

Speech Signal Assignemnt-3

- Srihari Bandarupalli

- 2021112006

Question 1

(a) UBM (Universal Background Model)

A UBM is a statistical model of the speech spectral features of a large corpus of speakers. It is used as a starting point for training speaker-specific models, or for performing speaker identification and verification tasks. It represents the statistical characteristics of all possible speakers or background noise in a given acoustic environment. In speaker recognition, it is used to model the variability of speakers' voices and is often employed in a GMM-UBM (Gaussian Mixture Model - Universal Background Model) framework to classify speakers.

(b) LP Residue (Linear Prediction Residue)

In speech processing, LP residue refers to the signal left after applying linear prediction analysis to speech. Linear prediction is a technique used to model the spectral envelope of a speech signal. The LP residue contains the fine-grained details and noise that are not captured by the linear prediction model, and it can be important for certain speech processing tasks like speech coding or enhancement.

(c) Spectral Subtraction

Spectral subtraction is a noise reduction technique used in audio processing to enhance the quality of a signal by reducing the background noise. It works by estimating the spectral profile of the noise and then subtracting it from the noisy signal in the frequency domain. This method is commonly used in applications like speech enhancement to improve the intelligibility of speech in noisy environments. It can introduce artifacts into the signal if the noise spectrum is not estimated accurately.

(d) Mel Filter Bank

A mel filter bank is a set of bandpass filters that are spaced according to the Mel scale. The Mel scale is a non-linear frequency scale that is designed to mimic the human perception of pitch. Mel filter banks are often used in speech recognition and speaker recognition systems. Mel filter banks help capture the relevant acoustic information in a more perceptually meaningful way by emphasizing certain frequency regions.

(e) Cepstrum

Cepstrum is a mathematical transformation commonly used in speech and audio signal processing. It involves taking the inverse Fourier transform of the logarithm of the magnitude of the Fourier transform of a signal. Cepstral analysis is useful for separating the characteristics of the vocal tract (source) and the vocal cord excitation (source-filter model) in speech signals. The cepstrum is often used in speech recognition and speaker recognition systems because it is less sensitive to noise and channel distortion than the magnitude spectrum.