MARATRON

책을 듣다

Name 이남기

Date 2019.03 ~ 2019.06

Git https://github.com/brainmining-for-bigdata

Index




Intro

Service

Model

Conclusion

Intro

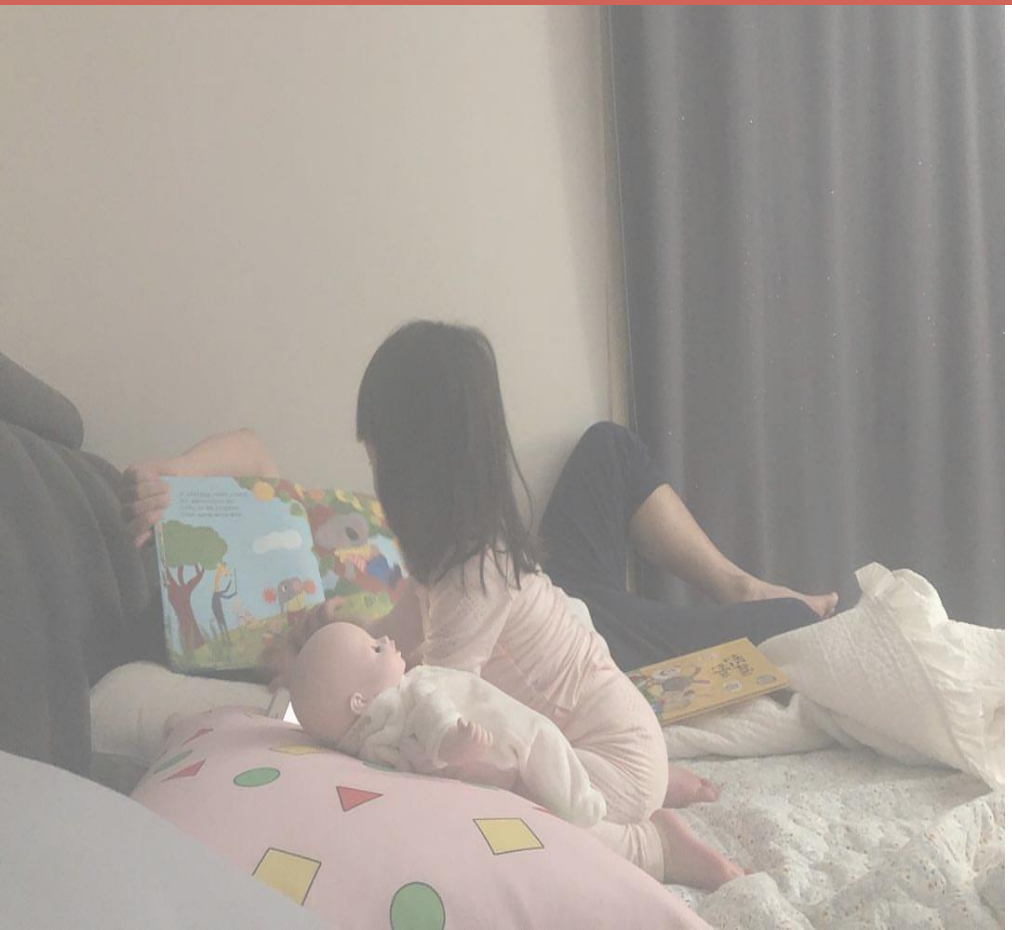
Background

발디딜 틈도 없는 지하철.
간신히 이어폰을 꺼내어 귀에 꽂습니다.

들려오는 목소리.....

「내가 읽고 싶은 책을 누군가 아름다운 목소리로 읽어준다면 얼마나 좋을까요?」

Background



퇴근 후 매일 밤 아이가 잠들기 전 책을 읽어주고 싶지만.

쏟아지는 피곤에 정작 먼저 잠드는건 아빠네요...

「아이에게 엄마아빠의 목소리로 책을 읽어 준다면? 」

Summary

프로젝트명 : MARATRON

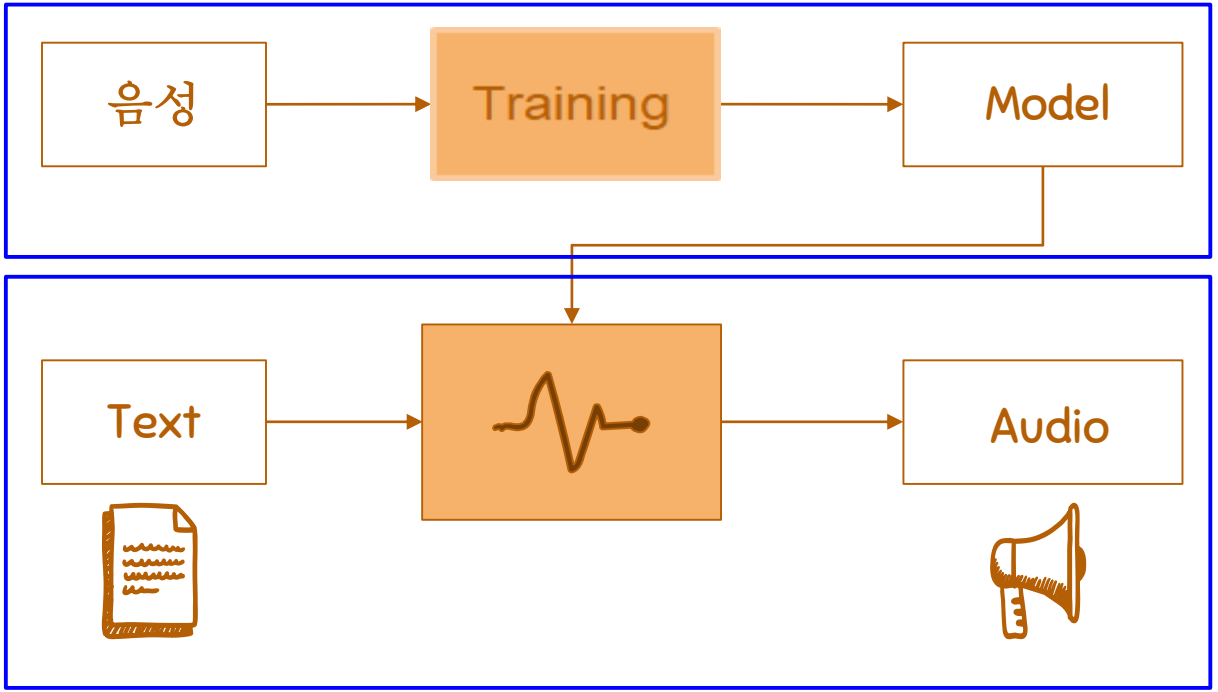
서비스명 : 책을 듣다.

