CS630: Speech Technology

LAB-1: Speech Signal to Symbol Transformation

OBJECTIVE:

Manually performing speech signal to symbol transformation

SEQUENCE OF STEPS:

(a) Collect speech files for one of the Indian languages.

(b) Display a speech file using the utility "wavesurfer".

(c) Transcribe it in ITRANS code.

(d) Convert your transcription into syllable-like units.

(e) Label the speech file with syllable-like units using "wavesurfer".

(f) Repeat the same procedure for other speech files.

(g) Write a brief report on your observation.

1 Displaying the speech data

Display the given speech data file using the utility ”wavesurfer”.
wavesurfer samplefile.wav

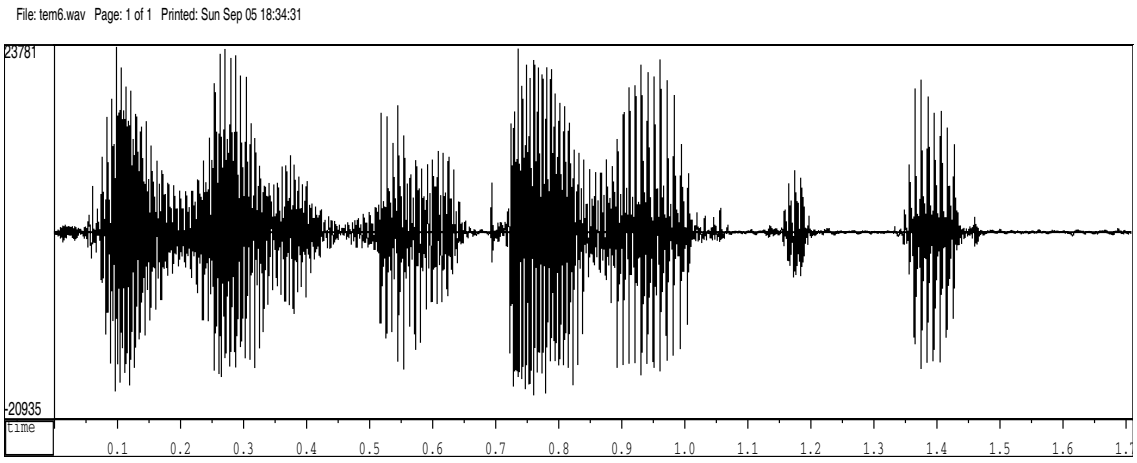

Figure 1: Speech waveform for the utterance ”Sri M. Venkaiahnayudu chepperu”

2 Transcription into known language

Play the speech file using ”wavesurfer”

By listening to the speech, write the utterance into the known (native) language.

3 Transcription into English using ITRANS code tables

Transcribe the utterance into English using ITRANS code tables.
ITRANS code: sri em ve.nkayyanayudu chepperu
(ITRANS code is the common transliteration code for Indian languages. Using these ITRANS code tables, any Indian language script can be transcribed into English. For each of the Indian language a separate ITRANS code table is used to transcribe the script from the given language into English)

4 Deriving the syllable-like units from English transcription

Split the utterance (text in ITRANS code) into syllable-like units. That is the ITRANS code text representation of the utterance is divided into subword units such as syllable-like units (text to symbol transformation, i.e., sentence to subword units)

Here syllable-like units are the symbols corresponds to the segments of the speech signal.

Reasons for choosing the syllable-like units as symbols against to phonemes (consonants or vowels (C or V)):

(1) It is very difficult to segment the speech data into phonemes.

(2) Due to coarticulation the effect of consonants is observed in the adjacent vowels.

(3) A character in an Indian language scripts is close to a syllable, and is typically one of the forms: V, CV, CCV, CVC, CCVC and CVCC, where C is a consonant and V is a vowel.

Utterance in ITRANS code: sri ve.nkayyanayudu chepperu
Syllable-like units of the utterance : sri em ve.n ka yya na yu du che ppe ru

5 Identifying the boundaries of the syllable-like units in speech waveform

Manually mark the boundaries of the syllable-like units by listening to the speech file segment by segment. The marked syllable boundaries and the associated waveform are shown in Figure 2. In the derived syllable boundaries some of the syllables present in the transcription are missing and some new syllables are marked. The expected syllable boundaries for a given speech file as per the transcription are shown in Figure 3.

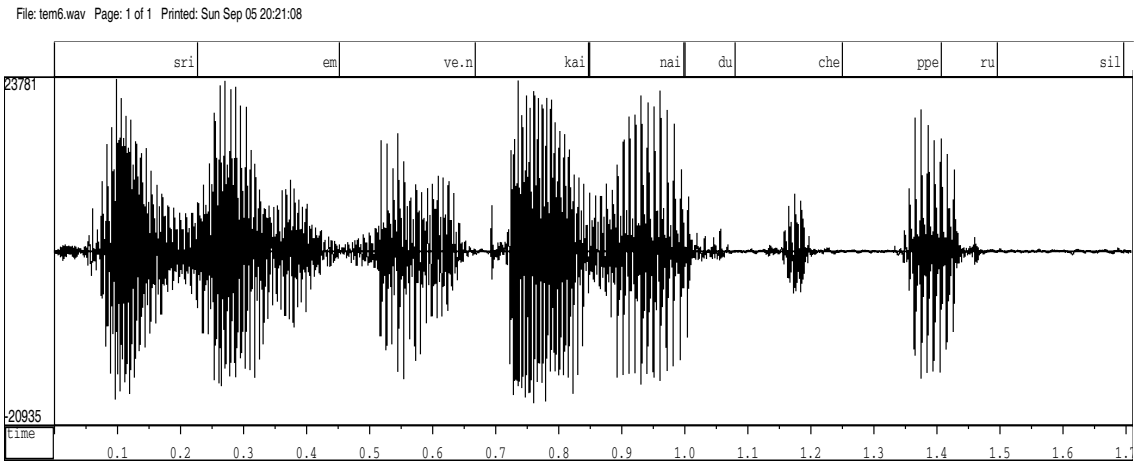

Figure 2: Speech waveform with the labels marked by listening to the segments of speech waveform

6 Observation

(1) While listening to the entire sentence, we can easily make out the message.

(2) Even the word boundaries can also be marked easily most of the times.

(3) While marking the boundaries of the syllables in a polysyllabic word a great difficulty is observed.

(4) Due to coarticulation effect, it is difficult to listen the particular syllable without the influence of the adjacent syllables in a polysyllabic word.

(5) In particular marking the boundaries of a syllable with semivowels found to be difficult.

(6) While listening to the segments of the speech data file, we are not able to perceive (completely missing) some syllables. But at the same time we can able to perceive a word clearly, where the syllables (missing) are present.

This observation shows the complexity involved in speech signal to symbol transformation. Even for human beings endowed with the natural speech

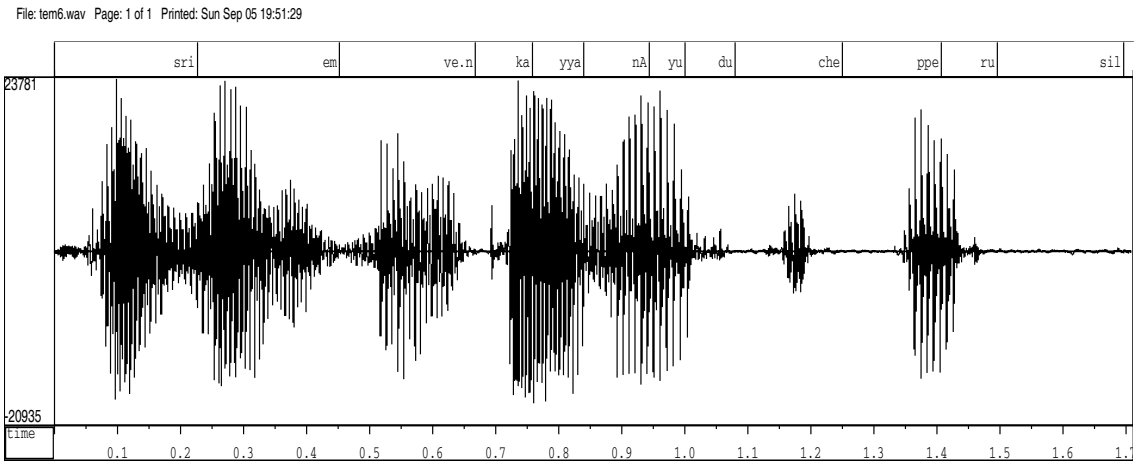

Figure 3: Speech waveform with expected labels

production and perception mechanism, found to be difficult to segment (label) the speech data into syllable-like units.

CS630: Speech Technology
LAB-2: Speech Production Mechanism

OBJECTIVE:

To express the characteristics of speech in terms of production characteristics.

SEQUENCE OF STEPS:

(a) Recording speech signal for a sentence and displaying waveform and spectrogram.

(b) Identifying/locating voiced/unvoiced/plosive/silence regions.

(c) Acoustic-phonetic description of the regions.

(d) Description of the time-varying excitation.

(e) Description of the time-varying system characteristics.

(f) Observing time-varying excitation and system characteristics from spectrogram.

(g) Writing a brief note on the observations.

1 Record and Play

Record a speech signal using the following linux command

brec -s 8000 -b 16 -t 2 -w samplefile.wav

"brec" is the linux command

"s" Sampling frequency

"b" Number of bits used to store the sample value

"t" Time interval used to record the speech signal

"w" format (wav) in which the speech signal to be represented

Display the speech waveform and its spectrogram using "wavesurfer".

Speech file can be played using a linux command

bplay -s 8000 -b 16 -w samplefile.wav

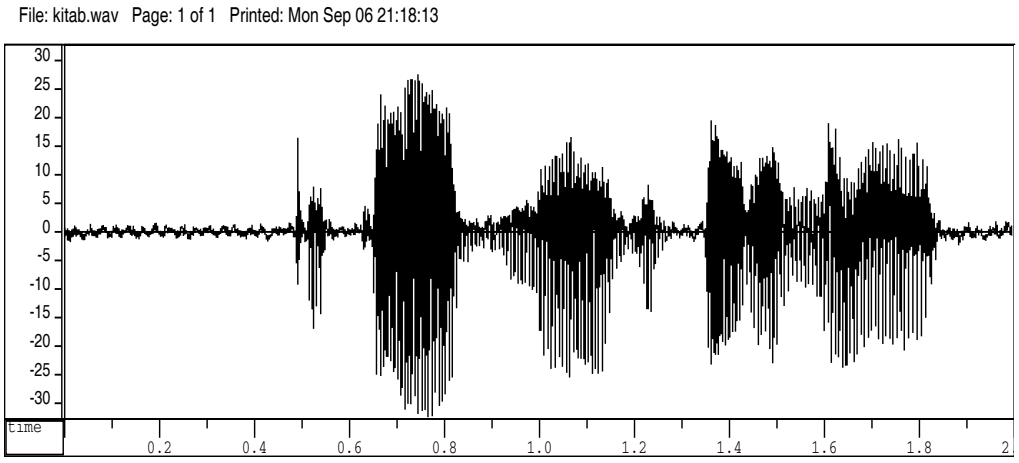

Figure 1: Speech waveform for the utterance "kitAb mEj par hai"

The spectrogram is used to represent the frequency components present in the speech signal. It is a three dimensional representation. X-axis represents the timing information, Y-axis shows the frequency components present in the speech signal and the darkness indicates the energy present in speech signal at that frequency. The dark bands in the spectrogram represents the resonances of a vocal tract system for the given sound unit. These resonances are also called as formant frequencies which represents the high energy por-

tions in the frequency spectrum of a speech signal. The shape of the dark bands indicates, how the vocal tract shape changes from one sound unit to the other.

2 Identifying the Voiced/Unvoiced/Plosive/Silence regions

From time domain waveform:

1. Voiced: quasiperiodicity and high amplitude regions
2. Unvoiced: nonperiodic and random like noise
3. Plosive: noise burst like signal indicates the sudden release of constriction at different places in vocal tract system
4. Silence: no speech signal (zero amplitude)

From spectrogram:

1. Voiced: In the case of vowels a regular formant structure (3 to 4 formant frequencies) and pitch harmonics (vertical striations in the case of wideband spectrogram) are used for identifying the voiced regions, where as nasals and voiced stops low frequency regions and pitch harmonics are used as clues.
2. Unvoiced: Energy at high frequency regions and no regular formant structure
3. Plosive: A silence bar followed by energy at high frequency regions.
4. Silence: no frequency components (white region)

3 Acoustic phonetic description

This description is based on the theory of acoustic phonetics. Acoustic phonetics deals with study of the physical properties of the speech sounds, as transmitted between mouth and ear. This description is provided using place

of articulation (POA) and manner of articulation (MOA). The following Tables 1 and 2 used to describe speech sounds (vowels and consonants (V and C)) using manner of articulation and place of articulation.

Example: kitAb mEj par hai

kitAb (/k/, /i/, /t/, /A/, /b/)

Unvoiced unaspirated velar stop followed by front vowel followed by unvoiced unaspirated dental stop followed by middle vowel followed by voiced unaspirated bilabial stop.

mEj (/m/, /E/, /j/)

Nasal followed by front vowel followed by voiced unaspirated stop.

par (/p/, /a/, /r/)

Unvoiced unaspirated bilabial stop followed by middle vowel followed by semivowel.

hai (/h/, /ai/)

Fricative followed by diphthong.

Table 1: Vowel classification

Vowel type	Sound units
Short vowels	a,i,u,e,o
Long vowels	A,I,U,E,O
Diphthongs	ai,au

4 Description of time varying excitation

The time varying excitation is described for the sound units present in the utterance using the Table 3.

Example: kitAb mEj par hai

/k/ : Release of velar constriction

/i/ : Vocal folds vibration

/t/ : Release of dental constriction

/A/ : Vocal folds vibration

/b/ : Vocal folds vibration + release of bilabial constriction

/m/ : Vocal folds vibration + lowered velum + closure of lips

Table 2: Consonant classification

Place of articulation	Manner of articulation				Nasals	Semi vowels	Fricatives
	Unvoiced		Voiced				
	Unaspirated	Aspirated	Unaspirated	Aspirated			
Velar	k	kh	g	gh	kn		h
Palatal	ch	chh	j	jh	chn	y	sh
Alveolar	T	Th	D	Dh	Tn	r	shh
Dental	t	th	d	dh	n	l	s
Bilabial	p	ph	b	bh	m	v	

/E/ : Vocal folds vibration
/j/ : Vocal folds vibration + release of alveolar constriction
/p/ : Release of bilabial constriction
/a/ : Vocal folds vibration
/r/ : Vocal folds vibration + turbulence at alveolar ridge
/h/ : Narrow constriction at velum
/ai/: Vocal folds vibration

5 Time varying system description

Description of time-varying system characteristics is nothing but specifying the positions of different articulators and shape of the vocal tract while producing the particular sound unit.

For production of vowels the time-varying system characteristics are described by the extent of opening of oral cavity, position of a tongue hump in a oral cavity and height of the tongue hump. The position and height of the tongue hump for different vowels can be described using the Figure 2.

The time varying system is described for the sound units present in the utterance using the Table 4.

Example: kitAb mEj par hai

/k/ : Complete closure at velum position

/i/ : Tongue hump is high and is in front position of the vocal tract (VT) system, VT system is narrowly open.

