# ECE 252A Speech Compression

# Term Project-
# LPC-10 coding


## Name: Cheng Qian
## PID: A53209561
## Email: chq019@eng.ucsd.edu


Academic Integrity Policy: Integrity of scholarship is essential for an academic community. The University expects that both faculty and students will honor this principle and in so doing protect the validity of University intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind.

By including this in my report, I agree to abide by the Academic Integrity Policy mentioned above.

# Content

# Instruction：

1- Record a sentence with a sampling rate of 8kHz.

2- Follow the guidelines described in Question 9.4 of Chu's book (Chu, Wai C. Speech coding algorithms: foundation and evolution of standardized coders. John Wiley & Sons, 2004) to implement an LPC-10 encoder.

Do not quantize the LPC parameters, pitch and gain values (you can store your data in a Matlab file).

3- Design also a decoder which can read stored data and synthesize the speech.

Submit

- A report describing (i) how you made voiced/unvoiced decisions for each frame; (ii) how you determined the pitch period; (iii) plot the time-domain waveform of a voiced frame.
- Submit the original speech and the encoded/decoded speech (audio files)
- Also, comment on the quality of the synthesized speech in your report.
- Submit your Matlab or c codes.
- You can put everything into a zip file.

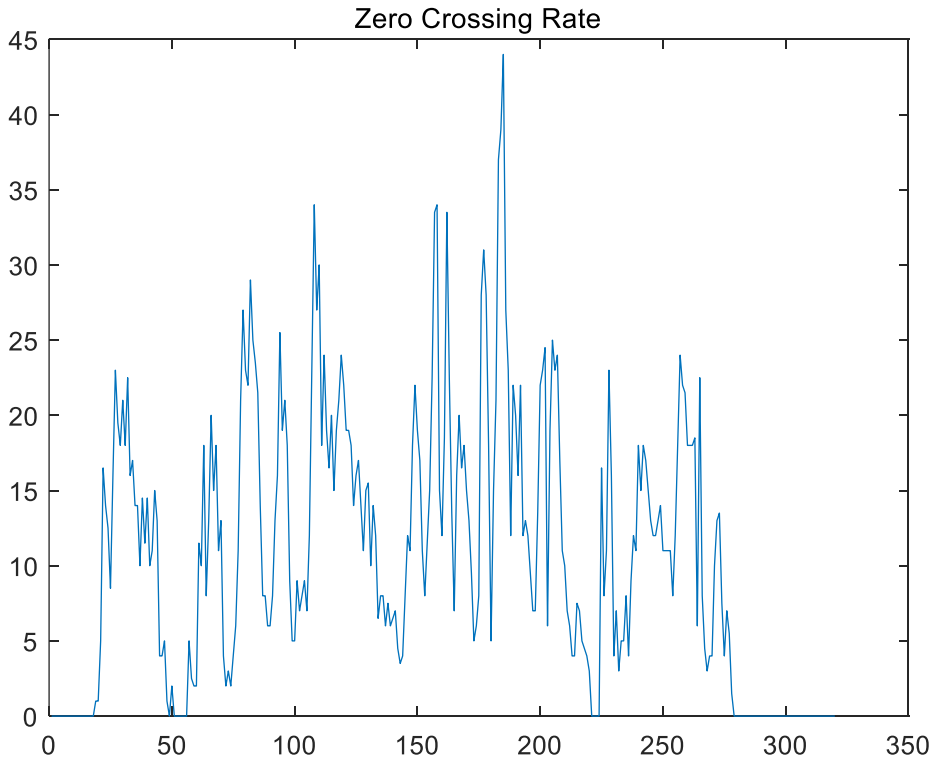
# Question:

(i) how you made voiced/unvoiced decisions for each frame

Energy for voiced speech tends to concentrate below 3kHz. Unvoiced speech energy is found at higher frequencies. Since high frequencies imply high zero crossing rates, the zero crossing rate is a good indicator of voiced/unvoiced frame. I calculate the zero –crossing rate within each frame as

$$Z = 1/L \sum_{n=0}^{L-1} sign(s[n]) - sign(s[n-1])$$

Then determine a maximum likelihood threshold such that $Zav, voiced < Zav, unvoiced$.



