

마비말 장애인을 위한 STT, TTS 어플 개발

프로젝트 수행 기간 : 2024.05 ~ 2025. 02

진행중 (25. 02월 05일 발표 예정), 배리어프리 앱개발 콘테스트, 현대오트모버

프로젝트 개요

- 65세 이상의 노인약자가 약 20%를 차지, 2050년에는 약 45% 이상 전망, 신체적 약화는 정신적으로 많은 문제를 야기, 고립 및 소외감으로 인한 우울증 및 고독사 발생
- 장애 인구의 고령화와 신경학적 장애로 인한 후천성 언어장애 발병증가.
- 뇌신경질환으로 인한 언어장애 및 뇌졸중 환자의 25~40%가 실어증을 경험. 제한된 언어 사용으로 인하여, 의사소통 능력의 부재 직장, 사회, 결혼생활 등의 다양한 대인관계의 소외 야기

어플리케이션의 기능 및 개발

어플리케이션 기능 소개

- React Native 활용 프론트엔드 설계. 마비말 장애인의 음성을 wav 파일로 변환, 앱에 변환된 TTS 출력.
- Fast API를 통해 백엔드 설계, wav 파일을 저장하고 user와 wav 파일의 경로를 저장 후, 텍스트를 발화자의 음성으로 변환
- Pytorch를 통해 음성 데이터를 불러오고 마비말장애 특화 음성 인식을 설정, 음성 합성기를 통하여 발음이 조정된 음성 파일을 출력

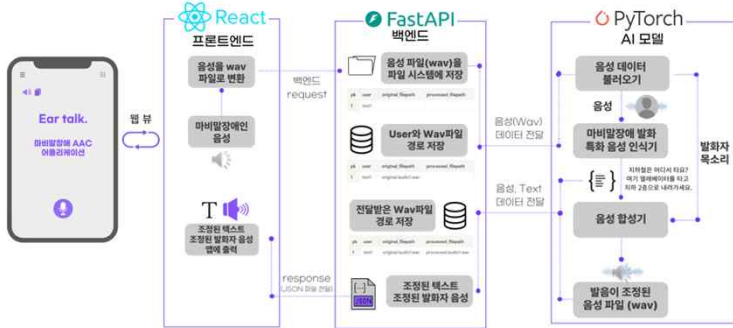


그림 1. 어플리케이션 기능 플로우차트

마비말 장애 특화 음성 인식기 개발

수집년도	모델 학습용 데이터	데이터 종류/양
2021	구음장애 음성인식 데이터	뇌신경 장애: 1,042시간 후두장애: 2,445시간 언어청각장애: 1,519시간
2022	파킨슨병 및 관련 질환 진단 음성 데이터	파킨슨(전) 200명 (후) 623명, 본태성 진전 (전)297명 다계통위축 (전) 170명, 진행성 핵상 마비, 피질기저핵변성 (전) 150명 정상 385명 총 1862 case
2023	구음장애인 명령어 데이터	뇌경색 867,343건, 뇌출혈 6,039건, 루게릭 203건, 청각장애 2,102,837건 합계 2,976,422건

표 1. 학습용 음성인식 데이터 통계

- Waveform을 Mel_Spectrogram 특징추출 기법을 활용하여 음성의 특징 추출
- Self Attention을 통하여 조정된 텍스트를 추출

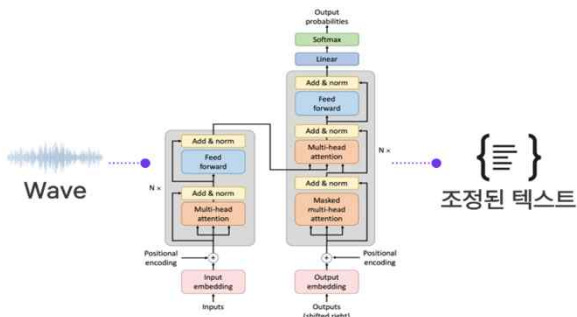


그림 2. Transfomer 기반 마비말 장애 특화 음성 인식기 개발

Zero-shot TTS 음성 합성기 사용

- 조정된 텍스트와 발화자의 목소리를 사용하면 발화자의 발음이 조정된 (명료도가 증가한) 음성 출력이 가능하다.

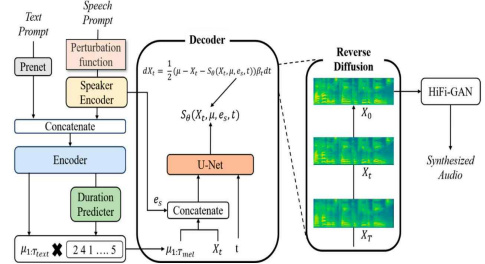


그림3. Zeroshot TTS 기반 음성 생성

데이터베이스 테이블 설계

User	Audio
id integer	id integer
username varchar(20)	user_id integer
password varchar(255)	original_filepath varchar(255)
age integer	processed_filepath varchar(255)
sex bool	processed_text text
	create_date datetime

그림 4. 데이터베이스 각 타입 설정

User Table	Audio
id user 테이블 기본(Primary Key)키	id 테이블 기본키
username 사용자의 ID, 최대 20자로 제한	user_id 사용자의 ID, 최대 20자로 제한
password 사용자의 암호화된 비밀번호, 최대 255자로 제한	original_filepath 웹 서버에 저장된 원본 음성 파일을 불러오기 위한 파일경로
age 사용자의 나이	processed_filepath AI 모델을 거쳐 처리된 오디오 파일 경로
sex 사용자의 성별(1:남성,0:여성)	processed_text 원본 음성에서 생성된 텍스트. 길이 제한 없음
	create_date 모든 처리가 다 끝나고 데이터가 생성된 시간

표 2. 데이터베이스 테이블, 오디오 처리 및 회원가입 구현

- 회원 가입 및 서비스 이용 약관을 통하여, 음성 녹음이 되고 있음을 명시.
- 버튼 하나로 녹음 및 목소리 Text변환 Text 에서 TTS까지 가능하도록 설계

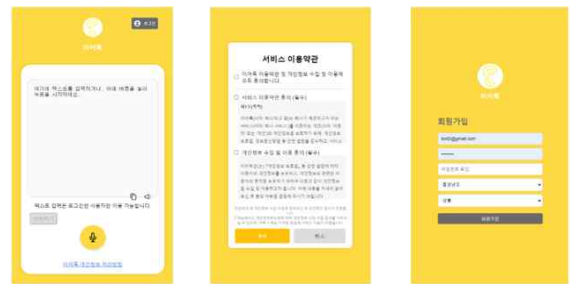


그림5. 어플리케이션 구동 확인

기대효과

- 배리어프리(Barrier-Free) 특화 고정된 음성인식 및 합성 기술
마비 말 장애 환자의 음성 특성을 파악하여 개인화된 음성인식 및 합성 기술을 개발. 최신 AI 기술로 실시간 음성 변환과 다양한 음성 합성을 제공, 효과적인 의사소통을 지원.
- 기술의 지속적인 발전과 확장 가능성
지속적인 데이터 수집으로 알고리즘 정확도가 향상되고 맞춤형 개인화 모델을 제공 가능. 마비 말 장애 환자 외에도 청각장애인과 아동에게 서비스를 확장.
- 마비말장애 환자의 사회 참여 및 다양성과 포용성 증진
AAC 도구를 통해 교육, 직장, 일상생활에서 환자의 사회 참여와 자립을 지원. 맞춤형 음성 생성 기술로 의사소통 장벽을 낮추고 사회적 소외감 해소에 기여.