

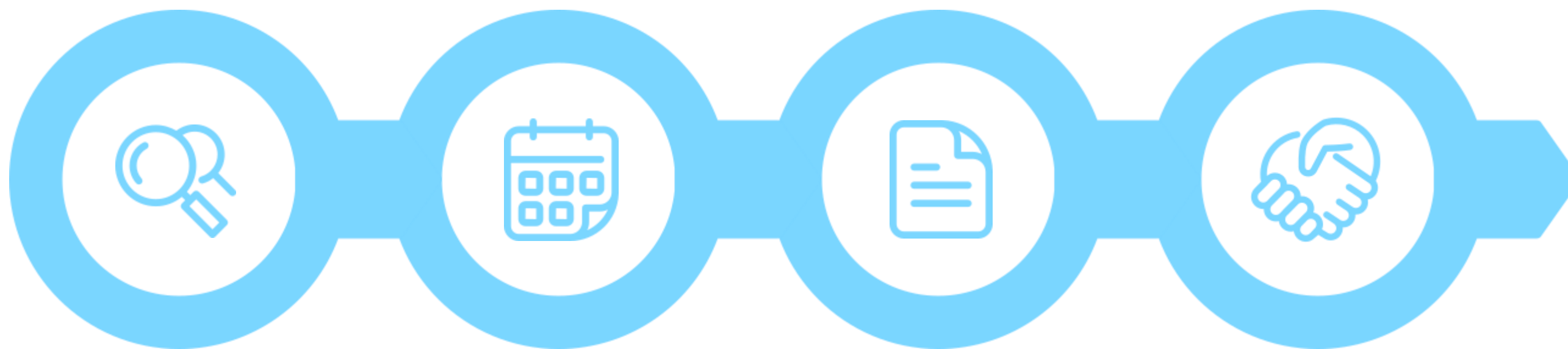
공공시설 음성 인식 키오스크



VOA(Voice Of All)



목차



STEP1

배경 및 개선점

STEP2

서비스 구성도

STEP3

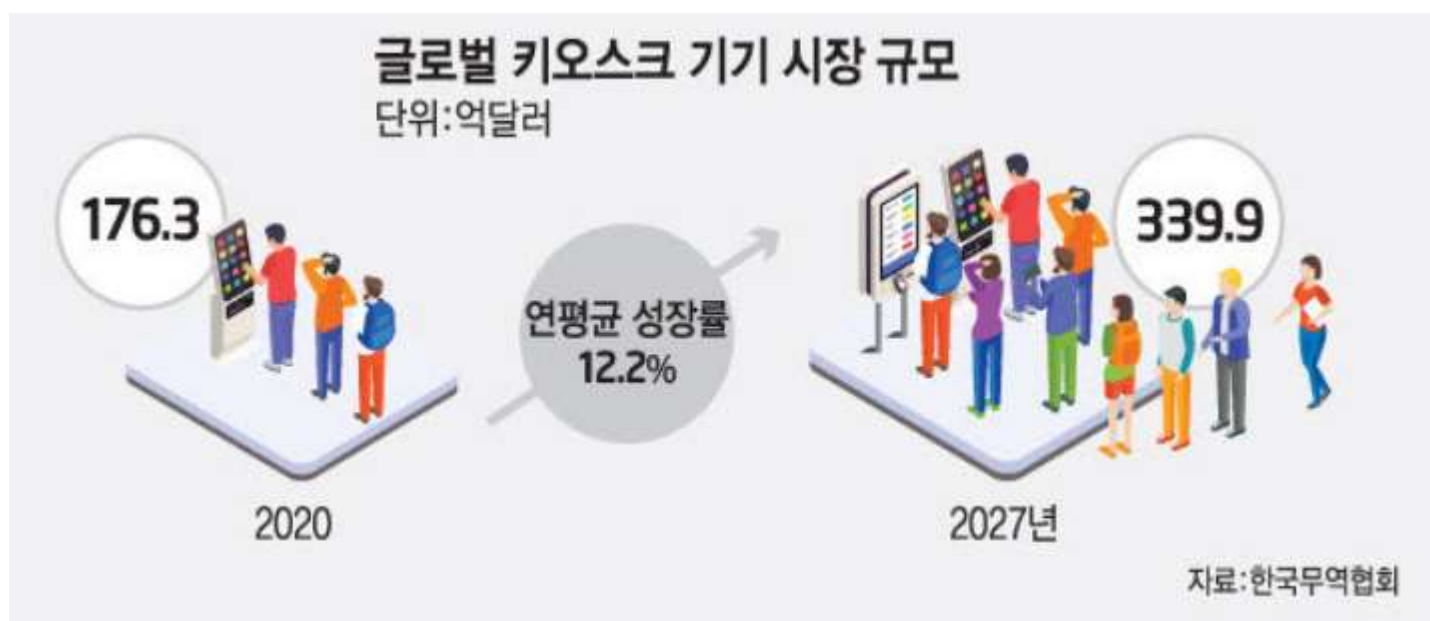
사용된 기술

STEP4

기대효과

주제 선정 배경

1. 키오스크 시장 규모 증가



- 키오스크 시장 규모가 매년 가파른 상승세를 보이고 있음
- 한국무역협회는 2020년에 176.3억 달러의 규모가 2027년에는 339.9억 달러의 규모로 성장할 것으로 예측했다.

2. 시각장애인의 정보 격차 현황

"일상화된 키오스크, 시각장애인에겐 차별"...인권위에 진정

장애인은 키오스크 3년 뒤에 써라? 또 '단계적 적용' 논란

장애인에겐 너무 먼 키오스크...앞으로 음성안내 제공해야

"일상화된 키오스크, 시각장애인에겐 차별"...인권위에 진정

- 키오스크는 일상화가 되었지만 시각장애인의 경우 그 혜택을 누리지 못하고 있다.
- 음성안내 기능이 삽입된 키오스크는 계속 미뤄지고 있다.

