

Automatic Speech Recognition

The automatic speech recognition (ASR) module recognizes text from the input audio files. For narrow band and wide band data, different models are used inside this module, respectively. Indices are also prepared for the following keyword spotting pipeline.

Input

Metadata file: containing input audio filename and corresponding genre information (i.e., TB, NB, CS)

Data directory: containing audio files

Output

(please also indicate if this component modifies the input in-place)

CTM file: containing the transcribed text and word confidence

Index directory for keyword spotting

*The inputs are not modified in-place

Docker Commands

```
docker run \  
  -v <parameter A> \  
  -v <parameter B> \  
  -v <parameter C> \  
  -e <parameter D> \  
  -e <parameter E> \  
  <parameter F>
```

- <parameter A>: input directory
- <parameter B>: metadata.tsv path
- <parameter C>: output directory
- <parameter D>: number of jobs
- <parameter E>: number of threads
- <parameter F>: docker image name

Examples

```
docker run --rm \  
-v input:/opt/app/input:ro \  
-v metadata.tsv:/opt/app/metadata/metadata.tsv:ro \  
-v output:/opt/app/output \  
-e NUMBER_OF_JOBS=8 \  
asr
```

-e NUMBER_OF_THREADS=4 \
material/asr-kk:v1

System Requirements

(list only the ones that are applicable)

- CPU
- RAM (10GB * CPUs / 4)

Standalone

yes

Approach

Hybrid LF-MMI system, semi-supervised training for wide-band data.