CS630: Speech Technology
LAB-6: Linear Prediction Analysis

OBJECTIVE:

To study the features of speech from linear prediction analysis.

SEQUENCE OF STEPS:

(a) Take a segment (200 msec) of voiced speech /a/ and a segment (200 msec) of unvoiced speech /s/.

(b) Compute short-time (20 msec) spectrum, inverse spectrum and 14^{th} order LP spectrum for a voiced segment (/a/).

(c) Compute short-time (20 msec) spectrum, inverse spectrum and 14^{th} order LP spectrum for unvoiced segment (/s/).

(d) Examine the LP residual for voiced and unvoiced segments.

(e) Compute autocorrelation function of signal and its LP residual for voiced and unvoiced segments.

(f) Obtain LP residual for the entire 200 msec of vowel and integrate to examine the glottal pulse shape

(g) Obtain LP spectrum for a voiced segment for different orders of LP p=14,10,6,3,1

(h) Obtain normalized error for different orders for a voiced and unvoiced segment

(i) Write a brief note on the observations

1 Collection of voiced (/a/) and unvoiced (/s/) speech segments

- Record a vowel /a/ for one second at 10 KHz sampling frequency with 16 bit quantization. From this recorded speech file collect 200 ms in steady portion of the waveform.
- Record an unvoiced segment /s/ for one second at 10 KHz sampling sampling frequency with 16 bit quantization. From this recorded speech file collect 200 ms in steady portion of the waveform.
- The short voiced and unvoiced speech segments are shown in Figure 1.
- Observation:
  In voiced speech segment (/a/) periodicity (pitch) is observed, where as in unvoiced speech segment (/s/) signal appears random in nature.

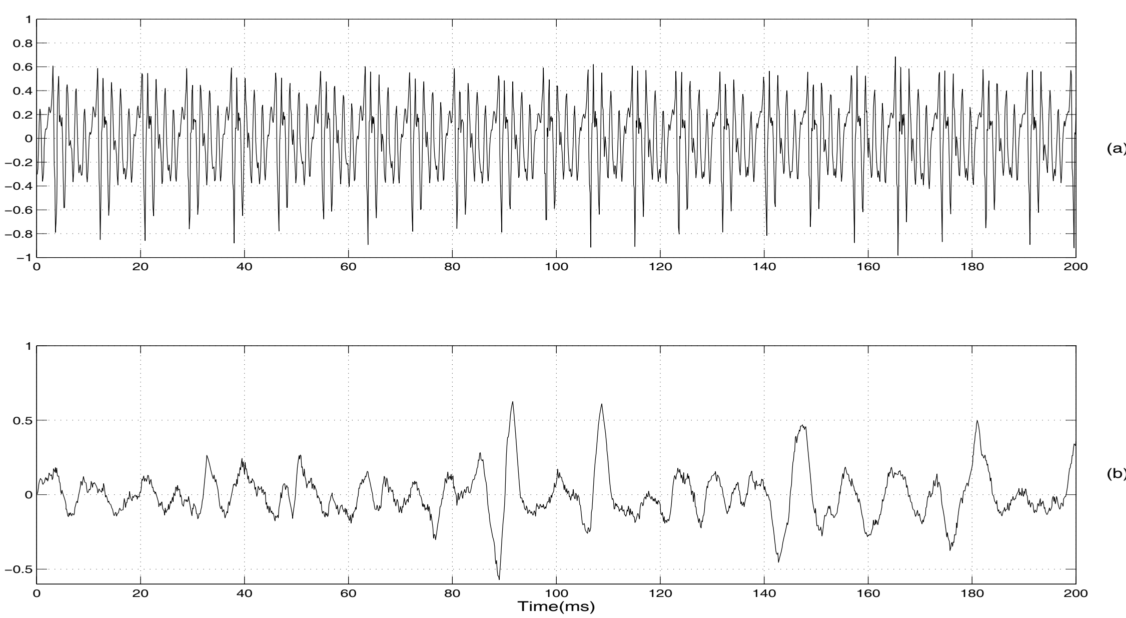

Figure 1: (a) Segment of voiced speech /a/ and (b) segment of unvoiced speech /s/

2 Short time spectrum, LP spectrum and Inverse spectrum for voiced segment /a/

2.1 Short time spectrum

The short time spectrum consists of range of frequencies (magnitude and phase components) that are present in a small segment (10-30 ms) of a signal. Short time spectrum is computed with the following procedure:

- Take 20 ms of voiced speech segment /a/ after preemphasis
  ```
  a=wavread('vowel.wav');
  a=diff(a);
  a200=a(501:700);
  ```

- Apply a hamming window over a voiced segment (a200), then compute the fast Fourier transform for the voiced segment (a200) and plot the magnetite of the spectrum.
  ```
  ham=hamming(200);
  a200ham=a200.*ham;
  a200hamspec=fft(a200ham,1024);
  y=abs(a200hamspec.*a200hamspec);
  logy=10*log10(y);
  figure;plot([1:512]*5000/512,logy(1:512));grid;
  ```

