


CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인 수강 학생 중 프로젝트 "Dalarm(달람)"를 수행하는 팀 "달달한 컴퓨터 달콤"의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 "달달한 컴퓨터 달콤"의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

캡스톤 디자인 I 종합설계 프로젝트

프로젝트 명	Dalarm (달람)
팀 명	달달한 컴퓨터 달콤
문서 제목	중간평가요약보고서

팀원	김 세희(조장)
	강 승균
	남 병욱
	위 광진
지도교수	김 준호 교수
Version	1.3
Date	2018-04-10

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

문서 정보 / 수정 내역


Filename	중간평가요약보고서-Dalarm(달람).doc
원안작성자	김세희, 남병욱, 위광진
수정작업자	강승군, 김세희

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2018-03-28	김세희	1.0	최초 작성	프로젝트 목표
2018-04-09	남병욱	1.1	최초 작성	수정된 연구내용 및 추진 방향, 향후 추진계획
2018-04-09	위광진	1.2	최초 작성	수행내용 및 중간결과, 건의사항
2018-04-10	강승군, 김세희	1.3	내용 수정	

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

목 차

1	프로젝트 목표	4
2	수행 내용 및 중간결과	5
2.1	계획서 상의 연구내용	5
2.2	수행내용	5
2.2.1	구체적인 수행내용	5
2.2.2	개발환경 구축 및 알람이 가능한 안드로이드 앱 구현, UI 디자인	6
2.2.3	프로젝트 진행 중의 고충 사항	7
3	수정된 연구내용 및 추진 방향	8
3.1	수정사항	8
3.1.1	추가	8
3.1.2	삭제	8
3.1.3	보완	8
4	향후 추진계획	9
4.1	향후 계획의 세부 내용	9
5	고충 및 건의사항	10


 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

1 프로젝트 목표

많은 사람들은 목소리가 좋은 사람에게 호감을 느끼게 된다. 심지어 목소리 하나만으로 인기를 유지하는 연예인이 있을 정도로 누군가에겐 목소리가 사람을 좋아하게 되는 이유가 될 수 있다. 또, 어떤 사람은 자신이 좋아하는 라디오DJ의 목소리를 들으면서 잠을 청하기도 하고, 목소리가 좋은 유튜브 스트리머나 인터넷 방송 BJ가 책을 읽어주는 방송을 듣기도 한다.

많은 사람들이 목소리가 좋은 사람을 좋아하고 찾게 된다는 사실을 통해 이와 관련된 음성합성(TTS)을 이용한 서비스를 만들어보려고 한다. TTS란 인위적으로 사람의 소리를 합성하는 시스템이고, 텍스트를 음성으로 변환하는 텍스트 음성 변환 (Text-To-Speech)시스템이다. 이 기술을 활용하여 시각 장애인을 위한 오디오북, 대화 인공지능 등 많은 분야에 이용되고 있으나 대부분 음성이 기계음이어서, 듣는 사람이 지루하고 집중이 안 될 것이라고 생각을 했다. 그래서 본 달람 팀은 자신이 좋아하는 연예인이나 유명 인물의 목소리로 아침에 알람을 울리게 하거나, 사용자의 일정을 읽어주는 서비스를 개발해 보기로 했다.

그리하여 요약하면, 기존에 존재하는 알람 어플의 기능에 딥러닝과 TTS를 결합하여 특정 유명인의 목소리를 사용한 알람의 기능을 구현하는 것을 목표로 한다.

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

2 수행 내용 및 중간결과

2.1 계획서 상의 연구내용

본 팀의 중간평가 이전까지의 계획은 알람 서비스를 제공하는 전체적인 틀을 개발하는 것이었다.

