Brain-Audio Internal Education for Consultant

Junhyeok Lee

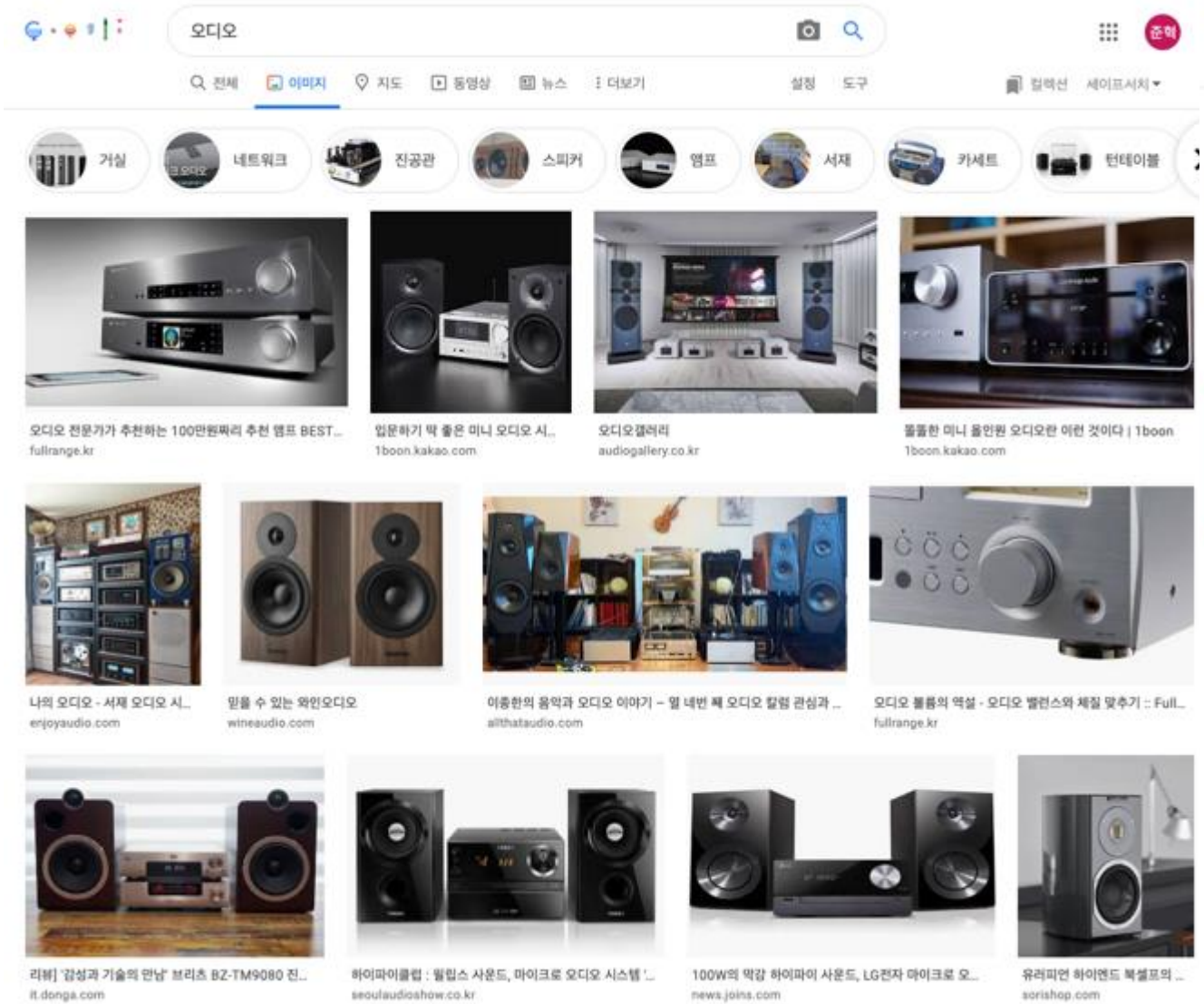
Brain-Audio

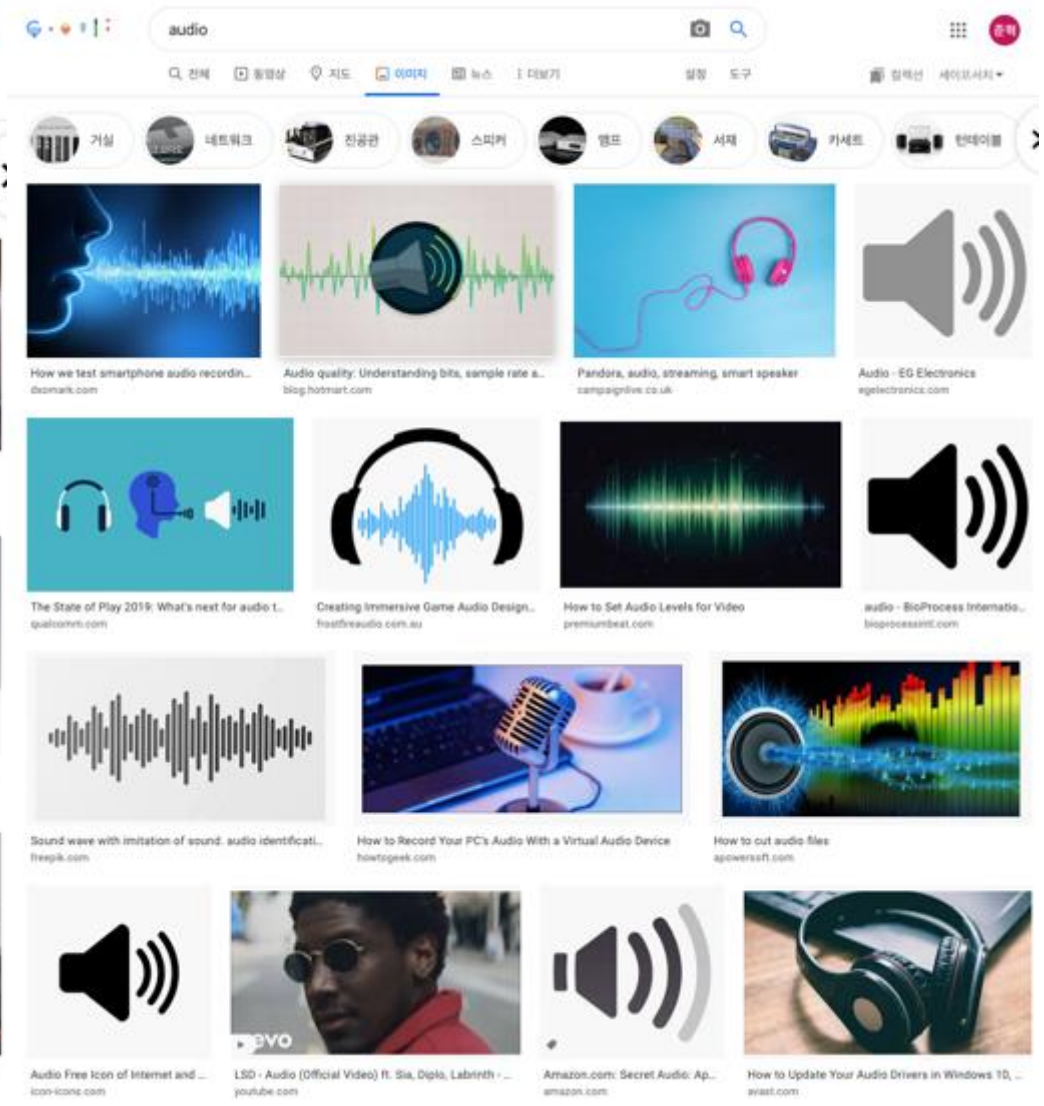
MINDs Lab

Objectives

- Understanding domain knowledge
- Reminding keywords
- Increase Googleability

Audio


오디오
전체
이미지
지도
동영상
뉴스
더보기
설정
도구
컬렉션
세이프서치
거실
네트워크
진공관
스피커
앰프
서재
카세트
턴테이블
오디오 전문가가 추천하는 100만원짜리 추천 앰프 BEST...
fullrange.kr
입문하기 딱 좋은 미니 오디오 시...
1boon.kakao.com
오디오갤러리
audiogallery.co.kr
똘똘한 미니 올인원 오디오란 이런 것이다 | 1boon
1boon.kakao.com
나의 오디오 - 서재 오디오 시...
enjoyaudio.com
믿을 수 있는 와인오디오
wineaudio.com
이종한의 음악과 오디오 이야기 – 열 네번 째 오디오 칼럼 관심과 ...
allthataudio.com
오디오 볼륨의 역설 - 오디오 밸런스와 체질 맞추기 :: Full...
fullrange.kr
리뷰] '감성과 기술의 만남' 브리츠 BZ-TM9080 진...
it.donga.com
하이파이클럽 : 윌립스 사운드, 마이크로 오디오 시스템 '...
seoulaudioshow.co.kr
100W의 막강 하이파이 사운드, LG전자 마이크로 오...
news.joins.com
유러피언 하이엔드 북셀프의 ...
sorishop.com


audio
전체
동영상
지도
이미지
뉴스
더보기
설정
도구
컬렉션
세이프서치
거실
네트워크
진공관
스피커
앰프
서재
카세트
턴테이블
How we test smartphone audio recordin...
dxomark.com
Audio quality: Understanding bits, sample rate a...
blog.hotmart.com
Pandora, audio, streaming, smart speaker
campaignlive.co.uk
Audio - EG Electronics
egelectronics.com
The State of Play 2019: What's next for audio t...
qualcomm.com
Creating Immersive Game Audio Design...
frostfireaudio.com.au
How to Set Audio Levels for Video
premiumbeat.com
audio - BioProcess Internatio...
bioprocessintl.com
Sound wave with imitation of sound. audio identificati...
freepik.com
How to Record Your PC's Audio With a Virtual Audio Device
howtogeek.com
How to cut audio files
apowersoft.com
Audio Free Icon of Internet and ...
icon-icons.com
LSD - Audio (Official Video) ft. Sia, Diplo, Labrinth - ...
youtube.com
Amazon.com: Secret Audio: Ap...
amazon.com
How to Update Your Audio Drivers in Windows 10, ...
avast.com

