

오프라인 한국어 음성인식 시스템

목차

1. <서론> 주제선정 및 가치판단

- 선정한 주제
- 가치
- 범위

2. 주제 구현 방법 및 보조자료

- 구현과정 이미지 및 Flow Chart
- 소프트웨어 구현방법

3. 개발 계획 및 기자재

- 제품 구현 시 필요한 하드웨어 부품
- 팀원간 업무 내용
- 작품 제작 추진 계획 및 일정표
- 지원 경비 사용 계획

1. <서론> 주제선정 및 가치판단

오프라인 한국어 음성인식 시스템

1-1 주제 개요

KALDI 라이브러리를 사용하여 오프라인 환경에서도 작동하는 한국어 음성 인식 시스템을 개발한다. 음성 인식 오류율을 측정하고, 가장 성능이 우수한 학습 알고리즘을 결정하는 것이 중요하다.

1-2 주제를 선택한 이유 및 근거

-시중의 성능이 좋은 온라인 음성인식 플랫폼들은 유료이기 때문에 비용 부담이 있다.

-오프라인 라이브러리 중 Pocket Sphinx 나 CMU Sphinx 같은 오픈 소스 STT 는 머신러닝을 활용하는 KALDI 와 같은 대규모 음성 데이터에 비해 정확도가 낮고 어휘 크기가 제한적이다.

-기존의 논문은 KALDI 를 사용하지만 사물인터넷 제품을 제어하는 간단한 명령어만 사용하고 있다.

1-3 문제 해결의 가치 및 범위

가치

-기존의 유료로만 사용할 수 있었던 음성인식 플랫폼을 대체하여 개발함으로써 무료로 사용이 가능하며, 오프라인 환경에서도 작동이 가능한 시스템이기 때문에 별도로 네트워크에 연결하지 않아도 사용이 가능하다. 오프라인 환경에서 무료로 동작 가능한 음성인식 서비스가 있다면 여러 분야에 활용도가 높을 것이다.

음성인식 오류율에 대한 기사(2023)

<https://www.cxtoday.com/speech-analytics/how-accurate-is-speech-to-text-in-2023-assemblyai/>

범위

-오프라인 엔진을 통한 음성인식을 개발한다. 기존 온라인 음성인식 제품들과의 오류율을 비교하여 오프라인에서 구동이 가능하면서도 오류율이 낮은 음성인식 시스템을 개발한다. STT 를 별도의 디스플레이에 출력하여 확인할 수 있도록 하고 Alhub 에서 한국어 음성 데이터셋을 이용해 KALDI 가 지원하는 3 가지 학습 알고리즘의 음성인식 오류율을 테스트하여, 가장 성능이 좋은 학습 알고리즘을 추정한다.

{GMM(Gaussian Mixture Model), SGMM(Subspace Gaussian Mixture Model),
nnnet(Deep Neural Network)}

참고 논문 <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07614090>

KALDI 를 이용한 라즈베리파이 기반 오프라인 음성 인식

음성 데이터셋

1. <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=130>
2. <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=123>

- 오프라인 논문 기재일이 2018 년 12 월이고 라즈베리파이 4 출시일이 2019 년 6 월
이므로 논문에서 사용된 라즈베리파이는 3 으로 추정된다.

