

1. 발음 기호와 문자 표현.

- phoneme : 음소. 가장 작은 소리의 단위.
- grapheme : 자소. 가장 작은 문자의 단위 (Character).

∴ phoneme \neq grapheme.

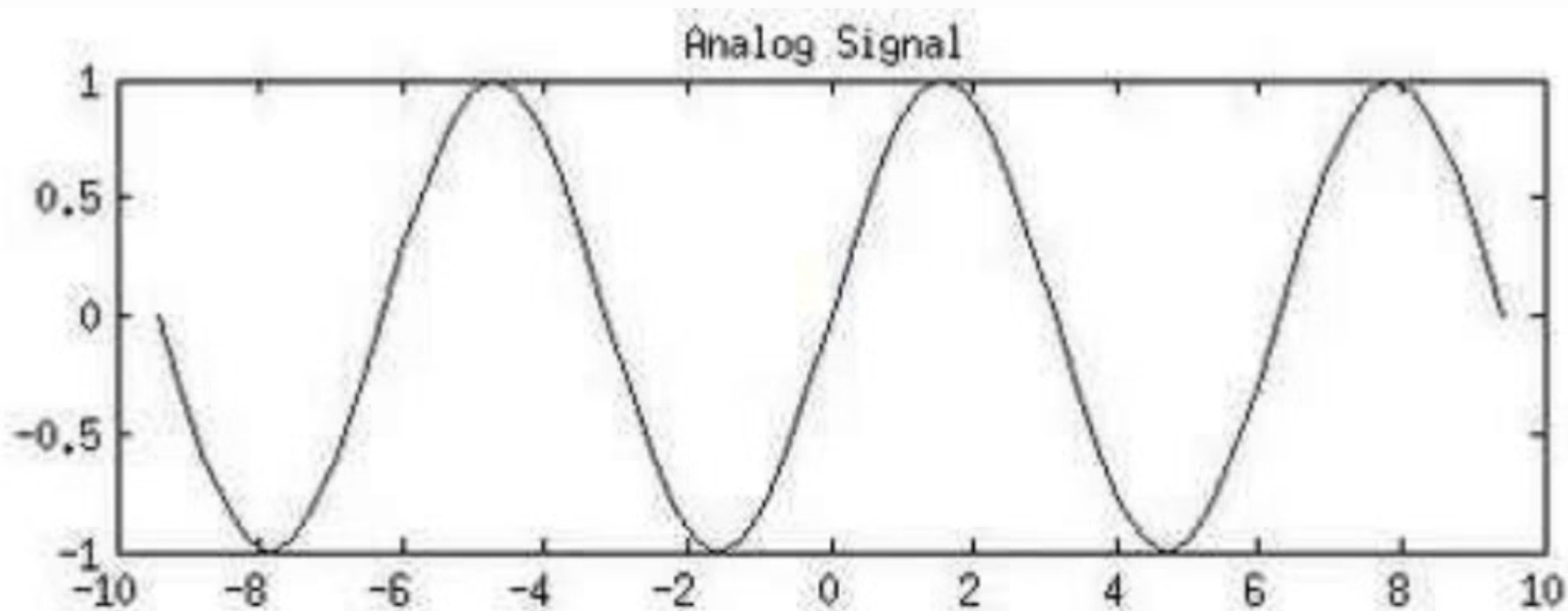
2. 음성파일의 기본 개념.

- 마이크의 측화 과정 : 소리 \rightarrow 마이크의 떨림판.
 \rightarrow +, -로 떨림판의 진동폭을 기록.