Other tasks

- Vision

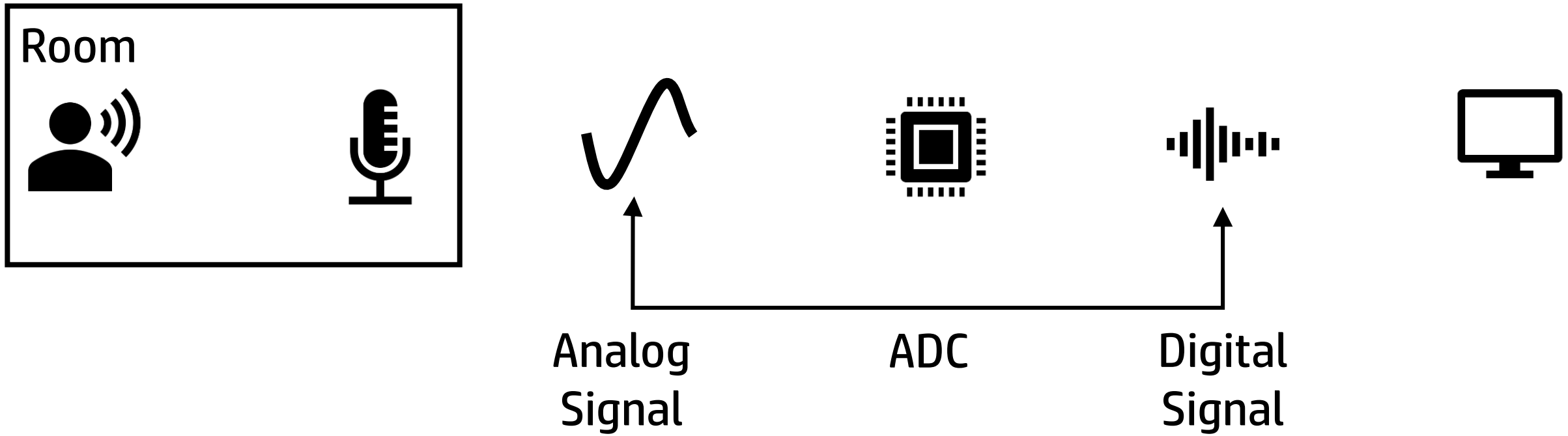
- NLP

Today

Data

Model

Audio Signal

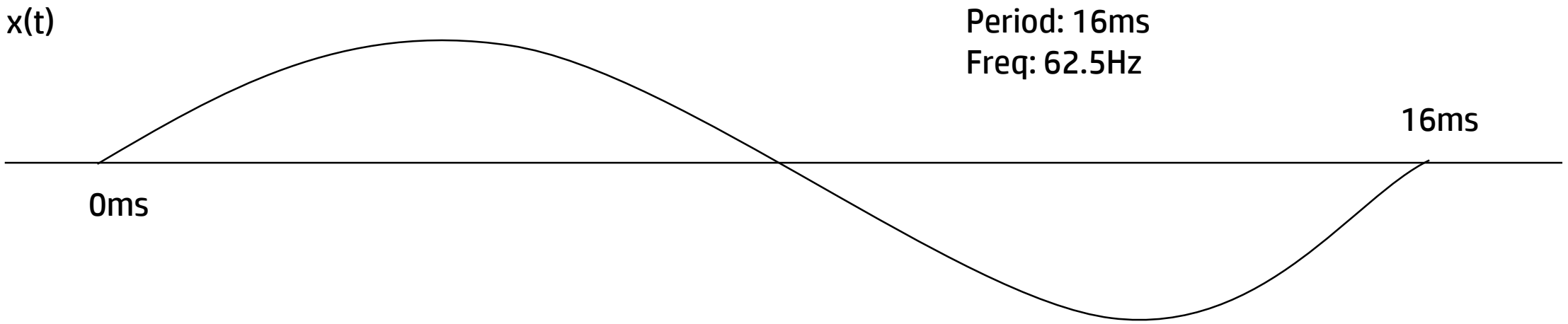


소리: 공기를 매질로 하는 압력의 진동
(공기가 아닌 매질/상온/상압/이 아닌경우 전달이 다르게 됨 e.g. 헬륨 보이스)

방마다 반사가 다름
마이크 특성을 탈수도?

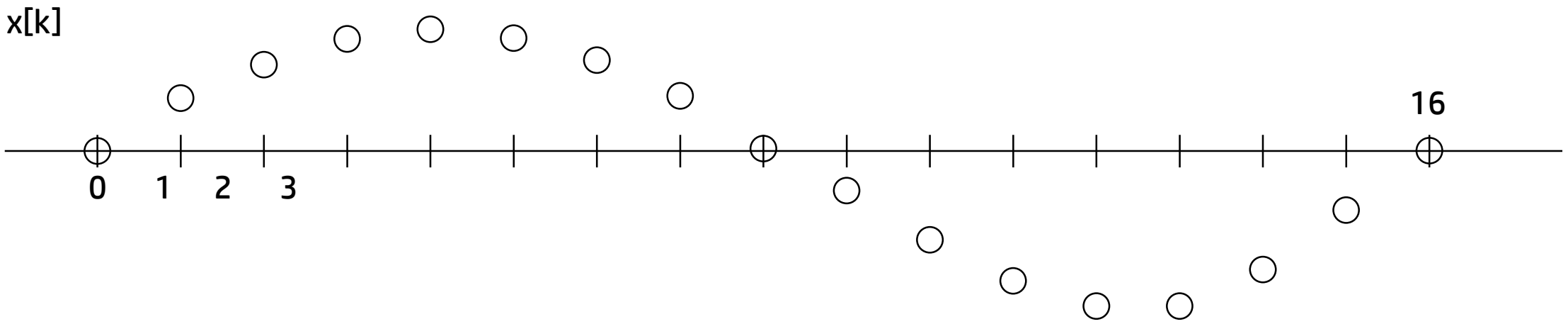
Analog(시간, 값이 모두 연속) Signal
↓
ADC(Analog-Digital Converter)
↓
Digital(시간, 값이 모두 불연속) Signal

Analog → Digital (Sampling)



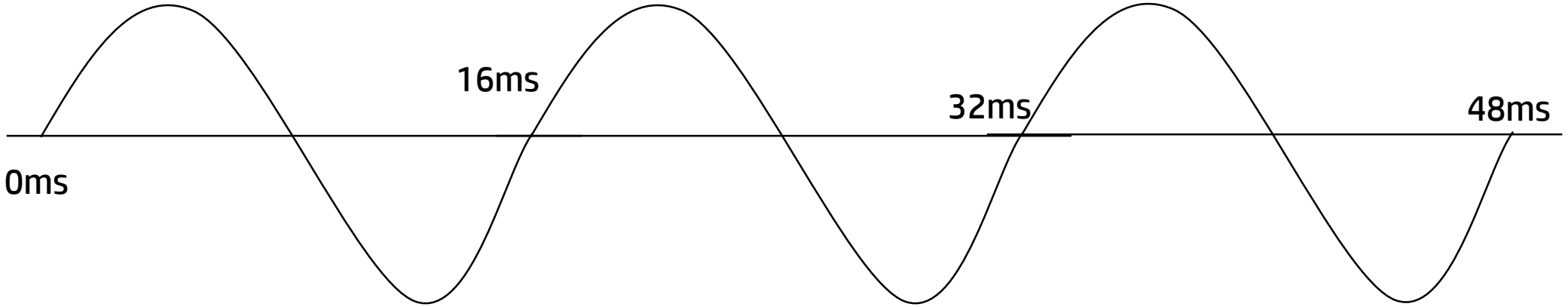
Sampling by sampling rate 1000Hz = sampling period 1ms

Frequency = 1/Period

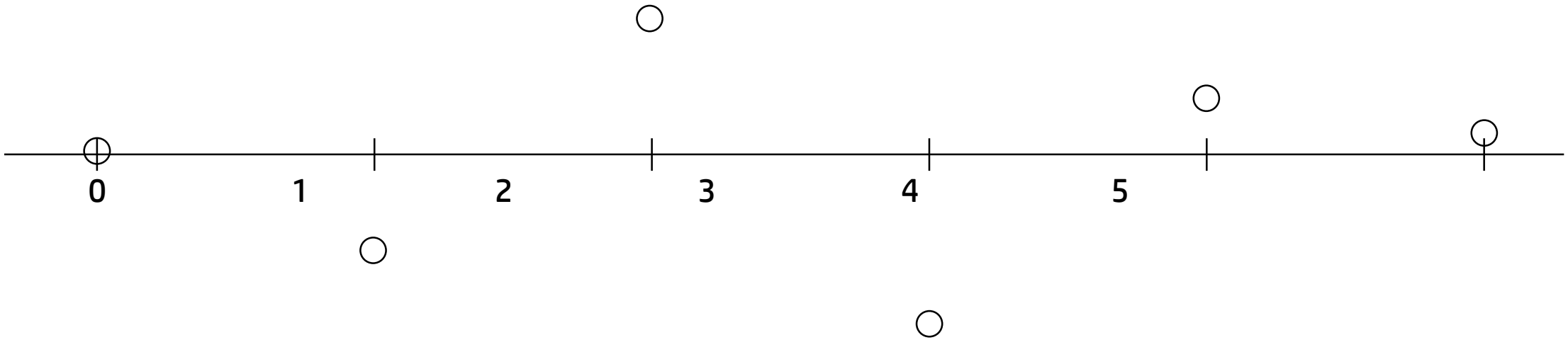


Analog → Digital (Sampling)

- Sampling rate is important!



Sampling by sampling rate 100Hz = sampling period 10ms



44.1kHz

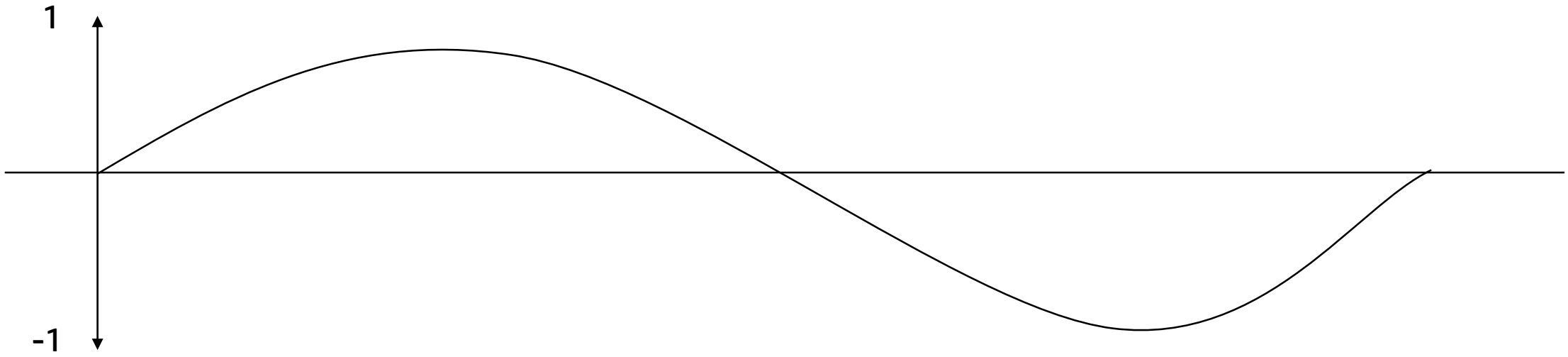
2kHz

400Hz

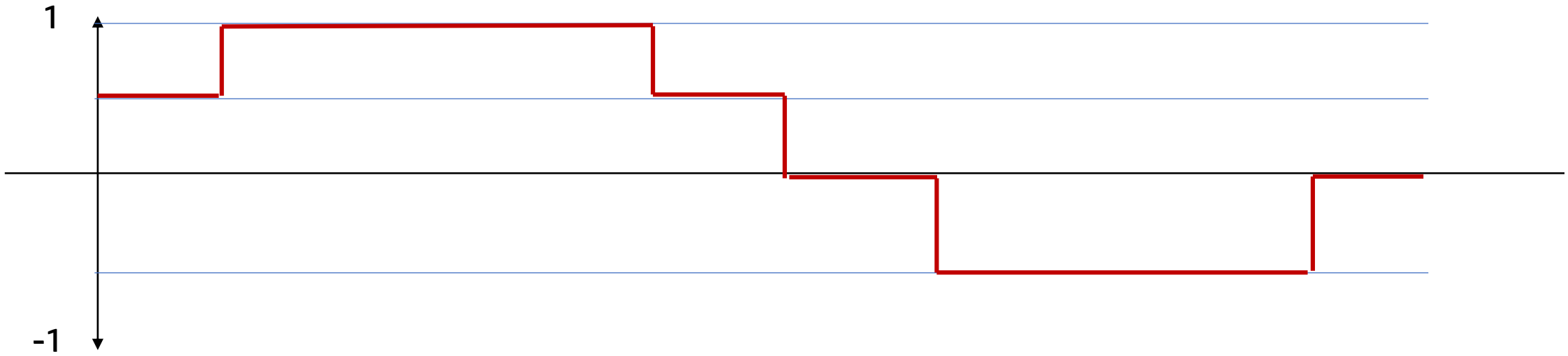
Sampling rate N Hz → 0~N/2 Hz에 해당하는 주파수 정보만 존재 (Nyquist frequency)
사람의 가청 주파수: ~20kHz → 44.1kHz/48kHz 가 효율적이면서 고음질 ~~(이거 보다 높다? 사기)~~