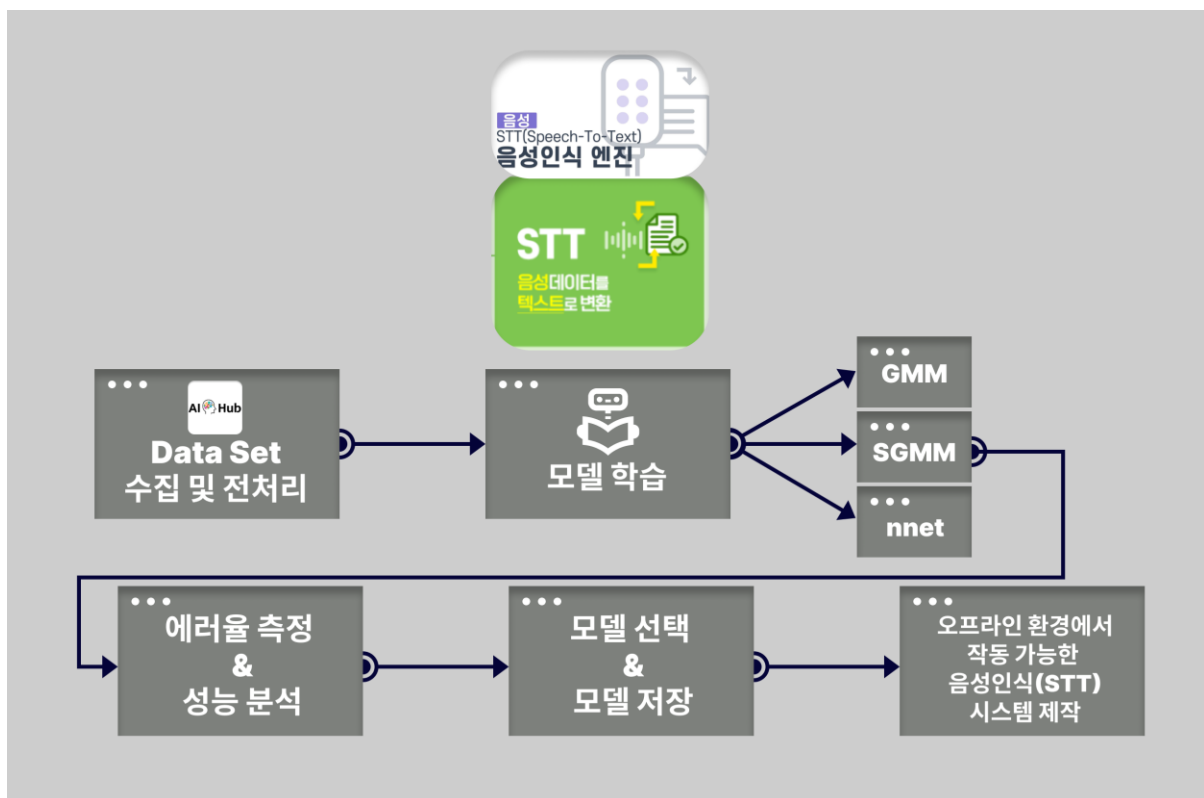
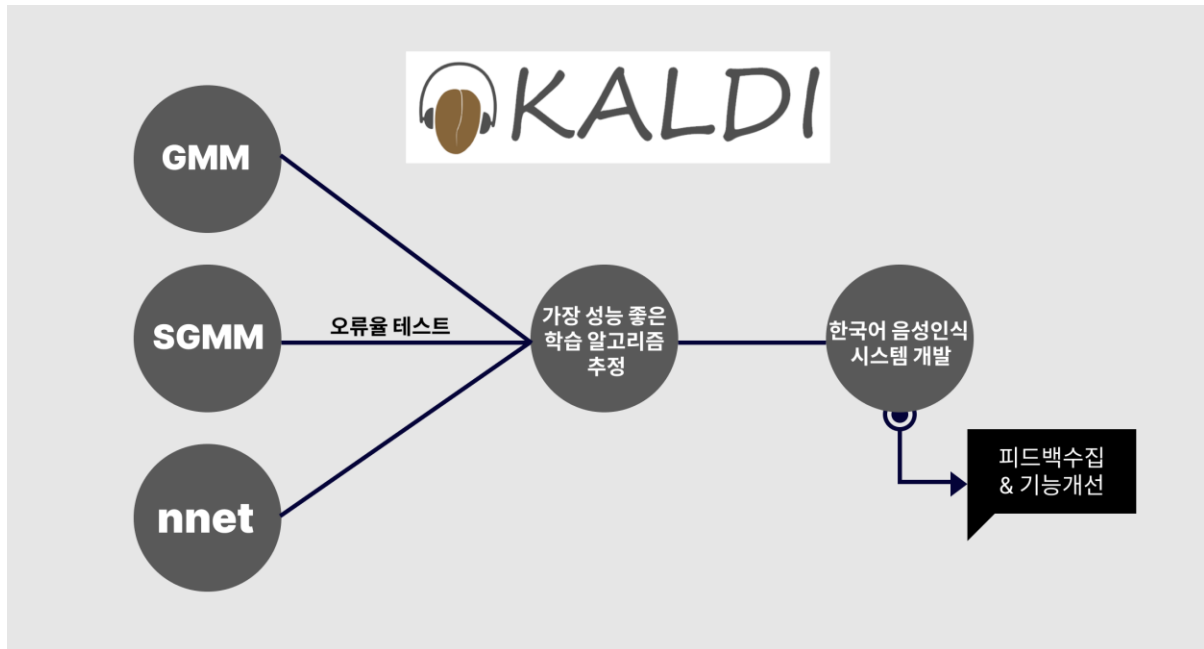
	라즈베리파이 3B+	라즈베리파이 4B	라즈베리파이 5
탑재 SoC	Broadcom BCM2837B0, Cortex- A53 (ARMv8) 64-bit SoC @ 1.4GHz	Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.8GHz	Broadcom BCM2712 2.4GHz quad-core 64- bit Arm Cortex-A76 CPU
캐쉬			512KB per-core L2 caches and a 2MB shared L3 cache
RAM	1GB LPDDR2 SDRAM	1GB, 2GB, 4GB or 8GB LPDDR4-3200 SDRAM (depending on model)	LPDDR4X-4267 SDRAM
GPU			VideoCore VII GPU, supporting OpenGL ES 3.1, Vulkan 1.2
이더넷	기가비트 이더넷	기가비트 이더넷	기가비트 이더넷
무선연결	2.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN, Bluetooth 4.2, BLE	2.4 GHz and 5.0 GHz IEEE 802.11ac wireless, Bluetooth 5.0, BLE	Dual-band 802.11ac Wi-Fi® Bluetooth 5.0 / Bluetooth Low Energy (BLE)
저장장치	Micro SD port for loading your operating system and storing data	Micro-SD card slot for loading operating system and data storage	microSD card slot, with support for high- speed SDR104 mode

