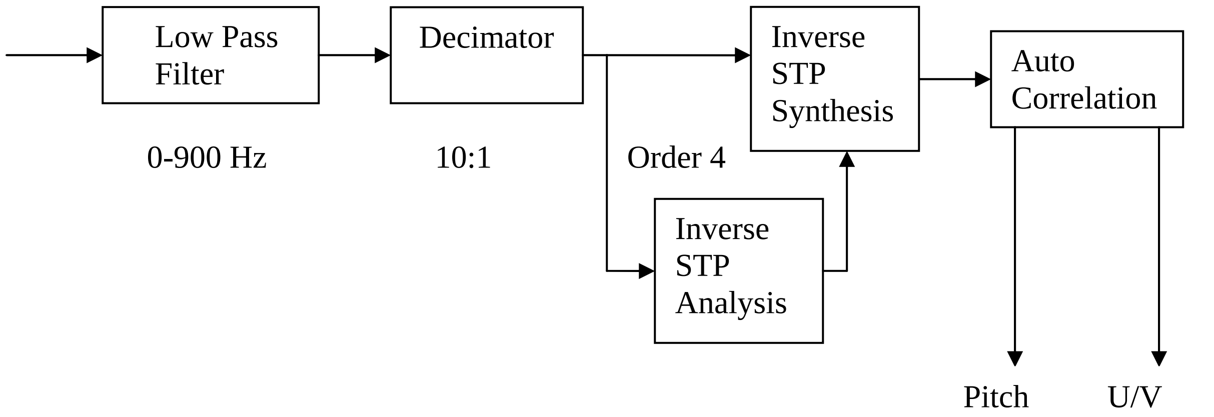


Diagram 2: SIFT Algorithm

- Low pass filtering and Decimation:

The input speech signal is low pass filtered to a bandwidth of 1kHz with a sampling rate of 2kHz.

- Inverse STP Analysis:

The low pass filtered signal is subjected to Inverse Short Term Prediction Analysis yielding the Linear prediction coefficients and gain parameter for the Inverse STP synthesis filter.

- Inverse STP Synthesis:

From the low pass filtered signal and parameters from the inverse STP Analysis, the inverse STP synthesis filter calculates the Linear prediction error.

- Autocorrelation:

Performing the autocorrelation of the linear prediction error sequence, pitch of the speech frame is estimated from the autocorrelation peak sample index. For U/V decision, the algorithm is given below.

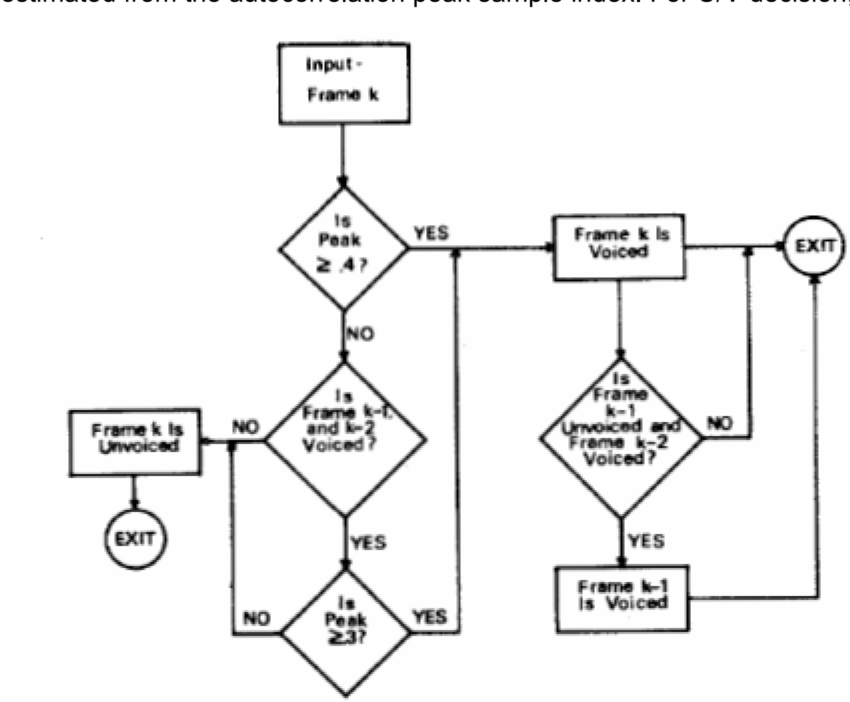


Diagram 3: Algorithm for Unvoiced/Voiced Decision.

