

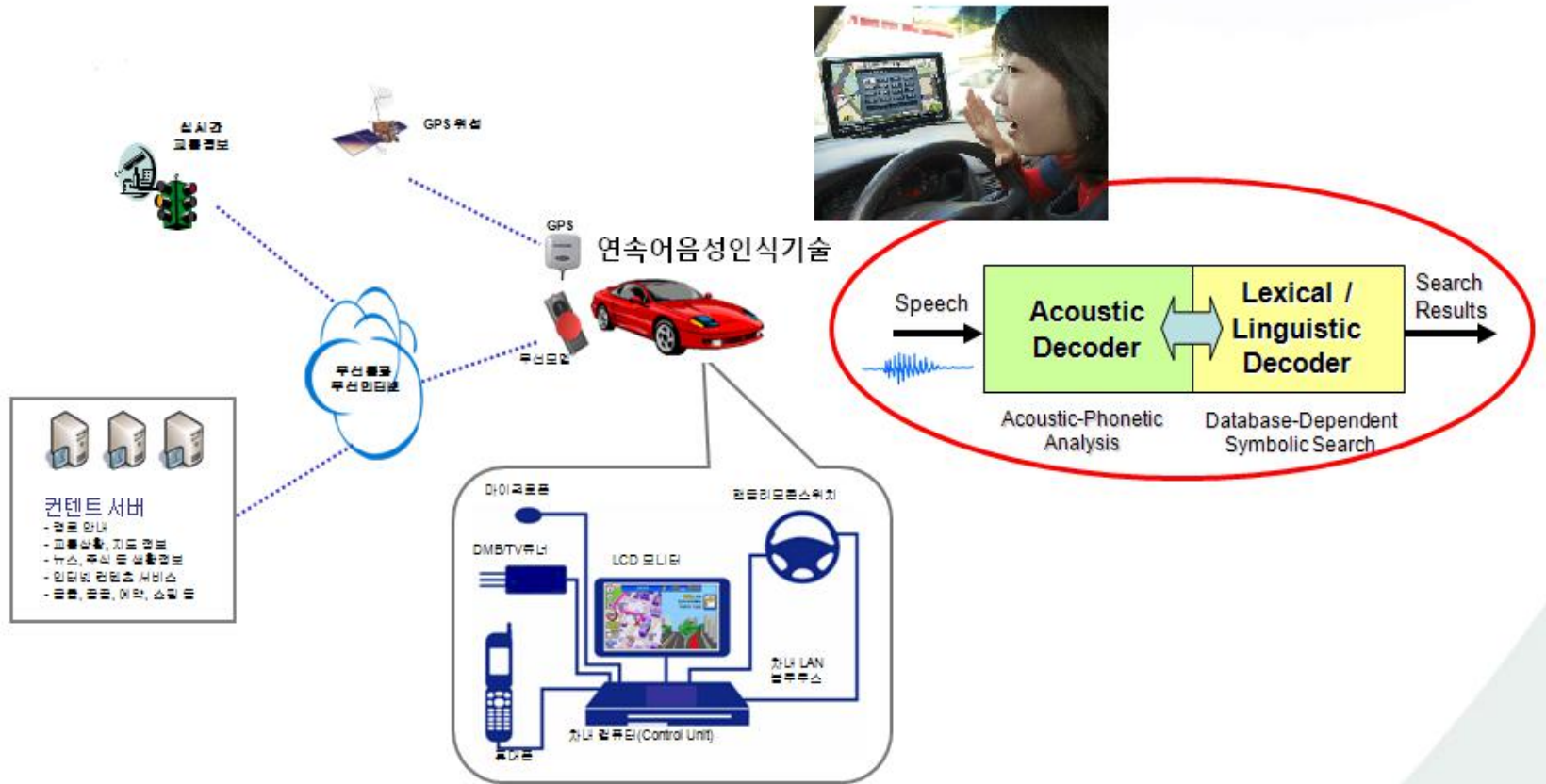
내비게이션 단말기용 연속어 음성인식 기술 ve rsion 2



2016. 1.