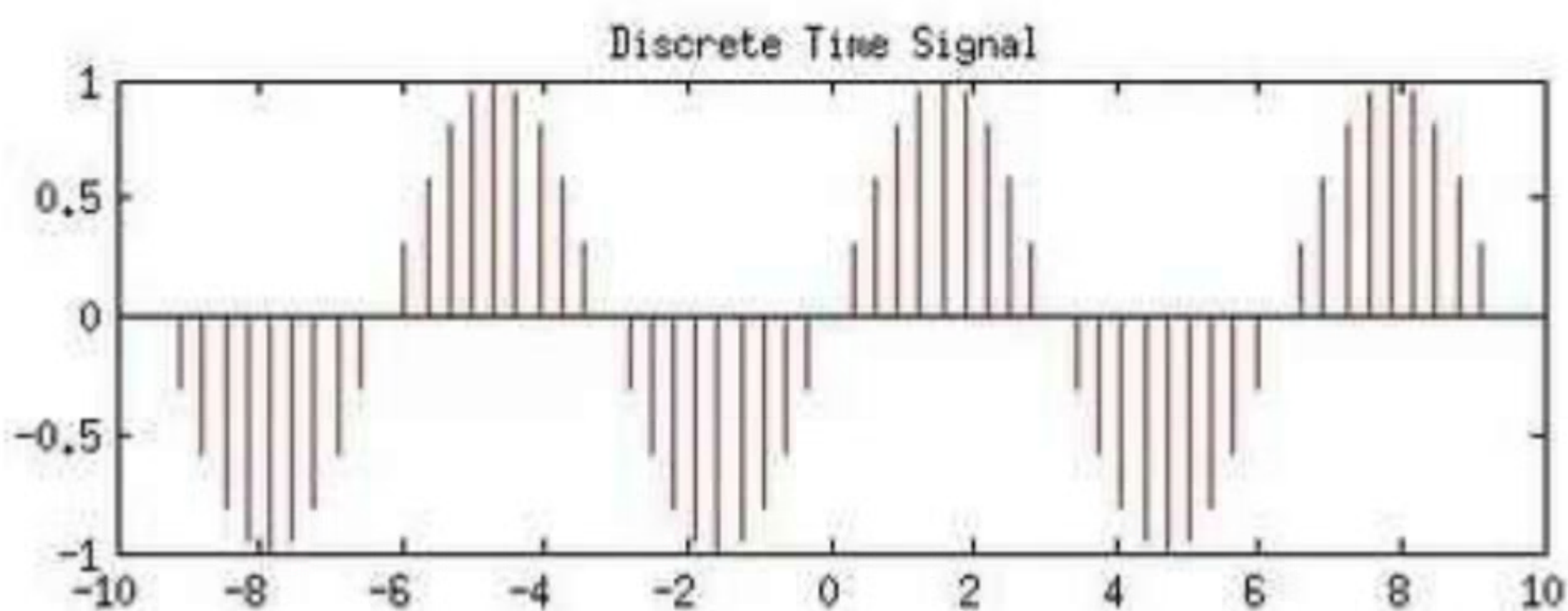
< analog 값 >

- wav 파일의 이해.

↳ analog 신호를 일정 시간 단위 (sampling rate)로 기록.
PCM. (Pulse - Code Modulation)



