



국민대학교
전자정보통신대학
컴퓨터공학부


캡스톤 디자인 I

종합설계 프로젝트

| | |
|--------|-------------|
| 프로젝트 명 | 달람 (Dalarm) |
| 팀 명 | 달달한 컴퓨터 달콤 |
| 문서 제목 | 최종보고서 |

| | |
|---------|------------|
| Version | 1.1 |
| Date | 2018-05-29 |

| | |
|----|-----------|
| 팀원 | 김 세희 (조장) |
| | 강 승군 |
| | 남 병욱 |
| | 위 광진 |

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |


CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 “달람 (Dalarm)”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.

문서 정보 / 수정 내역


| | |
|-----------------|----------------------|
| Filename | 최종보고서-달람(Dalarm).doc |
| 원안작성자 | 김세희 |
| 수정작업자 | 김세희 |

| 수정날짜 | 대표수정자 | Revision | 추가/수정 항목 | 내 용 |
|------------|-------|----------|---------------|-----|
| 2018-05-27 | 김세희 | 1.0 | 최초 작성 | |
| 2018-05-29 | 김세희 | 1.1 | 추가로 필요한 항목 작성 | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

목 차

| | | |
|---------|------------------------------------|----|
| 1 | 개요 | 4 |
| 1.1 | 프로젝트 개요 | 4 |
| 1.2 | 추진 배경 및 필요성 | 4 |
| 1.2.1 | 기술의 시장 현황 | 4 |
| 1.2.2 | 기 개발된 시스템의 문제점과 개발할 시스템의 필요성 | 5 |
| 2 | 개발 내용 및 결과물 | 6 |
| 2.1 | 목표 | 6 |
| 2.2 | 어플 소개 | 6 |
| 2.3 | 어플 장점 | 6 |
| 2.3.1 | 알람 기능 | 6 |
| 2.3.2 | 기호에 따른 유명한 목소리 선택 기능 | 6 |
| 2.3.3 | 기존 연예인 알람과의 차별성 | 7 |
| 2.4 | 연구/개발 내용 및 결과물 | 7 |
| 2.4.1 | 연구/개발 내용 | 7 |
| 2.4.1.1 | 모바일 어플리케이션 구현 | 7 |
| 2.4.1.2 | 웹서버 | 7 |
| 2.4.1.3 | 음성 데이터 추출 | 8 |
| 2.4.1.4 | 텍스트 데이터 생성 | 9 |
| 2.4.1.5 | 음성 딥러닝 학습 | 10 |
| 2.4.2 | 시스템 기능 요구사항 | 14 |
| 2.4.3 | 시스템 비기능(품질) 요구사항 | 14 |
| 2.4.4 | 시스템 구조 및 설계도 | 15 |
| 2.4.5 | 활용/개발된 기술 | 15 |
| 2.4.6 | 현실적 제한 요소 및 그 해결 방안 | 16 |
| 2.4.7 | 결과물 목록 | 17 |
| 2.5 | 기대효과 및 활용방안 | 18 |
| 3 | 자기평가 | 19 |
| 4 | 참고 문헌 | 20 |
| 5 | 부록 | 21 |
| 5.1 | 사용자 매뉴얼 | 21 |
| 5.2 | 테스트 케이스 | 26 |

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

1 개요

1.1 프로젝트 개요

많은 사람들은 목소리가 좋은 사람에게 호감을 느끼고, 관심을 표하게 된다. 심지어 목소리 하나만으로 인기를 유지하는 연예인이 있을 정도로 누군가 예쁜 목소리가 사람을 좋아하게 되는 이유가 될 수 있다. 또, 어떤 사람은 자신이 좋아하는 라디오DJ의 목소리를 들으면서 잠을 청하기도 하고, 요즘에는 목소리가 좋은 Youtube Streamer나 인터넷 방송 BJ가 책을 읽어주는 방송을 듣기도 한다.

이처럼 사람들이 목소리 좋은 사람을 좋아하고 또, 그 사람을 찾게 되는 이러한 사실을 통해 이와 관련된 음성합성(TTS)을 이용한 서비스를 만들어보려고 한다. TTS란 인위적으로 사람의 소리를 합성하는 시스템이자, 텍스트를 음성으로 변환하는 텍스트 음성 변환 (Text-To-Speech) 시스템이다. 이 기술을 이용하여 시각 장애인을 위한 audio book, 대화 인공지능 등 많은 분야에 활용하고 있으나 대부분 음성이 기계적이기 때문에, 듣는 사람이 지루하고 집중이 안 될 것이라는 생각을 했다.


그리하여 달콤한 컴퓨터 달콤팀은 자신이 좋아하는 연예인이나 유명 인물의 음성을 최대한 자연스럽게 구현하여, 아침에 알람을 울리게 하거나 필요한 시간에 사용자가 원하는 텍스트의 스케줄 알람을 읽어주는 서비스를 개발해 보기로 했다. 요약하자면, 기존에 존재하는 알람 어플의 기능에 딥러닝과 TTS를 결합하여 특정 유명인의 목소리를 사용한 알람의 기능을 구현하는 것을 목표로 한다.

1.2 추진 배경 및 필요성

1.2.1 기술의 시장 현황

음성인식의 반대라고 할 수 있는 음성합성 기술이 최근에 많은 연구가 되고, 실용화 되고 있다. 이렇게 음성합성에 대한 기술이 아주 많은 관심의 대상이 된 이유는 매우 편리하고 사용할 수 있는 분야가 광범위하다는 장점 때문이다.

대표적으로 네이버에서 자체적으로 개발한 nVoice엔진이 있다. nVoice는 먼저 특정 사람의 목소리를 녹음한 대량의 데이터를 시스템에 얹어서 nVoice엔진을

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

만들게 된다. 그리고 나중에 원하는 텍스트를 nVoice엔진에 입력하게 되면 그 사람의 음성으로 변환되어 나오게 되는 방식이다. 이런 음성 합성기술은 전문적인 목소리 교육을 받은 사람 없이도 필요할 때에 텍스트 내용을 바로 읽어줄 수 있기 때문에, 신속하게 오디오 콘텐츠를 만들어야 할 때 유용하다.