I set 25 as the boundary as the voiced frames. All the frame has a higher zero crossing rate then 25 will be distinguished as unvoiced frame. And those lower than 25 will be distinguished as voiced frame.

## (ii) how you determined the pitch period;

The auto-correlation function is

$$\mathrm{R}(\tau) = \int f(t) f(t-\tau)\, dt$$

If $\mathrm{f(t)}$ is a periodic signal with period T

$$\mathrm{f(t)} = \mathrm{f(t-T)}$$

Then the auto-correlation function of this signal will be

$$\mathrm{R}(\tau+\mathrm{T}) = \int f(t) f\big(t - (\tau+T)\big)\, dt$$

$$= \int f(t) f\big(t-\tau-T)\big)\, dt$$

$$= \int f(t) f\big(t-\tau)\big)\, dt$$

$$= \mathrm{R}(\tau)$$

Thus, the auto-correlation function is also periodic with the same period T.

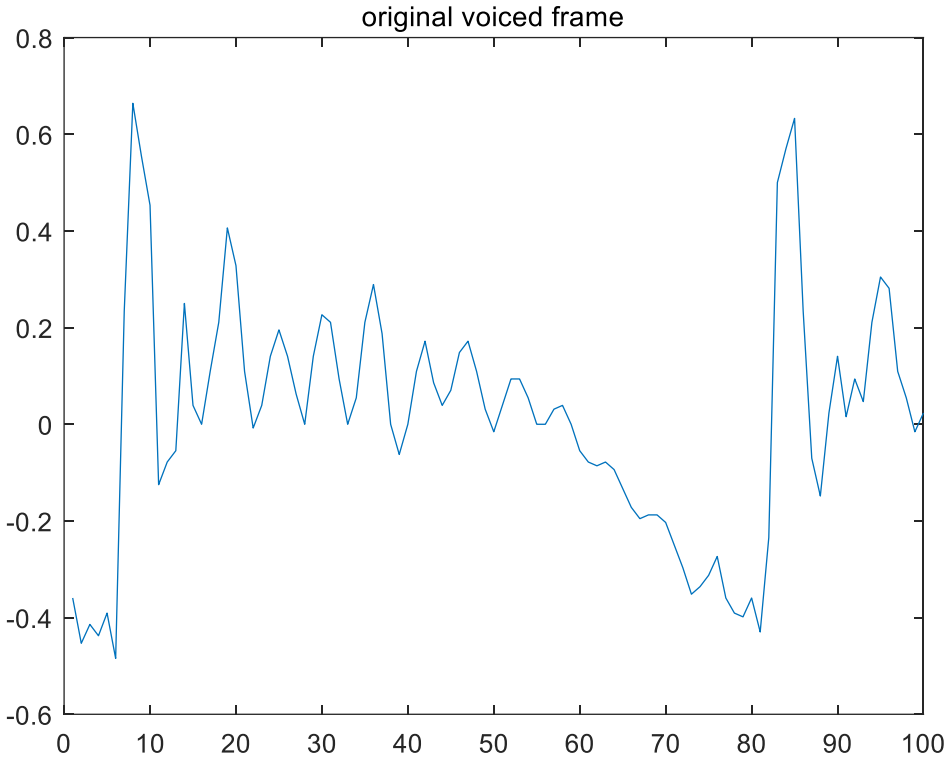
We can get the conclusion that if the function is periodic, its autocorrelation function must be periodic with the same period. The autocorrelation function of voiced signal has peak at the integral multiple of pitch period, while the autocorrelation function of unvoiced signal does not have obvious peak. So we need to find the peak of autocorrelation function.

