

# 졸업 작품 개발보고서

## ‘비전 토크’

[시각 장애인을 위한 객체 인식 및 설명 앱]

작품명	비전 토크
과목명	융합소프트웨어종합설계1, 2
지도교수명	전병태 교수님
제출일	2024. 04. 08
팀원	김태민 (팀장)

## 목 차

### 1. 작품명

### 2. 작품 개요

### 3. 시스템 개발 필요성 및 기존 시스템과의 차별성

#### 3.1 개발의 필요성

#### 3.2 기존 시스템과의 차별성

##### 3.2.1 기존 시스템

##### 3.2.2 차별성

### 4. 시스템 구현 방법 및 일정

#### 4.1 구현 방법

#### 4.2 구현 예정

#### 4.3 구현 일정

#### 4.4 구현 분담

### 5. 시스템 실험

#### 5.1 시스템 기능

#### 5.2 시스템 기능 시험

### 6. 결론

#### 6.1 시스템의 효용성 및 기대효과

#### 6.2 개선 방향

### 7. 참고문헌

# 1. 졸업작품명

비전 토크 (VisionTalk)

## 2. 작품 개요

‘비전토크’ 앱은 시각 장애인과 읽기 장애를 겪는 사용자들의 일상생활 속 정보 접근성을 개선하기 위해 개발되었다. 이 앱은 사용자가 찍은 사진 속의 물체를 정확히 인식하고, 음성으로 그 물체에 대해 설명해주는 객체 인식 기능을 제공한다. 또한, 사진 속의 주요 색상을 파악하여 이를 음성으로 설명해주는 색상 인식 및 설명 기능을 통해 색상 정보를 사용자에게 제공한다. 사진에 포함된 텍스트를 음성으로 변환해주는 기능도 포함되어 있어, 시각적 정보를 오디오 형식으로 전달함으로써 시각 장애인은 물론 어린이와 읽기 장애가 있는 사람들에게 계도 큰 도움을 줄 수 있을것이다. 사용의 편리성을 높이기 위해, "비전토크"는 음성 안내와 사용자의 필요에 따라 조절 가능한 음성 응답의 속도 및 톤 조절 기능을 포함하고 있다.

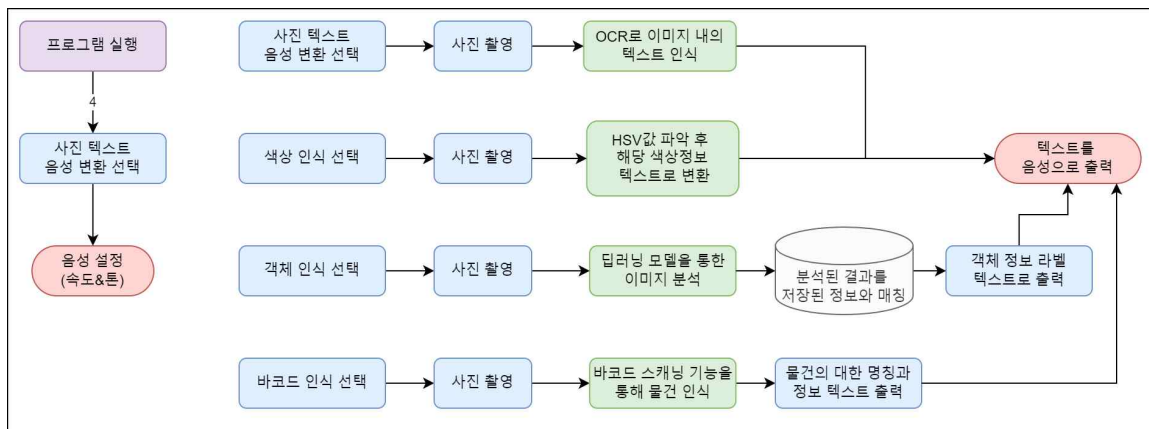


그림 1 비전 토크 플로우 차트 (후일 수정)