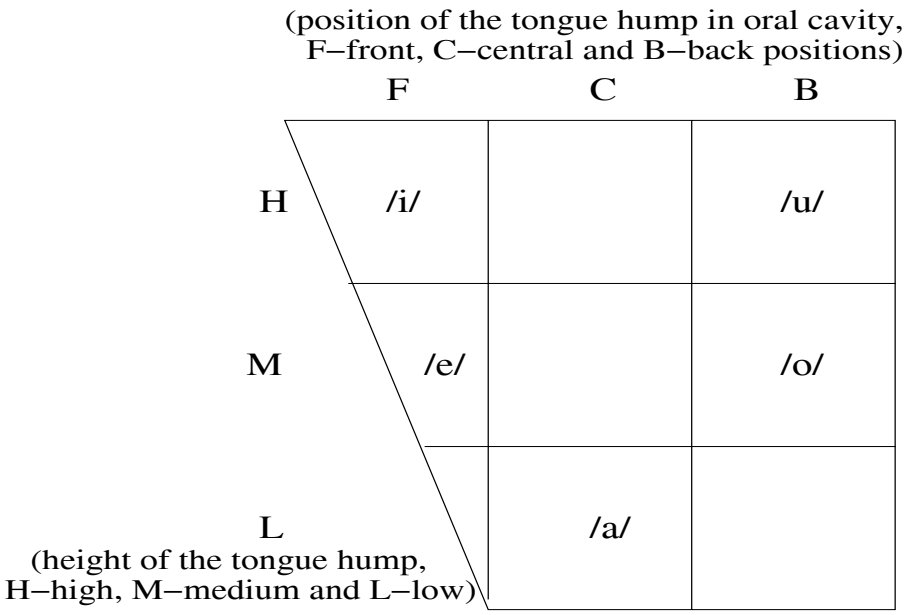

Figure 2: Position and height of the tongue hump for producing different vowels

/t/ : Complete closure at dental position

/A/ : Tongue hump is low and is in back position of the vocal tract (VT) system, VT system is widely open.

/b/ : Closure at lips

/m/ : Opening of velum and closure at lips

/E/ : Tongue hump is medium and is in front position of the VT system, VT system is moderately open.

/j/ : Narrow opening at alveolar ridge

/p/ : Closure at lips

/a/ : Tongue hump is low and is in central position of the VT system, VT system is widely open.

/r/ : Narrow opening at alveolar ridge

/h/ : Narrow opening at velum

/ai/: Tongue hump at alveolar ridge, narrow opening at alveolar ridge, VT system is narrowly open.

6 Observing the time varying system and excitation characteristics using spectrogram of the speech signal

So far we described the sound units interms of acoustic phonetics, time varying excitation and time varying system characteristics. Here we demonstrate the time varying excitation and system characteristics using the spectrogram of a speech signal. The speech waveform, its transcription and spectrogram are shown in Figure 3. Table 5 presents the spectral details of different sound units using spectrogram.

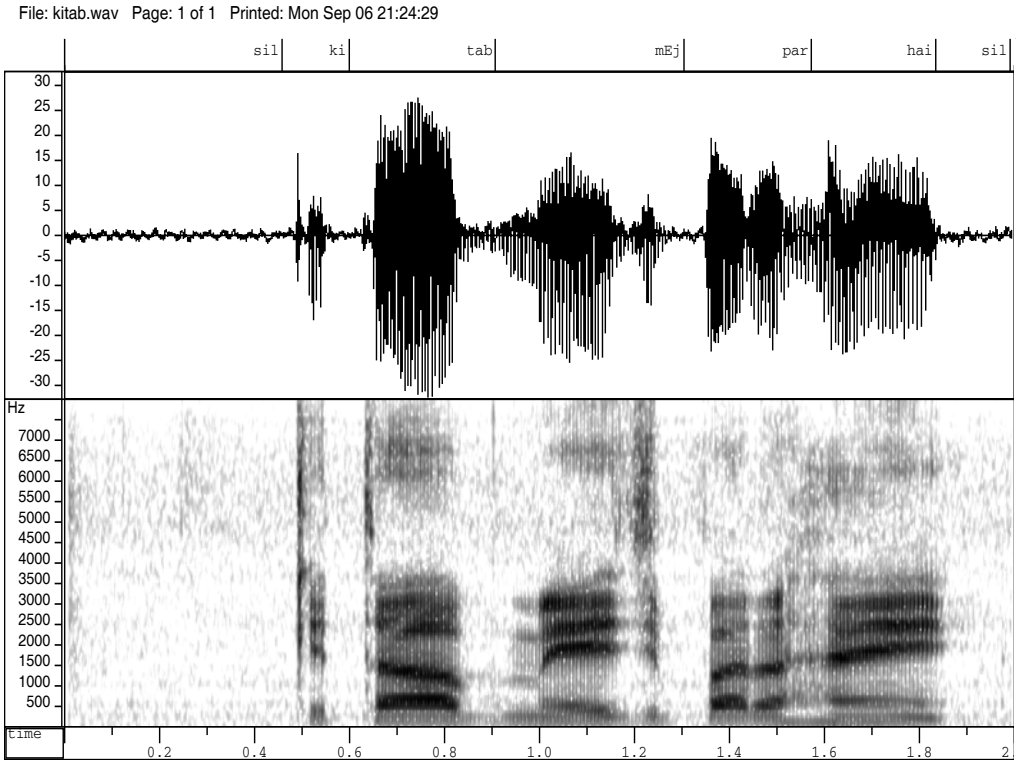

Figure 3: Speech waveform and its wideband spectrogram for the utterance "kitAb mEj par hai"

Time varying excitation characteristics from spectrogram :

/k/ : Silence bar before the burst is observed as no frequency components

/i/ : Vocal folds vibration can be observed interms of pitch harmonics (vertical striations) in spectrogram.

/t/ : Silence bar before the burst is observed as no frequency components

/A/ : Vocal folds vibration can be observed in terms of pitch harmonics (vertical striations) in spectrogram.

/b/ : Vocal folds vibration and closure at lips is observed as pitch harmonics at low frequency portion of the spectrogram.

/m/ : Vocal folds vibration, closure at lips and opening of velum is observed as pitch harmonics at low frequency portion of the spectrogram.

/E/ : Vocal folds vibration can be observed in terms of pitch harmonics (vertical striations) in spectrogram.

/j/ : Vocal folds vibration and closure at palatal is observed as pitch harmonics at high frequency portion of the spectrogram.

/p/ : Silence bar before the burst is observed as no frequency components

/r/ : Vocal folds vibration and narrow opening at alveolar ridge is observed as pitch harmonics at lower formants in spectrogram.

/h/ : no pitch harmonics are observed in spectrogram due random nature in time domain.

/ai/ : Vocal folds vibration can be observed in terms of pitch harmonics (vertical striations) in spectrogram.

Time varying system characteristics from spectrogram :

/k/ : Complete closure at velum and release of constriction at velar position are the system characteristics, these are observed in spectrogram as silence bar followed by energy at high frequency components (1700-4000 Hz).

/i/ : Tongue hump at front position of the vocal tract system and narrow opening of oral cavity is observed in spectrogram as regular formant structure and the formant frequencies observed to be f1=320 Hz,f2=1960 Hz and f3=2530 Hz.

/t/ : Complete closure at dental region and release of constriction at dental position are the system characteristics, these are observed in spectrogram as silence bar followed by energy at high frequency components. But the energy of the high frequency components has lower compared to /k/.

/A/ : Tongue hump at central position of the vocal tract system and wide opening of oral cavity is observed in spectrogram as regular formant structure and the formant frequencies observed to be f1=640 Hz,f2=1400 Hz and f3=2610 Hz.

/b/ : Closure at lips and radiation from cheeks and jaws is observed in spectrogram as low frequency components (less than 400 Hz).

/m/ : Opening of velum and closure at lips are the system characteristics and in spectrogram formant structure is observed, and the energy associated to the formants is observed to be much lower (around 25db less) compared to normal vowels. This formant structure may be due to the influence of its following vowel (/E/).

/E/ : Tongue hump at front position of the vocal tract system and moderate opening of oral cavity is observed in spectrogram as regular formant structure and the formant frequencies observed to be f1=520 Hz,f2=1880 Hz and f3=2440 Hz.

/j/ : In spectrogram a short discontinuity in formant structure is observed due to silence bar of the sound /j/. High frequency spectrum contains more energy over low frequency components. The influence of /E/ is observed in the spectrum of /j/ in the form of regular formants.

/p/ : Closure of lips is the system characteristics, in spectrogram no significant frequency spectrum is observed.

/a/ : Tongue hump at central position of the vocal tract system and wide opening of oral cavity is observed in spectrogram as regular formant structure and the formant frequencies observed to be f1=680 Hz,f2=1320 Hz and f3=2280 Hz.

/r/ : Narrow opening at alveolar ridge is the system characteristics, in spectrogram only two formants are observed f1=520 Hz and f2=1400 Hz. Intensity of these formants are less (about 30 db lower than the normal formants intensity).

/h/ : Narrow constriction at velum, in spectrogram due to coarticulation a thin traces of first two formants is observed. Some discontinuities are also observed in these two formants. The energy associated to these formants is very less. In general for the sound unit /h/ no significant frequency components are observed.

/ai/ : Tongue hump initially observed at central (due to /a/)and later at front position (due to /i/) of the vocal tract, oral cavity is initially wide opened and gradually reaches to narrow opening at the end of the sound unit. In spectrogram regular formant structure is observed. As the sound

unit is diphthong, a clear transition of the formant structure of vowel /a/ to vowel /i/ is observed. The formants at the initial region are found to be f1=680, f2=1720 and f3=2400, and at the final region f1=280, f2=2040 and f3=2520.

Table 3: Excitations and the corresponding sounds

Excitation type	Sound units
Vocal folds vibration	Vowels
Release of velar constriction	k,kh
Release of palatal constriction	ch,chh
Release of alveolar constriction	T,Th
Release of dental constriction	t,th
Release of bilabial constriction	p,ph
Release of velar constriction and vocal folds vibration	g,gh
Release of palatal constriction and vocal folds vibration	j,jh
Release of alveolar constriction and vocal folds vibration	D,Dh
Release of dental constriction and vocal folds vibration	d,dh
Release of bilabial constriction and vocal folds vibration	b,bh
Vocal folds vibration, velum is lowered and constriction at velum	kn
Vocal folds vibration, velum is lowered and constriction at palatal	chn
Vocal folds vibration, velum is lowered and constriction at alveolar	Tn
Vocal folds vibration, velum is lowered and constriction at dental	n
Vocal folds vibration, velum is lowered and constriction at lips	m
Vocal folds vibration and narrow constriction at palatal	y
Vocal folds vibration and narrow constriction at alveolar ridge	r
Vocal folds vibration and narrow constriction at dental	l
Vocal folds vibration and narrow constriction at lips	v
Narrow constriction at velum (turbulent)	h
Narrow constriction at palatal (turbulent)	sh
Narrow constriction at alveolar (turbulent)	shh
Narrow constriction at dental (turbulent)	s

Table 4: System characteristics and the corresponding sounds

Vocal tract system characteristics	Sound units
Tongue hump is low and it is in central position of the vocal tract (VT) system, VT system is widely open	a
Tongue hump is high and it is in front position of the VT system, VT system is narrowly open	i
Tongue hump is medium and it is in front position of the VT system, VT system is moderately open	e
Tongue hump is high and it is in back position of the VT system, VT system is narrowly open and cylindrical in shape	u
Tongue hump is medium and it is in back position of the VT system, VT system is moderately open and cylindrical in shape	o
Complete closure at velum	k,kh,g,gh
Complete closure at palatal	ch,chh,j,jh
Complete closure at alveolar	T,Th,D,Dh
Complete closure at dental	t,th,d,dh
Complete closure at lips	p,ph,b,bh
Complete closure at velum and opening of nasal cavity	kn
Complete closure at palatal and opening of nasal cavity	chn
Complete closure at alveolar opening of nasal cavity	Tn
Complete closure at dental opening of nasal cavity	n
Complete closure at lips opening of nasal cavity	m
Narrow constriction at velum	h
Narrow constriction at palatal	sh
Narrow constriction at alveolar	shh
Narrow constriction at dental	s
Partial closure of VT with tongue hump at palatal	y
Partial closure of VT with tongue tip at alveolar ridge	r
Partial closure of VT with tongue tip at dental	l
Partial closure of VT with lower lip and upper teeth	v

Table 5: Spectral details for different sound units

Sound unit	Spectrogram details
a	Regular formant structure (730,1090,2440), pitch harmonics
A	Regular formant structure (520,1190,2390), pitch harmonics
i	Regular formant structure (270,2290,3010), pitch harmonics
I	Regular formant structure (390,1990,2550), pitch harmonics
u	Regular formant structure (300,870,2240), pitch harmonics
U	Regular formant structure (440,1020,2240), pitch harmonics
e	Regular formant structure (530,1840,2480), pitch harmonics
E	Regular formant structure (660,1720,2410), pitch harmonics
o	Regular formant structure (570,840,2410), pitch harmonics
m,n (nasals)	Concentration of low frequency energy and midrange frequencies with no prominent peaks, pitch harmonics
s,sh	Concentration of high frequency energy
k,ch,T,t,p	Concentration of high frequency energy
g,j,D,d,b	Concentration of low frequency energy

CS630: Speech Technology
LAB-3: Nature of Speech Signal

OBJECTIVE:

To obtain the pitch and formant frequencies from the nature of the speech signal in the time and frequency domains respectively.

Tasks

(a) Record speech for three vowels in the context of h*ee*d (for vowel /i/), h*e*ad (for vowel /e/) and h*o*d (for vowel /a/).

(b) Note the average pitch using time domain representation.

(c) Note average formant frequencies (F_1, F_2 and F_3) from 10^{th} order LPC spectrum.

(d) Repeat for one more set of vowels.

(e) Compare the readings from the two sets (variability within a given speaker).

(f) Compare the average pitch and formant frequency values of one speaker with values from other speakers (variability among speakers).

(g) Write a brief note on the observations.

1 PROCEDURE

1. **Recording Speech Signal**

   - Record a speech signal for heed, head, hod using 'brec' command as
     brec -s 8000 -b 16 -t 2 -w samplefile.wav
     where "brec" is the linux command
     "s" is the sampling frequency in Hz
     "b" is the number of bits/sample used for quantization
     "t" is the time interval used to record the speech signal
     "w" format (wav) in which the speech signal to be represented

   - Play the signal back using 'bplay' command
     bplay -s 8000 -b 16 -w samplefile.wav

2. To obtain Average Pitch

   - Display the recorded speech waveform using the utility "wavesurfer".

     wavesurfer filename.wav

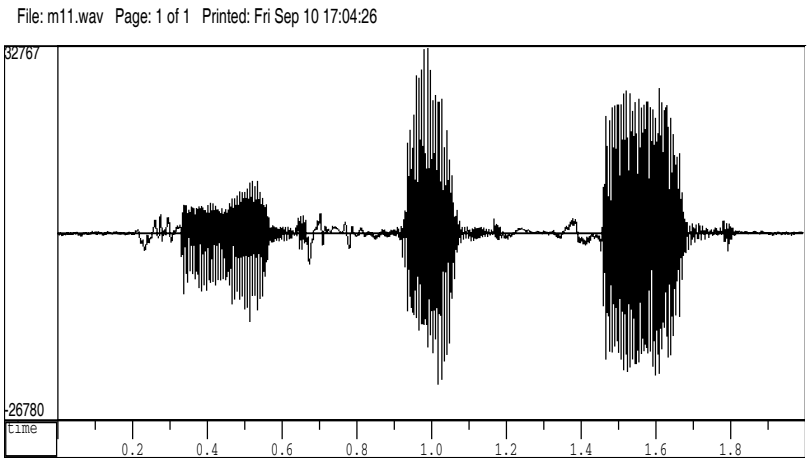

Figure 1: Speech waveform for the utterance "heed- head - hod"

- Zoom the vowel segment using 'zoom in' command from menu options 'view'. Asegment of the vowel /i/ is shown in the Figure 2.