비디오 출력	Full-size HDMI®	2 × micro-HDMI® ports (up to 4kp60 supported)	Dual 4Kp60 HDMI® display output with HDR support
	CSI camera port for connecting a Raspberry Pi camera	2-lane MIPI DSI display port	2 × 4-lane MIPI camera/display transceivers
	DSI display port for connecting a Raspberry Pi touchscreen display	2-lane MIPI CSI camera port	
오디오 출력	4-pole stereo output HDMI I2S	4-pole stereo audio HDMI I2S	HDMI I2S
GPIO	Raspberry Pi standard 40-pin header	Raspberry Pi standard 40-pin header	Raspberry Pi standard 40-pin header
USB	4 USB 2.0 ports	2 USB 3.0 ports 2 USB 2.0 ports	2 × USB 3.0 ports, supporting simultaneous 5Gbps operation
			2 × USB 2.0 ports
전원사용	5V 2.5A micro USB	5V 5A USB-C 타입	5V 5A USB-C 타입
기타			PCIe 2.0 x1 interface for fast peripherals (requires separate M.2 HAT or other adapter)

최신 버전 모델일수록 CPU 성능 및 클럭 수가 증가했으며, 특히 라즈베리파이 5의 경우에는 캐쉬와 GPU 등이 추가가 되었다. 메모리 용량도 지속적으로 증가했으며 PCIe 인터페이스가 추가되어 고용량 및 고속의 데이터를 처리하기 적합해 논문에 제시된 실험 결과보다 라즈베리파이 5에서 구현할 때 더 원활하고 정확한 환경에서 실험을 진행할 수 있다. 에러율 등 실험결과에 대한 부분도 논문에서는 100 명의 목소리를 대상으로 진행했는데 이번 음성 인식 시스템을 제작할 때는 2000 여개 이상의 음성 데이터셋을 활용하여 제작할 계획이기 때문에 에러율을 비롯한 실험결과들에 대해 더 정확한 실험결과를 도출할 수 있다.

2. 주제 구현방법 및 보조자료

<구현 과정에 대한 이미지 및 Flow Chart>



-소프트웨어 구현방법



프로젝트 목표:



KALDI 라이브러리를 사용하여 한국어 음성 인식 시스템을 오프라인으로 개발하는 것이 목표입니다.

음성 인식 에러율을 측정하고, 가장 성능이 우수한 학습 알고리즘을 결정하는 것이 중요합니다.

라이브러리 선택:

KALDI 를 선택하여 한국어 음성 인식 시스템을 개발합니다.



개발 환경 설정:

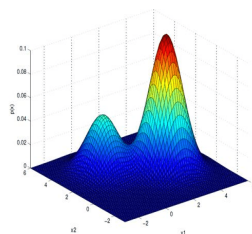
라즈베리 파이 5 에 KALDI 를 설치하고 설정합니다.
음성 데이터셋을 준비하고, KALDI 가 사용할 수 있도록 데이터를
전처리합니다.

데이터 수집 및 전처리:

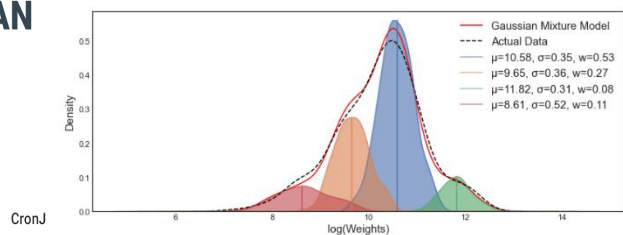
AIHub 의 한국어 음성 데이터셋을 사용하여 필요한 데이터를
수집하고 전처리합니다.