해상도(사진)

Analog → Digital (Value Quantizing)

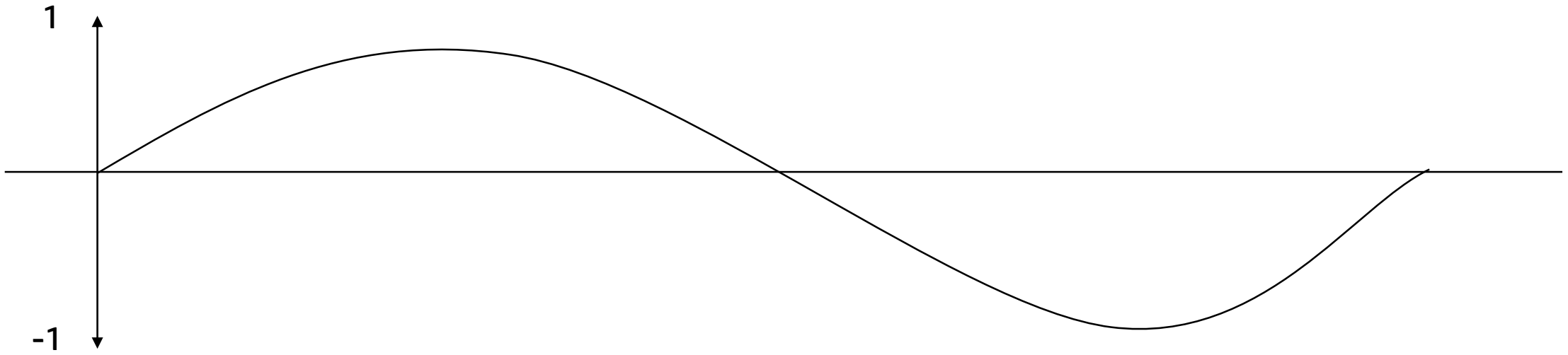


Quantizing with bit depth 2 = 4(2^2) levels

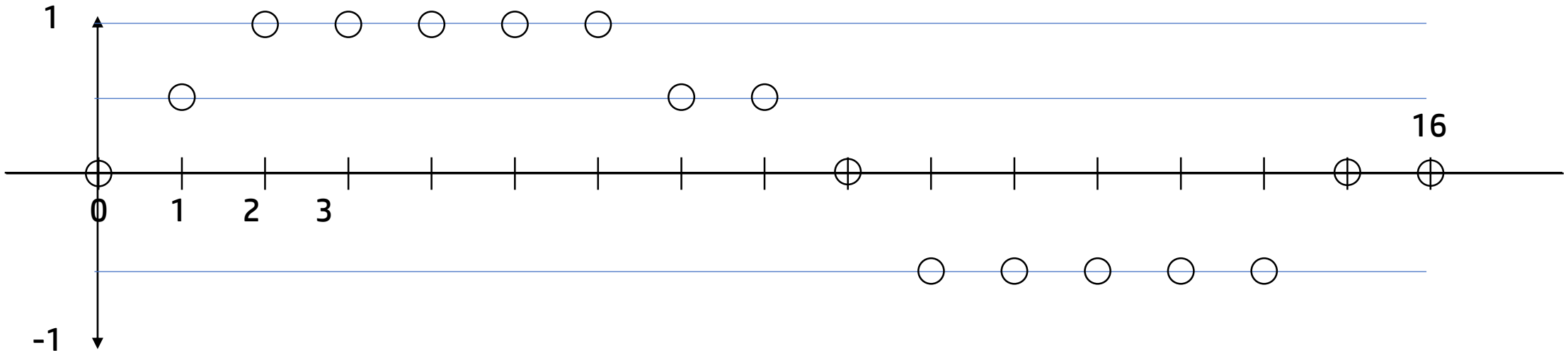


Analog → Digital (Sampling & Quantizing)

Analog Signal

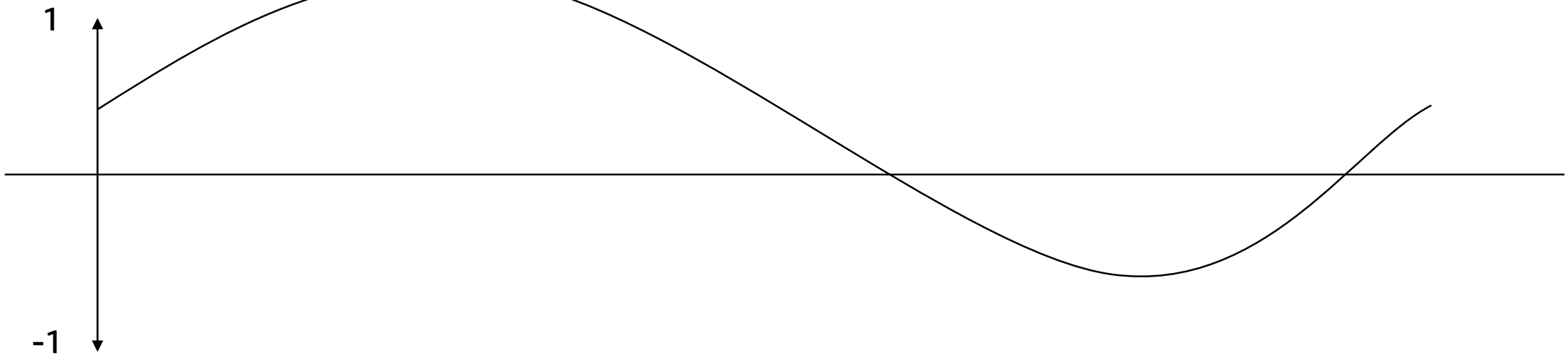


Digital Signal w/ sr: 1kHz, bd: 2

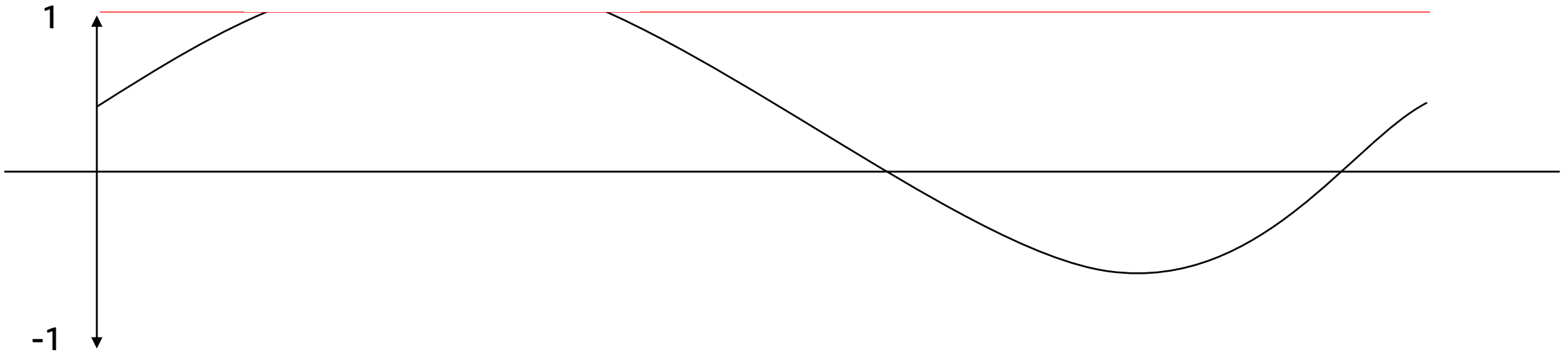


Clipping

Analog Signal



Clipped signal



Clipping 방지를 위해선 적당한 볼륨으로 녹음/저장 필요

퀀타이제이션(사진)

채도 400% 에서의 artifact
QD를 낮추면 비슷한 현상이 생김

Quantize depth: 1bit

클리핑(사진)

화이트밸런스 조정

턱 라인 실종→ 정보의 손실

Analog → Digital

- Analog signal → Sampling & Quantizing → Digital signal

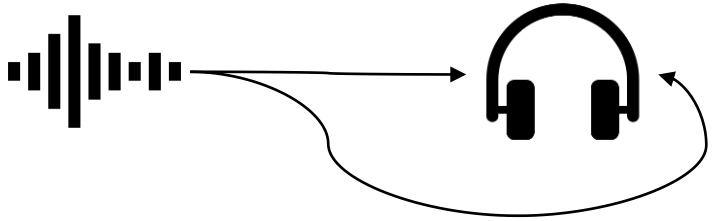
- Sampling rate: 8kHz, 16kHz, 44.1kHz, 48kHz
  - Sampling rate N Hz → Contain 0~N/2 Hz frequency

- Bit depth: 16bit, ~~24bit~~, 32bit
  - 16bit Int, 32bit Int
  - 16bit float, 32bit float
  - MP3 compress this!

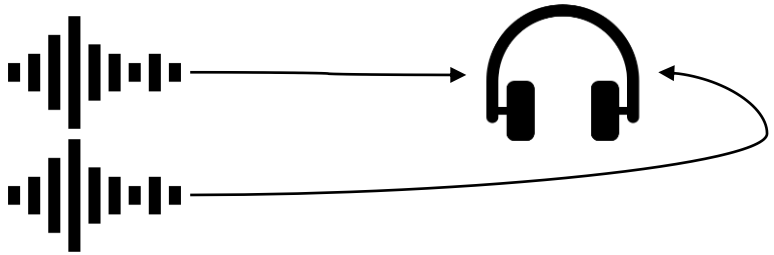
- 51s x 48 kHz = 2448000 samples

Channel

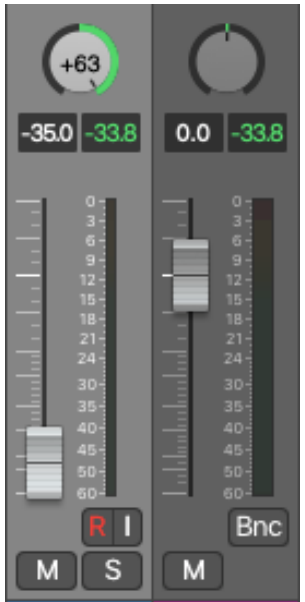
- Mono
  - Only one channel
  - Most of speech data



- Stereo
  - Two channel [Left, Right]
  - Music
  - Game
  - Phone call [spk1, spk2]
  - Some time mono data saved as stereo