프로젝트 정의 : 개인화 음성 합성 기술 TTS(Text to Speech)를 기반으로 목소리를 분석해 학습된 목소리로 글을 읽어주는 웹 서비스

활동기간 : 2019.03.12 ~ 2019.06.22

Concept

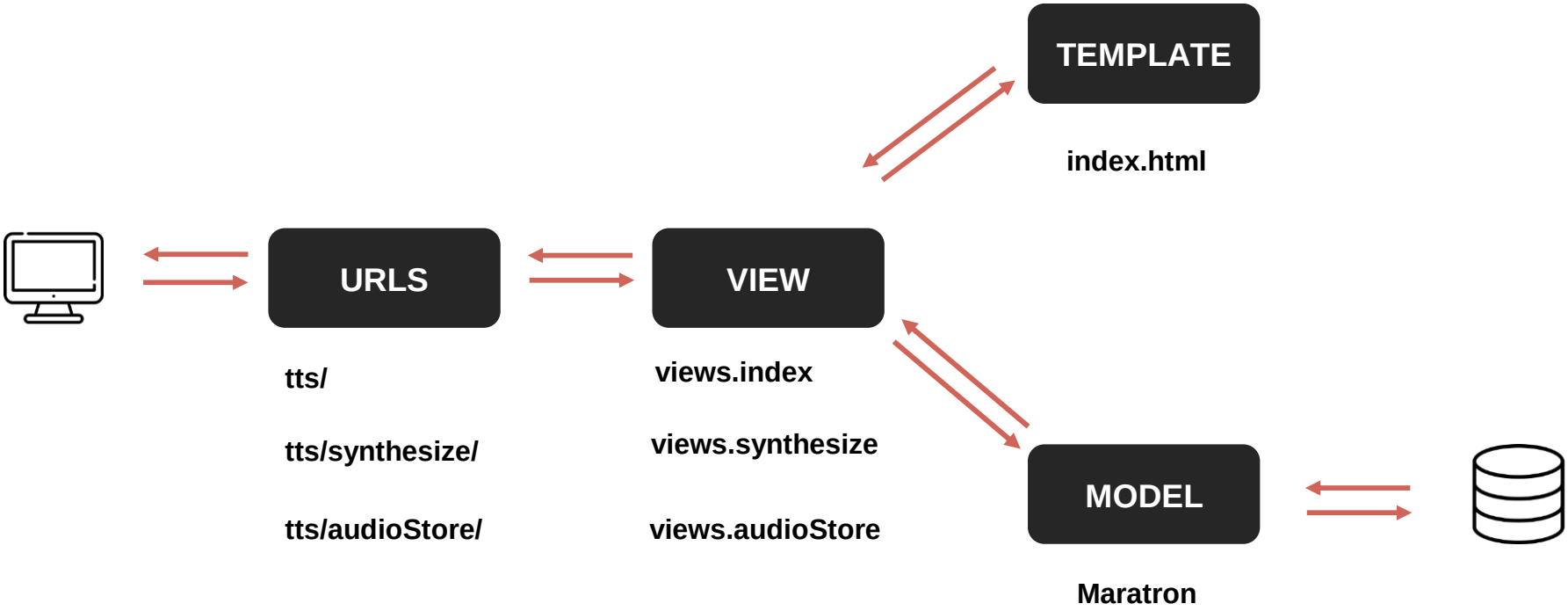
충분한 시간의 음성만
있다면 텍스트를
원하는 음성으로 합성



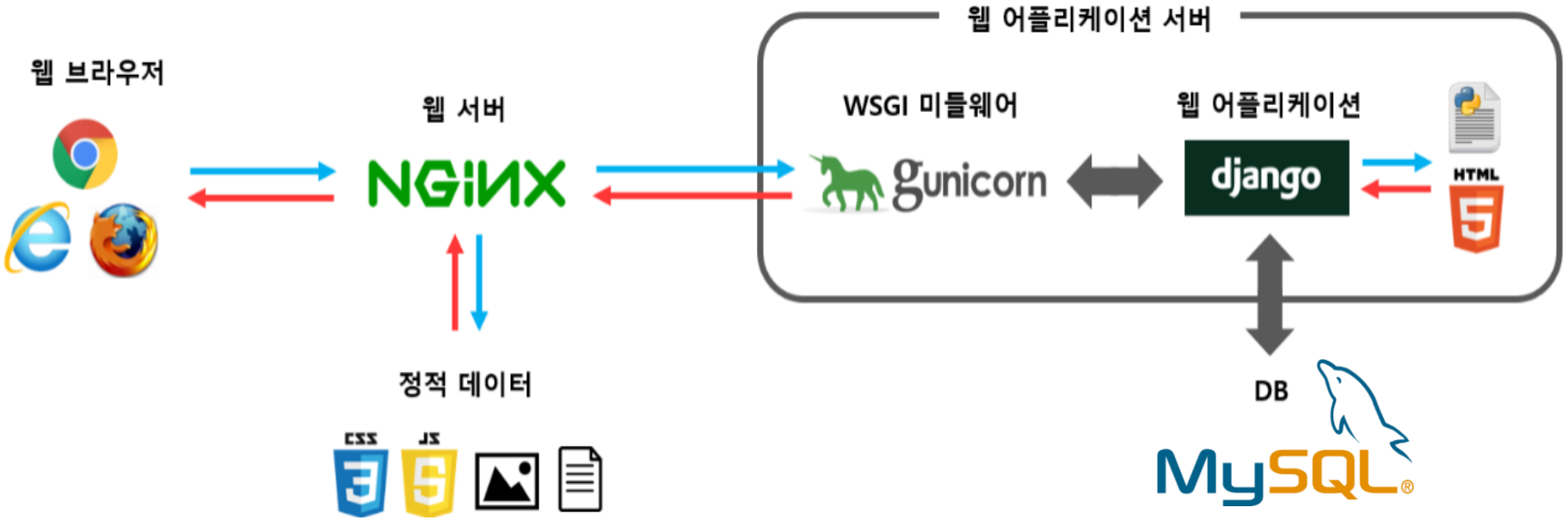
Technology Stack

Collaboration	slack, Google Drive, OvenApp.io, GitHub
Front-End	HTML, JS, CSS, AJAX, jQuery
Back-End	python, TensorFlow, django
Database	MySQL
Infra	docker, Google Cloud, NGINX
Development	Visual Studio Code