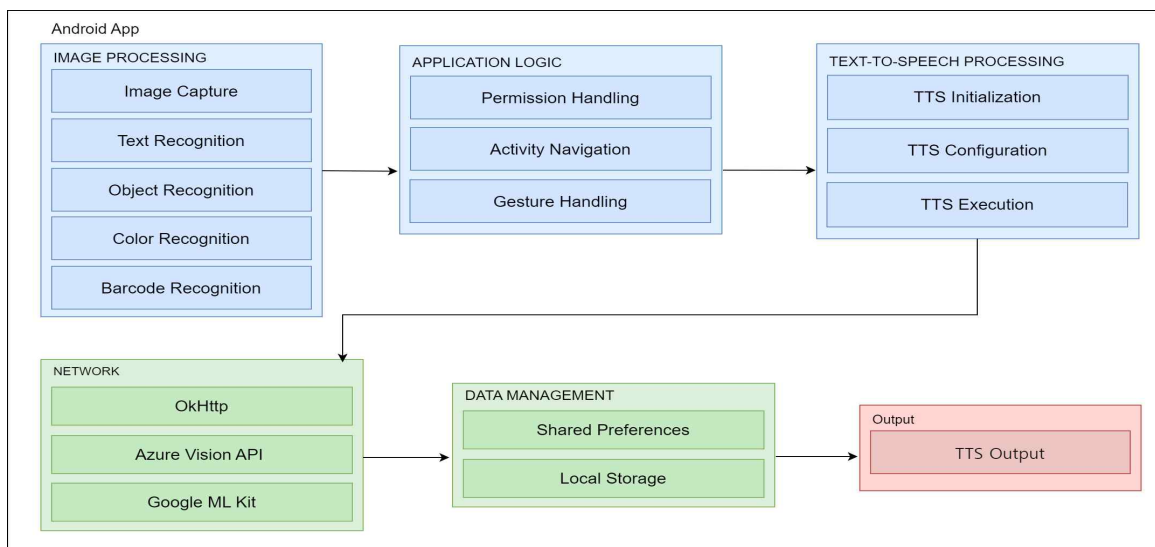


그림 2 비전 토크 시스템 구조 (후일수정)

### 3. 시스템 개발 필요성 및 기존 시스템과의 차별성

#### 3.1 개발의 필요성

"비전토크" 애플리케이션 개발의 필요성은 현대 사회에서 정보 접근성의 중요성이 점차 증가하면서 더욱 명확해졌다. 정보가 우리 일상 생활의 모든 측면에서 중요한 역할을 하며, 시각 장애인과 읽기 장애를 겪는 사람들에게 접근 가능한 정보는 자립과 독립성을 향상시키는 핵심 요소였다. "비전토크"는 이러한 사용자들이 시각적 정보를 오디오 형식으로 쉽게 변환할 수 있도록 하여, 일상 생활에서 겪는 정보 접근의 어려움을 해결하려 했다. 이 앱은 사진 속 객체와 색상을 인식하고 설명하며, 사진에 포함된 텍스트를 음성으로 변환하여, 사용자들이 주변 세계를 더 잘 이해하고 상호작용할 수 있도록 도울 것이다.

이러한 기술적 진보는 정보 접근성이 사용자의 자립성과 직접적으로 연결되어 있는 현대 사회에서 필수적인 것이다. "비전토크" 앱은 사용자들이 일상생활에서 마주하는 다양한 도전을 극복하도록 설계되었다. 예를 들어, 시각적으로 어려움을 겪는 사람들은 쇼핑을 할 때 제품의 색상을 알아야 하거나, 길을 찾을 때 방향을 알려주는 표지판의 정보가 필요했다. 이 앱은 그러한 정보를 음성으로 제공하여, 사용자가 더 독립적으로 활동할 수 있도록 지원했다.

더 나아가, "비전토크"의 개발은 팬데믹 이후 정보 접근성의 중요성이 새롭게 강조된 사회적 변화에 부응할 것이다. 팬데믹 기간 동안 우리는 디지털 기술의 중요성과 AR 환경에서의 접근성의 필요성을 새삼 깨달았다. "비전토크"는 이러한 사회적 요구에 발맞추어, 시각 장애인과 읽기 장애를 겪는 사람들이 디지털 시대의 정보에 더 쉽게 접근할 수 있도록 하며, 일상 생활에서의 독립성을 향상시키는 도구가 될 것이다.