- Or More



Recording

- 샘플링 레이트는 무조건 높거나 목표랑 같은게 좋다
  - 높은건 낮출 수 있음

- 샘플당 비트도 무조건 많을수록 좋다
  - 샘플링 레이트 보다는 덜 중요
  - 24bit 는 처리하기 힘드니 사용 X

- 클리핑이 없다는 가정하에 소리는 클수록 좋다

- 리버브가 없는 무향실에서 녹음하는게 좋다
  - 특정 방에서 실행되는게 목표인 경우 그 방에서 녹음해도 좋다

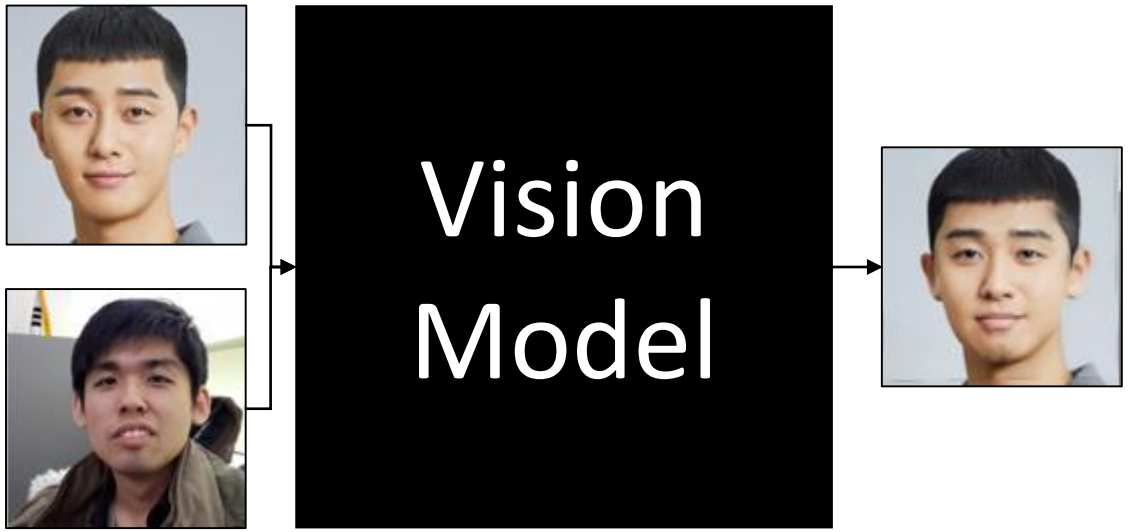
- 좋은 마이크 → 좋은 음질

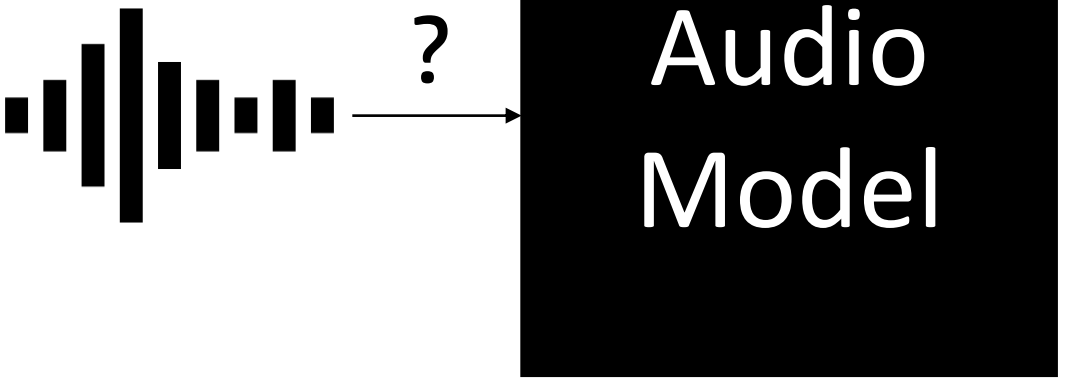
마이크 종류

- 컨덴서 마이크
  - 싸구려 마이크도 컨덴서일 수 있다
- +
  - 더 민감함
  - 더 넓은 범위의 주파수에 반응함
  - 음질이 좋다
- -
  - 추가 전원 필요
  - 비쌈
  - 자가 노이즈가 있다
  - 약하다(물리)

- 다이내믹 마이크
  - a.k.a 노래방 마이크
- +
  - 충격에 강하다
  - 싸다
  - 외부 전원이 필요 없다
  - 노이즈에 민감하지 않다
- -
  - 민감하지 않다
  - 크다
  - 반응이 느리다
  - 음질이 안좋다

Input?





Feature Extraction

- Raw Audio

- Fourier Transform
  - STFT
    - STFT Magnitude
  - Mel Spectrogram
  - ~~MFCC~~
  - CQT

- ~~Wavelet Transform~~

Fourier Transform

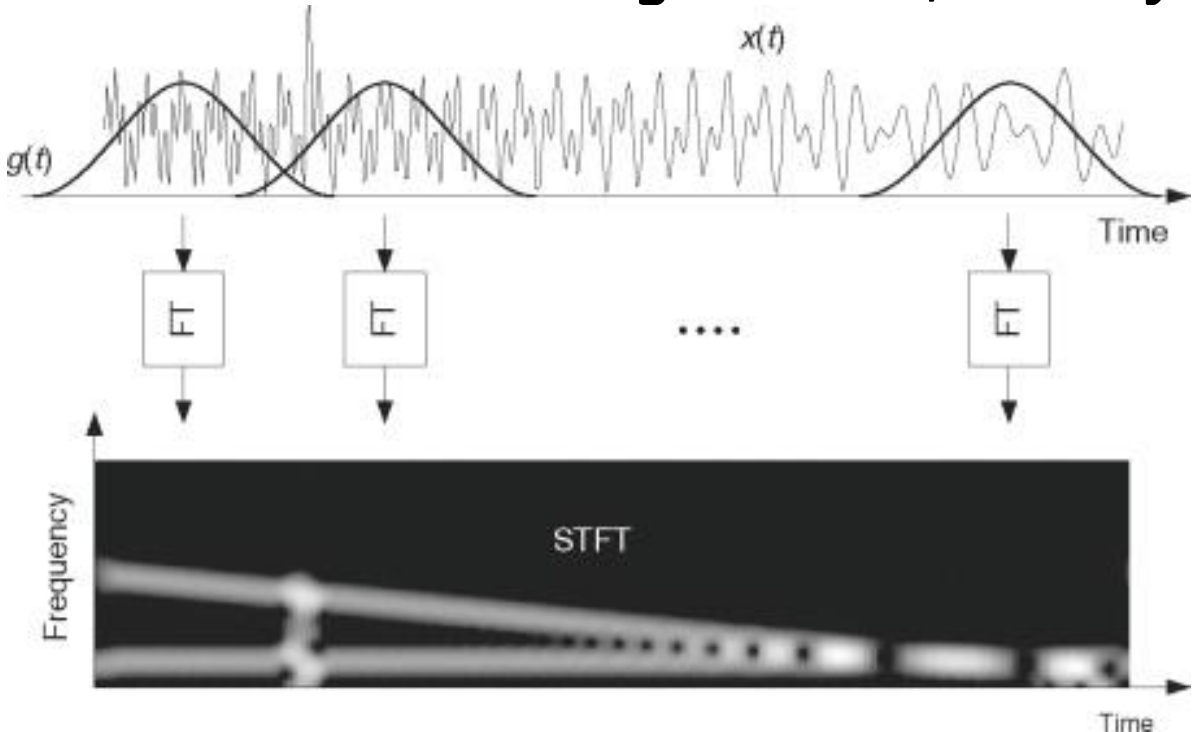
- Continuous Fourier Transform
- $F(f) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i f t} dt$
- Inverse CFT
- $f(t) = \int_{-\infty}^{\infty} F(f) e^{2\pi i f t} df$

- Linearity $\mathrm{F}\{a \cdot f(t)\} = a \cdot F(s), \mathrm{F}\{f(t) + g(t)\} = F(s) + G(s)$
- Invertible
- Differentiable

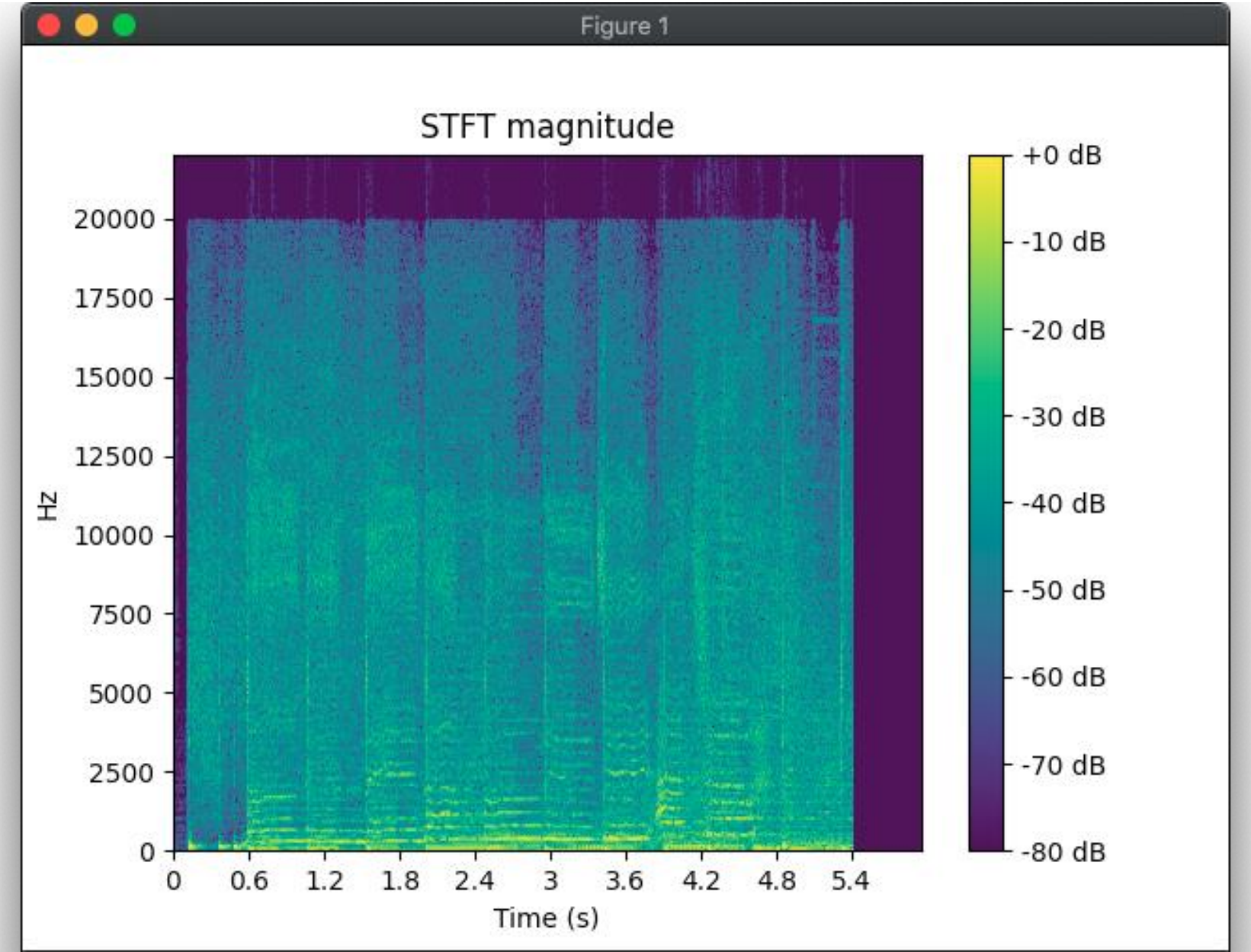
- Decompose signal to sum of periodic signals(sine/cosine)
- Time → (Time bin, Frequency bin)

