

박명수 목소리 만들기

휴먼명수체

중간고사 이전

음성 도메인 지식을 쌓기 위한 배경 지식 스터디

Mel Spectrogram

KUBIG 음성 프로젝트 세션
17기 진유석

Transformer

Attention Is All You Need

KUBIG 음성 프로젝트
17기 진유석

VITS

Conditional Variational Autoencoder with Adversarial Learning for
End-to-End Text-to-Speech

Jaehyeon Kim¹ Jungil Kong¹ Jubee Son^{1,2}

2018190313 진유석

중간고사 이후

프로젝트: 인공지능을 통한 가수 목소리 바꾸기(Voice Conversion)

* Voice Conversion: 음성에 포함되어 있는 발화자의 특징을 변환하여 타인의 음성을 생성하는 것

데이터 확보를 위해 라디오 진행 경험이 있는 박명수를 선택

Voice Conversion with 박명수

역할 분담

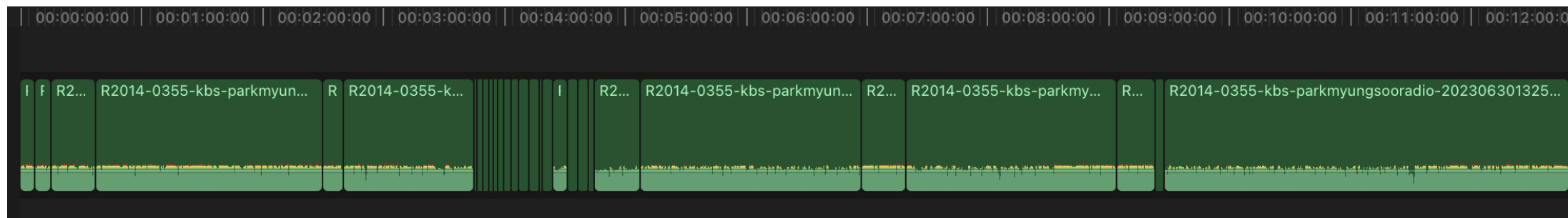
전처리: 안영지, 이은찬, 진유석

학습: 신인섭, 천원준

Voice Conversion with 박명수

전처리 방법 1

- ① 40분 라디오 파일 2편(총 80분) 준비
- ② 박명수 목소리만 나온 부분만 편집 (총 13분)
- ③ 가우디오 스튜디오를 통한 배경음악 삭제
- ④ 15초 간격으로 잘라 학습 데이터 준비



Voice Conversion with 박명수

전처리 방법 2

- ① 클로바 노트로 박명수가 나온 부분 확인
- ② pydub.AudioSegment를 통해 박명수 발화 부분만 편집
- ③ 스테레오 → 모노 리샘플링
- ④ 무음인 부분 제거
- ⑤ 일정 길이로 잘라 학습 데이터 준비

CLOVA Note®

새 노트 만들기

홈

전체 노트

공유 받은 노트

내 폴더

기본폴더

휴지통

300분 중 268분 남음

녹음은 앱에서 무제한 변환 가능하고,
사용 시간은 7월 28일에 갱신됩니다.

여러분의 의견을 들려주세요

audio_data

전체 노트 · 6.28 수 오후 3:29(6.28 수 수정) · 32분 12초

박명수

음성 기록

점유율

편집

박

박명수 00:00

한 주 시작 월요일이고요. 오늘은 이제 초대 손님이 있는 날이죠 직업의 섬세한 세계가 있는 날인데 상당히 어 이 스튜디오 안에 분위기가 좋습니다.
아주 잘생긴 모델 출신의 또 탤런트 배우 한 분 나오셨는데 이수혁 씨가 나오십니다.
이수혁 씨가 오늘 또 함께 하는데 이수혁의 모든 것들을 한번 파헤쳐보도록 하고요.

