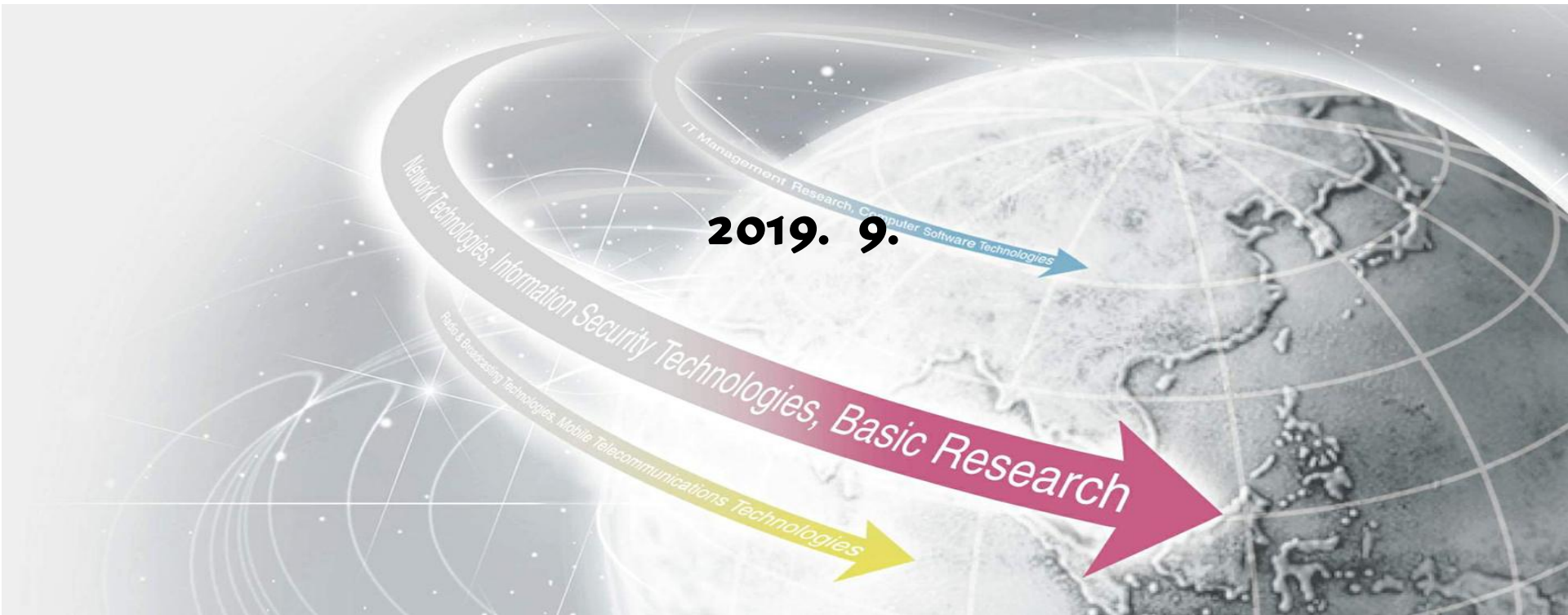


외국어 학습을 위한 음성인식 기반의 발음평가 기술(단말용)



기술 개요

● 정의

“외국어 말하기 학습을 위한 음성인식 기반의 발음평가 기술”

음성인식 기술에 기반한 말하기 중심의 외국어 학습 및 평가용 기술로 비원어민 학습자가 발성한 영어 또는 한국어 음성을 고속, 고성능으로 인식하는 기술과 발성된 음성의 발음 정확도를 평가하기 위한 기술로 구성

