



Tacotron2 model



tacotron2

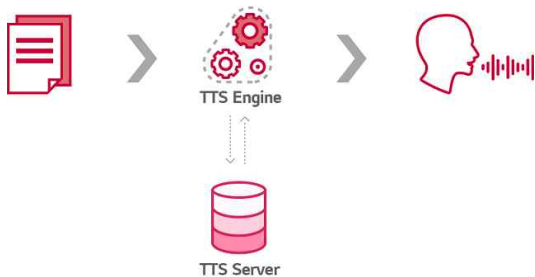
S시스

otron 2

2-1 tacotron2개념

2-2 tacotron2학습방법

3 tacotron2활용 방법



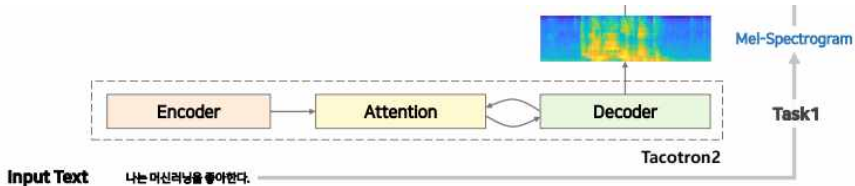
TTS - Text To Speech

- 텍스트를 입력하면 목소리로 변환해주는 기술
- 기계와 인간이 대화할 수 있도록 하기 위한 기초기술
- 영어, 숫자, 한글 등의 문자를 입력하면 자연스러운 인간의 음성으로 출력



- Mel - spectrogram으로부터 음성을 합성하는 단계

Tacotron2 - 개념 및 구조



Tacotron 2

: 음성 합성에 대표적인 모델로 고품질의 음성을 생성할 수 있는 딥러닝 기반 TTS 모델

: input - character / output - Mel-Spectrogram

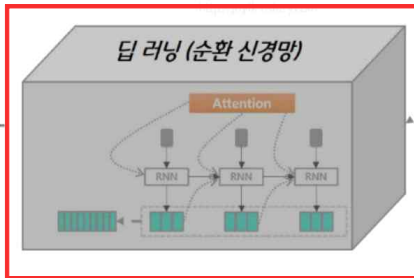
- Encoder: character로부터 음성에 특징을 추출하는 것
- Attention: 매 시점 Decoder에서 사용할 정보를 추출하고 할당
- Decoder: Attention에서 얻은 정보와 이전 시점에서 생성된 mel-spectrogram을 이용해
 - (1) 현재시점의 Mel-spectrogram을 생성
 - (2) 현재시점의 종료확률을 계산
 - (3) Mel-spectrogram의 품질을 향상

"6월 6일"

(텍스트)

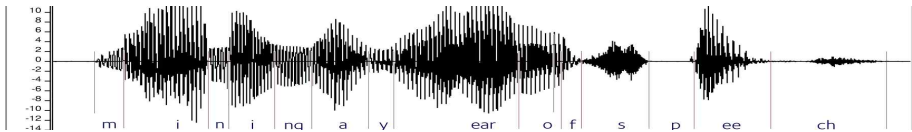


(음성)



순환 신경망(RNN)

: 시간의 흐름에 따라 변화하는 데이터를 학습하기 위한 인공지능망



KSS 데이터셋

KSS 데이터셋은 전문 여성 성우 한 분이 한글과 한영사전 4권의 예문을 읽은 약 12시간 분량의 데이터셋입니다. 이 데이터셋을 사전학습한 후 개발자의 녹음된 음성을 이용해 이전에 학습된 모델 가중치로부터 학습을 업데이트할 수 있습니다.

감사합니다