

2019년 대표성과

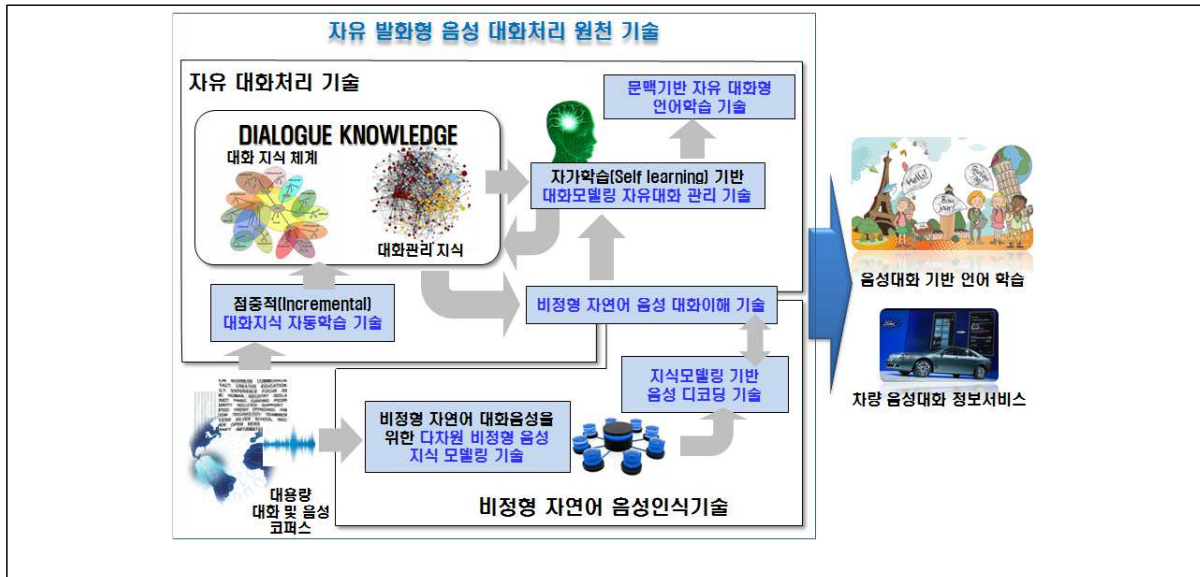
[1] 기본 정보

| | | | | |
|---------------------|--|--|----------------------|-------|
| 후보추천 Track | 미래선도 (), 산업육성 (<input checked="" type="radio"/>), 국가·사회문제해결 () | | | |
| 협약과제명 (협약과제번호) | 언어학습을 위한 자유발화형 음성대화처리 원천기술 개발 (18HS3700) 다중 화자간 대화 음성인식 기술개발 (19HS7700) | | | |
| 총연구기간 | 2015년 2월 ~ 2019년 2월 (18HS3700) 2019년 4월 ~ 2021년 12월 (19HS7700) | | | |
| 총연구비 (단위: 백만원) | 총 14,969 (18HS3700) 총 11,500 (19HS7700) | 정부:11,250, 민간:3,719 (18HS3700) 정부:10,120, 민간:1,380 (19HS7700) | | |
| 성과책임자 정보 | 연구자 성명 | 직할부서 | 연구본부/연구실 | 직위/직급 |
| | 박전규 | 인공지능연구소 | 지능정보연구본부/ 복합지능연구실 | 실장/책임 |
| 연구사업계획서 관련 성과목표명 | [성과목표1-1] 스스로 보고 듣고 읽으며 성장하는 범용 인공지능 원천기술 | | | |