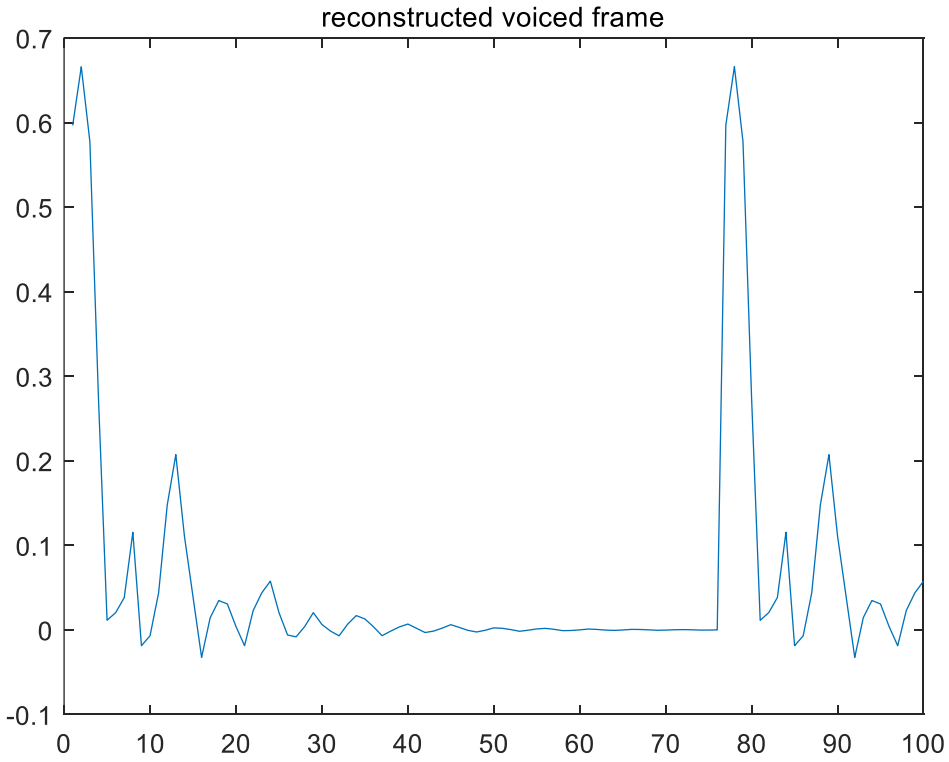
$$R(k) = \sum sign(n) * sign(n-k)$$

Actually we can estimate the pitch period by taking the index of the first max value of autocorrelation function. Before that, I need to discard the first 10 points to avoid misjudgment.

## (iii) plot the time-domain waveform of a voiced frame.

These two figures are the time-domain waveform of a voice frame respectively from original signal and reconstructed signal.






It can tell from the two figures that my system simply reconstruct the frequency and amplitude of original signal.

# Implementation step:

To run my script, first run record.m to record your sentence, then run encoder.m and finally run decoder.m. The two waves will be stored into original_speech and decoded_speech. You can listen to them with your audio player.

## 1. Encoder.

At encoder, we divde the frames in to voiced and unvoiced groups according to zero crossing rate. Then calculate pitch perid, LPC coefficient and gain.

## Voicing detector:

Energy for voiced speech tends to concentrate below 3kHz. Unvoiced speech energy is found at higher frequencies. Since high frequencies imply high zero crossing rates, the zero crossing rate is a good indicator of voiced/unvoiced frame. I calculate the zero –crossing rate within each frame as

$$Z = 1/L \sum_{n=0}^{L-1} sign(s[n]) - sign(s[n-1])$$

Then determine a maximum likelihood threshold such that $Zav, voiced < Zav, unvoiced$. I set 45 as the boundary as the voiced frames. All the frame has a higher zero crossing rate then 45 will be distinguished as unvoiced frame. And those lower than 45 will be distinguished as voiced frame.

## LP analysis:

Solve the autocorrelation normal equation to get the LPC coefficient for each frame.

as expressed in matrix form

$$\begin{bmatrix} R_{\hat{n}}(0) & R_{\hat{n}}(1) & . & . & R_{\hat{n}}(p-1) \\ R_{\hat{n}}(1) & R_{\hat{n}}(0) & . & . & R_{\hat{n}}(p-2) \\ . & . & . & . & . \\ . & . & . & . & . \\ R_{\hat{n}}(p-1) & R_{\hat{n}}(p-2) & . & . & R_{\hat{n}}(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_{\hat{n}}(1) \\ R_{\hat{n}}(2) \\ . \\ . \\ R_{\hat{n}}(p) \end{bmatrix}$$

$$\Re \boldsymbol{\alpha} = \boldsymbol{r}$$

with solution

$$\boldsymbol{\alpha} = \Re^{-1} \boldsymbol{r}$$

Use the formula below to calculate gain for each frame.

$$E^p = G^2 = \mathrm{R}(0) - \sum_{k=1}^{p} \alpha_k R(k)$$

In this project, I directly use the built in function [coefficient,gain] = lpc(signal,order). The gain that calculated is actuaclly the power of gain, so we need to square it to get the real gain.

