

적은 음성 데이터 바탕 TTS 서비스 앱 개발

20180953 컴퓨터공학과 박수빈
20180128 컴퓨터공학과 이해찬
20180119 컴퓨터공학과 최수아
20180462 컴퓨터공학과 채승현
20190065 컴퓨터공학과 강두경
20190431 컴퓨터공학과 최정원

연구 목적 및 방법

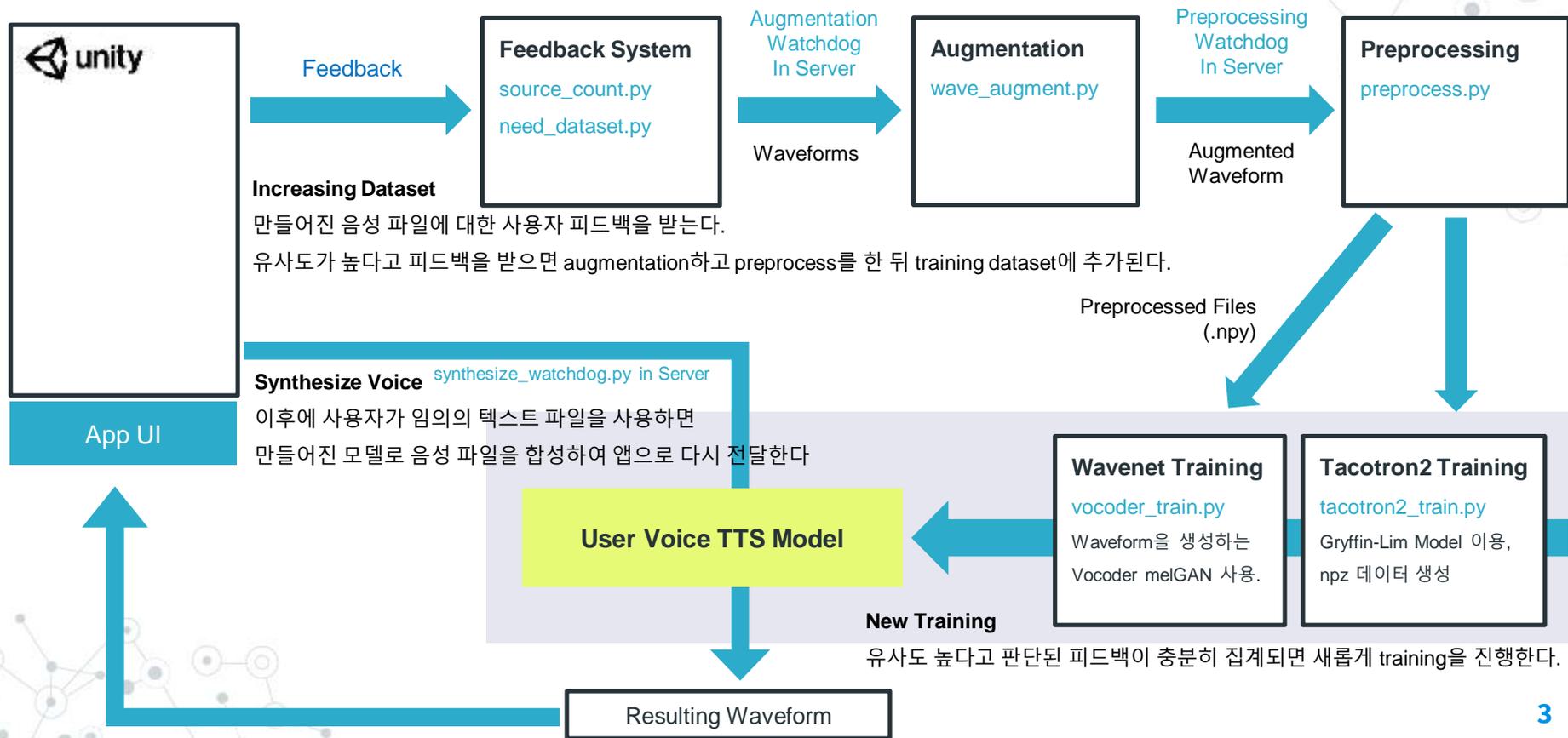
연구목적

말하기 불편한 분들을 위해 텍스트를 입력하면 그에 맞는 사용자의 목소리로 읽는 음성을 제작하고 적은 양의 음성데이터를 통해서도 유사도 높은 음성을 제작할 수 있는 앱 개발

