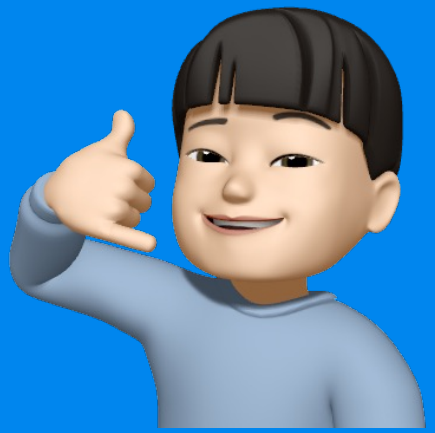
내 음색에 어울리는
음악 추천시스템

이지평 장성현 김보현 김종윤 김정하

팀원 소개

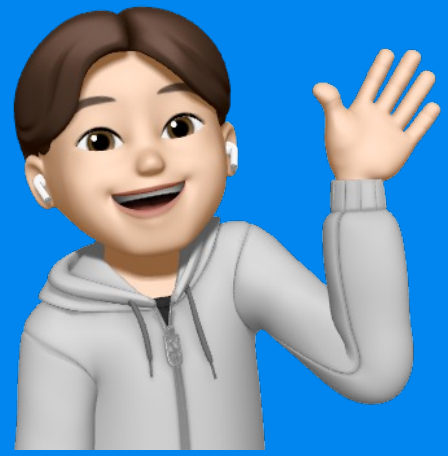
이지평

김정하

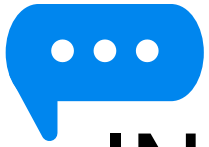


장성현

김보현



김종윤

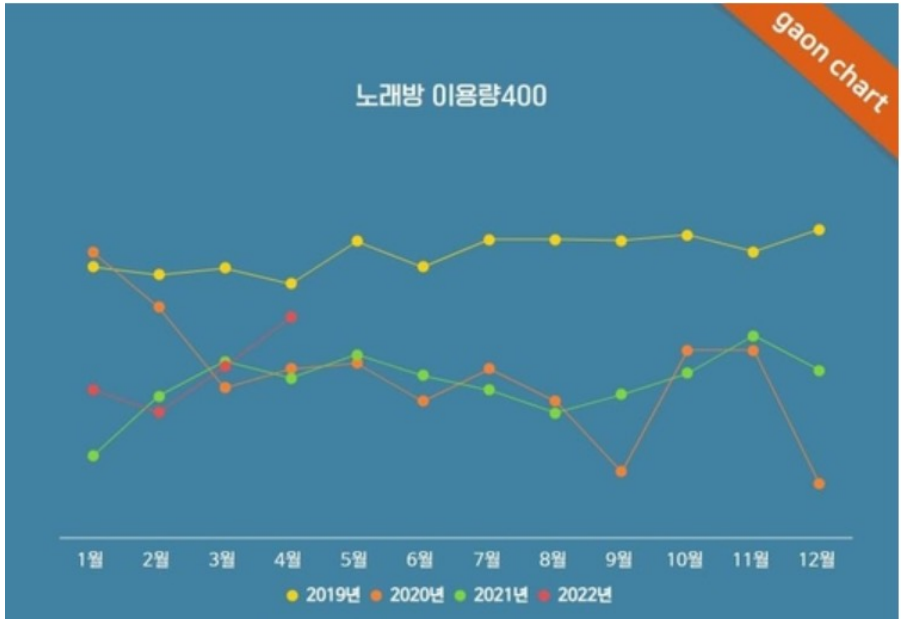


INDEX

- 주제 설명
- Contribution
- Framework
  - Framework
  - NSVB
  - Contrastive learning
- 진행 계획

