Whisper API

Automatic Speech Recognition Model (ASR)

- **ASR (Automatic Speech Recognition)**: 음성을 받아 텍스트로 변환하는 기술.
  - **STT (Speech to Text)** 라고도 부른다.
- ASR은 다양한 분야에서 활용됨: 자동 자막 생성 / 음성 명령 인식 / 회의 및 강의 녹취록 작성 / 다국어 번역 지원

Whisper API

Whisper

- **Whisper**는 OpenAI가 개발한 강력한 음성 인식 모델로, 다양한 언어의 음성을 텍스트로 변환할 수 있음.
- Whisper의 주요 특징:
  - **고품질의 음성 인식**: 다양한 억양과 배경 소음을 포함한 데이터로 학습되어 높은 정확도를 제공함.
  - **다국어 지원**: 여러 언어의 음성을 인식하고 번역할 수 있음.

Whisper API 사용법

`1_whisper.py` 코드를 살펴보자.

```
from openai import OpenAI
client = OpenAI(api_key=YOUR_KEY_HERE)

audio_file= open("./data/audio.wav", "rb")
transcription = client.audio.transcriptions.create(
    model="whisper-1",
    file=audio_file
)

print(transcription.text)
```

파일 경로의 표현

- `.`: 현재 디렉토리
- `..`: 부모 디렉토리
- `/`: 디렉토리 내부에 있는 파일 또는 폴더를 구분
- `filename.extension`: `.` 뒤에는 파일의 **확장자(extension)** 가 오며, 파일 형식을 나타낸다.
  - 오디오 파일 확장자: `.wav`, `.mp3`, `m4a`...
  - 이미지 파일 확장자: `.png`, `jpeg`, `jpg`...
  - 비디오 파일 확장자: `.mp4`, `.mov`...
  - 텍스트 파일 확장자: `.txt`...

> audio.wav 대신에 audio_noise.wav를 입력으로 하고 싶다. `1_whisper.py`코드를 수정해보자.

>

예제: 오디오 파일을 텍스트로 저장하기

`2_transcription.py`코드를 살펴보자.

```
from openai import OpenAI

client = OpenAI(api_key=YOUR_KEY_HERE)

audio_file = open("./data/audio.wav", "rb")
transcription = client.audio.transcriptions.create(
    model="whisper-1",
    file=audio_file
)

# 변환된 텍스트를 파일로 저장
with open("./output/transcription.txt", "w") as text_file:
    text_file.write(transcription.text)

print("Transcription saved to transcription.txt")
```

> 위 코드를 실행했을 때 발생하는 에러는 무엇이며, 에러를 고치기 위해서는 어떻게 해야 하는지 생각해보자.

퀴즈: 음성 인식 정확도 비교하기

실험 목표

- Whisper 모델을 사용하여 다양한 유형의 오디오 파일을 입력하고, 음성 인식의 정확도를 비교한다.
- 어떤 상황에서 Whisper가 더 잘 인식하거나, 오류를 발생시키는지 분석한다.

실험 방법

1. 다양한 음성 환경의 오디오 녹음

아래 4가지 유형의 오디오 파일을 직접 녹음하여 준비한다.

1. **영어 오디오**: 또박또박 읽기
2. **한국어 오디오**: 또박또박 읽기
3. **한국어 속삭이는 오디오**: 작은 목소리로 속삭이며 읽기
4. **한국어 여러 명이 동시에 말하는 오디오**

2. 추가적인 오디오 3개 녹음

실험을 확장하기 위해 **특수한 상황**의 오디오 3개를 추가로 녹음한다.

예시:

- **배경 소음이 있는 환경에서 말하기** (카페, 지하철 소음 등)
- **빠르게 읽기 or 느리게 읽기**
- **감정을 담아 말하기 (기쁨, 분노, 슬픔 등)**
- 학생들이 직접 아이디어를 내어 실험해도 됨

3. Whisper 모델을 사용하여 오디오 변환

준비한 7개의 오디오 파일을 `2_transcription.py`코드를 활용하여 텍스트로 변환한다.

4. 변환된 텍스트 비교 및 분석

- 원래 말한 내용과 Whisper가 변환한 결과를 비교한다.
- 어떤 상황에서 오류가 발생하는지 확인하고, Whisper의 강점과 한계를 분석한다.

결과 정리

아래의 표와 같은 형태로 결과를 정리해보자.

오디오 유형	예상 결과	실제 결과	오류 여부
영어 오디오	?	?	?
한국어 오디오	?	?	?
한국어 속삭이기	?	?	?
여러 명이 동시에 말하기	?	?	?
직접 추가한 오디오	?	?	?