주제 선정 배경

3. 장애인 차별 금지법

제15조(재화·용역 등의 제공에 있어서의 차별금지) ①재화·용역 등의 제공자는 장애인에 대하여 장애를 이유로 장애인 아닌 사람에게 제공하는 것과 실질적으로 동등하지 않은 수준의 편의를 가져다주는 물건, 서비스, 이익, 편의 등을 제공하여서는 아니 된다.

②재화·용역 등의 제공자는 장애인이 해당 재화·용역 등을 이용함으로써 이익을 얻을 기회를 박탈하여서는 아니 된다.

③ 재화·용역 등의 제공자는 무인정보단말기(터치스크린 등 전자적 방식으로 정보를 화면에 표시하여 제공하거나 서류발급, 주문·결제 등을 처리하는 기기를 말한다)를 설치·운영하는 경우 장애인이 장애인 아닌 사람과 동등하게 접근·이용할 수 있도록 하는 데 필요한 정당한 편의를 제공하여야 한다. <신설 2021. 7. 27.>

④ 제3항에 따른 재화·용역 등의 제공자의 단계적 범위 및 정당한 편의의 구체적인 내용 등 필요한 사항은 대통령령으로 정한다. <신설 2021. 7. 27.>

- 무인정보단말기(키오스크)에 대한 장애인의 접근성 보장을 명시한 장애인 차별 금지법 개정안이 2023년 1월 28일부터 시행됨
- 하지만 복지부의 단계적 적용 방침에 따라 시행령은 3년 후인 2026년부터 적용 예정

기존 배리어프리 키오스크



닷 사의 배리어프리 키오스크

- 배리어프리 : 장애인과 비장애인의 장벽을 허무는 일
- 사용자 인지 높이조절 시스템/음성안내/점자 디스플레이 등 다양한 기능이 탑재되어 있음

‘배리어프리 키오스크’ 1002대 중 1대... 주문조차 할 수 없는 장애인

- 하지만 일반 키오스크의 가격은 2~300만원인 것에 비해 배리어프리 키오스크의 가격은 2000만원이기 때문에 설치가 어려움

심층 인터뷰

성별, 연령

- 50대 남성

장애 등급

- 전맹 시각장애인



키오스크의 불편한 점

1. 위치를 찾는 데 어려움이 있음
2. 키오스크의 문구를 읽기 어려움
3. 음성 안내 기능이 제공되지 않음

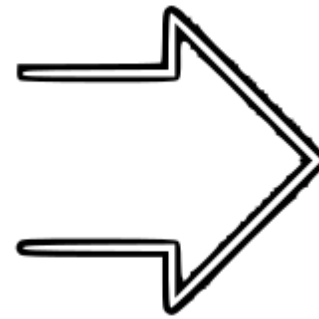
키오스크에 바라는 점

- 음성 안내 기능이 여러 번 반복되길 희망함

기존 키오스크와의 개선점

기존 키오스크의 문제점

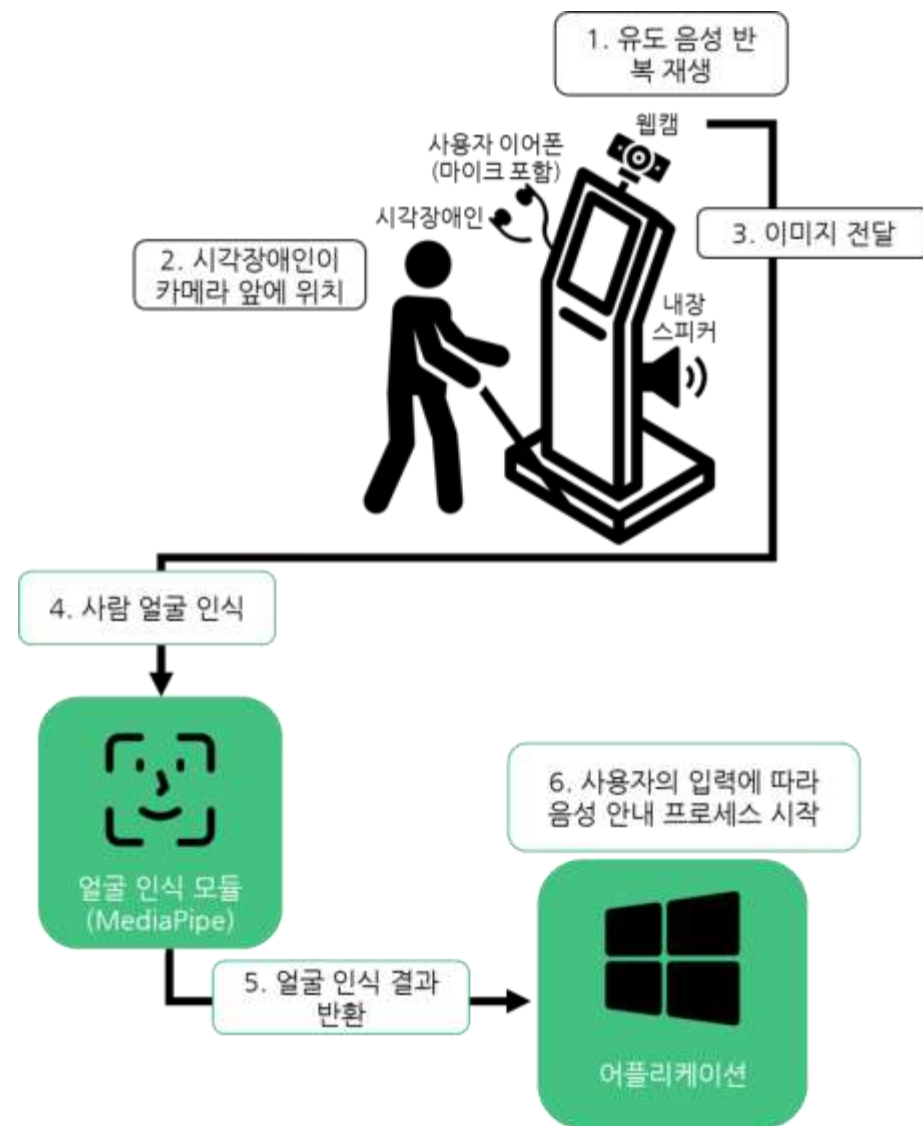
1. 위치를 찾기가 어려움
2. 키오스크의 문구를 읽기 어려움
3. 음성 안내 기능이 제공되지 않음



