
ME:MIC

- 음향 추출 및 분석을 통한 성대모사 도움 어플리케이션

Final Report / Manual



캡스톤디자인(1)

TEAM 자취단

소프트웨어학부 20191312 배준재

소프트웨어학부 20193988 이건희

소프트웨어학부 20190170 윤준서

목 차

1. 개요.....	3
2. Github 리포지토리	3
3. 팀 역할.....	3
4. 개발 동기	3
5. 콘텐츠.....	4
6. 알고리즘 상세와 근거.....	6
7. 매뉴얼	8
8. 마치며	16
9. 주요 참고문헌.....	16

1. 개요



ME:MIC - 음향 추출 및 분석을 통한 성대모사 도움 어플리케이션

'ME:MIC(미믹)'은 성대모사 가이드를 제공하는 웹 서비스입니다.

사용자가 원하는 음원 파일을 첨부하거나 유튜브 링크를 첨부하면 배경음을 분리한 음성 데이터를 얻어 분석에 사용합니다.

최종적으로 사용자가 녹음을 마치면 따라하려는 음성과 녹음된 음성의 전체 유사도를 제공하며, 구간 별 유사도가 가장 낮은 3 개의 구간에 대해 유사도 점수와 점수를 올릴 수 있는 피드백을 제공합니다.

사용자는 서비스 이용으로 성대모사 실력 향상의 경험을 가질 수 있습니다.

2. Github 리포지토리

프론트엔드 : <https://github.com/2023-CapstoneDesign-MEMIC/MEMIC-front-react>

백엔드 : <https://github.com/2023-CapstoneDesign-MEMIC/MEMIC>

3. 팀 역할

이건희 : 프론트엔드

배준재 : 백엔드(입력 데이터 가공 및 MFCC 알고리즘)

윤준서 : 백엔드(포먼트 분류 및 피드백 생성 알고리즘)

4. 개발 동기

- 성대모사의 높은 수요

: 세대를 불문하고 유행이 멈추지 않는 성대모사 개인기

많은 사람들이 성대모사를 배우며 비슷하게 가꾸는 것을 취미로 즐기고 있다.

성대모사는 창작자들에게 더 많은 창작 가능성을 제공하고 엔터테인먼트 능력을 높인다.

- 분석 지표의 부재

혼자 연습할 때 골도청각으로 인한 착각과 객관적인 지표가 없어 어려움이 있었고, 타인에게 들려주기에는 많은 부끄러움이 있을 수 있다.

5. 콘텐츠

일련의 워크플로우를 따름.

1. 음원부 (/fileupload)

2-1. 유튜브 링크

- 1) 링크 입력란
- 2) 임베드 된 유튜브 영상을 보고 구간 설정

2-2. 파일 업로드

- 1) 파일 선택 버튼
- 2) 음원 재생 기능과 함께 구간 설정

2. 녹음부 (/record)

: 녹음 버튼, 녹음한 음성 다시 듣기, 다시 시도하기, 제출하기 버튼

3. 피드백 출력 (/analyse)

3-1. 전체 유사도 점수

: 전체 구간에 대한 유사도 점수와 다시 듣기, 다시 시도하기 버튼

한 줄 평가는 다음과 같이 구성됨

90 점 이상 - "완벽에 가깝습니다! 당신은 성대모사의 달인!"

80 점 이상 90 점 미만 - "매우 훌륭해요! 조금만 더 연습한다면 완벽한 성대모사가 가능할 거예요."

70 점 이상 80 점 미만 - "좋은 시도예요! 일부 발음과 감정 표현에 조금 더 주의를 기울이면 좋겠어요."

60 점 이상 70 점 미만 - "잘 따라하고 있어요, 하지만 명확한 발음과 감정을 담아 더 연습해야 해요."

50 점 이상 60 점 미만 - "발음과 감정을 더 잘 캐치해 보세요. 꾸준한 노력이 필요해요!"

50 점 미만 - "아직 많은 연습이 필요해요. 원본과 비교하며 발음과 감정을 연습해 봅시다."