모델 학습:

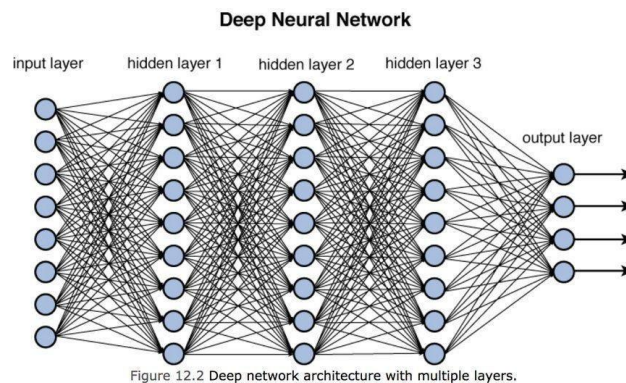
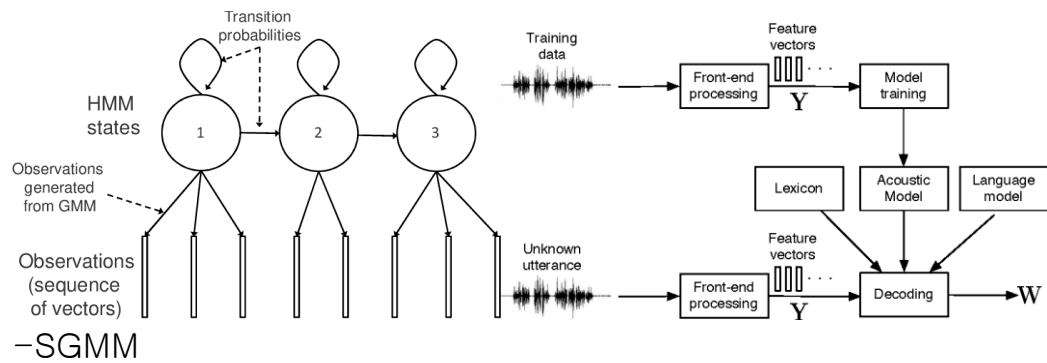
KALDI 를 사용하여 음성 인식 모델을 학습합니다.
GMM (Gaussian Mixture Model), SGMM (Subspace Gaussian
Mixture Model), nnet (Deep Neural Network) 알고리즘을 사용하여
모델을 학습하고, 각 알고리즘의 음성 인식 에러율을 측정합니다.



**GAUSSIAN
MIXTURE
MODEL**



-GMM



–nnet

성능 평가:

학습된 모델을 사용하여 테스트 데이터에 대한 음성 인식 성능을 평가합니다.

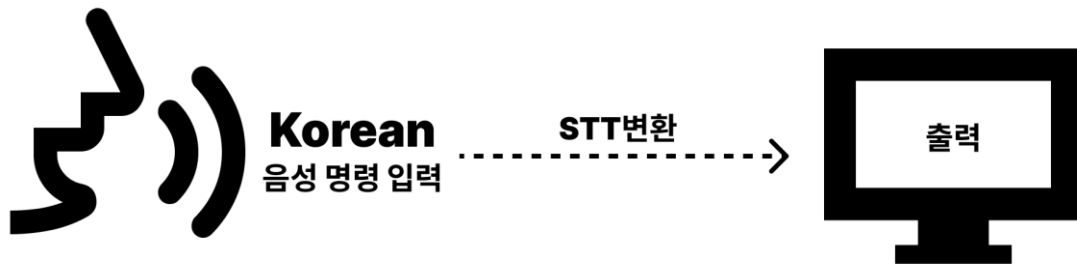
각 알고리즘의 음성 인식 에러율을 비교하여 성능이 우수한 알고리즘을 선택합니다.

음성 데이터셋:

<https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=130>

<https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=123>

프로젝트 목표:



목적 및 목표 설정:

목적: 라즈베리 파이 5 를 사용하여 한국어 음성 인식 시스템을 오프라인으로 개발하는 것입니다.

목표: 사용자가 한국어 음성 명령을 입력하면 시스템이 해당 명령을 인식하고 적절한 작업을 수행할 수 있도록 하는 것입니다.

사용자 요구사항 식별:

시스템은 한국어 음성을 인식할 수 있어야 합니다.

사용자의 음성 명령에 따라 적절한 작업을 수행할 수 있어야 합니다.

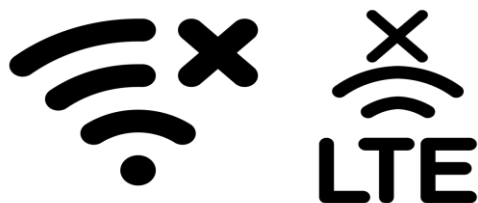
기능 요구사항 정의:

음성 입력을 받고 인식하는 기능이 필요

모델 학습 및 성능 측정을 위한 테스트 기능이 필요

제약 조건 및 제한사항 고려:

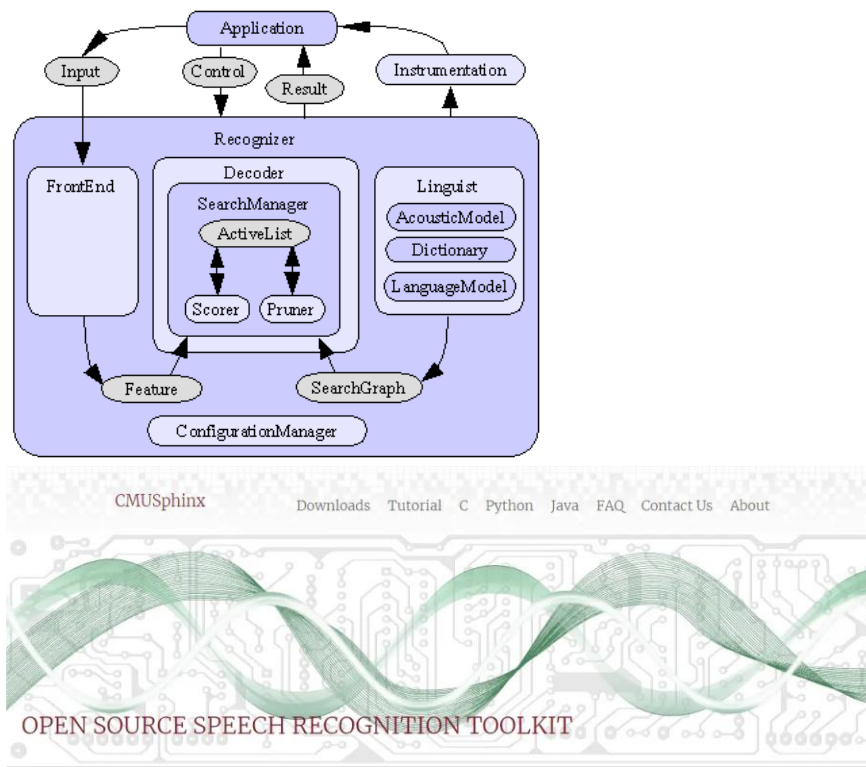
라즈베리 파이 5 의 하드웨어 성능 및 용량 제한을 고려해야 함
오프라인 환경에서 작동해야 하므로 인터넷 연결이 없어도 동작해야 함.



품질 요구사항 명시:

음성 인식의 정확도가 높아야 함
시스템의 응답 속도가 빨라야 함

라이브러리 선택:

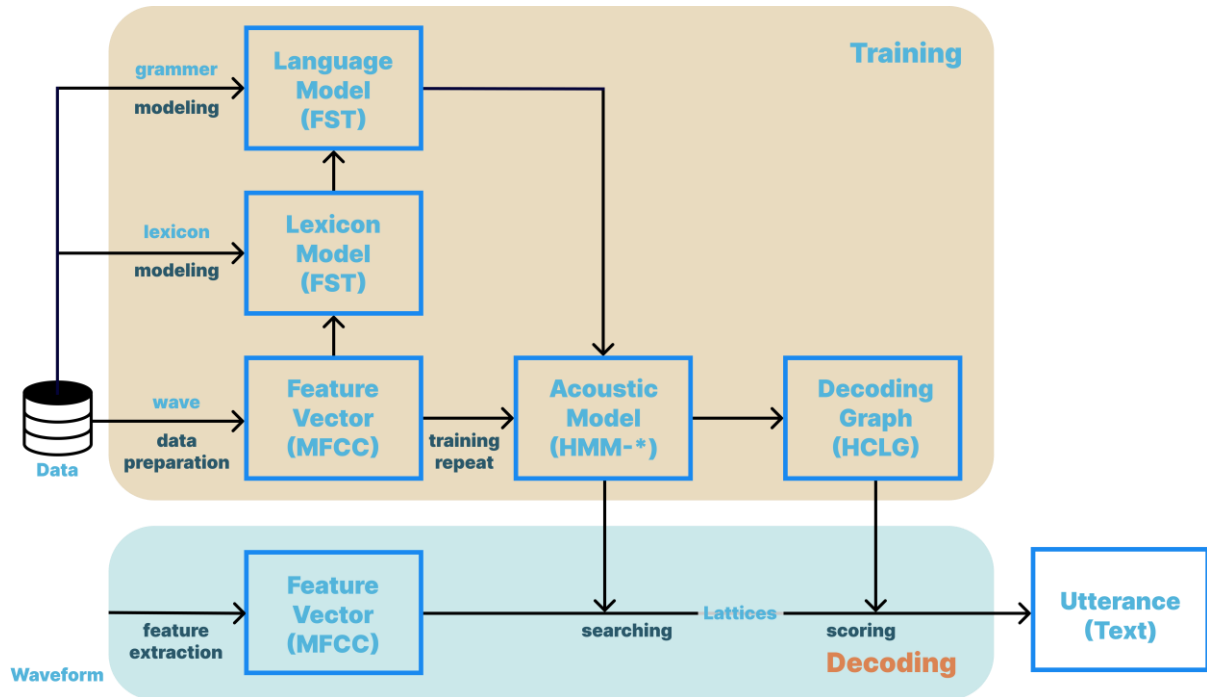


Oct 13, 2022

SphinxTrain 5.0.0 is released!