For a detailed treatment of the SIFT algorithm, refer to the paper "*The **SIFT** Algorithm for Fundamental Frequency Estimation*" by John D Markel, cited in the reference section.

Linear Prediction Synthesis Filter:

Using the

1. Short Term Linear Prediction Coefficients
2. Gain parameter
3. Pitch period
4. U/V Decision

the synthesis filter which is a IIR filter synthesizes the speech from linear prediction analysis.

$$y[n] = \sum_{1}^{p} A(k)y[n-k] + Gu[n]$$

where y[n] = speech sample at time 'n'.
A[k] = Linear prediction coefficient
Gu[n] = Excitation Sequence

Excitation Sequence Generator:

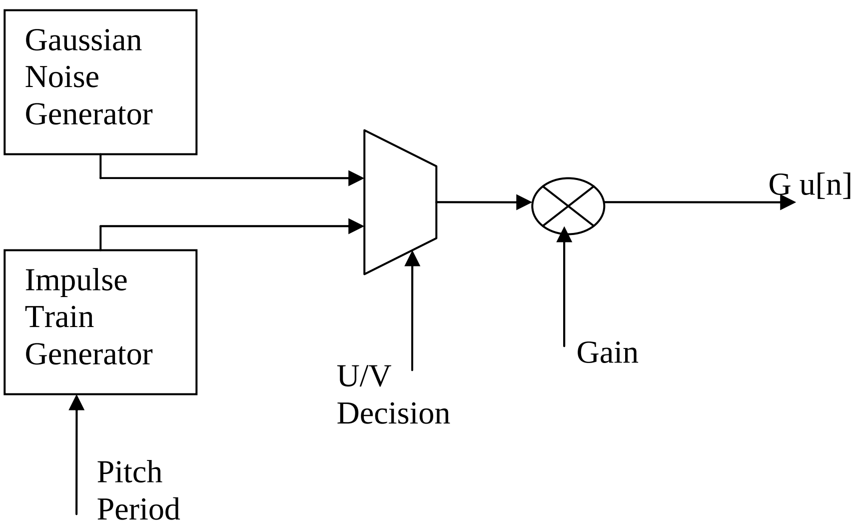


Diagram 4: Excitation Sequence Generator.

The residual speech is obtained as the difference between original speech and synthesized speech at the transmitter portion. The transmitter portion of the Vocoder transmits STP coefficients, gain, pitch period, U/V Decision and the residual speech.

Receiver portion of the Vocoder:

