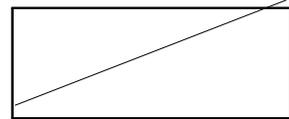


2016년 12월

16ZS1100-16-1124P



언어장벽 없는 국가구현을 위한  
자동통번역 산업경쟁력 강화 사업

Strengthening competitiveness of automatic translation  
industry for realizing language barrier-free Korea

# 인 사 말 씀

현재 세계 각국은 “과학기술의 발전만이 21세기에 선진국으로 살아남을 수 있는 길이다.” 라는 명제 하에 국운을 걸고 과학기술에 대한 인적, 물적 투자를 증가시키고 있습니다. 특히 21세기에는 지능형 지식정보화 사회가 될 것이라는 전망에서 볼 때, SW, 빅데이터, 클라우드, HCI 기술을 망라한 정보기술은 기술적, 산업적, 사회적 측면 등 모든 분야에서 그 중요성이 점점 커지고 있습니다.

이러한 지능형 지식정보 기술 중 컴퓨터와 인간사이의 가장 자연스러운 의사소통 방법인 말을 이용한 휴먼 인터페이스 기술은 지난 30여 년간 세계 각국에서 치열한 경쟁 속에 활발히 연구되어 왔습니다. 최근에는 보다 진보된 개념인 서로 다른 언어 간의 의사소통을 가능하게 해주는 자동통역 기술이 새로운 화두로 대두되었으며, 미국, 일본 등 선진국들은 이미 이 분야에 막대한 투자를 시작한 바 있습니다. 늦었지만 다행히 우리나라도 2008년 휴대형 한·영 자동통역 기술 개발을 시작으로 2012년 한/영 자동통역 대국민 서비스 실시, 2013년 한·일, 한·중까지 통역 대상 언어를 확대, 2014년 인천아시안게임, 2015년 광주유니버시아드 자동통역서비스 실시 등 활발한 연구를 진행하고 있습니다. 그러나 국내의 짧은 자동통역기술 개발 역사로 선진기술에 대비 경쟁력을 갖기 위해 음성인식, 자동번역 등 핵심기술 확보, 경쟁력 있는 연구개발 환경 구축, 시장창출을 위한 산업생태계 조성, Open R&D를 통한 인력 양성 등 노력해야 할 분야가 산재해 있는 실정입니다.

이 보고서는 이러한 관점에서 자동통역 실현에 꼭 필요하나 자동통역 기술 개발에 미흡한 부분을 강화하고 향후 자동통역 분야의 성공적인 수행에 밑거름이 되게 하자는 목표로 수행되었습니다.

2016년 12월

한국전자통신연구원 원장 이 상 훈



# 제 출 문

본 연구보고서는 주요사업인 "언어장벽 없는 국가구현을 위한 자동통번역 산업경쟁력 강화 사업"의 수행결과로서, 본 사업에 참여한 아래의 연구실이 작성하였습니다.

2016년 12월

사업책임자 :	책임연구원 김상훈	(자동통역언어지능연구부)
연구참여자 :	책임연구원 박 준	(자동통역언어지능연구부)
	책임연구원 박상규	(자동통역언어지능연구부)
	책임연구원 이수중	(자동통역언어지능연구부)
	책임연구원 김정세	(자동통역언어지능연구부)
	책임연구원 김승희	(자동통역언어지능연구부)
	선임연구원 김동현	(자동통역언어지능연구부)
	선임연구원 윤 승	(자동통역언어지능연구부)
	선임연구원 최무열	(자동통역언어지능연구부)
	책임연구원 이영직	(자동통역언어지능연구부)
	책임연구원 김영길	(자동통역언어지능연구부)
	책임연구원 최승권	(자동통역언어지능연구부)
	책임연구원 김창현	(자동통역언어지능연구부)
	책임연구원 서영애	(자동통역언어지능연구부)
	책임연구원 황금하	(자동통역언어지능연구부)
	책임연구원 최미란	(자동통역언어지능연구부)
	책임연구원 권오욱	(자동통역언어지능연구부)
	연구원 이민규	(자동통역언어지능연구부)
	연구원 이담허	(자동통역언어지능연구부)