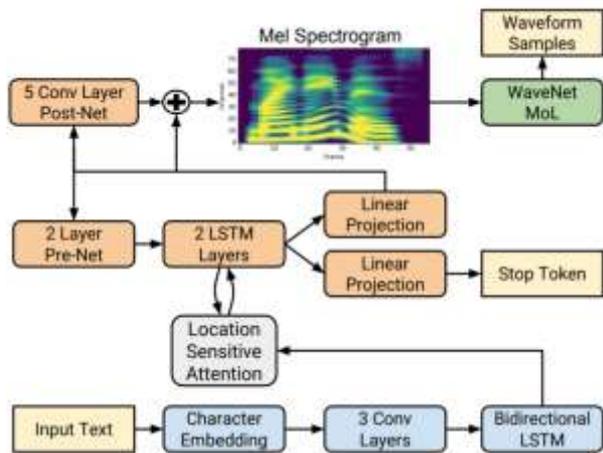
연구방법

1. TTS (Text-to-Speech), Audio Style Transfer 오픈소스 분석 및 활용
2. 학습 및 활용 데이터 분석 및 수집 (script)
3. 적은 음성 데이터로도 학습할 수 있도록 사용자 목소리와 생성된 음성간 유사도 측정 및 활용
4. 적은 음성 데이터로도 학습할 수 있도록 음성 augmentation 활용
5. TTS 서비스 제공용 앱 개발

연구 내용 (전체 프로젝트 구조)



연구결과 - 적은 음성 데이터를 활용한 유사도 높은 음성 제작



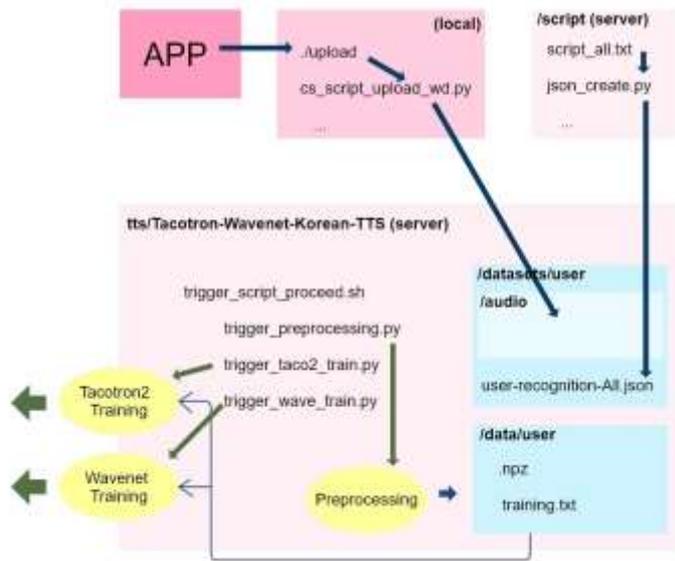
< tacotron2 system architecture >

음성 제작을 하고자 하는 해당 음성 데이터 확보를 위해 사용자에게 특정 script를 읽도록 함 (script_system)

앱을 통한 특정 문장들에 대해 사용자 음성 녹음 데이터는 서버로 전송된 후 데이터셋으로 변환

음성 제작을 위한 TTS 모델로 Tacotron2 사용, 그 중에서도 오픈소스로 제공된 한국어 모델을 활용

(<https://github.com/hccho2/Tacotron2-Wavenet-Korean-TTS>)