01

주제 설명



올해 4월 '노래방 이용량 400'은

코로나19 대유행 첫해인 2020년 4월보다는 30.6%,

사회적 거리두기 전면 해제 이전인 올해 3월보다는 28.9% 증가

→ 노래방 사용률 증가 추세

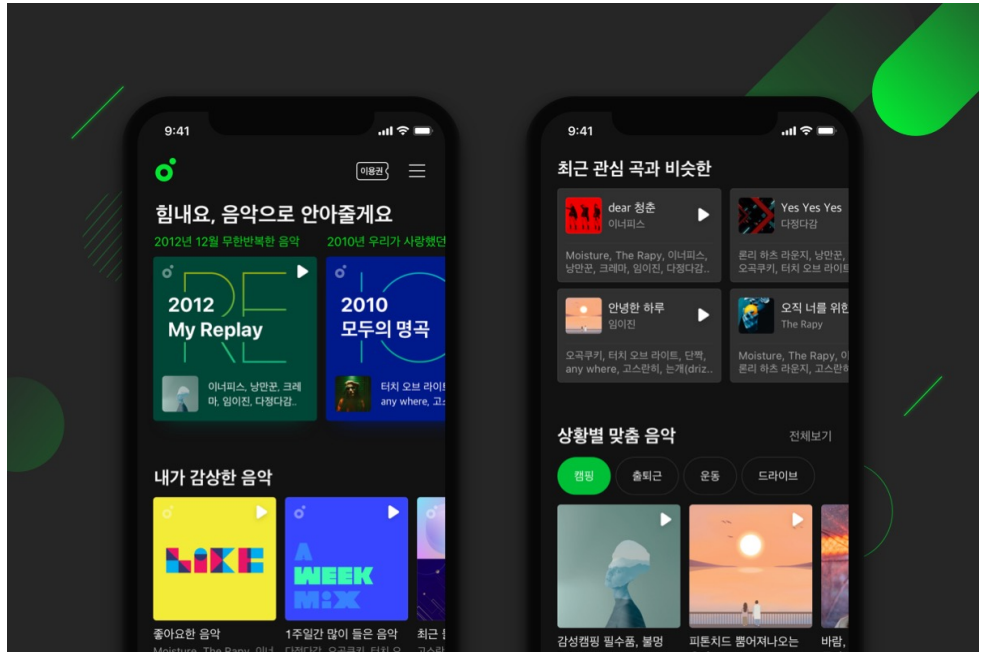
01

주제 설명


낭스월드
인기차트
가요
1위 — MV 포장마차
2위 — 조금취했어
3위 — MV OST 흔들리는꽃들속에서네샴푸향이느
4위 ↑01 사랑이란 멜로는없어
5위 ↓01 OST 안녕(호텔델루나OST)
MV 니소식
사랑이식었다고말해도돼
황인욱
임재현(Prod.2soo)
장범준
정상근
폴킴
송하예
먼데이키즈
HYNN(박혜원)

01

주제 설명



음악 추천시스템에 대한 기존의 연구들은

사용자의 인구통계학 정보,

개인취향 및 최근 관심 곡 등을 기반으로

듣는 음악에 대한 추천이 대부분

01

주제 설명

“ 듣는 음악 추천이 아닌 부르는 음악 추천 “

사용자의 목소리 하나만으로 추천을 해주는

콘텐츠기반의 추천시스템 모델을 개발하고자 함

02

Contribution

1. 기존 연구들의 한계였던 **음색, 보컬 스타일을 추출**하기 위해 **Contrastive Learning**을 사용한다.

2. Vocal note 조정을 통해 이성 노래 추천이 가능하다.

3. 해당 task에 적합한 **한국어 노래 데이터셋을 구축**하여 사용한다.