[2] 2019년 우수성과 내용

| |
|--|
| 1. 성과명 |
| 세계 최고 수준의 비정형 자연어 음성인식 기술 |
| 2. 성과내용 |
| 기술개발 목표달성도 |
| <input type="checkbox"/> 기술적 선점이 필요한 분야 <ul style="list-style-type: none"> <input type="radio"/> 컴퓨터가 인간의 역할을 대체 또는 보완하여 듣고(음성인식/이해), 말하고(대화 처리), 가르쳐주는(평가/튜터링) 자유발화형 음성인식 원천 기술 |
| <input type="checkbox"/> 기술개발 목표 <ul style="list-style-type: none"> <input type="radio"/> 세계 최고 수준의 비정형 자연어 음성인식 원천기술 개발 <ul style="list-style-type: none"> (목표 ①) 잡음 및 음원 분리, 다중 특징 고정밀 전처리 기술 (목표 ②) 비정형 자유발화 음성을 위한 다차원 지식 모델링 기술 개발 (목표 ③) 비정형 자연어 발화에 대한 발음 유창성 자동평가 기술 개발 |

<기술개발 개념도>



□ 기술개발 목표의 달성성과 및 핵심기술 확보

[개발목표 ①] 잡음 및 음원 분리, 다중 특징 고정밀 전처리 기술

➔ (핵심기술 확보)

- 딥러닝 잡음 제거 및 음원 분리 기반 다중 특징 통합 전처리 기술 개발
- 잡음·음성 분리 과정의 왜곡 최소화 및 딥러닝 기반 왜곡 정규화 기술 개발
- 언어 교육 및 평가용 다양한 부가 특징 기반 특징 최적화 모델 개발

[개발목표 ②] 비정형 자유발화 음성을 위한 다차원 지식 모델링 기술 개발

➔ (핵심기술 확보)

- 비정형 자연어 음성인식 성능 향상을 위한 음향모델링 기술 최적화
- 비문법/간투어 등에 강인한 언어모델링 기술 개선
- 다차원 지식모델 기반 멀티스트림 디코딩 기술 통합 개발
- 외국어 발음 변이 등에 강인한 비원어민 최적화 음성인식 기술 개발

[개발목표 ③] 비정형 자연어 발화에 대한 발음 유창성 자동평가 기술 개발

➔ (핵심기술 확보)

- 비정형 자연어 발화에 대한 발음 유창성 자동 평가 기술 개발
- 자유발화형 음성인식 기반 언어교육 서비스 개발
- 다양한 모국어의 비원어민 어학 학습 시스템 시범서비스 수행

[기술이전 및 사업화]

➡ **기술이전: 총 1,151.96 백만원**

- 착수기본료: 1,130백만원 / 11건
- 경상실시료: 21.96백만원 / 5건

➡ **사업화: 40+건, 매출액 65+억** (음성솔루션의 순매출액, 주요 10개 기술이전업체 조사)

- **콜센터 녹취음성인식: 22+건** (금융감독원, 오렌지라이프, 흥국화재, 삼성화재, 아주캐피탈, 현대카드, 현대해상화재, SGI서울보증, 우정사업본부 등)

※ 2014년 기술이전/사업화 이래 북미삼성전자 포함 국내외 100+개 사이트 사업화 성공

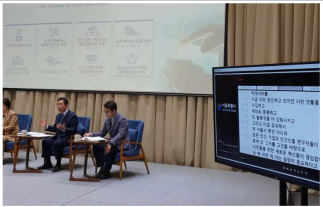



- **회의록/녹취록 및 정보 서비스: 10+건** (SRT, 고리원전, TTA, 서울시, 서울시교육청, 인천시교육청, 인천국제공항 2건, 해양경찰청, 제주시의회 등)

※ 2018년부터 법원(20개 법원 법정적용 시범서비스), 금감원(채권추심, 심사업무 본 서비스), 행안부(시범서비스), 주요 시도의회 등 민간 및 공공부문으로 적용 확장 중

- **외국어교육 서비스: 6+건** (삼성출판사삼성영어, 아이스크림에듀, 비상교육 등)

