

Speech Signal Processing: An Overview

S. R. M. Prasanna



Department of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati
December, 2012

Organization

- Introduction
- Sampling frequency and bit resolution
- Non-stationary nature
- Short term processing
- STFT and Spectrogram
- Energy and Pitch
- Cepstral analysis
- Linear prediction analysis
- Speech processing tasks
- Summary

Introduction to Speech Signal Processing

- **Speech:** Fundamental and effortless mode of communication among humans.
- **Speech communication:** Talker, listener and channel
- **Speech Production Process:** Message formulation, language coding, neuro-muscular commands, movement of speech production organs, acoustic pressure variations
- **Speech Perception Process:** acoustic pressure variations, movement of speech perception organs, neuro-muscular commands, message comprehension

What is present in Speech Signal?

- Message
- Speaker
- Emotion
- Language
- Dialect
- Sensor
- Channel
- How to analyze, extract and model these information

Sampling Frequency

- Acoustic pressure variations to electrical signal using microphone
- Digitization for storage, analysis and processing on a digital machine
- **Sampling, quantization and Encoding**
- **Sampling Theorem:** The sampling frequency should be greater than or equal to twice the maximum frequency
- Audio frequency range: 20 Hz to 20 kHz
- Speech components up to 14 kHz, but can consider the whole audio range.
- Min. Sampling frequency recommended is 40 kHz
- Including some guard band it is 48 kHz

Bit Rate

- Number of bits / sample
- Bit resolution
- Number of quantization levels
- Minimum 16 bits is recommended

Non-Stationary Nature

- Signal, system, and signals and systems
- Stationary vs non-stationary signal
- Significance of non-stationary nature of speech

Short Term Processing

- Need for short term processing
- Approach for short term processing
- Frame size and frame shift
- Short term time domain processing
- Short term frequency domain processing

Short Term Domain Parameters

- Short term energy
- Short term zero crossing rate
- Short term autocorrelation

Short Term Frequency Domain Parameters

- Short term Fourier transform
- DTFT, STFT, DFT, FFT
- Spectrogram
- Wideband spectrogram
- Narrowband spectrogram

Cepstral Analysis of Speech

- Separation of source and system components in cepstral domain
- Feature extraction stage of automatic speech processing systems
- Also in estimation of pitch
- Cepstrum pitch determination

Linear Prediction Analysis of Speech

- Separation of source and system components in time domain
- Filter coefficients for speech coding
- Pitch estimation by SIFT

Automatic Speech Processing Tasks

- Speech recognition
- Speaker recognition
- Speech synthesis
- Language identification