

Audio/Speech Data Characteristics



Temporal Structure

Time-dependent sequential information



Frequency Domain

Sound is combination of different frequencies



Sampling Rate

Typically 16kHz-44.1kHz for speech and music



Variable Length

Utterances and clips have different durations



Background Noise

Acoustic environment affects quality



Speaker Variability

Accent, pitch, speed vary across individuals



Required Time-Frequency Analysis

Spectrograms

MFCCs

Mel-Frequency Analysis