● 기술의 특성

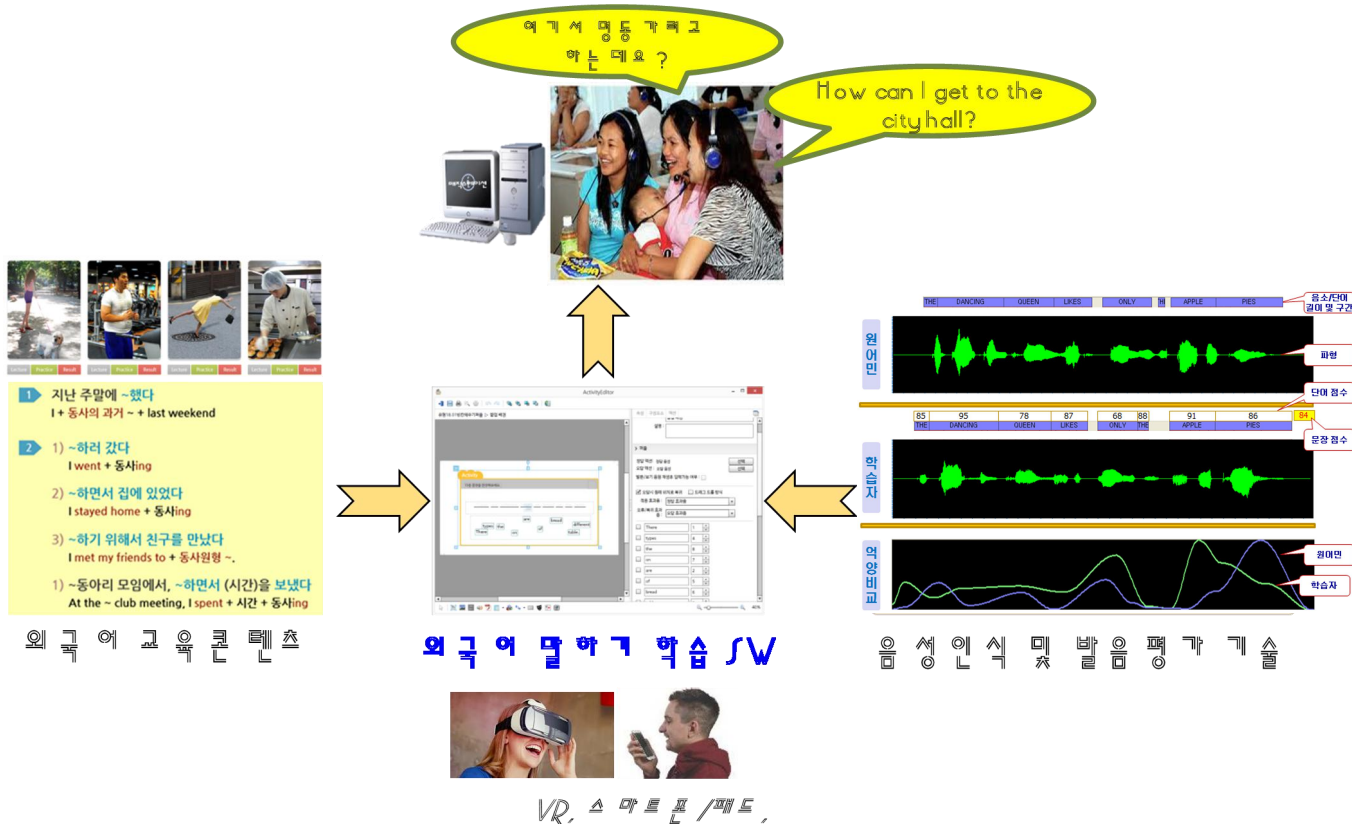
- 문장 형태의 음성인식을 위해 연결어 또는 연속어 음성인식 엔진을 사용
- 새로운 학습 콘텐츠에 대해 인식용 런타임 이미지 생성을 위한 도구를 제공
 - ✓ 음성인식용 발성사전을 위한 발음변환 도구 지원
 - ✓ 주어진 문장/문법 등에 따라 런타임 리소스를 생성하기 위한 이미지 생성 도구 지원

● 기술의 구성

- 전처리 기술: 신호처리 및 디코딩 정보를 사용하는 고성능 끝점 검출 및 특징 추출 기술
- 디코딩 기술: 음향 모델을 프레임에 동기하여 디코딩하는 기술
- 발음 평가 기술: 비원어민이 발성한 문장에 대해 발음의 정확도를 측정하는 기술

서비스 개요

- 외국어 말하기 학습 서비스를 위한 음성인식 기반의 발음 평가 기술
 - 안드로이드/iOS 기반 모바일 단말에서 외국어 학습 서비스용 애플리케이션 등에 적용 가능
 - 비원어민의 한국어 또는 영어 발성에서 발생하는 발음상의 특징을 반영한 음향모델을 탑재



기술이전 내용 및 범위

- 기술명 : 외국어 말하기 학습을 위한 음성인식 기반의 발음평가 기술
- 기술이전의 범위
 - 발음평가 엔진 SDK (object library) : 안드로이드 및 iOS 운영체제 지원
 - 사용자 지침서 [1122-2016-00247, 1122-2015-00641]
 - 예제 응용 프로그램 샘플 코드 [PG20170431]