3-2. 최악의 3 구간 보고서

- 1) 각 구간별 유사도 점수
- 2) 각 구간별 다시 듣기 기능
- 3) Action : 방법 피드백 – 혀의 높낮이, 혀끝의 위치, 입술의 모양, 음의 높낮이에 대한 실제적 피드백 제공

(예 : 더 높은 점수를 얻으려면, 혀의 높이를 더 낮게 내리고, 피치를 높이기 위해 성대를 조여 높은 음을 내세요.)

- 4) Reason : (상세보기) 구분된 Vowel 의 특성과 주파수 값, 높낮이의 근거가 되는 피치에 대한 설명 제공

(예 : 따라하려는 음성은 a 발음에 가깝고, 사용자님의 음성은 ε 발음에 가까워요.

따라하려는 음성의 피치(pitch)는 243.67Hz, 사용자님의 음성의 피치(pitch)는 164.02Hz 입니다.

따라하려는 음성의 a 는 저모음으로 분류되며, 혀의 높이를 매우 낮게 위치해야 하고, 전설모음으로 분류되며, 혀끝의 위치를 매우 앞쪽(입술에 가깝게)에 위치해야 합니다.

사용자님의 음성인 ε 는 중저모음으로 분류되며, 혀의 높이를 약간 낮게 위치한 상태고, 전설모음으로 분류되며, 혀끝의 위치를 매우 앞쪽(입술에 가깝게)에 위치한 상태입니다.)

6. 알고리즘 상세와 근거

- 포먼트(Formant)에 대한 논의

1) 어떤 포먼트를 사용하는가?

포먼트(Formant)는 공명 주파수로, 증폭된 Amplitude 구간의 주파수 집합이다. 낮은 주파수 대역으로부터 포먼트를 F1, F2, F3, ... 로 표기한다.

F1 은 혀의 높이에 대한 조음 표지로 개구도에 비례하는 값이며, F2 는 혀의 전후 위치에 대한 조음 표지로 혀가 전방에 위치할 수록 그 값이 커진다.

주로 음성학 연구에서 F1~F3 까지 음성 구분의 evidence material 로 고려되고 있으며, 더 상위의 포먼트는 높은 음질이 요구되며 사용자 녹음 환경과 파일 자체의 압축 포맷이나 코덱의 종류에 따라 편차가 심하게 발생될 것으로 예측되어 **F1, F2, F3** 를 **Vowel Classification** 의 **Feature** 로 채택하였다.

2) 포먼트 추출은 어떤 방식을 사용하는가?

‘양병곤(2008). 복합음과 대학생이 발음한 모음 포먼트 측정.’에 의하면 “프라트의 burg 방식에서 가장 적은 측정에러가 나왔다.”라고 소개되어 있다. 기존에 자체적으로 개발한 Peak Picking 의 알고리즘은 두 개의 포먼트를 하나로 추정하거나 F4 를 F3 로 추정하는 등의 오류가 발생하는 빈도가 높아 폐기하고 프라트의 burg 방식을 직접적으로 사용할 수 있는 parselmouth 라이브러리를 채택해 그 중 F1, F2, F3, F0(Pitch) 데이터틀 사용했다.

3) Formant-Vowel Classification 에 사용되는 데이터

Train Set 으로 Roy Becker-Kristal 에 의해 공개된 데이터셋을 사용했다.

해당 데이터는 Feature 로 국가, 사용 언어, 방언의 종류, 유전 계통, 발음한 모음, F1, F2, F3 를 갖는 3704 개의 샘플로 구성되어 있다.

해당 데이터셋은 깃허브 리포지토리(백엔드)에 업로드 되어있다.

사용자가 업로드하는 두 개의 음성은 DTW 된 시계열에 대해 burg 방식으로 얻은 F1, F2, F3 값을 사용하며, 결측값이 있는 경우(F3 이상의 포먼트가 측정되지 않은 손실된 음성) Drop 한다.