2.2 LP spectrum

- LP spectrum provides smoothed envelope of the short time spectrum, where only the formant frequencies (resonances) are observed. For realizing this linear prediction coefficients (LPCs) or filter parameters need to be computed from speech signal.
  ```
  ak=lpc(a200,14);
  lpspec=freqz(1,ak);
  y=abs(lpspec.*lpspec);
  logy=10*log10(y);
  figure;plot([1:512]*5000/512,logy);grid;
  ```

2.3 Inverse spectrum

- Inverse filter is realized by the inverse of LP filter. Therefore the spectrum of the inverse filter is computed as follows:

```
invspec=freqz(ak,1);
y=abs(invspec.*invspec);
logy=10*log10(y);
figure;plot([1:512]*5000/512,logy);grid;
```

The short time spectrum, LP spectrum and inverse spectrum for a segment of voiced speech are shown in Figure 2.

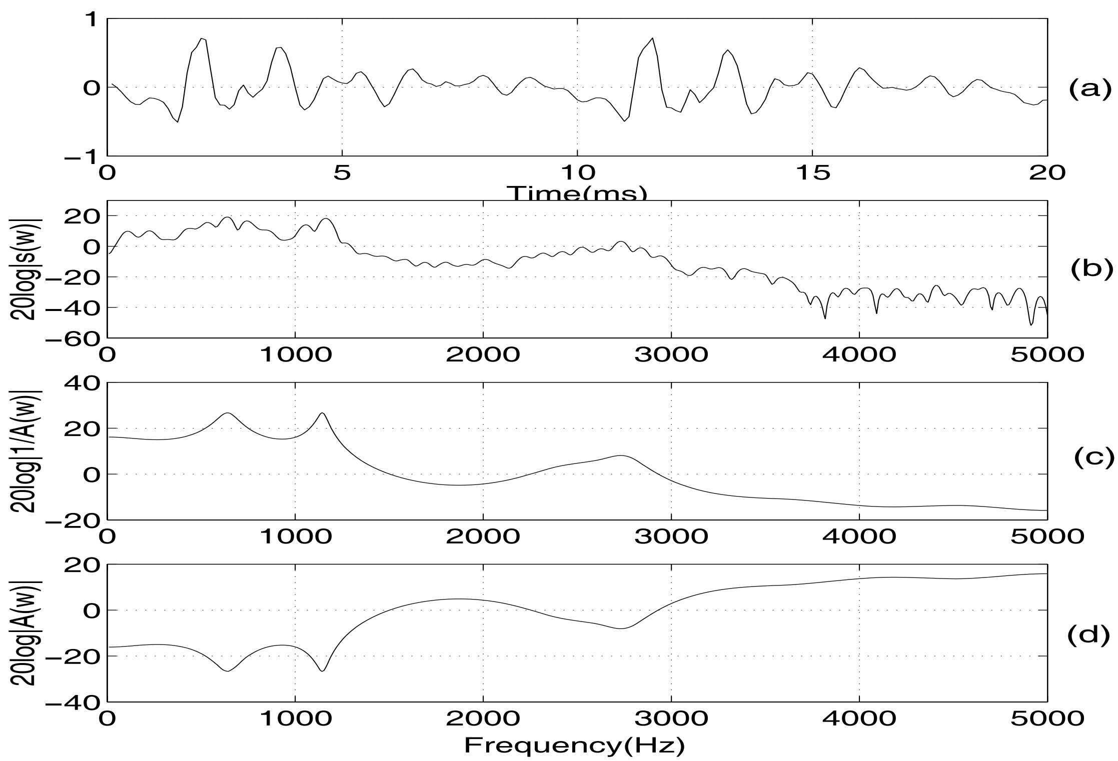

Figure 2: (a) Segment of voiced speech /a/ and its (b) short time spectrum, (c) LP spectrum and (d) inverse spectrum

- **Observation:**

- Short time spectrum gives both source and system information. The envelope of the spectrum gives system information (i.e., resonances in terms of formant frequencies) and spectral ripples (fine variations) give source information (i.e., pitch harmonics). It is a real and even function of ω.

- The linear prediction (LP) analysis models the vocal tract system. LP spectrum observed to be an envelope of short time spectrum (smoothed version of short time spectrum) and the peaks in LP spectrum indicate the formant frequencies (resonances) of the vocal tract system. With observation it is evident that the LP spectrum is derived from an all-pole filter.

- The inverse spectrum is observed to be reciprocal of the LP spectrum. Therefore we can observe the valleys correspond to the peaks in LP spectrum. It is represented by an all-zero filter.

3 Short time spectrum, LP spectrum and Inverse spectrum for unvoiced segment /s/

- For computing the short time spectrum, LP spectrum and inverse spectrum for an unvoiced segment /s/ the same procedure is followed as that of for voiced speech segment /a/.