또, 네이버와 YG엔터테인먼트가 함께해서 오디오북 기부 캠페인이라는 것을 하였는데, 목소리가 좋기로 유명한 유인나 배우가 책을 읽어주는 오디오북을 개발했다. 이 오디오북은 nVoice엔진을 이용해서 유인나 목소리를 학습시켜서 사용자들에게 유인나 목소리로 책을 읽어주는 서비스이다.

왕자와 거지

데미안

동물농장

"신분이 뒤바뀐 두 소년의 모험담 속 섬세한 묘사와 삶의 교훈을 느껴보세요" [책더보기](#)



전반부 유인나 낭독 오디오북



후반부 네이버 음성합성 오디오북

1.2.2 기 개발된 시스템의 문제점과 개발할 시스템의 필요성

기존의 TTS기술을 이용한 서비스 제품들을 살펴보면 대부분 음성이 기계음이어서 톤이 단조로워 사용자가 듣기에 지루하고, 말하는 방식도 어색해서 듣기 불편한 경향이 있었다.

또한, nVoice처럼 특정인물의 목소리를 음성합성 시킨 예도 있지만 목소리 종류가 다양하지 않는 단점이 있다. 그리고 우리가 개발하려는 알람 어플리케이션 부분에서는 일반적으로 시끄럽고, 자극적인 소리만 존재했다. 연예인 알람이 존재했지만 그것도 단순히 유명인물의 뮤직비디오나 동영상을 재생하는 것에 불과했다. 그래서 이러한 점들을 보완하면서 음성합성 기술을 이용해 유명 인물이 잠을 깨워주는 알람, 사용자의 스케줄을 읽어주는 서비스를 만들 필요성을 느꼈고,

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

본 프로젝트로 목적을 달성하려고 한다.

2 개발 내용 및 결과물

2.1 목표

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

2.2 어플 소개

기존의 알람 어플의 기능을 기본으로 딥러닝과 TTS를 사용하여 자신이 원하는 내용의 텍스트 메시지를 입력하면 그 내용을 특정 유명인의 목소리로 바꾸어서 원하는 시간에 맞춰 재생해 주는 어플이다. 기존에 존재하는 연예인 알람과는 다르게 미리 녹음되어 있는 것을 재생하는 것이 아니라 자신이 원하는 내용을 만들어서 재생할 수 있다.


2.3 어플 장점

2.3.1 알람 기능

본 어플은 기존에 있는 알람 어플의 기능을 모두 가지고 있다. 아침에 시끄러운 모닝콜뿐만 아니라 자신이 해야 할 스케줄을 미리 입력해 놓으면 설정한 시간에 적어놓은 메시지가 특정 유명인의 목소리로 재생된다.

2.3.2 기호에 따른 유명인 목소리 선택 기능

보통의 알람은 시끄러운 소리나 노래를 사용하는 반면에 평소 좋아하는 유명인의 목소리를 사용하여 알람을 설정하면 더 기분좋은 알람을 들을 수 있다.

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

2.3.3 기존 연예인 알람과의 차별성

기존에 존재하는 연예인 알람 어플의 경우엔 미리 녹음되어 있는 목소리를 구매하거나 유튜브에 있는 콘텐츠를 그대로 재생하여 알람에 사용된다. 이 어플의 경우에는 미리 녹음되어 있는 것이 아니라 내가 원하는 텍스트 메시지를 유명인의 목소리로 바꾸어서 재생하는 것이기 때문에 훨씬 더 자유롭게 사용자의 요구에 맞추어서 사용될 수 있다.

2.4 연구/개발 내용 및 결과물

2.4.1 연구/개발 내용

2.4.1.1 모바일 어플리케이션 구현

사용자가 직접 입력한 문장으로 원하는 유명인의 목소리를 통하여 알람을 들을 수 있다. 해당 기능은 유명인의 목소리를 직접 녹음한 것이 아니라 TTS (Text-To-Speech) 기술을 사용하고, 인물의 목소리를 딥러닝 기술을 이용해 학습한 것이다.

어플리케이션 작동 순서는 다음과 같다.

1. 어플을 설치 후 알람 설정 목록에서 알람이 필요한 시간대를 설정한다.
2. 알람으로 듣고 싶은 문장을 입력한다.
3. 사용자의 요구에 맞추어 해당 문장을 어떤 유명인의 목소리로 듣고 싶은지 선택하여 알람을 저장한다.
4. 알람을 설정한 시간대에 2번에서 저장한 문장이 사용자가 선택한 유명인의 음성으로 변환되어 제공된다.

2.4.1.2 웹서버

Python 서버 및 Apache 서버를 사용한다.

어플리케이션에서 알람이 울릴 Text를 전송하면, Python 서버가 학습된 목소리 모델의 음성파일을 생성한다. 해당 음성파일은 Apache 서버에 존재하게 되며, 알람이 울릴 시간이 되면 Apache

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

서버에서 Streaming 개념으로 접근하여 어플리케이션에서 음성파일을 재생한다.
두 서버는 아마존 웹서비스인 aws 위에 구축한다.

2.4.1.3 음성 데이터 추출

(RSS) Real Simple Syndication



The image shows a list of radio programs on the JTBC News website, each with an RSS icon. A red box highlights these icons. To the right, a navigation bar contains buttons for '뉴스홈', '다시보기', '오픈저널리움', and 'JTBC 뉴스룸'. Below this, a category bar lists '속보', '정치', '경제', '사회', '국제', '문화', '연예', '스포츠', and '날씨'. At the bottom, another navigation bar includes '공지사항', '자유게시판', 'JTBC뉴스앱', and 'RSS서비스', with the last one highlighted by a red box.