## Prediction-error filter and power computation:

In this system we do not transmit error signal.

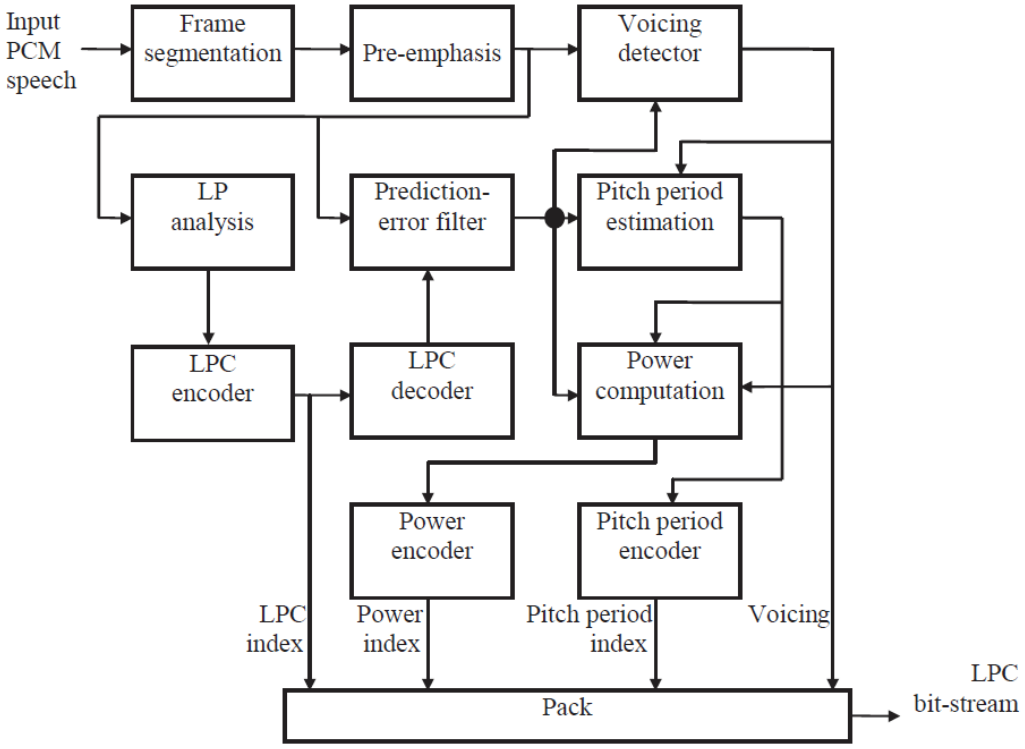
## Pitch period estimation:

If the function is periodic, its autocorrelation function must be periodic with the same period. The autocorrelation function of voiced signal has peak at the integral multiple of pitch period, while the autocorrelation function of unvoiced signal does not have obvious peak. So we need to find the peak of autocorrelation function.

$$R(k) = \sum sign(n) * sign(n-k)$$

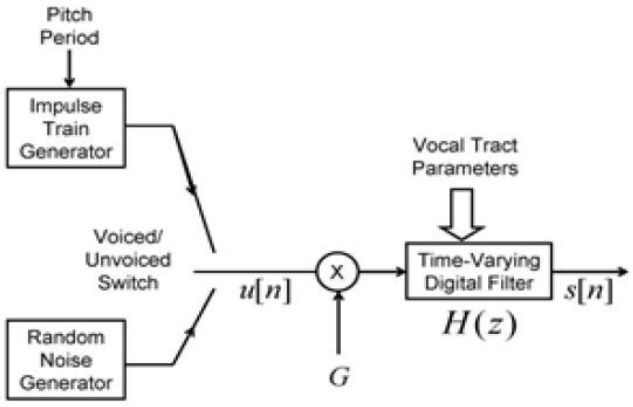
Actually we can estimate the pitch period by taking the index of the first max value of autocorrelation function. Before that, I need to discard the first 10 points to avoid misjudgment.

## Integration




After encoded, the related features(gains, lpc coefficients, zero crossing rates, pitch period) are stored into a matlab file.