# 요 약 문

## I. 제 목

언어장벽 없는 국가구현을 위한 자동통번역 산업경쟁력 강화 사업

## II. 연구목적 및 중요성

본 사업은 대국민 자동통번역 서비스 실시, 자동통번역 산업생태계 구축 및 개방형 R&D 추진 등을 목표로 하고 있다. 대국민 자동통번역 서비스는 한, 중, 영, 일 4개 언어를 대상으로 통역서비스를 제공하여, 2014년 아시안게임 및 2015년 광주유니버시아드 경기대회에 활용되었으며, 여기에 프랑스어, 스페인어, 독일어, 러시아어 4개 언어를 추가하여 2018년 평창 동계올림픽 지원에도 활용될 예정이다. 특히 대국민 자동통역 서비스를 통해 축적한 사용자 로그데이터는 자동통역 원천기술의 개발에 직접 활용될 예정이고, 이를 통해 최근 구글, 애플 등 다국적 기업에 의한 SW중속으로부터 국내 산업을 보호하고, 자동통번역 산업 경쟁력을 강화하고자 한다. 또한 본 사업을 통해 고품질의 자동통역 플랫폼 및 Open API를 국내 업체, 연구소, 대학 등에 지원하여 산업생태계 구축도 추진하고 있다.

## III. 연구내용 및 범위

- 한/스, 한/불 자동통역 서비스 시스템 구현 (통역률 ERR 10% 개선)
- 핸드프리 양방향 자동통역 원천기술 개발 (2차)
- 다국어 음성인식 기술 개발(독일어, 러시아어)
- 자동통역 플랫폼 기반 Open API 개선 (4차)
- 중소기업 시제품 제작 지원
- 다국어 음성언어연구기반 조성
- 2016년 평창 올림픽 테스트 이벤트용 5개 언어 자동통역 구현

#### IV. 연구결과

- 세계최초 Zero-effort UI/UX 통역용 어어셋 개발 성공 (2017.8월 ISO 국제 표준 예정)
- 세계수준 8개 언어 다국어 음성인식 기술 확보 (국내유일 다국어 기술 보유)
- 세계최초 발음사전 없는 음성인식 기술 및 topic LM 디코더 개발
- 실시간 방송음성 정제를 통한 대용량 학습용 DB 자동확보 기술 구현에 성공 (한국어, 독일어, 불어, 스페인어에 적용)
- 한컴인터프리 (ETRI연구소 기업 및 한컴그룹 자회사) ‘말랑말랑 지니톡’ 30만 앱 다운로드 기록 및 MWC2016 출품 (한컴은 단말통역기, 통역박스 등 다양한 제품으로 사업 확대 추진 중)
- 평창동계올림픽 테스트 이벤트에 5개 언어 자동통역 시범서비스 적용 (2016.11~2017.3)

#### V. 기대성과 및 건의

- ‘12 여수 세계엑스포, ‘14 인천 아시안게임, ‘15 광주 유니버시아드, ‘18 평창 동계올림픽에 자동통번역서비스 실시를 통해 언어장벽 없는 국가 실현 및 국격 제고
- ETRI의 자동통번역 플랫폼 및 Open API를 적극 지원하여 자동통번역 기반 새로운 BM 창조기회가 확대되고 이에 따라 대학, 중소기업, 응용SW 개발자, 수요업체(단말기, 포털사, 주력산업 제조업체 등)간의 산업생태계 조성
- ETRI와의 기술협력 및 향후 다국어 처리 기술지원을 통해 향후 3~4년 내 국내 SW기업의 기술력이 글로벌 경쟁력을 갖추게 될 것으로 기대



# ABSTRACT

## I . TITLE

Strengthening competitiveness of automatic translation industry for realizing language barrier-free Korea

## II . THE OBJECTIVES

This project aims at providing automatic speech translation public service, building up the eco-system of automatic speech/text translation industry, and constructing the open R&D environment. The automatic speech translation service will include Korean, English, Chinese, and Japanese, Spanish, French, German and Russian till 2018, and support the 2018 Pyeongchang Winter Olympics. We are utilizing the user log-data from the public service in improving the automatic translation technology, and thus try to strengthen the industry competitiveness against the global enterprises like Google. We will provide the automatic translation technologies along with its base platform and open API to the domestic industries, research institutes, and universities for industrial eco-system construction.