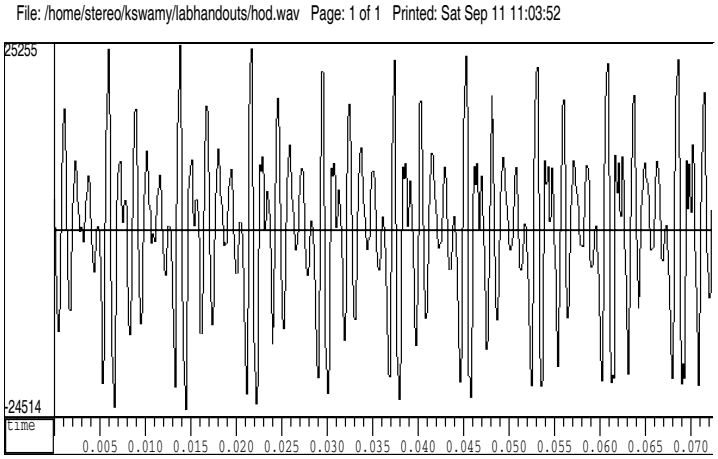

Figure 2: A part of the vowel /a/ in the speech segment of 'hod' obtained using 'zoom in' command in menu options

- Note the time in msec of the period for three or four cycles (say) and measure the average pitch period, preferably in the steady portion as illustrated in the Figure 3.
- As pitch period is not constant over the entire vowel segment, make four or five pitch period measurements.
- Mean of these values gives average pitch.

3. To obtain Average Formant Frequencies

   - Display voiced segment of the recorded speech segment /hod/ using wavesurfer.
   - Select the spectrum section in the menu popped up by right click of the mouse. Choose LPC section and set order to 10 and pre-emphasis to 1.0.This will display LPC spectrum as shown in the Figure 4. The peaks in the spectrum represents formant frequencies and are noted down in the Table **??**.

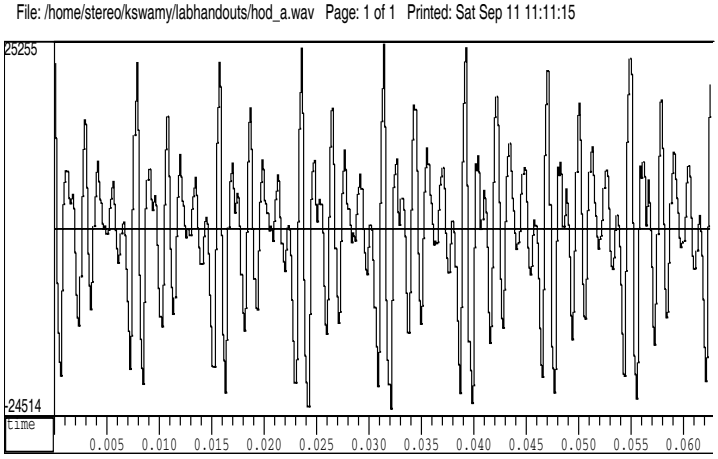

Figure 3: A part of the vowel /a/ in the speech segment of 'hod'. As shown in the Figure 3, pitch is measured as the interval between any two peaks

- The mean of each formants is calculated to obtain corresponding average values.
- Repeat the previous step for two or three short segments of the voiced region.
- Repeat the above step for recorded speech of /heed/ and /head/.
- Repeat the above procedure for recorded speech of different speakers.

4. OBSERVATIONS:

5. Conclusions

- The pitch for the same speaker varies with respect to time as well as varies across speakers. This clearly illustrates the variability of speech signal.
- For given sound-unit, the formant frequency values varies slightly for the same speaker and also across speakers. This again illustrates the variability of the speech in frequency domain for the same speaker.

Table 1:

Speaker	Vowel	Formants (Hz)						Pitch(msec)
		F1		F2		F3		
/i/								
/e/								
/a/								

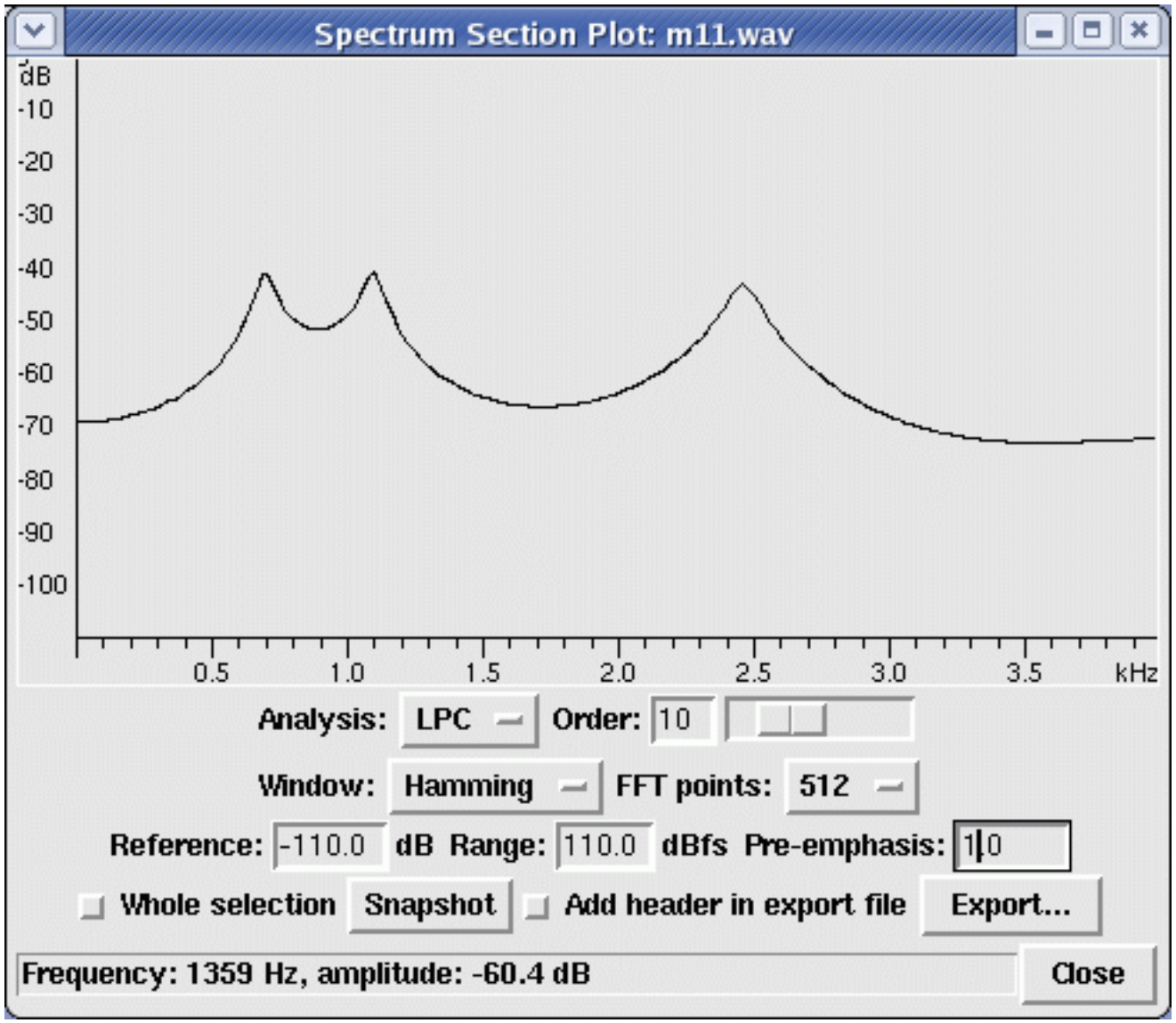

Figure 4: LPC spectrum of a short segment of vowel $/a/$ in the utterance of 'hod'. The spectrum is obtained using 10^{th} order LP analysis. As shown in the figure, the peaks of the LP spectrum represent the formant frequencies.

CS630: Speech Technology
LAB-4: Basics of DSP

OBJECTIVES:

To study quantization and aliasing effects.
To synthesize vowels /a/, /i/ and /e/

Tasks

(a) Record a short utterance of speech for 2 to 3 sec (8 kHz, 16 bits).

(b) Listen to speech at different bit rates (16 bits/sample, 8 bits/sample and 1 bit/sample).

(c) Listen to speech at different sampling rates (8 kHz, 4 kHz and 2 kHz).

(d) Synthesize vowels /i/, /e/ and /a/ for the actual pitch period, half the pitch period and twice the pitch period.

(e) Write a brief note on the observations.

1 PROCEDURE:

1. **Recording Speech Signal**

   - Record a short sentence of speech using 'brec' command brec -s 8000 -b 16 -t 2 -w samplefile.wav
     brec -s 8000 -b 16 -t 2 -w samplefile.wav
     where "brec" is the linux command
     "s" is the Sampling frequency in Hz
     "b" is the number of bits/sample used for quantization
     "t" is the time interval used to record the speech signal
     "w" format (wav) in which the speech signal to be represented

2. **Study of Quantization Effect**

   - The short segment of speech recorded is at a sampling frequency of $8kHz$ and resolution of 16 bits/sample.
   - Load this segment of speech into MATLAB workspace using
     $- >>$ p = wavread(filename1.wav);
     where wavread is MATLAB function which reads contents of a file in wave format into a array.
   - Change resolution from 16 bits/sample to 8 bits/sample
     $- >>$ wavwrite(p,8000,8,filename2.wav);
     $- >>$ $p1$ = wavread(filename2.wav);
   - Change resolution from 16 bits/sample to 1bit/sample $- >>$ $p2 = [p1 > 0]$;
   - Now p, p1, p2 respectively contains same speech at different resolutions 16 bits/sample, 8bits/sample and 1 bit/sample, but all at the same sampling frequency 8 kHz.
   - Play these sequences and note the observations:
     $- >>$ sound(p,8000); $- >>$ sound(p1,8000); $- >>$ sound(p2,8000);

3. **Study of Aliasing Effect**

- The short segment of speech recorded is at a sampling frequency of 8 kHz and 16 bits/sample resolution.
- Load this segment into MATLAB workspace using
  $- >> q1 = wavread(filename1.wav);$
- Choose every alternate sample from q1
  $- >> i = [1 : 2 : length(q1)];$
  $- >> q2 = q1(i);$
- Choose every alternate sample from q2
  $- >> j = [2 : 2 : length(q2)];$
  $- >> q3 = q2(j);$
- Now q1, q2 and q3 respectively contains $8000, 4000, 2000$ samples/sec. These are equivalent to sampling the speech respectively at 8 kHz, 4 kHz and 2 kHz (Low pass filtering is not done for studying the aliasing effect).
- Play these sequences and note the observations:
  $- >>$ sound(p,8000);
  $- >>$ sound(p1,8000);
  $- >>$ sound(p2,8000);

4. **Synthesizing vowels** $/a/, /i/and/e/$
   - Given formant frequencies F_1, F_2 and F_3 of the vowel, pitch period T_0 and sampling frequency F_s at which a vowel is to be synthesized.
   - The system used in the production of vowels is an all pole filter. Compute the parameters of the filter using formant frequencies and sampling frequency. The parameters are the coefficients of the numerator and denominator polynomials (b_k s anda_k s) of the system transfer function.
   - The source used for excitation is a periodic unit sample sequence.Generate unit sample sequence of required length, depending on the length of the speech segment to be generated.

- Obtain response of the system with excitation generated using
  $- >>$ y = filter(b, a, x) ;
  $- >>$ sound(y);

Illustration of Synthesizing Vowel $/a/$:

- Given,
  -Formant frequencies F_1 = 560 Hz, F_2= 1180 Hz, and F_3= 2480 Hz
  -Pitch period, $T_0 = 7.5$ msec
  -Sampling frequency, $F_s = 10000Hz$

- The system $H(z)$, shown in Figure 1 is an all-pole filter. When input excitation $x(n)$ is unit sample sequence, then its response $y(n)$ will be damped sinusoids as shown in Figure 2.

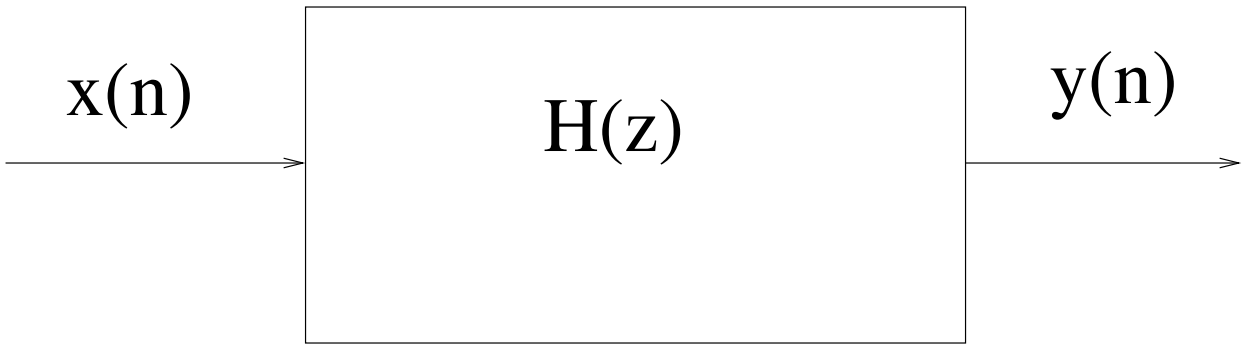

Figure 1: All-pole filter.