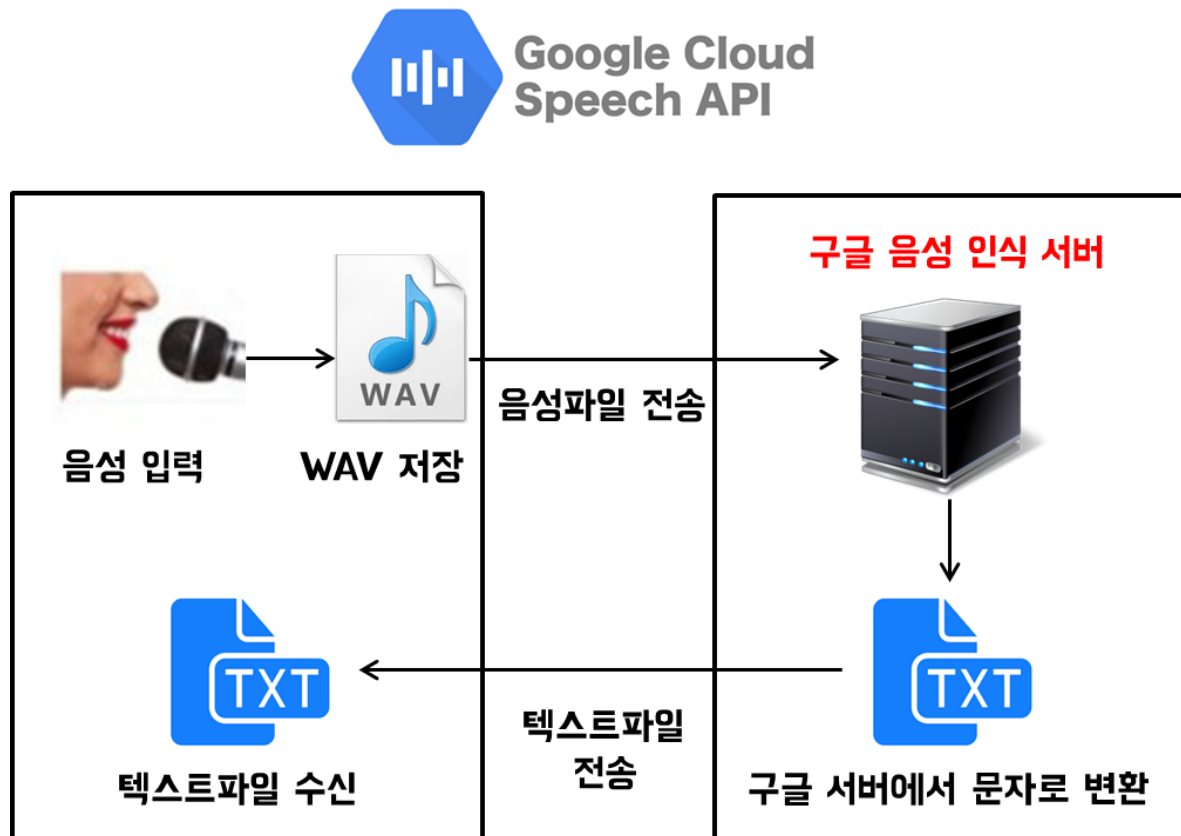
특정 인물의 목소리를 얻기 위해서는 대량의 음성 데이터셋이 필요하다. 이를 위해서 각 방송사의 팟캐스트에서 찾고자 하는 연예인이 진행하는 라디오 프로그램의 RSS주소를 제공하고 있다. RSS는 해당사이트를 직접 방문하지 않아도 RSS URL을 사용하여 자동 수집이 용이하게 해주기 때문에 이를 사용하여 오디오와 text 파일을 추출할 수 있다. 뉴스룸 에서도 이 RSS 기능을 제공하고 있다.

본 프로젝트는 KBS Cool FM의 박명수의 라디오쇼, 유인나의 볼륨을 높여요와 JTBC 뉴스룸의 음성 데이터를 추출하였다.

추가로 연예인의 라디오 프로그램의 음성보다 더 깨끗한 음성 데이터로 딥러닝 학습을 시작하기 위해 본 팀의 팀원들이 직접 20시간씩의 음성을 녹음하여 데이터를 생성하였다. 그러한 데이터로 만든 초기 음성 학습 모델 위에 손석희, 유인나, 박명수의 데이터를 학습하였다.

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |


2.4.1.4 텍스트 데이터 생성



2.4.1.3 에서 추출한 음성 데이터를 Google speech api를 이용하여 음성에 해당하는 텍스트 파일을 생성하였다.

연예인의 라디오 프로그램에서 추출해온 음성과 팀원들이 직접 녹음한 음성을 WAV 형식으로 저장하여 구글 음성 인식 서버에 전송한다. 그러면 구글 음성 인식 서버는 구글 자체 서버에서 문자로 변환하고, 변환된 텍스트파일을 생성한다.

추가로, 생성한 텍스트 파일은 기계에 의한 변환이기 때문에 어느 정도 오차가 존재한다. 그리하여 팀원들이 직접 음성 파일을 들으며 수기로 고치는 작업을 병행하였다.

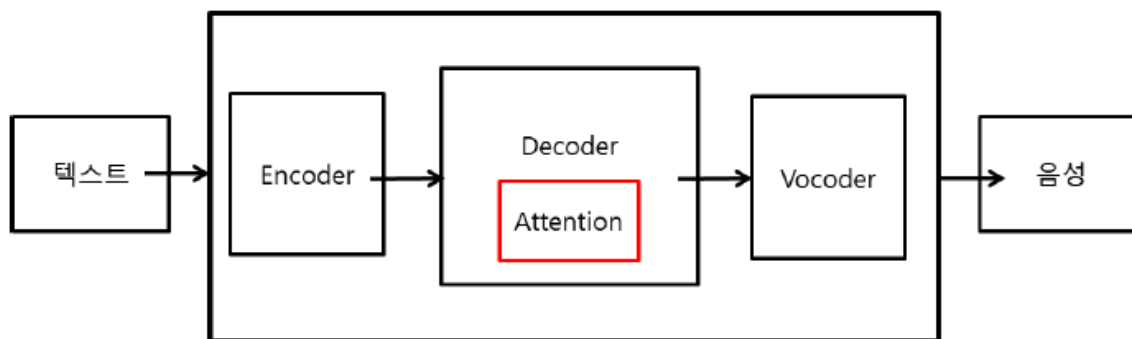
| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

2.4.1.5 음성 딥러닝 학습

- 음성 생성 시스템 'Tacotron' 을 이용


구글의 인공지능 관련 기업이자 알파고를 개발한 딥마인드(DeepMind)의 '타코트론(Tacotron)'을 이용한다. 타코트론은 2가지 심층신경망을 통해 문자를 인간과 거의 비슷하게 모방한 음성으로 변환한다. 우선 문자를 시각적으로 분석할 수 있는 스펙트로그램으로 변환한다. 스펙트로그램이란 소리, 파동 따위를 시간과 주파수축의 변화에 따라 농도와 표시 등으로 나타낸 것이다. 이후 변환된 스펙트로그램은 구글 딥마인드의 음성합성 신경망인 시스템 웨이브넷(WaveNet)에 업로드된다. 시스템은 이를 분석하여 해당 음성 요소들을 생성한다. 웨이브넷의 알고리즘은 분석을 통해 다른 목소리를 쉽게 배우며 자연스러운 호흡도 인공적으로 생성해낸다. 상기한 호흡 외에도 문구의 의미에 따른 발음 차이를 인식하고 구두점의 위치, 단어강조, 운율, 대명사구분 등을 배울 수 있다.