02

Contribution – 데이터셋 구축

솔로 가수 음원 크롤링

- 30곡 이상 (피쳐링 제외) 보유한 솔로가수 150명

음원 자르기

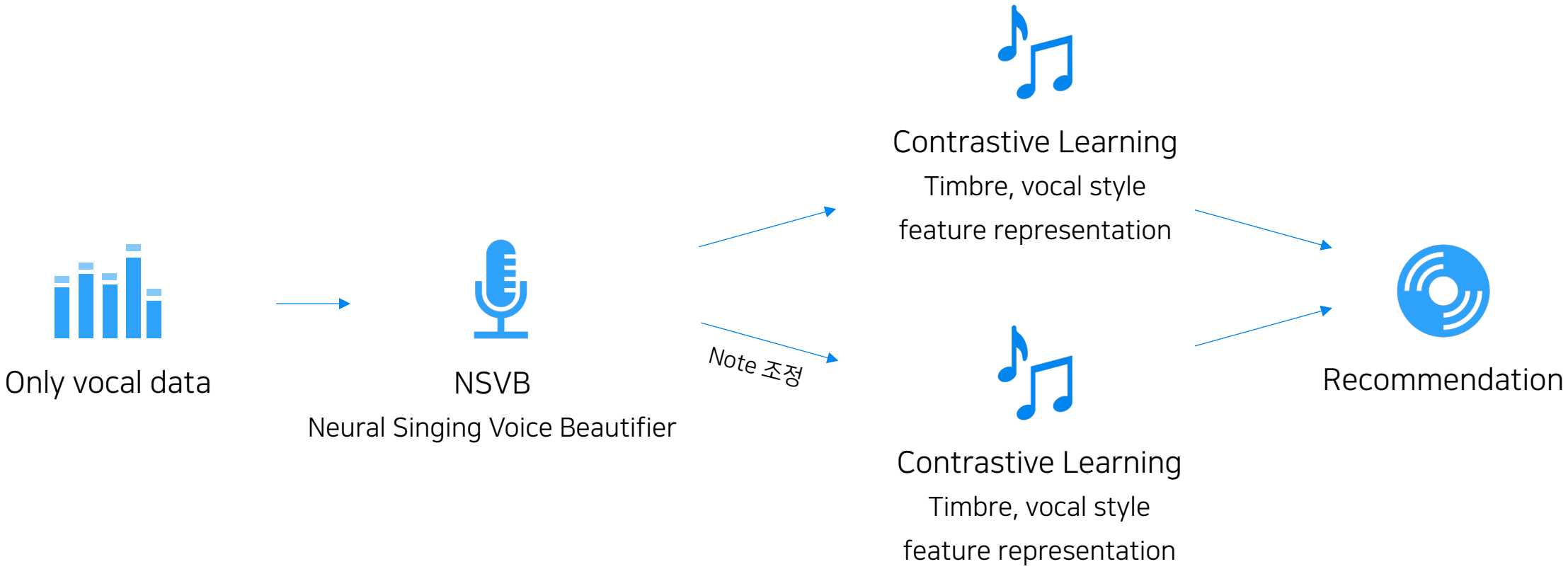
- 30초 ~ 2분 30초

보컬 추출 🤷🏻

- MR을 제거하여 보컬 데이터만 추출

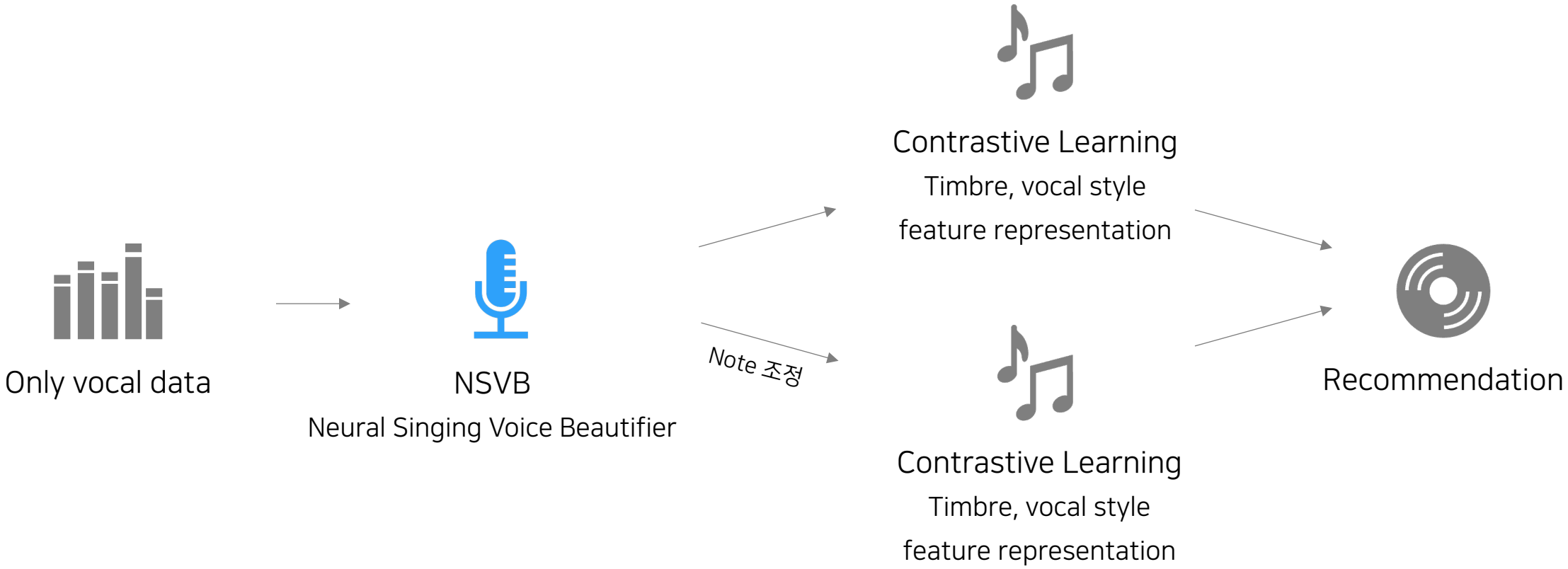
03

Framework



03

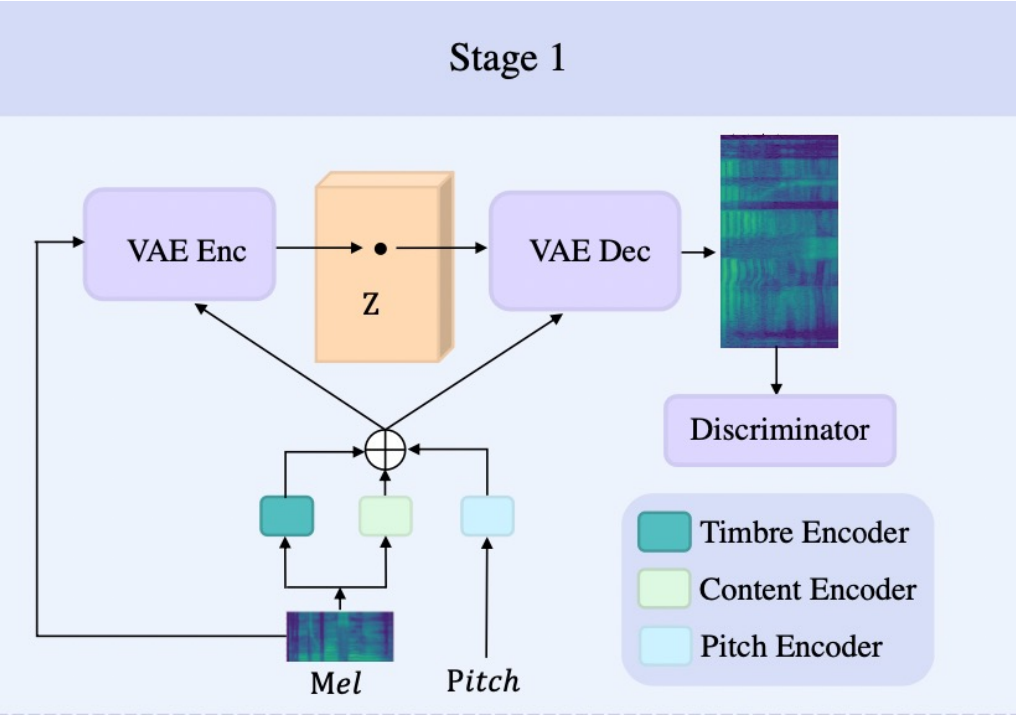
Framework



03

Framework

NSVB