- The short time spectrum, LP spectrum and inverse spectrum for a segment of unvoiced speech (/s/) are shown in Figure 3.

- Observation:
  In short time spectrum envelope no major peaks are observed. Pitch harmonics (periodic ripples) also not observed.
  In LP spec rum sharp peaks are not present and spectrum observed to be flat. In inverse spectrum also no major events are observed.

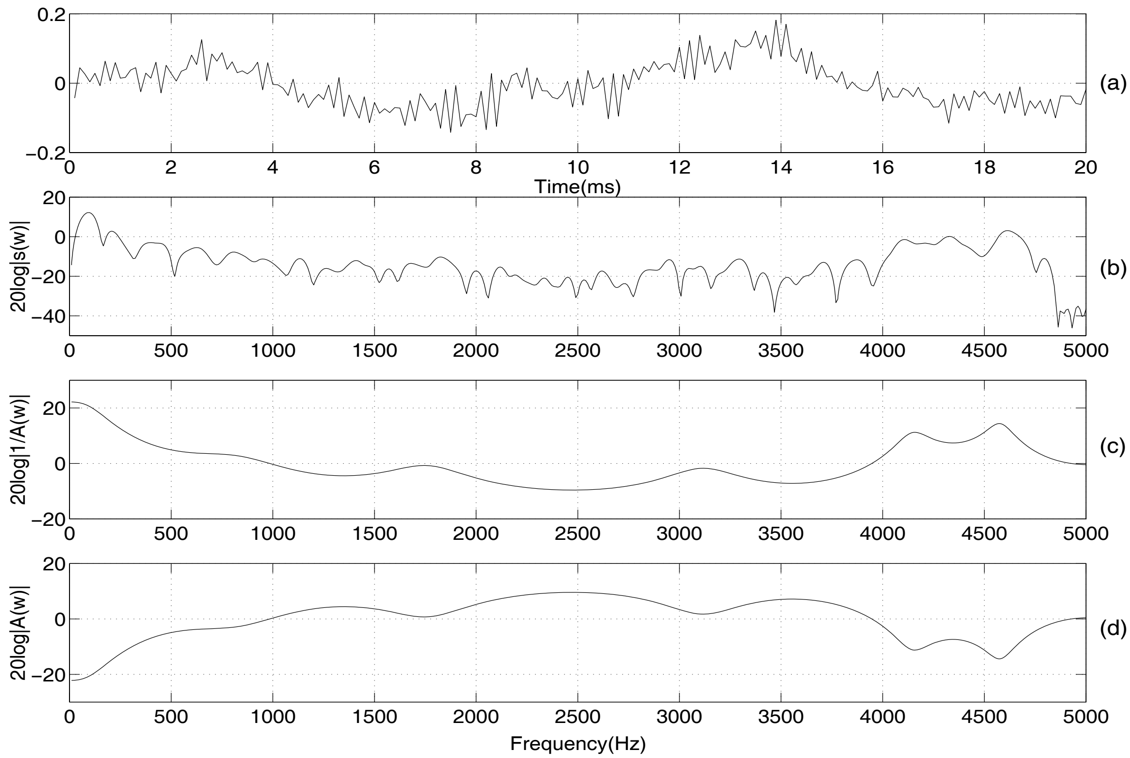

Figure 3: (a) Segment of unvoiced speech /s/ and its (b) short time spectrum, (c) LP spectrum and (d) inverse spectrum

4 LP residual for voiced and unvoiced segments

- LP residual signal is obtained by passing the speech signal through inverse filter designed with LP coefficients (LPCs). The block diagram of the inverse filter is shown in Figure 4.

- LP residual is computed for voiced and unvoiced speech segments using the following matlab commands:
  res=filter(ak,1,a200);
  figure;plot(real(res));grid;

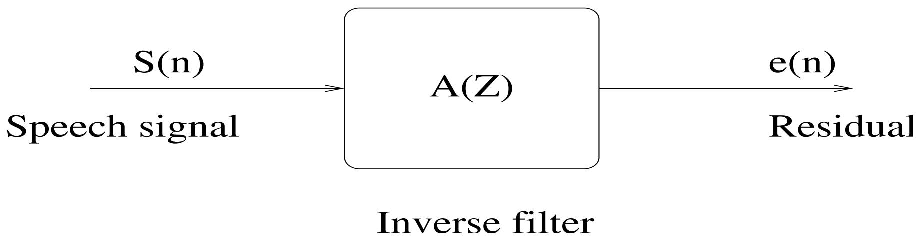

Figure 4: Inverse filter to obtain LP residual signal from speech signal

- Voiced and unvoiced speech segments and their LP residual signals are shown in Figure 5.
- **Observation:**
  For voiced speech segment its LP residual is observed to be periodic. In LP residual signal, peak amplitudes refers to closure of vocal folds (glottal closure), where the prediction is poor therefore its results as maximum error. The periodicity in LP residual also indicate the pitch information.
  LP residual is a result of passing the speech signal through inverse filter (i.e., removing the vocal tract information). This is also considered to be the excitation signal (source information).
  LP residual for unvoiced speech segment looks like random noise. As the unvoiced speech signal has no periodicity and looks like random noise (no relations among the samples), obviously its input also looks like noise.