➤ Tacotron의 구조



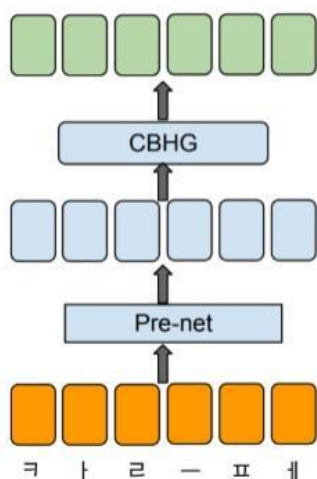
1. 데이터 프로세싱

특정 인물의 목소리를 데이터로 저장하기 위한 과정이다. 특정 인물의 음성을 추출하여 대량으로 얻어서 입력시킨다. 그 다음 각 문장을 자음과 모음의 단위로 나누어 주고, 텍스트 형식의 자음과 모음 데이터를 음성의 소리 발음과 매치해준다. 이런 프로세싱은 타코트론(Tacotron) 내부에서 진행된다.

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

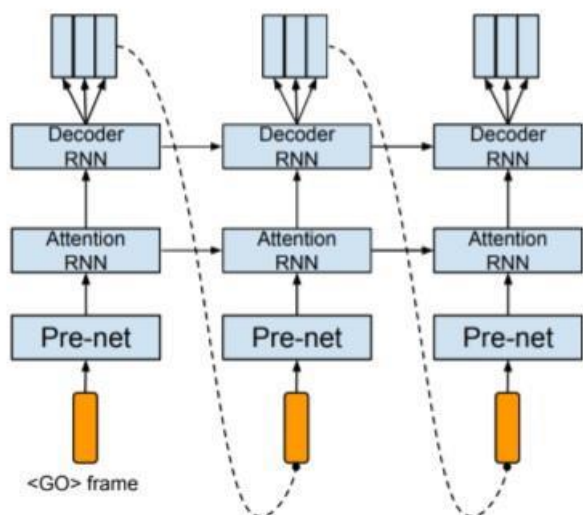
2. 인코더 (Encoder)


인코더에서 학습 데이터에 없었던 단어를 발음하기 위해 텍스트를 텍스트 정보를 잘 나타낼 수 있는 숫자로 변환(Character Embedding) 해준다. 음성의 스펙트럼 특징 벡터 열을 출력으로 지정한다.



3. 디코더 (Decoder)

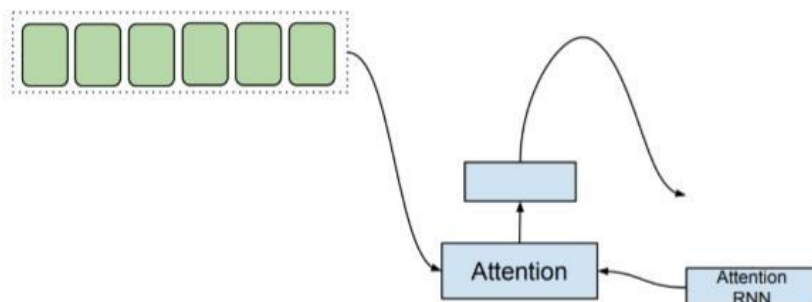
디코더에서 소리, 파동 따위를 시간과 주파수 축의 변화에 따라 농도와 표시 등으로, 즉 숫자로 나타낸 스펙트로그램을 만들어준다. 이때 입력과 출력의 길이 차이가 발생하게 된다.



| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

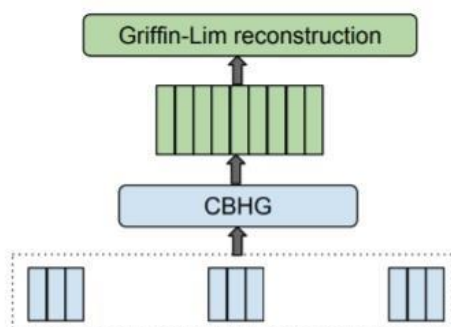
4. 어텐션 (Attention)

어텐션 이라는 신경망층을 도입해 음성을 만들 때 어디에 집중을 해 줄 것인가를 결정해주어서 입력과 출력 사이의 매핑(Mapping) 관계를 학습한다.

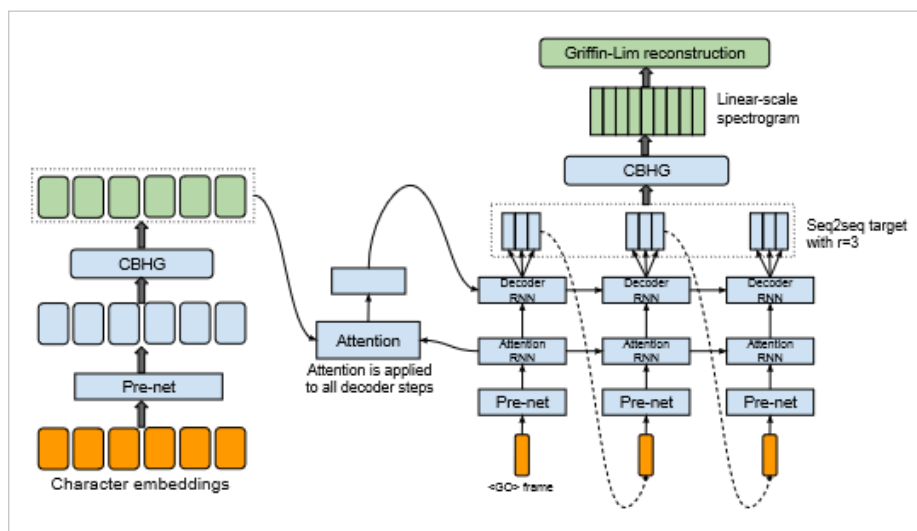



5. 보코더 (Vocoder)

보코더에서 스펙트로그램들을 사용하여 음성을 만들어준다.