## 2. Decoder.

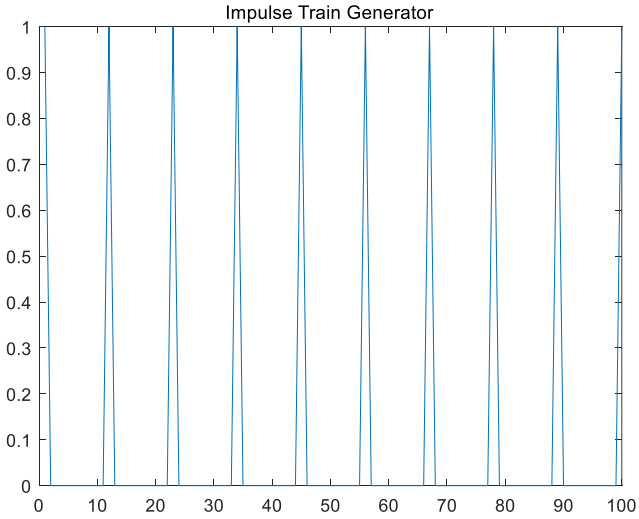



Accrding to zero crossing rate, the frames are divided into two groups: voiced and unvoiced frames. At decoder, we load from the previous matlab file to get

the related features. The excitation signal of voiced signal is generated by impulse train generator and unvoiced by random noise generator. Then the signal was multuplied by the lpc gain and filter by LPC inverse filter $H(z)$ to get the reconstructed signal.

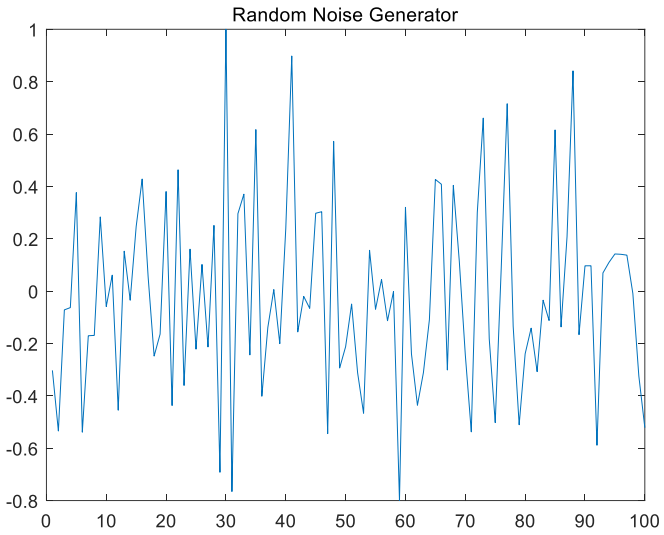
## Impulse train generator:

The excitation signal of voiced signal is an impulse train with period of pitch period of the corresponding frame that we have calculated before.




## White noise generator:

The excitation signal of unvoiced signal is a random noise sequence.




## Synthesis filter:

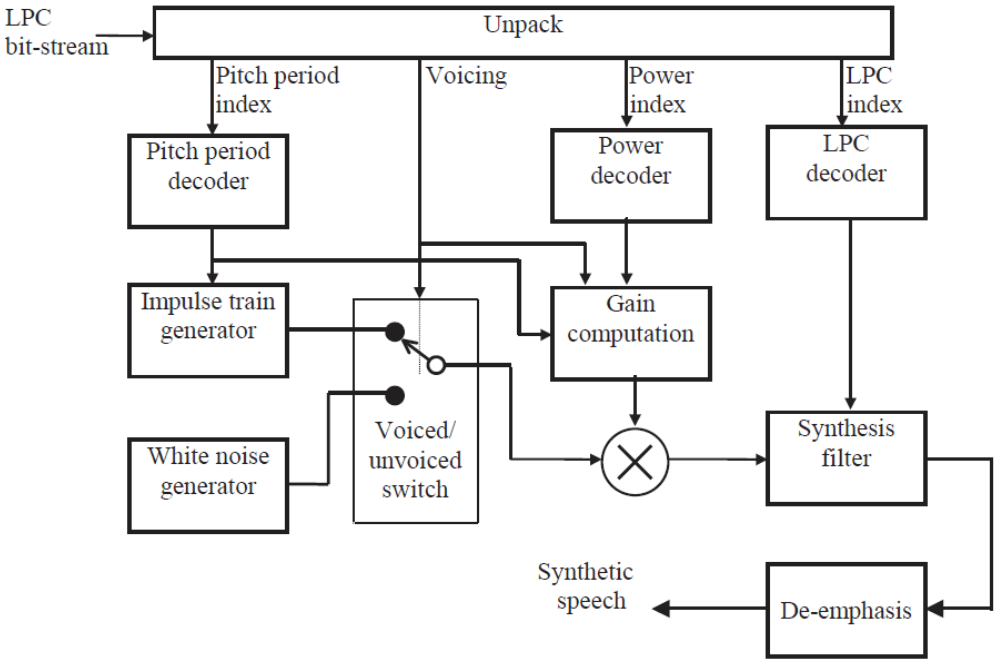
$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}$$