5 Autocorrelation function for voiced/unvoiced speech segments and their LP residuals

- Autocorrelation function of the signal x(t) is computed using the following formulation:
  $R(\tau) = \sum_{t=0}^{\infty} x(t)x(t+\tau)$
- The above formulation is implemented in matlab using the command xcorr(x(t)). Autocorrelation function for voiced and unvoiced segments

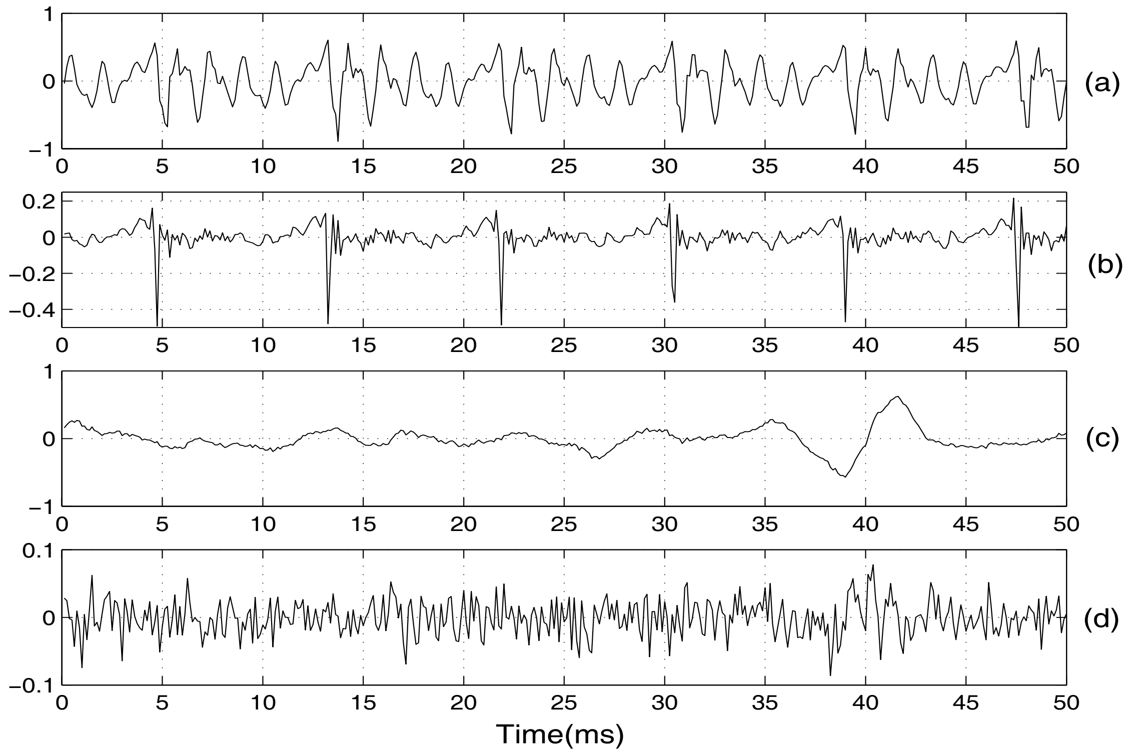

Figure 5: (a) Segment of voiced speech /a/ and its (b) LP residual signal, (c) segment of unvoiced speech /s/ and its (d) LP residual signal.

and their LP residuals is computed.
a200corr=xcorr(a200);

- The autocorrelation function for the voiced speech segment and its LP residual signal is shown in Figure 6.
- The autocorrelation function for the unvoiced speech segment and its LP residual signal is shown in Figure 7.
- **Observation:**
  The basic property of the autocorrelation function (even symmetry) is evident in all (voiced/unvoiced speech and LP residual signals) the plots.
  The samples in a voiced speech segment are highly correlated, there-