There is also an updated release of SphinxTrain, and the acoustic modeling tutorial has been updated to reflect the new and simplified usage. Still working on the other tutorials, sorry.

프로젝트의 요구사항과 목표에 부합하는 적절한 음성 인식 라이브러리를 선택해야 함. 오프라인 라이브러리 중 Pocket Sphinx 나 CMU Sphinx 같은 오픈 소스 음성인식은 머신 러닝을 활용하는 KALDI 와 같은 대규모 음성 데이터에 비해 정확도가 낮고 어휘 크기가 제한적이다. 따라서 주어진 요구사항에 따라 Kaldi 를 선택. Kaldi 는 오픈 소스로 제공되는 음성 및 음향 신호 처리 도구킷으로, 음성 인식 및 다양한 음향 모델링 작업에 사용된다. 다음은 Kaldi 를 선택하는 이유와 함께 라이브러리 선택의 요약이다:



Kaldi 선택 이유:

Kaldi 는 오프라인 음성 인식에 적합한 강력한 도구로 알려져 있습니다.

다양한 학습 알고리즘과 모델링 기법을 지원하여 음성 인식 정확도를 높일 수 있습니다.

널리 사용되는 오픈 소스 프로젝트이며, 커뮤니티 및 문서화가 잘 되어 있어 개발에 도움이 됩니다.

다양한 언어 및 데이터셋에 대한 지원을 제공하므로 한국어 음성 인식에도 적용 가능합니다.

라이브러리 선택 요약:

프로젝트의 목표와 요구사항을 고려하여 Kaldi 를 선택하였습니다.

Kaldi 를 사용하여 한국어 음성 인식 시스템을 개발할 것입니다.

Kaldi 를 활용하여 음성 데이터를 처리하고 모델을 학습하여 프로젝트의 목표를 달성할 계획입니다.

라이브러리 선택 단계에서는 프로젝트의 목표와 요구사항에 부합하는 라이브러리를 신중하게 선택하여 프로젝트를 원활하게 진행할 수 있습니다. Kaldi 를 선택함으로써 한국어 음성 인식 시스템을 개발하는 데 필요한 도구와 자원을 보다 효율적으로 활용할 수 있습니다.

개발 환경 설정:

라즈베리 파이 5 준비:

라즈베리 파이 5 를 준비하고 최신 버전의 Raspbian 또는 Raspberry Pi OS 를 설치합니다.
필요한 경우, 라즈베리 파이 5 에 마이크 및 스피커와 같은 오디오 장치를 연결합니다.

Kaldi 설치:

라즈베리 파이 5 에 Kaldi 를 설치합니다. 이를 위해 Kaldi 의 최신 버전을 다운로드하고 컴파일합니다.
컴파일 및 설치에 필요한 종속성을 해결하고 Kaldi 를 빌드합니다.

데이터 전처리:

AIHub 의 한국어 음성 데이터셋을 다운로드하고 필요한 전처리 작업을 수행합니다.
음성 파일을 형식에 맞게 변환하고 Kaldi 가 사용할 수 있는 데이터 디렉토리 구조를 생성합니다.

모델 학습 준비:

음성 데이터셋과 해당 데이터에 대한 텍스트 transcriptions 을 생성합니다.
Kaldi 의 스크립트 및 도구를 사용하여 데이터의 피쳐 추출과 모델 학습을 위한 준비 작업을 수행합니다.

Kaldi 설정 및 환경 변수 설정:

Kaldi 의 실행 파일 경로와 환경 변수를 설정하여 사용자가 언제든지 Kaldi 를 실행할 수 있도록 환경을 구성합니다.

데이터 수집 및 전처리:

데이터 수집 및 전처리 단계에서는 학습에 필요한 음성 데이터를 수집하고 이를 적절하게 전처리하여 모델 학습에 사용할 수 있도록 준비합니다. 아래는 데이터 수집 및 전처리의 주요 단계입니다:

데이터 수집:

AIHub 또는 다른 오픈 데이터 리포지토리에서 한국어 음성 데이터를 수집합니다.

데이터셋은 다양한 음성 명령 및 문장으로 구성되어야 하며, 여러 사람의 음성을 포함해야 합니다.

다양한 환경에서 녹음된 데이터가 포함되어 있어야 하며, 잡음이나 환경 소음이 포함된 경우도 고려해야 합니다.

