

길라잡이!

외국인을 위한 한류 콘텐츠 기반 발음 교육 서비스

고려대학교 데이터 캠퍼스 4조: 박근형, 이종현, 이정훈, 손소영, 정세연

CONTENTS:

1. Motivation

2. Methodology

3. Expected Effect

Motivation

한류와 함께 성장하는 한국어 교육의 수요



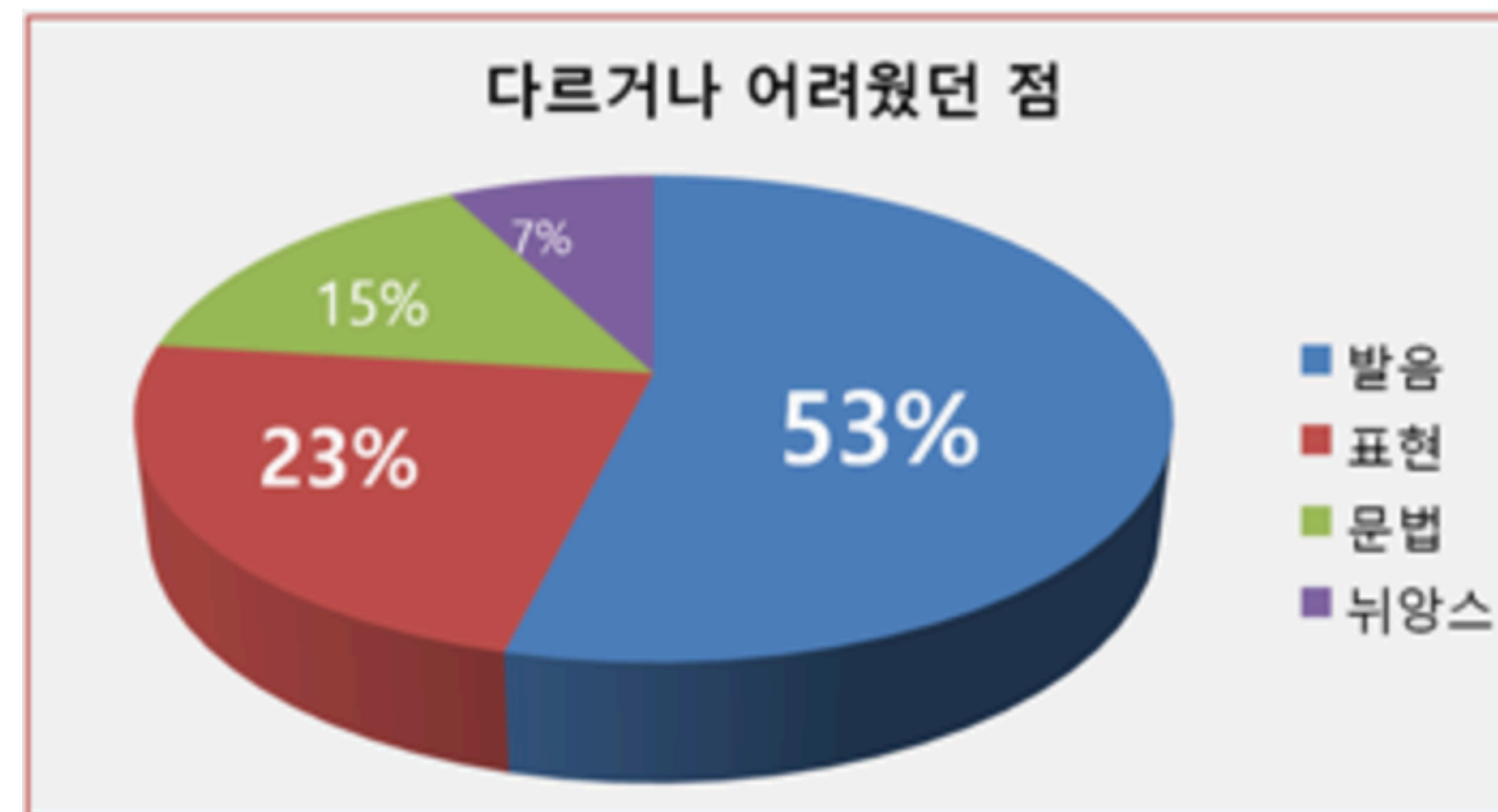
프랑스 대학 한국어 전공 지원 경쟁률 근황...