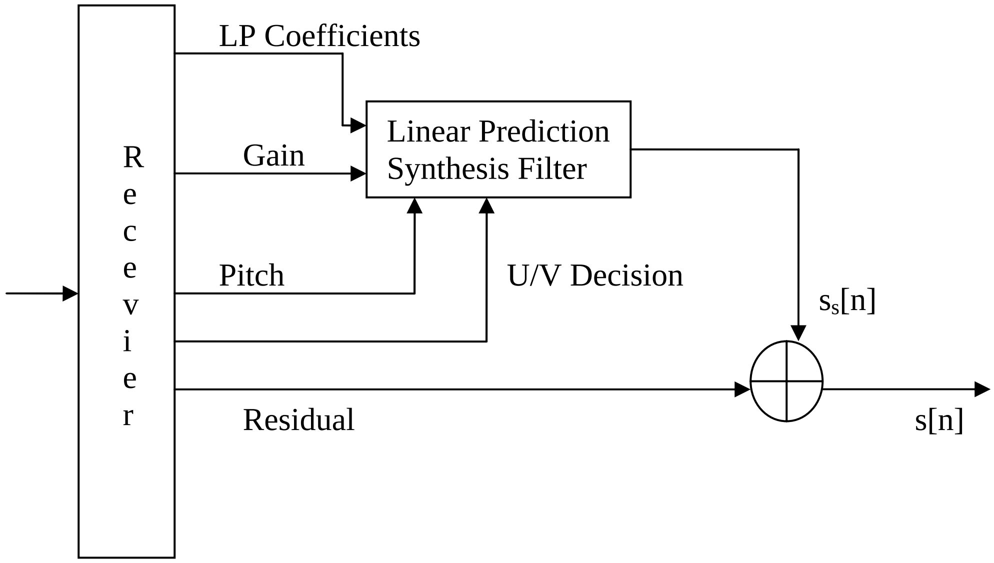


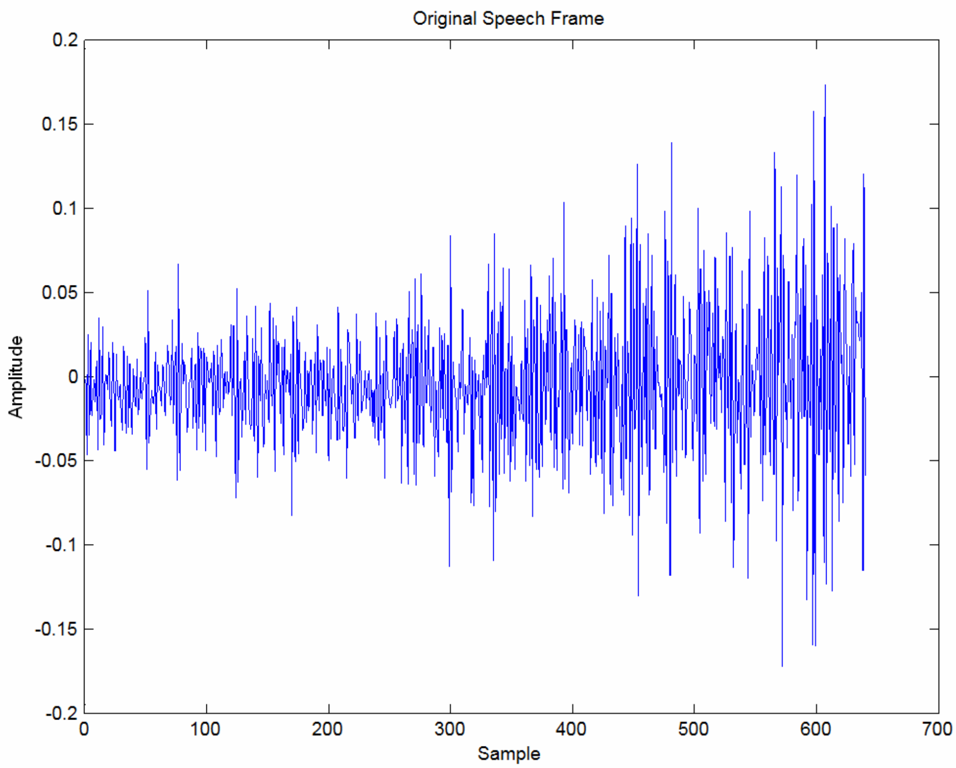
Diagram 5: Receiver portion of the Vocoder.

At the receiver portion, the linear prediction synthesis filter synthesizes the speech from STP coefficients, gain, pitch period, U/V Decision. The synthesized speech is combined with the residual speech to obtain the reconstructed speech frame.

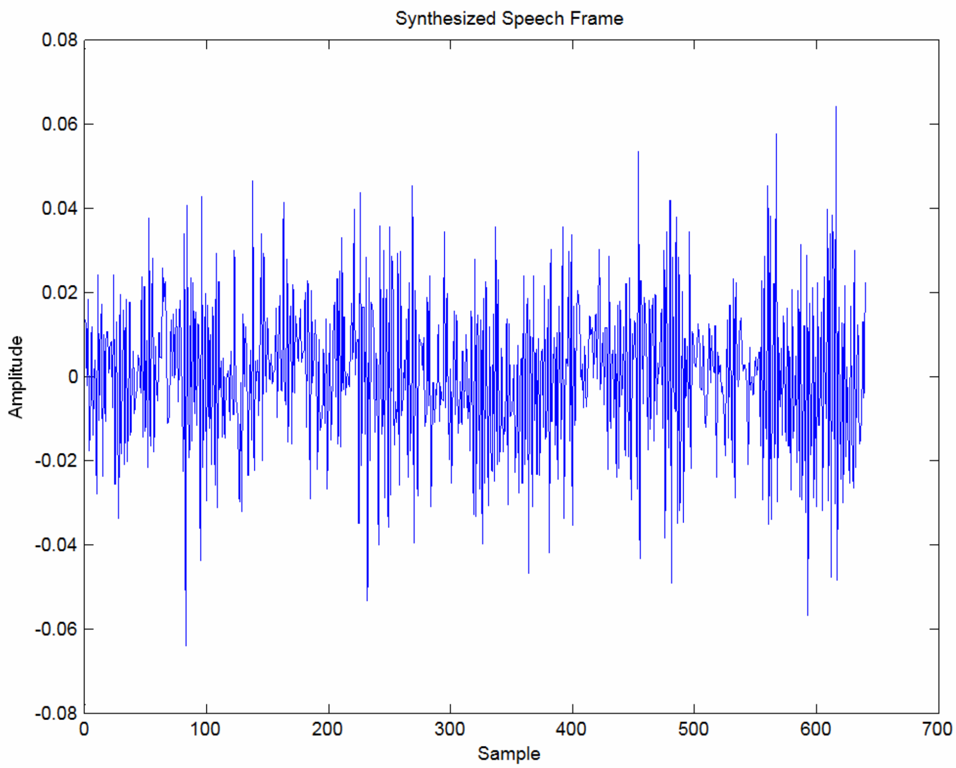
Values of the RELP Vocoder for a particular speech frame:

Sl	Parameter	Values
1	Linear prediction coefficients p = 10	-0.56394 0.13684 0.32025 -0.20456 -0.0008373 0.17523 0.26588 0.0056979 0.12529 0.092729
2	Gain	0.2052
3	Pitch period	9 samples
4	U/V Decision [1- voiced; 0- unvoiced]	1
5	Inverse Linear prediction coefficients p = 4	-0.027392 -0.092598 -0.06445 -0.44191
6	Inverse Gain	0.66706

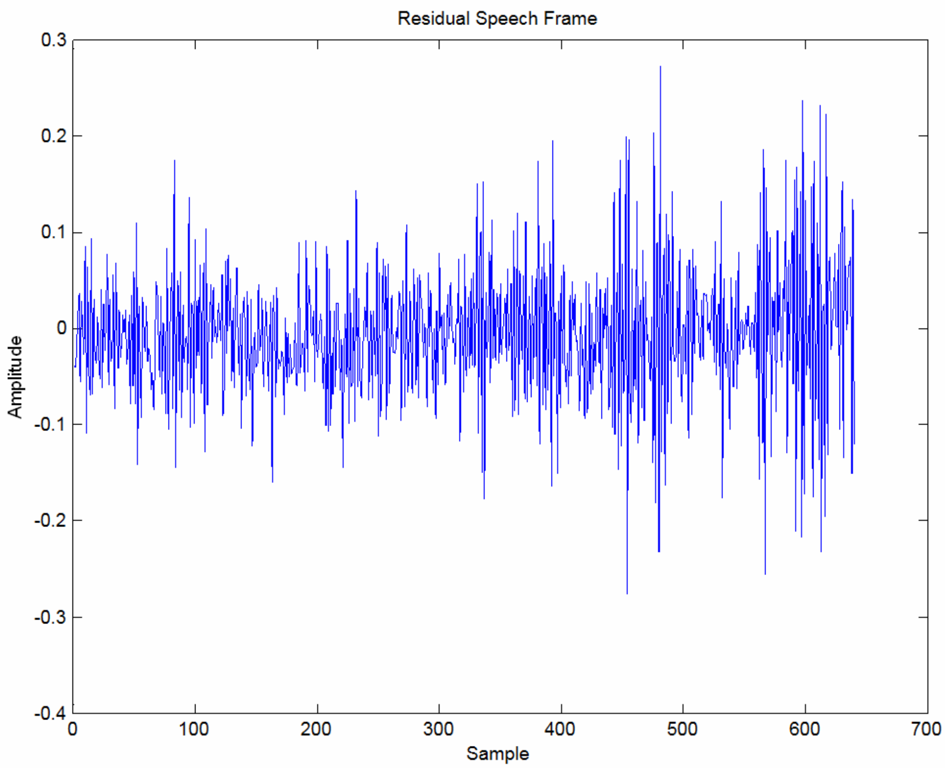
Original Speech Frame:



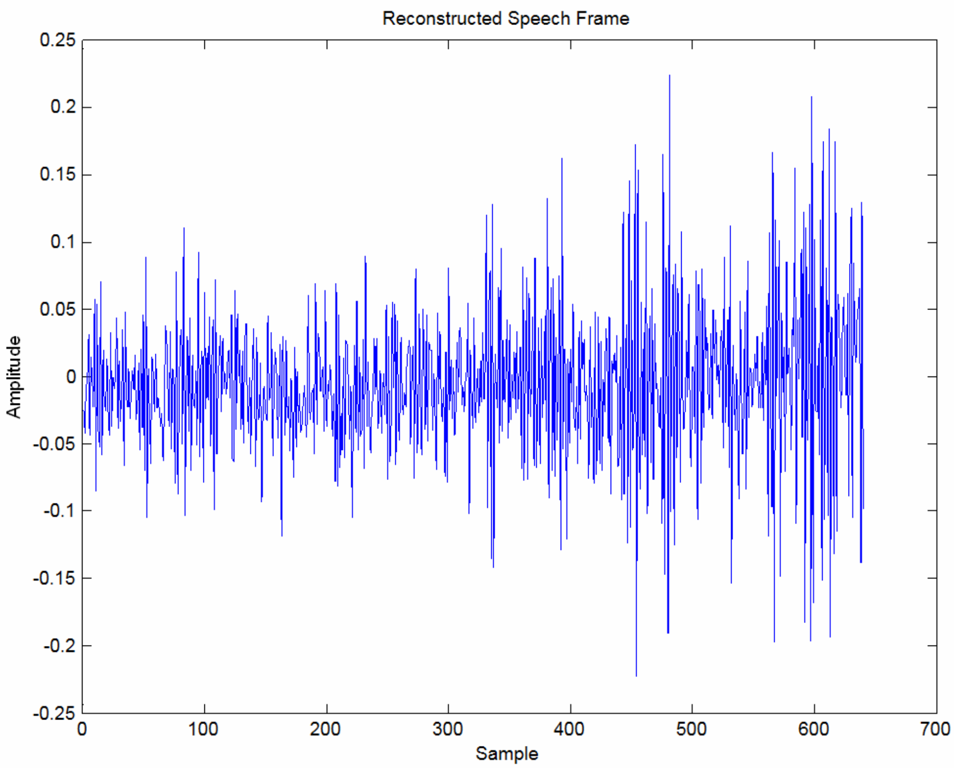
Synthesized Speech Frame:



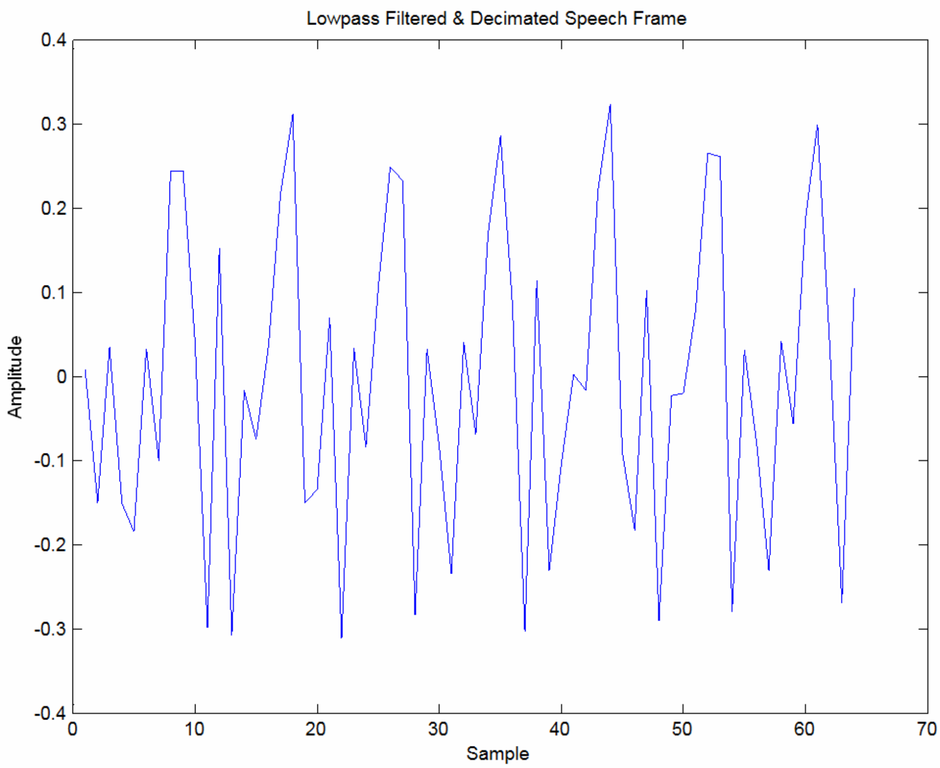
Residual Speech Frame:



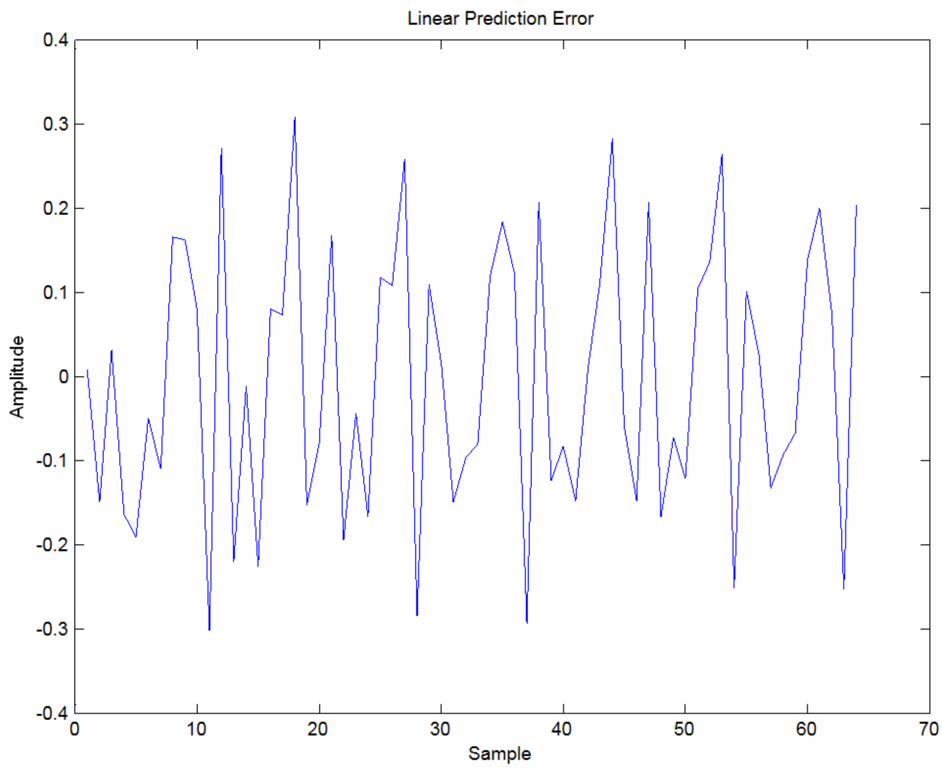
Reconstructed Speech Frame:



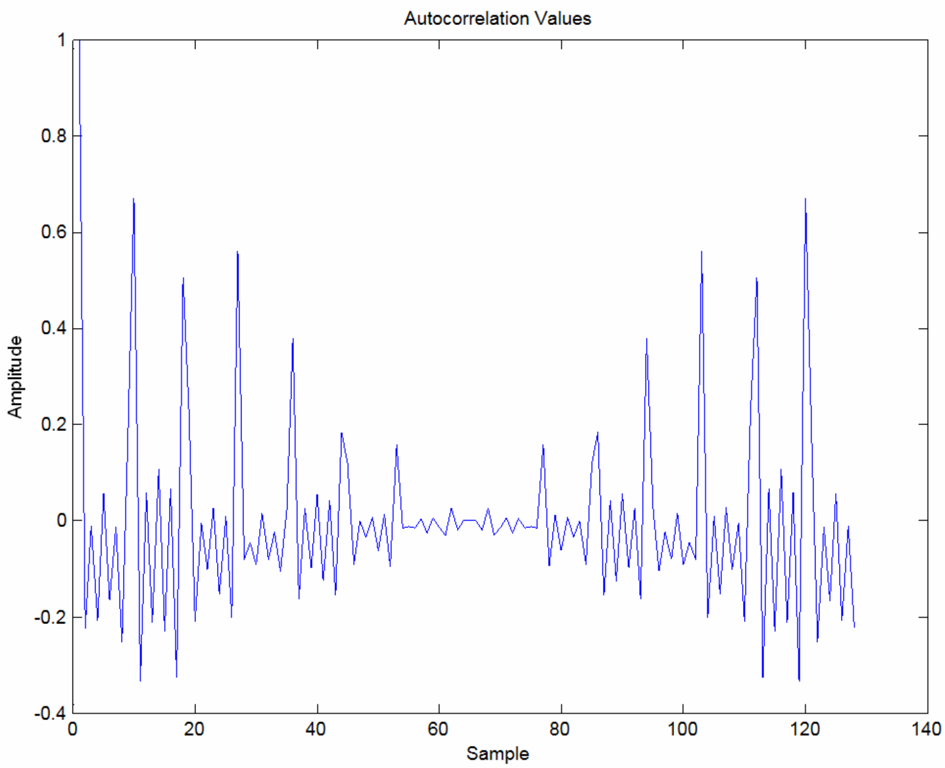
Lowpass Filtered & Decimated Speech Frame:



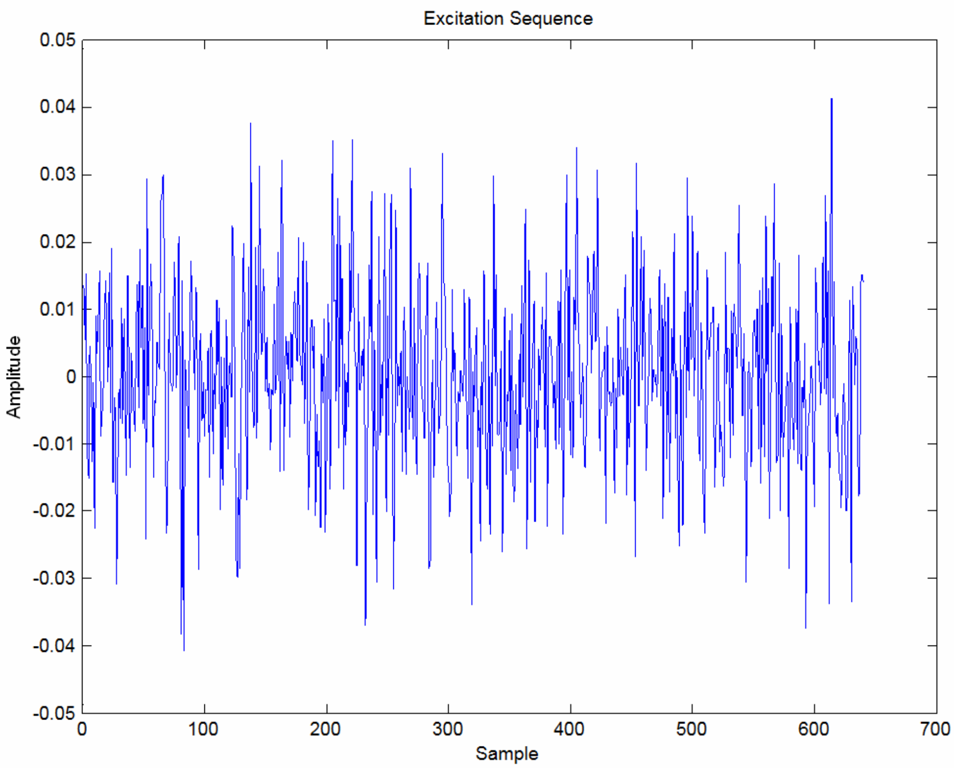
Linear Prediction Error of the SIFT algorithm:



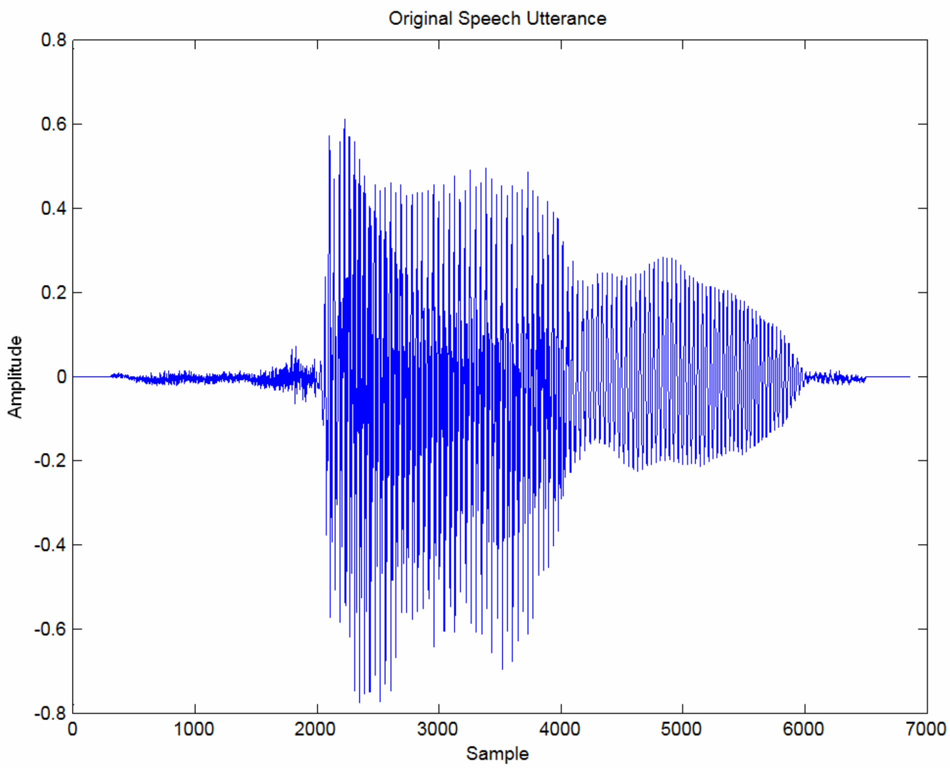
Autocorrelation Values of the SIFT algorithm:



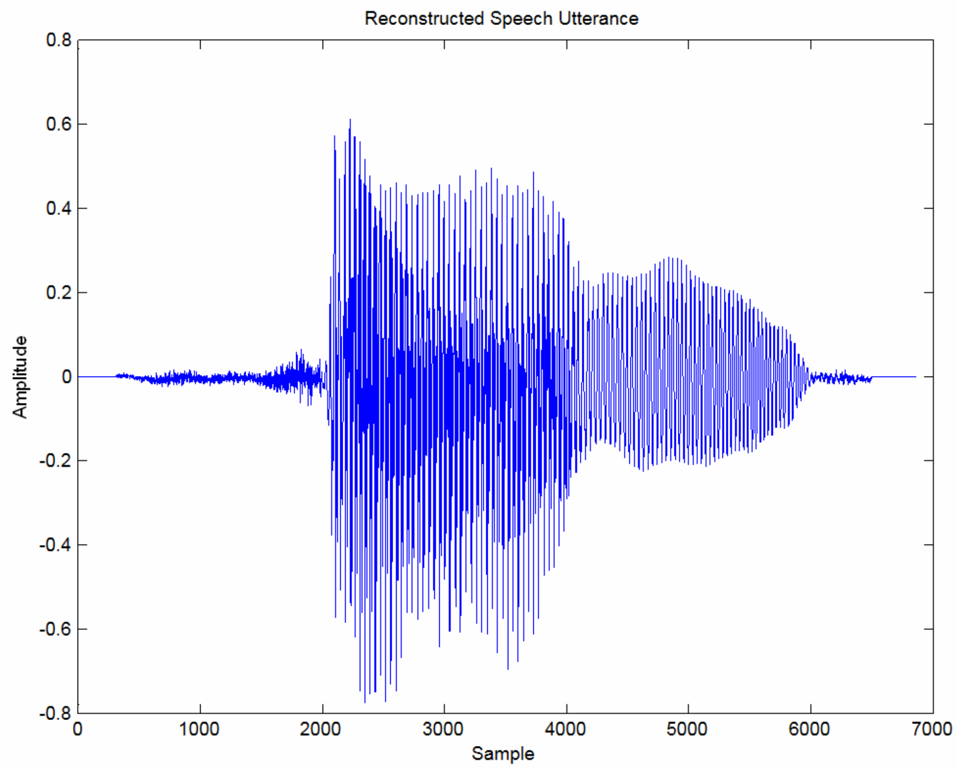
Excitation Sequence:



Original Speech Utterance:



Reconstructed Speech Utterance:



References:

[1] "*The **SIFT** Algorithm for Fundamental Frequency Estimation*" by John D Markel, IEEE Transactions **on** Audio and Electroacoustics, vol. no. **5,** Aug-20, December 1972.
[2] "Wireless Communications: Principles and Practice" Theodore S Rappaport, Pearson Education, Second Edition, 2002.
[3] "Voice Compression and Communciations: Principles and Applications for Fixed and Wireless channels" Lajos Hanzo etal, Wiley-IEEE press, 1st Edition, August 2001.