- 피드백 내용에 대한 논의

1) 피드백 근거로 모음 포먼트와 기본 주파수를 사용하는 이유

'안병섭, 박미영(2004), 모방 발화의 음향음성학적 연구(3) -전문 성대 모사자의 자료를 중심으로-'에 의하면, 성대모사자가 음성적으로 매우 유사한 특징인을 목표로 모방 발화한 경우에 음향적 분석 결과도 일치하는 것으로 보이며, 음성이 전혀 다른 목표 화자를 모방하였을 경우에는 모음 포먼트 값을 모방하는 데도 실패하고 기본주파수의 변화 양상이 두드러짐을 알 수 있다. 결론적으로 모방 발화는 목표 화자의 전반적인 음성적 특징을 모방하는 것이 아니라 가장 특징적인 음성 특징을 모방한다는 사실을 알 수 있다. 이 때, 음향적 분석으로 모음 포먼트의 삼각도 비교와 기본주파수 분포의 비교가 수행되었다.

이에 따라 모음 포먼트를 기반으로 사용자와 기존 화자의 조음 방법을 비교하고 기본 주파수의 경우는 사용자가 실천적으로 받아들이기 쉬운 높낮이를 조절하는 방식의 피드백을 제공한다.

2) 유사도 측정 방법에 대한 논의

두 음성에서 시간 인덱스의 MFCC 배열을 사용해 먼저 DTW 를 수행해 발화 속도의 오차를 없애고, 두 MFCC 에 대해 코사인 거리를 사용해 백분율화 한다. 유사도 척도로는 유클리드 유사도, 코사인 유사도, 자카드 유사도가 있는데, 벡터의 크기가 아닌 방향성(패턴)에 초점을 두는 코사인 유사도를 채택해 사용했다. 그러나 실제로 발화 속도가 중요한 성대모사는 고려하지 않은 것으로, 경우에 따라 유클리드 유사도와 코사인 유사도를 혼합한 유사도 점수 측정 방식을 고려해 보았으나 0.1s 구간에서 이루어지는 구간별 유사도 측정에서 심한 편차가 발생해 보류되었다.

7. 매뉴얼

- 권장 사양

OS : Windows 10, 11 / MacOS*

*MacOS 에서 Apple Silicon(M1, M2)의 경우 Spleeter 관련 오류가 발생할 수 있음.

<https://github.com/deezer/spleeter/issues/607#issuecomment-1021669444>

오류가 발생한다면 위 링크의 가이드를 따르면 정상 실행이 가능함.

브라우저 : Google Chrome

Django : Version >= 3.0

Python : Version 3.8 ~ 3.10

*최종 발표 당시 데모 실행환경 : MacOS, Chrome, Python 3.8

백엔드, 프론트엔드 각 Repository 에서 README 를 참조해 requirements 를 설치할 것.

배포 사이트가 완성되지 않아 각 서버를 로컬에서 구동하고 프론트엔드의 웹 뷰에서 실행 가능.