※ 특히 비상교육은 베트남, 중국, 남미, 캐나다 등에 영어/한국어 교육서비스를 수출

➡ **대표적인 사업화 사례**

| 사업화 (기업) | 사업화 | 비고 |
|-----------------------------|--|---|
| 서울시 (소리자바) | 서울시 인공지능회의록지원 시스템 구축 - 음성인식으로 회의록 내용을 실시간으로 확인, 회의록 작성업무 단축 으로 담당자 업무경감 구현, 신속/정확한 대시민 서비스 구현 |  |
| 인천공항 (소리자바) | 인천공항 인공지능 기반 음성인식 스마트사이니지 시범사업 - 음성인식 결과를 분석하여 스마트사이니지에 구현된 쇼핑/식음매장 및 편의시설 안내 다자간 음성인식 회의록 지원시스템 구축 |  |
| 서울시교육청, 인천시교육청 (소리자바) | - 학부모 운영위 등 다양한 다자간 회의에 대해 실시간 회의록 → 회의 종료 직후 검토/보완 회의록 완성, 막대한 업무 경감 구현 |  |
| TTA/방통위 (소리자바) | 멀티미디어 콘텐츠에 대한 음성인식 기반 자동 자막 생성 시스템 시제품 제작” - 인터넷 서비스되는 대부분의 영상 콘텐츠는 자막과 수어가 제공되지 않아 비장애인과 정보 격차가 심화, 장애인 지원을 위한 온디맨드형 음성인식 자막 생성 시스템 개발 |  |
| SRT (텍스토리) | SRT 열차 방송 녹취 서비스 - 열차내 방송 및 페이지 서비스에 대한 녹취, 대국민 서비스 개선, 품질개선 및 CRM 활용 |  |

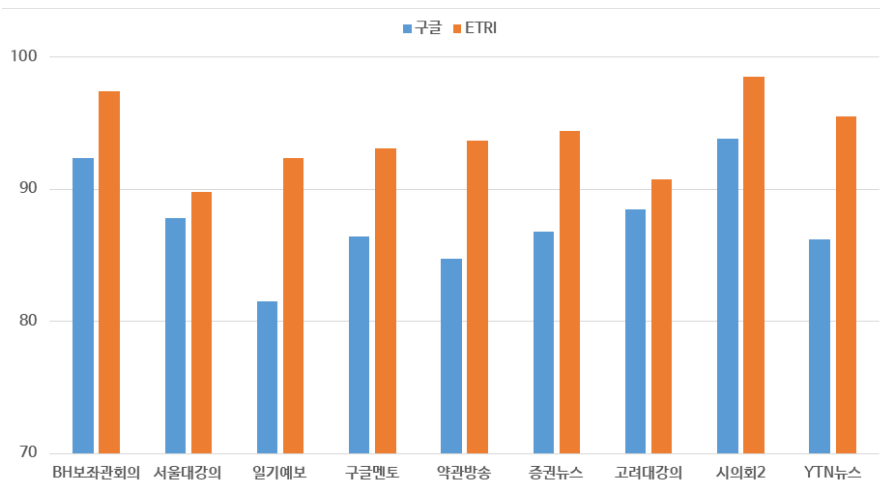
3. 우수성 및 차별성

기술수준 향상 성과

○ **세계 최고 수준의 비정형 자연어 음성인식 기술 확보** : 한국어에 대해 구글의 음성인식 정확도를 상회하는 성능 확보

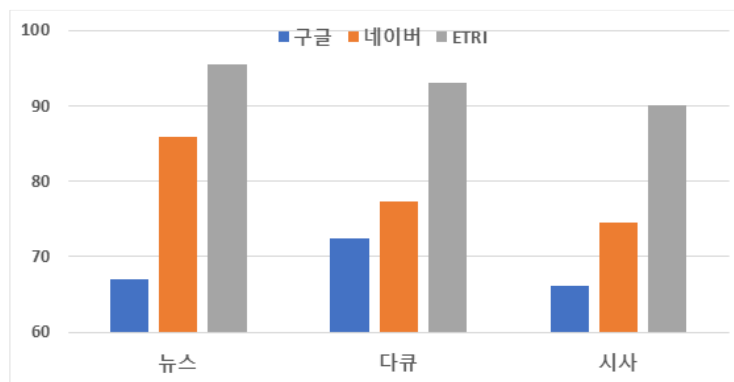
- 자유발화 코퍼스에 대한 수요자 평가 결과 ((주)소리자바, 2018)

