BIRZEIT UNIVERSITY

Electrical and Computer Engineering

Spoken Language Processing, Fall 2017

Dr. Abualsoud Hanani

كل هادي الصفحة عالدوكس محلولة

Question sheet 1 (Review questions which you should now be able to answer)

1. What is a phoneme? Minimal unit of sound

2. What is the difference between the phonemes /b/ and /p/? 'b' is voiced, 'p' is unvoiced

3. What is the name given to the class of phonemes which includes /s/, /f/ and /sh/ (as in /sheet/)? unvoiced, fricatives, consonants

4. What is the name given to the class of phonemes which includes /p/, /t/ and /k/? unvoiced, stops, consonants

5. What is the source-filter model of speech production? Draw a diagram.

6. What is a formant?

7. What is the difference between a narrow-band and wide-band spectrum? What is the difference in terms of DFT analysis window?

8. Given a segment of speech waveform, how can you estimate the fundamental frequency?

9. What is autocorrelation function and what we used it for?

10. How do the glottal source spectrum and vocal tract transfer function combine to produce speech?

11. What is short-term cepstrum? What information is carried by the lower cepstral coefficients?

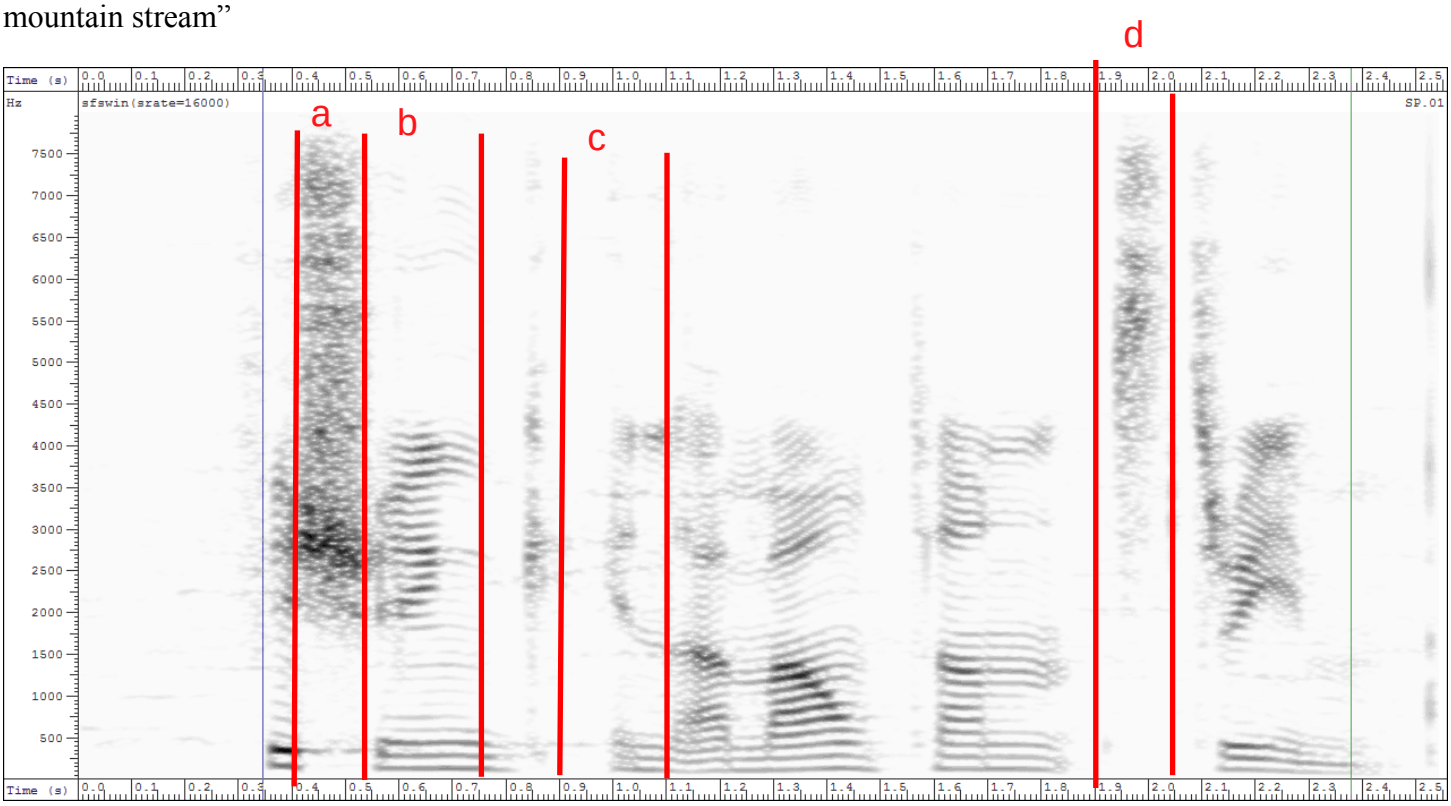
12. Draw a waveform and short-term spectrum of a voiced and unvoiced sound.