Short Time Fourier Transform

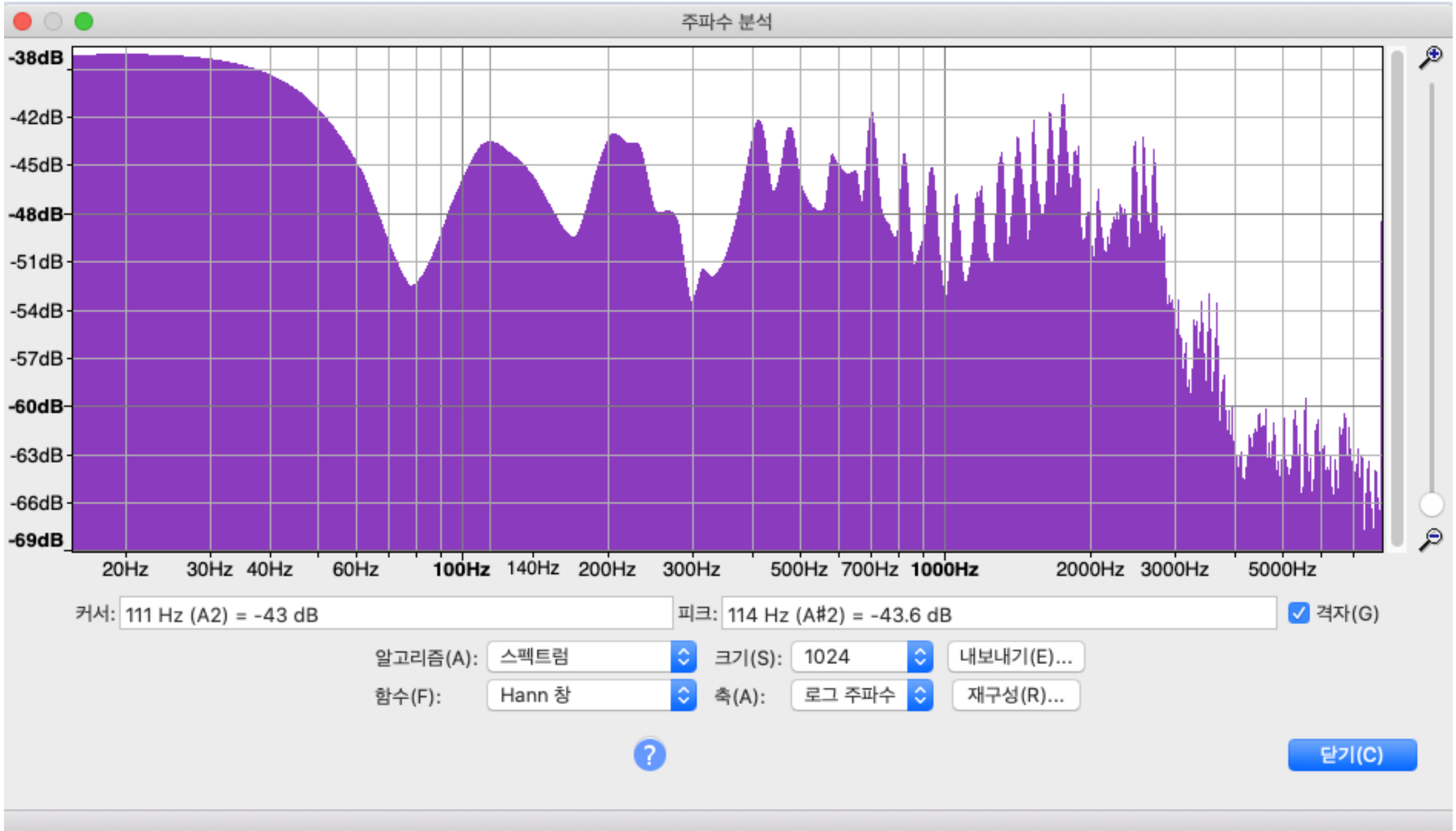
- STFT
- $X(\omega, m) = \sum_{n=-\infty}^{\infty} x[n]w[n - h \cdot m]e^{-j\omega n}$
- NFFT(n): 1024, 512, 256
- Hop length(h): usually quarter of NFFT or 10ms
- Window(w): Hanning window, usually same size w/ NFFT or 25ms



Short Time Fourier Transform



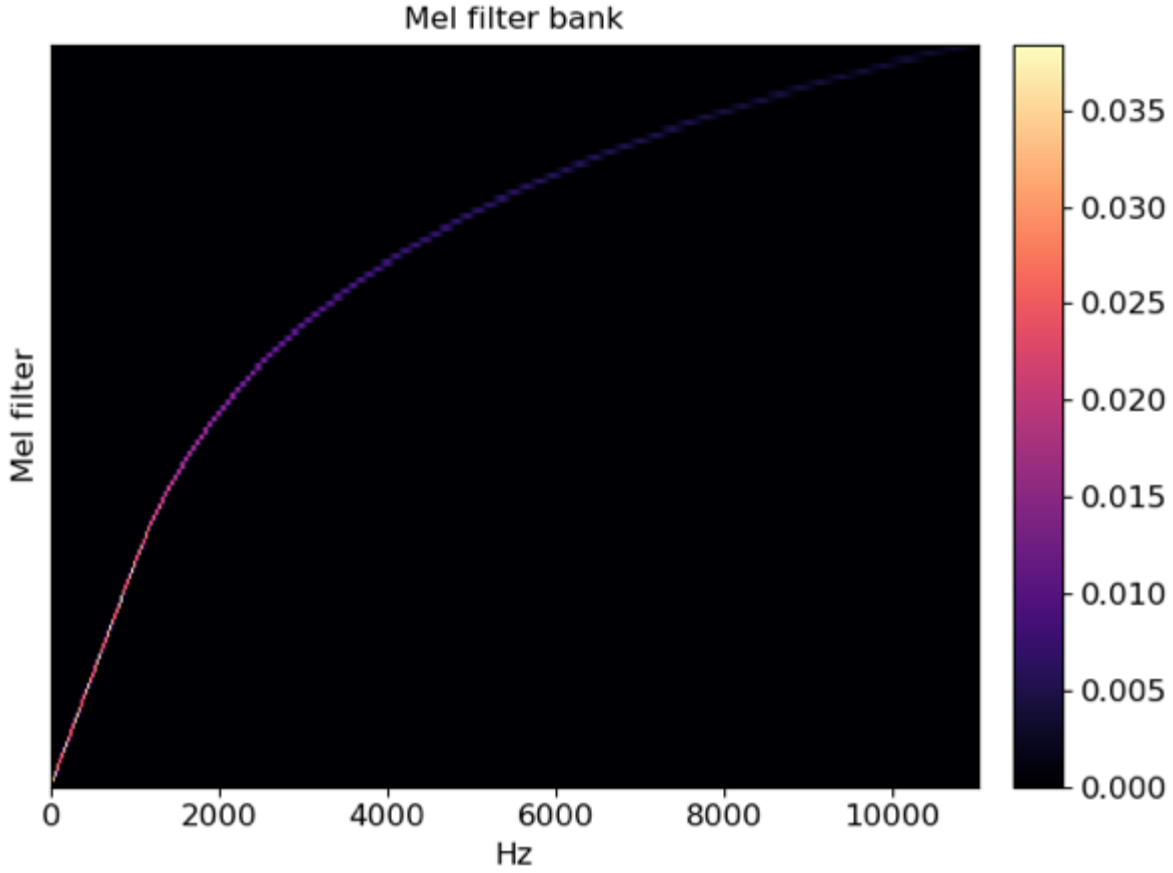
Short Time Fourier Transform (RMS)



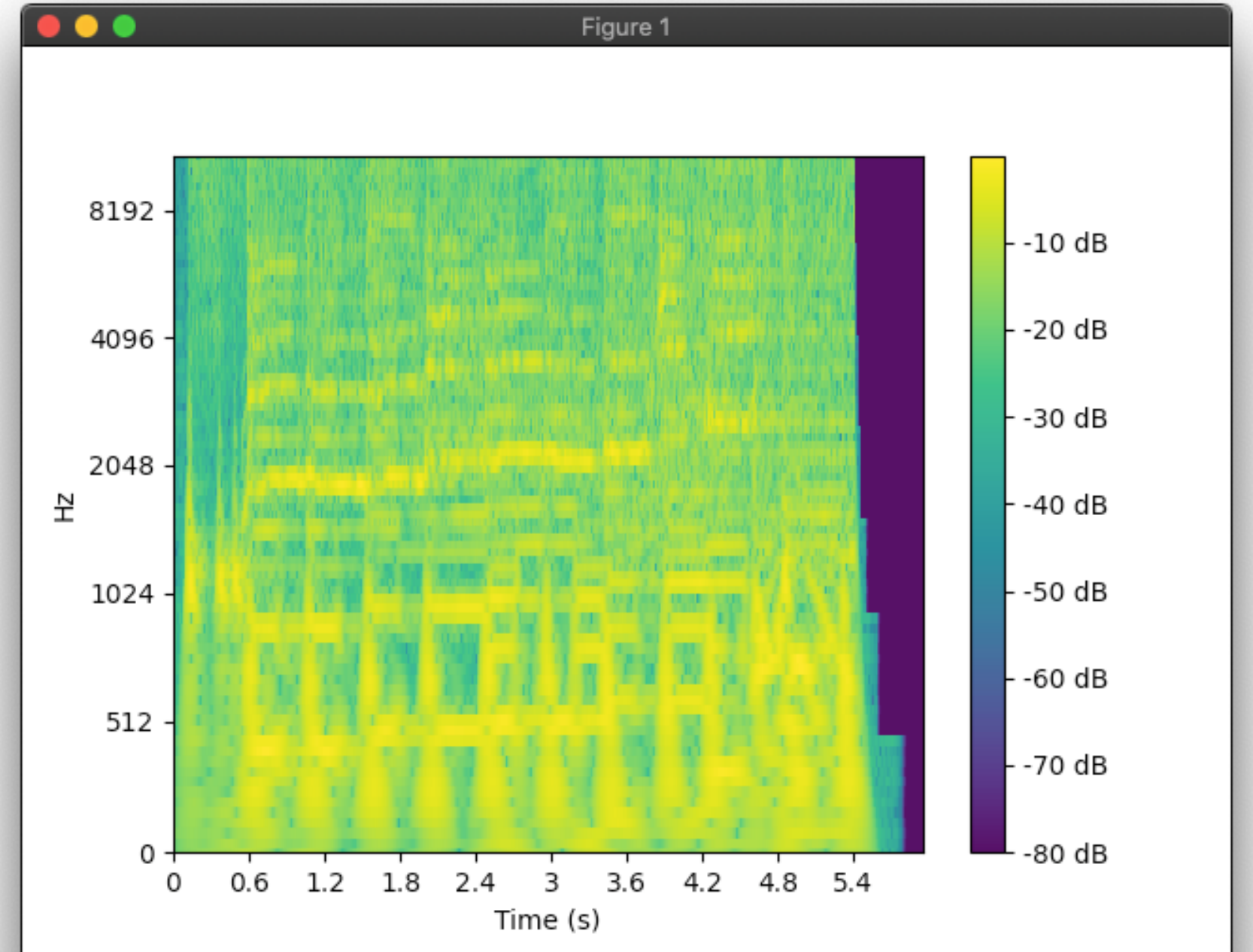
Mel Spectrogram

- STFT's y-axis: linear scale Hz
- MelSpec's y-axis: log scale Hz
- N_mel: 40, 80(tacotron)

