BIRZEIT UNIVERSITY

Electrical and Computer Engineering

Spoken Language Processing, Fall 2017

Dr. Abualsoud Hanani

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

1. What is a phoneme?

2. What is the difference between the phonemes /b/ and /p/?

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

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

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?

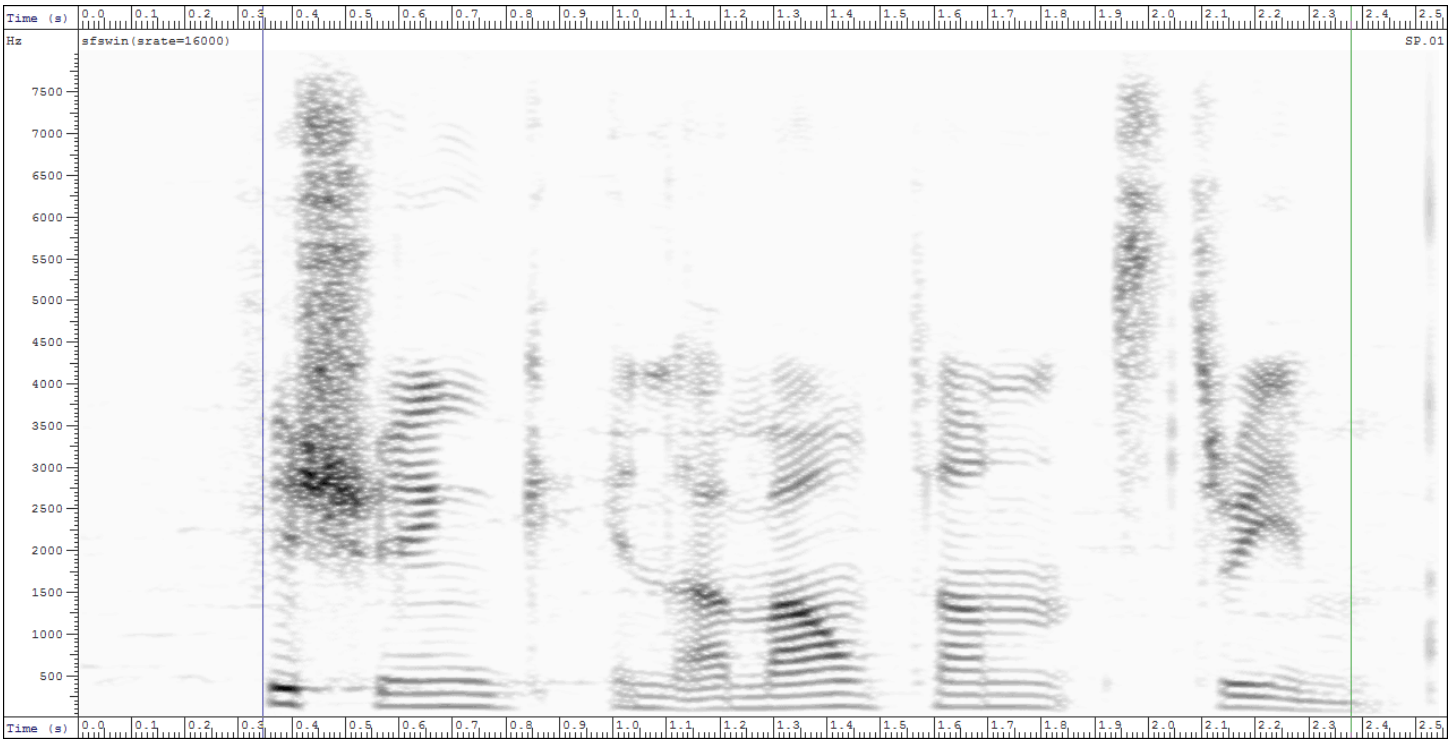
12. Draw a waveform and short-term spectrum of a voiced and unvoiced sound.

