Indian Institute of Technology, Kharagpur

Centre for Educational Technology

**Mid Semester Examination 2013**

Subject**: INTRODUCTION TO DIGITAL SPEECH PROCESSING** Code: ET60007

Time: 2:00 Hours PART-A:-10\*2=20

PART-B:-5\*6=30

**Full Marks =50**

***Answer all the questions of PART-A and PART-B***

*(Please enclose the Annexure-1 along with the answer script)*

**PART-A**

1. Table 1 shows an *F1* and F2 of the vowels of a language plot the vowels in *F1* and *F2* plane and mark the axis intern of tongue height tongue position.

Table-1

|  |  |  |
| --- | --- | --- |
| Vowels | F1 [Hz] | F2[Hz] |
| /u/ | 325 | 1035 |
| /o/ | 378 | 1025 |
| /ɔ/ | 543 | 1019 |
| /a/ | 866 | 1530 |
| /æ/ | 591 | 1846 |
| /e/ | 383 | 1978 |
| /i/ | 309 | 2131 |

1. In case of sound perception why we need two ears?
2. Acoustic intensity of an audio system is *5W/m2*. Represent it in *dB* and find out the Loudness (L) in Sones.
3. Write the phonetic transcription for the following words. How many syllables are present in each of the above word?
4. “Production” (b) “Perception”
5. Write the place and manner of articulation of the following phonemes?

/gh/, /u/, /m/, /ɖ/

1. First formant frequency of a steady state vowel is *F1=250Hz*. Consider that the vowel is produced using a single lossless acoustic tube. What will be the length of the vocal tract? How long is the impulse response of the tube? Where the speed of sound *c=350m/s*.
2. Two source of speech signal producing a vowel /o/. The formant frequencies of the sources are differing by 3-4%. If a human being perceives the two sounds as same sound explains why this happen?
3. *2 sec*. speech segment is encoded using LPC coefficient and the LPC coefficient are extracted for each frame (frame length(L)= 5 pitch period) with a frame rate *100 frame/s*. Determine the required order of the LPC analysis and how many frame’s LPC coefficient can be extract from the above speech signal. where the *F0* of the speech segment is *250 Hz* and sampling frequency *Fs=16 kHz*
4. Define critical band for human auditory system. What will be centre frequency of the 7th critical band in bark scale? Where the bark is as define below.
5. Which of the following pair of tones is perceived as louder tone?

(a) *25dB* level at *200Hz* and *25 db* at *600 Hz* (b) *5dB* level at *7 KHz* and *5dB* level at *2 KHz*

**PART-B**

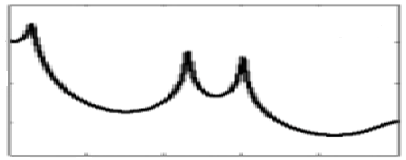
1. (a) An analog signal *xa(t)=6sin(320t)+4cos(820t)* is sampled at *600 Hz*. Dertermine the *F0* of the signal. How many sampled will be present in one fundamenal period?

(b) A discrete signal *x[n]=6.35 cos(n/5)* is quantized with aresolution *=0.01.* how many bits are required in A/D conversion?

(c) If the above degitized signal is passthrough an digital to analog converter, derive the expresion of the reconstracted signal.

1. Linear prediction analysis is used to obtain a 6-order all-pole model for a segment of voiced speech that was sampled at a rate of *FS = 10000 Hz*. Figure -1 shows the LPC spectrum of the above extracted LPC coefficient. Determine the root magnitude and root angle of the pole corresponding to first three formant*. F1=288Hz , F2=719Hz, F3,=2294Hz BW1=92 Hz, BW2=65Hz, BW3=50Hz*

F1



BW1

BW2

BW3

Frequency

F3

F2

Figure-1

1. Bilabial unvoiced fricative sound is produce by constrictions in vocal tract at lips. Consider the vocal tract is model using a lossless uniform tube of length *l*. The output sound pressure at lips to produce the above fricative sound is *PL()ejt.* Using the wave equation derive the transfer function *H()* (*where H()=UL()/PL())*

Using the radiation load *RL=jLrRr/Rr+jLr* derive the expression of sound pressure at front of lips.

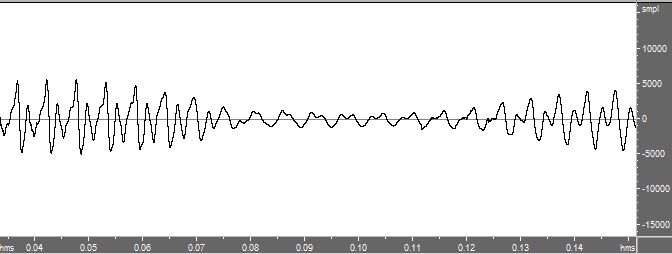
1. (a) Figure-2 (a) and (b) in annexure-1 represents waveform and spectrogram of a VCV speech segment where C represent consonant and V represent Vowel. Mark the occlusion period, burst and VOT part? Write the manner of articulation of the consonant represented by the figure-2.

(b) Figure-3 in annexure-1 represent wideband spectrograms of a speech segment mark the following region your knowledge of acoustic phonetics i) voiced segment ii) unvoiced segment iii) any one consonant to vowel transition segment

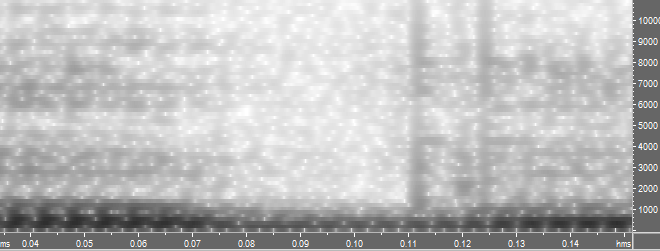
1. (a) Draw the LPC decoder and encoder block diagram

(b) A signal is sampled at *16 KHz, 16 bit*, encoded with 18th order LPC. Each of the LPC coefficients is encoded with *2 byte*, Gain in *2 byte*. Voiced unvoiced *F0* information is encoded using *1 byte*. What is the compression ratio if frame rate is *100 frame /sec*?

Annexure-1

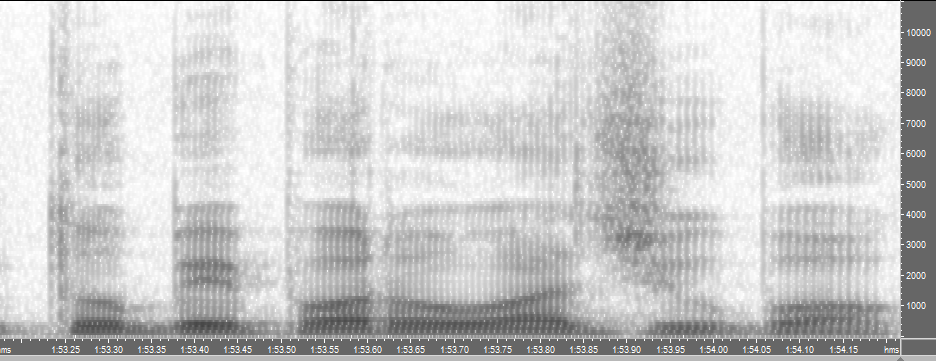


(a)



(b)

*Figure-2*



*Figure-3*