이 프로젝트는 단순히 앱 개발을 넘어서, 시각 장애인과 읽기 장애가 있는 사용자들이 사회의 완전한 구성원으로서 활동할 수 있도록 하는 데 기여할 것이다. "비전토크"를 통해, 이러한 사용자들은 자신의 능력을 최대한 발휘하고, 일상 생활의 다양한 상황에서 보다 효과적으로 정보를 접근하고 활용할 수 있게 될 것이다. 이는 디지털 접근성을 향상시키는 것뿐만 아니라, 우리 사회 전반에 걸쳐 포용성과 동등한 기회를 증진시키는 데 기여하는 중요한 단계일 것이다.

## 3.2 기존 시스템과의 차별성

### 3.2.1 기존 시스템

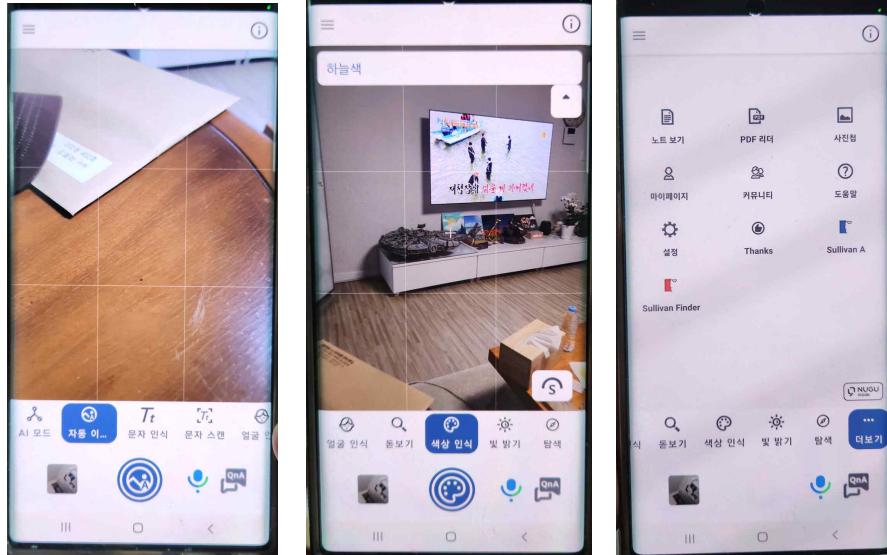


그림 3 설리번 플러스 앱 사용 화면

그림 3은 현재 시중에서 유통되고 있는 시각 장애인을 위한 앱인 설리번 플러스의 사용 화면이다. 이 앱은 시각 장애인을 위한 다양한 기능을 포함하고 있다. 기능적인 측면에서는 좋은 성능을 가지고 있지만, 사용 편의성 측면에서 전맹 시각자를 충분히 배려하지 않은 점이 아쉽고, 표정 묘사 등 여러 가지 불필요한 기능이 포함되어 있다. 전반적으로 사용법과 기능들이 전맹 시각자보다는 저시각자를 위한 앱에 가깝다.