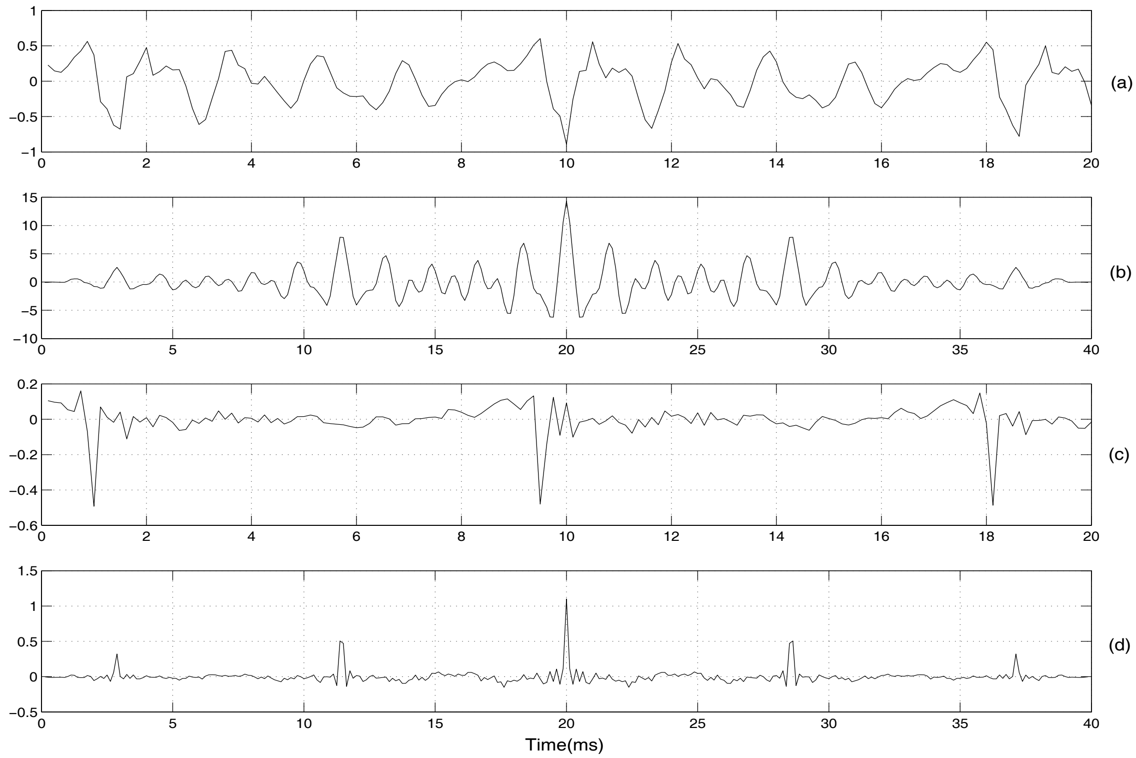

Figure 6: (a) Segment of voiced speech /a/ and its (b) autocorrelation function, (c) LP residual for the voiced speech segment and its (d) autocorrelation function.

fore we will observe the peaks other than center are also prominent. As voiced speech is periodic, it is inherited in its autocorrelation function also.

In LP residual, the correlation among the samples is less, therefore its autocorrelation function contains the peaks at pitch rate. Hence autocorrelation function of a LP residual is useful for pitch computation.

The autocorrelation function of an unvoiced speech segment shows a major peak at the center and other peaks are not significant, since unvoiced speech signal appears like random signal.

The autocorrelation function for the LP residual of an unvoiced speech segment shows a dominant peak at the center and no other peaks in rest of the portion. This is because, unvoiced speech itself looks like random (no correlation among the samples) and its residual reflects