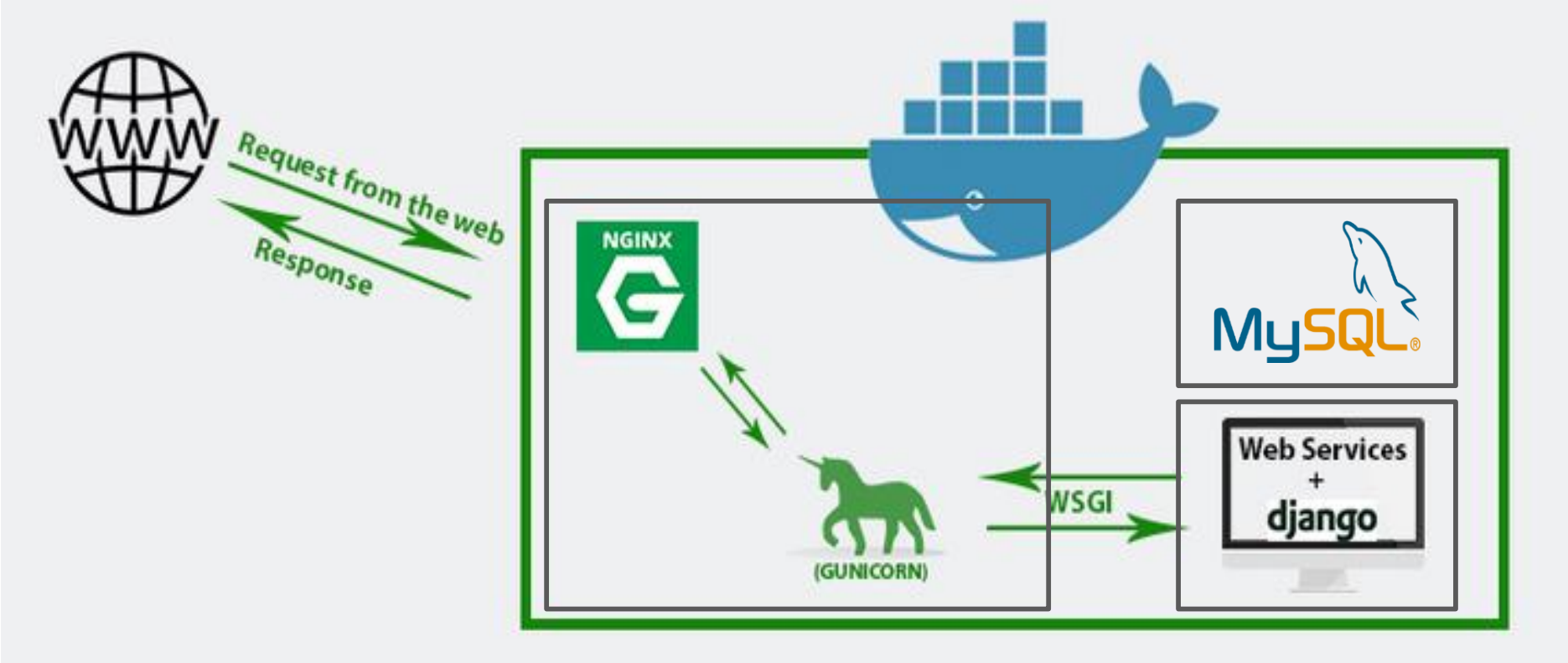
Django MVT Pattern



Deployment : Django Server Architecture



Deployment : Docker 사용 - 3개의 container로 구성

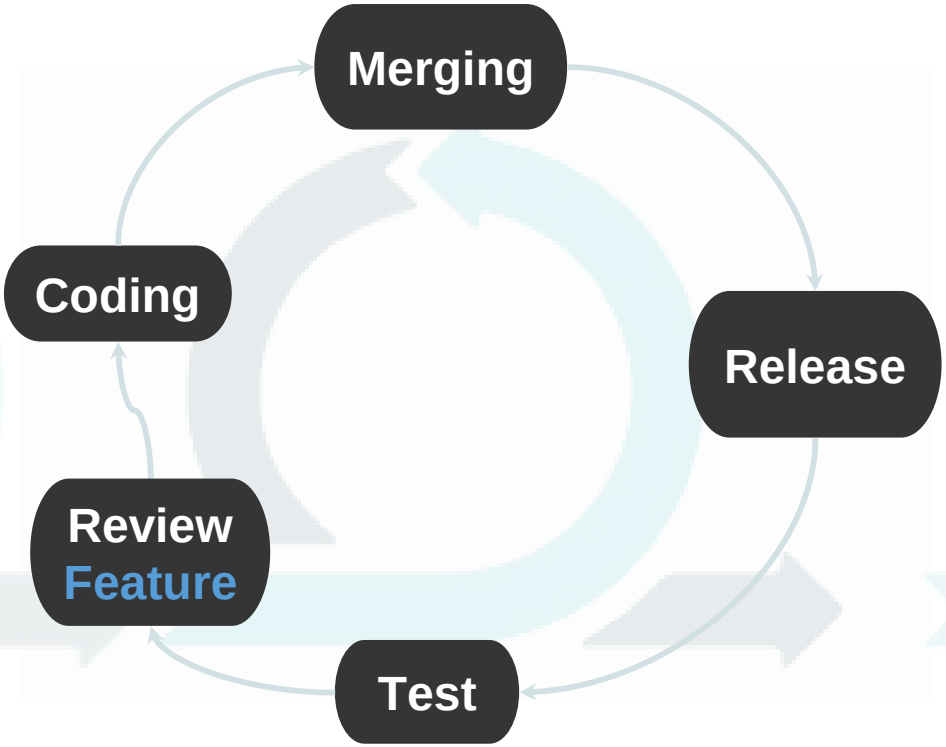


Service Release : 배포 플랫폼 선정

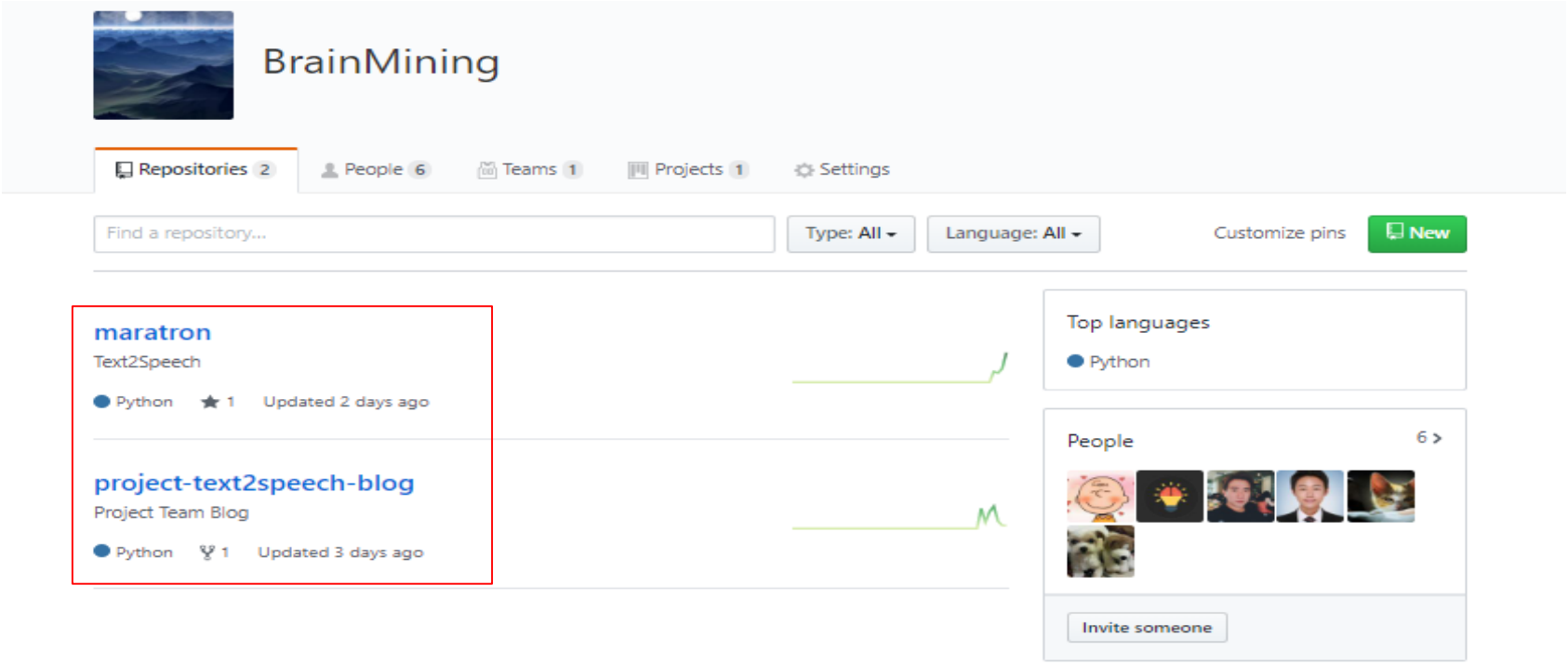
python™

django

Management : Agile



Management : GitHub Organization account

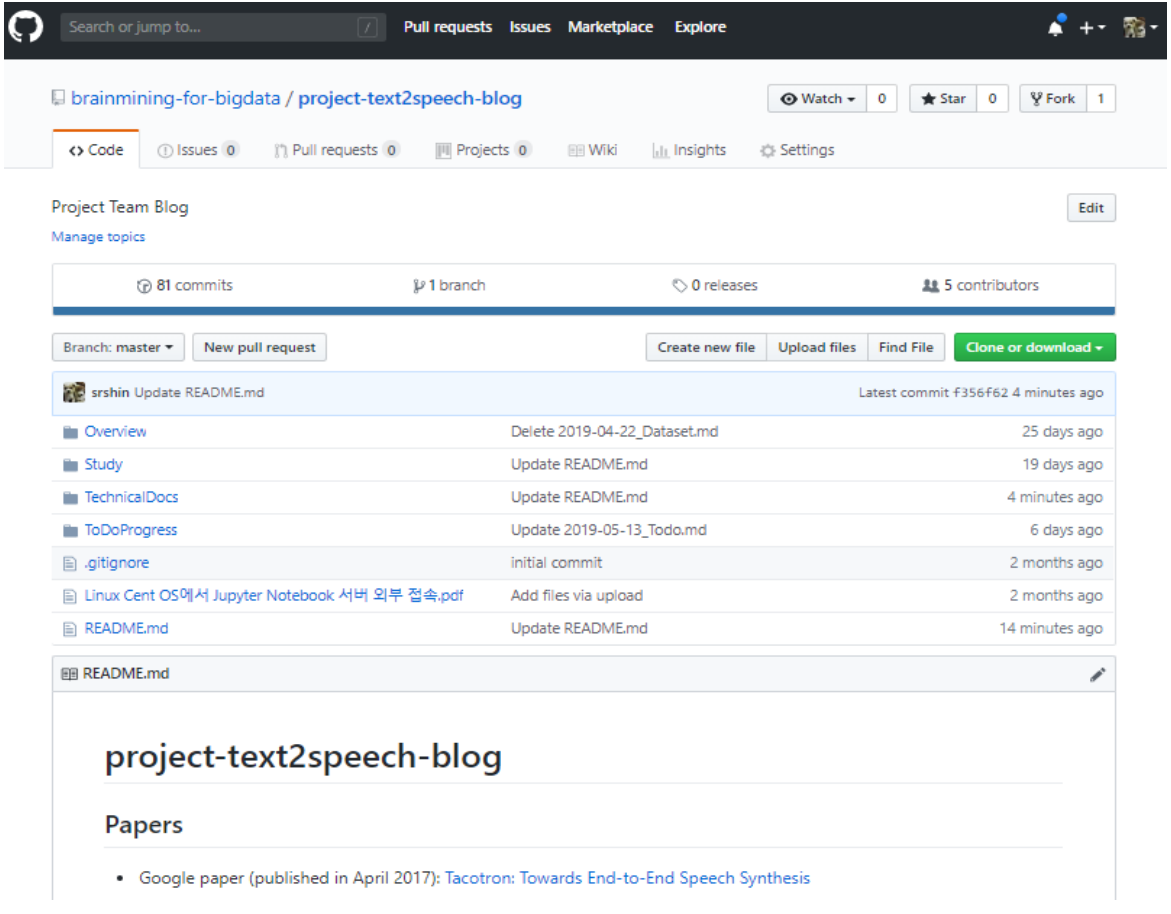


