



Electrical and Computer Engineering

Spoken Language Processing – Fall 2017

Midterm Exam

Name:

ID:

---

**Question 1: [11 marks]**

(a) Two of the criteria used for classification of speech sounds are 'manner of articulation' (i.e. how the sound is made) and 'type of excitation'. Give the category name to which the phonemes **/K/**, **/SH/**, **/AH/**, and **/M/** belong to according to each criteria. [4pts]

(b) Describe, in general term, what the waveforms look like for speech sounds belonging to the same category as phonemes **/K/**, **/SH/** and **/AH/**. [3pts]

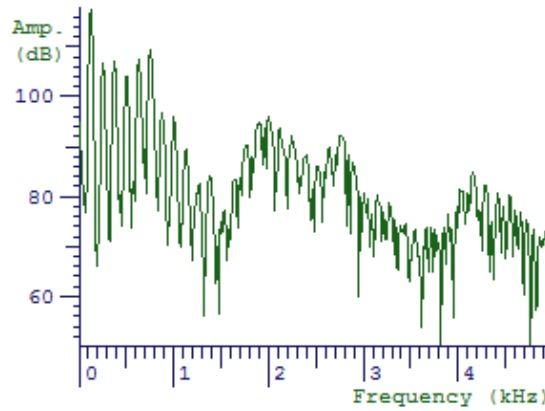
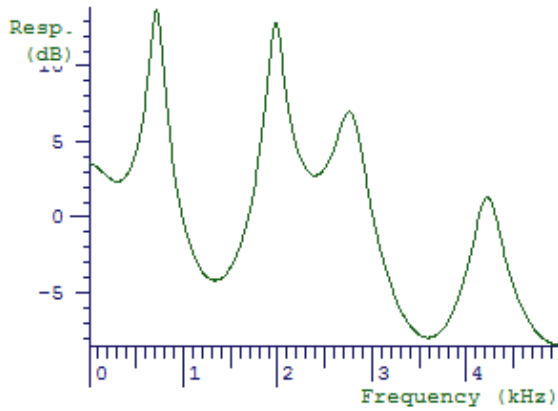
**(c)** What is spectrogram? Compare between narrow and wide –band spectrograms in term of window length, frequency resolution and temporal resolution? [4pts]

**Question 2:[12 marks]**

**(a)** Draw a block diagram for source-filter model of speech production and explain its correspondence to the human speech production process. [3pts]

(b) Give an equation for each block and for the overall speech production model in z-domain. [3pts]

(c) The figure below shows the 'filter response' [left] and 'output spectrum' [right] for a short speech frame. What are these two graphs? Explain how the right-hand graph is related to the left-hand graph? Is this voiced or un-voiced frame? Explain? [4pts]



(d) Mention two benefits of applying discrete cosine transform (DCT) in the Mel-frequency Cepstral coefficient (MFCC) feature extraction technique? [2pts]

**Question3:[10 marks]**

(a) Given the following windowed speech segment,

$S(n) = [-0.35, 0.0, 3.15, -2.5, -3.54, 2.8, 0.0, -0.28]$ , with sampling frequency of **600** sample/sec.

Find the following basic features (show the equation for calculating each one):

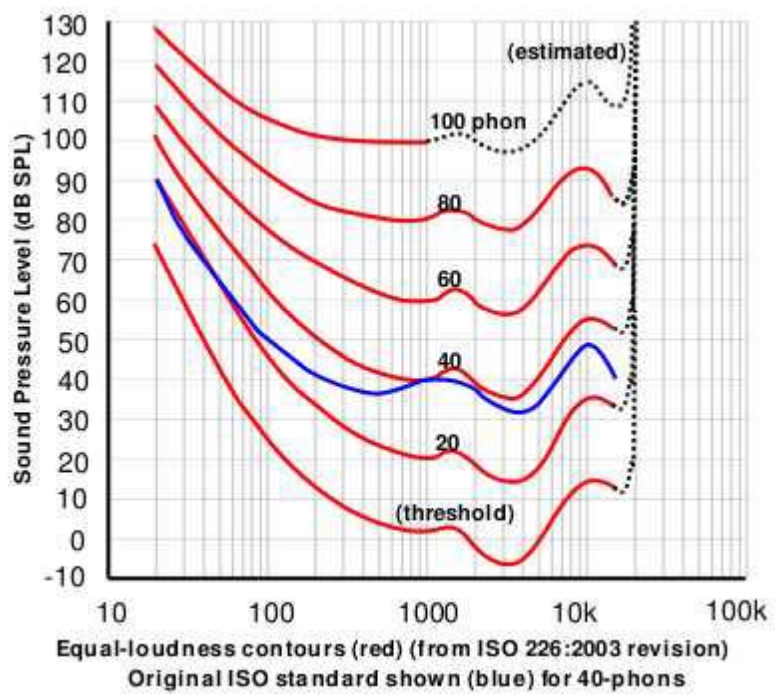
(i) Energy in decibels (dB). [1pt]

(ii) Zero-crossing rate [1pt]

(iii) Use autocorrelation method to find Pitch period  $T$ , if we assume the fundamental frequency ( $F_0$ ) is in the range 150-300Hz. [2pts]

**(b)** Explain how the basic features, in part (b), can be used for voiced/unvoiced classification of the speech segment? Is the speech segment given in (b) above voiced or unvoiced? [3pts]

(c) Use the equal loudness curves shown in the following figure to answer the following questions. The line labeled “threshold” is the line below which humans typically cannot hear.



(1) Can humans typically hear a 100 Hz sound at 10 DB SPL? [1pt]

\_\_\_\_\_

(2) Which sounds louder: a tone of 100 Hz at 25 DB SPL or a tone of 1000 Hz at 20 DB SPL? [1pt]

\_\_\_\_\_

(3) According to the graph above, which frequency range best corresponds to the range at which humans are most sensitive to loudness? [1pt]

- A. 3kHz to 10kHz
- B. 2kHz to 5kHz
- C. 500 Hz to 1000 Hz

End of the Exam