- Windows 11 에서의 실행 예시

1) 요구사항 설치

백엔드(MEMIC) : pip install -r requirements.txt

pip install django-cors-headers

프론트엔드(MEMIC-front-react) : npm install

2) 각 프로젝트 실행

백엔드(MEMIC) : python manage.py runserver

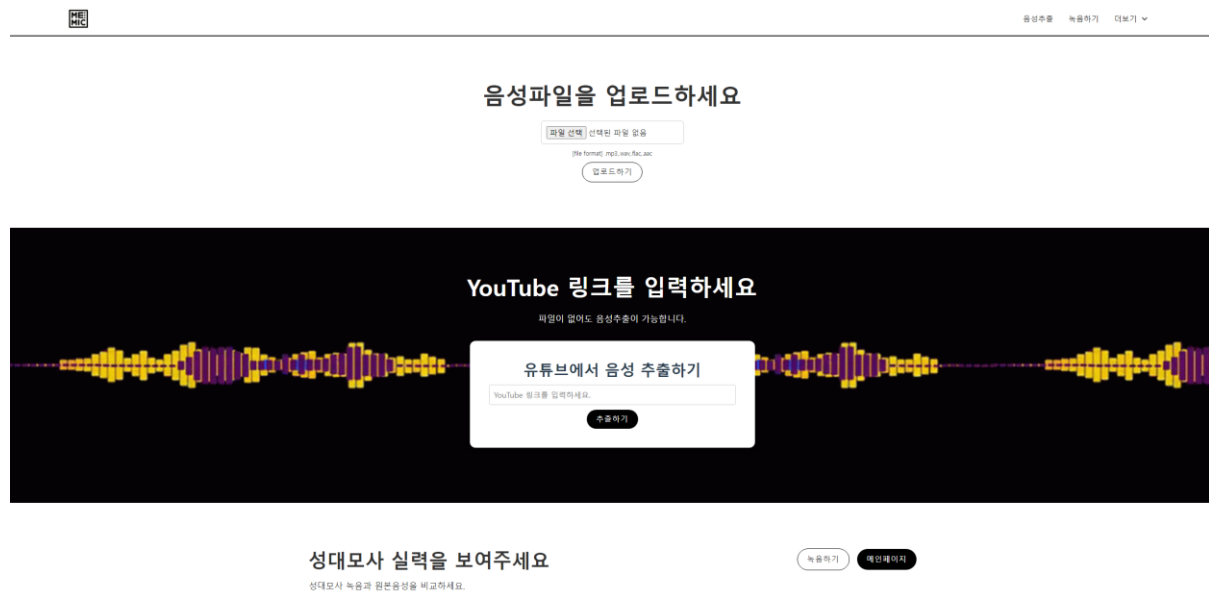
프론트엔드(MEMIC-front-react) : npm start

3) 메인 페이지 (localhost:3000)



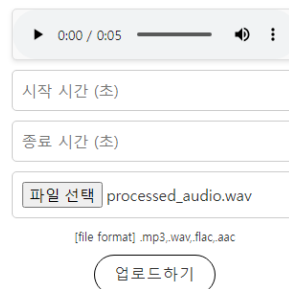
하단 [시작하기] 버튼 클릭

4) 따라하려는 음성 업로드하기 (<http://localhost:3000/fileupload>)



4-1) 상단 : 음성파일 업로드

음성파일을 업로드하세요



[파일 선택] 버튼을 눌러 로컬 파일을 업로드.

업로드 후 해당 파일을 들을 수 있는 오디오 플레이어 생성되며,

시작 시간과 종료 시간을 입력할 수 있음.


설정을 마치고 [업로드하기] 버튼을 클릭하면 5 단계로 이동.

4-2) 유튜브 링크를 입력해 업로드

YouTube 링크를 입력하세요

파일이 없어도 음성추출이 가능합니다.

유튜브에서 음성 추출하기



다음 작품에 봄은 올까..?
기다렸지. 기다렸지.

동영상 더보기

0:39 / 4:01

최대한 정확한 시간을 입력해주세요!

링크를 입력하면 유튜브 동영상 플레이어가 생성되며,

시작 시간과 종료 시간을 입력할 수 있음.

설정을 마치고 [추출하기] 버튼을 클릭. 5 단계로 이동.

5) 녹음하기 (<http://localhost:3000/record>)

성대모사 실력을 보여주세요

성대모사 녹음과 원본음성을 비교하세요.