2021-07-01 16:00

한국 관련 전공 있는 프랑스 대학 지원 경쟁률 수준

2021년 기준 프랑스 대학(보르도) 한국어 전공 경쟁률 28대 1까지

“발음”



외국인 대상 한국어의 어려움에 대한 설문조사

“한류 콘텐츠” + “발음교육”

한류 콘텐츠를 활용한 한국어 발음 교정 서비스

Methodology

How To Service?

1. 한류 콘텐츠 클립 보여주기 - 예능, 드라마, 뮤직비디오
2. 클립의 한국어, 로마자, 영어 자막 출력
3. 사용자가 자막을 따라읽으면 음성을 인식
4. 사용자의 발음을 분석하여 피드백 진행
5. 청각 구두식 교수법에 기반한 반복학습

영상의 음성 한국어 자막 변환방법: STT

Project: Zeroth

- Kaldi-based Korean ASR open-source project
- Official recipe now (https://github.com/kaldi-asr/kaldi/tree/master/egs/zeroth_korean/s5)
- License: Apache 2.0
- forum: <https://groups.google.com/forum/#!forum/zeroth-help>

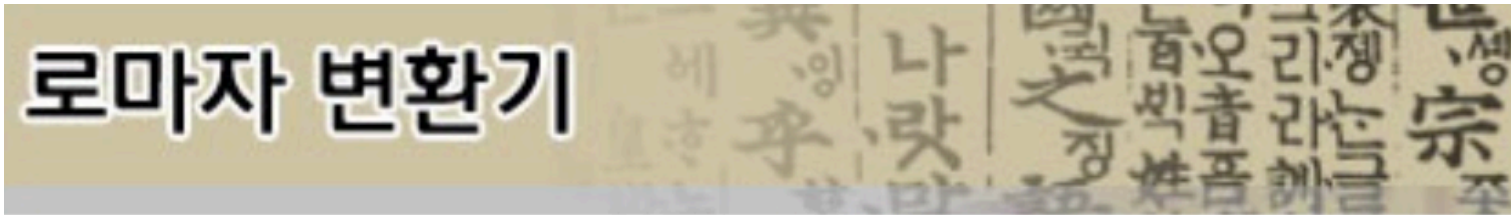
Zeroth is an open source project for Korean speech recognition implemented using the Kaldi toolkit.

This project was developed as part of Atlas Labs's (<https://www.atlaslabs.ai>) Language AI platform, which enables enterprises to add intelligence to their B2C communications.

추출한 한국어 자막을 영어로 번역: PAPAGO API



추출한 한국어 자막을 로마자로 번역: Romaji Converter



[ENGLISH](#)

[의견 보내기](#)

한국어/로마자 변환기는 서울대 이상억 교수님의 자문을 받아
부산대학교 정보컴퓨터공학부 [인공지능연구실](#)과
[\(주\)나라인포테크](#)가 공동으로 만들고 있습니다.

▶ 변환할 타입을 선택해 주십시오.

