Speech Signal & Spectrum

Librosa(0.8.1)

- python package for music and audio analysis.

https://librosa.org/doc/0.8.1/index.html

- librosa.load

librosa.load(path, sr=None)

Load an audio file as a floating point time series.

sr: target sampling rate. If 'None' uses the native sampling rate.

- librosa.display.waveplot

librosa.display.waveplot(y, sr=*16000*, x_axis=*'time'*)

Plot the amplitude envelope of a waveform.

STFT(Short Time Fourier Transform)

➢ Librosa.stft

- The **STFT** represents a signal in the time-frequency domain by computing discrete Fourier transforms (DFT) over short overlapping windows.

- **librosa.stft**(y, n_fft=*2048*, hop_length=*None*, win_length=*None*, window= *'hann'*, center=*True*, dtype=*None*, pad_mode= *'reflect'*)

- **Parameters:**

y: np.ndarray [shape=(n,)], real-valued
input signal
n_fft: int > 0 [scalar]
length of the windowed signal. The number of rows in the **STFT** matrix D is (1 + n_fft/2). The default value, n_fft=2048
hop_length: int > 0 [scalar]
number of audio samples between adjacent **STFT** columns. Default value, win_length // 4.
win_length: int <= n_fft [scalar]
Each frame of audio is windowed by window of length win_length and then padded with zeros to match n_fft.
defaults to win_length = n_fft.

Returns:
Dnp.ndarray [shape=(1 + n_fft/2, n_frames), dtype=dtype]
Complex-valued matrix of short-term Fourier transform coefficients.

Windowing(Framing)

➢ Frame

- 사람의 시간영역에서 입력된 소리 신호는 지속적으로 변화함.
- 일반적으로 프레임의 길이는 20~40ms 정도로 함.
- Sampling Rate에 따라 샘플의 수는 달라지겠지만, 44.1KHz 기준으로 프레임 당 약 천 개의 데이터 포함.