※ (주)소리자바: 국내 최대의 속기 서비스 업체



- 방송·인터넷 콘텐츠에 대한 수요자 평가 결과 (TTA, 2019)

※ TTA(한국정보통신기술협회)에서 수행하는 방통위 사업 수행 중 평가 결과



○ **세계 최고 수준의 비원어민 음성인식 정확도 및 유창성 평가 기술 확보** : 비원어민 한국어 발성에 대해 ETS SpeechRater와 동등 이상의 유창성 평가 성능 확보

※ ETS: TOEFL, TOEIC, GRE 등 세계적인 교육용 검사개발 및 측정연구기관

- 낮은 유창성의 비원어민 한국인 발성 영어에 대한 음성인식 정확도 89.44%

※ ETS는 다중국적 비원어민 발성에 대해 82.9% (MC=multiple choice read aloud task, ETRI 평가방식에 비해 현저히 낮은 난이도임), 출처: ETS Research Report No. RR-15-31.

- 한국인 발성 영어에 대해 전문가와 기계간 평가점수의 상관도 $r=0.77$ 수준

※ ETS는 다중국적 비영어권인 발성에 대해 0.74, 출처: ETS Research Report No. RR-15-31.

| 구분 | 리듬 (rhythm) | 발성속도 (speech rate) | 분절음 (segmental) | 억양 (intonation) | 대표점수 (holistic) |
|-------|-------------|--------------------|-----------------|-----------------|-----------------|
| 선형회귀법 | .71 | .66 | .73 | .67 | .75 |
| 딥러닝 | .73 | .68 | .74 | .68 | .77 |

기술수준 공인 성과

- 영어 및 한국어 비정형 자연어 음성인식 정확률 평가: TTA (2019.1.)

| 평가 항목 | 단위 | 목표 | 성과 |
|--------------------------------|-----------|------|-------|
| 영어 비정형 자연어 음성인식 정확률 | 단어인식률(%) | 85 | 89.44 |
| 한국어 비정형 자연어 음성인식 정확률 | 음절인식률(%) | 90 | 91.77 |
| 잡음 환경에서의 음성인식 오류개선 성능 | 오류 개선률(%) | 25 | 25.1 |
| 발음 유창성평가 성능 (인간-기계간 평가점수의 상관도) | 상관계수(0~1) | 0.72 | 0.74* |

* TTA 평가는 2019년 1월 평가결과이며 성능개선을 거쳐 현재 0.77을 달성함

- 상용화시 경쟁 기관 대비 우수한 성능 평가 결과로 서비스 채택

- 네이버 영어학습 앱 케이크(Cake) : 비영어권인 한국어 영어 발성에 대해 우수한 성능평가 결과로 기존 구글 엔진에서 ETRI 엔진으로 교체하여 서비스 중 (2018.12.~)

※ 케이크는 영상 콘텐츠 기반의 영어회화 학습 앱으로 동일 마켓 세그먼트에서 사용자 수 및 보급수 측면에서 최강자임 (구글 플레이 기준 설치수 1,000,000+)

4. 성과의 활용도 및 파급효과

경제 활성화 효과

기업 경쟁력 향상

- (콜센터 녹취 음성인식 기술) 현재 활발히 사업화되고 있는 콜센터 녹취 음성인식 기술에 대한 홍보/기술이전 기업 추가 발굴, 이전기업에 대한 사업화 지원

※ 2014년부터 북미삼성전자 포함 국내외 100+개 사이트 사업화 성공. 특히 사업화 과정에서 세계적인 콜센터/컨택센터 솔루션업체인 Verint, NICE, Nuance 등과의 BMT 결과 성능우수에 따라 채택

- (회의 및 속기를 위한 녹취 음성인식 기술) 현재 사업화 초기 단계에 있는 회의 및 속기 음성인식 기술에 대한 홍보 및 기술이전, 사업화 지원

※ 2018년부터 법원(20개 법원 법정 시범서비스중), 금감원(채권추심, 심사업무 본 서비스중), 행안부(시범서비스 완료), 주요 시도의회(속기록 지원서비스) 등 민간 및 공공부문으로 활발하게 적용중