13. What is a critical bandwidth?

14. What is the mel frequency scale?

15. How many mels equal 1kHz?

16. The figure below shows an SFS display of a speech spectrogram for an example of a phrase “fishing in a mountain stream”



(a) for each of the following start times and end times, identify the voicing classification (voice/unvoice) and the manner of articulation classification (fricatives, vowel and plosives).

(i) start time 0.4s, end time 0.54s fricative, unvoiced.

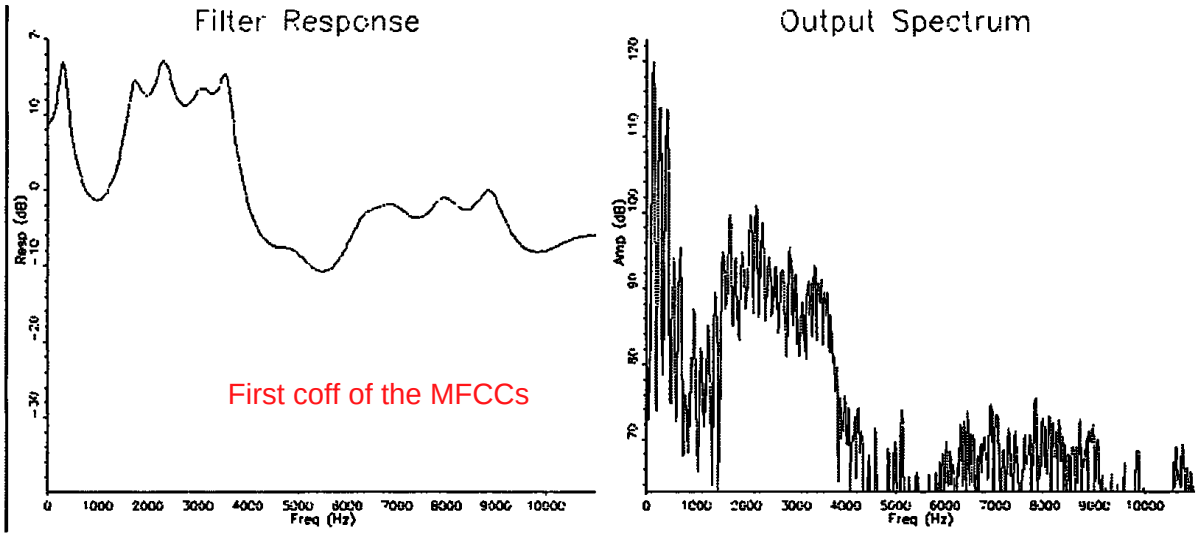
(ii) start time0.54s , end time 0.76s vowel, unvoiced

(iii) start time 0.9s, end time 1.1s plosive, voiced

(iV) strst time 1.9s, end time 2.05s fricative, unvoiced

(b) What is meant by ‘source-filter’ model of speech production? Your answer should include a diagram.

(c) the figure below shows the ‘filter response’ and ‘output spectrum’ at time 2.2s . what are these two graphs? Explain in detail how the right-hand graph is related to the left-hand grapg.



(d) What is the short-term cepstrum? Explain in detial how it can be used to recover the left-hand grapg from the right-hand graph in part (c).

17.

(a) Draw a diagram of the 'source-filter' model of speech production and explain its correspondence to the human speech production process.

مكرر

(b) What is the short-term cepstrum? MFCC for each frame

مكرر

(c) Consider a 30ms segment of a vowel sound pronounced by a female person. Draw an example of the short-term spectrum and short-term cepstrum of such a signal. Explain in detail what information is contained in the figures.