● 최종 알고리즘 구조



| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

- 실제 서버에서의 딥러닝

```

Administrator: C:\Windows\system32\cmd.exe - python train.py --data_path=datasets/yuinna --load_path
Step 51600 [1.584 sec/step, loss=0.09118, avg_loss=0.09279]
Writing summary at step: 51600
Step 51601 [1.598 sec/step, loss=0.09385, avg_loss=0.09284]
Step 51602 [1.593 sec/step, loss=0.09472, avg_loss=0.09283]
Step 51603 [1.596 sec/step, loss=0.09122, avg_loss=0.09283]
Step 51604 [1.573 sec/step, loss=0.09459, avg_loss=0.09286]
Step 51605 [1.574 sec/step, loss=0.08848, avg_loss=0.09284]
Step 51606 [1.597 sec/step, loss=0.09563, avg_loss=0.09286]
Step 51607 [1.615 sec/step, loss=0.09410, avg_loss=0.09286]
Step 51608 [1.614 sec/step, loss=0.09466, avg_loss=0.09288]
Step 51609 [1.652 sec/step, loss=0.09192, avg_loss=0.09293]
Step 51610 [1.647 sec/step, loss=0.09034, avg_loss=0.09289]
Step 51611 [1.640 sec/step, loss=0.08873, avg_loss=0.09282]
Step 51612 [1.647 sec/step, loss=0.09340, avg_loss=0.09286]
Step 51613 [1.647 sec/step, loss=0.09474, avg_loss=0.09286]
Step 51614 [1.652 sec/step, loss=0.09451, avg_loss=0.09289]
Step 51615 [1.652 sec/step, loss=0.09462, avg_loss=0.09292]
Step 51616 [1.653 sec/step, loss=0.09242, avg_loss=0.09293]
Step 51617 [1.641 sec/step, loss=0.09185, avg_loss=0.09290]
Step 51618 [1.630 sec/step, loss=0.09499, avg_loss=0.09291]
Step 51619 [1.647 sec/step, loss=0.09625, avg_loss=0.09294]
Step 51620 [1.647 sec/step, loss=0.09165, avg_loss=0.09293]
Step 51621 [1.650 sec/step, loss=0.09180, avg_loss=0.09291]
Step 51622 [1.651 sec/step, loss=0.08996, avg_loss=0.09290]
Step 51623 [1.639 sec/step, loss=0.08805, avg_loss=0.09284]
Generated 32 batches of size 16 in 3.672 sec
Step 51624 [1.618 sec/step, loss=0.09319, avg_loss=0.09281]
Step 51625 [1.632 sec/step, loss=0.09416, avg_loss=0.09280]
Step 51626 [1.622 sec/step, loss=0.09412, avg_loss=0.09280]

```

실제로 달람 서버에서 학습을 돌리고 있는 화면이다.

한 스텝당 약 1.6초가 소요되며 한 목소리 모델 당 약 10만 스텝 이상 러닝을 진행하였다.

러닝을 시키는 조건 중 하이퍼 파라미터를 조절하는 부분이 있었는데, 가장 최적화 되는 파라미터를 찾아서 사용하게 되었다.

또한, 수집한 데이터의 양이 적거나 데이터의 질이 떨어져서 러닝이 잘 되지 않을 때는 데이터를 더 수집하거나 음질이나 발음이 더 깨끗한 데이터를 수집하여 러닝을 다시 시작하여 업그레이드 하였다.

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

2.4.2 시스템 기능 요구사항

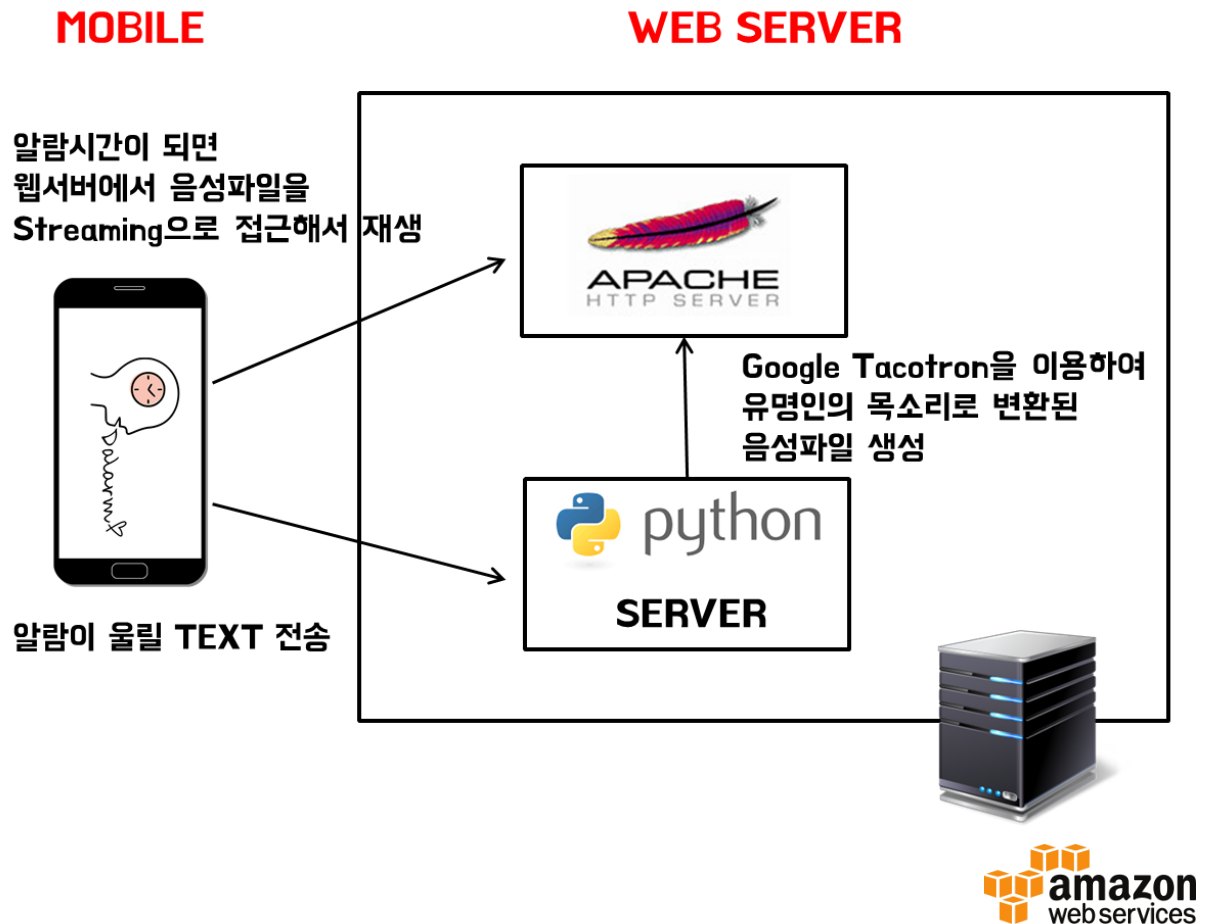
| 대분류 | 분류 | 내용 | 비고 |
|-----------|-------|---------------------------|-----|
| 기능적 요소 | 주요 기능 | 기본적인 알람 기능 | 완료 |
| | | 알람으로 듣고 싶은 텍스트 입력 기능 | 완료 |
| | | 기호에 따른 유명인 목소리 선택 기능 | 완료 |
| | | 유명인의 목소리로 음성 변환 기능 | 완료 |
| | | Google calender api 연동 기능 | 미완료 |