- (외국어 말하기 학습 기술) 영어 및 한국어 말하기 학습 기술에 대해 시장진입 민간 및 공공 부문에서 활용, 수출 지원

※ 2013년이래 엔씨소프트, 삼성출판사, EBS 등에서 해외경쟁자인 Nuance, SRI 등과 BMT 후 채택

산업 경쟁력 향상

- 자유대화형 언어학습 기술은 플랫폼 및 언어 독립적인 기술로서, 한국어 및 영어뿐만 아니라 중국어, 일본어, 유럽어 등 다른 언어를 위한 언어교육 및 정보 서비스 시스템 및 SW에 쉽게 이식 및 응용 가능함
- 음성대화 핵심원천기술은 차량 정보서비스, 콜센터 녹취 및 고객응대서비스, 노인 말벗서비스를 비롯한 개인비서서비스, 동시통역시스템 등의 성능개선 및 신규 서비스 창출, 확장이 가능함
- 언어학습서비스 및 지능형 정보서비스 영역에서 풍부한 사업화 경력이 있는 참여기업과의 긴밀한 공조를 통해 단계별 사업화 추진 (1단계:핵심기술 응용 사업화, 2단계:언어확장 및 글로벌 사업화)

경제적 파급효과

- 딥러닝을 비롯한 인공지능 알고리즘은 최근 몇 년의 짧은 기간에 음성인식 정확도를 크게 향상시켰으며 이러한 기술적 진보에 힘입어 음성인식 세계 시장 규모는 2017년 91.2억불 규모이며, 2024년 1275.8억불 규모로 고속 성장할 것으로 예상하고 있음(출처: Grand View Research, 2018)
 - 주요 적용영역별 시장 규모 예측



국가사회적 파급효과

○ 해결해야 할 국가사회문제

- 국내 음성인식 원천기술의 수준은 세계 최고 수준과 근소한 격차를 유지하고 있고 한국어는 외국대비 동등 이상의 경쟁 우위를 보이고 있으나, 다국어 기술의 열세는 세계 시장은 물론 국내 시장 경쟁력 저하 요인임. 특히 수출형 제품인 스마트폰, TV, 자동차 등의 UI/UX를 위해 다국어 솔루션 개발이 절실함
- 대화형 AI 외국어 튜터링 서비스를 통해 국가적, 사회적 문제로 대두되고 있는 과도한

외국어 사교육비 문제를 해소하고, 고품질 영어교육을 전 국민에게 제공하여 잉글리시 디바이드(English Divide) 해소에 기여 가능함. 또한 동일 수준의 한국어 서비스는 다문화 가정의 언어소통 문제 해결에 기여 가능함

○ 성과에서 개발된 기술적 솔루션

- 음성인식 기술은 제한된 영역 대상의 시비서, 고객-상담원이 대화하는 콜센터 음성인식 등의 영역에서 인간 수준을 넘어서는 성능 수준을 보이고 있으며 국내에서 활발히 파급 확산되어 관련 시장 성장을 주도하고 있음
- 음성인식의 표준적 성능평가 태스크인 스위치보드 등에 대해 이미 사람의 능력을 넘어서는 수준을 확보하고 있는 등 딥러닝 기술의 발전에 따라 음성인식은 획기적인 성능 개선을 이룬 바 있으나, 심한 잡음환경, (초)원거리 환경, 다화자 발성겹침 환경에 대해서는 성능병목이 심하여 집중적인 기술 개발이 필요함

○ 국가·사회적 파급효과

- 음성인식에 사용되는 딥러닝 및 신경망 기술은 본 과제에서 다루지 않는 언어지능, 시각지능 등 다른 인공지능 분야의 기반 기술 방법론으로 활용가능하여 인공지능 기술의 국가적 고양에 가능함
- 비원어민을 위한 고도한 어학학습 기술은 한국어/영어 발음 교육에서부터 회화 연습, 독해 내용 토의에 이르는 다양한 교육 서비스가 가능하며 특히 언어에 종속받지 않는 원천기술 특성으로 인해 국민의 의사소통능력 고양과 한국어 위상 제고가 가능함