딥러닝 기반의 서버형 음성인식 기술

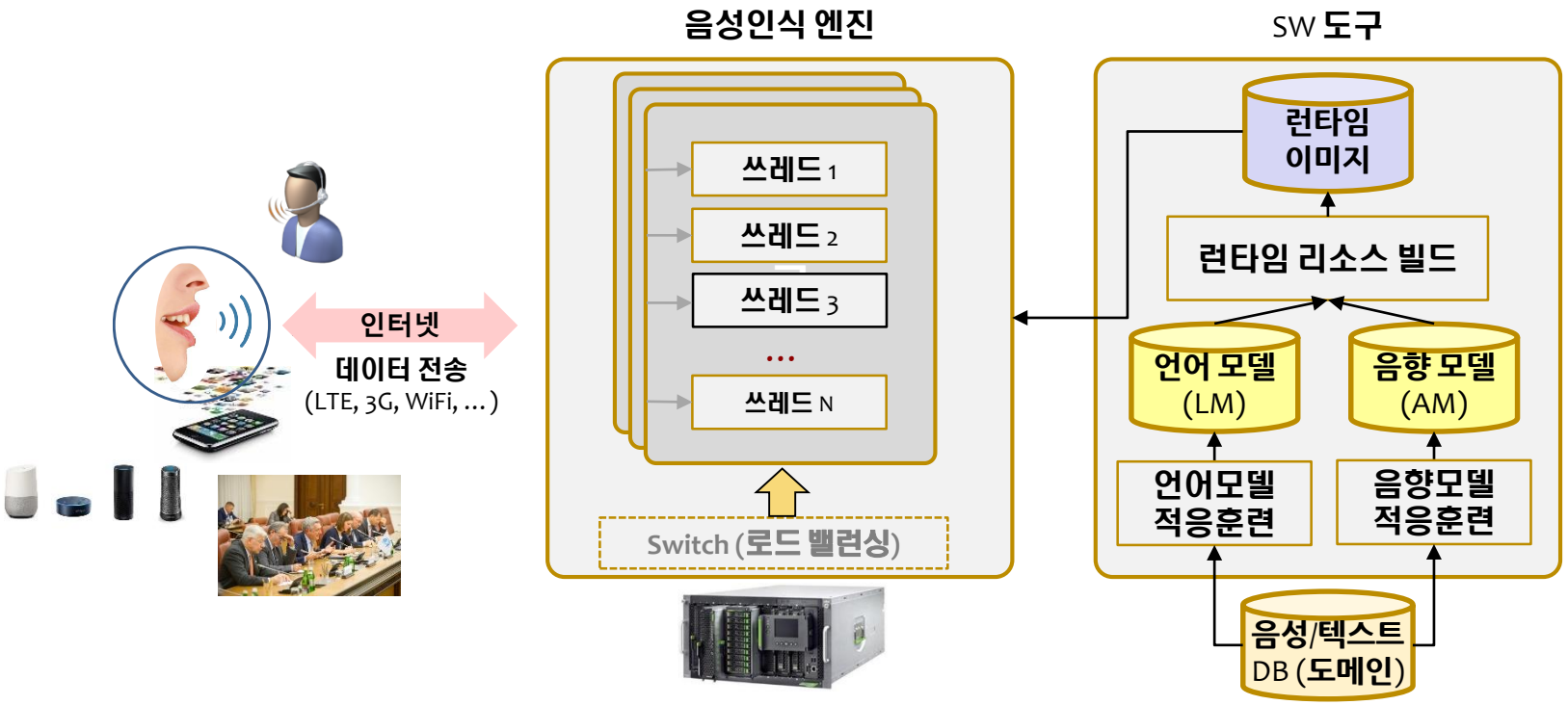


2018. 7.