목적 : 주어진 표현(Condition)을 가지는 노래의 잠재적인 특징(Z)을 찾는 것

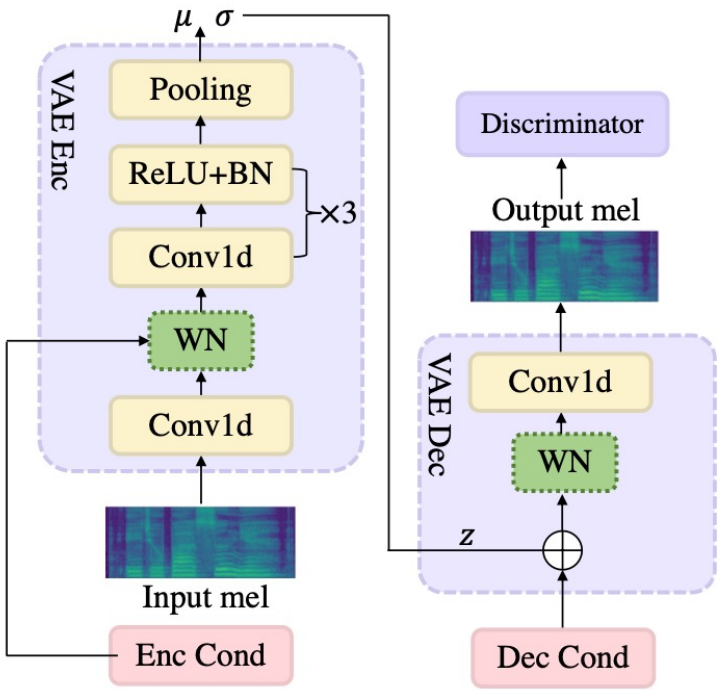
- Condition : Timbre, Content, Pitch
- Encoder/ Decoder 모두 Condition이 주어짐
- 아래의 loss function을 최소화하면서 모델을 최적화한다.

$$L(\phi, \theta) = -\mathbf{ELBO}(\phi, \theta) + \lambda L_{adv}(\phi, \theta),$$

03

Framework

NSVB