- The Transfer function $H(z)$ is given by,

$$H(z) = \frac{\sum_{k=0}^{M-1} b_k.z^{-k}}{\sum_{k=0}^{N-1} 1 + a_k.z^{-k}} = \frac{1}{1 + \sum_{k=1}^{6} 1 + [a_k z^{-k}]} \qquad (1)$$

- For a given formant frequency F_i, let its bandwidth be $B_i = 0.1 * F_i$
- For each (F_i, B_i), system parameters are computed using

$$H_i(z) = \frac{1}{1 - 2.exp(-pi.B_i.T).cos(2.pi.F_i.T).z^{-1} + exp(-2.pi.B_i.T).z^{-2}} \quad (2)$$

where T is sampling rate = $\frac{1}{Fs}$ = 100*10^{-6} sec

- Substituting $F_1 = 560$ Hz, $B_1 = 56$Hz in Equation 1 and simplifying we get

$$H_1(z) = \frac{1}{1 - 1.8447 * z^{-1} + 0.9654 * z^{-2}} \quad (3)$$

- Substituting $F_2 = 1180$ Hz, $B_2 = 118$ Hz in equation 1 and simplifying we get

$$H_2(z) = \frac{1}{1 - 1.4213 * z^{-1} + 0.9285 * z^{-2}} \quad (4)$$

- substituting $F_3 = 2480$ Hz, $B_3 = 248$ Hz in Equation 1 and simplifying we get

$$H_3(z) = \frac{1}{1 - 0.0232 * z^{-1} + 0.8557 * z^{-2}} \quad (5)$$

- The system transfer function $H(z)$ is given by, $H(z) = H_1(z).H_2(z).H_3(z)$
- Substituting for $H_1(z), H_2(z), H_3(z)$ and simplifying we get

$$H(z) = \frac{1}{1 - a_1 z^{-1} + a_2 z^{-2} - a_3 z^{-3} + a_4 z^{-4} - a_5 z^{-5} + a_6 z^{-6}} \quad (6)$$

where a = $[1, a_1, a_2, a_3, a_4, a_5, a_6]$ is shown below:

- From Equation 6 we have b= [1];
  a= $[1, -3.2892, 5.4471, 4.8318, -2.6604, 0.7669]$;

- We can also compute values of a_k s from $H_1(z), H_2(z) and H_3(z)$ by convolution
  $- >>$ t = conv([1,-1.8447,0.9654], [1, -1.4213, 0.9285]);
  $- >>$ a = conv(t,[1, -0.0232, 0.8557]);

- The values obtained by convolution are ,
  $a = [1, -3.2892, 5.472, -5.9844, 4.8321, -2.6606, 0.7670]$;
  This gives the system parameters.

- Generate the unit sample sequence to be used as the excitation at the actual pitch period (7.5 msec i.e., around 60samples at 8 kHz) as
  − >>x1 = zeros(1, 4000);
  − >>for i =1:60:length(x1)
  x1(i) = 1;
  end;

- Similarly, excitation sequences at half the pitch period and twice the pitch period are generated as follows :
  − >>x2 = zeros(1, 4000);
  − >>for i =1:30:length(x1)
  x1(i) = 1;
  end;
  − >>x3 = zeros(1, 4000);
  − >>for i =1:120:length(x1)
  x1(i) = 1;
  end;

- Now using the excitation sequence, the output sequence is generated using − >>y1 = filter(b,a,x1);
  Figure 3 shows the excitation sequence and the signal synthesized for the half pitch period case. − >>y2 = filter(b,a,x2);
  Figure 4 shows the excitation sequence and the signal synthesized for the pitch period case.
  − >>y3 = filter(b,a,x3);

  Figure 5 shows the excitation sequence and the signal synthesized for twice the pitch period case.

- Play the output sequence using sound command − >>sound(y1);
  − >>sound(y2);
  − >>sound(y3);
  Figure 6,7,8 shows the synthesized signal for the sound unit /a/ and the corresponding spectrograms for excitation with a period of half the pitch, pitch, twice the pitch respectively.

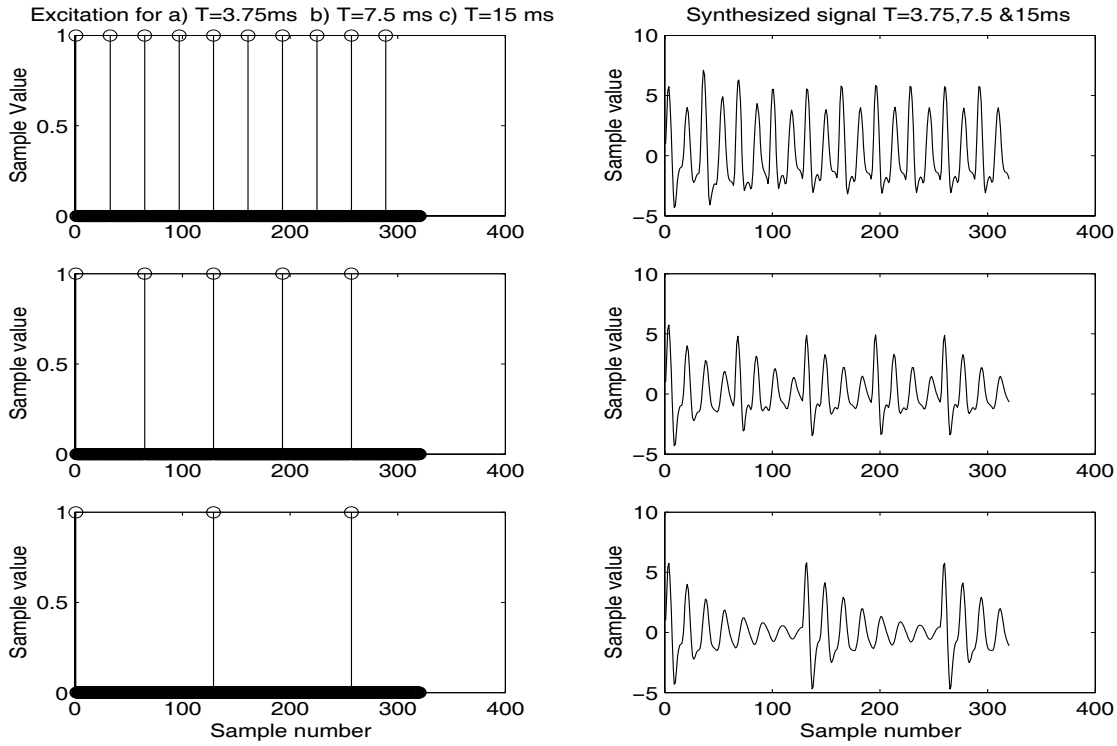

Figure 2: The excitation and the response of the all-pole filter for impulse sequence excitation with a period of 3.75ms, 7.5ms and 15ms respectively.

2 CONCLUSIONS

- The quality of speech is directly proportional to the number of quantization levels (that is, number of bits used for quantization).Though quantization results in loss of information the human perception mechanism can still get the information present in the speech signal.
- It is interesting to note that even with 1 bit/sample, most of the speech is intelligible. Hence we can conclude that information lies in the sequence and not in the set of numbers.
- The aliasing effect is perceived as distortions in the output signal. This is due to the effect of overlapping of frequency components.

- The synthesized vowel sounds as that of vowel except at the cost of naturalness. The formants are clearly defined and there is no variability in the signal as it is a synthetic siganl.

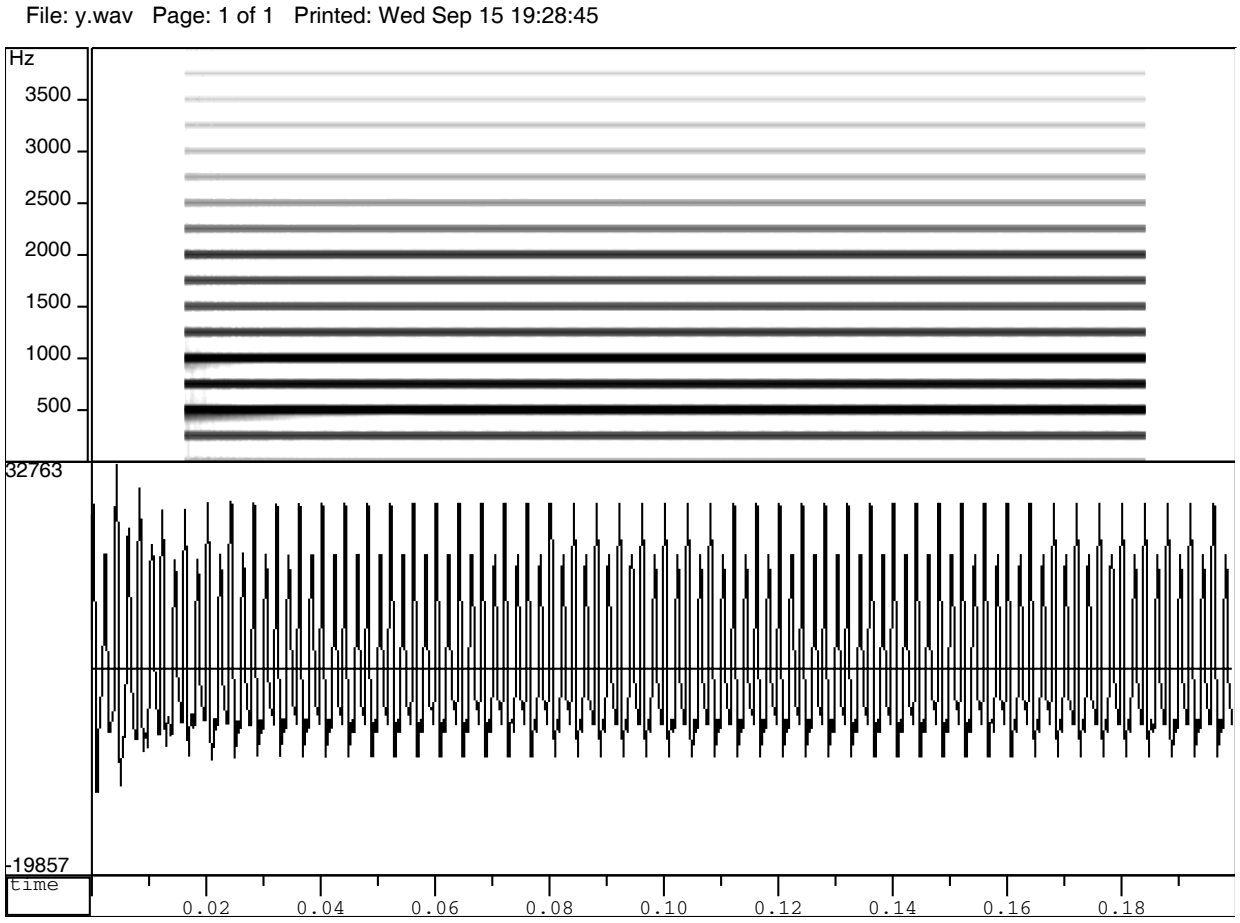

Figure 3: Waveform of the the synthesized vowel /a/ and its spectrogram for excitation at half the pitch period.

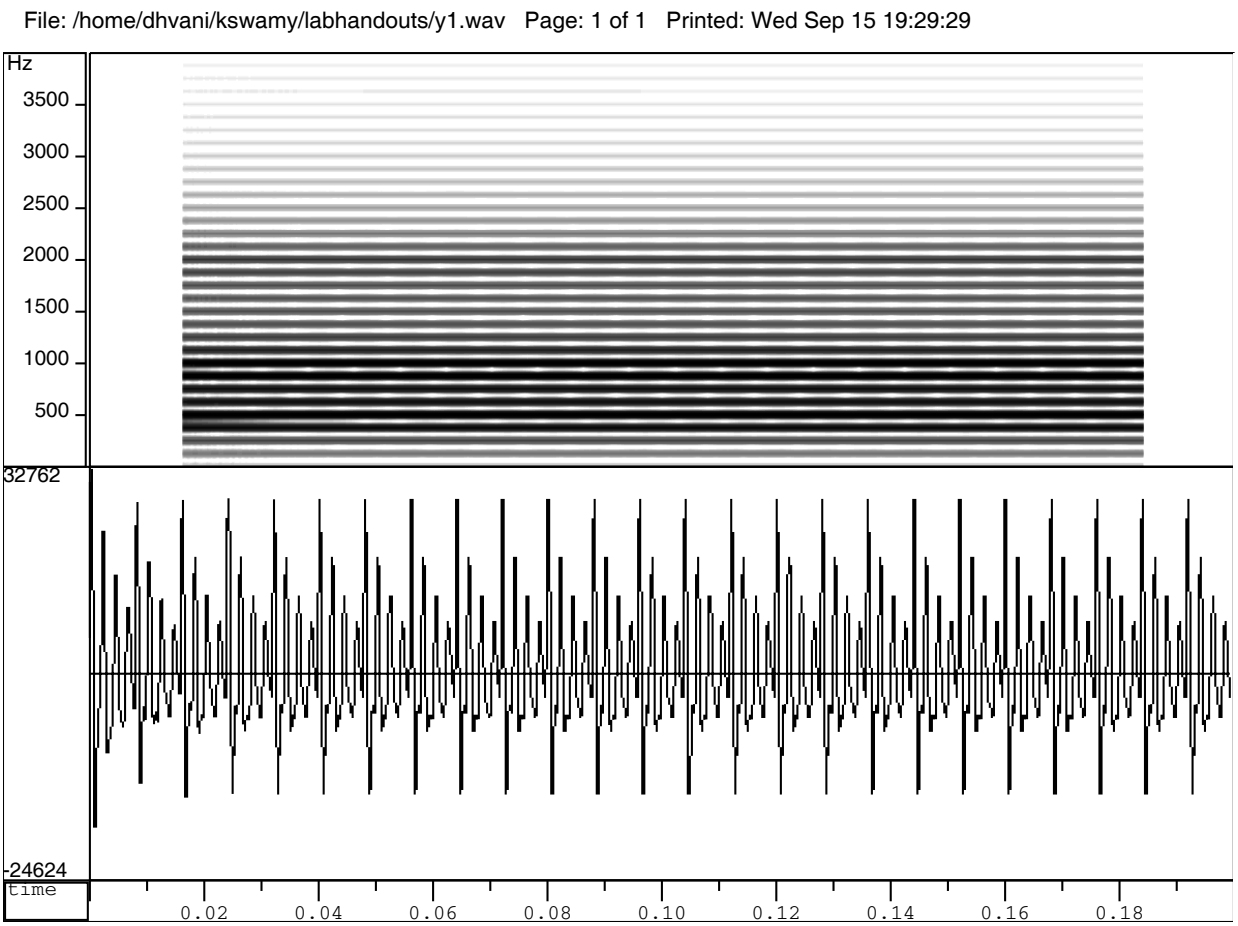

Figure 4: Waveform of the the synthesized vowel /a/ and its spectrogram for excitation at the pitch period.