기술 개요

● 내비게이션 단말용 연속어 음성인식 소프트웨어



기술미전 내용 및 범위

- 기술명 : 내비게이션 단말기용 연속어 음성인식 기술 version 2
- 기술미전의 범위
 - 단말기 내장형 연속어 음성인식 기술
 - 전국 약 300만 POI 대상 One Shot 행선지 음성인식 기능 (※ POI(Point Of Interest)의 수는 적용 CPU의 성능에 따라 가변적임)
 - 전국단위 주소인식 기능 (신규주소 체계 적용)
 - 단말기 내장형 음성인식 엔진 SDK
 - 오브젝트 라이브러리(object library) 파일: Windows CE, Android, iOS, Linux 등 적용 OS 및 플랫폼에 따라 사용자화(customization) 지원
 - 인식용 이미지 파일 및 이미지 파일 생성 도구
 - 사용자 지침서
- 기술미전 특이사항
 - 기존 “내비게이션 단말기용 연속어 음성인식 기술”의 업그레이드 기술임
 - 심층신경망(DNN; Deep Neural Network) 기반의 심층학습(Deep Learning) 기술에 따른 성능 고도화된 음향 모델과 관련 API가 추가됨
- 상세한 기술 구성은 <별첨> 참조

기술료 제안

- 기술개발 기간 : 2012. 3. 1 ~ 2015. 11. 30

- 착수기본료

| 구분 | | 중소기업 | 중견기업 | 대기업 |
|-----------|-------------|----------|---------|---------|
| 기술료 조건 | 착수기본료 (천원) | 50,000 | 100,000 | 100,000 |
| | 매출점률사용료 (%) | copy당 적용 | | |

- ※ **착수기본료 면제 조건**: 본 건의 하위 버전 기술을 전수받은 기관에 대해서는 상기 착수기본료의 20%인 1천만원을 업그레이드 비용으로 납부하고 계약을 추진함
 • 하위버전 기술명: “내비게이션 단말기용 연속어 음성인식 기술” (계획번호: 7040-2011-0004)

- 매출점률사용료

| 연간 판매 대수 (단위: 대) | 중소기업 | 중견기업, 대기업 |
|------------------|----------------|----------------|
| 1~40,000 | 2,300원 / 1copy | 4,600원 / 1copy |
| 40,001~100,000 | 1,800원 / 1copy | 3,600원 / 1copy |
| 100,001~ | 1,300원 / 1copy | 2,600원 / 1copy |

- ※ **계약조건**: 누적 판매 대수는 년 단위로 reset됨

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)

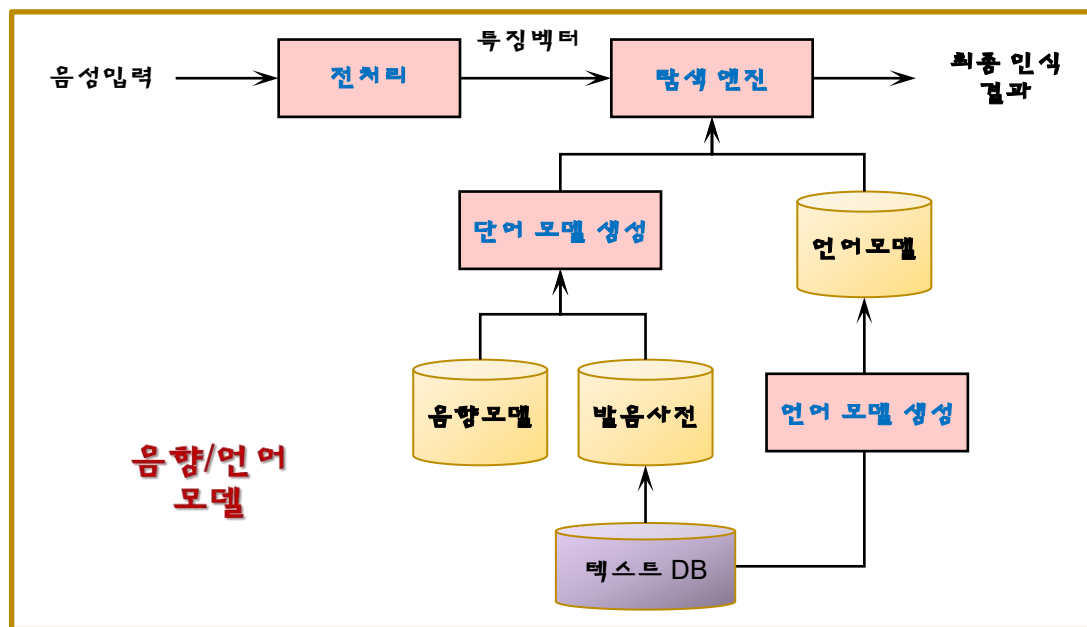
- 음향 및 언어 모델 등의 지식 베이스에 기반하여 고속/고성능 음성인식을 수행