첫째, 기본적인 개발환경 및 알람기능이 가능한 안드로이드 어플리케이션 구현 및 UI 디자인개발

둘째, 안드로이드 어플리케이션과 데이터베이스간의 통신을 해서 서버에서 TTS 음성파일을 생성하는 것

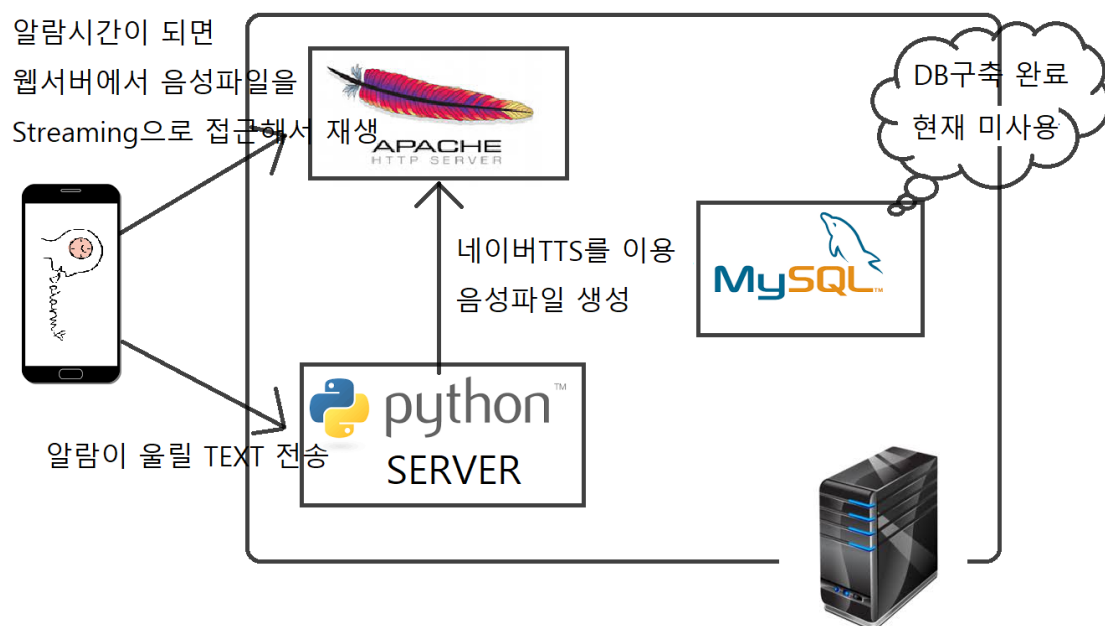
셋째, 기초적인 딥러닝 개념학습 및 tacotron 논문 분석이 있었다.


최종목표는 딥러닝으로 학습된 모델을 이용하여 특정 유명인의 목소리로 음성파일을 생성하는 것으로 이것은 중간평가 이후로 계획했다.

2.2 수행내용

2.2.1 구체적인 수행내용

우선, 계획서에 기재한 중간평가 이전까지의 목표들은 모두 수행했고 추가적으로 몇가지 기능을 더 수행하거나 기존 계획을 약간 수정한 부분이 있다.



	중간평가요약보고서		
	국민대학교	Dalarm (달람)	
	컴퓨터공학부	달달한 컴퓨터 달콤	
	캡스톤 디자인 I	Confidential Restricted	Version 1.3 2018-04-10

2.2.2 개발환경 구축 및 알람이 가능한 안드로이드 앱 구현, UI 디자인

A. 개발환경

개발환경은 server와 client 두가지로 구성되었다.

- server

OS : windows10

cpu : intel i7

gpu : gtx980

tensorflow 1.3

CUDA v8.0

cudnn v6.0

mysql

apache web server

python server

- client

OS : windows10

IDE & language : android studio, java

B. 안드로이드 어플리케이션과 서버와의 통신

안드로이드 app과 DB간의 통신을 이용해서 서버에서 음성파일을 만드는 작업이었다. 저희는 서버에 mysql, apache서버를 설치해서 app과 DB 통신을 가능하게 했다. 현재 딥러닝 모델의 음성파일을 사용하지 않으므로 DB는 구축은 했지만 미사용이다. client에서 보낸 텍스트와 목소리모형을 어떻게 처리를 할 것인지 고민하다가 python server를 구축해서 처리하게끔 했다. python server는 client에서 보내온 텍스트와 목소리모형을 전달받아서 Naver Clova Speech Synthesis api를 사용하여 음성파일을 apache서버경로에 저장하고 client는 apache서버에 접근해서 알람을 울리게 하는 방식으로 동작하게 만들었다.


C. 딥러닝 개념학습과 tacotron 논문분석

딥러닝 공부와 음성합성을 위한 라이브러리인 tacotron 논문분석이었는데 지도교수님인 김준호 교수님께서 추천해주신 youtube cs231n의 강의를 듣고 회의시간에 팀원들과 같이 공부를 진행했고 마찬가지로 tacotron 논문도 같이 공부를 했다. 추가적으로 기존에 tacotron을 이용해서 음성 학습을 시도한 개발자분이 공개하신 코드를 테스트하는 부분도 같이 진행했다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

2.2.3 프로젝트 진행 중의 고충 사항

현재 어렵고 잘 진행되지 않은 부분이 유명인의 목소리를 학습시키는 부분이다. text와 voice의 쌍을 생성할 때 voice에 해당하는 text를 google speech api 를 사용하는데 올바르게 한국말을 인식하는 것이 가끔 부족할 때가 있고, 학습시킬 data 크기가 적어서 현재 학습하는데 어려움을 겪고 있다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

3 수정된 연구내용 및 추진 방향

3.1 수정사항

3.1.1 추가


처음 제안서에는 기존에 존재하는 목소리 모델을 사용하여 어플을 만들고 후에 새로운 목소리 모델을 추가할 계획이었으나 기존에 존재하는 모델이 악용 될 수가 있다는 이유로 만들어 놓은 모델을 전부 github에서 삭제 하였기 때문에 우리가 처음부터 목소리 모델을 만들어야 했다. 따라서 그에 따른 새로운 Tacotron을 사용한 목소리 모델 학습에 관한 내용이 추가 된다. 학습에 관한 내용은 주로 Tacotron논문과 'Devview2017 - 책읽어주는 딥러닝 - 김태훈'의 동영상과, 자료를 참고하여 학습과정을 인코더, 디코더, 보코더, 어텐션의 단계에 따라 각 과정에서 일어나는 일, 각 과정이 어떻게 연결되어 영향을 미치는지에 대해서 추가해 갈 예정이다.