عالدوكس محلول

(d) What is meant by the term 'critical bandwidth'? Describe an experiment that could be performed to demonstrate the critical bandwidth. How does the concept of critical bandwidth influence the front-end for automatic speech recognition?

masking

(e) What is meant by the term 'delta cepstrum', and what is the motivation for its use in automatic speech recognition.

نحسب سرعة التغير بين كل تنين ورا عض

speech signal is not constant (slope of formants, change from stop burst to release).
• So we want to add the changes in features (the slopes).

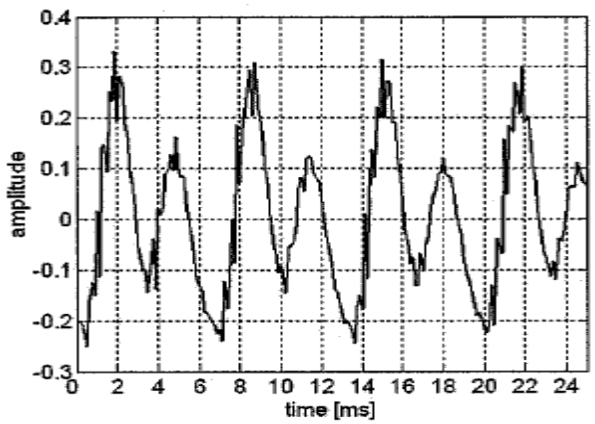
18.

مكرر

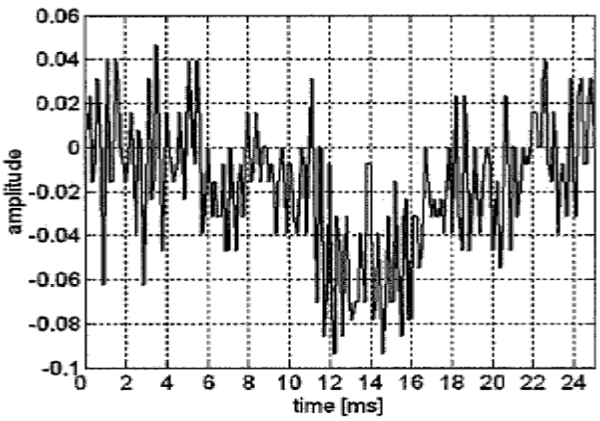
(a) Describe the stages involved in transforming a speech waveform into a sequence of mel-frequency cepstral coefficients (MFCCs) using a Discrete Fourier Transform.

(b) Suggest two reasons why a cepstral representation typically gives better results than a spectral representation in a speech recognition system? reducing dimensionality and redundancy of data[12 features only], separate vocal tract from source

(c) Figure 1 depicts the waveform of two speech sounds. The x-axis depicts time in [ms], the y-axis is the amplitude of the signal.



(a)



(b)

a)voiced,periodic,high power
b)unvoiced,non periodic, low power

(i) Classify the sounds based on the type of excitation (voiced/unvoiced)? Justify your answer.

(ii) Estimate the fundamental frequency of the voiced sound as accurately as you can? Explain your calculations? 15-9 ms = 6ms

(d) Draw a figure illustrating an example of the frequency characteristic of the vocal tract when producing vowel sound? Explain the figure? الفارمنتس بنزلو

(e) Consider the IPA classification of speech sounds based on manner of articulation (i.e. how sound is made). Which categories do the speech sounds /m/ and /p/ belongs to? For one of these sounds, explain how the sound is made? Give properties of the sound and give more examples of phonemes belong to the same category?

m=> nassle,consonant ==> n
p=>unvoiced,plosive ==> t

19.

(a) Noise corruption in speech signals can be categorized into two general types. additive,convolutive

(i) Give their names and give an example illustrating how each type of noise might occur in practice.
conv from channel. Additive from the surrounding env

(ii) What is the effect of each type of noise at each stage of the front-end processing producing MFCCs? (your answer should include a block diagram of the front –end processing).

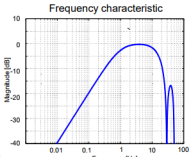
(iii) Describe the techniques called 'Cepstral Mean Subtraction' and 'Rasta filtering'.