녹음하기

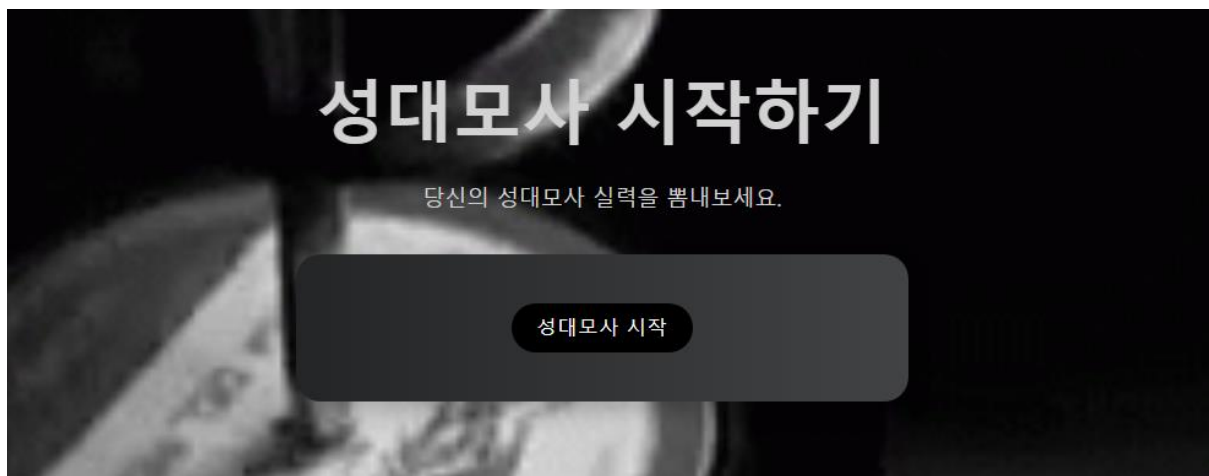
메인페이지



© 2023 MEMIC. All rights reserved.

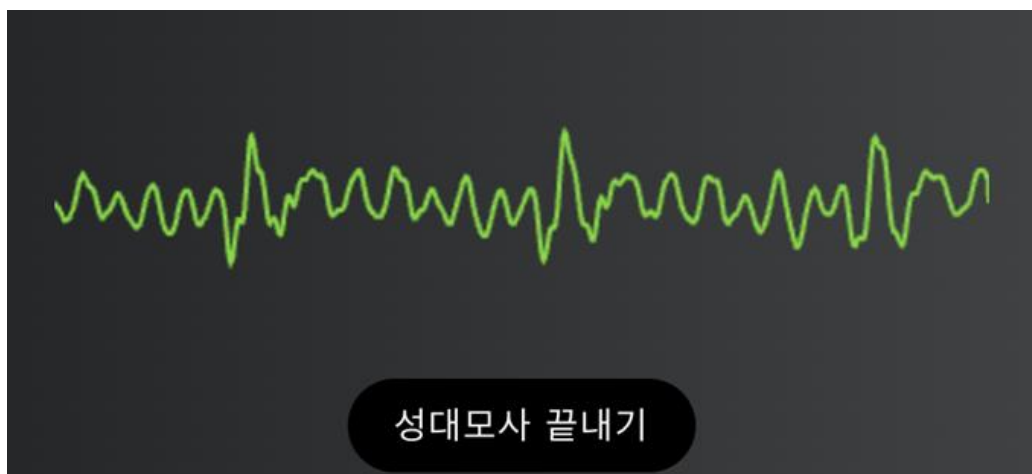


4 단계를 마치고 하단 [녹음하기] 버튼을 클릭 -> 페이지 이동



[성대모사 시작] 버튼을 누르면 즉시 녹음이 시작됨.

마이크 권한 설정 팝업이 나타나면 [허용] 버튼을 클릭.



녹음 진행 중에는 다음과 같은 인터페이스가 나타남.

[성대모사 끝내기] 버튼을 누르면 녹음이 종료되며,

녹음한 내용을 다시 들을 수 있는 오디오 플레이어가 생성됨.

다시 성대모사를 시도하려면 [다시하기] 버튼을,

마치려면 [제출하기] 버튼을 클릭.

결과보기

모든 준비가 끝났습니다.
이제 결과보고서를 받아보세요.

결과보기

파일 업로드부터 다시하기

[결과보기] 버튼을 눌러 6 단계로 이동.

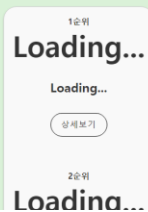
6) 성대모사 평가 및 피드백 보고서 (localhost:3000/feedback)

성대모사 평가



세부분식

피드백 보고서



로딩을 기다리면 다음과 같은 내용이 출력됨.