3.1.2 삭제

원래는 Tacotron과 DeepVoice를 모두 사용하는 계획이었지만 목소리 모델을 만들어서 알람에 적용하는데 Tacotron만 사용하여도 무리가 없다는 판단으로 DeepVoice를 사용하지 않게 되어서 DeepVoice에 관련한 내용을 삭제 한다.

3.1.3 보완

처음 제안서에 Tacotron을 사용한 딥러닝 과정 및 데이터 셋 수집 과정 등에 관련된 내용이 부족하여 그에 관련한 내용을 보완 한다. 딥러닝의 과정은 Tacotron논문과 따로 공부한 자료들(cs231n, cs224n 등)을 참고하여 어떤 hyperparameter가 존재하고 그것을 어떻게 설정할 계획인지에 대해서 설명하고 학습 데이터 셋(텍스트, 음성)을 어떻게 수집할지에 대한 내용은 우선 목소리 모델이 될 인물을 설정하고 수집방법(예를 들어 손석희의 경우 jtbc홈페이지에서 뉴스영상에서 음성파일을 추출, 추출한 음성파일을 google speech를 통해 그 음성을 텍스트로 변환 하는 것과 유사한 방식)에 대한 내용을 보완 할 예정이다.


 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

4 향후 추진계획

4.1 향후 계획의 세부 내용

현재 알람 어플의 기본 기능인 정해진 시간에 알람이 울리는 기능과 알람을 설정 할 때 텍스트와 목소리 모델을 선택하면 그 텍스트와 목소리 모델에 대한 정보가 DB로 전달되고 그 정보를 파이선 서버에서 네이버 TTS서비스를 사용하여 기계음성 파일을 생성하고 그 파일을 정해진 시간에 스트리밍으로 재생을 하는 단계까지 완성을 했다.

그에 따라 이제 Tacotron을 사용한 딥러닝 학습에 전념하여 새로운 목소리 모델을 만들고 그 모델을 네이버 TTS서비스를 대체하여 기계음성 파일이 아닌 유명인의 목소리 파일이 생성되어서 재생될 수 있도록 할 것이다. 구체적인 계획으로 먼저 모든 팀원들(강승균, 김세희, 남병욱, 위광진)은 현재 진행중인 Tacotron의 논문을 이해와 Youtube를 통한 cs231n, cs224n 등을 공부하고 이전부터 계속 Tacotron으로 목소리 모델을 만들고 있었던 위광진을 도와서 우선은 데이터 셋(텍스트, 음성파일)이 공개되어있는 문재인 대통령의 목소리 모델을 만들고 그것을 바탕으로 완전히 새로운 인물을 설정하여 데이터 셋(텍스트, 음성파일)을 수집, Tacotron을 사용하여 새로운 목소리 모델을 만들 계획이다. 새로운 인물의 목소리 모델을 만드는 것에는 딥러닝 학습 시간(현재는 약 1step당 4~5초), hyperparameter 등에 문제들이 있는데 학습 시간의 경우 현재 주어진 서버(하드웨어)에 의한 것이므로 어쩔 수 없지만 hyperparameter를 적절히 설정하여 최대한 빠른 학습으로 고품질의 목소리 모델을 만들어 그에 따른 소요시간에 따라 2~3개 정도의 새로운 목소리 모델을 만들 계획이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간평가요약보고서		
	프로젝트 명	Dalarm (달람)	
	팀 명	달달한 컴퓨터 달콤	
	Confidential Restricted	Version 1.3	2018-04-10

5 고충 및 건의사항

1. 캡스톤 디자인을 수강하는 전체 팀이 30팀이 넘는데 그 많은 팀들이 각자 프로젝트를 수행할 수 있는 공간이 부족하다.
2. 딥러닝을 위한 고성능 서버제공이 있었으면 좋을 것 같다. 음성합성을 위해 학습시키는데 소요되는 시간이 굉장히 오래 걸려서 수차례 학습을 시켜야 제대로 된 결과물이 나올 수 있는데 그렇지 못하는 상황이다. 학교 측에서 따로 고성능 GPU가 탑재된 서버를 캡스톤 디자인팀들에게 수요조사를 해서 학기가 시작할 때 제공해 주면 훨씬 수월한 개발이 될 것이다.