https://github.com/brainmining-for-bigdata

Management: Documentation

Project Blog

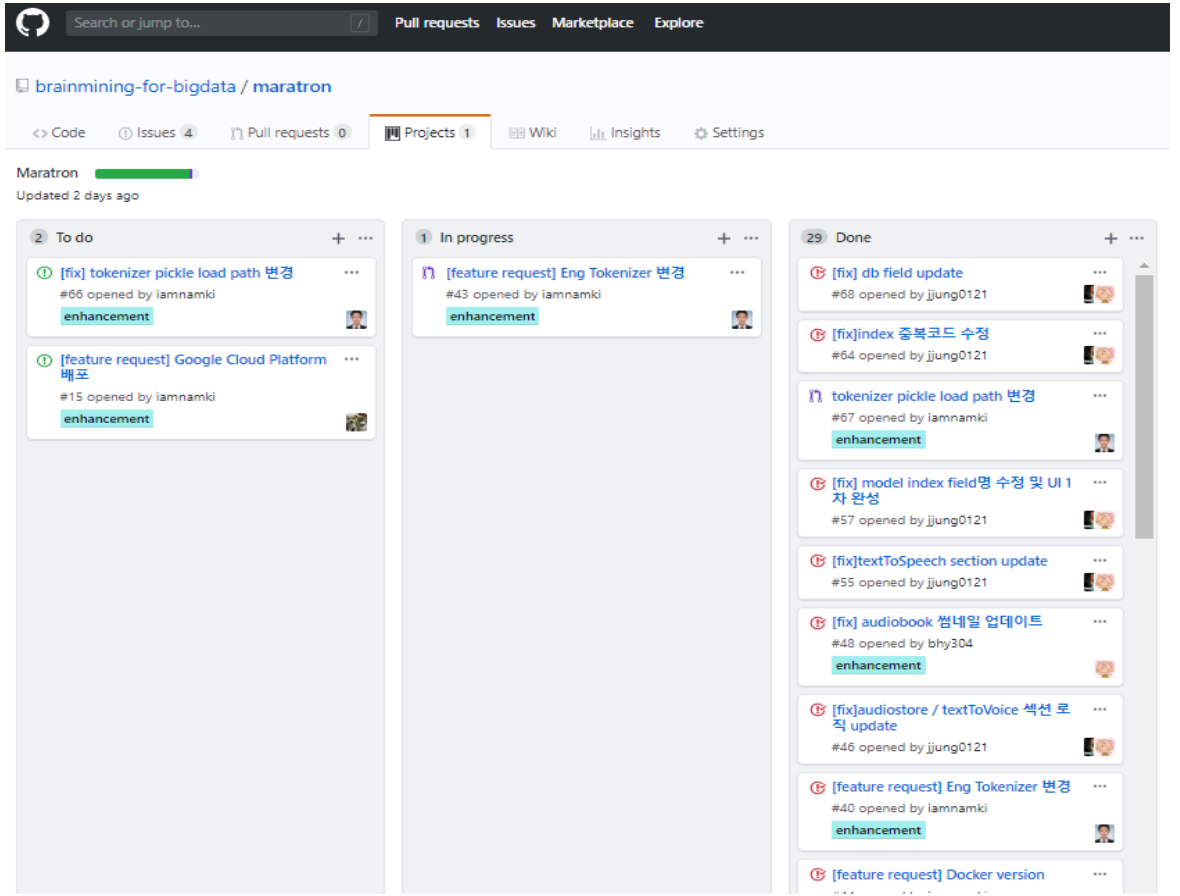
- Overview
- Study
- Technical Docs
- ToDoProgress : 회의록



Management: Project Management

Kanban System

- Todo
- In progress
- Done



Management : Issue Tracking

Issue Type

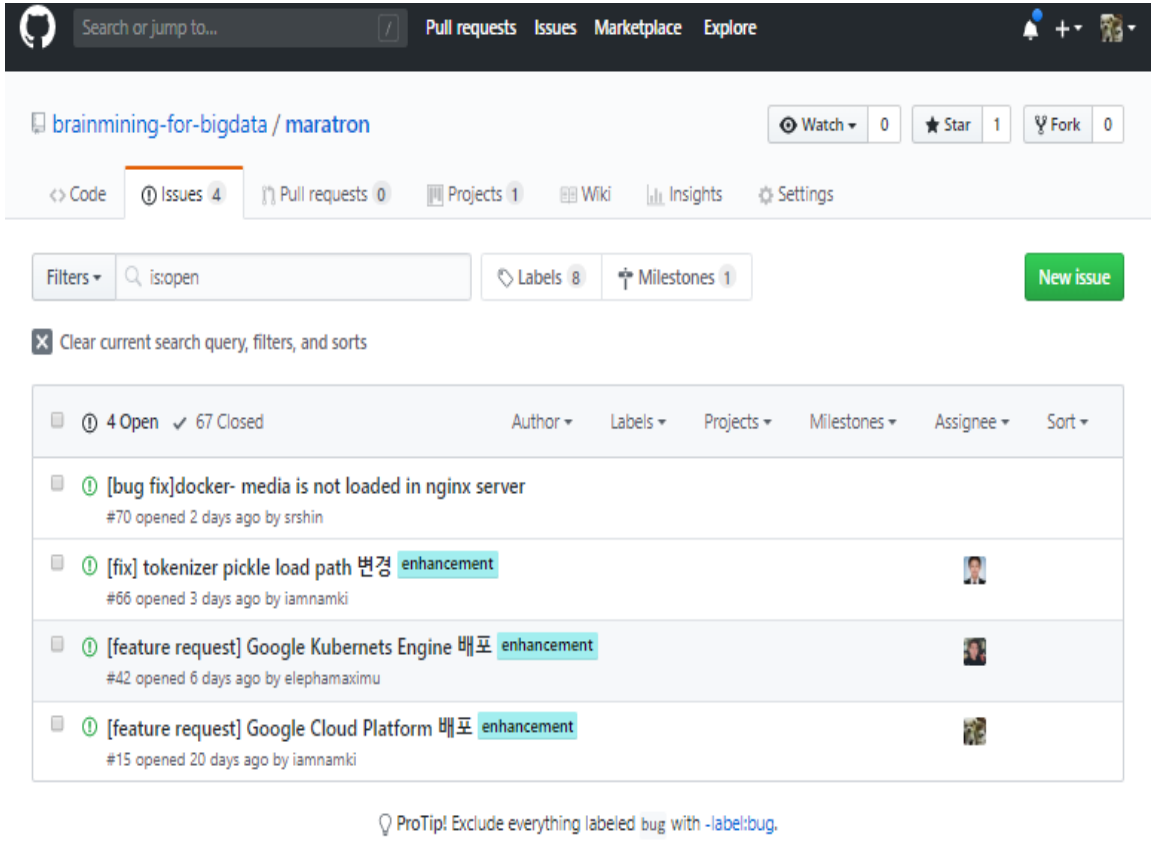
- [Feature Request]
- [Bug Fix]

