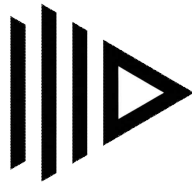


1. Conceptualization

Open Media Player



Open Media Player

Student No.	22313530
Name	BAE WON IL
E-Mail	baewonil@yu.ac.kr

[Revision history]

Revision date	Version #	Description	Author
03/14/2025	1.00	First Draft	BAE WON IL
03/23/2025	1.01	Logo 추가	BAE WON IL
30/04/2025	1.02	Live-Support 추가	BAE WON IL

= Contents =

1. [Business purpose](#)
2. [System context diagram](#)
3. [Use case list](#)
4. [Concept of operation](#)
5. [Problem statement](#)
6. [Glossary](#)
7. [References](#)

1. Business purpose

1.1. Project Background

현대의 멀티미디어는 다양한 형태로 제공되고 있고

이러한 멀티미디어를 사용함에 있어

디지털 취약계층과, 청각장애우와 같이 접근성에 제약이 있는 사용자에게
자막생성 및 LM을 사용한 Helper Chat-bot과 동작 자동화를 통해
접근성을 개선하고, 제한된 사용경험을 보완하는것이 목표이다.

1.2. Target Market

일반 가정 사용자 : 기본적인 미디어 재생 및 자동자막

청각 장애 사용자 : 자막 생성을 통한 접근성 강화

1.3. Goals

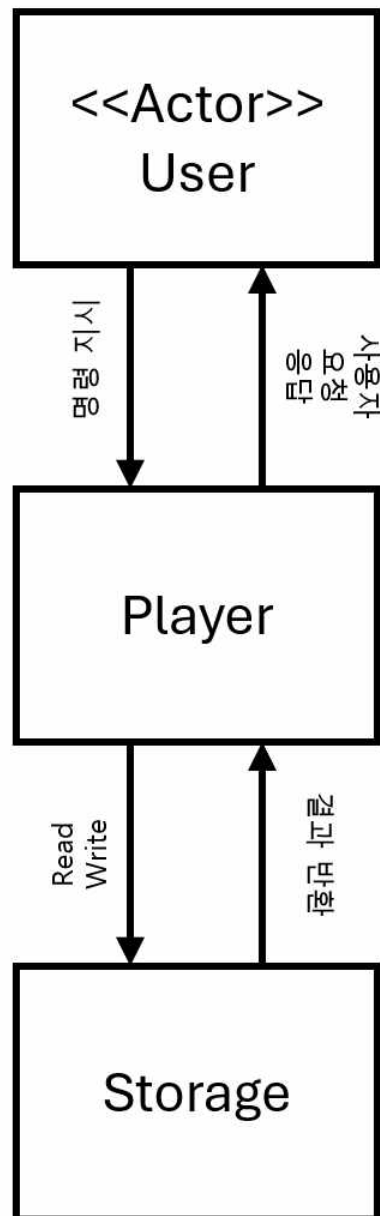
기본적인 미디어 재생 기능

영상 미디어의 음성 데이터로부터 STT를 진행하여 자막 생성

2. System context diagram

- 미디어 파일 재생/일시정지
- 볼륨 / 화면 조정
- 자막 생성 설정

- 오디오/비디오 디코딩
- 재생목록 관리
- 자막 생성



3. Use case list

1) Open Media File

Actor	User
Description	사용자가 저장소에서 미디어 파일을 선택하여 재생

2) Playback Control (Play/Pause/Stop)