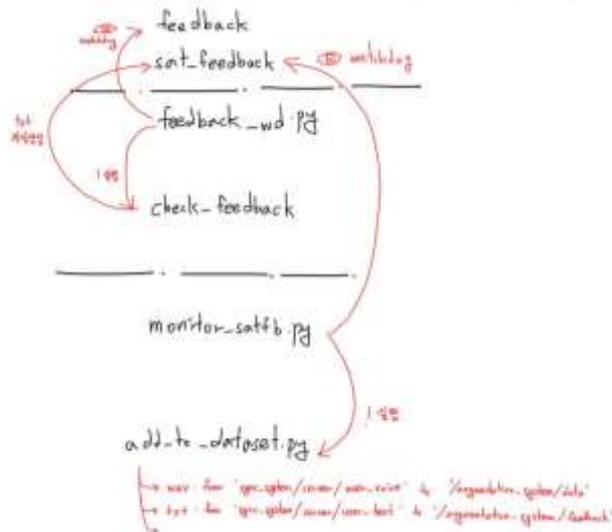
< script_system 흐름도 >

연구결과 - 피드백, augmentation을 활용한 데이터셋 확충

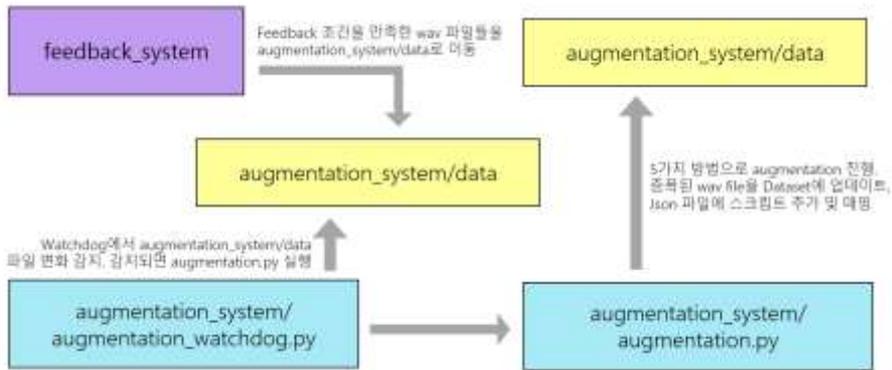
feedback_system을 통해 새로 생성된 음성 파일의 모방 능력에 대한 사용자 만족도(피드백)을 0~10 사이로 받음

