

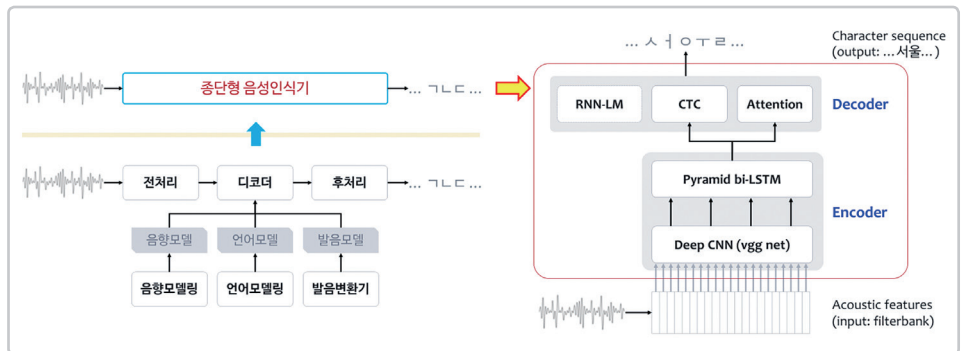
실시간 스트리밍 종단형 음성인식기술(한국어, 다국어 및 발성 평가용) 및 화자 분리 기술(Ver.1.3)

기술개요

- 음향모델과 언어모델, 발음사전 등을 별도로 학습하는 기존의 음성인식 모델을 종단형(End-to-End) 방식으로 통합 학습하도록 개선함으로써 음성인식 성능을 대폭 향상시키고 실시간 스트리밍 음성 인식을 가능하게 하여 활용 범위 대폭 향상
- 회의 환경등과 같이 다수 화자가 같은 장소에서 발생하는 경우 각 화자별로 발생한 구간을 추정하는 화자분리 기술도 사용 가능

기술특성

- 원어민/비원어민을 포함한 대용량의 한국어 및 다국어 음성 DB와 자체 기술을 활용하여 미리 훈련된 실시간 스트리밍 종단형 음성인식 모델로 별도 DB 및 훈련 과정 없이 각 언어별 높은 성능의 음성인식 시스템 구현 가능
- 종단형으로 통합 학습함으로써 높은 정밀도의 실시간 스트리밍 음성-문자 변환(speech-to-text conversion) 가능(오픈소스 활용 및 추가 도구 제공)
- 한국어 및 영어, 중국어, 일본어, 스페인어, 프랑스어, 독일어, 러시아어, 아랍어, 베트남어, 태국어, 인도네시아어, 말레이어 등의 다국어 지원
- 음성인식 적용 도메인의 DB를 기반으로 적응학습 등을 통해 음성인식 성능 제고 가능(오픈소스 활용 및 추가 도구 제공)
- 비원어민의 한국어 또는 영어 발성에 대한 음성인식을 통한 발성 평가가 가능
- 화자 분리 기술을 이용하여 다수의 화자가 발생하더라도 회의록 녹취 등을 효율적으로 진행 가능
- Phrase Hint 기능, Fixed-Grammar(종단형)를 지원하여 도메인 튜닝의 용이성 대폭 향상



적용분야

- ▶ 다양한 실시간 스트리밍 음성인식 응용 서비스(자막 생성, 지능형 컨택센터, 자동 통역, 외국어 교육, 회의록 녹취 등)
- ▶ 음성인식 기반 지능형 스피커
- ▶ 대화형 음성정보 서비스



음성인식 SW

- 방송, 유튜브 자막 서비스
- 컨택센터 실시간 보조
- 실시간 회의록 녹취
- 디지털 사이니지



인공지능 비서

- 인공지능 스피커
- 지능형 로봇
- 스마트폰 App

기술완성도 (TRL)

▶ 9단계 : 사업화



기술이전 내용

▶ 세부기술 1 | 종단형 음성인식기술(Ver. 1.3.)

- 종단형 음성인식 엔진(PyTorch 기반 리눅스, 윈도우 및 MacOS 지원)
- 엔진 튜닝 도구(리눅스, Ubuntu16.04이상, 바이너리, 또는 스크립트 제공)

▶ 세부기술 2 | 실시간 스트리밍 종단형 음성인식기술(Ver. 1.3.)

- 종단형 음성인식 엔진 및 실시간 스트리밍 종단형 음성인식 엔진(PyTorch 기반 리눅스, 윈도우 및 MacOS 지원)
- 엔진 튜닝 도구(리눅스, Ubuntu16.04이상, 바이너리, 또는 스크립트 제공)

▶ 세부기술 3 | 화자분리기술(Ver. 1.3.)

- 회의 등과 같이 2명 이상의 사람이 무작위적 순서로 발생하는 환경에서 녹음된 단일채널 신호로부터 각 화자별로 발생한 구간 정보를 추정하여 출력하는 기술
- 화자분리 엔진 (리눅스, CentOS 7 이상, cuda 10.0 이상, 샘플링 주파수 16kHz 지원)

지식재산권 현황

No.	출원·등록번호	특허명	상태
1	2022-0001723	도메인특화 음성인식 모델 구성 방법 및 장치와 이를 이용한 종단형 음성인식기	출원

기술이전 문의

- ▶ ETRI 연구성과확산실 | 042-860-4881 / etri_tco@etri.re.kr