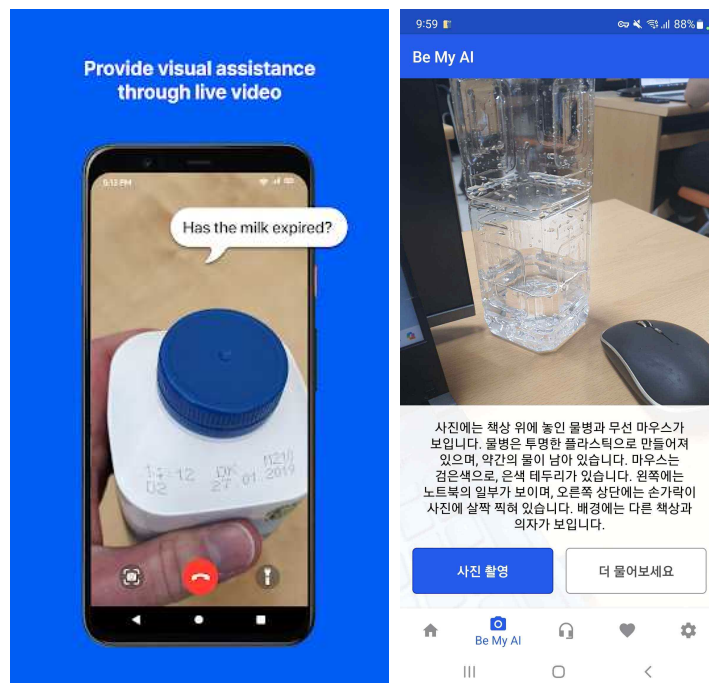


그림 4 be my eyes 사용 화면

그림 4는 그림 3과 마찬가지로 시각 장애인을 위한 앱인 Be My Eyes 앱이다. 이 앱은 시각 장애인과 봉사자를 연결하여, 시각 장애인이 카메라를 통해 보여주는 물체나 상황을 봉사자가 전화 연결을 통해 설명해주는 방식으로 문제를 해결한다. 기능 특성상 매우 범용적으로 사용할 수 있으나, 인터넷 연결 의존성, 특정 시간대나 지역에서의 봉사자 의존성, 프라이버시 문제 등의 단점을 가지고 있다. 또한, 해당 앱에는 AI 인식 기능도 포함되어 있으며, 매우 디테일한 기능을 제공하지만 인식에 시간이 오래 걸리고 불필요한 내용이 많이 나오는 문제가 있다.

### 3.2.2 차별성

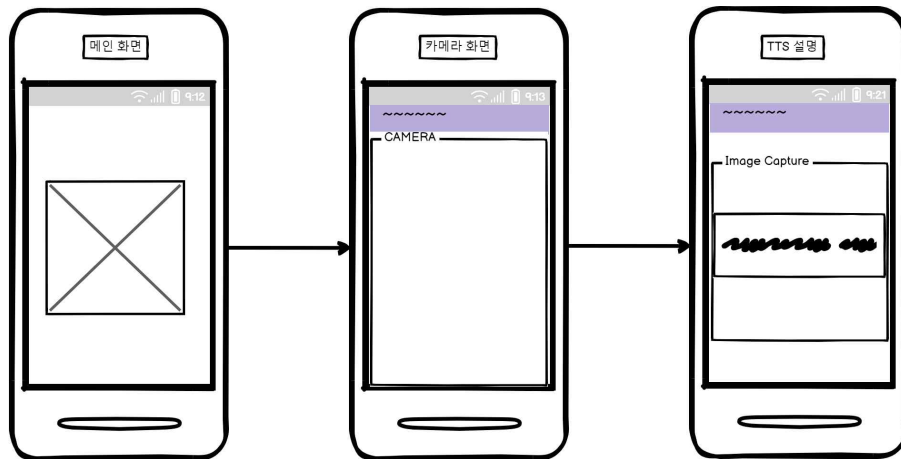


그림 5 본 프로젝트 개발 앱의 사용 예상도 1 (와이어 프레임)

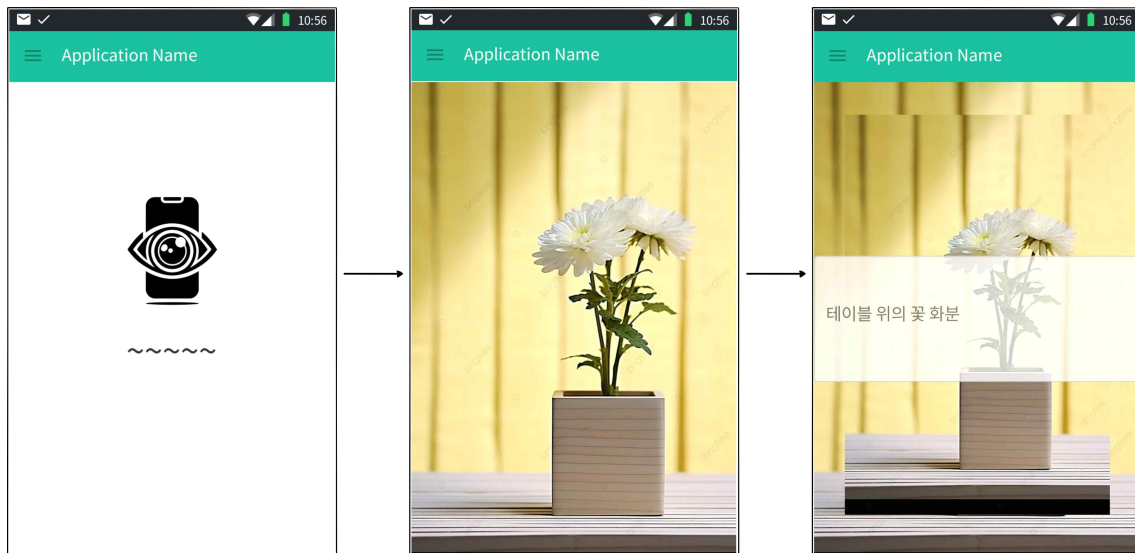


그림 6 본 프로젝트 개발 앱의 사용 예상도 2 (프로토타입)

