جامعة بيرزيت
BIRZEIT UNIVERSITY

Electrical and Computer Engineering

Spoken Language Processing – Spring 2022

Midterm Exam

Name: Solution ID:

Question 1: [20 marks]

(a) Choose the correct answer: [12pts]

1- A speech signal was recorded at 16 kHz. The length of the window for acoustic analysis is set to 30ms. The length of the window in samples is

A – 160 samples

B – 440 samples

C- 480 samples

D- 320 samples

2. Which of the following digital signal processing techniques is used to estimate the pitch frequency (F0)?

A. Average Magnitude Difference Function (AMDF).

B. Linear Predictive Coding (LPC)

C. Cross-correlation

D. Fast Fourier Transform (FFT)

3. Which of the following digital signal processing techniques is used to construct spectrograms?

A. RMS Amplitude

B. Short-time energy.

C. Autocorrelation

D. Fast Fourier Transform (FFT)

4. When conducting Fast Fourier Transforms, longer window sizes, as compared to shorter window sizes, result in

A. discontinuous breaks at the edges of windows.

B. better amplitude resolution, but worse time resolution.

C. better time resolution, but worse frequency resolution.

D. better frequency resolution, but worse time resolution.

5. Sound A is a tone played at 1200 Hz, sound B is played at 2000Hz, sound C is played at 12kHz, sound D is played at 13kHz. All are played at 60dB. Which sounds like a greater increase in pitch?

A. The increase from Sound A to Sound B.

B. The increase from Sound C to Sound D.

C. They pitch increase from Sound A to Sound B is perceived as the same as the pitch increase from Sound C to Sound D.

6. What kinds of speech sounds can easily be identified by the shapes of their formants?

A. Consonants.

B. Vowels.

C. Syllables.

D. All phonemes.

7. Which is NOT an example of variability in the speech signal?

A. The lack of boundaries between words and phonemes

B. People with different sized/shaped vocal tracts

C. People talking in noisy environments

D. Dialect /accent differences

8. From the point of view of speech PERCEPTION, coarticulation of speech sounds helps us to:

A. figure out words quickly

B. speak faster

C. create phonemic categories quickly

D. perceive F1 and F2 quickly

9. A whispered sound may be said to be _____ .

A. voiced

B. unvoiced

C. stopped

D. a fricative

10. Normal speech has an intensity of around _____.

A. 30-35 dB

B. 40-50 dB

C. 55-65 dB

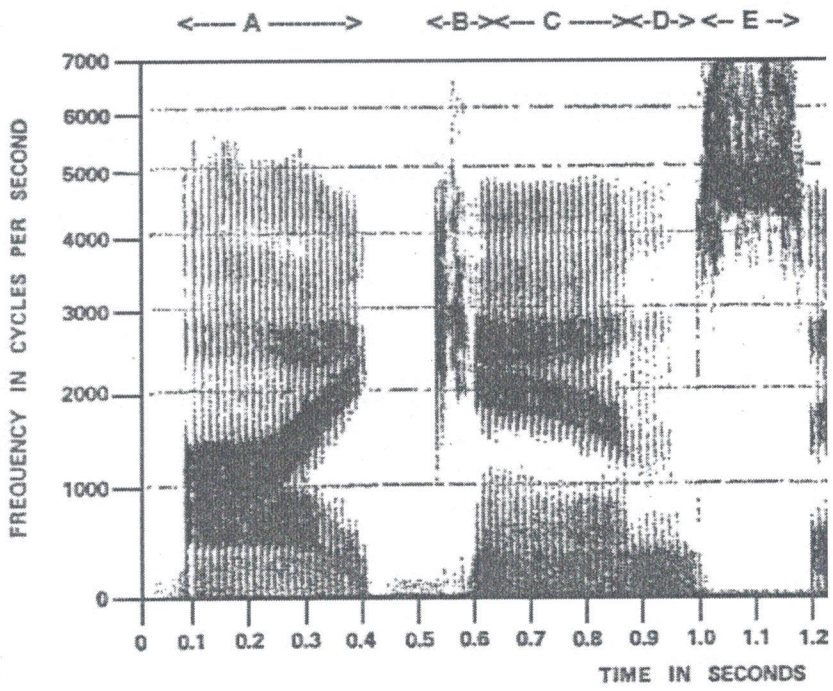
D. 75-85 dB

Question 1 answers:

1	2	3	4	5	6	7	8	9	10
C	A	D	D	A	B	A	B	B	C

Question 2: [14 marks]

The figure below shows a spectrogram od a speech segment.