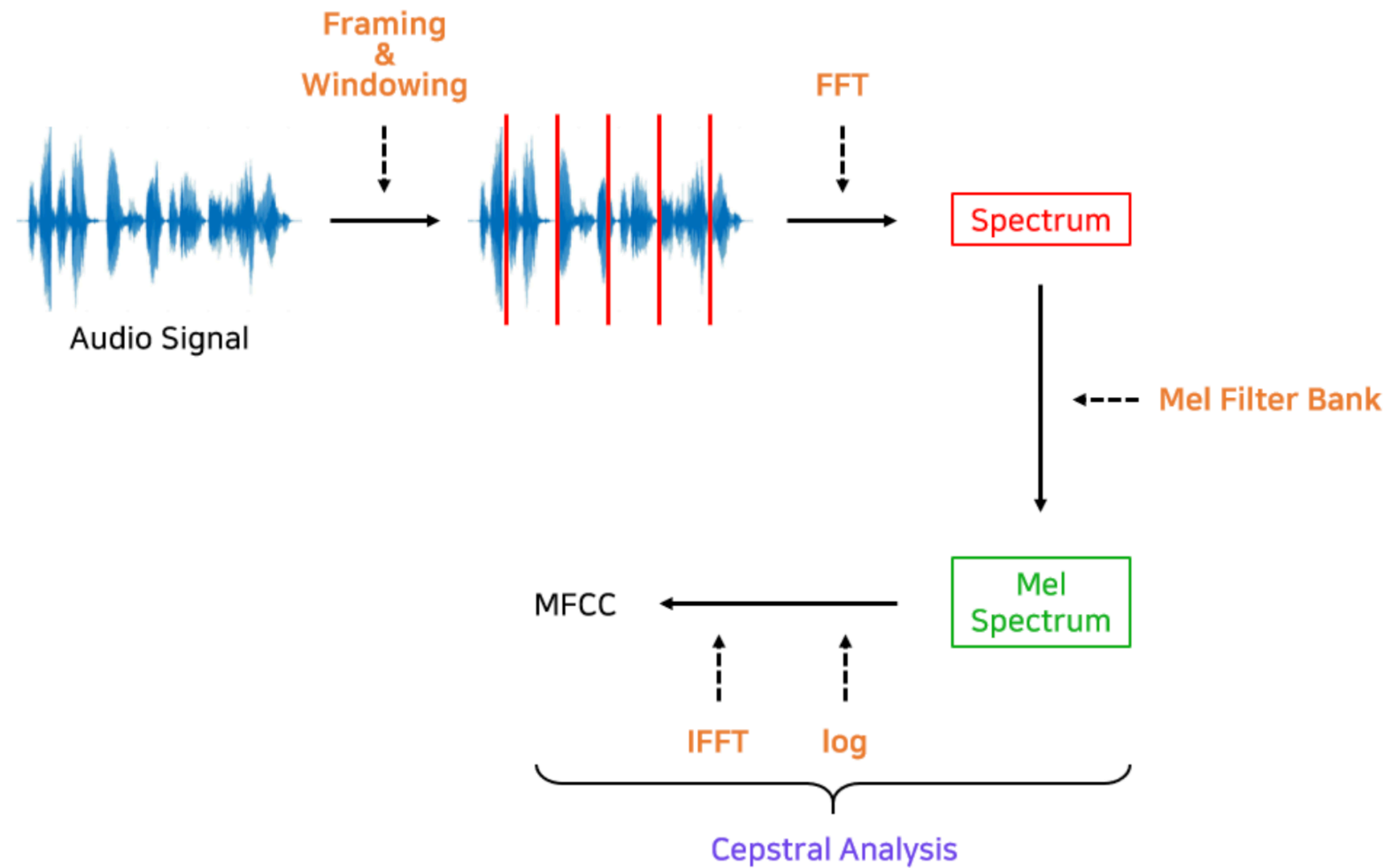
- ☒ 1. 한글 → 로마자 표기법
- ☐ 2. 현행 로마자 표기법 → 매륜-라이샤워 표기법
- ☐ 3. 매륜-라이샤워 표기법 → 현행 로마자 표기법

▶ 변환할 낱말이나 문장을 넣어 보십시오.