개선점

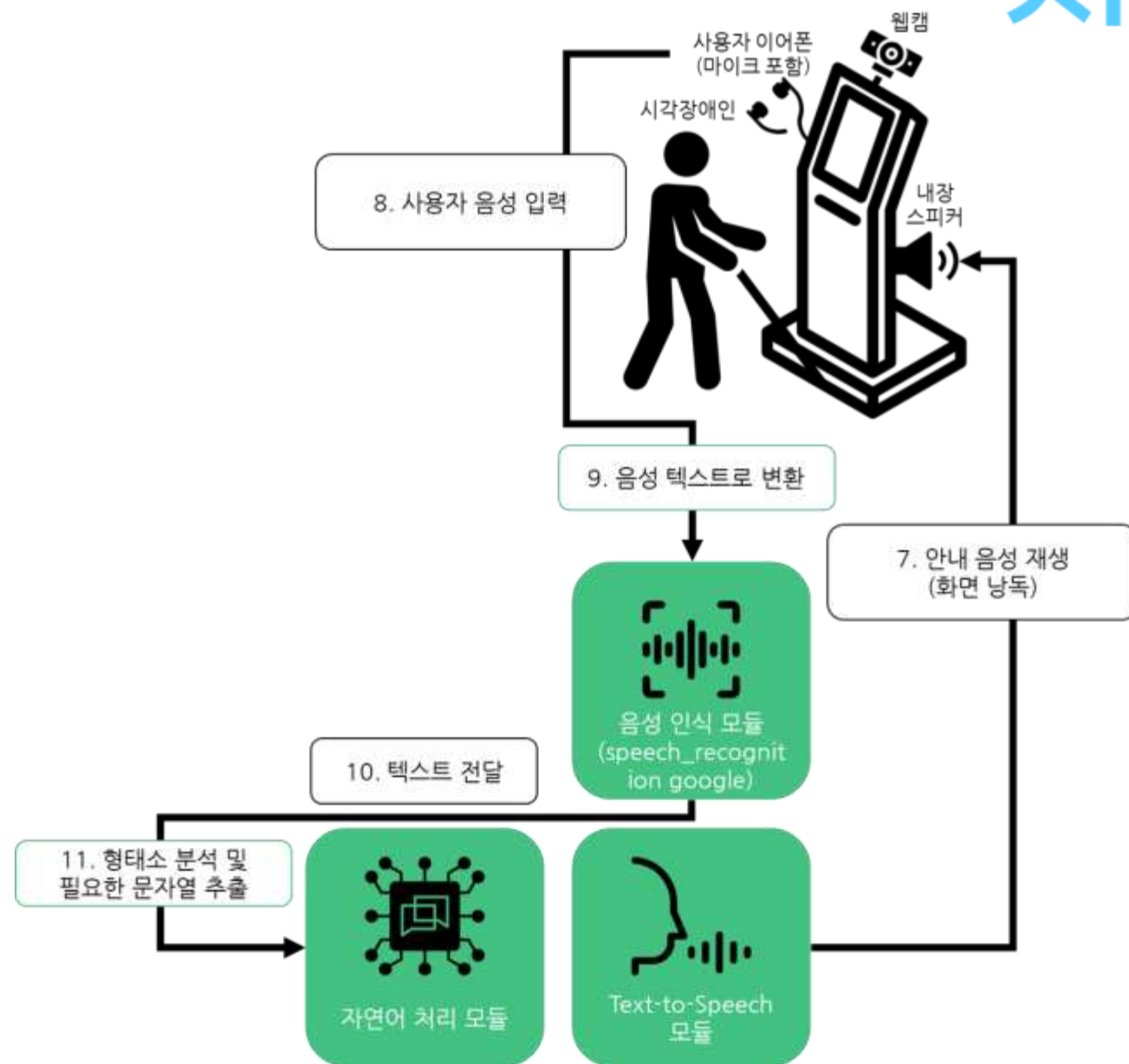
1. 10초마다 키오스크의 위치를 안내하는 유도 음성을 재생함
2. 간단한 터치 이외에는 모든 조작을 음성으로 진행함
3. 구글의 TTS를 통해 화면의 내용을 낭독함