Issue Status

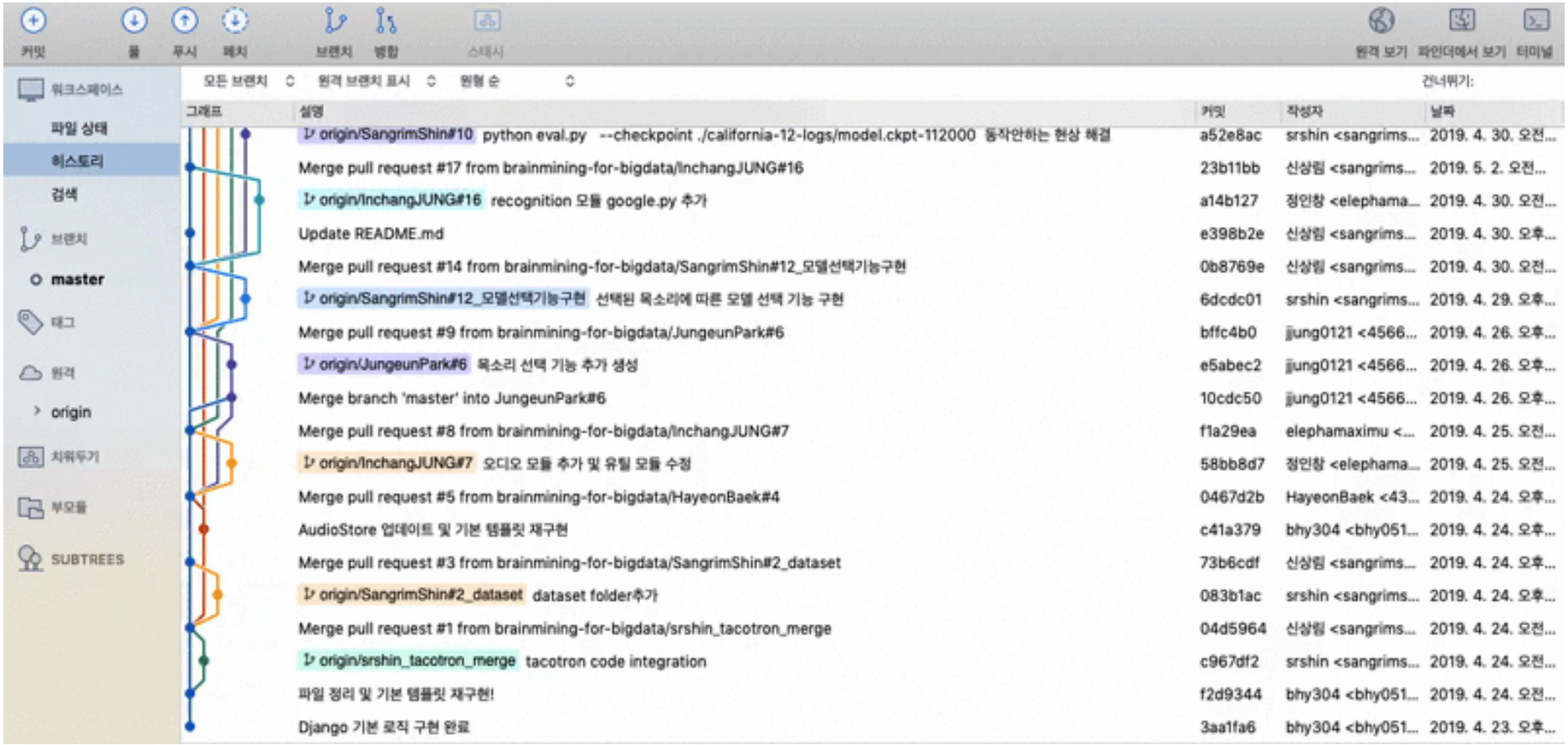
- Open
- Closed

Integration

- Kanban 과 연동
- Commit과 연계

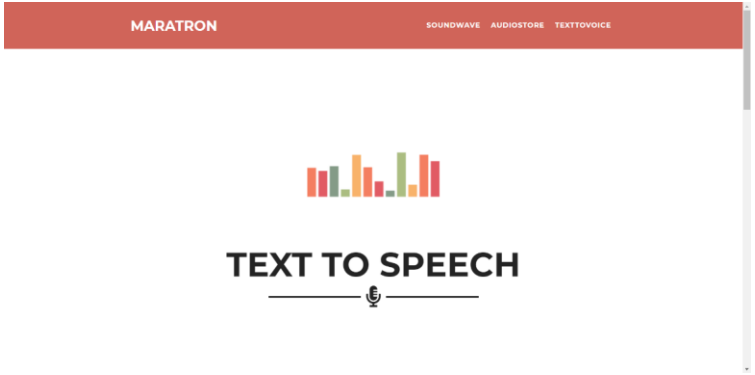


Management : Issue Tracking



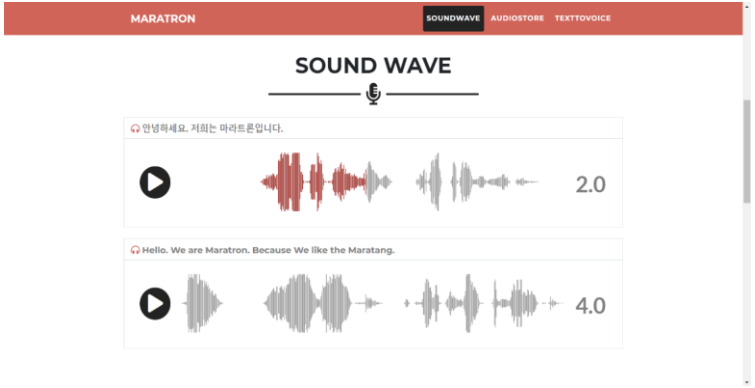
Service

Implement



MARATRON

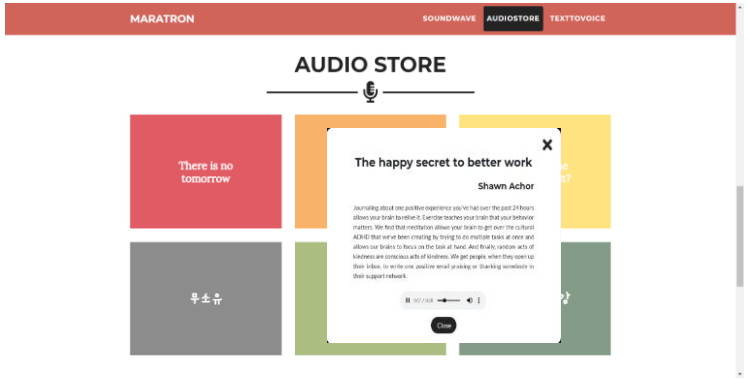
AUDIO STORE , SOUND WAVE , TEXT TO VOICE 메뉴 선택 가능.



SOUND WAVE

Training되어 있는 목소리의 샘플 오디오를 통해 목소리를 확인할 수 있다.

Implement



AUDIO STORE

미리 등록되어 있는 오디오 컨텐츠를 들을 수 있다.

TEXT TO VOICE

Voice Type을 선택 - 읽어주길 원하는 Text 입력 - Synthesize - 오디오제공

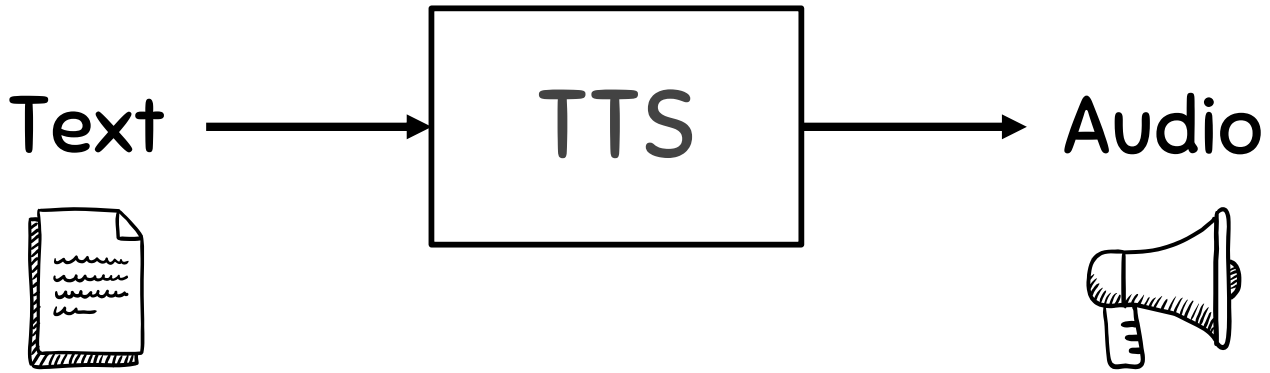
Model

Model : Dataset

Voice Type	English_Female	English_Male	Korean_Female
Total Size	2.6 GB	567 MB (zip)	3 GB
Total Length	24 hours	20 hours	12 hours
Number of audio-clips	13,100	1 audio file	12,853
Training Check-Point	40,000	112,000	64,000

Model

TEXT-TO-SPEECH : Concept



Model : Paper

TACOTRON: TOWARDS END-TO-END SPEECH SYNTHESIS

Google paper (published in April 2017)

TACOTRON: TOWARDS END-TO-END SPEECH SYNTHESIS

Yuxuan Wang*, RJ Skerry-Ryan*, Daisy Stanton, Yonghui Wu, Ron J. Weiss†, Navdeep Jaitly,

Zongheng Yang, Ying Xiao*, Zhifeng Chen, Samy Bengio†, Quoc Le, Yannis Agiomyrgiannakis,