## III. THE CONTENTS AND SCOPE OF THE STUDY

- Providing Korean-Spanish/French automatic speech translation trial public service to public.
- Developing the fundamental technologies for hands-free, bidirectional spontaneous speech translation (2nd phase).
- Multilingual speech recognition technology development (German, Russian)

- Implementation of the automatic speech translation platform and Open API (4th phase).
- Implementation of the automatic speech translation platform and Open API.
- Set-up of the multilingual spoken language research infrastructure.
- Development of speech translation system (supporting 5 languages) for 2016 PyeongChang Olympics test event.

#### IV. RESULTS

- Development of the new ear-set for the zero-effort UI/UX speech translation for the first time (to be fixed as ISO International Standard in August 2017.8.)
- The state-of-the-art multilingual (8 languages) speech recognition technology
- Development of the world's first speech recognition system without pronunciation dictionary and the topic LM-based decoder
- automatic collection technology of the large-scale training DB by refining the real-time broadcast speech signal(Korean, German, French, and Spanish).
- Hancom-interfree, the Hancom-ETRI joint investment company, developed the MalangMalang GenieTalk system which recorded 300,000 downloads and was displayed at MWC2016.
- Applying the speech translation system (supporting 5 languages) to the 2016 PyeongChang Olympics test event. (2016.11~2017.3)

## V. EXPECTED RESULT & PROPOSITION

- Series of the automatic speech translation services at the 2012 Yeosu Expo, the 2014 Incheon Asian Games and the 2018 Pyeongchang Winter Olympics will make the language barrier-free nation and enhance the dignity of our homeland, Korea.
- Service platform and its Open API will encourage the new BM developmen, which will again expedite the advent of the industrial eco-system
- Multi-lingual translation technology cooperation between ETRI and the associated industries will make contribution to intensify the global competitiveness of the automatic translation industry



# CONTENTS

CHAPTER 1 Importance and Problems	23
SECTION 1 Importance of the project	23
SECTION 2 Restrictions (or Challenge) of the project	23
SECTION 3 Expected effects	24
CHAPTER 2 Status and Approach	26
SECTION 1 Status of at home and abroad	26
SECTION 2 Key components and approach	32
SECTION 3 Innovativeness and creativity	34
CHAPTER 3 Overview	35
SECTION 1 Definition of the project	35
SECTION 2 Concept of the technology	35
SECTION 3 road-map of technology development	36
CHAPTER 4 Targets and Scope of the project	37
SECTION 1 Final target	37
SECTION 2 Targets and scope per each year	37
SECTION 3 Annual plan	38
SECTION 4 Major performance objective and plan	41
CHAPTER 5 Detailed Outcomes	44
SECTION 1 Qualitative outcomes	44
SECTION 2 Quantitative outcomes	79
CHAPTER 6 Utilization Plan	81
SECTION 1 Technical evaluation	81
SECTION 2 Possibility of commercialization	81
SECTION 3 Market and competition	83
SECTION 4 Transferable technologies	84

SECTION 5 Commercialization plan for applications	84
CHAPTER 7 Implementation System and Method	85
SECTION 1 Implementation system	85
SECTION 2 Implementation method	85
CHAPTER 8 Expected Effects	86
SECTION 1 Technical aspects	86
SECTION 2 Economic industrial aspects	86
CHAPTER 9 Future Work	88
[Appendix 1] Evaluation method for translation success rate	89
[Appendix 2] Evaluation measure for speech recognition	91
[Appendix 3] Acronym table	92