본 앱의 주요 장점은 전맹 사용자를 비롯하여 모든 시각 장애인들이 쉽게 사용 가능하게 설계되었다는 점이다. 최대한 간결하게 제작된 사용법과 조작법을 통해 누구나 쉽게 앱을 이용할 수 있으며, 모든 메뉴 선택 과정과 기능 사용 방법을 TTS(Text-to-Speech)로 안내

내해 사용자의 편의를 높였다. 전면적인 TTS 지원으로 앱의 모든 내용을 음성으로 설명해 주어 시각 장애인 사용자들에게 큰 도움이 될 것이다. 또한, 이 앱은 사용자의 실제 필요에 맞춘 기능만을 포함하여 불필요한 요소를 최소화하였다.

## 4. 시스템 구현 방법 및 일정

### 4.1 구현 방법 (후일 수정)

VisionTalk 앱은 안드로이드 스튜디오를 이용하여 코틀린 언어로 개발되었다. 주요 구현 방식은 다음과 같다.

스플래시 액티비티는 앱 실행 시 첫 화면으로, TTS(Text-to-Speech) 기능을 초기화하고 필요한 권한을 확인하는 역할을 수행한다. 권한이 부여되지 않은 경우 사용자에게 권한을 요청하고, 모든 권한이 부여되면 메인 액티비티로 이동하도록 구현되었다.

메인 액티비티는 다양한 기능을 탐색하고 선택할 수 있도록 구현되었다. 사용자는 제스처를 통해 각 기능을 탐색하고, 두 번 터치하여 원하는 기능으로 이동할 수 있도록 하였다.

음성 설정 액티비티에서는 사용자가 TTS 기능의 음성 속도와 톤을 제스처를 통해 조절할 수 있도록 구현하였다. 조절된 값은 SharedPreferences에 저장되며, 사용자가 화면을 좌우로 슬라이드하여 속도를, 상하로 슬라이드하여 톤을 조절할 수 있게 하였다. 설정이 완료되면 화면을 두 번 터치하여 메인 화면으로 돌아갈 수 있도록 구현하였다.

텍스트 음성 변환 액티비티에서는 Google ML Kit의 한국어 텍스트 인식 기능을 사용하여 촬영한 이미지에서 텍스트를 추출하고, 이를 TTS를 통해 음성으로 안내하는 기능을 제공한다. 사용자는 화면을 두 번 터치하여 다시 촬영하거나, 상하 슬라이드 제스처로 메인 화면으로 돌아갈 수 있도록 하였다.

객체 인식 기능은 Azure의 비전 API를 사용하여 촬영한 이미지를 분석하도록 구현하였다. 인식된 객체 정보를 TTS 기능을 통해 음성으로 안내하며, 사용자가 화면을 두 번 터치하여 다시 촬영하거나, 상하 슬라이드 제스처로 메인 화면으로 돌아갈 수 있도록 하였다.

색상 인식 기능은 이미지의 중앙 픽셀의 HSV 값을 분석하여 색상을 인식하도록 구현하였다. 분석된 색상 정보를 텍스트로 변환한 후 TTS 기능을 통해 음성으로 출력하며, 사용자가 화면을 두 번 터치하여 다시 촬영하거나, 상하 슬라이드 제스처로 메인 화면으로 돌아갈 수 있도록 하였다.

지폐 인식 기능은 Azure의 비전 API를 사용하여 촬영한 지폐 이미지를 분석하도록 구현하였다. 금액과 화폐 종류를 인식하고 이를 TTS 기능을 통해 음성으로 출력하며, 사용자가

화면을 두 번 터치하여 다시 촬영하거나, 상하 슬라이드 제스처로 메인 화면으로 돌아갈 수 있도록 하였다.