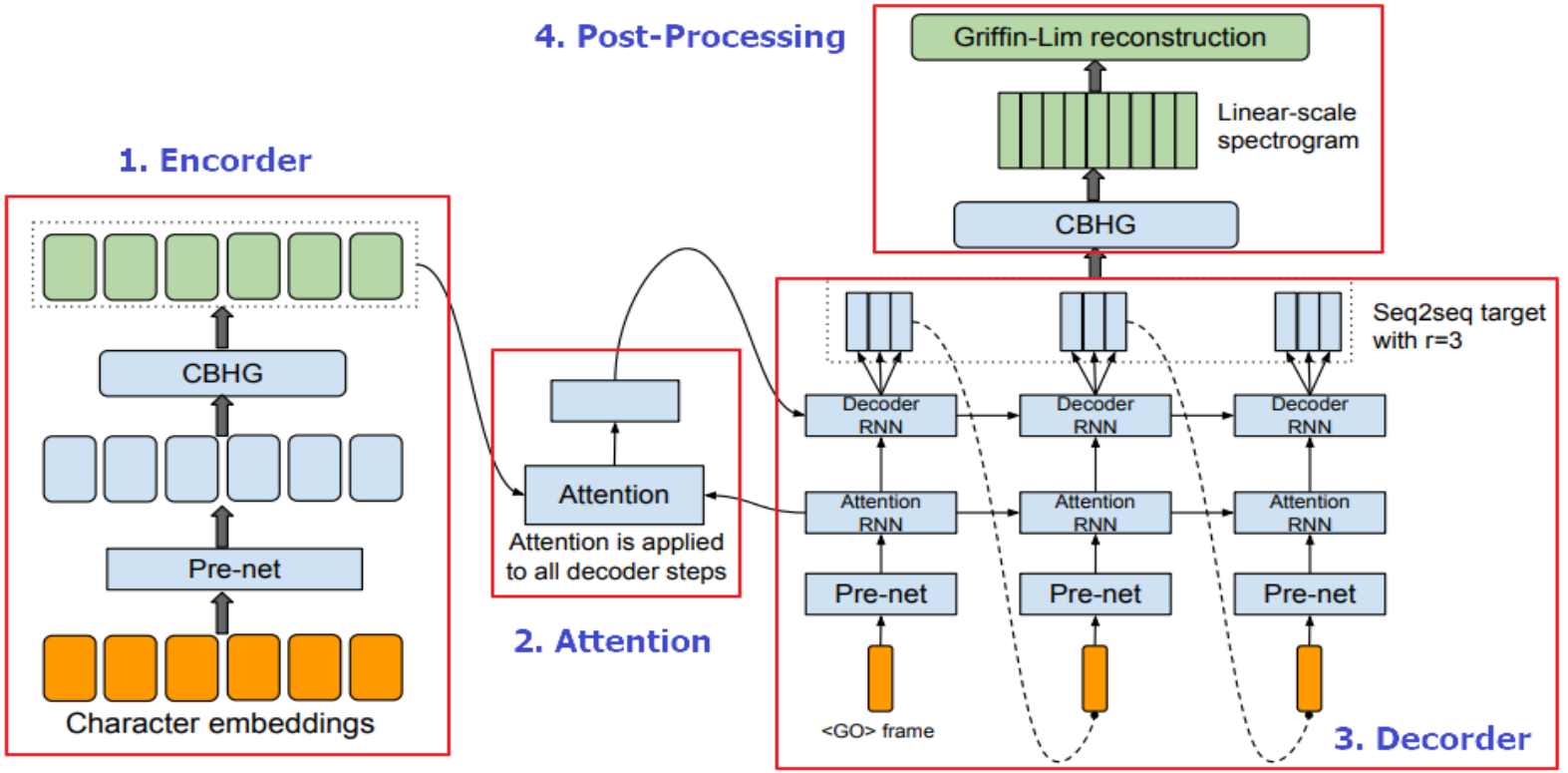
Rob Clark, Rif A. Saurous*

Google, Inc.
{yxwang,rjryan,rif}@google.com

ABSTRACT

A text-to-speech synthesis system typically consists of multiple stages, such as a text analysis frontend, an acoustic model and an audio synthesis module. Building these components often requires extensive domain expertise and may contain