Condition에 관한 특징을 학습하기 위한 Encoder/Decoder

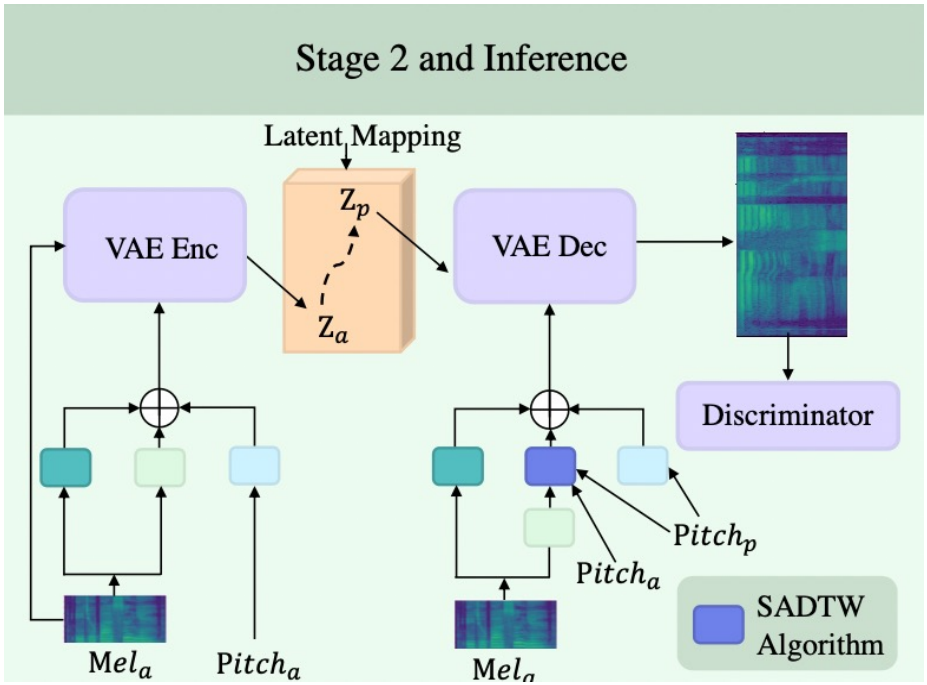
목적: Condition을 나타내는 최적의 잠재적인 특징(latent space, Z)를 찾는 것

- Encoder's output : Condition의 분포에서 추출된 평균과 표준편차
- 추출된 평균과 표준편차를 통해 latent space 생성
- Stage1의 Timbre, Content, Pitch Encoder는
  각 Condition의 잠재적인 특징(latent space,Z)을 output으로 갖는다.