한국전자통신연구원
Electronics and Telecommunications Research Institute

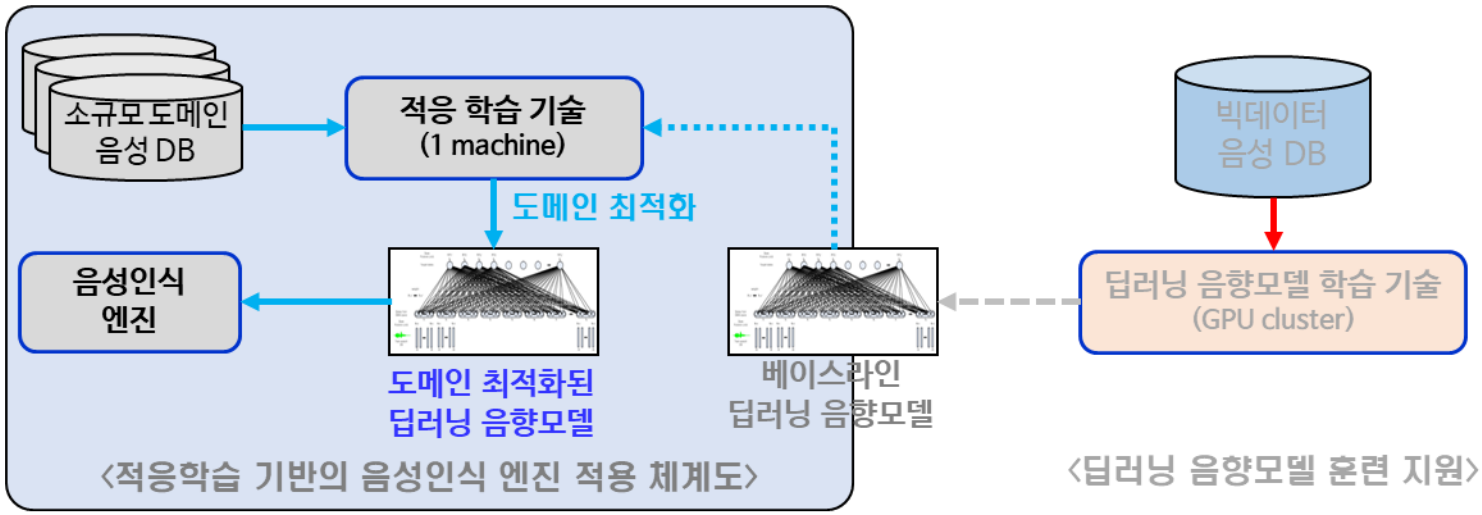
기술 개요

- **딥러닝 기반의 서버형 음성인식 기술**
  - **다양한 이동 환경**(스마트폰, 자동차 등 포함), **고객센터**(유무선 전화 기반의 콜센터, 온라인 및 오프라인 기반의 제반 고객센터 포함), **각종 기록물**(연설, 회의, 발표, 방송 등 포함) **등을 대상으로 하는 고성능 자연어 음성인식 기술**
  - **정보검색, 고객응대, 회의물/오디오물/영상물에 대해 음성-문자 변환** (STT: speech-to-text conversion) **서비스 및 시스템 개발에 활용**



기술 특성

- **딥러닝 기반의 서버형 음성인식 기술**
  - 단일쓰레드 또는 멀티쓰레드 기반의 동시접속 환경에서 실시간(online) 또는 일괄처리(batch) 방식의 음성인식을 위한 리눅스 기반의 서버-클라이언트형 음성인식 기술 지원
  - 최신 딥러닝 기술에 기반하는 서버 기반의 음성인식 기술을 적용하여 높은 정밀도의 음성-문자 변환 인터페이스를 제공
  - 음성인식 적용 도메인의 음성 및 텍스트 DB를 기반으로 적응학습을 통해 음성인식 성능을 제고하는 도구를 제공