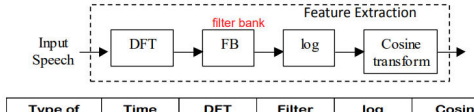
- Cepstral Mean Subtraction (CMS)
  - mean (over a num of frames) subtraction
  - lowpass filtering
  - eliminates communication channel spectral shaping

•Filtering log filter bank output (or equivalently cepstral) temporal trajectories by band pass filter

•Remove slow changes to compensate for the channel effect (≈CMS over 0.5 sec. sliding window)

•Remove fast changes (> 25Hz) likely not caused by speaker with limited ability to quickly change vocal tract configuration





Type of Noise	Time Domain	DFT	Filter Bank	log	Cosine Transform
Additive	+	+	~ +	? ~ max	?
Convolutive	conv*	×	~ ×	~ +	~ +

20.

(a) What is meant by Linear Prediction Coding (LPC) in context of speech production modeling?

طريقة نقرب فيها السامبل من خلال السامبلز اللي قلبها, بالتالي بنستعمل هادي الطريقة عشان نطلع معاملات الفوكل تراكت من خلال تقليل الايرور في منطق الصفر للبلص ترين

(b) Draw a block diagram of LPC processing?

(c) Explain briefly how do we estimate the vocal tract response using LPC technique?

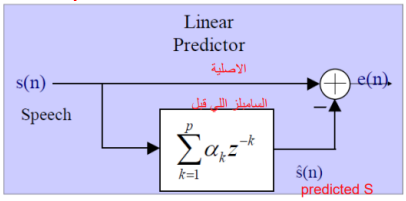
بنقلل الايرور

(d) Explain how LPC can be used for estimating pitch period?

الفرق بين المناطق اللي الايرور فيها اعلى اشي

(e) What is a pre-emphasis filter and why it is used in LPC analysis?

يشيل اخر بول من الجلوتل



21.

(a) Explain two methods for calculating short-time energy for speech signal. [4]

(b) Given the following short speech segment,

S (n) = [3.4 -3.5 0.4 -2.2 1.2 -2.4 1.8 3.0 6.8 -0.8], with sampling frequency of **600** sample/sec.

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

(i) Energy [4] بنربع كل سامبل وبنجمع

(ii) Zero-crossing count [4] اختلاف بالاشارة

(iii) Pitch period **T**, if we assume the fundamental frequency (F_0) is in the range 100-300Hz. [9] F0 = Fs/P[in samples]

(c) Explain how the basic features, in part (b), can be used for voiced/unvoiced classification of the speech segment? [4] voiced: high energy, low ZCC
Unvoiced : low energy,high ZCC

22.

(a) Draw a block diagram for the Mel-frequency Cepstral Coefficients (MFCC) feature extraction? [10]

(b) Describe, briefly, the functionality of each block? [5]

(c) Mention two benefits of using Discrete Cosine Transform (DCT) in the MFCC feature extraction? [5]

-DCT produces highly uncorrelated features
-Since the logpower spectrum is real and symmetric, inverse DFT reduces to DCT(better in computation)

23.

LPC:

$$\sum_{k=1}^{p} \alpha_k R(k-i) = R(i) \quad \text{for } i = 1,2,3 \ldots\ldots p$$

(a) Explain how autocorrelation can be used for estimating LPC parameters?

(b) What is Durbin's algorithm? Why it is used in LPC analysis? (no need to write algorithm steps)

(c) Explain how LPC can be used for speech coding?

(d) Explain how LPC can be used for speech synthesis?

(e) Compare between LPC and MFCC feature extraction techniques? Which is better for speech recognition?

(f) Given autocorrelation coefficients R(k) = {1, -3.2, 2.4}, for k=0,1,2. Write the matrix equation of finding LPC parameters? Use Durbin's algorithm to find LPC parameters (i.e. a1, a2, a2).

Toeplitz Matrix,

بقلل عدد السامبلز

بعد ما نوجد البارامیترز تبعون الفوکال تراکت, بنستخدمهم بالسنتسس

$$\begin{bmatrix} R0 & R1 \\ R1 & R0 \end{bmatrix} \begin{bmatrix} a1 \\ a2 \end{bmatrix} = \begin{bmatrix} R1 \\ R2 \end{bmatrix}$$

24. Auditory hearing system

(a) Human ear consists of three main parts, give their names, and name two components in each part.

(b) What Auditory masking means? What are the two type of masking? Explain briefly what is the difference between two types of masking?

(c) What is meant by intra-band and inter-band masking? Support your answer by examples.

(d) what is meant by post-masking and pre-masking? Give examples?

(e) explain a simple experiment for estimating the masking effect on hearing threshold.