

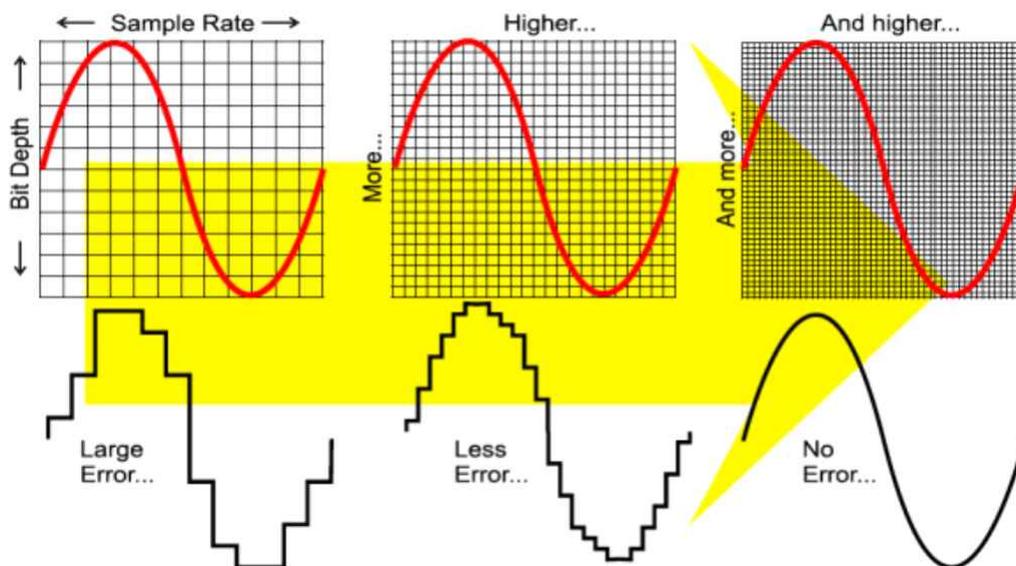
Chap. 8 딥러닝 영상처리를 이용한 음성 처리와 인식

▶ 수업 내용

- Time-Amplitude 차원의 음성 파일을 2차원 이미지로 변환하여 딥러닝을 이용하여 인식하는 방법
- rhksfus YouTube 영상 시청

1. 소리 및 음성 파일 기초

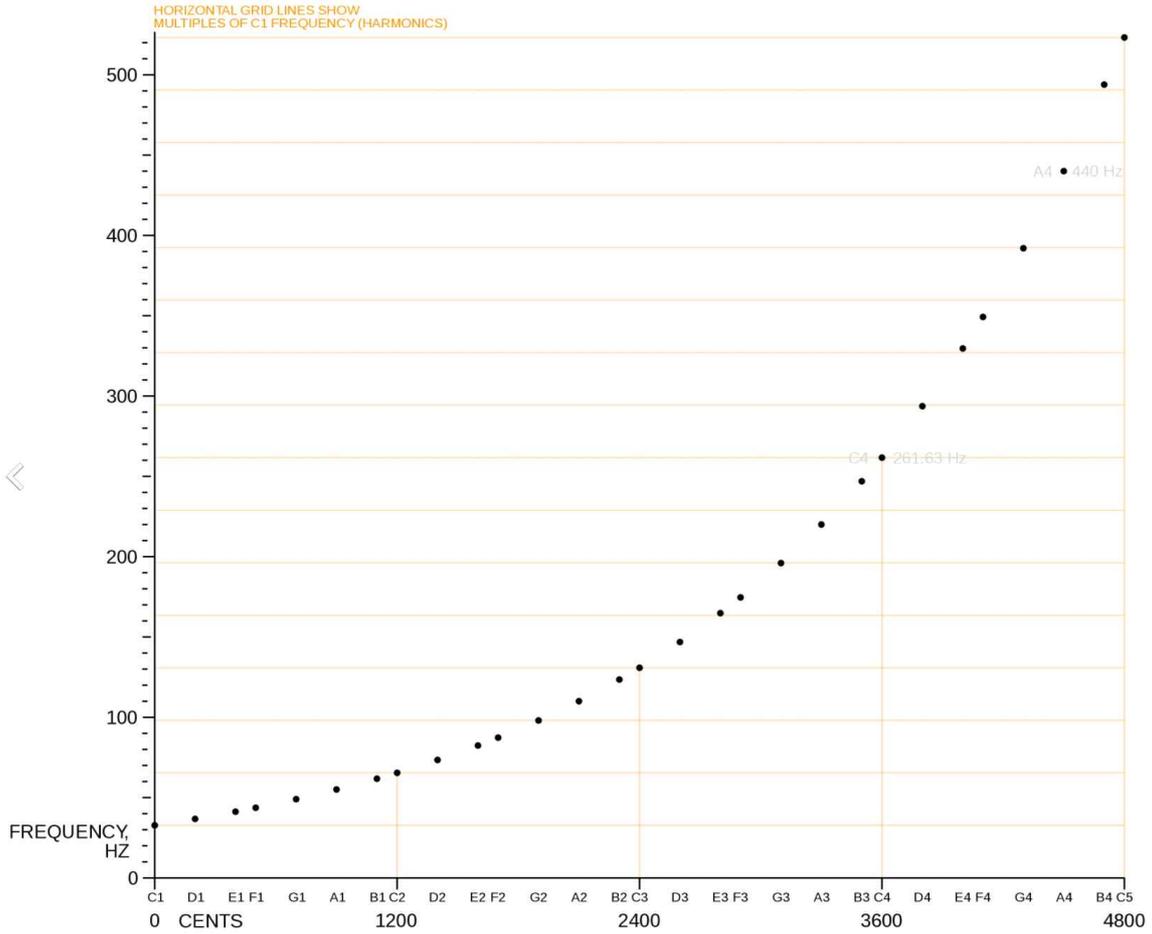
- ◎ 소리 : 공기를 매질로 하는 파동 에너지.
파동의 진동수가 작으면 낮은 에너지의 저음이 되고, 진동수가 크면 많은 에너지를 가진 고음이 된다.
- ◎ 소리의 녹음과정 : 파동에너지가 진동판을 떨리게 하고, 진동판의 진동폭을 전기적(아날로그) 신호로 변환해서 + - 로 기록, 샘플링을 통하여 디지털 신호로 변환
- ◎ Wav 파일 생성 : 소리 파동의 높이 값(Amplitude)을 일정한 시간 간격(sampling rate, time)마다 기록한 실수값의 배열. 이런 저장방식을 PCM(Pulse Code Modulation) 이라 함
- ◎ Sampling Rate : 1초당 디지털화 하는 횟수 (예: 44.1KHz)