서비스 구성도



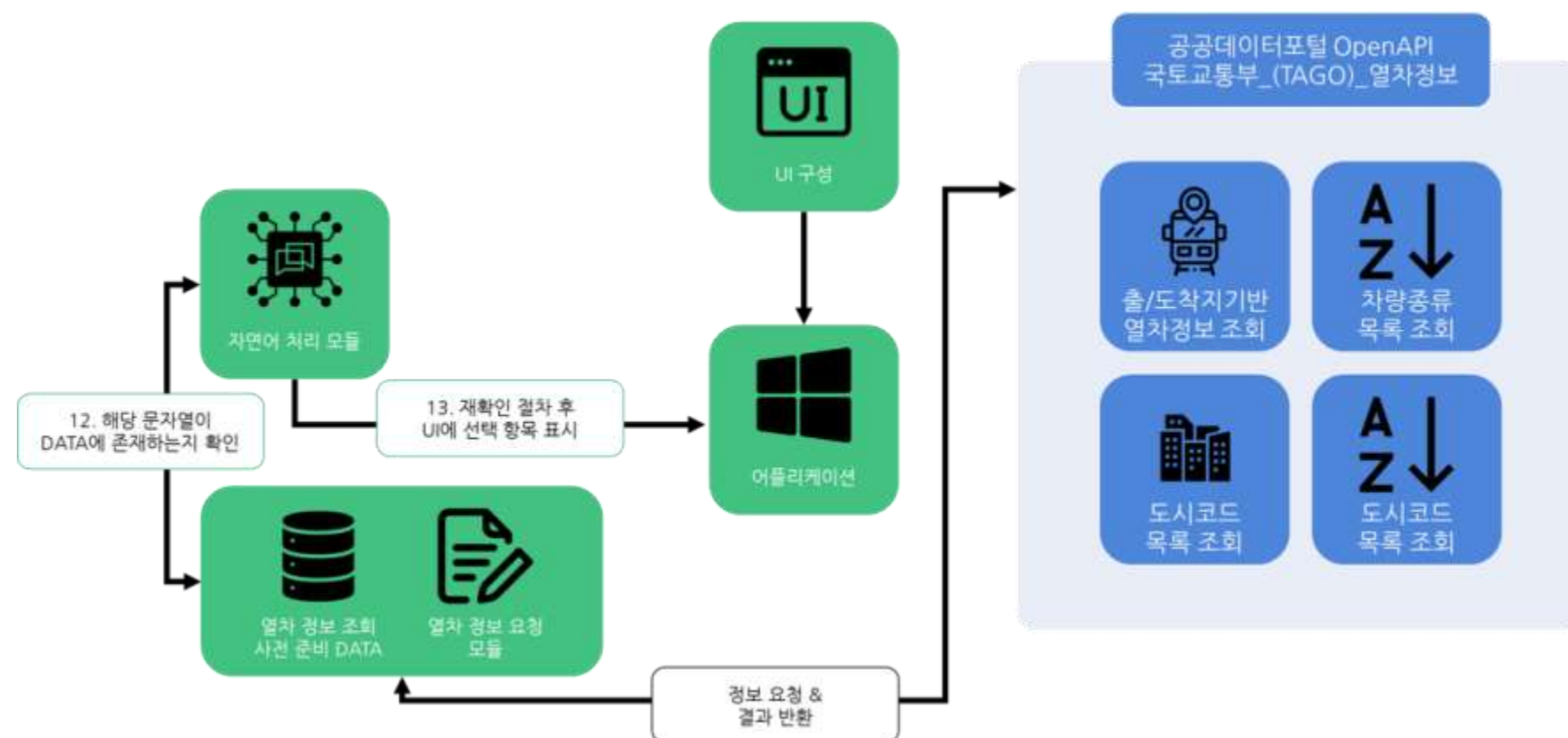
1. 키오스크에서 10초 간격으로 유도음성 반복 재생
2. 사용자가 키오스크에 접근
3. 키오스크의 웹캠에서 이미지를 인식해 MEDIAPIPE로 전달
4. MEDIAPIPE로 사람 얼굴 인식
5. 얼굴 인식 결과 반환
6. 사용자의 입력에 따라 음성 안내 프로세스 시작

서비스 구성도



- 7. 키오스크에서 안내 음성 재생
- 8. 사용자가 음성으로 답변
- 9. GOOGLE STT에서 입력된 음성을 텍스트로 변환
- 10. 텍스트를 자연어 처리 모듈로 전달
- 11. 자연어 처리 모듈에서 형태소 분석 및 필요한 문자열 추출

서비스 구성도



12. 해당 데이터가 기존에 설정한 값과 일치하는 지 확인

- 역 정보 : OPENAPI로부터 얻은 정보를 미리 로컬에 저장하고 이와 대조
- 시간, 가격, 열차 등 정보 : OPENAPI로 서버에 요청해 실시간으로 결과 반영

13. 사용자가 입력한 정보가 적합하다면 사용자에게 입력된 정보를 낭독해 의사를 재확인 후 의사결정

서비스 화면 및 기능

1

승차권 구매

승차권 환불

예약표 찾기

예약 취소

2

뒤로 가기 승차권 예매 처음으로

출발 도착

서울 서울

서울 용산 광명

영등포 수원 평택

천안아산 천안 오송

조치원 대전 서대전

김천구미 구미 동대구

1. 진입 화면

- 10초 간격으로 유도음성을 재생시켜 시각장애인 사용자가 키오스크 위치 인식을 용이하게 함
- 사용자가 카메라 앞에 서면 얼굴 인식 후 음성인식 모드로 전환
- 터치 입력이 들어오면 음성 안내 종료

2. 기차역 선택 화면

- 저시력 시각장애인을 고려한 UI로 변경
- 자연어 처리를 통한 문장 인식
- 인식 실패할 경우 레벤슈타인 거리 계산으로 발음이 유사한 기차역으로 인식

서비스 화면 및 기능

3

뒤로 가기 승차권 예매 처음으로

출발 서울 도착 부산

2023년 11월 01일 (수요일) 05:53

수 목 금 토 일

1 출발임 02 03 04 05 06

00시 01시 02시 03시 04시 05시 06시 07시

조회

4

뒤로 가기 승차권 예매 처음으로

출발 서울 도착 부산

2023년 11월 3일 (금)

