# 한 문장으로 검산하기

내 강의 음성, 로컬 TTS로 만들고 전사로 검산하기

파일을 처음부터 듣기 전에 원고와 전사를 비교하세요. 숫자, 단위, 부정어처럼 뜻이 달라지는 곳을 먼저 찾습니다.

## 01 실습 묶음으로 먼저 듣기

첫 수업의 실습 묶음을 풀고 version.wav를 재생하세요. script.txt와 transcript.txt를 나란히 열어 달라진 첫 문장을 찾습니다. 검토 페이지의 실습실에서는 같은 예제를 바로 불러올 수 있습니다.

| 낭독 원고 | 실제 Whisper 전사 |
| --- | --- |
| 버전은 삼 점 일입니다. | 버전은 3.1입니다. |
| 다시 한 번 확인합니다. | 다시 한 번 확인합니다. |

## 02 판정 예시를 한 줄로 기록

| 파일과 시각 | 달라진 내용 | 판정 |
| --- | --- | --- |
| version.wav / 0:00–0:01.40 | 삼 점 일 → 3.1 | 숫자 표기 차이 |

원음에서 들리는 숫자와 전사의 숫자를 확인하세요. 삼 점 일과 3.1은 같은 값을 뜻합니다. 소수점이 빠져 31로 적혔다면 원음에서 값이 달라졌는지 다시 판단합니다.

## 03 일부러 차이를 심은 연습 예시

‘입력은 저장하지 않습니다’가 ‘입력은 저장합니다’로 적힌 상황을 가정합니다. 원음을 들을 수 없다면 발화 오류인지 인식 오류인지 판정을 보류합니다.

# 로컬에서 WAV 만들기

Apple Silicon Mac과 Python 3.12, FFmpeg 기준입니다. 모델 다운로드와 라이브러리 설치를 마치고, 다음 합성과 전사는 로컬 파일로 실행합니다.

## 01 모델과 실행 환경 준비

```
python3.12 -m venv .venv
source .venv/bin/activate
pip install mlx-audio==0.5.9 soundfile 'huggingface_hub[hf_xet]'
hf download mlx-community/Qwen3-TTS-12Hz-1.7B-CustomVoice-8bit \
  --local-dir qwen3-customvoice-8bit
```

## 02 묶음의 코드로 같은 원고 합성

압축을 푼 폴더의 synthesize.py와 script.txt를 사용합니다. 제공된 원본을 보관하고, 새 파일명으로 다시 합성할 수 있습니다.

```
python synthesize.py --model ./qwen3-customvoice-8bit \
  --output version_rebuilt.wav
```

기본 화자는 Aiden입니다. 영상에서는 같은 코드로 생성한 version.wav를 사용합니다. 새로 합성한 파일의 길이는 환경과 실행 조건에 따라 달라질 수 있습니다.

## 03 제공된 원본을 별도로 전사

```
pip install openai-whisper
whisper version.wav --model small --language ko --device cpu \
  --fp16 False --condition_on_previous_text False \
  --output_format txt --output_dir transcript
```

모델과 패키지 설치에는 네트워크가 필요할 수 있습니다. transcript/version.txt를 script.txt와 나란히 놓습니다. 묶음의 transcript.txt는 강의에서 얻은 전사입니다.

# 들어볼 곳을 기록하기

짧은 인사·질문·설명 중 하나를 골라 먼저 검산하세요. 모델과 화자, 난수, 온도를 적어 두면 변경 전후를 비교할 수 있습니다.

## 청취 검수표

| 칸 | 이번에 남길 기록 |
| --- | --- |
| 파일명·들을 시각 | _____ |
| 원고 | _____ |
| 전사 | _____ |
| 실제로 들은 말 | _____ |
| 판정과 근거 | 발화 / 인식 / 보류 |
| 수정본·자막 | _____ |

## 쉼과 자막의 순서

1. 문장 사이 긴 쉼을 0.10초로 정리합니다. 2. 문장 안은 0.50초를 넘는 쉼만 0.50초로 줄입니다. 컷 앞 0.10초와 꼬리 0.08초는 별도입니다.

version.wav와 version_edited.wav를 순서대로 재생하세요. 예제의 문장 사이 쉼은 0.56초에서 0.10초로 줄였습니다. 같은 문장의 앞뒤 음절을 확인하고 편집본에서 자막 시각을 다시 맞춥니다.

음성 범위는 본인 또는 사용을 허락받은 음성입니다. 기본 화자를 사용했다면 모델과 화자 이름을 기록합니다.

# 올릴 곳의 표시 확인

플랫폼 원문 확인일 2026년 11월 6일. 게시할 때 공식 도움말과 게시 화면을 함께 확인하세요.

| 게시처 | 확인한 기준 |
| --- | --- |
| 인프런 | TTS는 AI 콘텐츠 제작으로 분류. 온라인 강의의 AI 활용 여부를 체크. |
| Instagram | 일반 게시물에서 현실적으로 들리는 디지털 생성·변경 음성은 공개 도구로 표시. |
| YouTube | 본인 음성 복제 해설·더빙은 공개 불필요 예시. 다른 현실적 합성 유형은 자체 기준 적용. |
| TikTok | 현실적인 AI 생성 음성이 들어간 콘텐츠 표시. |

이 강의의 제작 고지 문구는 '로컬 Qwen3-TTS 기본 화자로 합성한 내레이션을 사용했습니다.'입니다. 설명란 문구와 플랫폼 체크 항목은 각각 확인합니다.

## 공식 원문과 실행 참고

인프런 영상 제작 FAQ

인프런 강의 생성 안내

Meta 합성 콘텐츠 표시 안내

YouTube 생성형 AI 콘텐츠 공개

TikTok AI 생성 콘텐츠

Qwen3-TTS 공식 저장소

사용한 MLX 모델 카드

Whisper 공식 저장소