박

박명수 00:21

아 왜 여기 나왔는지도 물어보겠습니다.
저랑 같이 프로그램도 하고 있지만 이수혁 씨 대에서 예 진짜 많은 것들이 좀 감춰져 있습니
다.
그래서 오늘 좀 확실하게 이렇게 좀 알아보도록 하고 예 이수혁 씨 모시기 전에 간단하게 이수
혁 씨의 모습을 잠깐 좀 소개해 드리면 바지를 여름 바지를 입고 오셨어요.
굉장히 좀 아버지 같은 걸 입고 오셨는데 조금 그 직업 세계에 계신 분들이 입는 스타일을 입
고 오셨는데 상당히 좋게 입었습니다.
그래서 좀

박

박명수 00:47

일단 몸매가 워낙 좋아서 그렇긴 하지만 여름 바지를 입고 오셨는데 잠시 후에 한 번 보이는
라디오를 통해서 공개하도록 하겠습니다.
이수혁 씨와 한 시간 이수혁의 모든 걸 알아보구요.
궁금해서 부들부들 사 팔고있고 자문 배워 다음 오신 쿤과 마이케이르 이수혁 씨한테 궁금하

Voice Conversion with 박명수

모델: Diff-SVC

Singing Voice Conversion을 위한 Diffusion 기반 모델

논문의 Contribution

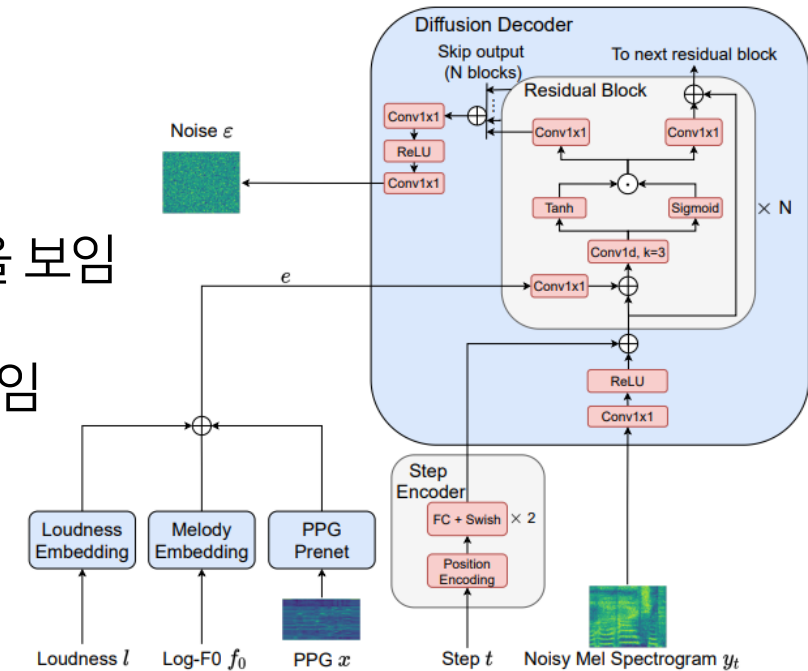
- ① Diffusion probabilistic model이 SVC task에 효과적으로 적용 가능함을 보임
- ② 목소리의 자연스러움과 유사성에서 이전 SVC 시스템보다 우수한 성능을 보임

DIFFSVC: A DIFFUSION PROBABILISTIC MODEL FOR SINGING VOICE CONVERSION

Songxiang Liu^{*1}, Yuwen Cao^{*1}, Dan Su², Helen Meng¹

¹Human-Computer Communications Laboratory, The Chinese University of Hong Kong

²Tencent AI Lab



Voice Conversion with 박명수

학습 방법

① 깃허브에 구현되어 있는 Diff-SVC 코드 사용 (<https://github.com/UtaUtaUtau/diff-svc>)
* Sampling rate: 44.1kHz

② 전처리를 통해 얻은 zip 파일 형식의 dataset 준비

③ Pretrain된 모델 파라미터 준비
(https://github.com/MLo7Ghinsan/MLo7_Diff-SVC_models)

④ 학습 시작

박명수김태호라디오_9.wav
박명수김태호라디오_10.wav
박명수김태호라디오_11.wav
박명수김태호라디오_12.wav
박명수김태호라디오_13.wav
박명수김태호라디오_14.wav
박명수김태호라디오_15.wav
박명수김태호라디오_16.wav
박명수김태호라디오_17.wav
박명수김태호라디오_18.wav
박명수김태호라디오_19.wav
박명수김태호라디오_20.wav
박명수김태호라디오_21.wav
박명수김태호라디오_22.wav
박명수김태호라디오_23.wav
박명수김태호라디오_24.wav
박명수김태호라디오_25.wav
박명수김태호라디오_26.wav
박명수김태호라디오_27.wav
박명수김태호라디오_28.wav
박명수김태호라디오_29.wav
박명수김태호라디오_30.wav
박명수김태호라디오_31.wav
박명수김태호라디오_32.wav
박명수김태호라디오_33.wav
박명수김태호라디오_34.wav
박명수김태호라디오_35.wav
박명수김태호라디오_36.wav
박명수김태호라디오_37.wav
박명수김태호라디오_38.wav
박명수김태호라디오_39.wav
박명수김태호라디오_40.wav