※ 제약조건

- 본 기술은 “외국어 말하기 학습을 위한 서버형 음성인식 기반의 발성 평가 기술”의 파생 기술로 반드시 “외국어 말하기 학습을 위한 서버형 음성인식 기반의 발성 평가 기술”이 선행되어 또는 동시에 계약되어야 함
- 입력 음성 파일은 16kHz의 샘플링 주파수로 인코딩되어 있어야 함
- 적용 언어: 한국어 및 영어를 지원하며 언어별로 별도 기술이전 계약함

기술료 제안

● 예상기술료

구분		일반 기업		
		중소기업	중견기업	대기업
경상기술료	착수기본료 (단위: 천원)	50,000	100,000	100,000
	매출정률사용료 (단위: %)	1.25	3.75	5

* 기술미전 특이사항

1. 본 기술은 “외국어 말하기 학습을 위한 서버형 음성인식 기반의 발성 평가 기술”의 파생 기술로 반드시 “외국어 말하기 학습을 위한 서버형 음성인식 기반의 발성 평가 기술”이 선행되어 또는 동시에 계약되어야 함
2. 상기 기술미전 조건은 하나의 언어를 대상으로 할 경우의 조건이며, 한국어와 영어를 모두 사용하고 자하는 경우 별건으로 계약하여야 함.

별첨

1. 음성인식 엔진(ESTk-laser) 개요

- ❑ ESTk-laser: ETRI Speech Toolkit - Large Scale Speech Recognizer
- ❑ ESTk-laser is developed to recognize very large scale of recognition domain on both high-end servers and resource-limited embedded devices.
- ❑ Technical features
 - Language independency
 - Platform independency
 - Single channel speech enhancement
 - Noise-robust endpoint detection
 - Speaker and environment adaptation
 - Speaker and channel normalization
 - Deep Learning (deep neural network) support

2. 음성인식 엔진의 구성

□ 탐색 엔진 (search engine 또는 decoder)

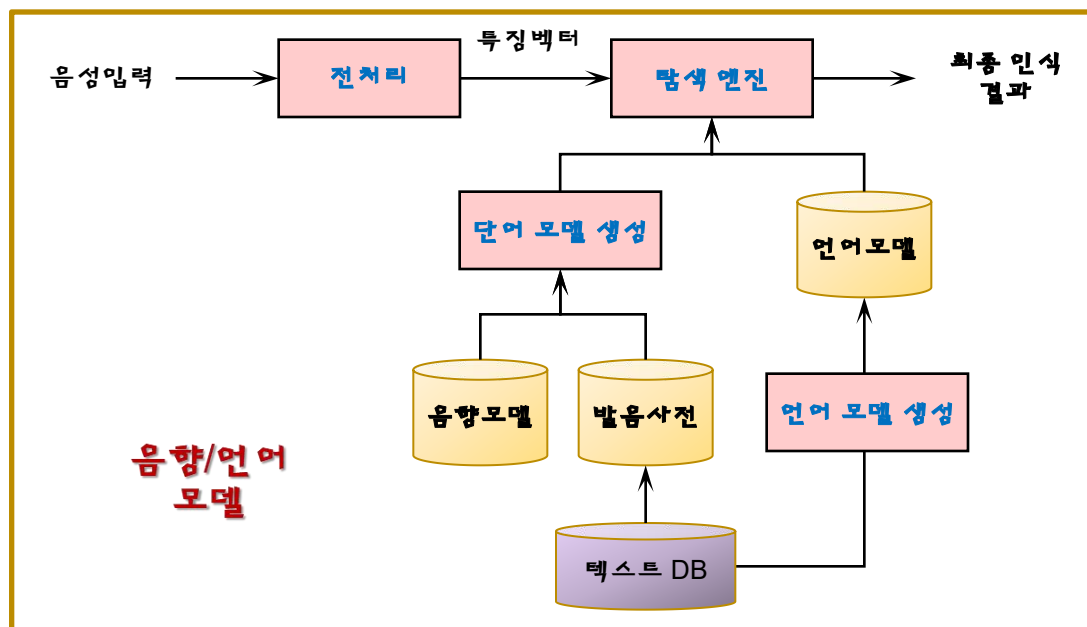
- 음향 및 언어 모델 등의 지식 베이스에 기반하여 고속/고성능 음성인식을 수행