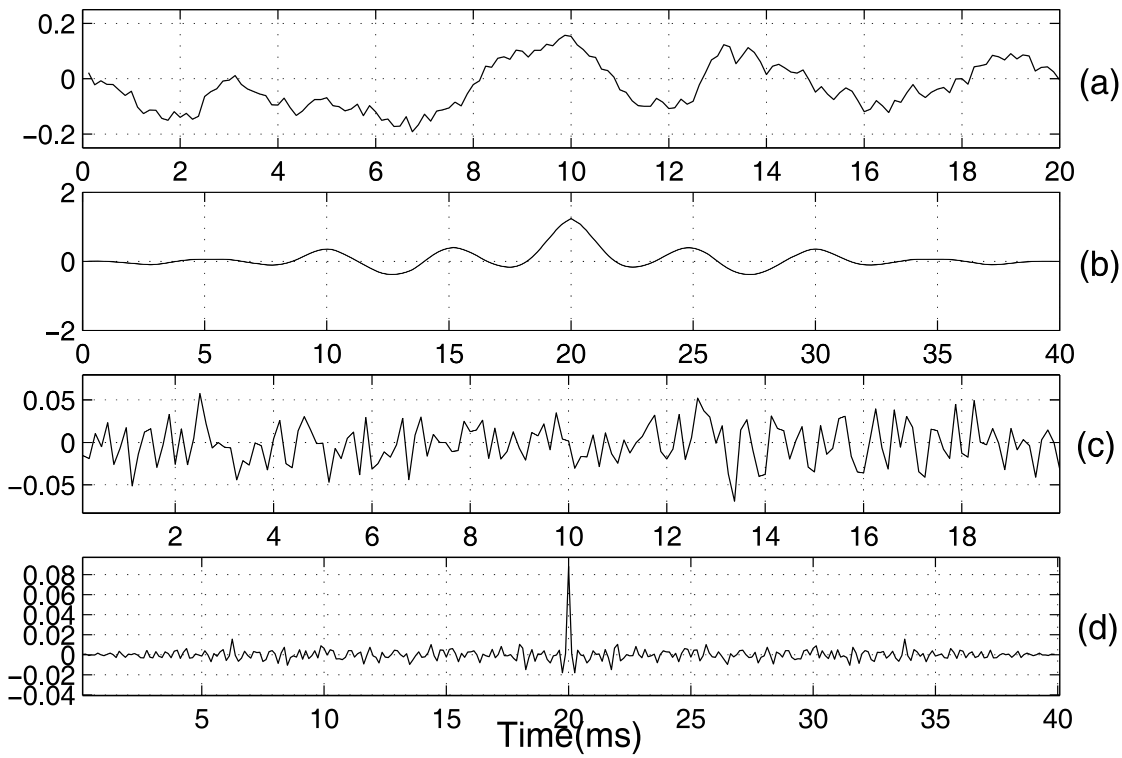

Figure 7: (a) Segment of unvoiced speech /s/ and its (b) autocorrelation function, (c) LP residual for the unvoiced speech segment and its (d) autocorrelation function.

still random.

6 Glottal pulse shape in voiced portion of a speech signal

- By integrating the LP residual we can obtain the glottal pulse shape, it is also known as glottal volume velocity.
- The integration function is implemented in matlab with a function *cumsum*.
  gp=cumsum(res);

- A segment of voiced speech its LP residual and glottal pulse (glottal volume velocity) waveforms are shown in Figure 8.

- **Observation:**
  This gives the information about the air pressure build up near vocal folds from lungs, which cause the vibration of vocal folds resulting in its open/closure. Glottal pulse shape shows the change in volume of air. It is also referred to as glottal volume velocity. From the glottal pulse waveform, it is observed that volume of air and its pressure will be maximum at the instant of closure and then vocal folds will open, with that air pressure will decrease.

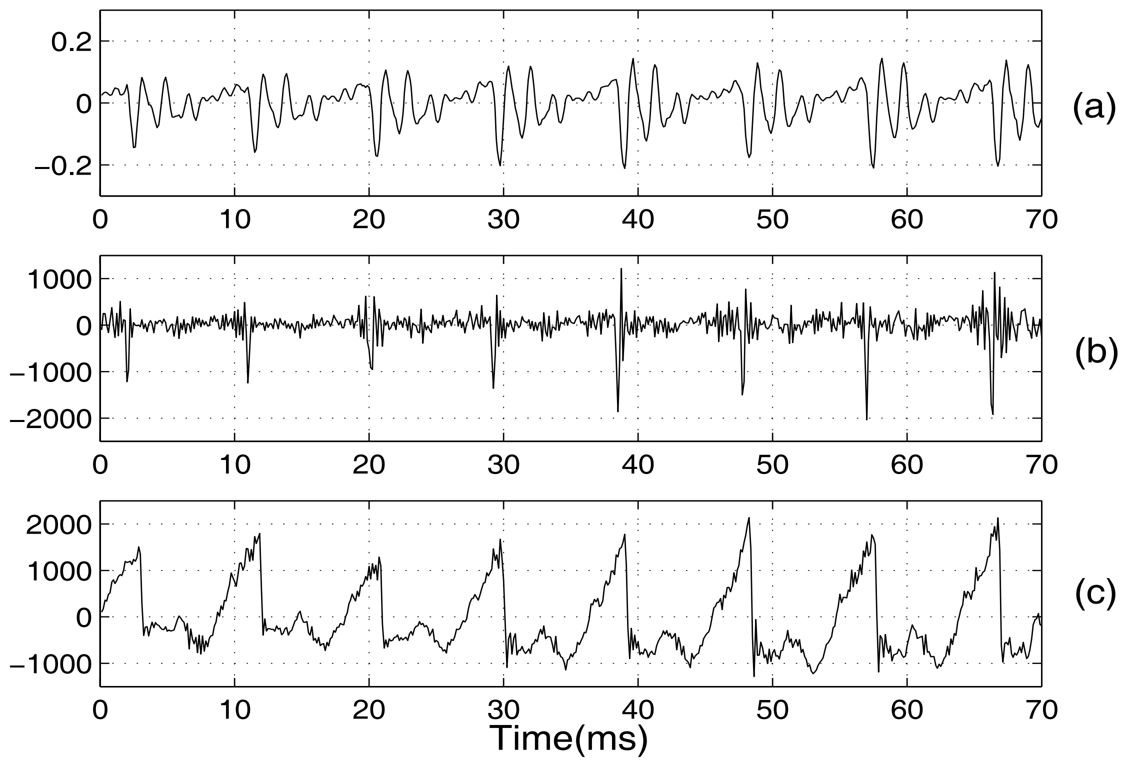

Figure 8: (a) Segment of voiced speech /a/, its (b) LP residual and (c) glottal pulse waveform.

