

Speech signal processing (additional questions)

1. What is speech activity detection (SAD)?
 - a. What is the main purpose?
 - b. Describe a time domain method to detect speech activity (give the key steps).
2. What are voiced and unvoiced sounds? Give an example. How can we differentiate between voiced and unvoiced sounds through time domain processing?
3. What is autocorrelation? How can it be used to “automatically” estimate pitch frequency (describe the key steps)?
4. We know that the speech signal is produced by excitation of the vocal tract system by the vibration of the vocal cords.
 - a. What are formants? To which component of the speech production process are they related to?
 - b. What is pitch frequency? To which component of the speech production process it relates to?
 - c. In the short-term spectrum of the speech signal (resulting after Fourier transform), how does the formant information and the pitch frequency information manifest? Illustrate it through a figure for the case where the pitch period is T seconds.
5. 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 the residual signal 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?
 - e. How can we estimate the formant frequencies using linear prediction?
 - f. How can we estimate pitch frequency using linear prediction?
 - g. 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)?
6. Linear prediction modeling is also known as analysis-synthesis modeling.
 - a. During analysis, what is given (or known) and what is estimated?
 - b. During synthesis, what is given (or known) and what is estimated or synthesized?
 - c. How is linear prediction analysis and synthesis applied in telephony to transmit speech signal at a reduced bit rate?
7. How can we deconvolve or separate excitation source and vocal tract system information through
 - a. time domain processing only?
 - b. frequency domain processing?