Actor	User
Description	사용자가 Audio / Video 재생 상태를 제어

3) Seek

Actor	User
Description	재생 도중 특정 구간으로 이동

4) Playlist Management

Actor	User
Description	등록된 재생목록을 확인, 수정하여 원하는 순서대로 미디어 재생

5) Volume Control

Actor	User
Description	볼륨 조절, 음소거 등으로 오디오 출력 설정

6) Subtitle Load / Config

Actor	User
Description	자막파일(smi/srt...)를 불러오거나 싱크를 설정해 표시

7) Preferences

Actor	User
Description	Player 세부 설정을 변경하고 저장

8) Generate Subtitle

Actor	User
Description	재생 시 자막을 생성할지 결정

9) Natural Language Helper

Actor	User
Description	자연어로 도움을 요청하면 그에 대응하는 응답 및 작업 수행

4. Concept of operation

1) Open Media File

Purpose	사용자가 미디어 파일 재생을 위해 파일을 선택해 불러온다
Approach	사용자가 파일 열기 메뉴 또는 버튼을 클릭한다. 파일 선택 창에서 원하는 미디어 파일을 선택한다 해당 파일 형식을 판단 후 코덱을 선택하여 준비를 완료한다.
Dynamics	사용자가 플레이어를 실행해 파일을 선택하고 열게되면 파일을 디코딩하고, 재생 가능한 상태로 Load하여 UI에 재생준비가 완료되었다고 표시한다.
Goals	파일선택 및 미디어 디코딩 기능을 구현한다.

2) Playback Control (Play / Pause / Stop)

Purpose	사용자가 미디어를 원하는 시점에 재생·중지 가능하게 한다.
Approach	사용자가 재생 버튼을 누르면, 디코딩된 프레임을 불러와 화면과 스피커로 출력한다. 일시정지 버튼을 누르면 현재 Frame, Timestamp를 기록한다. 정지 버튼을 누르면 Stream-Buffer를 해제한다.
Dynamics	사용자가 UI버튼 또는 단축키를 누르면 즉시 반응하여 현재 재생상태 변경과 그 결과를 UI에 통지하고, 화면과 오디오도 변경된 상태에 따라 갱신한다.
Goals	미디어 재생 기능을 구현한다.

3) Load Subtitle and Config

Purpose	자막파일의 내용을 연동하여 현재 미디어에 표시한다.
Approach	<p>사용자가 자막을 키거나 외부파일을 불러 올 수 있는 메뉴를 선택한다.</p> <p>플레이어는 현재 재생중인 영상 Timestamp에 맞춰 자막을 표시한다.</p> <p>사용자는 자막의 크기·위치·언어·싱크보정 등을 설정 메뉴에서 조정한다.</p>
Dynamics	<p>영상이 재생되는 동안 현재 Frame의 Timestamp에 맞게 자막 텍스트를 동기화 한다.</p> <p>일시정지 또는 재생 시에도 자막이 갱신된다.</p>
Goals	싱크조절기능 및 자막 표시 기능을 구현한다.

4) Generate Subtitle

Purpose	현재 재생중인 미디어로부터 자막을 생성한다.
Approach	<p>사용자가 자막생성 버튼을 선택한다.</p> <p>플레이어는 현재 Load된 미디어의 음성 Data로부터 STT를 수행하여 자막을 생성하고 화면에 표시한다.</p>
Dynamics	<p>Load된 영상에서 음성 Data를 분리하고 분리된 음성 Data에서 STT를 진행한 다음 Transcribe된 Text를 재생중인 영상의 Frame Timestamp에 맞게 동기화하여 표시한다.</p>
Goals	<p>재생중인 영상에서 음성을 추출하는 기능을 구현한다.</p> <p>추출된 음성에서 STT를 수행하는 기능을 구현한다.</p> <p>Transcribe된 Text를 자막표시 하는 기능을 구현한다.</p>

5) Natural Language Helper

Purpose	자연어 프롬프트를 입력받아 응답 및 작업을 수행한다.
Approach	사용자가 Help 메뉴를 선택하여 프롬프트 입력을 시작한다. 프롬프트 입력창에 특정한 작업을 요청한다. 해당 작업을 요청하면 local SLM이 요청을 처리한다.
Dynamics	사용자가 특정한 질문/작업 사항을 입력하면 local SLM이 해당 프롬프트를 분석하고 결과를 도출한다. 도출된 결과에 따라 특정한 작업을 수행하거나, 질문에 대한 응답을 제공한다.
Goals	local SLM의 동작을 위한 Fine-Tuning이나 system-instruction(message)를 설정한다. 도출 결과가 behavior인 경우 작업 수행을 위한 매크로를 구현한다.

5. Problem statement

1) Overview

'본 프로젝트'는 Windows를 기반으로 다양한 멀티미디어 콘텐츠를 하나의 애플리케이션에서 재생하고 부가기능을 통합 제공 함으로써, 사용자 편의성과 접근성을 강화한다.

Problem 1 – Compatibility of various file and codec

다양한 파일형식과 코덱의 지원에 있어 복잡성이 증가하여 구현상 어려움이 있을 수 있다.

plug-in 방식으로 구현하여 가능한 많은 파일 형식을 지원할 수 있도록 시도한다.