7 LP spectrum for different LP orders

- Compute LPCs for different LP orders (14, 10, 6, 3 and 1), and compute LP spectrum for each set of LPCs.

- A segment of voiced speech and its LP spectrum for different LP orders (14, 10, 6, 3 and 1) are shown in Figure 9.

- **Observation:**
  LP order determines to some extent the accuracy with which speech production mechanism is modeled.
  It uses an all-pole model to characterize the vocal tract system by capturing the resonances with spectrum and source information with LP residual (inverse filter i.e., all zero filter).
  LP order determines the number of resonances that can be captured by the model. The maximum number of resonances captured by the model with LP order P is $P/2$.
  The length of the vocal tract from glottis to lips is approximately 17 cm. This can generate four to five prominent resonances in 0-4 KHz range. These resonances can be captured with the LP model of order 10. We also should take care of radiation and windowing effects. Therefore with LP order 10-14 we can model the system by capturing required resonances.
  System with LP order more than 14 will introduce the spurious resonances, which leads improper representation of the vocal tract system.

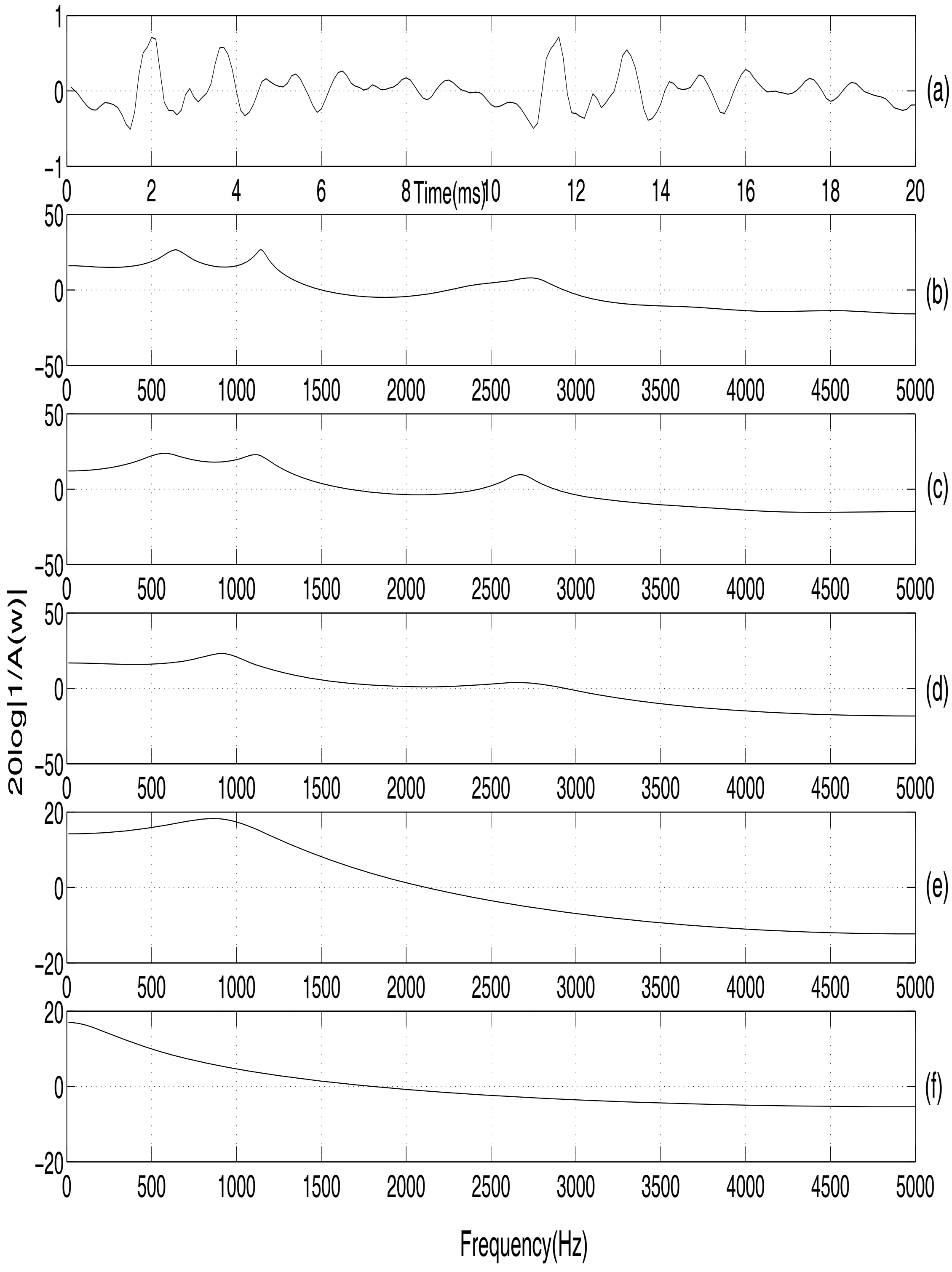

Figure 9: (a) Segment of voiced speech /a/, its LP spectrum for the LP order (b) 14, (c) 10, (d) 6, (e) 3 and (f) 1

8 Normalized error for different LP orders for voiced/unvoiced speech segments

- Normalized error is obtained by normalizing the LP residual energy with respect to speech signal energy.