03

Framework

NSVB



목적: **프로 가수의 Condition**을 갖는 새로운 mel-spectrogram 생성

- Decoder의 Condition:
  아마추어의 timbre, 아마추어의 content와 보정된 pitch, 프로의 pitch

- SADTW Algorithm :
  기존 아마추어의 pitch를 보정하여 향상된 pitch를 만듦

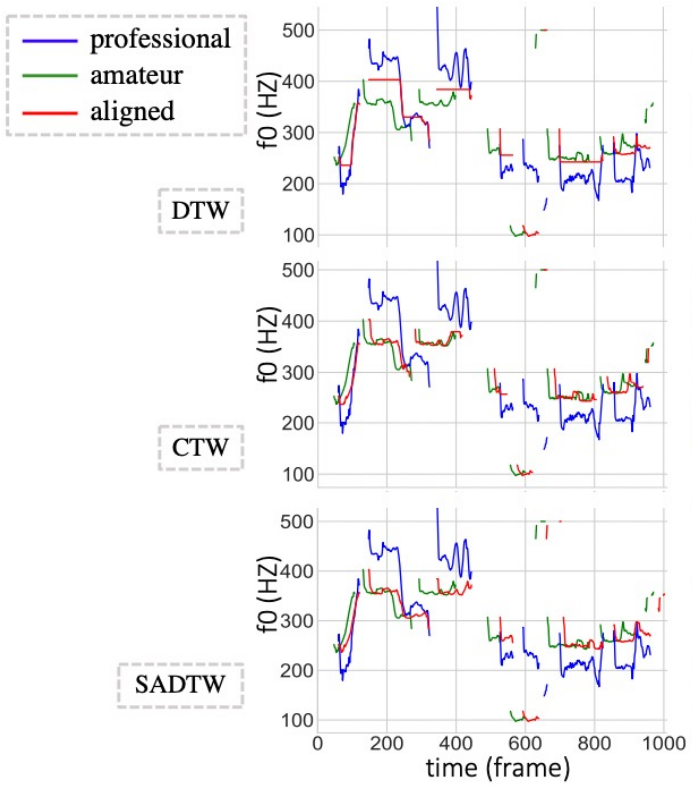
- Latent Mapping :
  아마추어의 latent space → 프로의 latent space

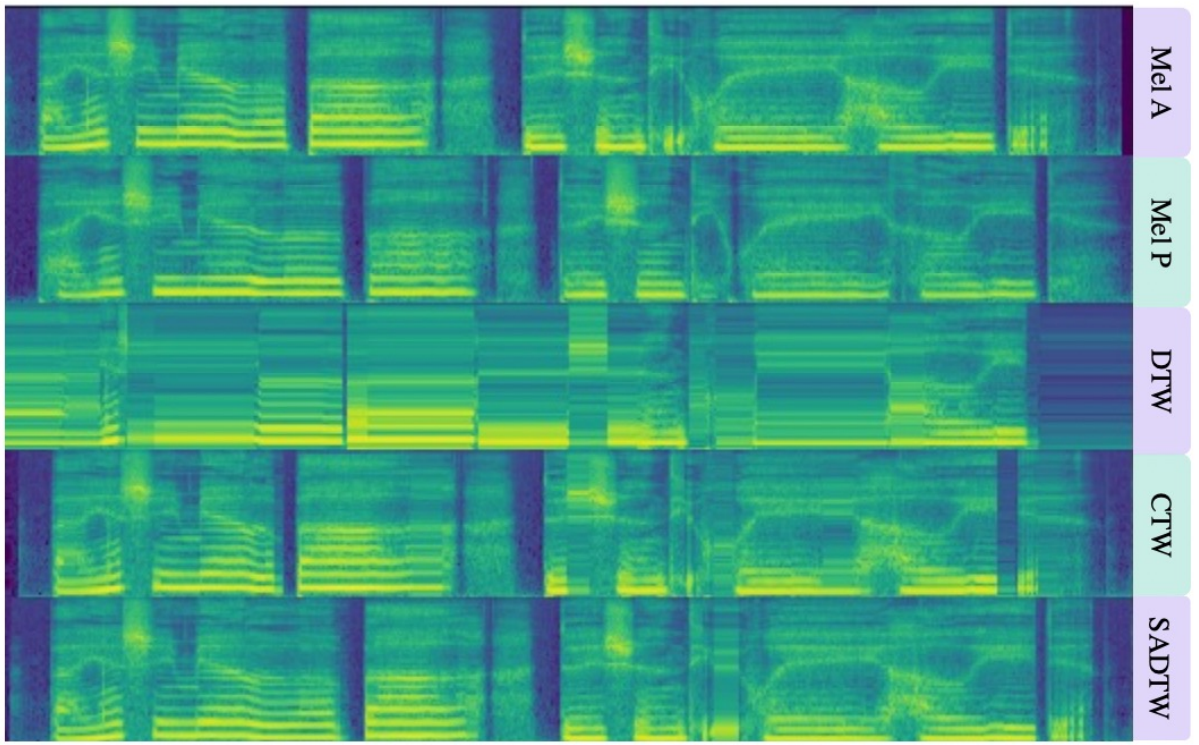
- 즉, Decoder를 통해 아마추어의 timbre는 유지하되
  프로의 pitch와 노래표현을 갖는 새로운 mel-spectrogram 얻음

