Vocoder

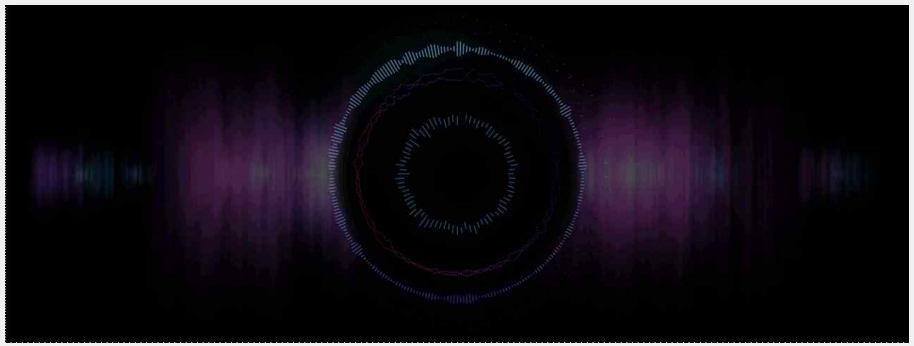
컴퓨터과학과
201511049 이현구

Artist - Kian Mosharaf

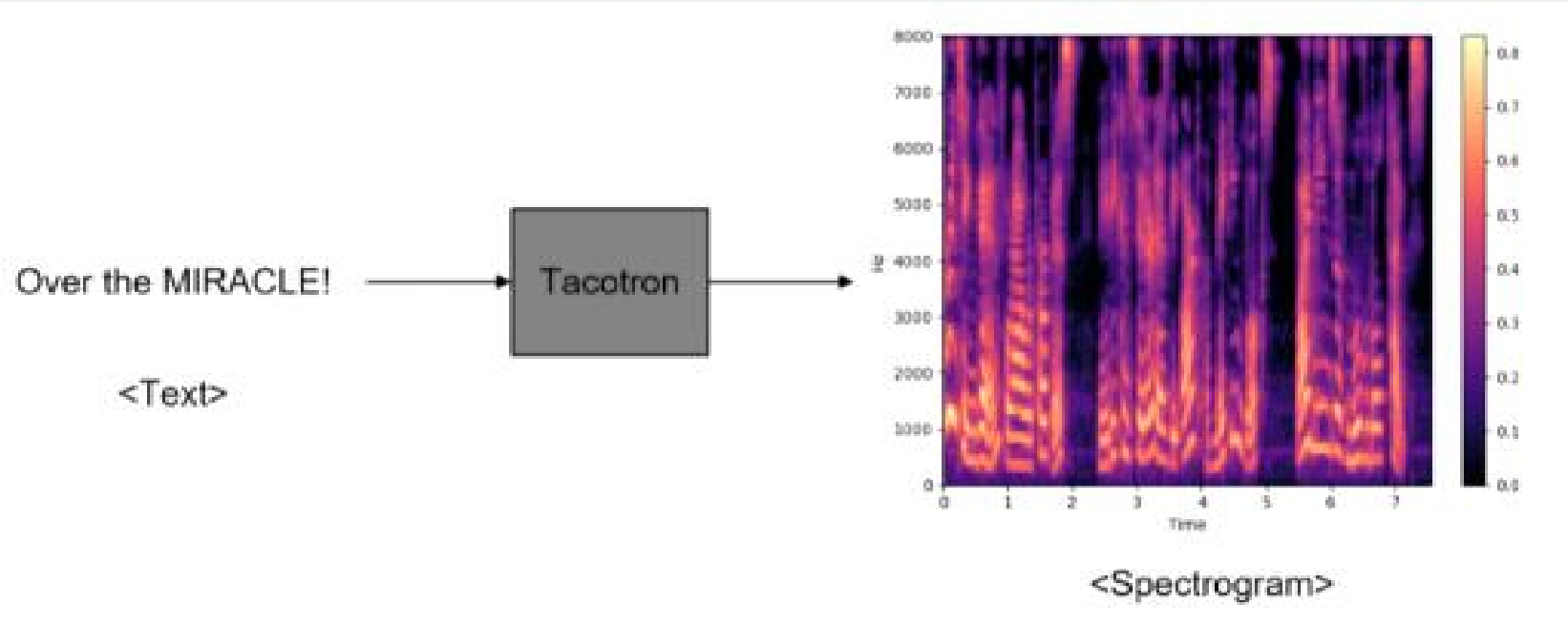
Vocoder → 신호 처리 알고리즘

시간에 따라 스펙트럼 특성이
어떻게 변하는 지 측정하여 음성을 검사

합성된 음성의 음질에 직접적인 영향을 준다
알맞은 Vocoder를 찾는 것이 중요하다



TTS? **Vocoder?** **Mel?**



TTS?　　Vocoder?　　Mel?