열차	출발	도착	요금
KTX 17	08:57 서울	11:15 부산	59800
ITX-새마을 1003	08:48 서울	13:42 부산	42600

3. 날짜 선택 화면

- 현재 시간으로부터 이전 시간은 선택되지 않음
- 먼저 날짜를 선택하고 시간을 선택함

4. 탑승 기차 선택 화면

- 선택된 도착 기차역, 도착 시간을 바탕으로 코레일의 기차 정보 API를 조회하여 가장 빠른 기차 2개를 추천함
- 화면을 터치하면 가장 가까운 기차가 선택되고 터치하지 않으면 두번째 기차가 선택됨

서비스 화면 및 기능



5. 인원 선택 화면

- 성인, 어린이, 경로, 중증장애인의 경우에 따라 가격이 다르게 계산됨
- 음성 안내 모드의 경우 탑승 인원은 1명으로 한정됨

6. 좌석 선택 화면

- 창측, 통로측을 선택한 후 좌석이 랜덤으로 배정됨
- 통로측 좌석으로 기본 설정되어있고, 화면을 터치하면 창측 좌석으로 선택됨

서비스 화면 및 기능

7

승차권 예매

출발: 서울(08:48)
도착: 부산(13:42)
열차: ITX-새마을 1003
가격: 42600원
성인(42600원): 1
어린이(21300원): 0
경로(29800원): 0
중증장애인(21300원): 0
선택한 좌석(1호차): 2A

결제하기

8

휴대폰 번호를 입력해주세요.

1	2	3
4	5	6
7	8	9
지우기	0	전부 지우기
확인		

7. 선택 정보 확인 화면

- 웹캠으로 사람을 인식하여 15초 이상 자리 비움 시 초기 화면으로 돌아감
- 화면에 표시된 사용자가 선택한 정보를 사용자에게 음성으로 안내하여 선택한 정보가 맞는 지 확인

8. 개인정보 입력 화면

- 사용자가 음성으로 휴대폰 번호 입력
- 입력된 번호가 -을 제외하고 11자리인지 확인
- 유효하지 않은 번호라고 판단되면 재입력 또는 사용자의 의사에 따라 입력 종료
- API를 이용하여 사용자가 입력한 번호로 예매 정보 문자 발송

기술 선택 기준

1. 한국어 STT 성능 비교 (단위:CER)

api/데이터셋	콜센터 데이터	저음질 데이터
openai whisper	7.51	17.27
Google STT	8.37	14.11
ETRI	8.36	15.46

기차역 특성 상 저음질 환경에서 가장 정확한 성능을 내는 GOOGLE STT를 선택

$$CER = \frac{S + D + I}{N}$$

S : 음성 인식된 텍스트에 잘못 대체된 음절 수
D : 음성 인식된 텍스트에 잘못 삭제된 음절 수
I : 음성 인식된 텍스트에 잘못 추가된 음절 수
N : 정답 텍스트의 음절 수

기술 선택 기준

2. 얼굴 인식 모듈 기능 비교

프레임워크/기능	얼굴 인식	포즈 인식	크로스 플랫폼	실시간 처리	다양한 모델
MediaPipe	O	O	웹, 모바일, 데스크톱	O	여러 모델
OpenPose	O	O	데스크톱	O	한 가지 모델
DeepPose	X	O	데스크톱	X	한 가지 모델
PoseNet	X	O	웹, 모바일	O	한 가지 모델