- Normalized error is computed for both voiced and unvoiced segments of speech with different LP orders.
  $\eta = \frac{residualenergy}{signalenergy}$
  for i=1:15
  ak=lpc(a200,i);
  res=filter(ak,1,a200);
  $\eta(i) = \frac{sum(res.*res)}{sum(a200.*a200)}$;
  end
  figure;plot(η);grid;

- Normalized error for voiced and unvoiced segments of speech for different LP orders is shown in Figure 10

- **Observation:**
  Normalized error for voiced speech signal reduces as the LP order is varied from 0 to 15, since the vocal tract system (speech production mechanism) is modeled more accurately as LP order varying from 0 to 15.
  P=0, no approximation, therefore maximum error
  P=1, only one coefficient used for prediction, therefore error is slightly less compared to that of P=0
  P=10, model correctly approximates the resonances of the vocal tract system, which leads to minimum error
  P $>$ 10 also results the correct modeling of the vocal tract system, which leads to similar error as that of model with P=10
  For unvoiced speech signal, as the signal and residual energies remains reasonable same, change in error as function of LP order is relatively insensitive. Unvoiced speech signal itself appears like random noise, therefore the prediction will remains poor even though if we increase the LP order. Therefore both unvoiced speech signal and its LP residual appears like noise, hence the normalized error for unvoiced speech signal won't depend on the LP order.

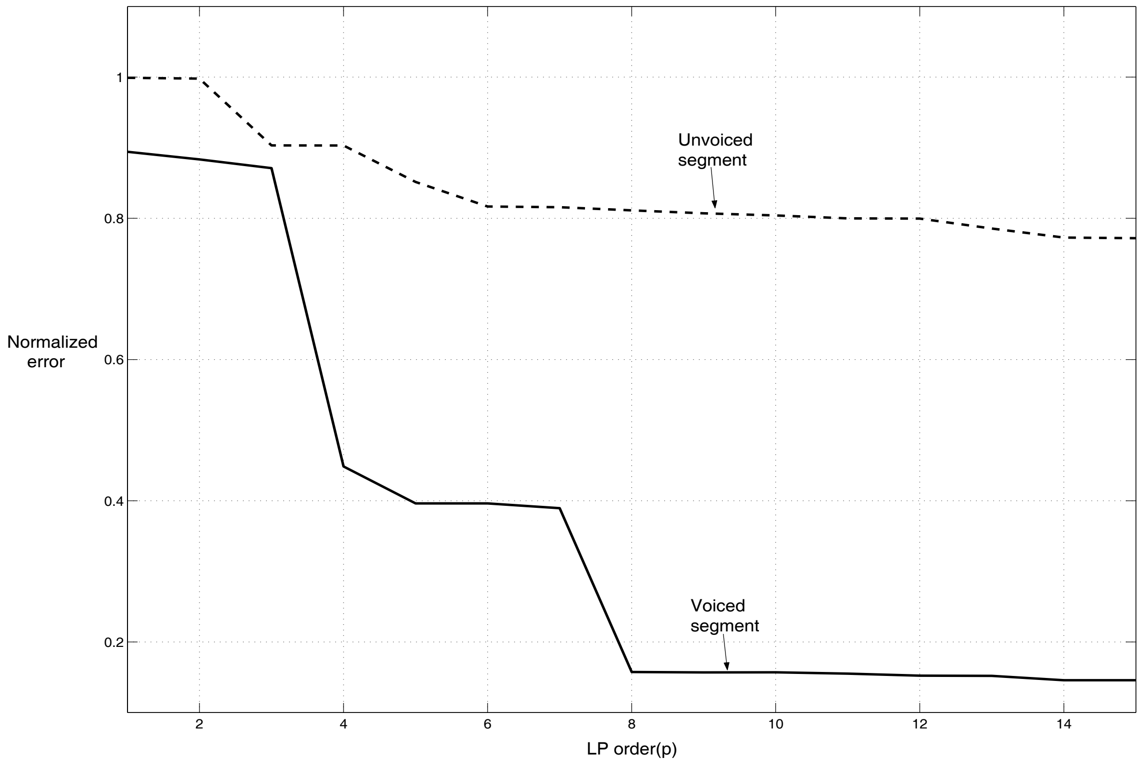

Figure 10: Normalized error for voiced and unvoiced speech segments for different LP orders.