Tacotron, Tacotron-2, Deep Voice (1, 2, 3), DC-TTS

문장을 입력으로 받아 직접적으로 오디오 파형을 출력하는 것이 아닌
음성의 **Feature**들(주로 **Mel Spectrogram**)을 반환한다.

Vocoder는 이를 받아서 실제 음성의 파형으로 변환하는 역할을 한다.

Vocoder의 종류

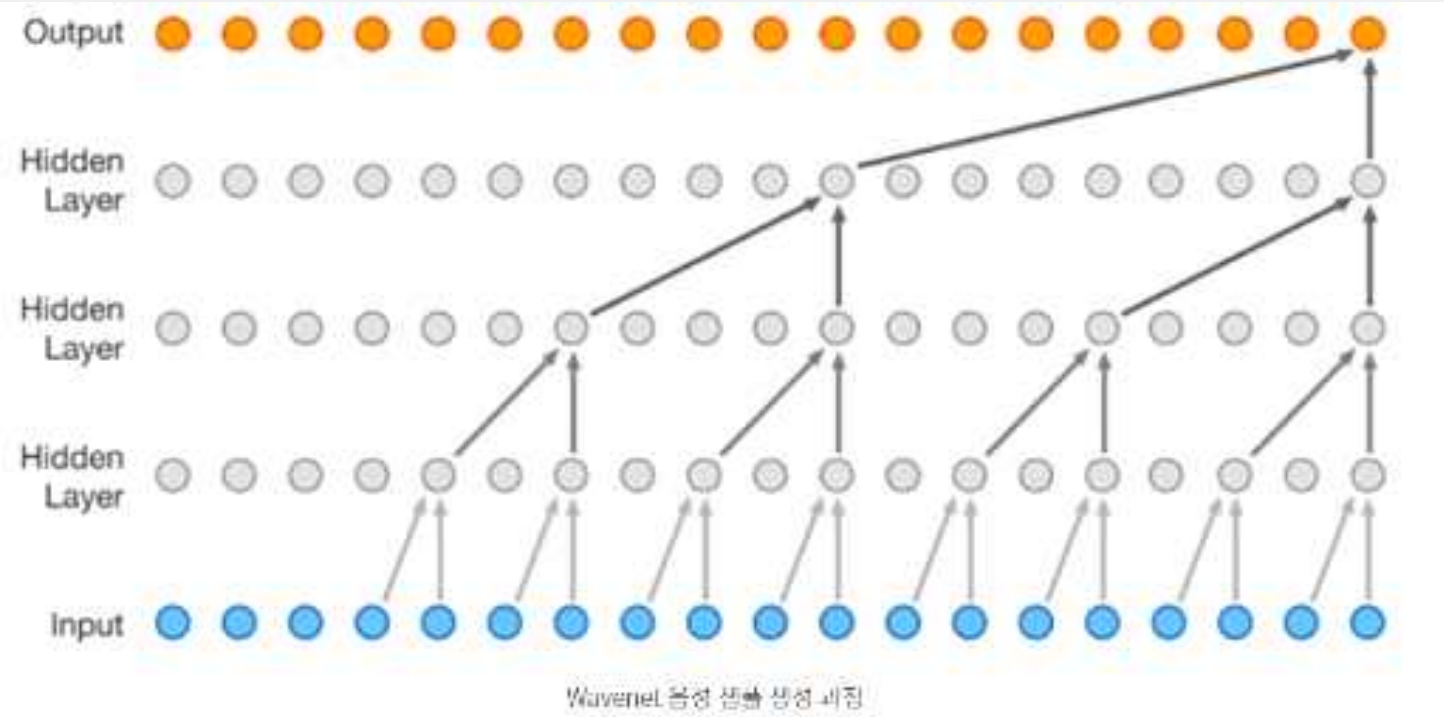
wavenet

2016년에 구글에서 발표한 뉴럴 보코더 모델

음성 샘플들 간의 순차적 특징을 이용하는 자기회귀(Autoregressive) 모델

이전 샘플들을 이용해 다음 샘플을 예측하는 방법으로 고품질의 음성을 합성