- Not invertible!
  - dim 256/512 → 40/80 compressed
  - Why we use vocoder

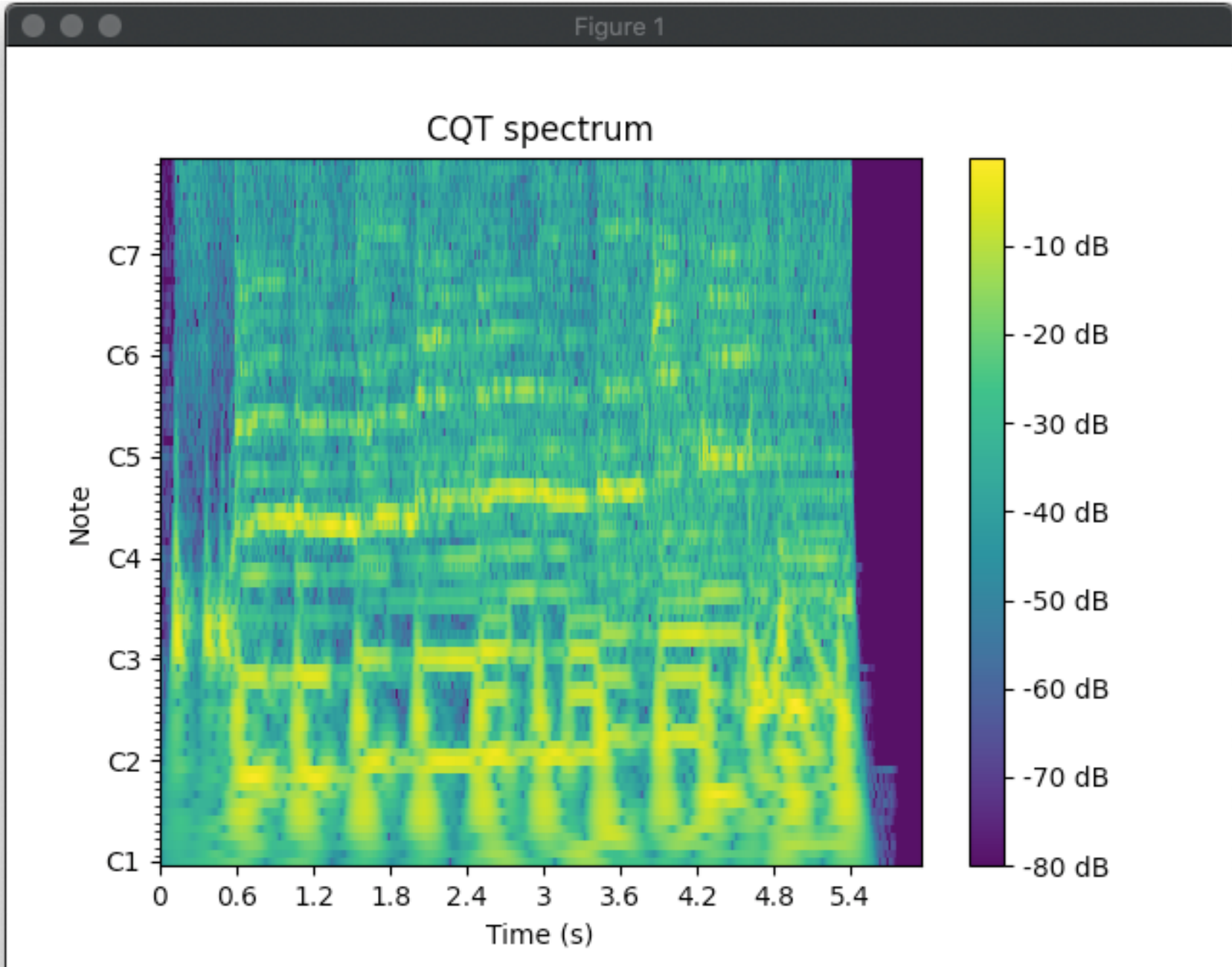


Mel Spectrogram



CQT

- Constant Q Transform
- Fit to music tasks



Thank You

Brain-Audio Internal Education for Consultant

Junhyeok Lee

Brain-Audio

MINDs Lab

Today

Data

Model

음성

음성 생성
Speech Generation, Text-to-Speech

실제 그 사람의 목소리 그대로 자연스럽게, 세계 최고 수준의 음질과 실시간 합성 속도를 제공합니다.

Voice Filter
Voice Filter

내 목소리와 다른 사람의 목소리가 겹쳐 있는 파일을 입력하면 내 목소리를 분리해냅니다. 마인즈랩이 구글에 이어 세계 최초로 구현에 성공한 엔진입니다. (2019년 6월)

음성 인식
Speech Recognition, Speech-to-Text

음성을 텍스트로 변환하는 엔진으로, 다양한 학습모델을 활용할 수 있고 높은 인식률과 빠른 처리 속도를 제공합니다.

음성 정제
Denoise

음성에 섞여있는 배경음과 같이, 음성 내의 다양한 잡음을 제거합니다.

화자 인증
Voice Recognition

사람의 음성 데이터를 Vector화하고 그 값을 대조하여 목소리를 인식합니다.

언어

문장 교정
Bert Correction

잘못된 한글 문장을 문맥에 맞게 교정해줍니다.

한글 변환
Konglish

영어 단어 또는 한글과 영어가 혼용된 문장에서 영어 단어들을 외래어 표기법에 가까운 한글로 변환시켜줍니다.

AI 독해
Machine Reading Comprehension

주어진 텍스트를 독해하여 문맥을 이해하고, 질문에 맞는 정답의 위치를 찾아내서 정답을 제공합니다.

텍스트 분류
eXplainable Document Classifier

뉴스 기사를 입력하면 기사의 주제를 정확하게 분류해 냅니다. 더불어 분류의 근거를 문장 단위와 단어 단위로 제공하는 '설명 가능한 AI'입니다.

자연어 이해
Natural Language Understanding

문장을 입력하면 형태소 분석과 개체명 인식 결과를 제공해 줍니다.

시각

AI Avatar
Face-to-Face Translation

동영상 내 인물의 얼굴 움직임을 포착하여, 사진 속 특정 인물이 이를 따라 움직이는 영상을 만드는 엔진입니다.

AI 스타일링
Text-to-Image for fashion

패션에 대한 설명 텍스트를 입력하면 이를 이미지로 생성해 냅니다.

텍스트 제거
Text Removal

이미지에 있는 텍스트를 찾아 내어 제거해줍니다.

도로상의 객체 인식
AI Vehicle Recognition (AVR)

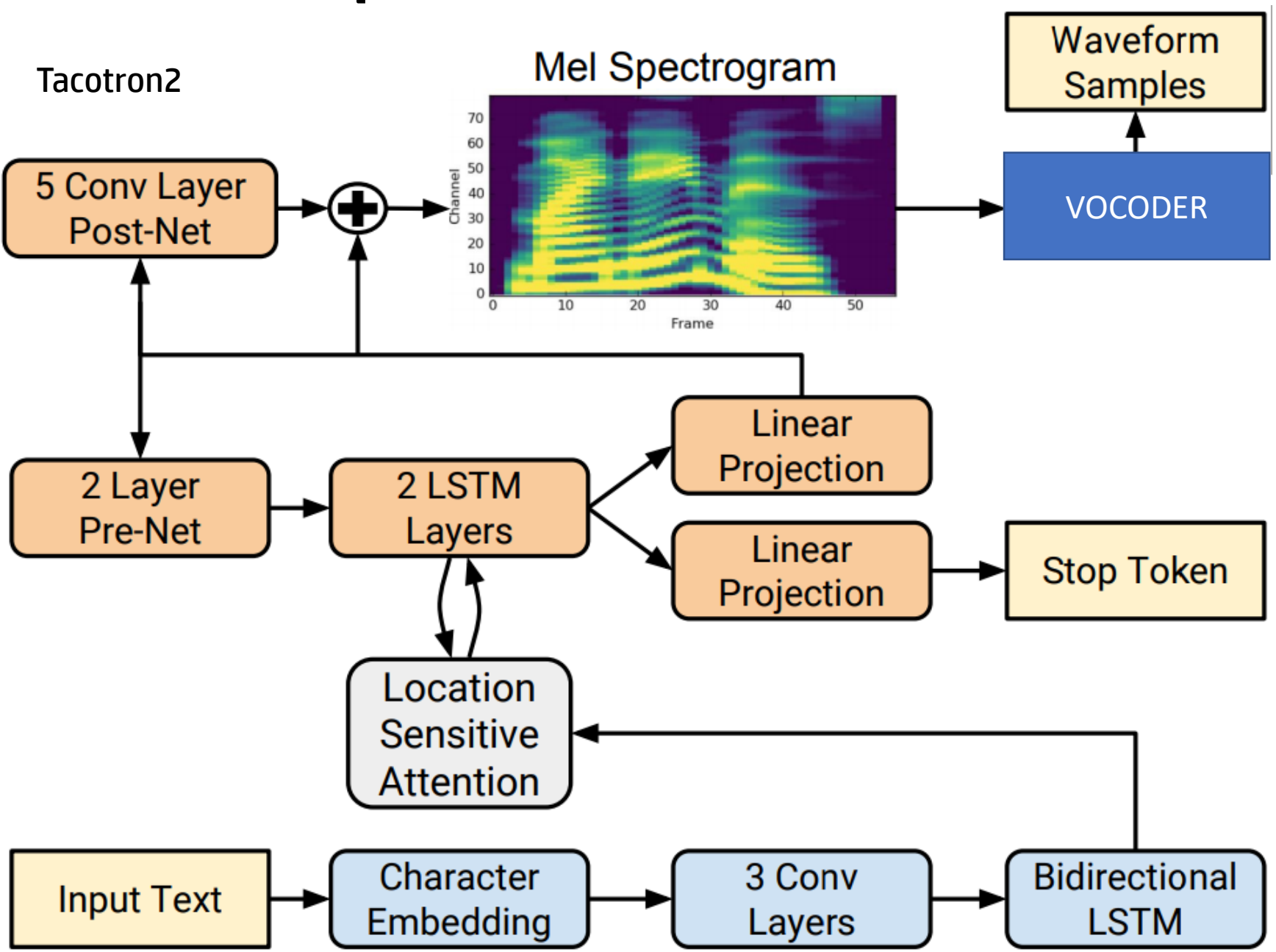
도로 상에서 달리는 차량의 이미지를 입력하면 창문의 위치, 차안에 있는 사람의 위치 그리고 번호판의 위치를 표시해줍니다.

얼굴 인증
Face Recognition

사람의 얼굴 데이터를 Vector화하고 그 값을 대조하여 얼굴을 인식합니다.

Text To Speech

Tacotron2

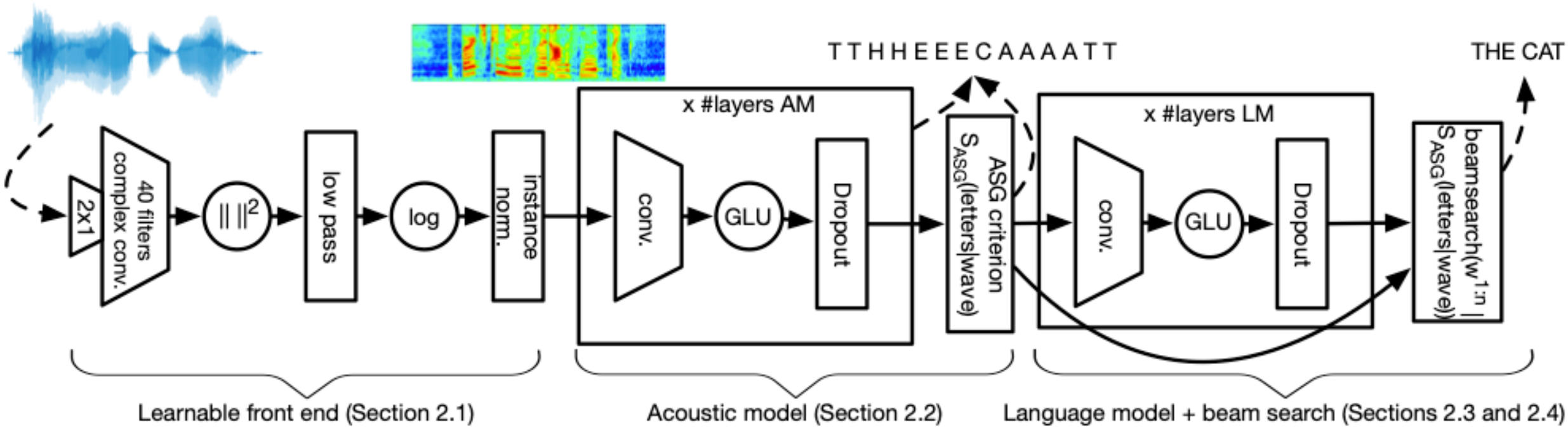


Tacotron2
Data: Speech(clean) + Text
Text 를 인풋으로 Mel spectrogram 생성

너무 짧거나 긴 음성은 잘 안되는 경향
텍스트 마사지 필요 (G2P 등등)
언어마다 전처리 상이

Vocoder
(MelGAN, WaveGLOW, VocGAN)
Data: only speech(clean)
Mel에서 Raw audio를 생성하는 모델
별도 학습