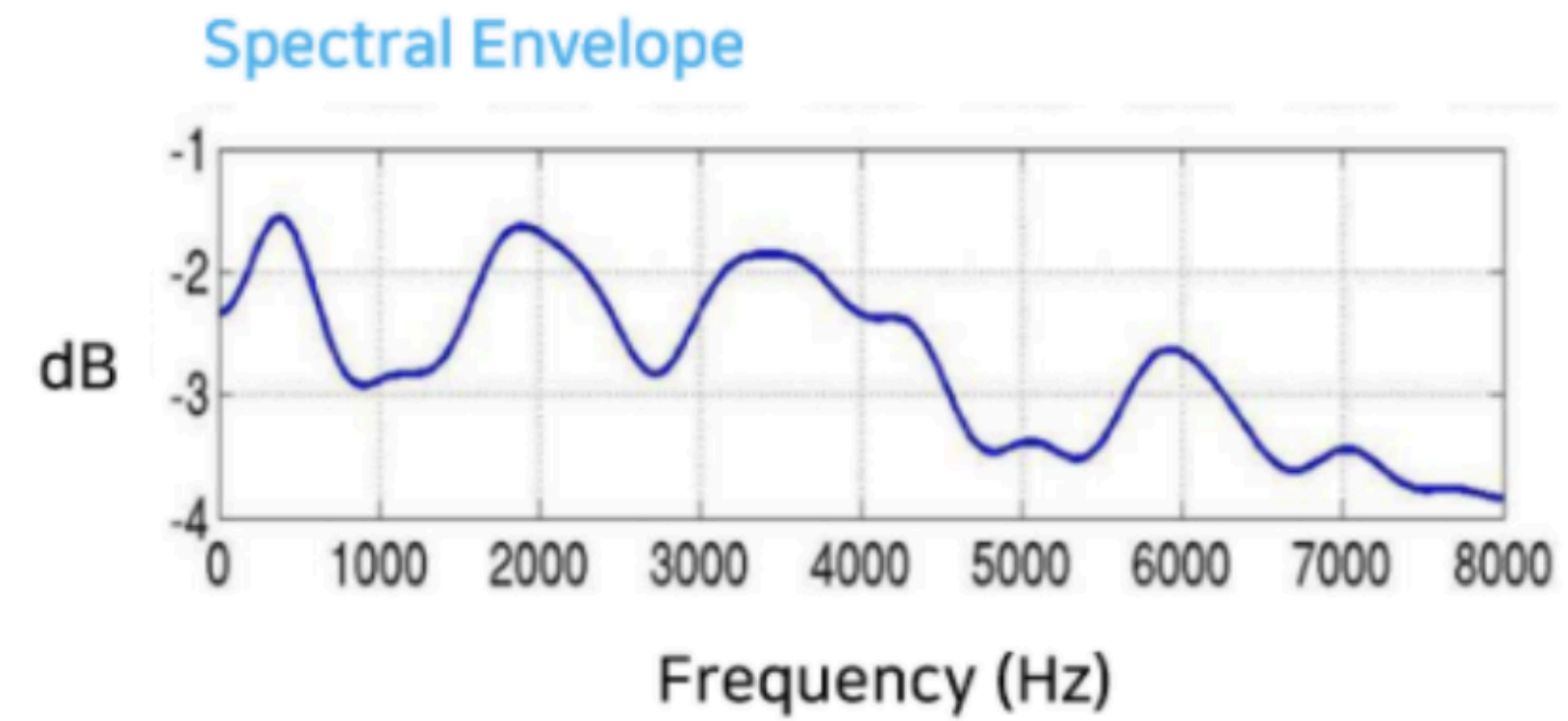
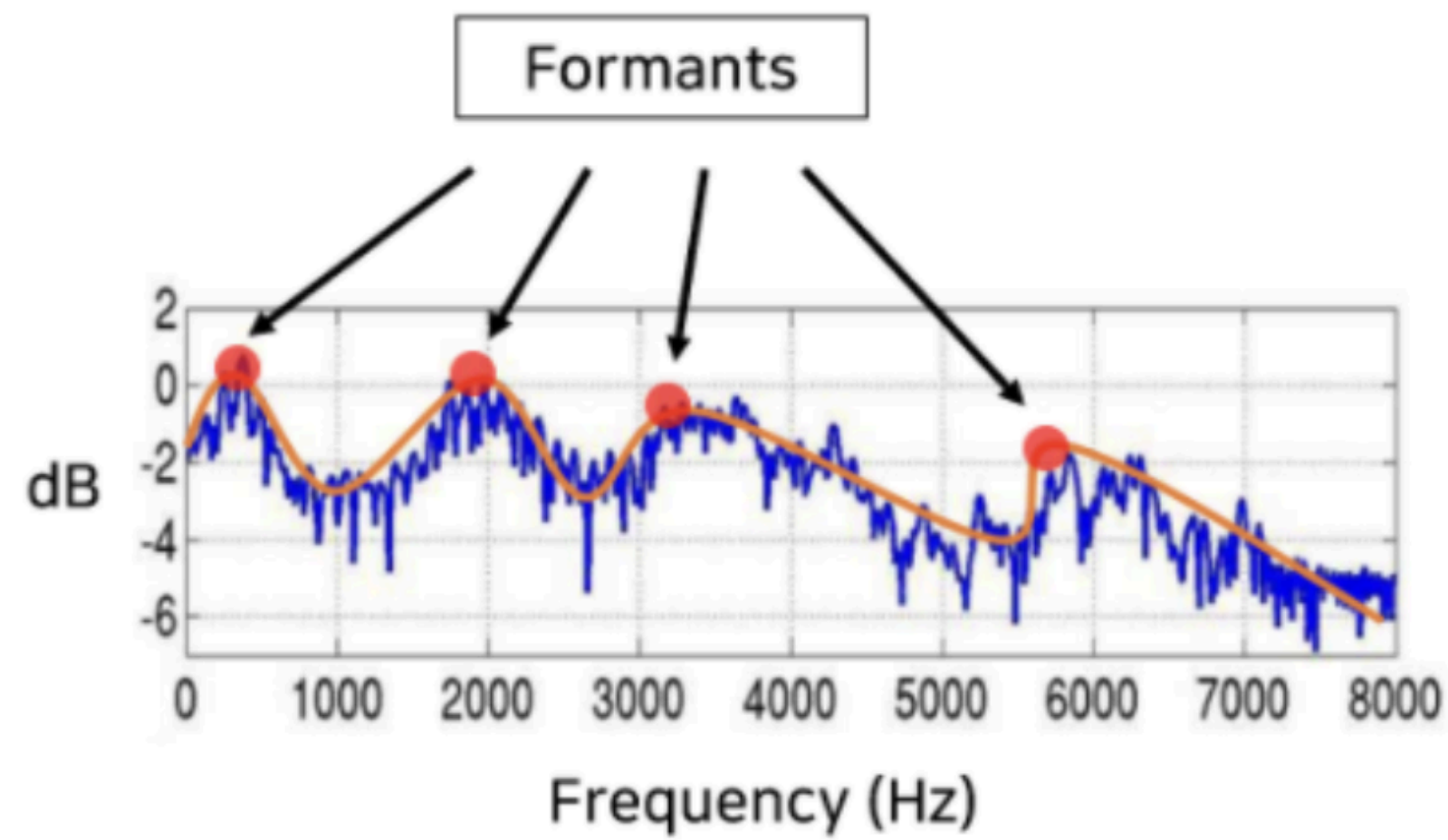
- ☐ 인명
- ☐ 고유명사
- ☐ 행정구역
- ☐ 일반
- ☐ 학술응용
- ☒ 모두