기술 선택 기준

3. 역 인식을 못할 경우 유사 역 반환 알고리즘 발전 과정

1. 코사인 유사도

문서에서 각 단어가 문서에 나타나는 횟수로 표현되는 벡터값



코사인 유사도가 단어의 빈도만 고려하기 때문에 발음의 유사도와 무관하게 '상봉역'만 반환하는 문제 발생

2. 레벤슈타인 거리

문자 삽입, 삭제, 교체 등의 연산을 최소한으로 사용하여 한 문자열을 다른 문자열로 변환하는 데 필요한 비용



한 문자열의 기준을 한 음절로 계산했기 때문에 '마산역'과 '용산역'과 같이 한 글자만 다른 역들의 계산 결과가 동일한 문제 발생

기술 선택 기준

3. 유사한 역 반환 알고리즘 발전 과정

3. 자모 분리+레벤슈타인 거리

한 문자열의 기준을 음소(초성, 중성, 종성)
으로 나눠서 레벤슈타인 거리 계산

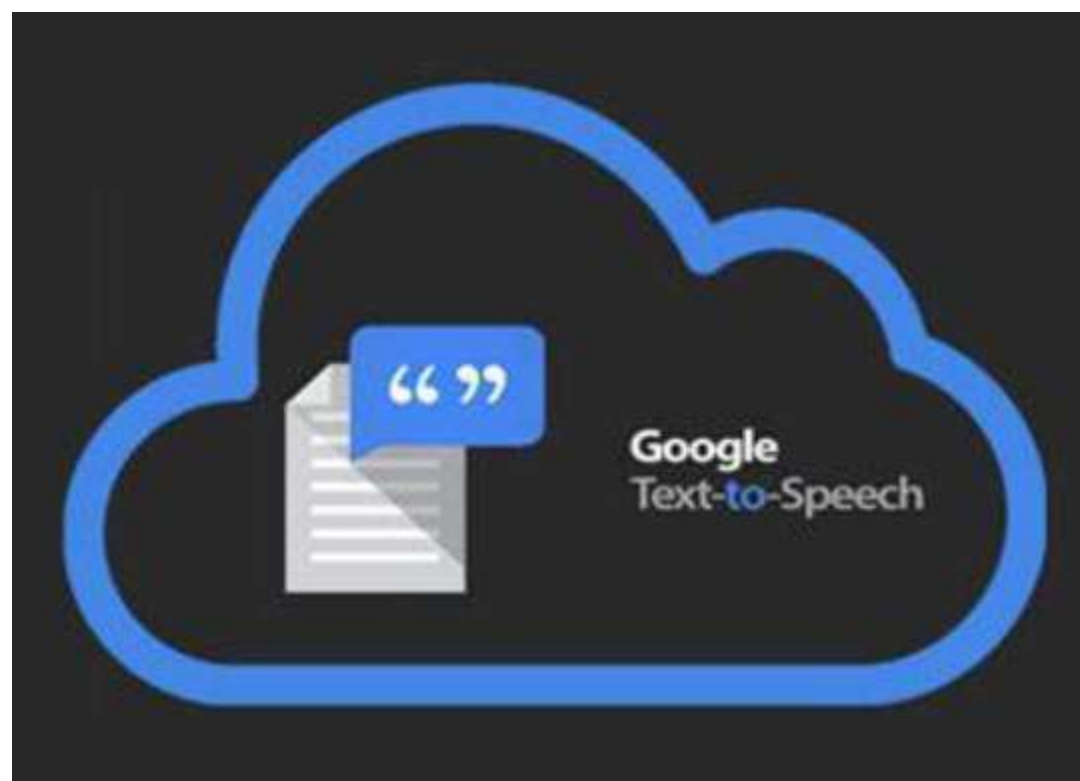


'감곡장호원역'과 같이 발음을 인식하기 어려운
역도 알고리즘을 통해 정확히 인식하게 됨

사용한 기술

1. GOOGLE SPEECH-TO-TEXT

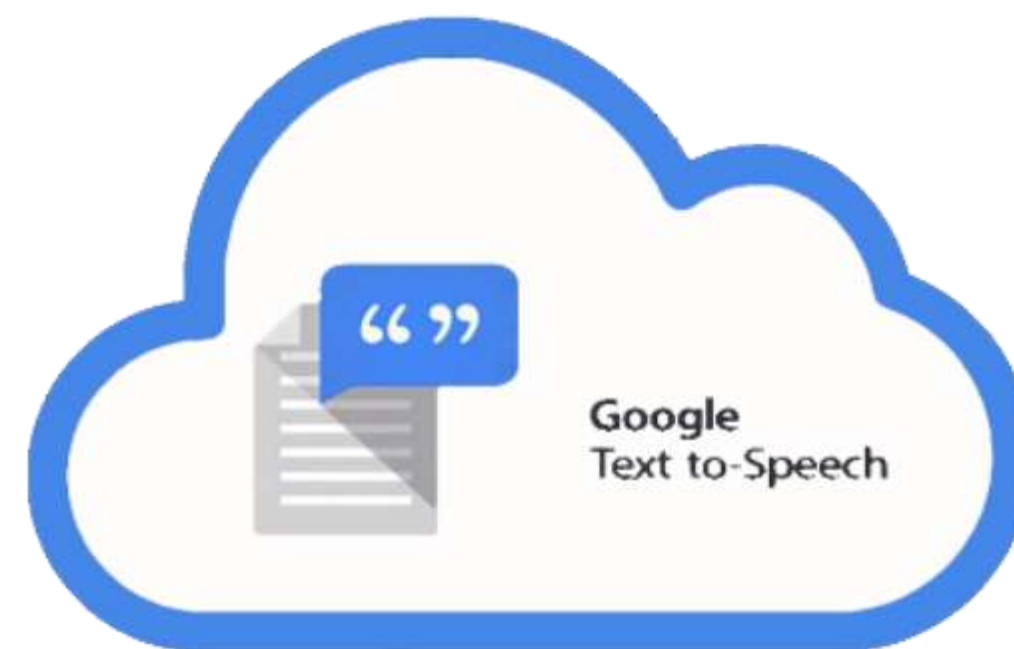
- 음성 인식 API를 제공하여 사용자가 음성 데이터를 텍스트로 변환할 수 있게 함
- 다양한 언어에 대해 거의 인간 수준의 이해 제공
- 매달 수십억 분의 음성 데이터 처리 능력 제공