Voice Conversion with 박명수

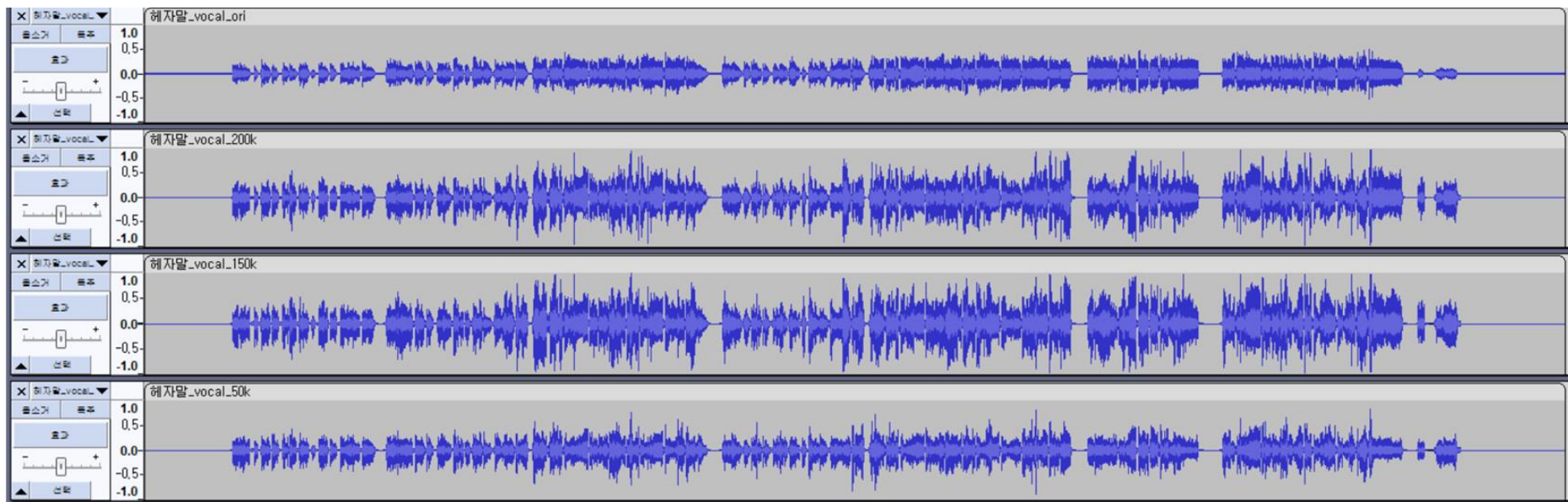
Inference 과정

- ① Train 때와 동일한 모델 사용
 - ② Fine-tuning된 모델의 checkpoint 및 config 파일 load
 - ③ Voice Conversion의 대상이 되는 오디오 load
 - ④ 출력 결과에 대한 파라미터 세부 설정
 - ⑤ zip 파일로 저장
-

Voice Conversion with 박명수

학습 과정 분석

- ① 학습이 적게 이루어질 때 (~50,000 step) 노이즈는 적지만 Voice Conversion이 제대로 이루어지지 않음
- ② 고음보다 저음 학습이 빨리 이루어짐
- ③ 학습이 이루어질수록 노이즈가 생기지만, 노래의 고음 부분 학습이 잘 이루어짐



Voice Conversion with 박명수

학습 과정 한계 & 어려웠던 점

- ① 가성 부분이 제대로 학습되지 않음
→ 학습 데이터 상에 박명수의 가성이 없어 생기는 한계
 - ② 박명수(남성)에 대한 학습이 진행되었기에 여성 대상 inferenc시에 제한적인 모습을 보임
(옥타브 자체가 낮아짐 / 멜로디의 변형 등)
 - ③ 두 번째 한계와 비슷하게, 음성의 key를 올렸을 때 박명수의 특성이 많이 사라짐
 - ④ 학습 및 추론 데이터의 형식(Sampling rate 등)이 맞지 않았을 때 결과물의 질 저하
 - ⑤ f0 calculation 알고리즘 사용 시 결과물의 질 저하
-

감사합니다