2.4.3 시스템 비기능(품질) 요구사항

| 대분류 | 분류 | 내용 | 비고 |
|-------------|----|---|----|
| 비 기능적 요소 | 성능 | 사용성 - 사용하기 쉬운 형태의 UI 구성과 알람을 받기까지의 간단한 절차 | 완료 |
| | | 정확성 - 사용자가 요구하는 시간에 원하는 메시지를 정확하게 알려줌 | 완료 |
| | | 신속성 - 알람이 빠른 시간 내에 생성되어 재생 | 완료 |

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

2.4.4 시스템 구조 및 설계도



2.4.5 활용/개발된 기술

- 개발된 기술
 1. 알람 기능 기반의 모바일 어플리케이션
 2. 원하는 텍스트로 음성 생성
 3. 기호에 따른 연예인 음성으로의 알람

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

- 활용된 기술

1. 음성에서 텍스트 데이터 추출 시 Google speech api 이용
2. Tacotron 을 이용한 음성 딥러닝

2.4.6 현실적 제한 요소 및 그 해결 방안

- 하드웨어

1. 스마트폰 기기가 안드로이드 기반이어도 각각의 스마트폰 종류에 따라 소프트웨어에 차이가 있다.

- 해결방안 : Android studio에서 개발 시에 Minimum SDK를 API 14로 설정하여 개발하면 대부분의 장치에서 정상적으로 작동이 가능하다.


- 소프트웨어

1. 딥러닝과 기존의 TTS 시스템의 한계 : 기존의 TTS 시스템은 목소리를 구현 시에 사람이 직접 말을 하는 것과 같이 자연스럽게 구현하는데 어려움이 있다.

- 해결방안 : 기존 방법인 parametric TTS, concatenative TTS 과는 다르게 본 팀이 사용한 구글이 개발한 타코트론(Tacotron) 알고리즘은 심표의 위치에 따라서도 문장을 읽는 높낮이와 속도가 달라지는 알고리즘이다. 그리하여 기존의 TTS 시스템보다는 자연스러운 학습이 가능하다.

2. 운영체제 제한: Android 기반의 OS에서만 작동이 가능하다.

- 해결방안 : 우선적으로 android 기반으로 구현하고 사용자의 요구에 따라 차후에 IOS에서도 개발할 수 있도록 한다.

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |


● 기타

1. 저작권 및 초상권 문제: 연예인 음성 사용에 대한 초상권 및 저작권 문제가 발생할 수 있다.

- 해결방안: 상업적인 사용을 하지 않고 교육성 프로젝트 선에서 마무리 하되, 차후에 상업적 목적을 가지고 사용을 하게 될 시에는 먼저 저작권 법규에 따라 조치를 취한다.

2.4.7 결과물 목록

| 대분류 | 소분류 | 기능 | 형식 |
|-----|---------|--------------------------------|--------|
| 파일 | 음성파일 생성 | 텍스트를 음성으로 바꾼 파일을 생성한다. | Python |
| | 음성파일 열기 | 음성으로 바꾼 파일을 연다. | Python |
| 서버 | 웹서버 | 음성파일을 Streaming 으로 재생한다. | 웹서버 |
| 입력 | 스마트기기 | 사용자가 원하는 텍스트와 알람 시간 등을 입력 받는다. | 입력장치 |
| 출력 | 스마트기기 | 연예인 음성으로 변환된 알람을 출력한다. | 출력장치 |

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

2.5 기대효과 및 활용방안

현재까지 대부분 사용되는 알람은 사용자가 설정한 시간에 기본 알람 소리와 진동으로 사용자를 깨워주는 알람이다. 근래 활용된 알람 중에는 사용자에게 특정 미션을 주어 사용자가 미션을 성공했을 때 알람 소리가 꺼지는 알람 등이 있었지만 본 어플처럼 사용자가 텍스트를 제공했을 때 그것을 연예인의 음성으로 변경하여 알람을 해주는 알람 어플은 출시되지 않았다는 점에서 신선함이 있다.


본 어플을 사용하는 사용자들이 가질 수 있는 기대효과 및 활용방안은 다음과 같다.

● 기대 효과

1. 원하는 연예인의 음성 알람을 받을 수 있다.
2. 연예인이 수동으로 음성을 직접 녹음하여 재생하는 기존의 방법을 깨고 내가 원하는 텍스트 메시지를 직접 읽어준다.
3. 기존 TTS의 불완전한 음성보다 훨씬 자연스러운 음성으로 들을 수 있다.