사용한 기술

2. GOOGLE TEXT-TO-SPEECH

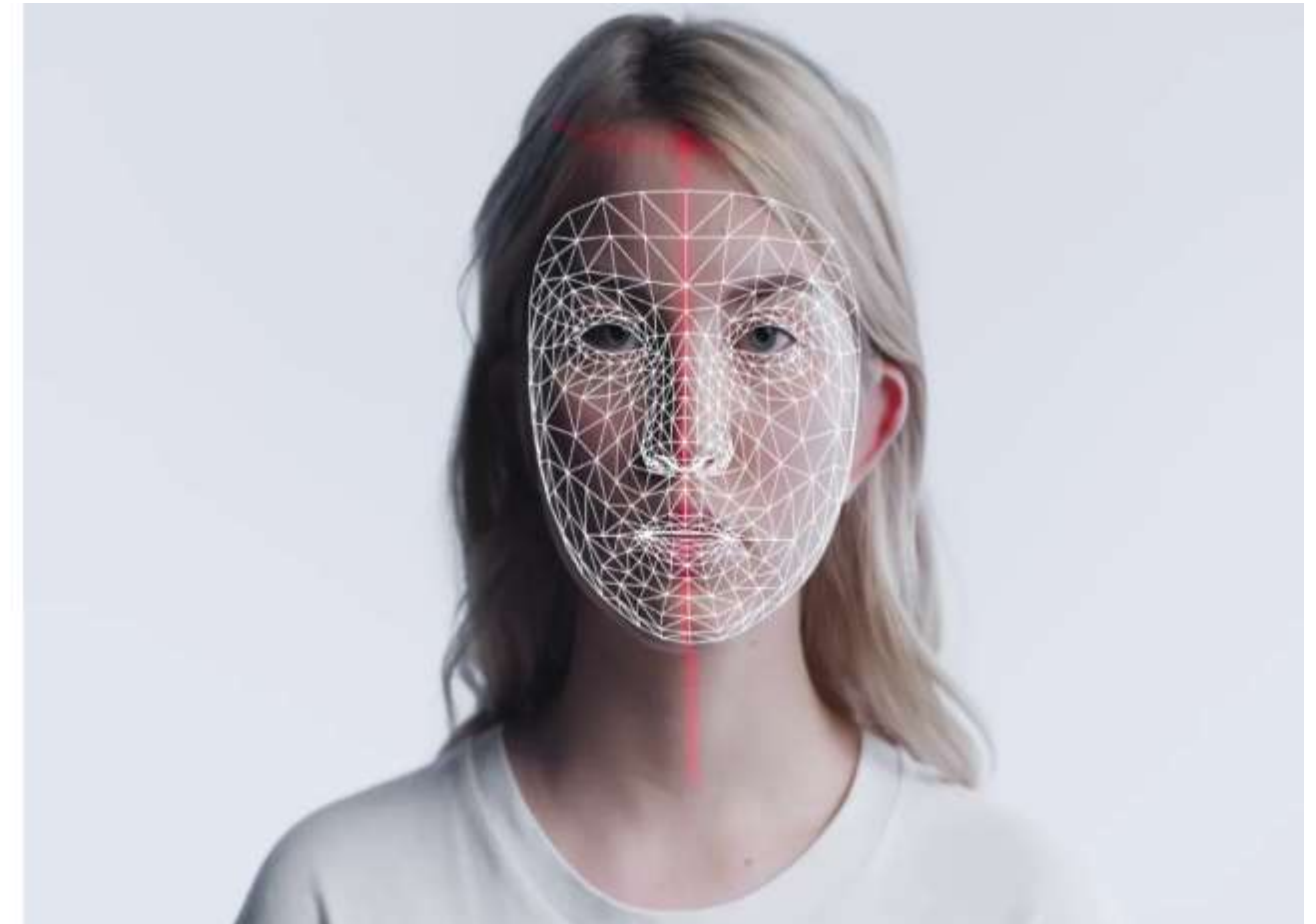
- 디지털 텍스트를 읽어주는 보조 기술 유형
- GOOGLE TTS API에서 글로 쓰여진 정보를 청각 정보로 변환하는 데 사용
- 자연스러운 발음과 다양한 언어 및 방언을 지원
- WAVENET 음성을 통해 사람의 음성 and 거의 구별할 수 없는 자연스러운 음성 합성 제공
- 오디오를 실시간으로 스트리밍하고 애플리케이션과 통합하여 사용



사용한 기술

3. MEDIAPIPE

- 머신러닝 모델을 이용하여 얼굴 인식, 다중 손 추적, 객체 인식 및 추적 등과 같은 최첨단 기능을 제공
- GPU 또는 CPU를 활용하여 복잡한 인식 파이프라인에 대한 가속 추론을 특별히 설계
 - 사용자가 키오스크 앞 2M 이내에 위치하면 얼굴을 인식해 음성 안내 여부를 선택하도록 함
 - 사용자가 키오스크에서 벗어나면 이를 감지하고 음성 안내 프로세스를 자동으로 종료함

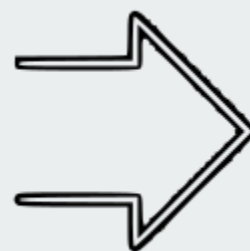


사용한 기술

4. 레벤슈타인 거리 알고리즘

$$\text{lev}_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1, j) + 1 \\ \text{lev}_{a,b}(i, j-1) + 1 \\ \text{lev}_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

- 두 문자열 A와 B 사이의 편집 거리
- A를 B로 변환하기 위해 필요한 최소 편집 연산(삽입, 삭제, 교체)의 수



- 역 이름 중 한글자만 다른 역의 경우 알고리즘이 이를 인식하지 못하는 문제 발생
→ 문제 해결 위해 문자열을 자모(초성, 중성, 종성)로 분리하여 거리 계산

사용한 기술

4. 레벤슈타인 거리 알고리즘 적용 후 정확도 비교

[테스트 방법]

1. 데이터베이스에 존재하는 총 288개의 역 중 무작위로 하나의 역 지정
2. 실험자가 마이크에 대고 해당 역 낭독
3. 구글의 STT로 음성을 텍스트로 변환
4. 분석된 결과를 레벤슈타인 거리 알고리즘을 통해 한 번 더 보정
5. 결과가 사용자가 말한 역과 일치하는지 확인
6. 일치하면 성공, 불일치하면 실패한 것으로 기록