03

Framework

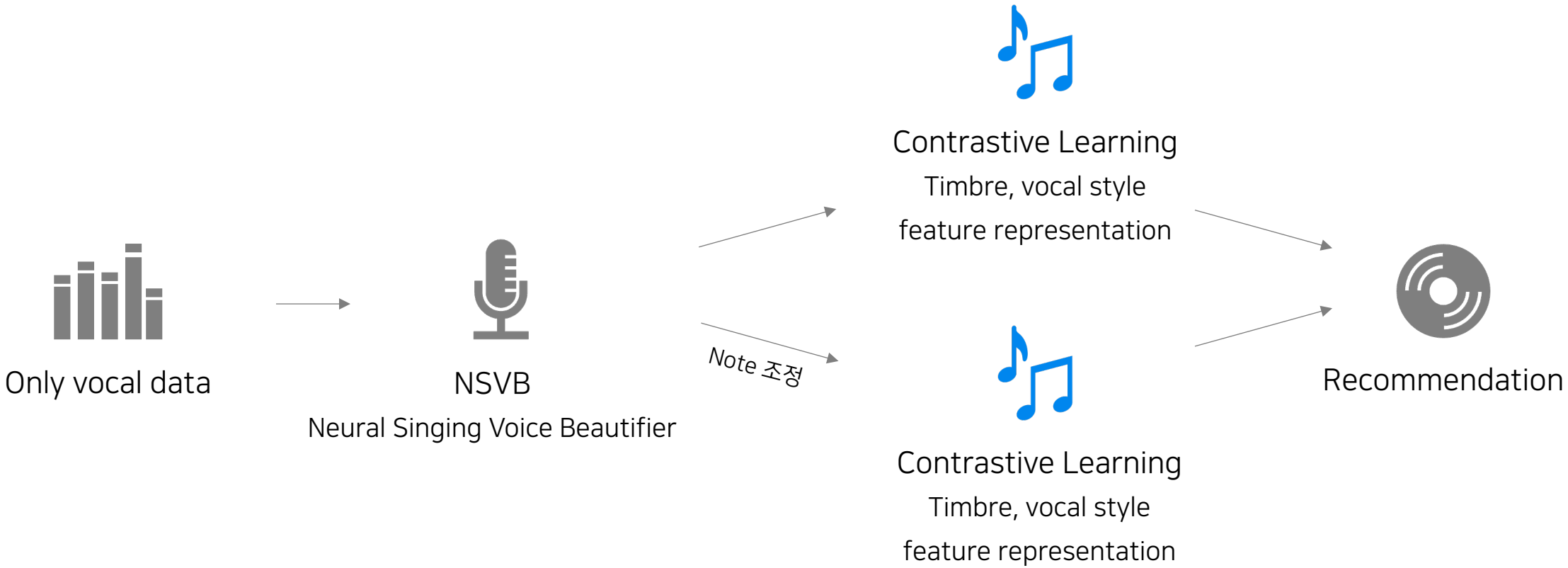
NSVB - SADTW Algorithm





03

Framework



03

Framework

Contrastive Learning

Self-Supervised Contrastive Learning for Singing Voices

노래하는 목소리 데이터를 활용하여 음색 관련된 특징을 추출하고자 함

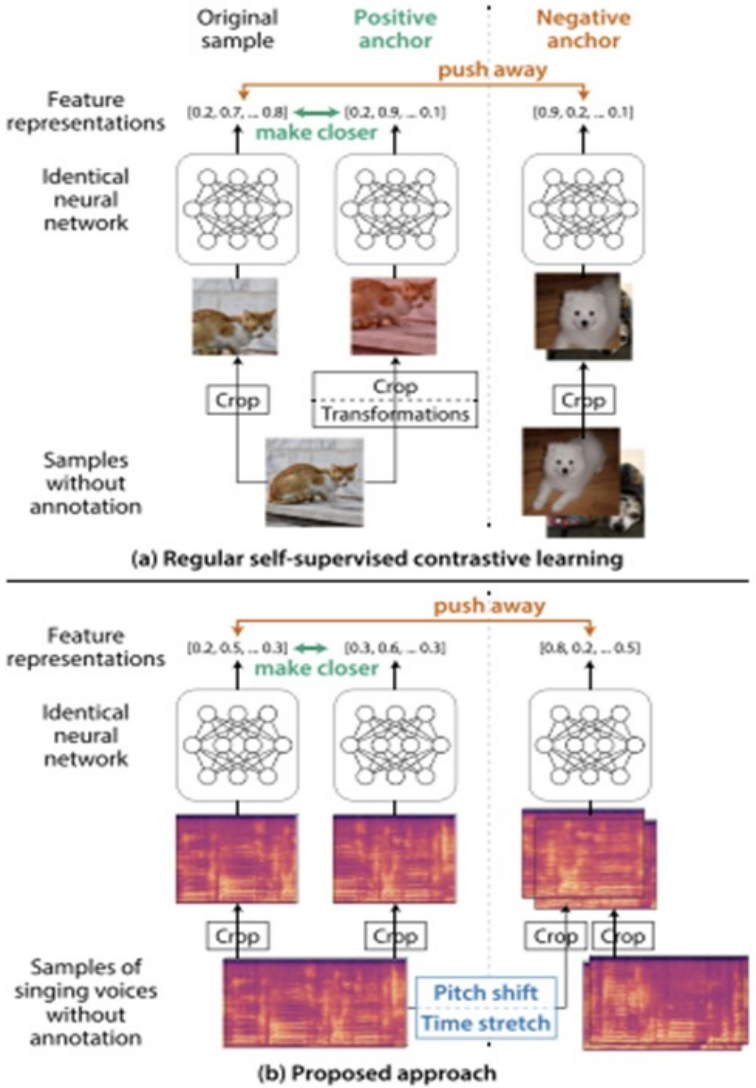
주요 기법

- Representation Learning
- Self-Supervised Contrastive Learning

위 기법들을 활용해 기존의 연구에서 추출해내기 어려웠던 음색을 뽑아내는데 성공함

Negative Pair 생성 방법

- Pitch Shift : 음역대에 관한 특징 표현을 학습할 수 있음
- Time Stretching : 시간을 늘려서 비브라토와 같은 템포와 상관없는 특징을 학습할 수 있음



03

Framework

Contrastive Learning

1. Representation Learning :

데이터에 적절한 표현을 찾도록 학습하는 과정을 의미

2. Self-Supervised Contrastive Learning :

- 라벨링이 되어 있지 않은 데이터로 학습을 진행하는 Self-Supervised Learning과정에서 생기는 비용을 단축하기 위해 Contrastive Learning 방식을 통해 보완을 하도록 함

- 동일한 Class의 데이터를 Augmentation을 이용해 Positive Pair를 만들고, 다른 class의 데이터를 가지고 Negative Pair를 만듦

- 여기서는 Negative Pair를 만들 때 Pitch-shift와 Time-stretching을 이용했음

04

진행 계획

1. NVSB 학습

2. Sub modeling – TCAE

3. Gender classification

4. 최종 추천시스템 모델 구축

Reference

Liu, Jinglin, et al. "Learning the Beauty in Songs: Neural Singing Voice Beautifier." arXiv preprint arXiv:2202.13277 (2022).

Yakura, Hiromu, Kento Watanabe, and Masataka Goto. "Self-Supervised Contrastive Learning for Singing Voices." IEEE/ACM Transactions on Audio, Speech, and Language Processing 30 (2022): 1614-1623.

감사합니다.

이지평 장성현 김보현 김종윤 김정하