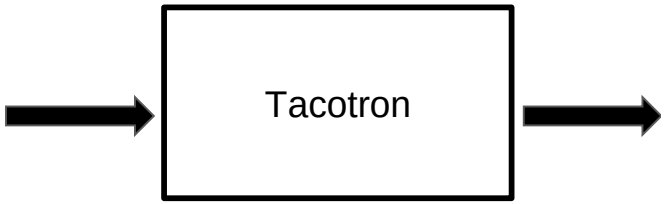
Model : Tacotron

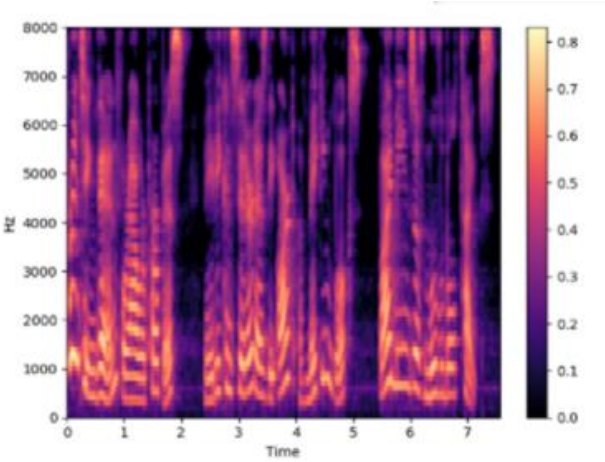


Model : Training Model

Over the MIRACLE

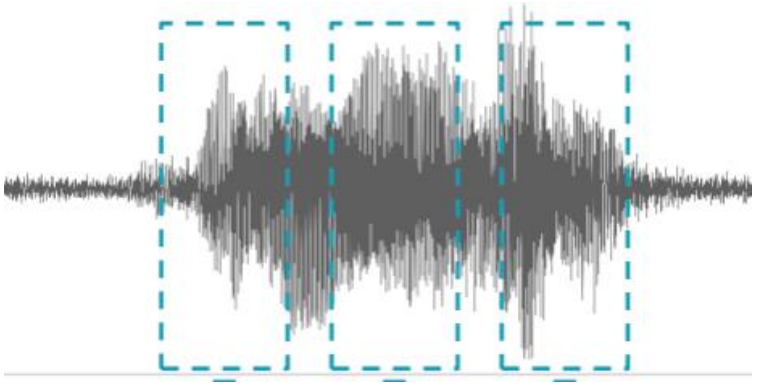
<Text>

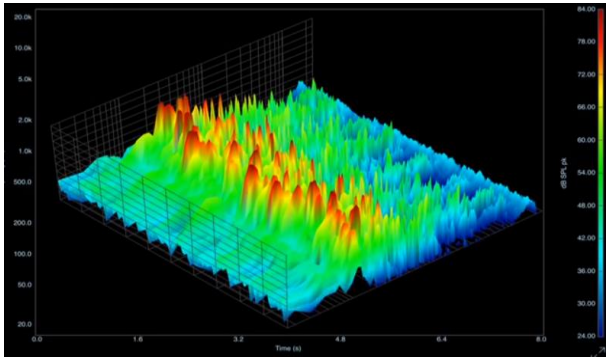


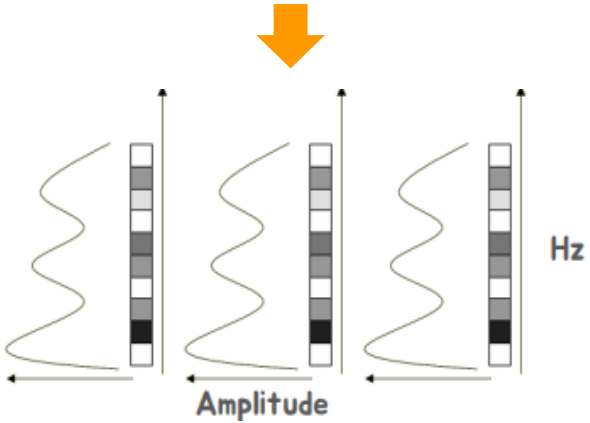


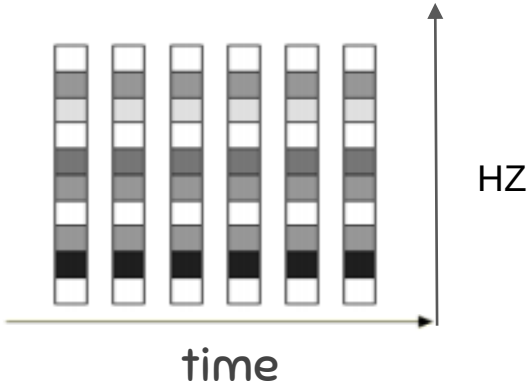
<spectrogram>

Model : Spectrogram - 소리를 형상화

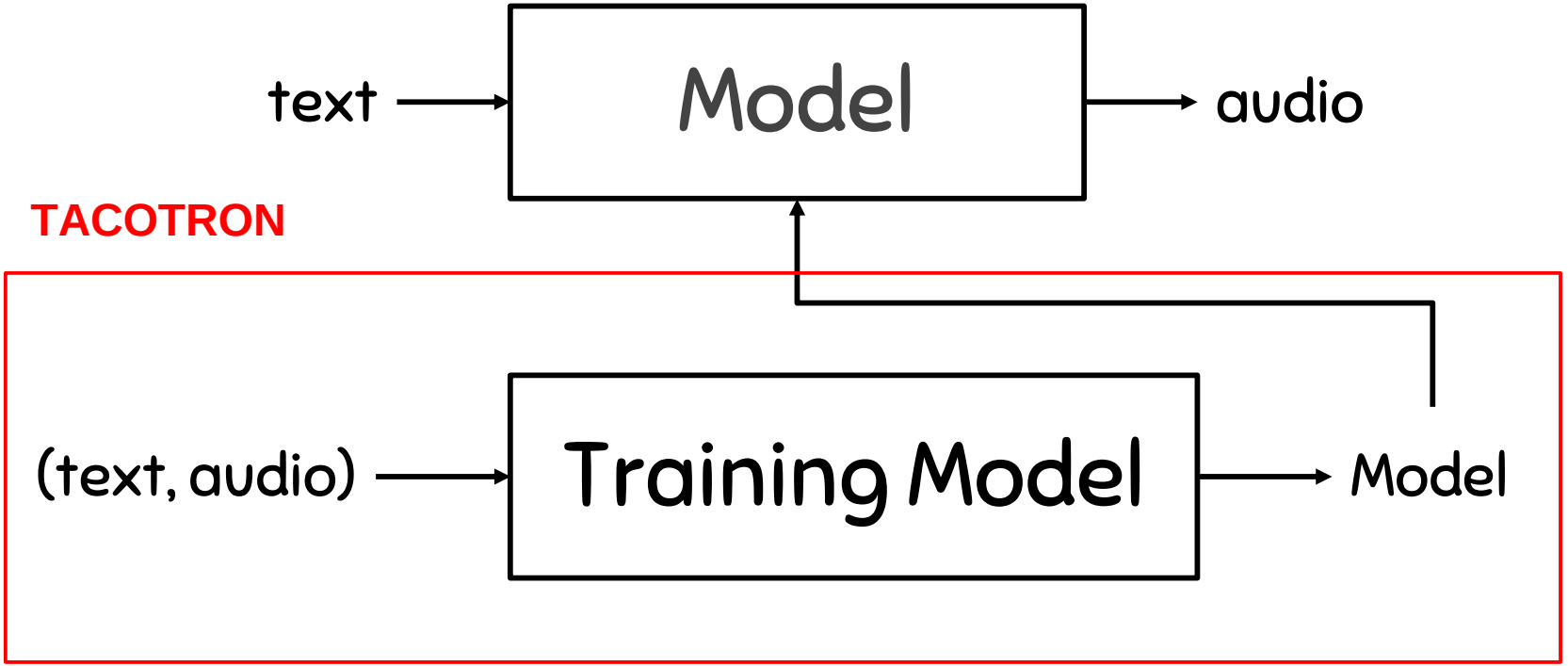




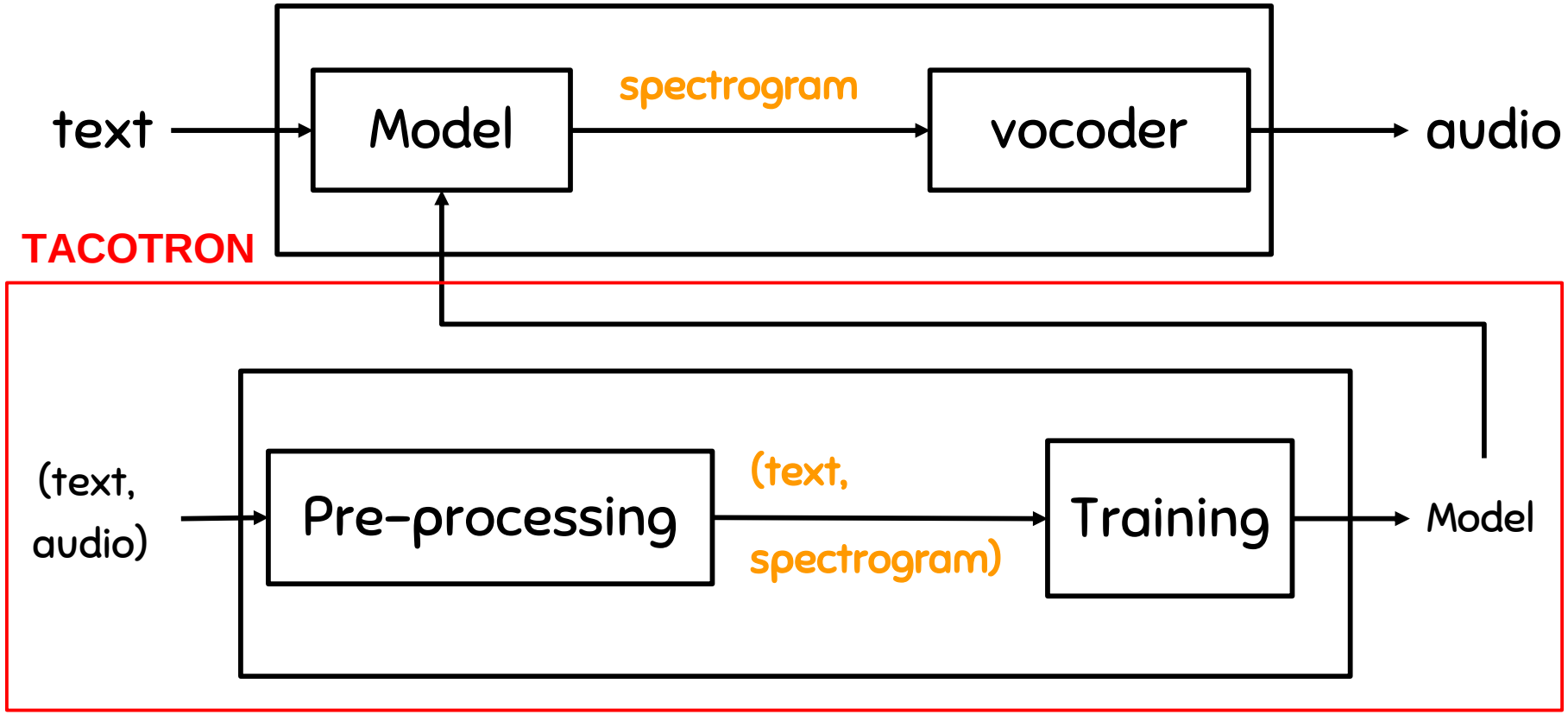




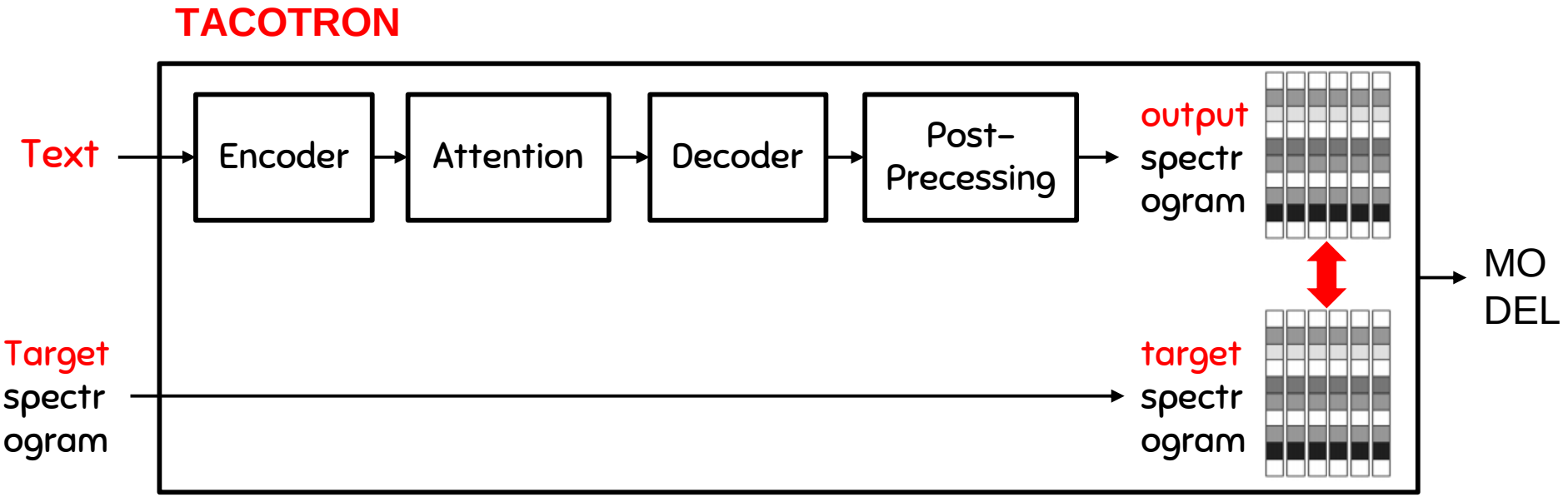
Model : structure



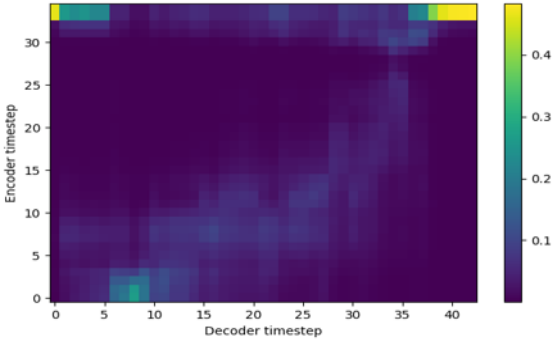
Model : structure



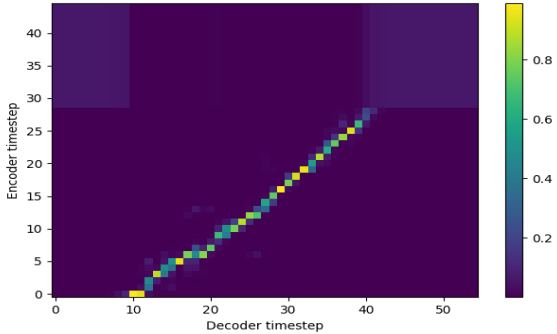
Model : Tacotron



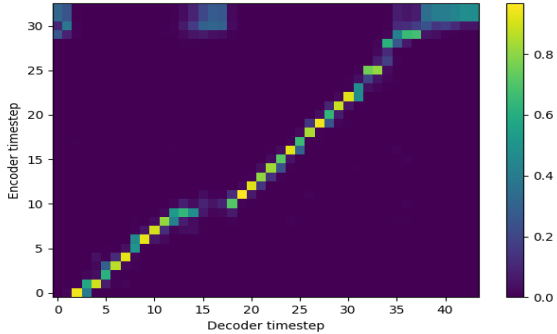
Training Model : linear Spectrogram



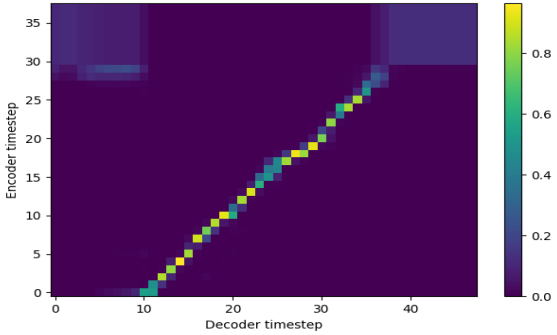
tacotron, None, 2019-05-03 18:51, step=5000, loss=0.12108