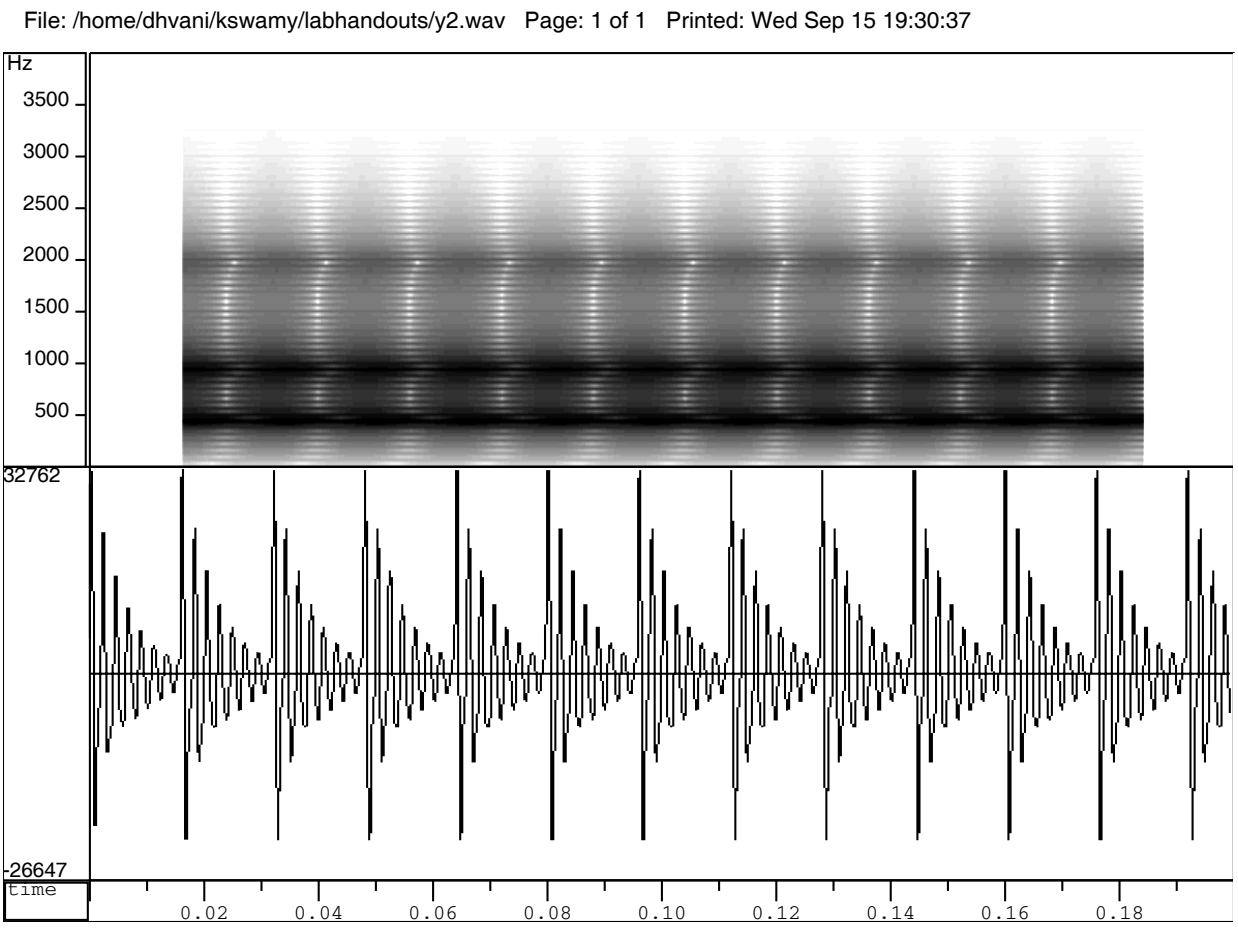

Figure 5: Waveform of the the synthesized vowel /a/ and its spectrogram for excitation at twice the pitch period.

CS630: Speech Technology
LAB-5: Short-Time Spectrum Analysis

OBJECTIVE:

To study issues in short-time spectrum analysis of speech.

SEQUENCE OF STEPS:

(a) Record (about 2 sec) of vowel /a/ and fricative /s/.

(b) Study the effects of convolution and correlation.

Consider $x(n)$, 128 samples of /a/ and $h(n)$, 5 samples of unit values Perform

- Linear convolution of $x(n)$ and $h(n)$
- Circular convolution of $x(n)$ and $h(n)$ with (N=128)
- Linear correlation of $x(n)$ and h(n)
- Linear autocorrelation of $x(n)$
- Linear autocorrelation of $h(n)$

(c) Short-time spectrum - Effects of size and shape of window.

- Consider $x(n)$, 160 samples of /a/.
- Use 512 point DFT to get $X(k)$.
- Plot log spectrum log—X(k)—^{2}.
- Study the effect of size of window - 5 msec, 20 msec, 50 msec.
- Study the effect of shape of window on 20 msec data namely, Rectangular, Hamming and Hanning windows.

(d) Short-time spectrum of voiced and unvoiced speech.

- 20 msec of $x(n)$, Hamming window and 512 pt DFT
- Plot log spectrum for voiced and unvoiced segments

(e) Spectrograms.

- Record a short utterance
- Observe features of WB and NB spectrograms

(f) Write a brief note on the observations.

1 Procedure

(1) **Recording of required speech utterances**

- Record the sound units vowel /a/ and fricative /s/ using command
  brec -s 8000 -b 16 -t 2 -w filename1.wav
  where s is the sampling rate, b is the number of bits/sample,
  t is the duration of speech utterance and w is the fileformat.

- Take a segment of duration 200 msec of vowel /a/ and a segment of duration 200 msec of fricative /s/.
  Using Matlab, read the speech signal in $filename1.wav$ into an array.
  a=wavread('filename1.wav');
  plot(a);
  a200=a(10501:10660);
  s=wavread('filename2.wav');
  plot(s);
  s200=s(6501:6660);

(2) **Effect of Convolution and Correlation**
The convolution of two sequences x(n) and h(n) is defined as

$$y(n) = \sum_k x(k)h(n-k) \tag{1}$$

The convolution sum is used in filtering a signal where h(n) is a filter. The filter suppresses some frequency components of x(n).
Consider x(n) as 128 samples of /a/ given by
a128 = a(10501:10628); and h(n) as 5 samples of unit values given by
h5 = [1; 1; 1; 1; 1];

- **Linear convolution of x(n) and h(n)**
  The convolution of two vectors can be obtained by using the function conv as

linconv = conv(a128,h5);
plot(linconv);
The resulting vector is of length equal to[length(x(n))+length(h(n))-1] and is shown in Figure 1.

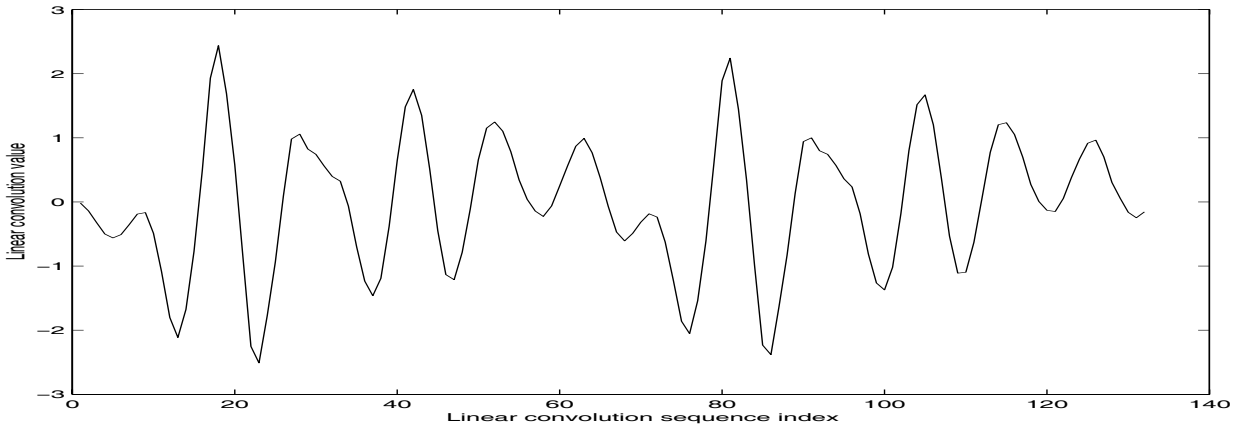

Figure 1: The linear convolution of x(n) and h(n).

- **Circular convolution of x(n) and h(n) by considering N=128**
  The circular convolution of two vectors are done by finding, the FFT of two vectors x(n) and h(n),
  followed by multiplying the two resulting vectors and finally finding the IFFT as given below.
  fftx=fft(a128,128);
  ffth=fft(h5,128);
  fftprod=fftx.*ffth;
  circonv=real(ifft(fftprod));
  plot(circonv);
  The result of circular convolution is shown in Figure 2.

  Observation: It is evident from figures 1 and 2 that the length of resulting vector in linear convoltion is 132 (length(x(n)) + length(h(n)) -1), where as the length of resulting vector in circular convolution is only 128 (max(length(x(n)),length(h(n)) if N is not specified). Circular convolution of x(n) and h(n) is the aliased version of the linear convolution if the sequence lengths are not properly extended.

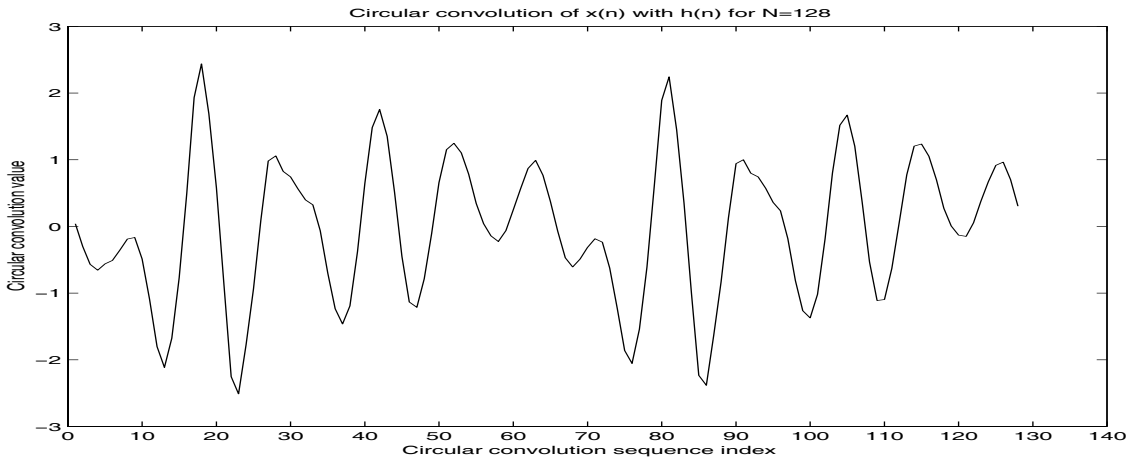

Figure 2: The circular convolution of x(n) and h(n).

- **Linear correlation of x(n) and h(n)** The correlation basically finds the relationship among two sequences. The correlation of two sequences x(n) and h(n) is given by
  $r(n) = \sum_k x(k)h(n+k)$
  If there exists no relationship among the sequences then r(0), r(1), r(2),....., are all will be zero. The correlation function is used for cross-correlation function estimation. The linear autocorrelatrion of x(n) and h(n), where x(n) is a vector of length 128 and h(n) is a vector of length 5 is obtained by using **'xcorr'** matlab builtin function. This function returns the length 255 cross-correlation sequence. Since h(n) is of length 5, it is zero padded to make its length equal to the length of x(n) which is 128. Thus in general, if largest length of either vectors is M then autocorrelation function returns a cross-correlation sequence of length $2 * M - 1$. The linear autocorrelation of x(n) and h(n) is computed as follows:
  lcrxnhn=xcorr(a128,h5);
  plot(lcrxnhn);
  The result of linear correlation is shown in Figure 3.

- **Linear autocorrelation of x(n)**
  The autocorrelation of x(n) is computed as
  lautocorrxn=xcorr(a128);

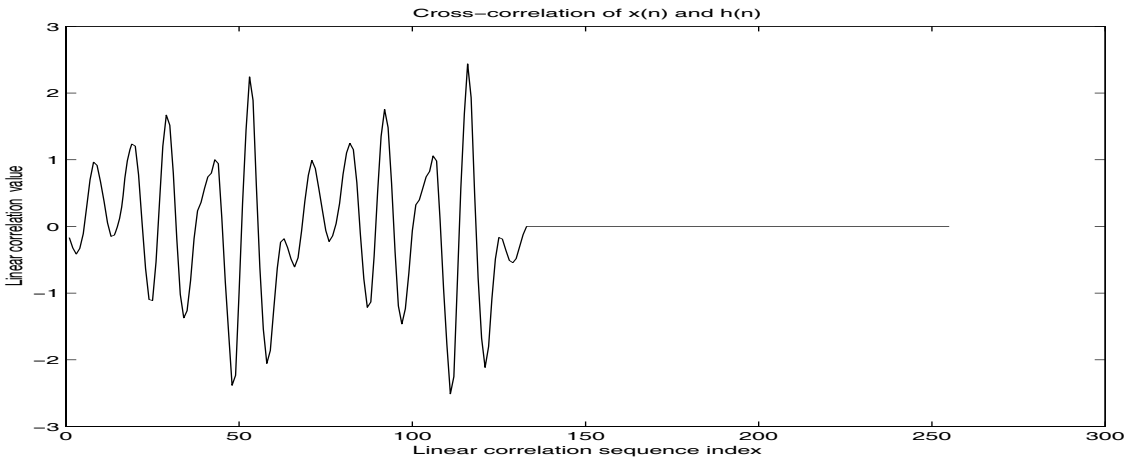

Figure 3: The cross-correlation of x(n) and h(n).

plot(lautocorrxn);
The result of linear autocorrelation of x(n) is shown in figure 4.
The autocorrelation of x(n) at n=0 gives energy of the signal.

- **Linear autocorrelation of h(n)**
  The autocorrelation of h(n) is computed using
  lautocorrhn=xcorr(h5);
  plot(lautocorrhn); The result of linear autocorrelation of x(n) is shown in figure 4.

(3) **Study the effect of size and shape of the window**

- **Effect of size of window a) 5 msec b) 20 msec c) 50 msec** -
  The effect of rectangular window of size 5 msec(assuming 512-point DFT) is illustrated by computing the short-time spectrum of the windowed signal as follows:
  a40 = a(10501:10540);
  X(k) = fft(a40,512);
  plot(10∗log10(abs(X(k).∗X(k))));
  The resulting log spectrum is shown in Figure 6. Since the spectrum of real valued signal is even symmetric, spectrum is plotted for only

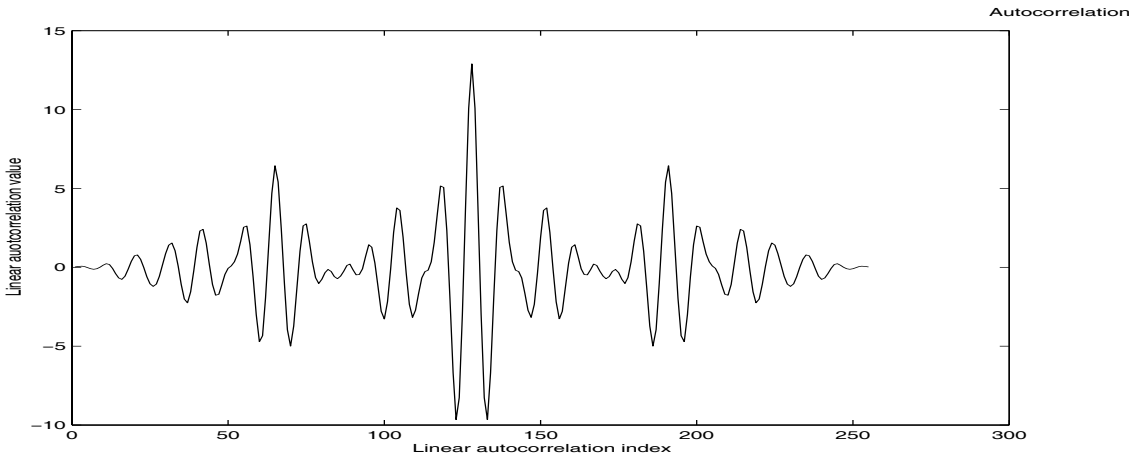

Figure 4: The autocorrelation of x(n).

half of the index values(k). This convention is followed in plotting the subsequent log spectrums. - The effect of rectangular window of size 20 msec(assuming 512-point DFT) is illustrated by computing the short-time spectrum of the windowed signal as follows:
a160 = a(10501:10660);
X(k) = fft(a160,512);
plot(10∗log10(abs(X(k).∗X(k))));
The resulting log spectrum is shown in Figure 7.