(i) Estimate the bandwidth of this signal in Hz, as accurate as you can? [4pts]

from the spectrogram:

B.W. is 7000Hz or 7kHz

(ii) for each of the identified areas in the figure, A, B, C, and D as (voiced/unvoiced), and as (consonent, vowel or nasal):

	Voiced/unvoiced	consonant/vowel/nasal
(i) Area A:	Voiced	Vowel
(ii) Area B:	Unvoiced	Consonent
(iii) Area C:	Voiced	Vowel
(iv) Area D:	Voiced	Nasal
(v) Area E:	Unvoiced	Consonent

Question2:[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.

n: 0 1 ... 7

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

(i) Energy in decibels (dB). [3pt]

$$\text{Energy(dB)} = 10\log_{10}\sum_{n=0}^{7}[s(n)]^2 = 10\log_{10}\sum_{n=0}^{7}(s(n))^2$$

$$= 10\log_{10}\left[(-0.35)^2 + (3.15)^2 + (-2.5)^2 + (-3.54)^2 + (2.8)^2 + (-0.28)^2\right]$$

$$= 10\log_{10}[36.745] = 15.652 \text{ dB}$$

(ii) Zero-crossing rate [3pt]

$$ZCR = \frac{1}{N}\sum_{n=0}^{N-1} 0.5\,|\,\text{sign}(s(n)) - \text{sign}(s(n-1))\,|$$

$$= \frac{1}{8}[4] = 0.5$$

Zcc = 4 crossings.

(b) Explain how the basic features, in part (a), can be used for voiced/unvoiced classification of the speech segment? Is the speech segment given in (a) above voiced or unvoiced? [4pts]

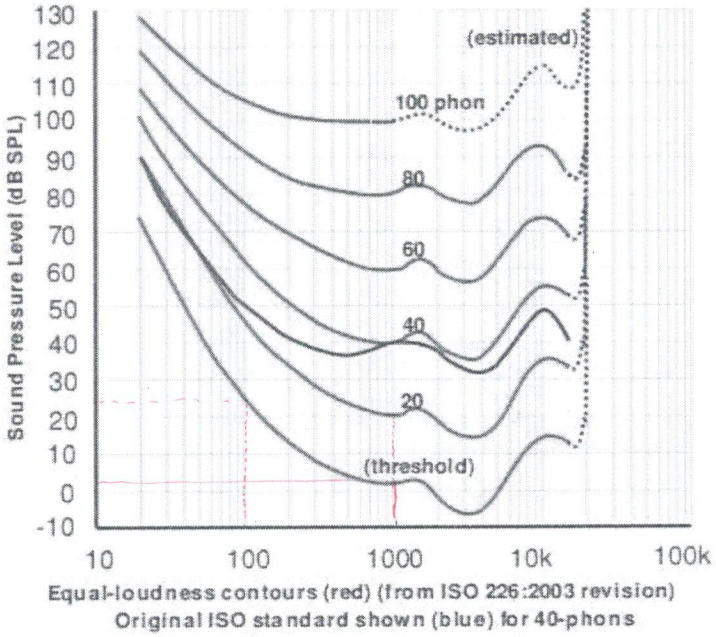
High Energy and low Zcc ⇒ Voiced speech.

Low Energy and High Zcc ⇒ UnVoiced speech.

In this speech segment ⇒ high energy and relativily low Zcc ⇒ This segment looks like a Voiced speech seg.

Question 4 [6 marks]

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 1000 Hz sound at 10 dB SPL? [2pt]

Yes, Because Min. threshold at 1 kHz is less than 10dB.

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

1000 Hz at 20 dB received louder, because Min. threshold of 1kHz is much lower

(3) According to the graph above, which frequency range best corresponds to the range at which humans are most sensitive to loudness? [2pt]

A. 3kHz to 10kHz

B. 2kHz to 5kHz

C. 500 Hz to 1000 Hz