기술이전 내용

- **탐색 엔진** (search engine, decoder)
  - 음향 및 언어 모델 등의 지식 베이스에 기반하여 고속/고성능 음성인식 기능을 수행함
  - 문장형태의 음성인식을 위해 연속어 음성인식 엔진을 사용함
  - 음성인식 결과에 대해 단어별 시간경계 정보를 출력함
    ※ WFST 기술 특성상 시간경계 정보가 실제 경계와 완전히 일치하지 않을 수 있음
- **음향 모델**
  - 사무실 및 모바일 환경의 일반적 음성 및 잡음이 반영된 기본 베이스라인 모델을 제공함
  - 베이스라인 모델에 대해 적용 영역의 소규모 데이터를 사용하여 학습을 수행하는 적응학습 도구를 제공함
  - 심층신경망(Deep Neural Network) 모델을 지원함
- **언어 모델**
  - 발음 변환 도구, 통계적 언어 모델 생성을 지원함으로써 새로운 음성인식용 태스크 또는 콘텐츠 입력 자료에 대해 음성인식용 발성사전과 문법 생성 환경을 제공함
  - 기본으로 제공되는 베이스라인 통계 언어 모델에 대해 적용 영역의 데이터를 사용하여 학습을 수행하는 적응학습 도구를 제공함
- **인식 후처리 도구**
  - 음성인식 엔진에서 출력되는 형태소열에 대해 어절을 복원하고 문장단위로 출력하도록 지원함
    ※ 단, 통계 및 예제 기반의 어절 복원 논리 및 알고리즘을 적용하므로, 신조어나 알려지지 않은 형태소열에 대해서는 완전한 후처리가 어려울 수 있음
- **이미지 생성 도구**
  - 음성인식 엔진에서 인식에 필요한 외부파일에 대해 이미지 형태로 생성하기 위한 이미지 생성 도구를 제공함
- **지원 언어**: 한국어, 영어 (※ 언어별로 별건 계약함. 언어별로 언어 처리 도구 및 자원, 음향 및 언어 모델이 상이함)
- **지원 샘플링주파수**: 8kHz, 16kHz (※ 샘플링주파수별로 별건 계약함. 샘플링주파수별로 훈련용 빅데이터 및 음향 모델이 상이함)

기술이전 범위

- 기술명: 딥러닝 기반의 서버형 음성인식 기술

- 기술이전의 범위
  - 리눅스 환경에서 실행 가능한 Library 형태의 오브젝트 파일과 인식용 이미지 파일 생성 도구 (리눅스 Centos 7.x 이상)
    - 서버 기반의 음성인식 엔진 SDK
    - 런타임 이미지 생성 도구
    - 개발자용 지침서
  - 음향 및 언어모델 및 그 적응 학습도구 (리눅스 Centos 7.x 이상)
    - 기본 음향모델 및 적응 학습 도구
    - 기본 언어모델 및 적응 학습 도구

※ 제약조건
  - 지원 샘플링주파수: 8kHz 및 16kHz를 지원함. 샘플링주파수별로 별건으로 기술이전 계약함
  - 지원 언어: 한국어 및 영어를 지원함. 언어별로 별건으로 기술이전 계약함

기술의 사업성

- 예상 응용 제품 및 서비스
  - 음성인식 기반의 인공지능 스피커 및 비서
  - 외국어 말하기 학습 서비스
  - 콜센터/고객센터 녹취 음성인식 서비스
  - 차량 인포테인먼트 서비스

- 사업성
  - 딥러닝 음성인식 기술은 종래 HMM 기반의 음성인식 기술에 비해 오류개선율(error reduction rate)이 매우 높음
  - 음성인식 성능 향상에 따라 인공지능 스피커, IoT, 스마트홈, 스마트카 등 관련 제품 및 서비스 시장이 급속히 확산 및 성장할 것으로 기대함

기술료 제안

● 경상기술료

구분	중소기업	중견기업	대기업
착수기본료 (단위: 천원)	120,000	240,000	240,000
매출정률사용료 (단위: %)	1.25	3.75	5

* 기술이전 특이 조건

1. **언어 및 샘플링주파수 조합에 별건으로 계약함**
   - 상기 착수기본료는 한국어 또는 영어에 대해 8kHz 또는 16kHz 엔진을 최초 계약할 경우의 조건임
   - 최초 계약 이후 다른 언어와 샘플링주파수 조합을 추가 계약할 시의 착수기본료는 50,000천원(중소기업), 100,000천원(중견기업/대기업)으로 계약함
2. **선행기술에 대한 추가 계약**
   - 본 기술에 대해 선행기술을 계약한 경우 착수기본료없이 위 '1. **언어 및 샘플링주파수별 계약**' 조건을 적용함
   - 선행기술의 정의
     1) 콜센터 서비스를 위한 한국어 음성인식기술 (8kHz)
     2) 콜센터 서비스를 위한 영어 음성인식기술 (8kHz)
     3) 모바일환경 고객지원용 서버기반 음성인식 기술 (16kHz, 한국어/영어)

❖ **계약 예시** (중소기업 예시)

✓ **최초 계약 또는 선행기술이 '한국어, 8kHz'인 경우** 120,000천원의 착수기본료를 납부함. **이후 또는 병행하여 '한국어, 16k', '영어, 8k', '영어, 16k' 등 3가지 경우를 추가 계약할 경우에 대해 각각** 50,000천원으로 계약함

감사합니다.