In matlab, I implemented this filter with

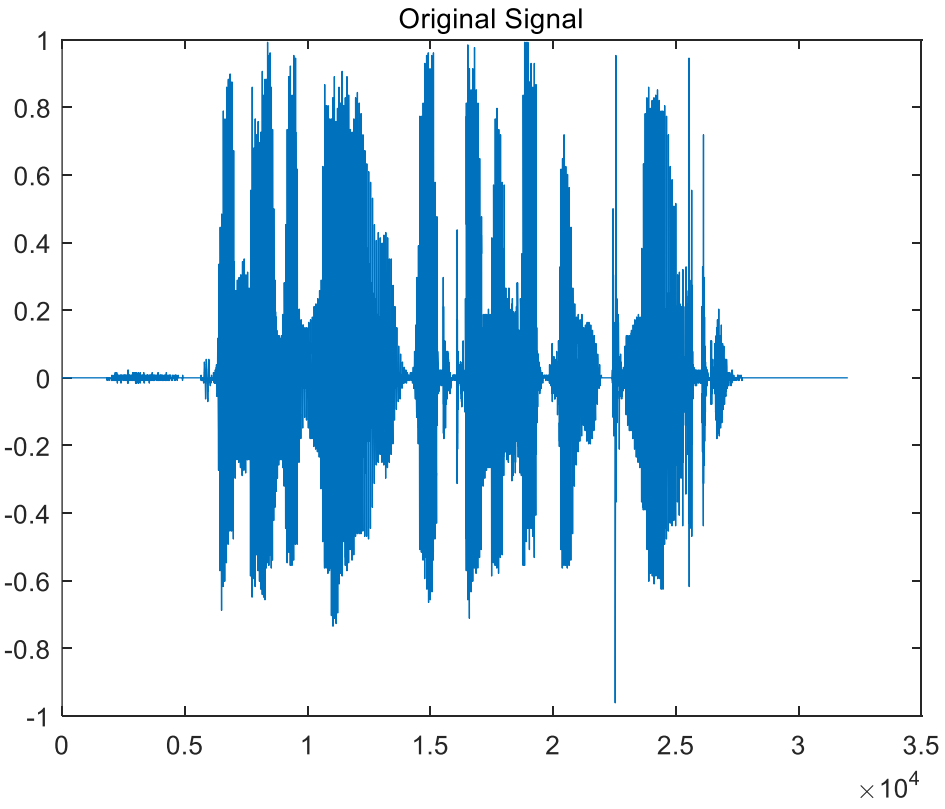
$$H(z) = \frac{gain}{coefficient(frame,:)}$$