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

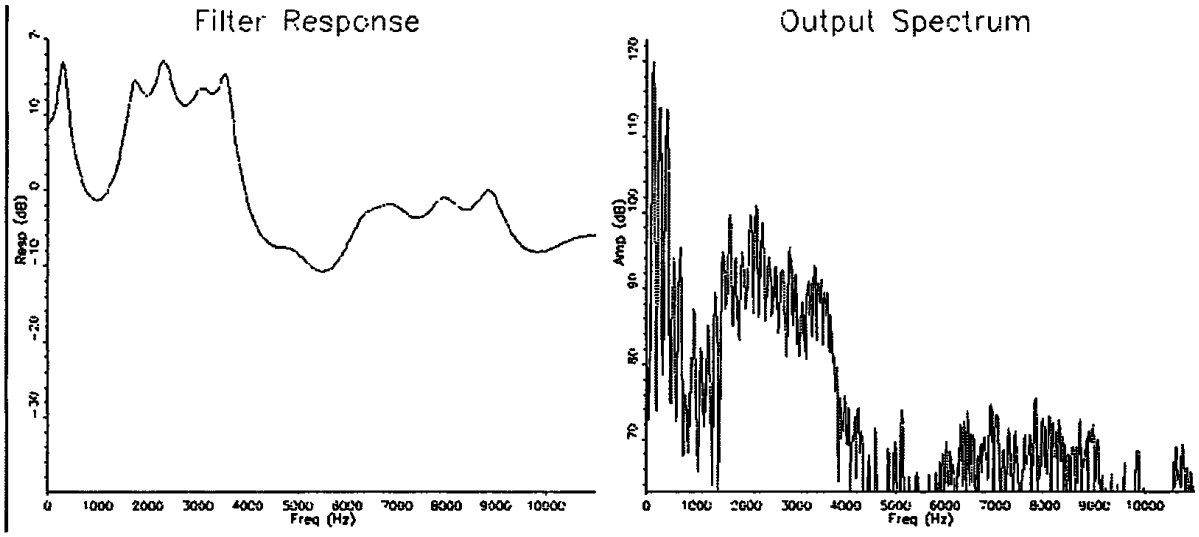
(ii) start time0.54s , end time 0.76s

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

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

(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.

(b) What is the short-term cepstrum?

(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.

(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?

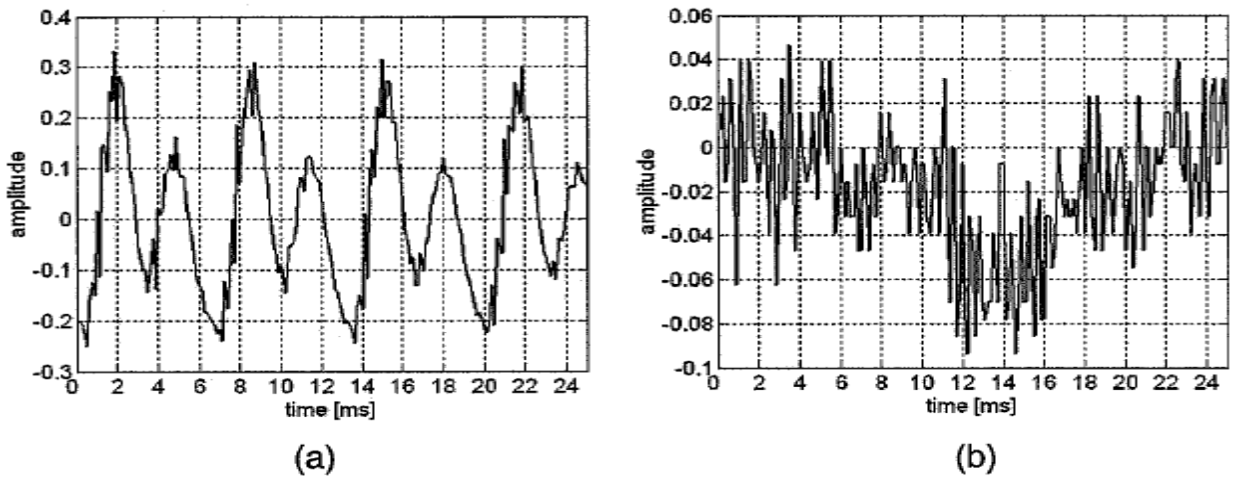
(e) What is meant by the term 'delta cepstrum', and what is the motivation for its use in automatic speech recognition.

18.

(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.

(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.



(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?

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

(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?

19.

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

(i) Give their names and give an example illustrating how each type of noise might occur in practice.

(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’.

20.

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

(b) Draw a block diagram of LPC processing?

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

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

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

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]

(ii) Zero-crossing count [4]

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

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

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]

23.

LPC:

(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, a3).

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.