● 활용방안

1. 기존에 존재하는 연예인 알람 어플은 연예인이 직접 음성을 녹음하여 그 음성을 알람으로 재생해 주는 방법으로 제작 되어있다. 미리 녹음된 음성 안에서 알람이 서비스되어 알람의 폭 자체가 크게 제한될 수 밖에 없었고, 때문에 연예인 어플 시장은 사용자들에게 크게 호응을 얻지 못하였다. 본 어플은 기존 연예인 알람 어플의 구조와 알람 서비스의 제한성을 모두 보완하여 활용하는 방식으로 알람 앱 시장에서의 활약이 기대된다.
2. 기존 TTS 방식의 음성 품질을 보완하는 Tacotron 알고리즘을 사용하여 자연스러운 음성을 반환할 수 있다.

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

3 자기평가

| | |
|----------|--|
| 최종 결과물 | 원하는 텍스트를 연예인의 음성으로 재생하는 알람 어플리케이션 |
| 주요 평가 기준 | 1. 기존의 TTS 방식보다 자연스러운 음성으로 재생되는가? |
| 자기 평가 | <p>기존의 TTS 방식은 인간 개개인의 본연한 말투나 억양, 성조 등을 반영하지 못하고 그저 주어진 문장을 기계적으로 읽는 것에 비해 달람의 TTS 음성은 학습한 유명인의 말투 스타일을 비슷하게 재연하였다.</p> <p>하지만 네이버 오디오북 등에서 출시한 굉장히 자연스러운 인공지능 음성까지는 따라가지 못한다. 입력하는 모든 텍스트를 확실하게 발음할 수 있는 수준까지는 아직 부족하다. 그런 이유로 알람에 필요한 텍스트를 선별해 오버피팅하는 작업을 진행했다.</p> |

| | |
|----------|---|
| 최종 결과물 | 원하는 텍스트를 연예인의 음성으로 재생하는 알람 어플리케이션 |
| 주요 평가 기준 | 2. 선택한 연예인의 실제 음성과 비슷하게 학습되었는가? |
| 자기 평가 | <p>두 팀원의 목소리를 각자 10 시간동안 학습하고, 완성된 모델에 연예인 음성을 3 시간씩 학습시키는 작업을 진행했다.</p> <p>결과물은 실제 연예인 음성과 비슷하게 재연되었다.</p> |

| | |
|----------|---|
| 최종 결과물 | 원하는 텍스트를 연예인의 음성으로 재생하는 알람 어플리케이션 |
| 주요 평가 기준 | 3. 특정 텍스트의 알람이 빠른 시간 내에 생성되는가? |
| 자기 평가 | <p>알람 추가 화면에서 특정 텍스트를 입력하여 알람을 설정하면 약 30 초 이내에 음성이 생성되므로 충분히 빠른 시간 내에 기능을 사용할 수 있다.</p> |

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

4 참고 문헌

| 번호 | 종류 | 제목 | 출처 | 발행년도 | 저자 | 기타 |
|----|------|---|----------------------------|--------|---------------------|----|
| 1 | 서적 | 개발자도 알아야 할 안드로이드 UI 디자인 | acorn+PACKT | 2017 년 | 제시카 쏠비 | |
| 2 | 서적 | 이것이 mysql 이다 | 한빛미디어 | 2016 년 | 우재남 | |
| 3 | 서적 | 강샘의 안드로이드 프로그래밍 | 루비페이퍼 | 2017 년 | 강성윤 | |
| 4 | 서적 | 안드로이드 프로그래밍 정복 1 | 한빛미디어 | 2016 년 | 김상형 | |
| 5 | 서적 | 안드로이드 프로그래밍 정복 2 | 한빛미디어 | 2016 년 | 김상형 | |
| 6 | 기사 | http://www.inven.co.kr/webzine/news/?news=191900 | 인벤 | 2018 년 | 강승진 | |
| 7 | 기술문서 | TACOTRON: TOWARDS END-TO-END SPEECH SYNTHESIS | Google | 2017 년 | Yuxuan Wang 외 | |
| 8 | 웹페이지 | http://yongtech.tistory.com/100 | tistory | 2016 년 | | |
| 9 | 웹페이지 | https://medium.com/@saxenauts/speech-synthesis-techniques-using-deep-neural-networks-38699e943861 | Medium | 2017 년 | | |
| 10 | 웹페이지 | https://guillaumegenthial.github.io/sequence-to-sequence.html | Guillaume genthial blog | 2017 년 | | |
| 11 | 웹페이지 | https://help.github.com/articles/resolving-a-merge-conflict-on-github/ | Github Help | | | |
| 12 | 웹페이지 | https://www.youtube.com/watch?v=3QjGtOliViI&list=PL1Kb3QTCLIVtyOuMgyVgT-OeW0PYXI3j5 | Youtube | 2017 년 | | |

| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

5 부록

5.1 사용자 매뉴얼

1. 초기 화면

애플리케이션에 접속하게 되면 나오는 초기 화면이다.

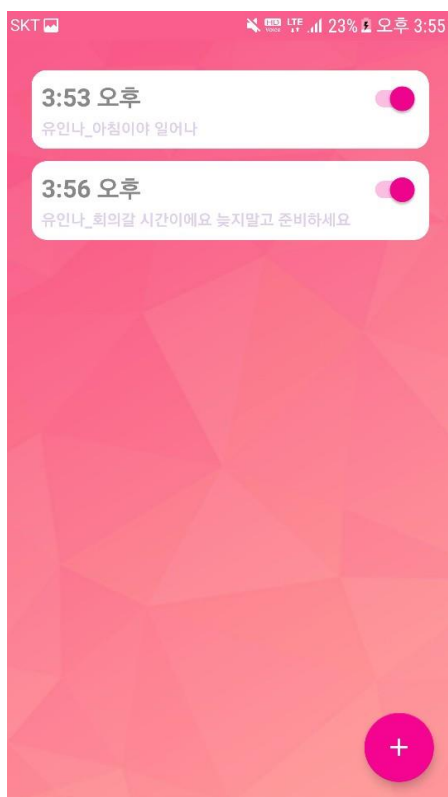


| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

2. 알람 목록 화면

초기 화면을 지나면 알람 목록을 볼 수 있다. 오른쪽 하단의 분홍색 원형 버튼을 누르면 알람 추가 화면으로 이동한다.