Speech To Text/Automatic Speech Recognition



Wav2Letter

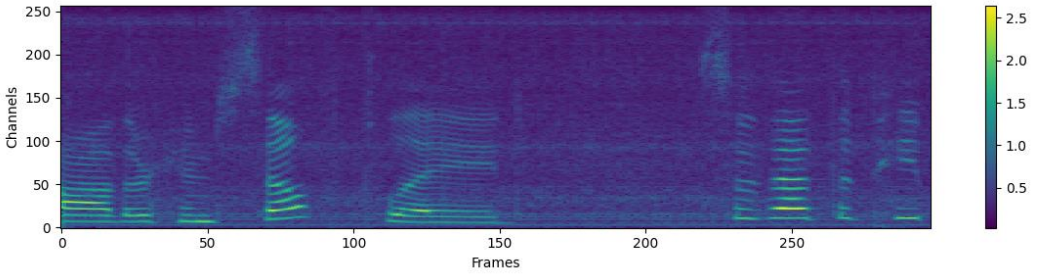
AM: Acoustic Model
LM: Language Model(한국어는 AM만으로도 잘된다)

Data: Speech(noisy ok) + text
+ noise(for augmentation)

IRM(Ideal Ratio Mask)

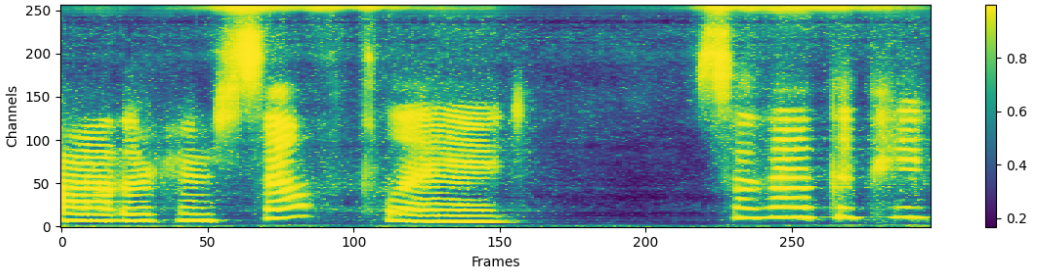
Only use magnitude of STFT spectrogram

Noisy

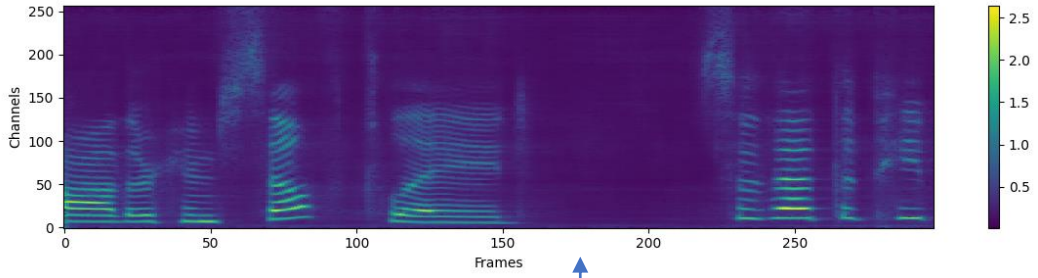


Estimate mask

Ideal Ratio Mask

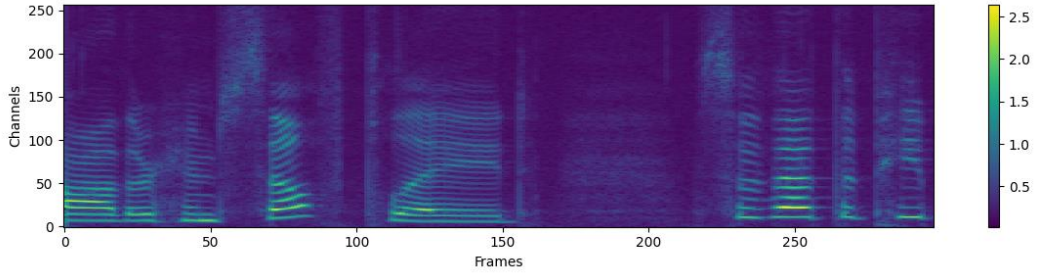


Masked

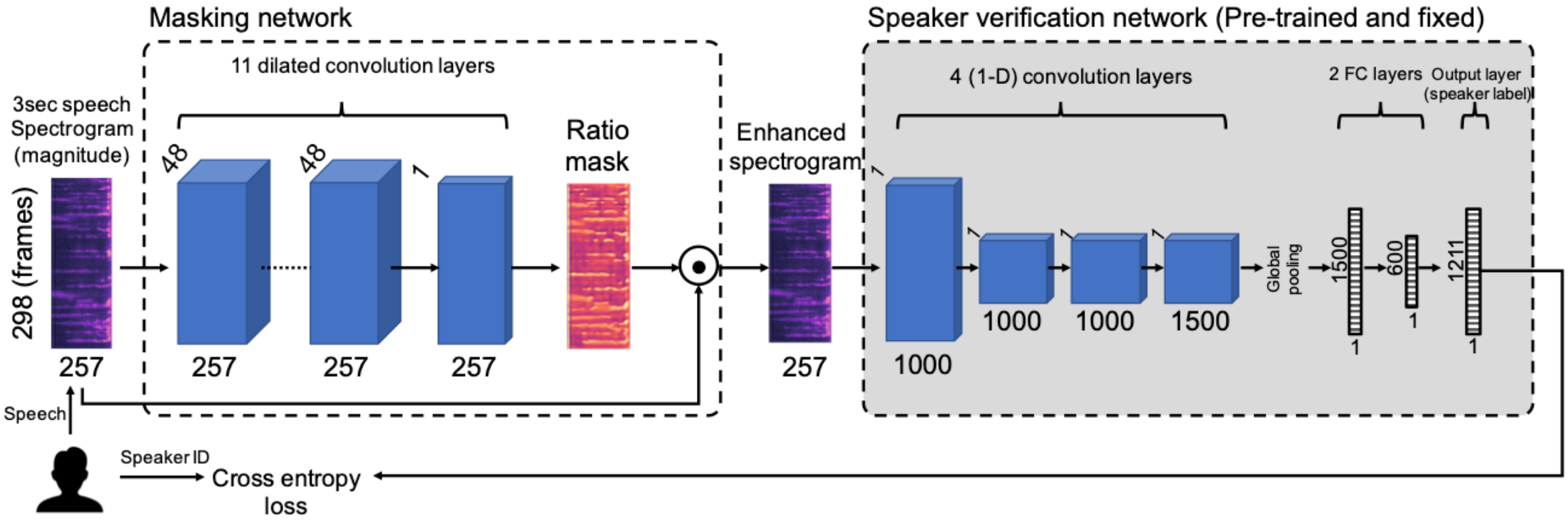


Calc loss

Target(clean)



Speech Enhancement

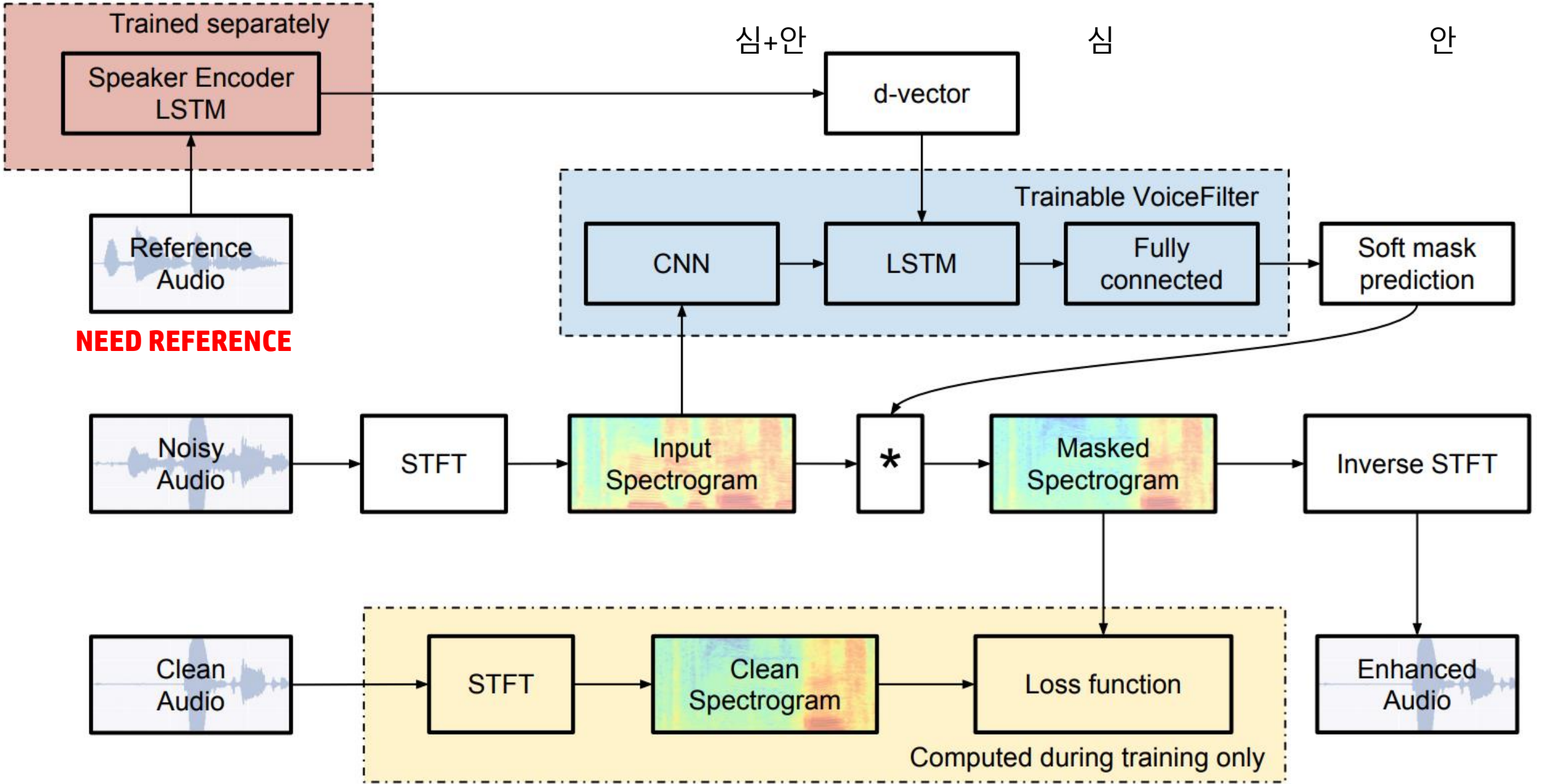


- Data: Clean Speech + Noise
  (Noisy speech + Noise is also possible)

