Speech Signal Acquisition and Analysis

1. Suppose a stereo recording of speech of duration 3 minutes is made at CD quality. How much memory in terms of bits is needed to store that recording?
2. Suppose we are given a mix of telephone quality, microphone quality and CD quality speech data. How can we go about dealing with sampling frequency differences?
3. Suppose we would like to acquire and process multiparty interaction (two or more people interacting) speech data. What kind of speech acquisition system would we need?
4. Describe and illustrate, what is
    a. An analysis window: definition, typical length, for what etc.
    b. A power spectrum
    c. A spectrogram
    d. On power spectrum as well as on spectrogram, what are the typical properties of the speech signal that can be observed?
5. What are voiced and unvoiced sounds? Give a few examples. How can we differentiate between voiced and unvoiced sounds using autocorrelation? Two signals can be differentiated by measuring zero crossing rate (ZCR), number of times the signal cuts the time axis (+ to -/- to +). How can we differentiate between voiced and unvoiced sounds using ZCR?
6. Describe the basic mechanism of speech production
    a. speech signal can be seen as the result (convolution) of two phenomena, what are they?
    b. Define pitch frequency and formants.
    c. What kind of information does pitch frequency convey? What kind of spectrogram is best suited for observing pitch frequency?
    d. What kind of information does formants convey? What kind of spectrogram is best suited for observing formants?
    e. Illustrate and explain the differences in the auto-correlation function and the power spectrum for speech signals with pitch frequency,
    (i) F_0 = 100Hz, and
    (ii) F_0 = 200 Hz
7. What is linear prediction?
    a. Why is linear prediction modeling well suited for speech signal processing?
    b. What do the linear prediction coefficients model? How are they estimated (cost function)?
    c. What does the residual signal or the linear prediction error signal model? How is it obtained?
    d. How can we heuristically select the order of linear prediction? Suppose we want to model at least 4 formants. What is the “minimal” linear prediction order needed to achieve that? Suppose the linear prediction order is set as 3. How many formants can then be modeled?
    e. How can linear prediction analysis be applied to classify,
        i. different phonemes?
        ii. voiced and unvoiced speech sounds?
        iii. speaker gender (male and female classification)?

8. What is the goal of speech coding? What is the input to the coding system and what is the output? What is the input bit rate? What is the transmission bit rate? What is output bit rate? Calculate with an example.
9. In the Introduction lecture, we briefly dealt with distant speech (Slide 13), namely, clean, reverberant, and reverberant+noise. Mathematically formulate these three scenarios.
10. How does choice of microphone affect speech acquisition? Would change of microphone affect the performance of speech processing system? Justify your answer with a mathematical formulation. How can we deal with the effect of microphone?