# Table List

Table 1 Competitiveness of automatic translation technology	29
Table 2 Detailed Outcomes	41
Table 3 Qualitative Outcomes	42
Table 4 Industrialization Performance	43
Table 5 French Punctuation Restoration Performance	46
Table 6 Performance evaluation of multilingual punctuation restoration technology	46
Table 7 Performance evaluation result of Korean/Spanish, Korean/French similar sentence retrieval technology	48
Table 8 Comparison of performance before and after automatic refinement of broadcasting voice DB (WA%)	50
Table 9 SR Performance comparison using Spontaneous Broadcast Voice DB	51
Table 10 Spanish voice data status	57
Table 11 Spanish Speech Recognition Results	57
Table 12 French voice data status	58
Table 13 French Speech Recognition Results	58
Table 14 Improvement of English speech recognition performance	59
Table 15 Phone speech recognition vs. grapheme speech recognition	60
Table 16 The corpus used in the experiment	62
Table 17 Word2Vec training example (Blue dot: input, orange dot: output)	63
Table 18 Dual embedding comparison for 'Wedding'	64
Table 19 Dual embedding comparison for 'attitude'	65
Table 20 Support for automatic translation BM and prototype development for small and medium-sized enterprises	67
Table 21 ETRI-University Cooperation Status	68
Table 22 Multilingual caption data collection status	69
Table 23 List of paper publication	79
Table 24 List of patents	80
Table 25 List of technology transfer	80

# Figure List

Figure 1 Automatic translation service road map (2012~2018)	25
Figure 2 Speechalator	27
Figure 3 VoiceTra developed by NICT	28
Figure 4 Microsoft-Skype translator	29
Figure 5 Speech translation research trends	32
Figure 6 How to construct speech translation industrial ecology	33
Figure 7 Automatic speech translation system	35
Figure 8 Plan for speech translation service from Jeju to Pyeong-Chang	36
Figure 9 Final project target	37
Figure 10 Targets and scope per each year	38
Figure 11 Automatic speech translation system structure	44
Figure 12 Flow diagram of the sentence mark recovery based on deep learning	45
Figure 13 Screen display of Korean-Spanish, Korean-French automatic speech translation	47
Figure 14 Similar sentence detection block structure	48
Figure 15 Concept diagram of the broadcast media data automatic extraction	49
Figure 16 Automatic construction of speech DB	50
Figure 17 Convenience-enhanced hands-free automatic speech translation UI compared with the conventional UI	52
Figure 18 2 channel speech signal processing on hands-free ear-set for automatic speech translation	53
Figure 19 Hands-free ear-set prototype for automatic speech translation	54
Figure 20 Trial performance of the hands-free bidirectional automatic speech translation	55
Figure 21 crowd sourcing platform app live screen	56
Figure 22 Word2Vec base model	61
Figure 23 Corpus features	61
Figure 24 CBOW(continuous bag of words) and SKIP-gram model structure	62
Figure 25 Speech recognition performance comparison	66
Figure 26 Concept diagram of GenieTalk-based industry ecology construction	69
Figure 27 Domain-tailored Korean-Spanish/French Scenarios	71

Figure 28 Trial service of PyeongChang Window Olympics Test Event	72
Figure 29 Automatic speech translation trial service flow	73
Figure 30 Automatic speech translation trial service strucure	74
Figure 31 Overall flow diagram of automatic speech translation trial service	75
Figure 32 Screen-shot of the log-data monitoring system	75
Figure 33 On-site demonstration at MWC 2016	77
Figure 34 Actual use scene in Spain	78
Figure 35 Overall R&D performance system	85