하지만 생성 속도가 매우 느림
1초의 음성을 생성하는데 1시간이 넘게 걸린다고.. (GTX 1080Ti기준)

wavenet



Parallel Wavenet

2017년 구글에서 개발한 모델

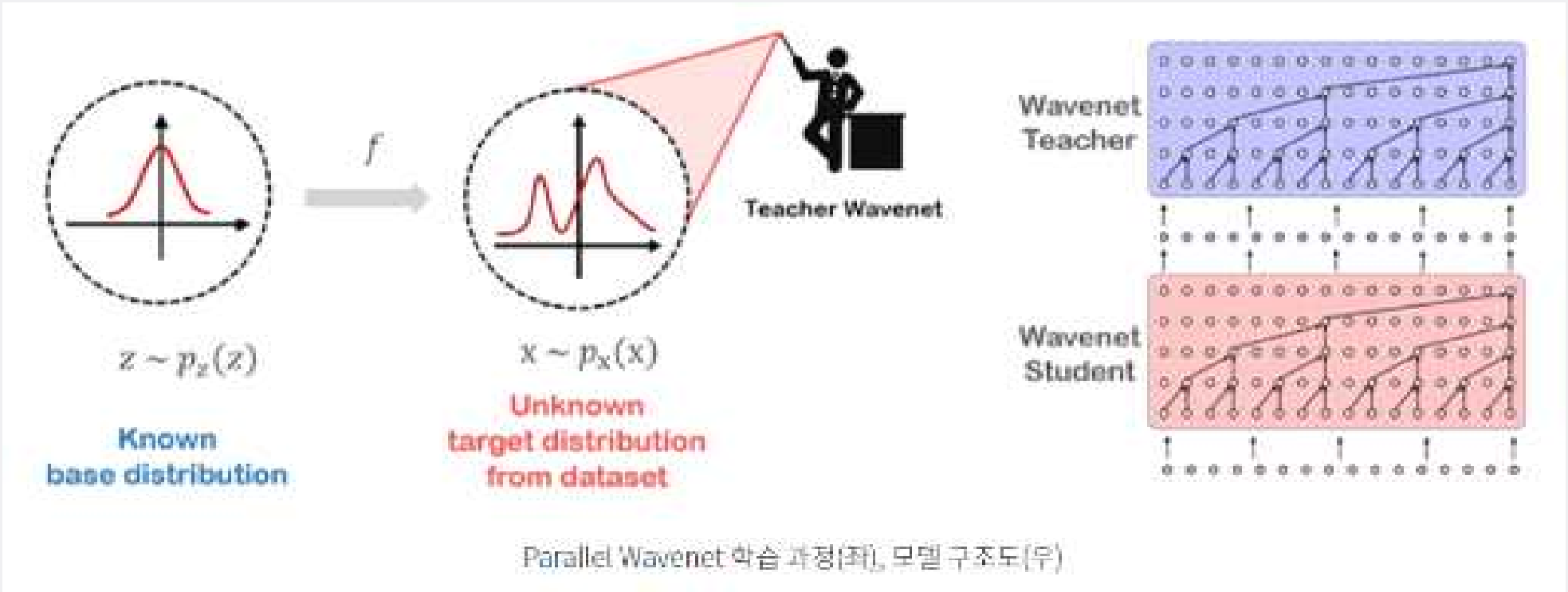
Wavenet의 느린 샘플 생성 속도를 보완하기 위해 고안

IAF(Inverse Autoregressive Flow) 모델을 이용해 음성을 합성

IAF 모델 → student network

Wavenet 모델 → teacher network

Parallel Wavenet



Parallel Wavenet 학습 과정(좌), 모델 구조도(우)