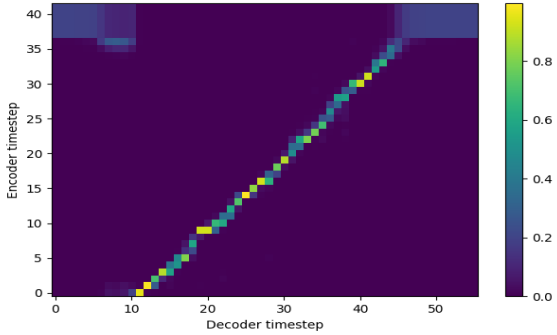
tacotron, None, 2019-05-04 01:44, step=10000, loss=0.10240



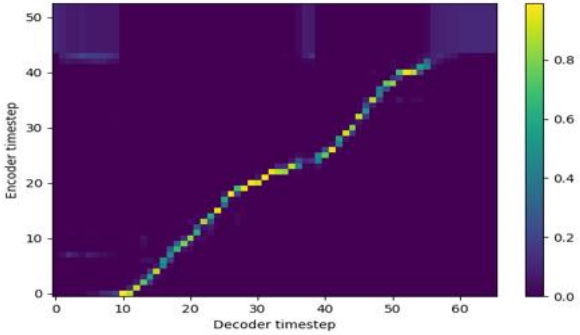
tacotron, None, 2019-05-05 05:49, step=30000, loss=0.09098



tacotron, None, 2019-05-05 20:09, step=40000, loss=0.08470

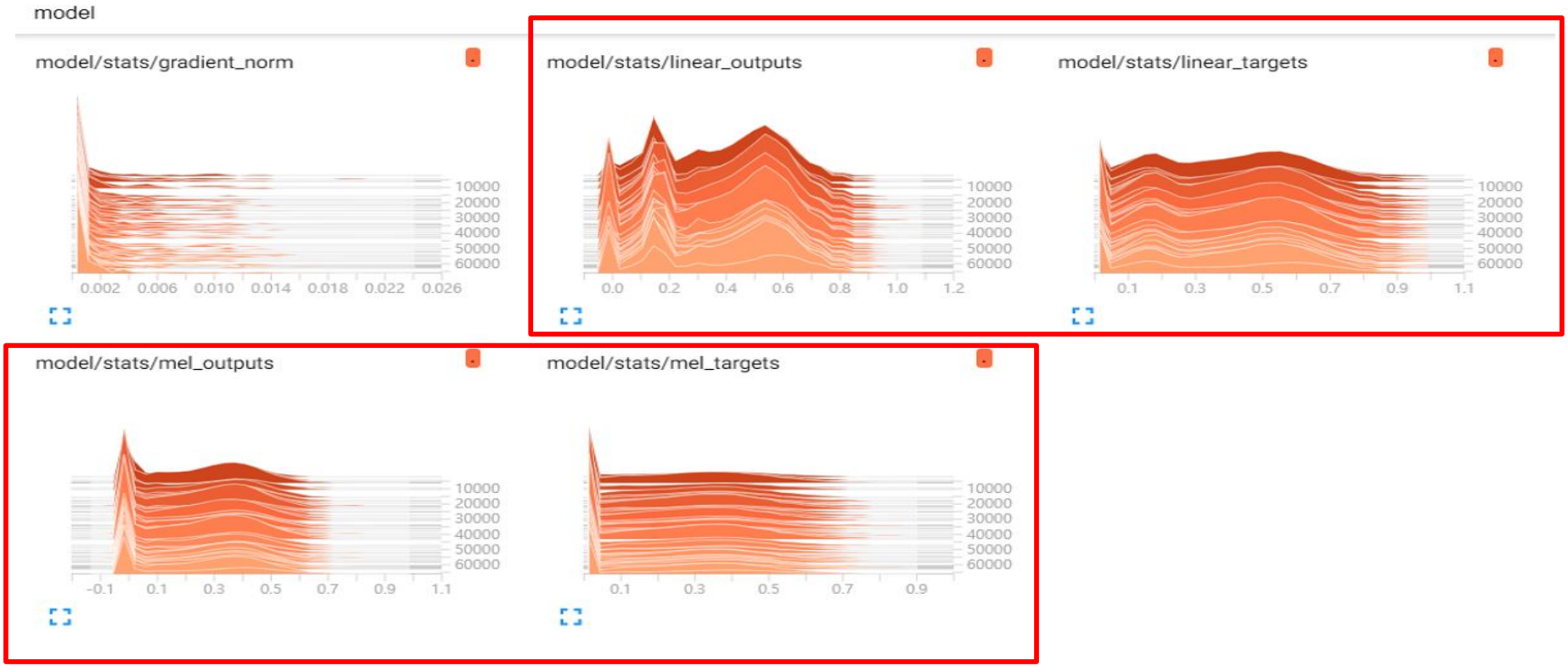


tacotron, None, 2019-05-06 10:42, step=50000, loss=0.08184

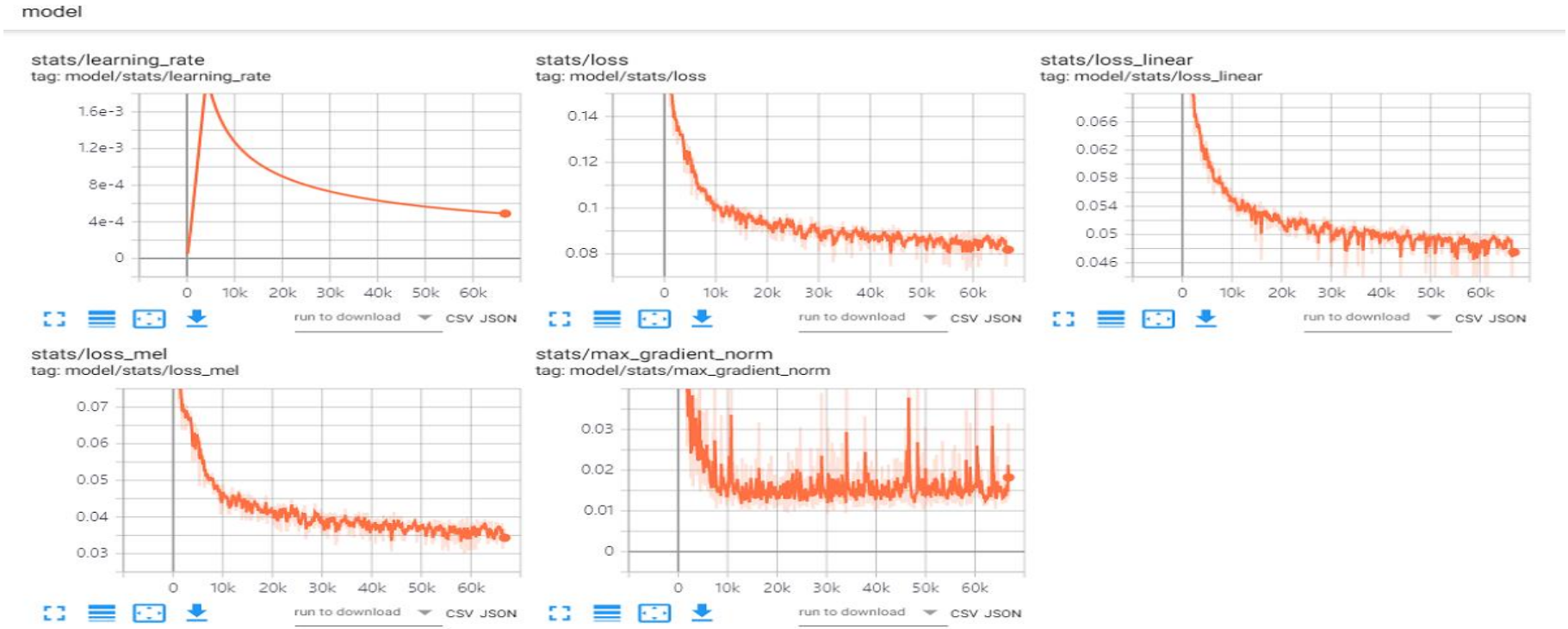


tacotron, None, 2019-05-07 09:17, step=65000, loss=0.08697

Tensorboard



Tensorboard



감사합니다.