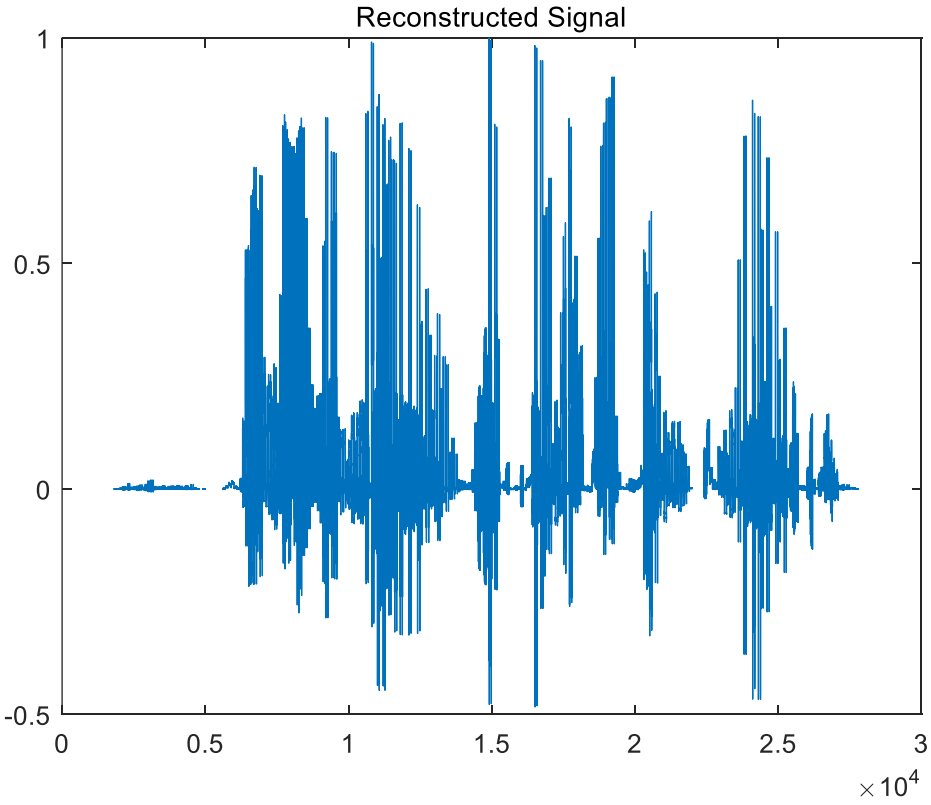
filter(gain,coefficient(frame,:),excitation_signal)

## Integration:

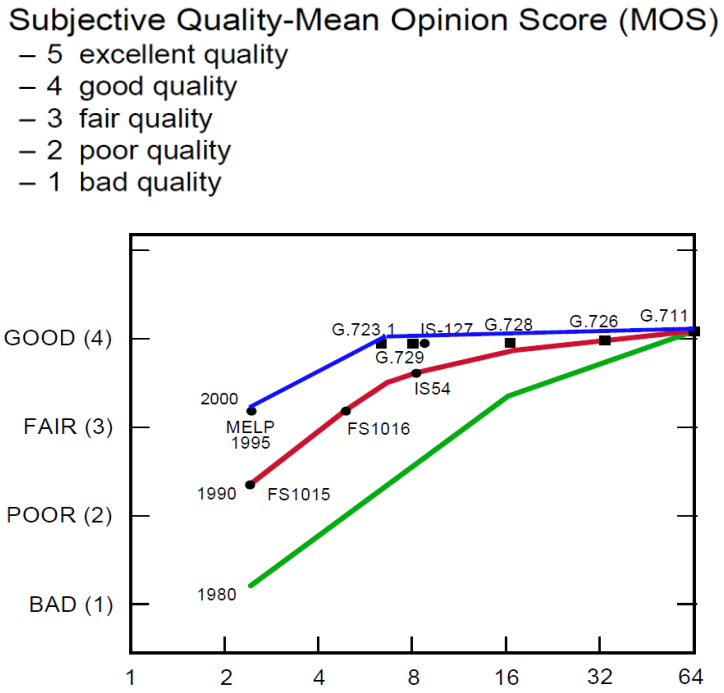



3. Test of encoder–decoder operation.






The mse between original and reconstructed signal is 0.0317




Using the system developed so far, speech is input to the encoder with the resultant parameters processed by the decoder. The intelligibility of the synthetic speech is generic, but only ‘intelligible’. There are some clip noise at the

boundary of voiced and unvoiced frames. And the pitch sounds same through all the frames. The performance of linear predictive coding is actually very poor. According to the speech coder subjective quality that the professor showed in the lecture, our LPC-10 coder's performance is between bad(1) and poor(2).

## 4. Incorporation of parameter-quantizers.

We are not using quantizer in this project according to the instruction.

## 5. Overall test and system improvement.

I tried to use power as the feature to determine voiced/unvoiced signal. However, the perfomance is same as the one using zero crossing rate. This may not be a feasible way to optimize.