데이터 정제:

수집된 음성 데이터를 오디오 형식으로 확인하고 정제합니다. 필요에 따라 음성 파일의 노이즈나 중복을 제거하고 데이터의 일관성을 유지합니다.

음성 텍스트 정렬:

각 음성 파일에 대한 텍스트 transcriptions 을 작성합니다. 음성과 해당 텍스트 간의 정확한 매핑을 위해 음성 파일의 각 구간에 대해 해당하는 텍스트를 작성합니다.

데이터 형식 변환:

Kaldi 에서 사용할 수 있는 형식으로 음성 파일과 해당 텍스트를 변환합니다.

오디오 파일을 WAV 형식으로 변환하고, 각 음성 파일의 텍스트 transcriptions 을 Kaldi 의 데이터 형식에 맞게 변환합니다.

데이터 디렉토리 구성:

Kaldi 의 데이터 디렉토리 구조에 따라 데이터를 구성합니다.
음성 파일과 해당 텍스트 파일을 각각의 디렉토리에 저장하고,
각각의 파일에 대한 리스트를 작성합니다.

피쳐 추출 준비:

Kaldi 에서 사용할 수 있도록 각 음성 파일에서 피쳐를 추출하기
위한 준비 작업을 수행합니다.
필요한 피쳐 추출 스크립트를 작성하고 데이터를 피쳐 추출에
필요한 형식으로 변환합니다.

모델 학습:

모델 학습 단계에서는 전처리된 데이터를 사용하여 음성 인식 모델을
학습시키는 과정을 진행합니다. 이를 위해 Kaldi 에서 제공하는 다양한
학습 알고리즘을 사용할 수 있습니다. 주어진 프로젝트의 경우, GMM
(Gaussian Mixture Model), SGMM (Subspace Gaussian Mixture Model),
그리고 nnet (Deep Neural Network) 알고리즘을 사용하여 각각의 음성
인식 모델을 학습시키는 과정을 진행할 수 있습니다. 아래는 모델 학습의
주요 단계입니다:

모델 준비:

Kaldi 에서 사용할 수 있는 기본적인 음성 인식 모델을
준비합니다.
각 학습 알고리즘에 따라 모델을 초기화하고 설정합니다.

모델 학습:

준비된 데이터셋을 사용하여 각 학습 알고리즘에 대한 모델을
학습시킵니다.
GMM, SGMM, nnet 알고리즘을 순차적으로 적용하여 각각의
모델을 학습시킵니다.
각 모델의 파라미터를 조정하고 최적화하여 음성 인식의 정확도를
향상시킵니다.

학습 평가:

학습된 모델의 성능을 평가합니다.
테스트 데이터셋을 사용하여 각 모델의 음성 인식 정확도를 측정하고 비교합니다.
각 학습 알고리즘에 대한 어려움을 측정하고 성능을 분석합니다.

모델 선택:

학습된 모델 중에서 성능이 가장 우수한 모델을 선택합니다.
각 알고리즘의 성능을 비교하여 가장 적합한 모델을 선정합니다.
선정된 모델을 이후의 음성 인식 시스템에 적용합니다.

모델 저장:

학습이 완료된 모델을 저장합니다.
각 모델에 대한 파라미터 및 상태를 저장하여 나중에 사용할 수 있도록 합니다.

성능 평가:

성능 평가 단계에서는 학습된 음성 인식 모델의 성능을 정량적으로 측정하고 분석합니다. 이를 통해 모델의 정확도와 효율성을 평가하고 필요한 경우 모델을 개선할 수 있는 방안을 찾을 수 있습니다. 아래는 성능 평가의 주요 단계입니다:

평가 데이터 선택:

모델의 성능을 평가할 데이터를 선택합니다. 이 데이터는 학습에 사용되지 않은 독립적인 데이터셋이어야 합니다.
테스트 데이터셋이나 교차 검증 데이터셋을 사용합니다.

모델 평가:

선택한 데이터셋에 대해 학습된 모델을 적용하여 음성 인식을 수행합니다.
모델이 예측한 결과를 정답과 비교하여 정확도를 계산합니다.

에러율이나 정확도 등의 성능 지표를 사용하여 모델의 성능을 평가합니다.

성능 지표 계산:

다양한 성능 지표를 사용하여 모델의 성능을 측정합니다.

일반적으로 사용되는 지표에는 다음이 있습니다:

정확도 (Accuracy): 모델이 정확하게 예측한 비율을 측정합니다.

에러율 (Error Rate): 모델이 잘못 예측한 비율을 측정합니다.

혼동 행렬 (Confusion Matrix): 모델의 예측 결과와 실제 값 간의 관계를 시각화하여 분석합니다.