(<https://www.izotope.com/en/learn/digital-audio-basics-sample-rate-and-bit-depth.html>)

◎ **Pitch** : 음의 높낮이

아래 그림처럼 사람의 소리 인식은 저주파에 더 민감하고 고주파로 갈수록 둔감하다 (Log Scale)

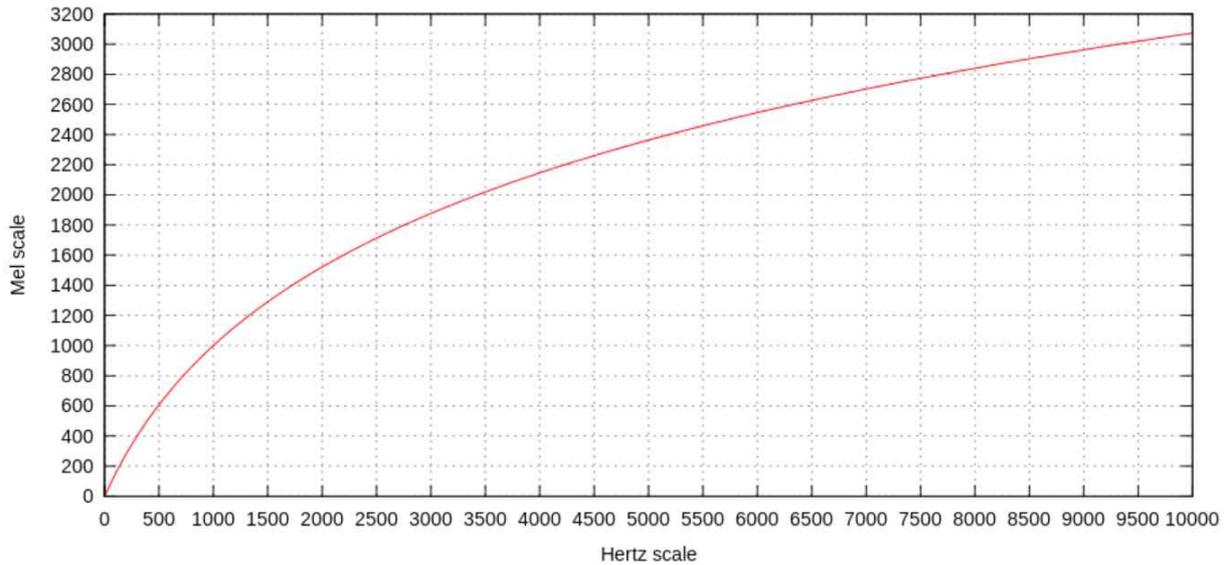


([https://en.wikipedia.org/wiki/Pitch_\(music\)](https://en.wikipedia.org/wiki/Pitch_(music)))

◎ **Decibel** : 소리의 높이에 대한 사람의 민감도가 Log Scale 이므로 이를 편하게 나타낸 방법

$$dB = 10 * \log_{10}(\text{에너지})$$

◎ Melody-Scale, Mel-Scale : 주파수(Hz) 대 Pitch Mel-scale



(https://en.wikipedia.org/wiki/Mel_scale)

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

2. YouTube 강의 시청

https://www.youtube.com/watch?v=m3Xbqflij_Y

Audio File, Fourier 변환, MFCC 설명

https://www.youtube.com/watch?v=Oa_d-zaUti8&t=1251s

파이썬으로 FFT, STFT 구현, MFCC 추출 방법, Librosa 사용

<https://www.youtube.com/watch?v=szyGiObZymo>

음악 장르 분류

---- 다음 강의 (Chap. 9)----

3. 음성 변환 실습 (voicerecog.py 설명)

4. 딥러닝 영상처리를 이용한 음성인식