← analog signal



← pcm signal

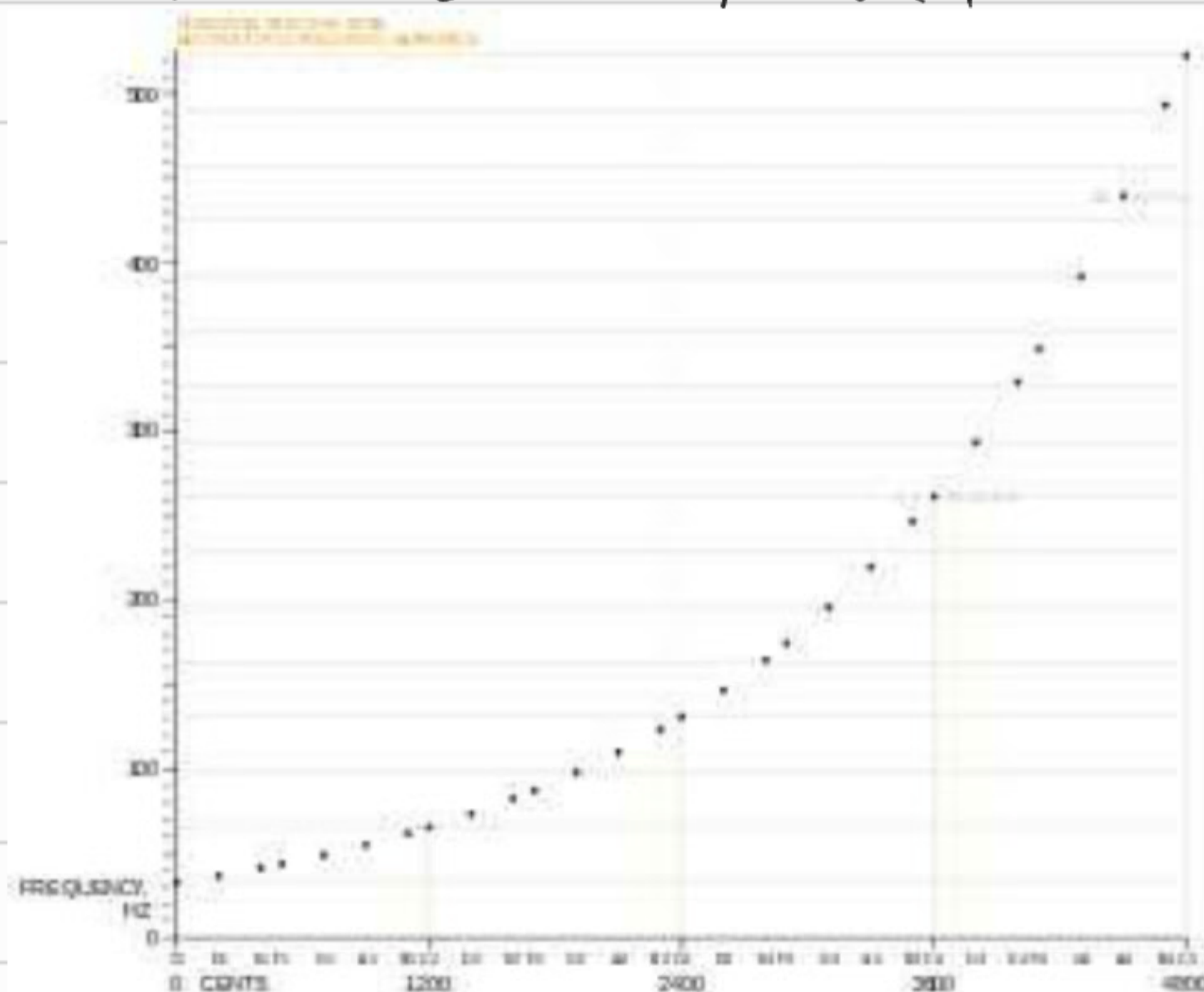
- Mono vs Stereo : Left Mono + Right Mono \Rightarrow stereo.
- Sampling rate : 초에 몇번 기록할지.
- Bit depth : y축을 기준으로 해상도 설정.
- Bit Rate : Sampling rate \times Bit depth. 초당 비트 전송량.

- Acoustic Model : 오디오 \rightarrow phoneme.
- Phoneme Dictionary : word \rightarrow phoneme (발음 기호 사전)
- G2P model (Grapheme to. Phoneme) : 발음 예측.

\Rightarrow undefined word에 대해서도 처리가능.

- Pitch. : 음의 높낮이에 따른 사람의 음성 인지.

저주파 \rightarrow 인지 \uparrow / 고주파 \rightarrow 인지 \downarrow



\rightarrow 사람이 인지하는 음의 높낮이는 시간과 exponential 한 관계.

- log - scale : 값에 log를 취해 normalize 하는 방식.
(i.e. mel - scale)

* mel - scale : $\left\{ \begin{array}{l} \text{① spectrogram 이미지에서 Amplitude 값에 대한 log - scale.} \\ \text{② spectrogram의 y축, frequency에 대해서 log - scale.} \end{array} \right.$

- mel - scale (melody - scale) : pitch에서 음을 인지하는 기준.

threshold를 반영한 변환 값.

\rightarrow exponential 값을 Linear 하게 변환함.

$$\text{Mel}(f) = 2595 \left(1 + \frac{f}{700} \right)$$

- Mel - filter bank.

: mel - scale에서 Linear 하게 N_{Hz} 나누어 주면.