Precision, Recall, F1-score: 클래스별로 모델의 성능을 측정하는 지표입니다.

결과 분석:

성능 평가 결과를 분석하여 모델의 강점과 약점을 파악합니다.

혼동 행렬을 통해 주요 오류 패턴을 확인하고 모델의 개선 방향을 결정합니다.

모델 개선:

성능 평가를 통해 발견된 모델의 약점을 개선하기 위한 다양한 방법을 모색합니다.

데이터 추가 수집, 모델 파라미터 조정, 피처 엔지니어링 등의 방법을 사용하여 모델을 개선합니다.

평가 보고서 작성:

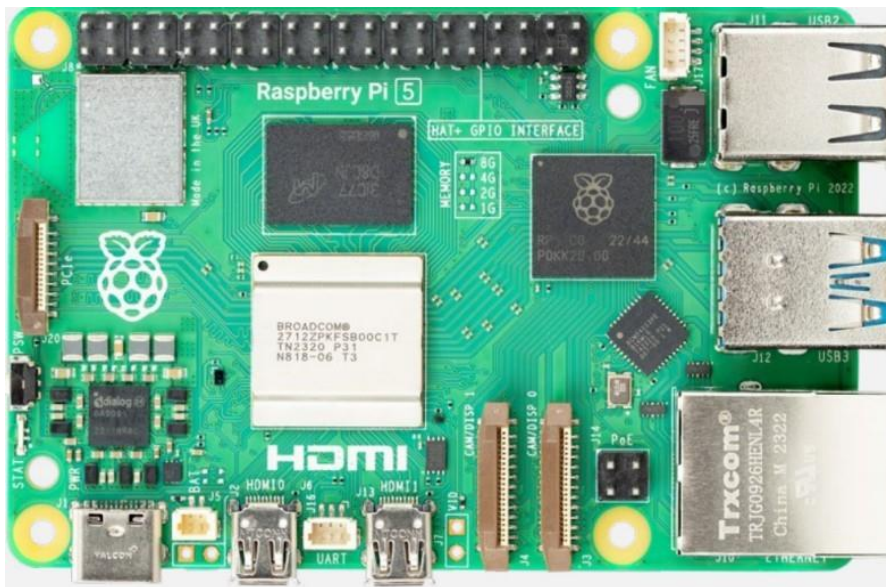
성능 평가 결과를 요약하고 분석 내용을 기록한 평가 보고서를 작성.

보고서에는 사용된 데이터셋, 성능 지표, 결과 분석, 개선 계획 등을 포함.

3. 개발 계획 및 기자재

-제품 구현 시 필요한 하드웨어 부품

(1) 라즈베리파이 보드



라즈베리파이 5

148,170 원

(2) 디스플레이

[samgu electro] 라즈베리파이 5 호환 Pcle M.2 SSD 2280 확장 보드

14,500 원

[라즈베리파이 5 호환 PCIe M.2 SSD 2280 확장 보드 \[TSC-RP503\]](#)
[/ 디바이스마트 \(devicemart.co.kr\)](#)

(4) 마이크로폰 센서



[Seeed] ReSpeaker 4-Mics Linear array Kit for Raspberry Pi

40,400 원

[ReSpeaker 4-Mic Linear Array Kit for Raspberry Pi \[107990056\]](#) /
[디바이스마트 \(devicemart.co.kr\)](#)

3. 팀원 간 업무 내용

길진성

회의록 작성
자재 구입
제품 기획
음성 데이터 수집
음성 데이터 전처리

김현욱

제품 기획
음성데이터 수집
음성 데이터 전처리
모델 학습

김태성

제품 기획
회로 구현
모델 학습

정선진

제품 기획
회로 구현
음성 데이터 수집
모델 학습

4. 작업 제작 추진 계획 및 일정

	4 주	5 주	6 주	7 주	8 주
	물품 구입 및 자세한 계획 수립 & 구조 장치의 하드웨어적 설계도(디자인)제작		Kaldi 설치와 음성 데이터셋 전처리		중간 고사 기간

	9 주	10 주	11 주	12 주	13 주	14 주	15 주	16 주
	전처리한 음성 데이터를 통해 모델 학습				작동 여부 테스트와 성능 향상			

5. 지원 경비 사용 계획

NO.	물 품 이 름	금 액
1	<i>라즈베리파이 5</i>	148,170 원
2	[Sunfounder] 라즈베리파이 3.5 인치 TFT LCD 디스플레이	37,000 원
3	[samgu electro] 라즈베리파이 5 호환 Pcle M.2 SSD 2280 확장 보드	14,500 원
4	[Seeed] ReSpeaker 4-Mics Linear array Kit for Raspberry Pi	40,400 원
5		
6		
7		
8		
9		
	합 계	240,070 원