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

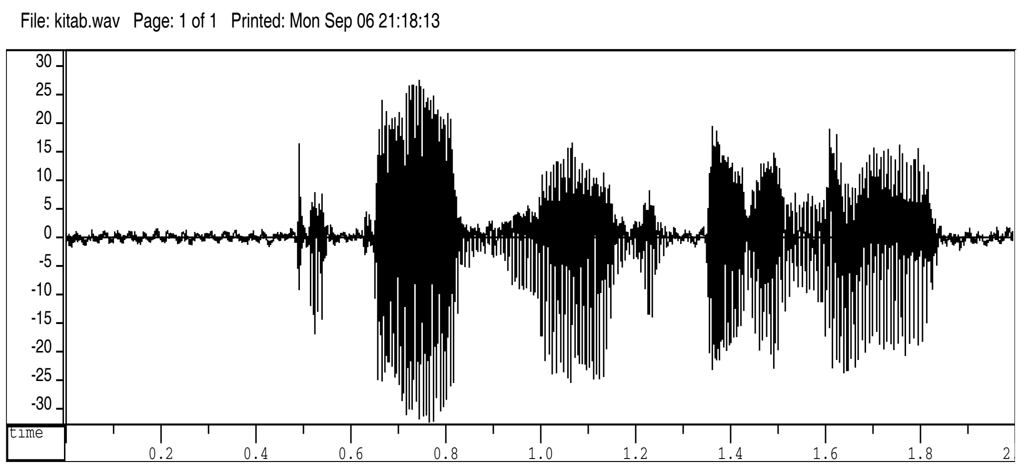

Figure 1: Speech waveform for the utterance "kitAb mEj par hai"

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 portions 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	Frica-tives
	Unvoiced		Voiced				
	Unaspi-rated	Aspi-rated	Unaspi-rated	Aspi-rated			
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.

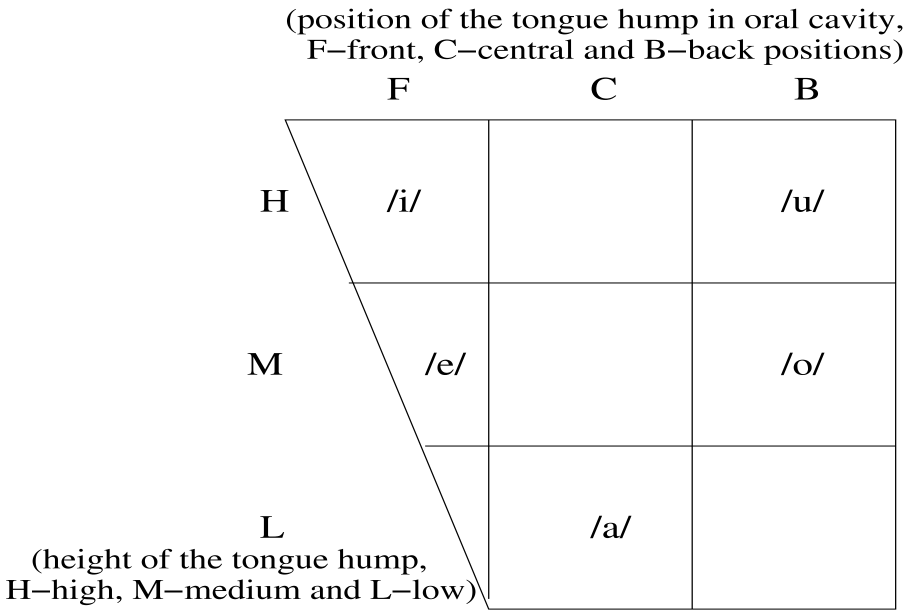

Figure 2: Position and height of the tongue hump for producing different vowels

/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.

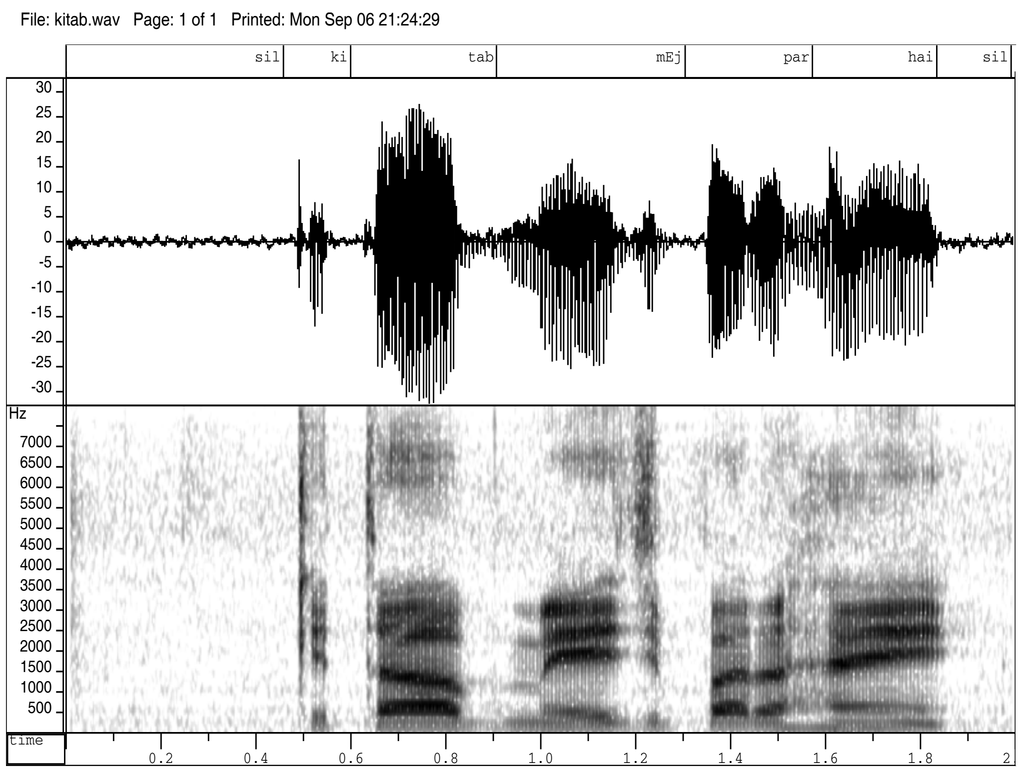

Figure 3: Speech waveform and its wideband spectrogram for the utterance "kitAb mEj par hai"

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