알람 목록에는 알람 시간, 사용자가 선택한 연예인 모델, 사용자가 알람으로 들으려는 텍스트가 보여진다. 알람을 키고 끌 수 있는 On-Off 버튼이 오른쪽에 존재한다.



| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

3. 알람 추가 화면

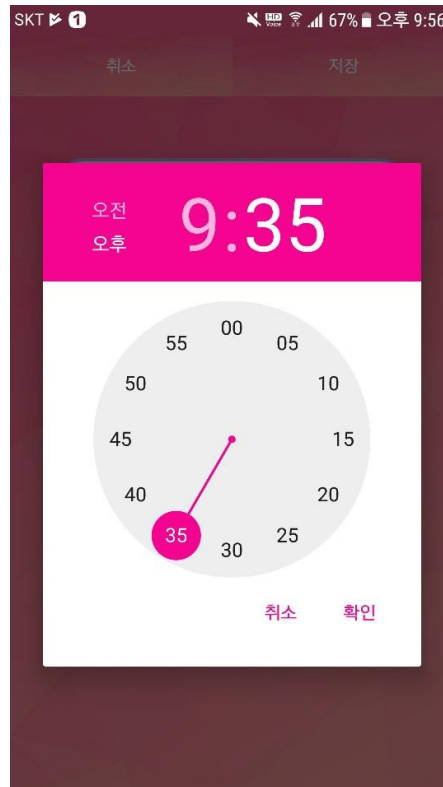
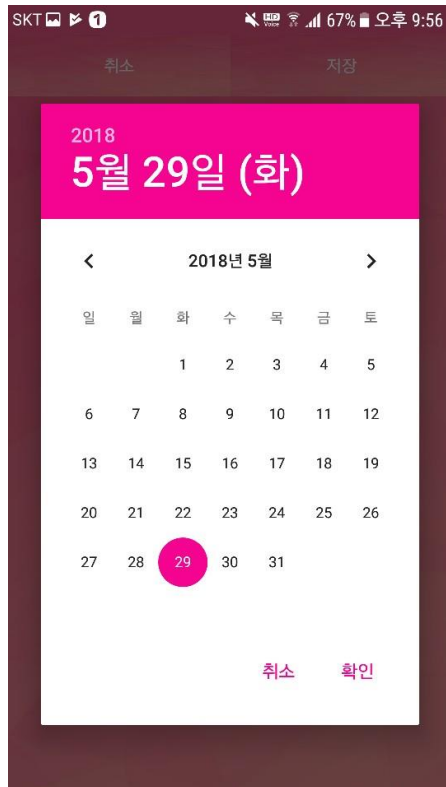
알람 목록 화면에서 오른쪽 하단의 분홍색 원형 버튼을 누를 시 이런 알람 추가 화면이 뜬다. 상단에 알람으로 설정한 날짜와 시간이 보여진다. 하단에서는 목소리 모델을 선택할 수 있다. 모델은 손석희, 유인나, 박명수 중 한 명을 고를 수 있다. 또한, 사용자가 알람으로 들을 텍스트를 입력할 수 있다.



| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

3-1. 날짜 선택 화면 & 시간 선택 화면

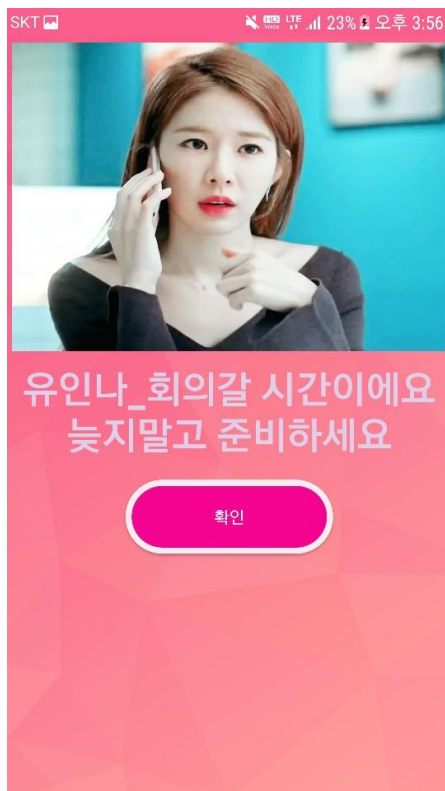
알람 추가 화면 상단에 보여지는 알람 설정 날짜와 시간 버튼을 누르면 아래와 같이 날짜와 시간을 선택할 수 있다.



| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

4. 알람 작동 화면

알람을 설정했다면 사용자가 선택한 모델과 텍스트로 알람이 작동된다. 화면에는 알람 모델 사진과 텍스트가 보여진다. 확인 버튼을 누르면 알람을 중지시킬 수 있다.



| | | | |
|---|-------------------------|-------------|------------|
|  국민대학교 컴퓨터공학부 캡스톤 디자인 I | 결과보고서 | | |
| | 프로젝트 명 | 달람 (Dalarm) | |
| | 팀 명 | 달달한 컴퓨터 달콤 | |
| | Confidential Restricted | Version 1.1 | 2018-05-29 |

5.2 테스트 케이스

| 대분류 | 소분류 | 기능 | 테스트 방법 | 기대 결과 | 테스트 결과 |
|-----|---------|--------------------------------|----------------------------------|---|--------|
| 파일 | 음성파일 생성 | 텍스트를 음성으로 바꾼 파일을 생성한다. | 앱에서 알람 설정 시 특정 텍스트를 입력한다. | 서버에 해당 음성 파일이 생성된다. | 성공 |
| | 음성파일 열기 | 음성으로 바꾼 파일을 연다. | 앱에서 알람 설정 시 특정 텍스트를 입력한다. | 서버에서 해당 음성 파일이 열린다. | 성공 |
| 서버 | 웹서버 | 음성파일을 Streaming 으로 재생한다. | 앱에서 알람을 설정했을 때 음성파일이 재생되는지 확인한다. | 서버에서 앱으로 음성파일이 전송된다. | 성공 |
| 입력 | 스마트기기 | 사용자가 원하는 텍스트와 알람 시간 등을 입력 받는다. | 앱에서 알람 시간을 설정하고 특정 텍스트를 입력한다. | 앱에서 알람시간과 특정 텍스트를 인식한다. | 성공 |
| 출력 | 스마트기기 | 연예인 음성으로 변환된 알람을 출력한다. | 앱에서 알람을 설정할 때 기호에 따라 연예인을 선택한다. | 서버에 있는 특정 연예인의 모델에서 음성 파일이 생성되어 앱으로 전송된다. | 성공 |