- The effect of rectangular window of size 50 msec(assuming 512-point DFT) is illustrated by computing the short-time spectrum of the windowed signal as follows:
a400 == a(10501:10900);
X(k) = fft(a400,512);
plot(10∗log10(abs(X(k).∗X(k))));
The resulting log spectrum is shown in Figure 8.

OBSERVATION:
The size of the window (rectangular in this case) determines the temporal resolution. Increase in size of the window decreases the temporal resolution as evident from the Figures 6, 7 and 8.

(4) **Effect of shape of the window**

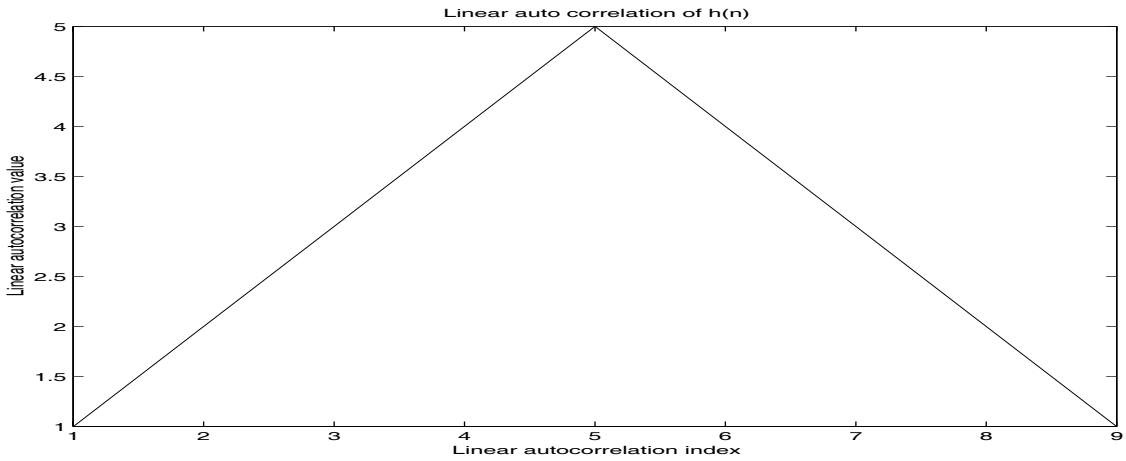

Figure 5: The autocorrelation of h(n).

Consider x(n) as 160 samples of /a/(20 msec).
a160 = a(10501:10660);
- **rectangular window**
The effect of rectangular window of 20 msec on the short-time spectrum is obtained as follows:
rectwindow = boxcar(160);
a160rectwindow= a160.*rectwindow;
X1= fft(a160rectwindow,512);
X=X1(1:256);
plot(10.∗log10(abs(X).*(X)));

The effect of rectangular window is shown in the Figure 9. **Hamming window**
The effect of hamming window on the short-time spectrum is obtained as follows:
hammingwindow = hamming(160);
a160hammingwindow = a160.*hammingwindow;
X2=fft(a160hammingwindow,512);
X=X2(1:256);
plot(10.*log10(abs(X.*X)));
The effect of hamming window on the log spectrum of a segment of speech signal is shown in Figure 10.

Figure 6: The Log spectrum of x(n), using rectangular window of size 5 msec.

Hanning window
The effect of hanning window on the short-time spectrum is obtained as follows:
hanningwindow = hanning(160);
a160hanningwindow = a160.*hanningwindow;
X2=fft(a160hanningwindow,512);
X=X2(1:256);
plot(10.*log10(abs(X.*X)));
The effect of hanning window on the log spectrum of a segment of speech siganl is shown in Figure 11.

OBSERVATION :
Sidelobe effect is dominant in the rectangular window.
Pitch harmonics can be clearly seen when Hamming or Hanning window is used.
Sidelobe attenuation is more with Hamming and Hanning windows and also the main lobe width increases.

(5) **Short-time spectra of voiced and unvoiced speech**

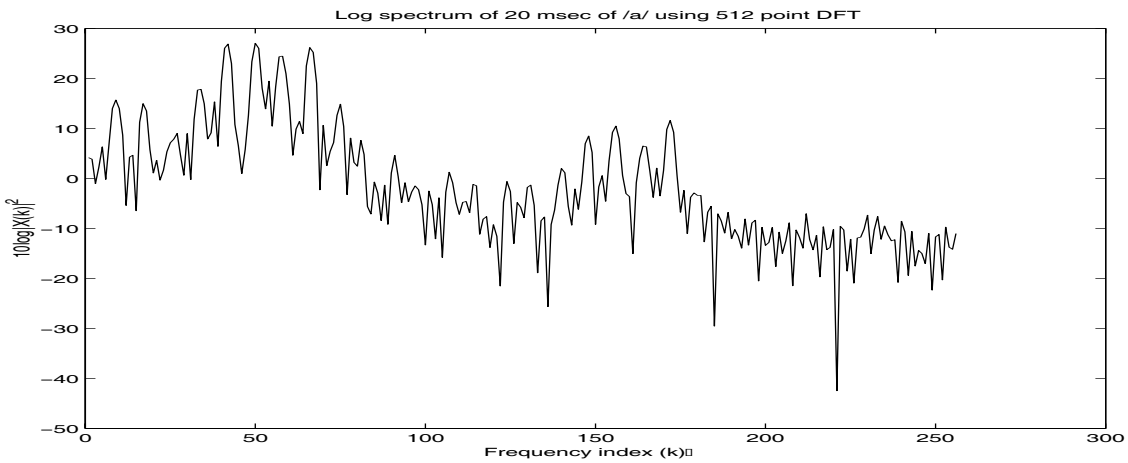

Figure 7: The Log spectrum of x(n), using rectangular window of size 20 msec.

The short-time spectra of 20 msec data of voiced speech segment of /a/, using Hamming window and 512-point DFT is computed as shown below.

```
hammingwindow = hamming(160);
a160hammingwindow = a160.* hammingwindow;
X=fft(a160hammingwindow,512);
X1 = X(1:256);
plot(10.*log10(abs(X1.*X1)));
```

The resulting log spectrum is shown in the Figure 12 (a).

The log spectrum of 20 msec data of unvoiced speech segment of /s/, using Hamming window and 512-point DFT is computed as follows.

```
s = wavread('s.wav');
s160 = s(6501:6660);
hammingwindow = hamming(160);
s160hamming = s160 .* hammingwindow;
X = fft(s160hamming, 512);
X1 = X(1:256);
plot(10.*log10(abs(X1.*X1))) ;
```

The resulting log spectrum is shown in the Figure 12(b).

OBSERVATION :

Voiced speech log spectrum shows clearly the harmonic (source feature)

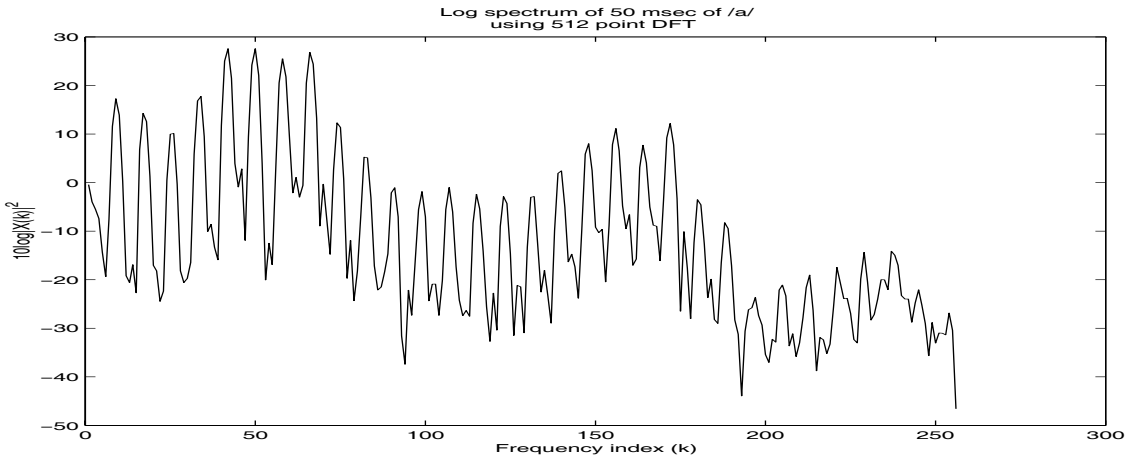

Figure 8: The Log spectrum of x(n), using rectangular window of size 50 msec.

and the formant structure (system feature) whereas there is no defined structure in the log spectrum of Unvoiced speech segment.

(6) **Spectrograms** Record a short utterance that contains both voiced and unvoiced components using the command :
brec -s 10000 -b 16 -t 2 -w filename.wav
where 's' is the sampling rate (Hz), 'b' is the number of bits/sample, 't' is the duration and w is the wave format.
The wideband and narrowband spectrograms for a given speech utterance is obtained by using the wavesurfer utility.
In wideband spectrogram a small time window (typically of duration 5 msec) is used. In Narrowband spectrogram
realtively larger time window (typically of duration 50 msec) is used. The wideband spectrogram is characterized by the vertical lines and the narrowband spectrogram by the horizontal lines.The resulting spectrograms are shown in Figure 13.

OBSERVATION Wideband spectrogram provides a better temporal resolution (vertical striations) and narrowband spectrogram provides a better spectral resolution (horizontal striations).

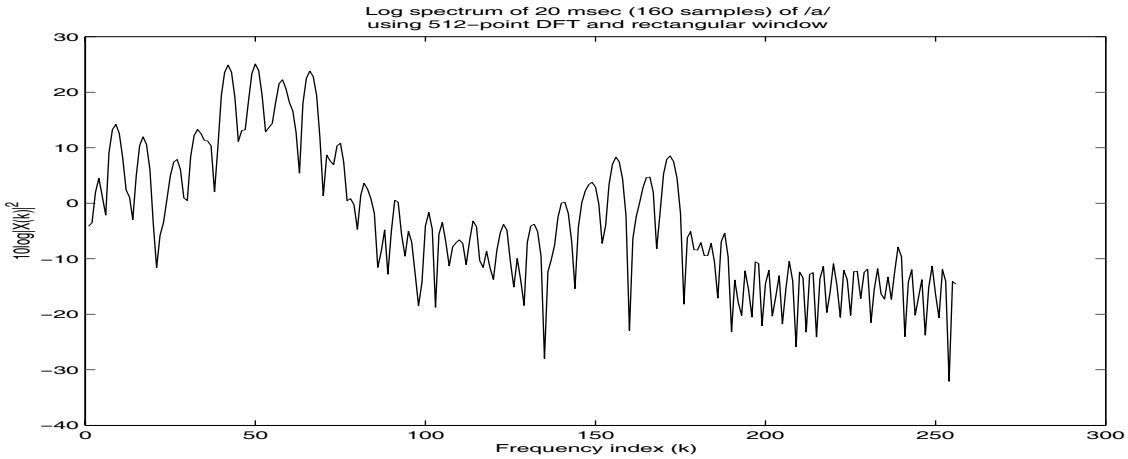

Figure 9: The Log spectrum of x(n), using rectangular window of size 20 msec.

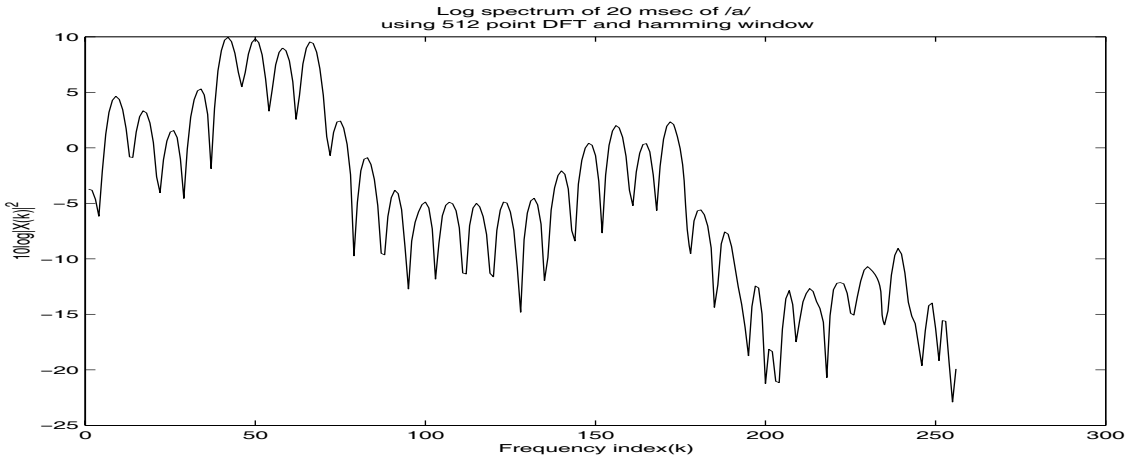

Figure 10: The Log spectrum of x(n), using hamming window of size 20 msec.

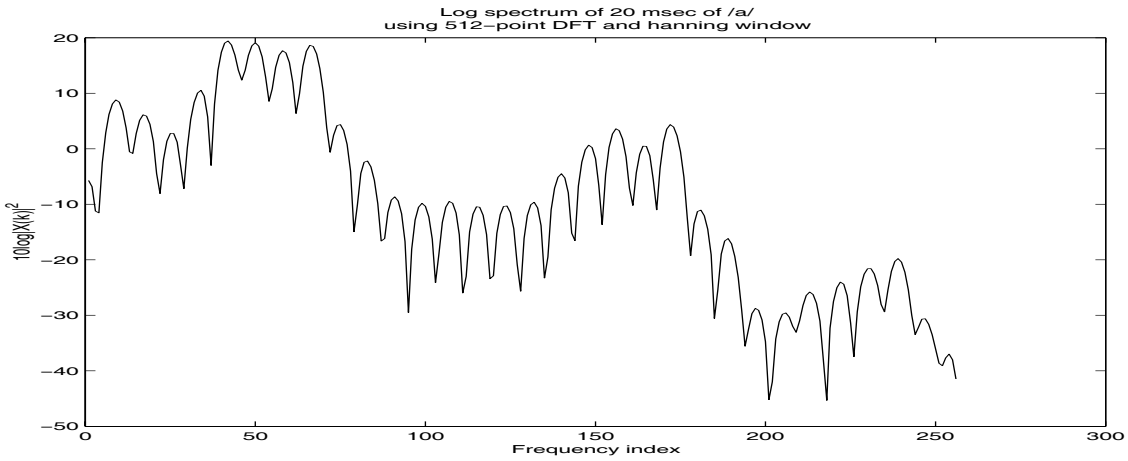

Figure 11: The Log spectrum of x(n), using hanning window of size 20 msec.