Parallel Wavenet

Wavenet보다 음성 합성 속도가 빠르다

Wavenet보다 합성 음성의 품질이 떨어진다

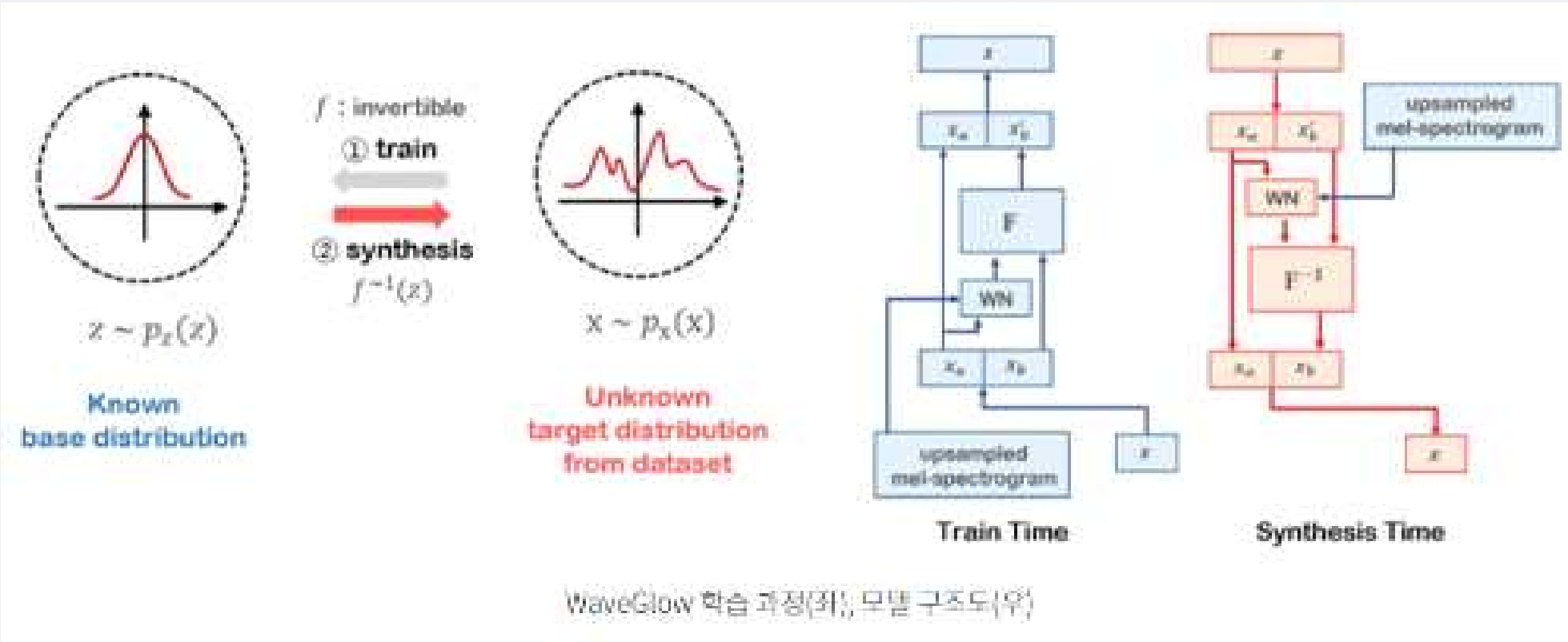
잘 학습된 teacher network를 먼저 학습시켜야 한다

WaveGlow

NVIDIA가 WaveNet에 Generative Model인 Glow를 합쳤다

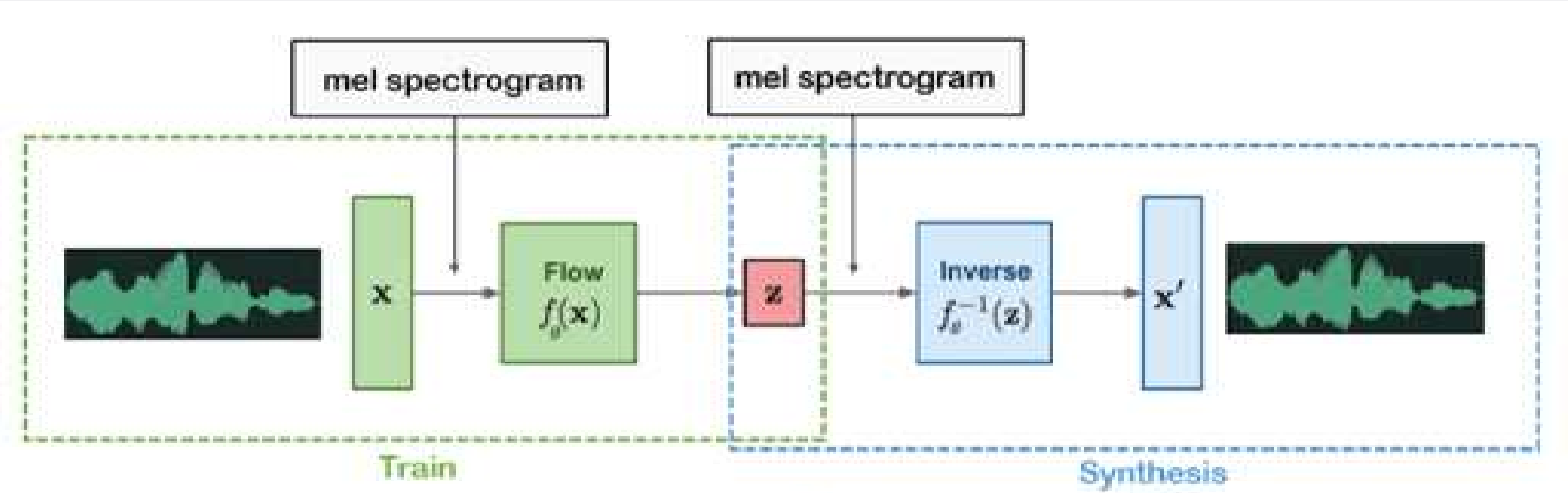
Normalizing Flow 기반의 뉴럴 보코더

WaveGlow



WaveGlow 학습 과정(좌), 모델 구조도(우)

WaveGlow



WaveGlow의 전체 과정

WaveGlow

WaveGlow는 Parallel Wavenet과 달리
사전에 학습된 teacher network가 필요가 없으며 **빠르게 음성을 합성할 수 있다**

분포를 기반으로 하는 손실(Loss) 함수를 이용하기 때문에
합성된 **음성의 품질이 다소 떨어질 수 있다**
TTS 시스템과 결합될 경우, 텍스트로부터 예측한 멜 스펙트럼의 품질에 따라
합성된 음성의 품질이 좌우되는 문제가 생긴다

GAN 기반 vocoder

MelGAN,

ParallelWaveGAN , VocGAN <- 오픈소스 없음

melgan은 mel spectrogram을 입력받아서 오디오 신호를 생성해내는 GAN 기반 보코더

Generative Adversarial Network = 인공지능 알고리즘

git 자료

wavenet

https://github.com/r9y9/wavenet_vocoder

waveglow

https://github.com/NVIDIA/waveglow

tacotron + deepvoice

https://github.com/carpedm20/multi-speaker-tacotron-tensorflow

MelGAN

https://github.com/descriptinc/melgan-neurips

음성합성 실습 강의자료

실습 영상 (SKplanet Tacademy)

https://youtu.be/vfzjfuZwRTY

실습 코드(Google Colab)

https://drive.google.com/file/d/1N6_joigWXgoDHfE1rm4rL2kBkmCpXW2H/edit

음성합성 실습 강의자료


[T아카데미] 음성합성 실습 (full code)
+ 코드 + 텍스트
필수 library 설치
PyTorch (1.5.1)
torchaudio
!pip install unidecode
!pip install torchaudio
Code (WaveGlow)
Reference
WaveGlow (implemented by NVIDIA): https://github.com/NVIDIA/waveglow

참고 논문

VocGan 논문: https://arxiv.org/abs/2007.15256

WaveGlow 논문: https://arxiv.org/abs/1811.00002

GAN 논문: https://arxiv.org/abs/1406.2661

MelGAN 논문: https://arxiv.org/abs/1910.06711

Parallel WaveGAN 논문: https://arxiv.org/abs/1910.11480