Request → Crawling

데이터의 전처리: MFCC

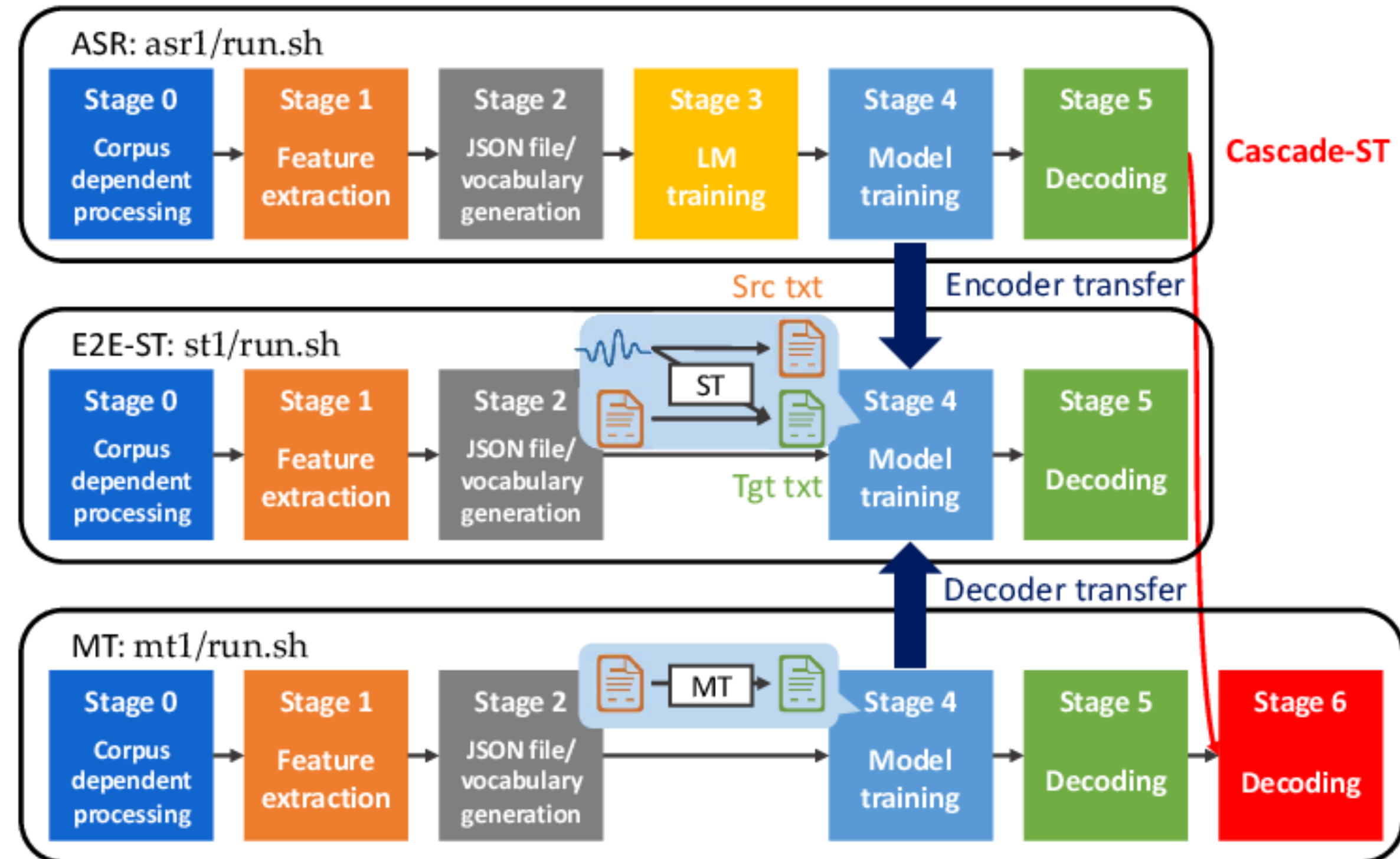


데이터의 전처리: Cepstral Analysis



발음 평가 모델: ESPNet

```
root
├── egs
│   ├── corpus-name
│   │   ├── asr1 (ASR recipe)
│   │   ├── st1 (E2E-ST recipe)
│   │   │   ├── conf (configuration files)
│   │   │   ├── local (corpus-dependent)
│   │   │   ├── data (pre-processed data)
│   │   │   ├── dump (speech features)
│   │   │   ├── exp (model directories)
│   │   │   └── run.sh (main script)
│   │   └── mt1 (MT/Cascade-ST recipe)
│   ├── espnet (model definition)
│   └── tools (tool installation)
```



ESPNet Training Data

음성/자연어
자유대화 음성(노인남녀)

텍스트 오디오

2020

음성/자연어
자유대화 음성(일반남녀)

텍스트 오디오

2020

음성/자연어
차량 내 대화 및 명령어 음성

텍스트 오디오

2020

음성/자연어
한국어 강의 음성

텍스트 오디오

2020



Expected Effect

Interests

다양한 한류 콘텐츠를 통해 한국어를 배움으로써 더욱 흥미를 가지며 학습 가능

Active Communication

가수들과 K-pop 해외 팬들간의 소통이 활발해 지면서 더 많은 코어팬 형성 가능성 증가

Culture

자연스럽게 한국이라는 나라의 문화에 대한 관심도 증가

Q&A, Feedback