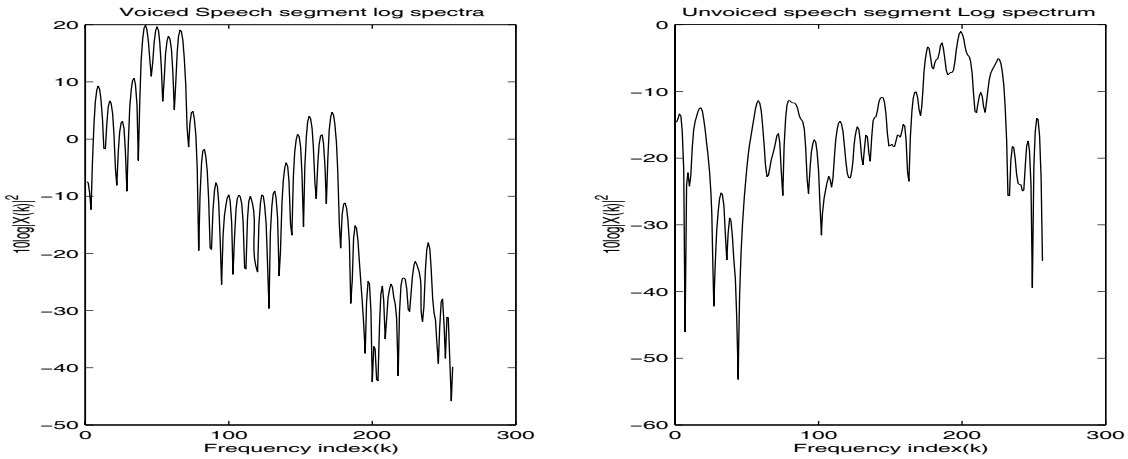

Figure 12: The Log spectrum of a) voiced speech segment b)Unvoiced speech segment.

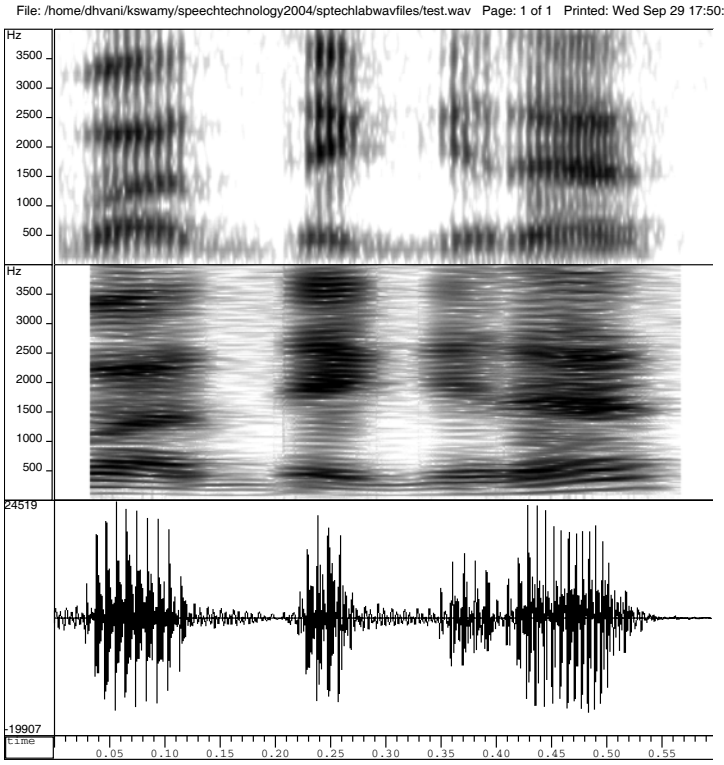

Figure 13: a. Wideband Spectrogram b. Narrowband spectrogram c. Segment of Speech signal

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.

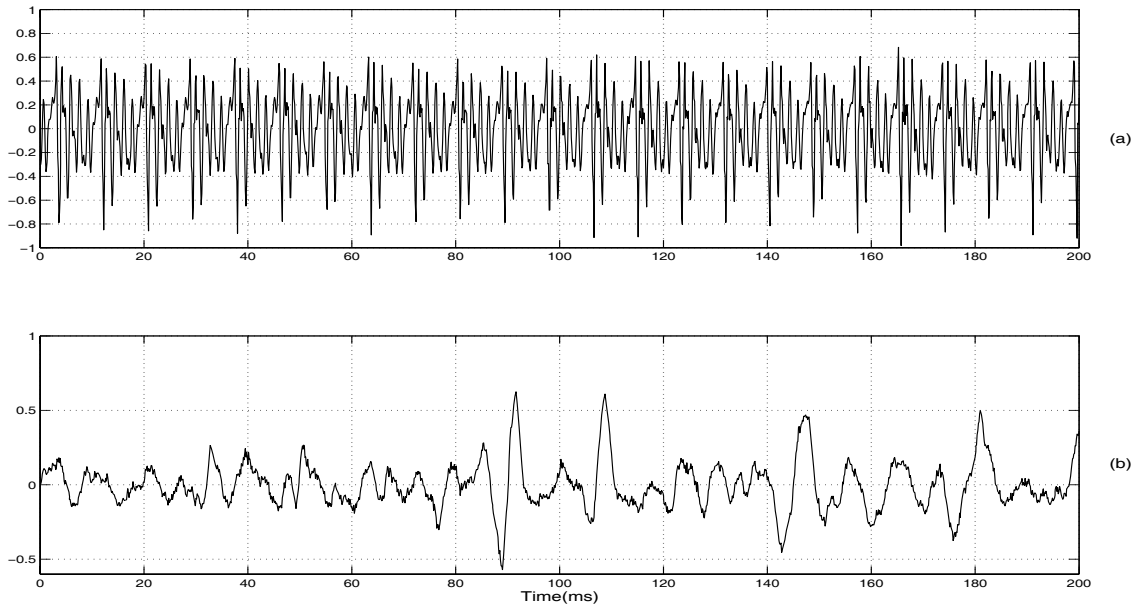

Figure 1: (a) Segment of voiced speech /a/ and (b) segment of unvoiced speech /s/

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.

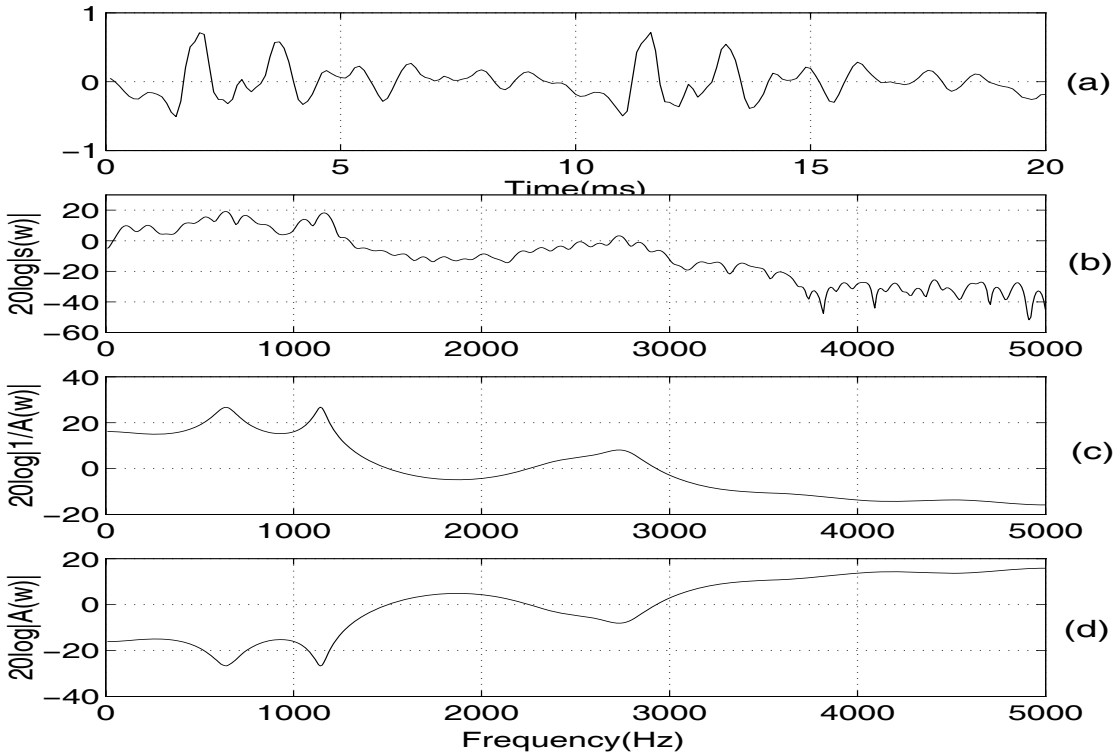

Figure 2: (a) Segment of voiced speech /a/ and its (b) short time spectrum, (c) LP spectrum and (d) inverse spectrum

- **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.

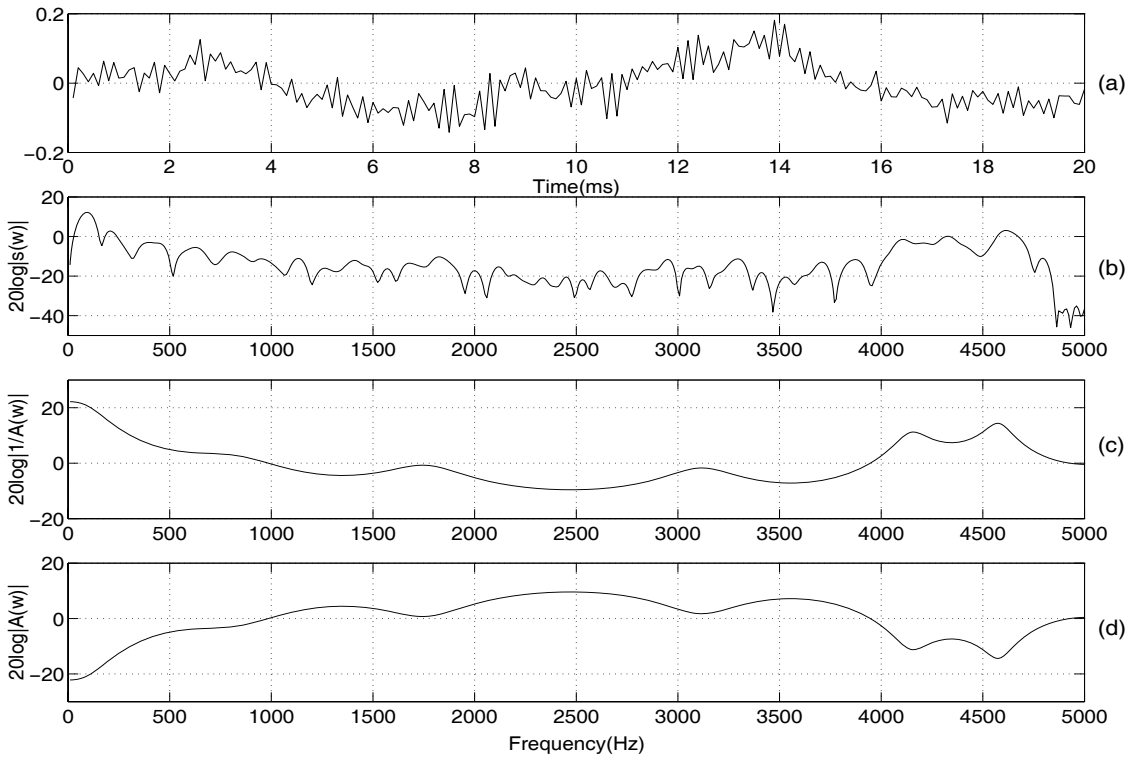

Figure 3: (a) Segment of unvoiced speech /s/ and its (b) short time spectrum, (c) LP spectrum and (d) inverse spectrum

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;

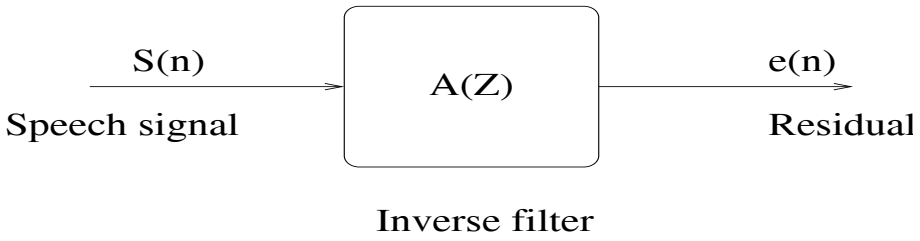

Figure 4: Inverse filter to obtain LP residual signal from speech signal

- 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

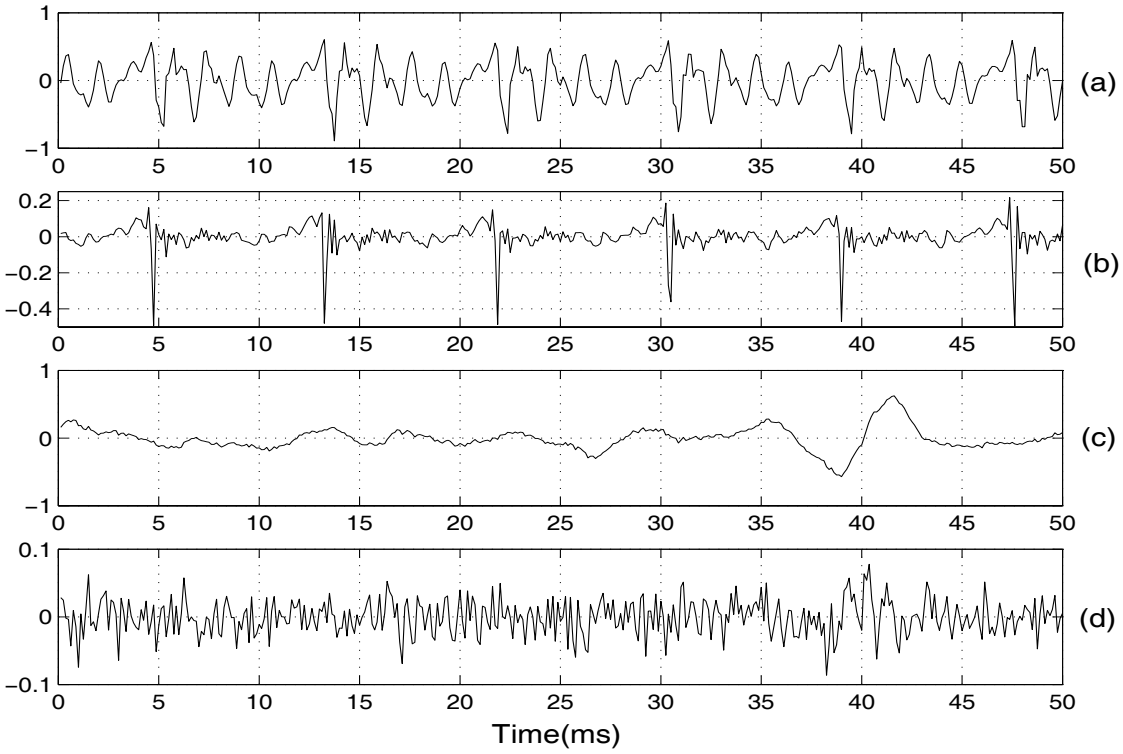

Figure 5: (a) Segment of voiced speech /a/ and its (b) LP residual signal, (c) segment of unvoiced speech /s/ and its (d) LP residual signal.