성대모사 평가



상단 화면에서 전체 유사도 점수를 확인할 수 있음. (사용자의 목표)

그래프의 하단에 점수에 대한 한 줄 평가가 나타남.

이하 하단의 [따라하려는 음성 듣기], [내 목소리 듣기] 버튼을 누르면 하단에 각 음성의 오디오 플레이어 생성되고,

[다시 녹음하기] 버튼을 누를 경우, 5 단계 녹음하기 페이지(/record)로 이동해 따라하려는 음성을 다시 업로드할 필요 없이 녹음부터 수행 가능.

[세부분석] 버튼을 누르거나 스크롤을 내리면 다음과 같은 피드백 보고서를 확인할 수 있음.



피드백 보고서는 1 순위, 2 순위, 3 순위의 총 세 개의 피드백으로 구성됨.

먼저, 각 구간의 유사도 점수와 함께, 따라하려는 음성과 내 음성이 몇 초에서 평가된 것인지 확인할 수 있으며, [상세보기] 버튼을 누르면 다음과 같은 피드백 문장과 오디오 플레이어 생성됨.

더 높은 점수를 얻으려면, 혀의 높이를 더 낮게 내리고, 피치를 높이기 위해 성대를 조여 높은 음을 내세요.

따라하려는 음성은 a 발음에 가깝고, 사용자님의 음성은 ε 발음에 가까워요.
따라하려는 음성의 피치(pitch)는 243.67Hz, 사용자님의 음성의 피치(pitch)는 164.02Hz입니다.
따라하려는 음성의 a는 저모음으로 분류되며, 혀의 높이를 매우 낮게 위치해야 하고, 전설모음으로 분류되며, 혀끝의 위치를 매우 앞 쪽(입술에 가깝게)에 위치해야 합니다.
사용자님의 음성인 ε는 중저모음으로 분류되며, 혀의 높이를 약간 낮게 위치한 상태고, 전설모음으로 분류되며, 혀끝의 위치를 매우 앞 쪽(입술에 가깝게)에 위치한 상태입니다.

상단 굵은 글씨가 피드백의 Action, 하단의 얇은 글씨가 피드백의 Reason에 해당함.

▶ 0:00 / 0:03

시작 시간: 0.07 종료 시간: 3

구간반복 OFF

그 아래에는 따라하려는 음성의 구간반복 기능이 제공되는 오디오 플레이어가 생성되며 시작 시간은 자동으로 입력됨. (바로 아래에, 사용자 음성에 해당하는 플레이어가 생성됨.)

8. 마치며

사용자는 종합 유사도 점수의 상승을 목표로 성대모사의 발전을 꾀할 수 있다. 사용자는 70 점 이상의 유사도에서 긍정적인 한 줄 평가를 받을 수 있어, 높은 점수를 받겠다는 의욕을 고취시킨다.

효과적인 성대모사 능력의 향상을 위해 우리는 사용자의 약점을 알려준다. 보완 범위를 최악의 3 구간으로 좁혀 사용자가 집중할 수 있도록 했고, 개인차가 있을 수 있으나 실천가능한 피드백을 제공해 직관성을 높이도록 했다. 사용자는 약점을 반복적으로 보완하며 점진적으로 능력을 높일 수 있다.

해당 서비스는 음향음성학 연구에 기초하여 피드백의 근거를 마련했으며, 또한 음향음성학 연구에 주로 사용되는 메이저한 오픈소스 라이브러리를 활용했다. 피드백의 타당성과 신뢰도를 높이기 위해 노력했다.

9. 주요 참고문헌

안병섭, 박미영(2004). 모방 발화의 음향음성학적 연구(3) -전문 성대 모사자의 자료를 중심으로-'

양병곤(2008). 복합음과 대학생이 발음한 모음 포먼트 측정

이주현, 장현숙, 정한진(2005). 한국어 음소의 주파수 특성에 관한 연구

Honglin Cao^{a,b,c} &Volker Dellwoc(2019). THE ROLE OF THE FIRST FIVE FORMANTS IN THREE VOWELS OF MANDARIN FOR FORENSIC VOICE ANALYSIS.