□ 음향 모델

- 사무실 수준의 환경 잡음 반영

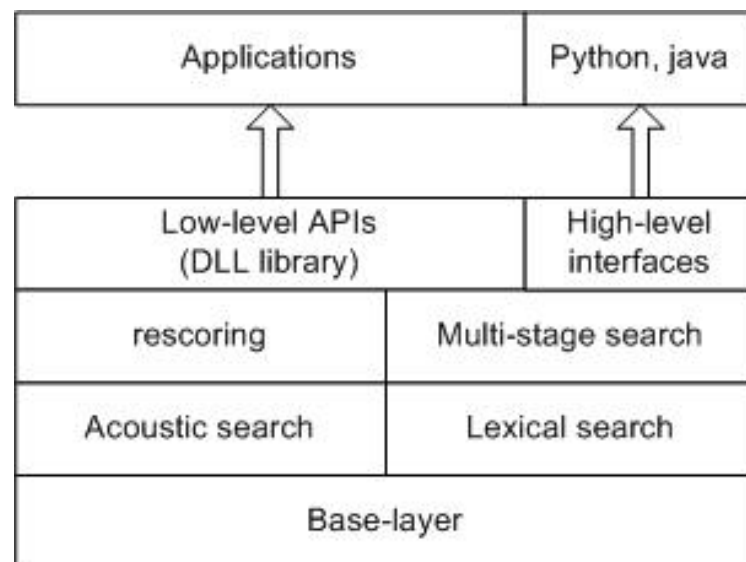
□ 언어 모델

- 단어(POI) 및 문장(주소 등) 반영



3. LASER 구조

- Base layer
 - Wrapper for platform independency
- Decoding layer
 - Acoustic search
 - Lexical search
 - Rescoring
 - Multi-stage search
- Interface layer
 - low-level APIs : DLLs
 - Script-level interfaces : python, java



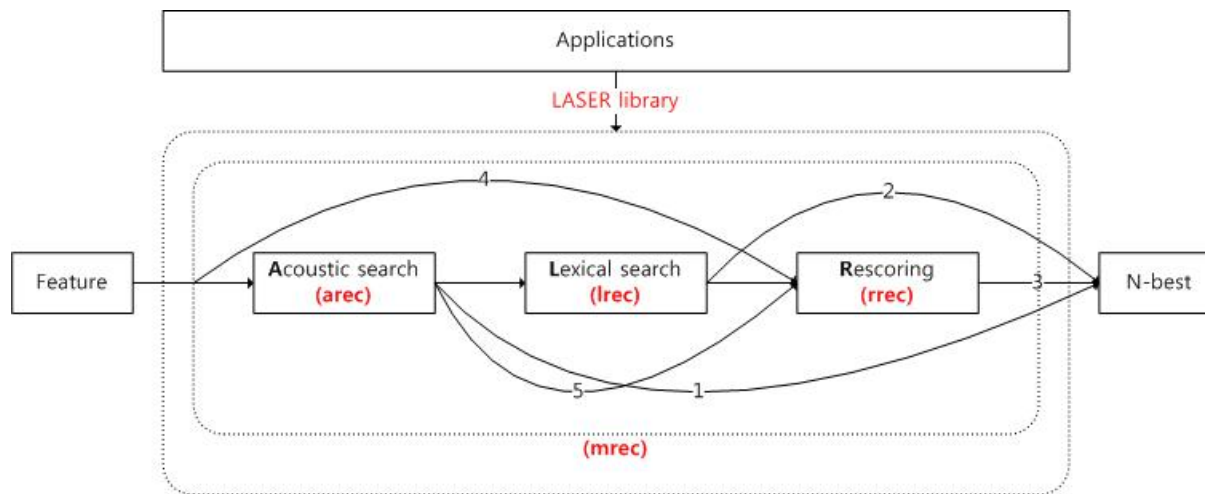
3. LASER 구조

□ Core 2 search components

- Acoustic search
 - Weighted finite state transducer-based speech recognizer
- Lexical search :
 - discrete HMM-based lexical level noisy channel decoder

□ Recognition modes

- Various recognition modes for different domain and system configurations



4. LASER 사양

Consideration		High-end device	Low-end embedded device
Language	Supporting languages	Korean, English	Korean, English
Platform	Supporting platforms	Linux, Windows	Windows, Android, iOS, Nucleus, etc.
Recognition Mode	Continuous	Vocabulary size	>100K (140M trigrams)
		RTF	1.0xRT
	One-shot	Vocabulary size	-
		RTF	-
Minimum H/W requirements	CPU		2.6 GHz
	Storage memory		30GB
	Running memory		40GB
etc	Grammar definition	ARPA, BNF, JSGF	ARPA, BNF, JSGF