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-

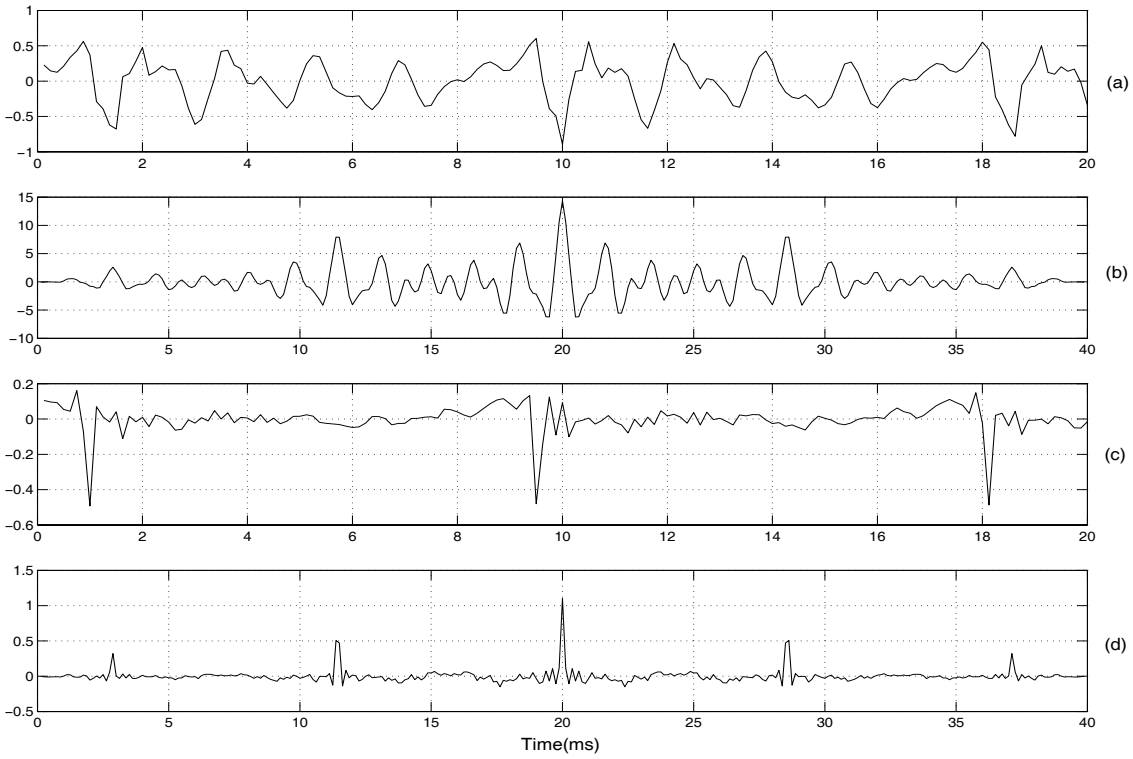

Figure 6: (a) Segment of voiced speech /a/ and its (b) autocorrelation function, (c) LP residual for the voiced speech segment and its (d) autocorrelation function.

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

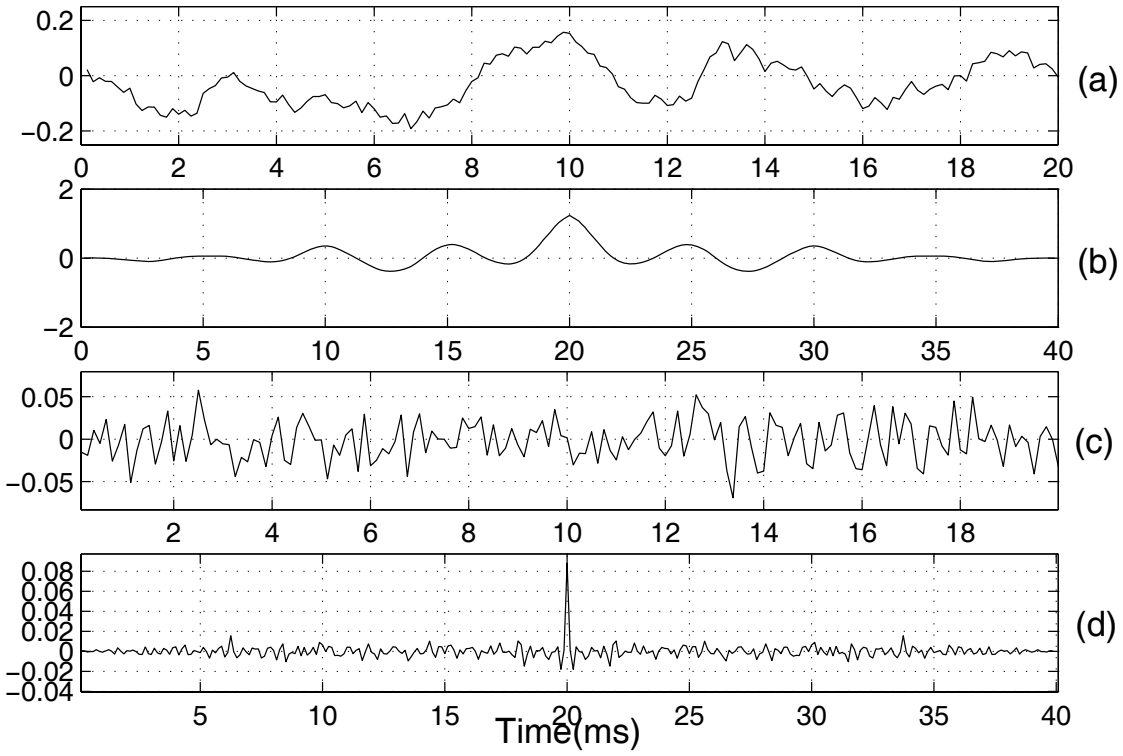

Figure 7: (a) Segment of unvoiced speech /s/ and its (b) autocorrelation function, (c) LP residual for the unvoiced speech segment and its (d) autocorrelation function.

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.

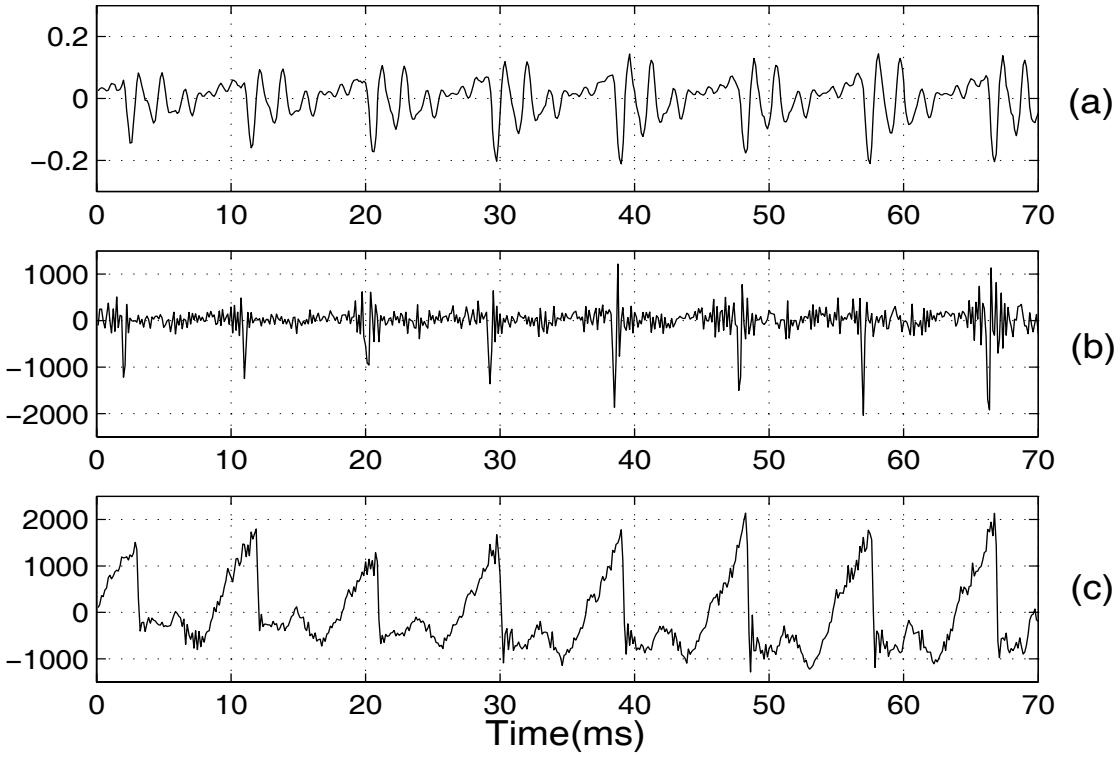

Figure 8: (a) Segment of voiced speech /a/, its (b) LP residual and (c) glottal pulse waveform.

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.

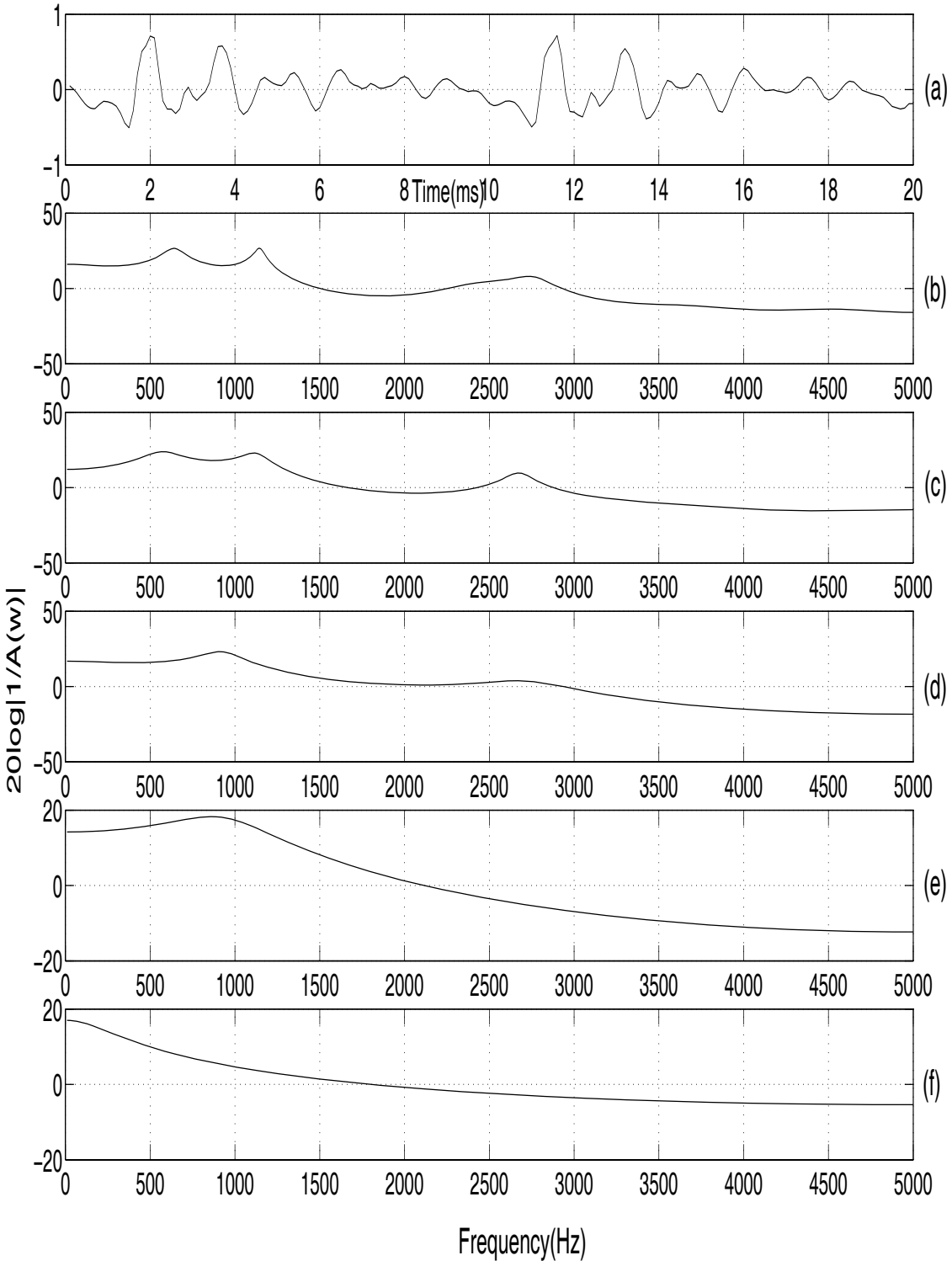

Figure 9: (a) Segment of voiced speech /a/, its LP spectrum for the LP order (b) 14, (c) 10, (d) 6, (e) 3 and (f) 1

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.

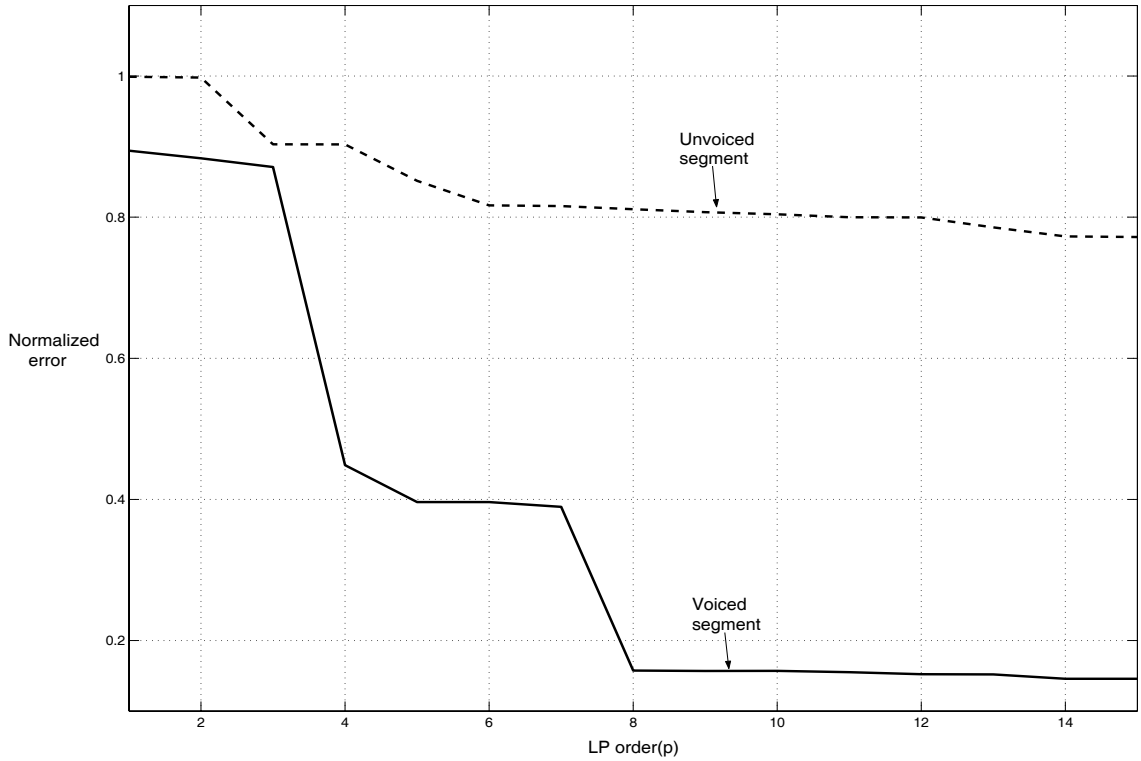

Figure 10: Normalized error for voiced and unvoiced speech segments for different LP orders.