특정 기준(9점)을 넘는 경우, 해당 음성의 training dataset에 생성된 음성이 추가됨

5개 이상의 새로운 음성 데이터가 데이터셋에 추가된 경우, 새롭게 트레이닝을 진행함



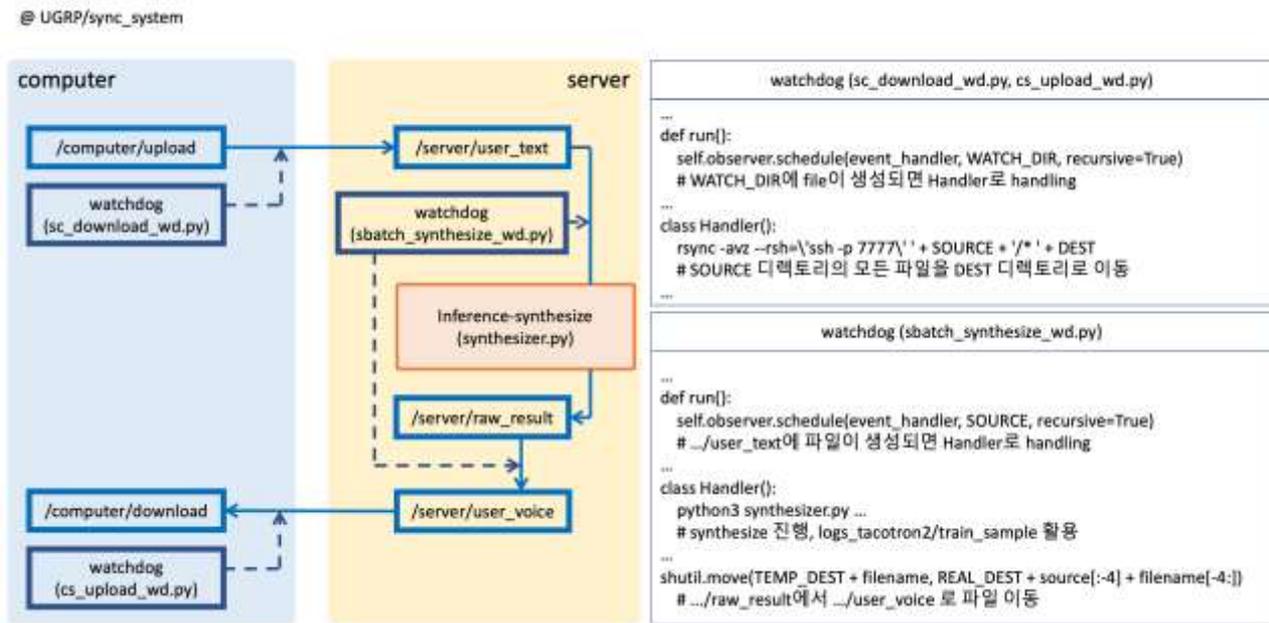
<feedback_system architecture>



<augmentation_system architecture>

만족스럽다고 판단된 음성이 바로 추가되기 전에 음성 데이터 augmentation을 통해 데이터의 양을 증폭시킴 (augmentation_system)

연구결과 - 앱, 데이터, 트레이닝/합성을 연결하는 시스템



<sync_system architecture>

script, feedback, augmentation_system 및 음성 생성 과정이 이뤄지는 앱, 컴퓨터, 서버를 watchdog Python 모듈과 rsync를 통해 연결/연동 (sync_system)

연구 결과 - 원활한 서비스 사용을 위한 앱 개발

씬	기능
초기 로딩 씬	서버와의 연결 대기, 연결되면 초기 메인 씬으로 이동
초기 메인 씬	스크립트 녹음 씬으로 이동
메인 씬	텍스트 전송 씬으로 이동
	녹음 씬으로 이동
스크립트 녹음 씬	화면에 보여주는 스크립트를 녹음
	녹음 파일 재생하여 확인
	녹음 파일 서버에 전송
	스크립트 녹음 종료 후 로딩 씬으로 이동
로딩 씬	서버에서 학습이 끝날 때까지 대기, 학습이 끝나면 메인 씬으로 이동
텍스트 전송 씬	텍스트를 입력하여 서버에 전송
	원하는 목소리로 변환된 음성을 서버로부터 수신
	음성 파일 재생
	음성 파일 다운로드
	평가 UI 생성, 유사도를 1~10 점으로 평가 후 서버로 전송



<초기 로딩 씬>



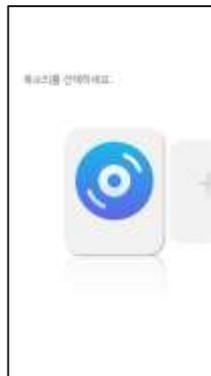
<초기 메인 씬>



<스크립트 녹음 씬>



<로딩 씬>



<메인 씬>



<텍스트 전송 씬-1>



<텍스트 전송 씬-2>



<텍스트 전송 씬-3>

연구 결과 (데모)



접근성이 용이한 스마트폰 앱으로 일반 대중에게 TTS 서비스를 제공할 수 있고, 데이터 augmentation 기술의 활용으로 비교적 적은 데이터셋을 확장하여 유의미한 학습 결과를 낼 수 있도록 한다.

모든 연구 결과는 <https://github.com/shchae7/UGRP>, https://github.com/botrote/UGRP_APP에서 확인 가능

연구 활용방안

1. 신체적 제한 등의 이유로 소통하기 어려운 경우
2. 목소리를 듣고 싶지만 듣기 힘든 상황에 있는 경우
3. 임의의 사람의 목소리를 복잡한 전처리 과정 없이 간단히 담고 싶은 경우
4. 데이터가 많지 않은 사람들의 목소리를 담고 싶은 경우

등 위와 같은 경우/상황에서 유용하게 사용할 수 있을 것이라고 예상