VoiceFilter



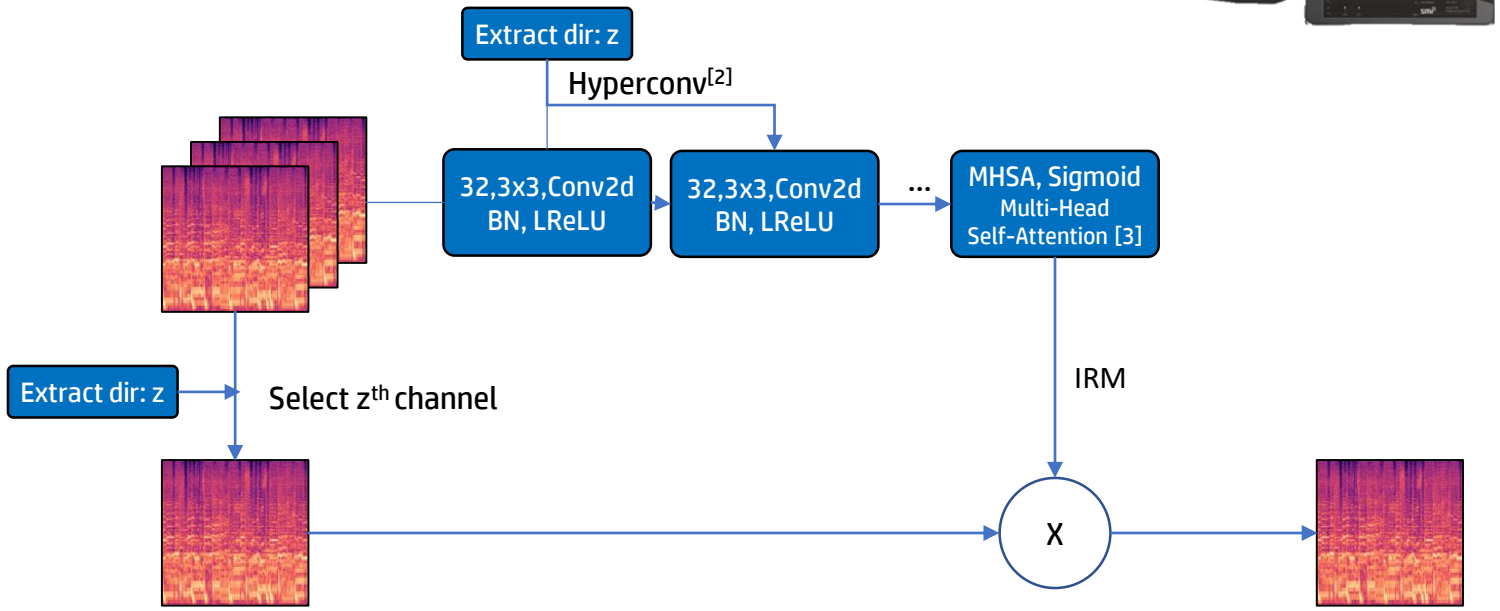
심+안 심 안

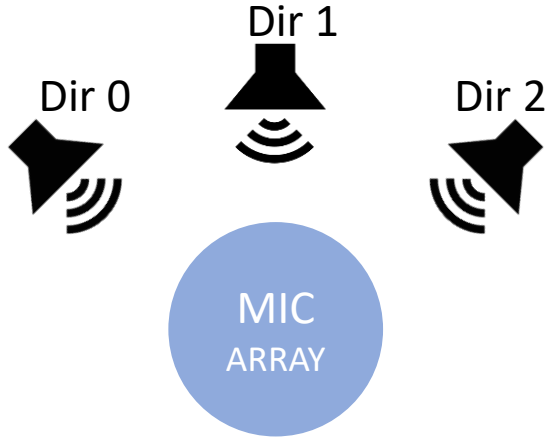


Speech Separation

- Blind Source Separation: 진행 예정

- Non-Blind Source Separation:
  - Directional Separation(w/ SMI)





Set 22802 (Unseen)

Dir 0

Mixed Masked Target

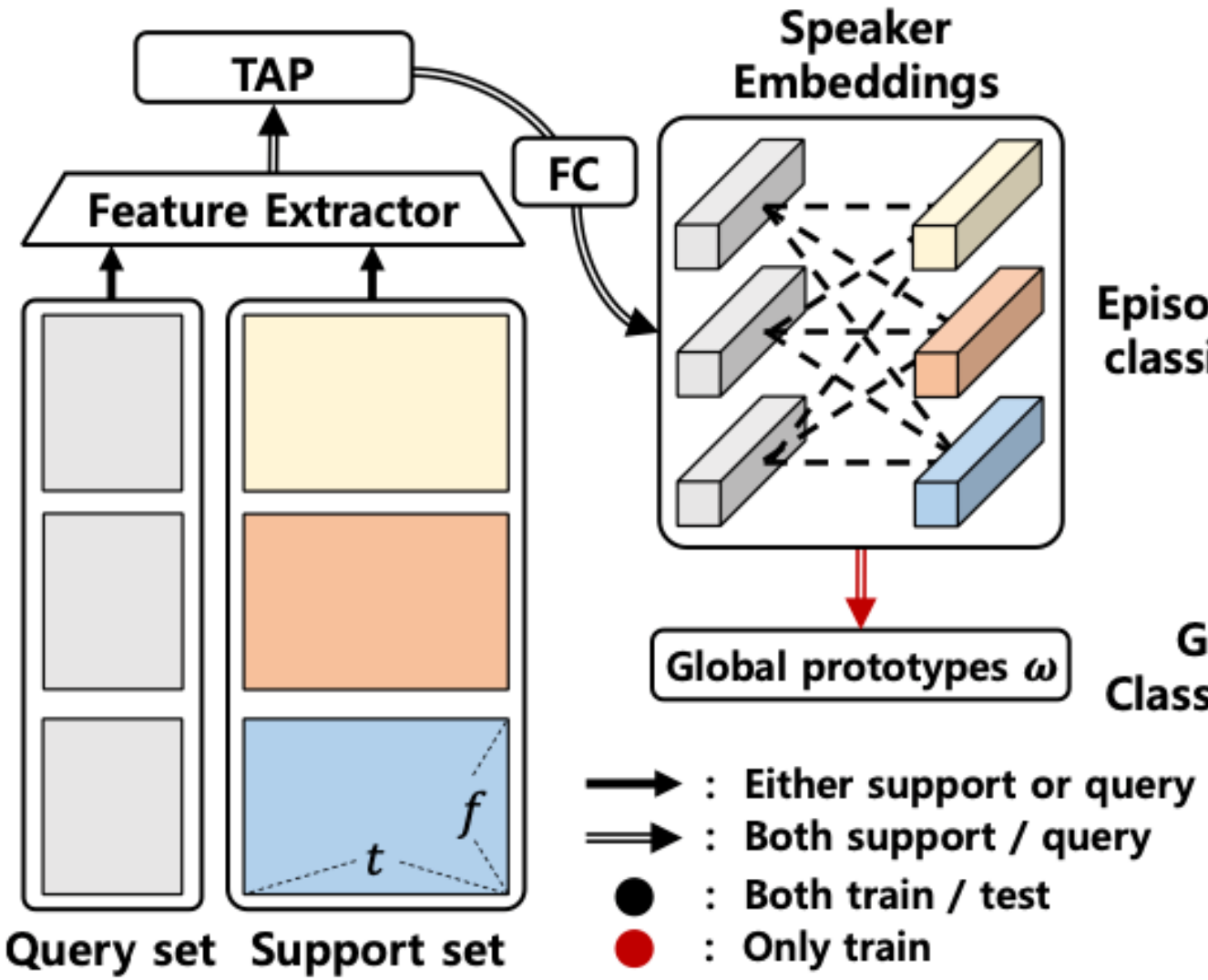
Dir 1

Mixed Masked Target

Dir 2

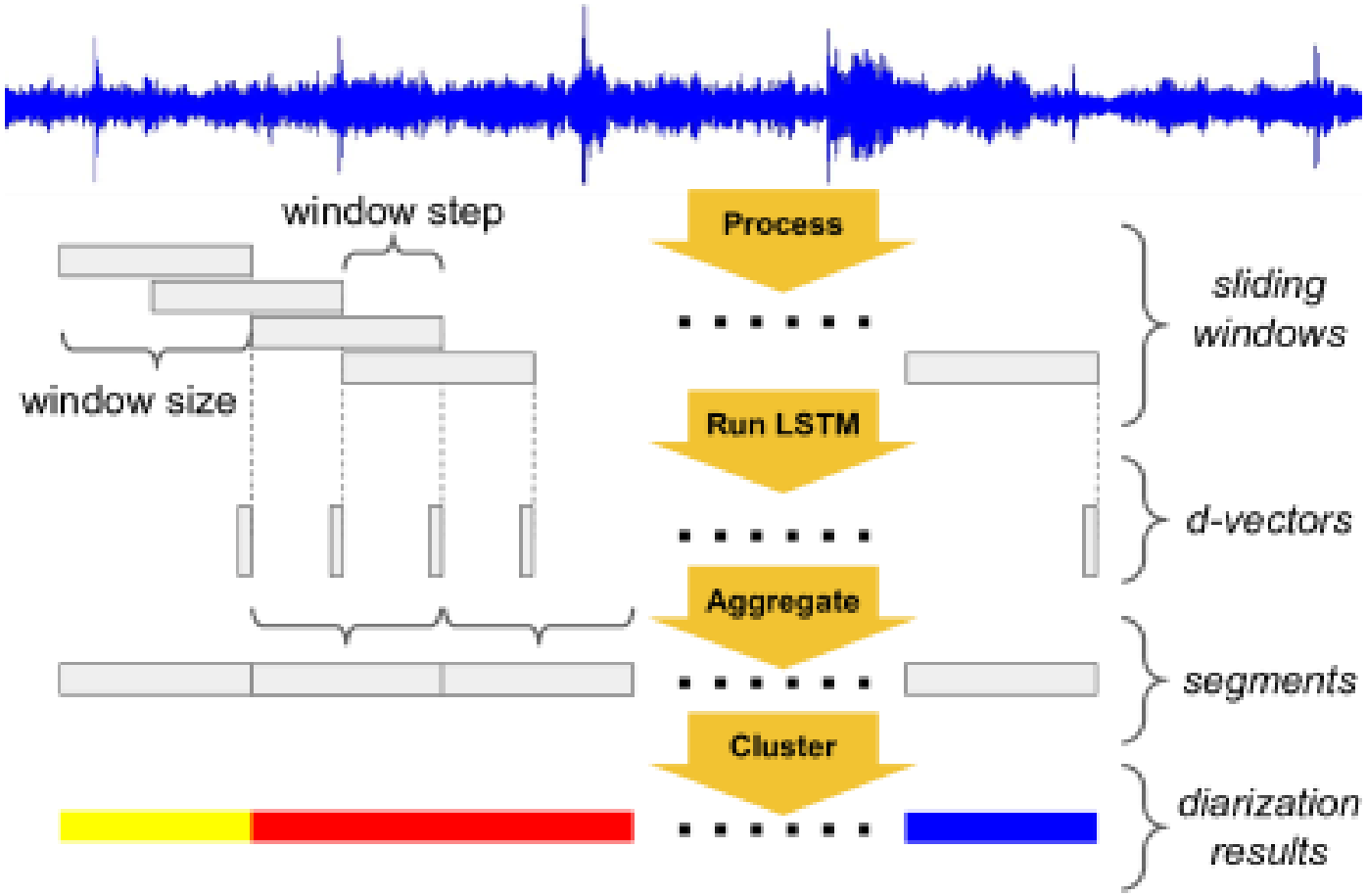
Mixed Masked Target

Speaker Verification/Recognition



- Speech -> Speaker Vector
- Compare 2 speaker is identical or not
- Data: speech + id
- (Speech w/o id could be used by self-supervised learning)

Speech Diarization



Speech → Time stamp of "who speaks when"

Current model: low acc
Need modification

Clustering d-vectors

Current system does not including learning part