Problem 2 – Security and License Issue

외부 라이브러리 / 프로젝트 사용 시 오픈소스 라이선스를 정확하게 준수해야한다.

라이선스 세부사항을 확인하고 준수한다.

Problem 3 – STT Function

STT기능 개발 시 인식정확도나 처리속도가 느린 등의 문제가 생길 수 있다.

인식 정확도를 높이기 위해서 추가적인 데이터셋의 확보나 Fine-Tuning이 필요할 수 있으며

처리 속도의 경우 Performance and Accuracy 와 Processing Time간 Trade-off 관계에 있으므로

사용자에게 선택권을 부여하거나, 기준을 정하여 적절하게 타협한다.

Problem 4 – LM Automation

SLM을 이용한 자동화 기능의 경우,

사용자 프롬프트와 답변 생성에 규제가 필요하며 system instruction이나 fine-tuning 또는 규제 계층을 별도로 추가하여 부적절한 응답을 막는다.

NFR

1) 성능

재생지연 : 앱을 실행하고 미디어 재생 시 짧은 대기 후 재생이 시작되도록 하고,
STT를 진행하거나, 용량이 큰 미디어를 불러오는 경우 프로그램이
응답없음 상태에 진입하게 되는 경우 waiting을 구현하여
응답없음 상황에서의 안내를 할 수 있도록 한다.

[로딩중입니다...] [처리중입니다...]

2) 안정성 : 재생 도중 갑자기 종료되거나, 시스템 자원을 과도하게 점유하지 않도록
에러 핸들링과 자원 해제에 유의한다.

3) 사용성 : 기본기능을 찾기 쉽게 배치하고, 고급옵션은 별도로 분리하여
최대한 혼동을 줄인다.

6. Glossary

Codec	멀티미디어 데이터를 압축(Encode)하고 풀어내는 (Decode) SW나 HW 알고리즘 (H264, AAC ...)
Subtitle	영상에 포함되거나 외부 파일로 제공되는 텍스트 정보
Playlist	여러 미디어 파일을 원하는 순서대로 모아놓은 목록
STT	Speech-To-Text의 약어로 음성을 텍스트로 변환(Transcribe)하는 기술
SLM	Small Language Model의 약어로, 거대한 크기의 Large Language Model과 달리 비교적 소규모의 parameter를 사용해 edge-device에서도 실행이 용이하도록 축소된 언어 모델

7. References

Open-AI Whisper(paper) : <https://arxiv.org/abs/2212.04356>

Open-AI Whisper(git) : <https://github.com/openai/whisper>

FFMPEG : <https://www.ffmpeg.org/documentation.html>

MS DirectX API: <https://learn.microsoft.com/en-us/windows/win32/directx>

MS Win32 API :

<https://learn.microsoft.com/en-us/windows/win32/apiindex/api-index-portal>

MS .NET API : <https://learn.microsoft.com/en-us/dotnet/api/>

WPF .NET:

<https://learn.microsoft.com/en-us/dotnet/desktop/wpf/overview/?view=netdesktop-9.0>

참고문헌

LLM-Powered GUI Agents in Phone Automation: Surveying Progress and Prospects

<https://arxiv.org/pdf/2504.19838>

B-MOCA:BENCHMARKING MOBILE DEVICE CONTROL AGENTS

ACROSS DIVERSE CONFIGURATIONS

<https://arxiv.org/pdf/2404.16660>

LLM-Powered GUI Agents in Phone Automation: Surveying Progress and Prospects

B-MOCA:BENCHMARKING MOBILE DEVICE CONTROL AGENTS ACROSS

DIVERSE CONFIGURATIONS

위 문헌들을 참고하여 추가된 부분이다.

LLM-Powered GUI Agents in Phone Automation에서

Language Model을 사용한 벤치마크 및 사용자 동작 자동화를 진행할 수 있음에
영감을 받았으며

국내 KAIST에서 진행했던 B-MOCA에서 Mobile-Benchmark를 진행함에 있어

Language Model을 사용해 기존의 정적인 매크로 보다 일반 사용자에게 있어
기반지식 없이도 자연어 요청을 통해 매크로 수행(자동화 동작)이 가능하다는 점에서
접근성을 개선 시킬 수 있을것으로 기대된다.