바코드 인식 기능은 Google ML Kit의 바코드 스캐닝 API를 사용하여 촬영한 이미지에서 바코드를 인식하도록 구현하였다. 인식된 바코드 정보를 TTS 기능을 통해 음성으로 안내하며, 사용자가 화면을 두 번 터치하여 다시 촬영하거나, 상하 슬라이드 제스처로 메인 화면으로 돌아갈 수 있도록 하였다.

이러한 구현 방식을 통해 VisionTalk 앱은 시각 장애인들이 시각 정보를 쉽게 인식하고 활용할 수 있는 유용한 도구가 되도록 개발되었다.

## 4.2 구현 예정

개발 내용	추진일정(월)								가시적 결과물
	3	4	5	6	7	8	9	10	
아이디어 회의									주제 선정 및 보고서 작성
개발 환경									개발 환경 선정 및 환경 구축
처리 및 학습 모델 개발									이미지 처리 및 모델 훈련
계획 회의									개선사항 추가 및 수정
앱 인터페이스 개발									인터페이스 개발 및 기능구현
통합 및 시스템 테스트									평가 및 개선사항 토의
부족한 점 보완									보완 및 마무리작업

기간		일정	내용
3월	3.15 ~ 3.31	아이디어 회의	프로젝트 개발 주제 선정, 보고서 작성
4월	4.01 ~ 4.24	개발 환경	개발 환경 선정 및 환경 구축
5월	5.01 ~ 5.31	이미지 처리, 기계 학습 모델 개발	이미지 처리 파이프라인 구축 및 CNN 모델 훈련
6월 ~8월	6.01 ~ 6.07 7.01 ~ 7.07	계획 회의	개선사항 추가 및 수정, 추가 기능 개발에 대한 세부 계획 수립
	5.01 ~ 8.31	앱 인터페이스 개발 및 기능 구현	인터페이스 개발과 동시에 기능 구현
9월	9.01 ~ 9.30	통합 및 시스템 테스트	모든 기능과 인터페이스 통합, 종합적인 시스템 테스트 및 버그 수정
10월	10.01 ~ 10.31	부족한 점 보완	부족한 부분 보완 및 마무리 작업



### 4.3 구현 일정

분류	내용	2024. 03				2024. 04				2024. 05				2024. 06			
		1주	2주	3주	4주	1주	2주	3주	4주	1주	2주	3주	4주	1주	2주	3주	4주
아이디어 회의	프로젝트 주제 선정																
	보고서 작성																
개발 환경 선정 및 구축	필요 소프트웨어 및 라이브러리 설치																
	개발 도구 설정 및 초기 프로젝트 설정																
이미지 처리 및 보고서 작성	이미지 전처리 작업																
	주제 재탐색 및 보고서 개선																
모델 훈련 및 평가 준비	docker로 TensorFlow 준비																
	Flask 간이 서버 제작 및 테스트																
	COCO 모델																
	YOLOv5 준비																
메인 화면 임시 제작	앱 아이콘 및 스플래시 화면 제작																
TTS 속도와 톤 조절 기능	제작 기간																
	제작 완료																
OCR TTS 기능	제작 기간																
	제작 완료																
색상 인식 기능	제작 기간 & 완료																
보고서 작성	보고서 작성																
객체 인식 기능	제작 기간																
	제작 완료																

### 4.3 구현 분담

2010280010\_김태민

- 기획, 보고서작성, 개발환경 및 DB 구축, 이미지 처리 및 기계 모델 학습 개발, 앱 인터페이스 개발, 통합 및 시스템 테스트

## 5. 시스템 실험

### 5.1 시스템 기능

- 5.1.1 객체 인식 및 음성 설명
- 5.1.2 색상 인식 및 음성 설명
- 5.1.3 이미지 내 텍스트 음성 변환
- 5.1.4 바코드 인식 및 음성 설명
- 5.1.5 음성 응답의 속도나 톤 조절

### 5.2 시스템 기능 시험(시연) (후일 작성)



그림 7 객체  
인식 아이콘



그림 8 객체 인식  
기능 시연

## 6. 결론 (후일 작성)

- 6.1 시스템의 효용성 및 기대효과
- 6.2 개선 방향

## 7. 참고문헌 (후일 작성)

모델 평가와 성능 지표 확인 진행