□ 음향 모델

- 차량 환경의 잡음 환경 반영

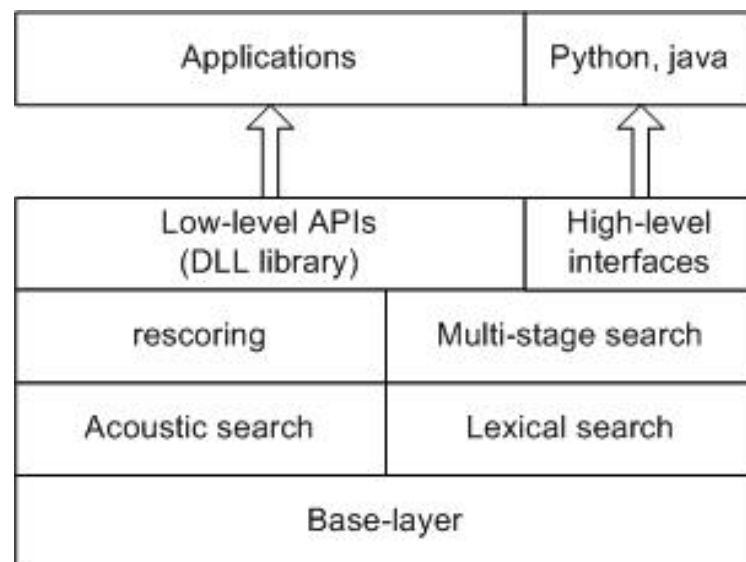
□ 언어 모델

- 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



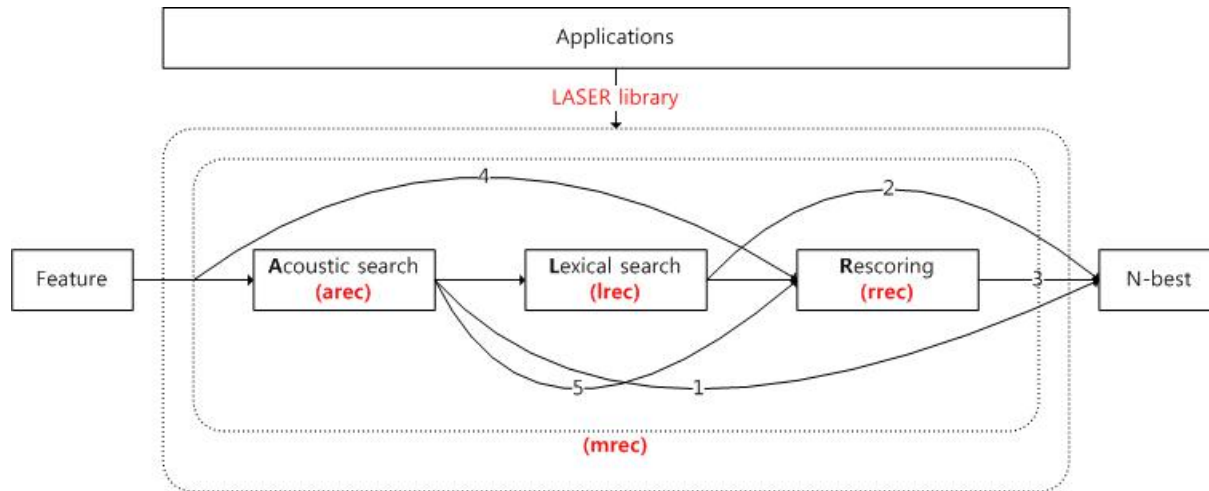
3. LASER 구조

□ 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 |