	정확도
무보정	78
레벤슈타인	88
자모 분리+레벤슈타인	95

기대 효과 및 개선점

기대효과

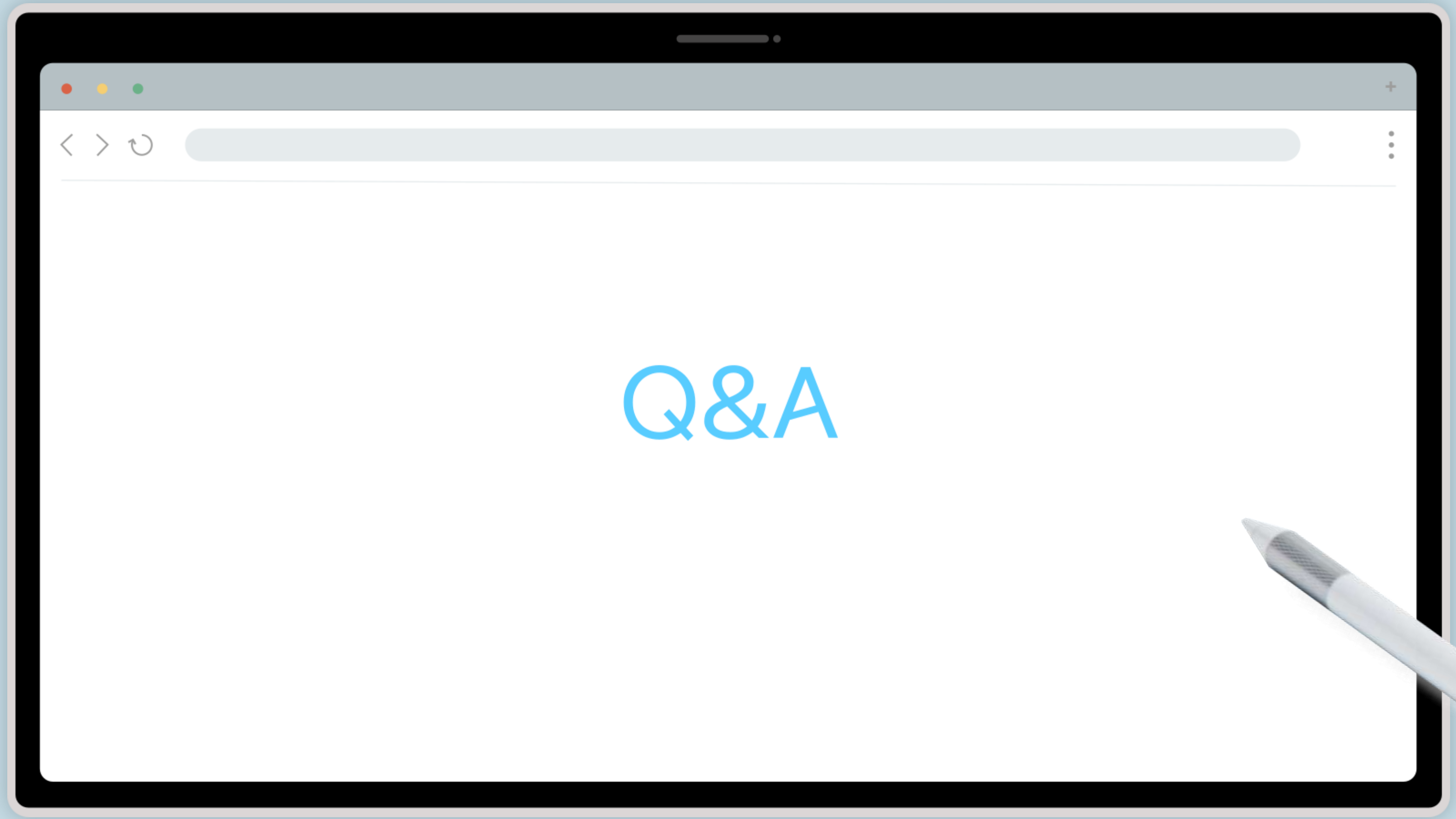
- 시각장애인 사용자에게 서비스 접근을 용이하게 함
- 자동화된 서비스 제공으로 인력 비용 절감
- 시각장애인 및 기타 장애를 가진 개인을 위한 접근성 향상으로 사회적 인식과 포용성 개선

개선점

- 키오스크가 배치될 환경의 소음을 고려하여 노이즈 필터링 기술을 개선한다면 현장 적용이 용이할 것
- 현재는 비대화형 인터페이스이지만 CHATGPT 기술을 도입하여 대화형 인터페이스로 발전시키면 사용성이 증대될 것

레퍼런스

- [HTTPS://RTZR.GITHUB.IO/AWESOME-KOREAN-SPEECH-RECOGNITION/](https://rtzr.github.io/awesome-korean-speech-recognition/)
- [HTTPS://CLOUD.GOOGLE.COM/TEXT-TO-SPEECH?HL=KO](https://cloud.google.com/text-to-speech?hl=ko)
- [HTTPS://CLOUD.GOOGLE.COM/SPEECH-TO-TEXT?HL=KO](https://cloud.google.com/speech-to-text?hl=ko)
- [HTTP://WWW.IMEDIALIFE.CO.KR/NEWS/ARTICLEVIEW.HTML?IDXNO=40346](http://www.imedialife.co.kr/news/articleview.html?idxno=40346)
- [HTTPS://WWW.YNA.CO.KR/VIEW/AKR20221118043700530](https://www.yna.co.kr/view/AKR20221118043700530)
- [HTTPS://WWW.BEMINOR.COM/NEWS/ARTICLEVIEW.HTML?IDXNO=23564](https://www.beminor.com/news/articleview.html?idxno=23564)
- [HTTPS://WWW.YNA.CO.KR/VIEW/AKR20211014121700004](https://www.yna.co.kr/view/AKR20211014121700004)
- [HTTPS://HUGGINGFACE.CO/SPACES/EVALUATE-METRIC/CER](https://huggingface.co/spaces/evaluate-metric/cer)
- [HTTPS://WWW.BEMINOR.COM/NEWS/ARTICLEVIEW.HTML?IDXNO=23506](https://www.beminor.com/news/articleview.html?idxno=23506)





감사합니다