# 목 차

1. 중요성 및 문제점	23
가. 연구개발과제의 중요성	23
나. 연구개발과제 수행의 제약요인	23
다. 연구개발과제 수행결과 기대효과	24
2. 현황 및 접근방법	26
가. 국내·외 현황	26
나. 핵심요소 및 접근방법	32
다. 혁신성과 독창성	34
3. 사업 개요	35
가. 사업 정의	35
나. 기술 개념	35
다. 기술개발 로드맵	36
4. 사업 목표	37
가. 최종 연구목표	37
나. 단계별 연구목표	38
다. 연차별 수행내용	38
라. 주요 성과목표 및 계획	41
5. 세부 추진실적	44
가. 정성적 추진실적	44
나. 정량적 추진실적	79
6. 활용(산업화) 방안	81
가. 기술평가	81
나. 활용(상용화) 가능성	81

다. 시장 및 경쟁	83
라. 이전가능한 기술목록	84
마. 자동통역 응용서비스의 상용화 계획	84
7. 추진체계 및 전략	85
가. 추진체계	85
나. 추진전략	85
8. 기대성과	86
가. 기술적 측면	86
나. 경제 산업적 측면	86
9. 향후계획	88
10. 부록	89
가. 번역이해도/통역성공률 평가방안	89
나. 음성인식률 산출 방법	91
다. 약어표	92

## 표 목차

표 1 자동통번역 국내외 기술 수준	29
표 2 성과목표달성실적	41
표 3 정량적 실적	42
표 4 산업화 실적	43
표 5 프랑스어 문장부호 복원 성능	46
표 6 다국어 문장부호 복원 기술의 성능 평가 결과	46
표 7 한국어/스페인어, 한/불어 유사문장 검색 기술의 성능평가 결과	48
표 8 방송음성 DB의 자동 정제 전/후 성능 비교 (WA%)	50
표 9 자유발화 방송음성 DB를 이용한 음성인식 성능비교 결과	51
표 10 스페인어 음성데이터 현황	57
표 11 스페인어 음성인식 평가결과	57
표 12 프랑스어 음성데이터 현황	58
표 13 프랑스어 음성인식 평가결과	58
표 14 영어 음성인식 성능 개선	59
표 15 Phone 음성인식 vs. grapheme 음성인식	60
표 16 실험에 사용된 말뭉치	62
표 17 Word2Vec 학습 예 (파란색 점: 입력벡터, 주황색 점: 출력벡터)	63
표 18 ‘결혼’ 의 dual embedding 비교	64
표 19 ‘태도’ 의 dual embedding 비교	65
표 20 중소기업 대상 자동통역 BM발굴 및 시제품 제작 지원 내용	67
표 21 ETRI-대학 협력 현황	68
표 22 다국어 자막 데이터 수집 현황	69
표 23 논문리스트	79
표 24 특허리스트	80
표 25 기술이전리스트	80

## 그림 목차

그림 1 자동 통번역 對국민 서비스 로드맵 (2012~2018)	25
그림 2 Speechalator	27
그림 3 일본 NICT에서 개발한 VoiceTra	28
그림 4 MS-스카이프 트랜스레이터	29
그림 5 국내외 자동통역 연구동향 및 수준	32
그림 6 자동 통번역 산업생태계 구축방안	33
그림 7 자동통역시스템 구성도	35
그림 8 제주~평창까지 자동통역 서비스 실시 계획	36
그림 9 최종 연구목표	37
그림 10 단계별 연구목표	38
그림 11 자동통역시스템 구조	44
그림 12 딥러닝 기반 문장부호복원 구조도	45
그림 13 한/스, 한/불 자동통역 화면	47
그림 14 유사문장 검색기 구조	48
그림 15 방송 미디어데이터 기반 자동 DB구축 개념도	49
그림 16 음성DB구축 자동화 방법	50
그림 17 기존 스마트폰 통역대비 제안된 핸드프리 통역의 사용자편의성 향상	52
그림 18 핸드프리 통역용 이어셋 2채널 음성신호 처리 구조	53
그림 19 핸드프리 통역용 이어셋 샘플	54
그림 20 실제 핸드프리 양방향 자동통역 시연 장면	55
그림 21 크라우드 소싱 플랫폼 앱 실행 화면	56
그림 22 Word2Vec 기본 모델	61
그림 23 말뭉치 특성	61
그림 24 CBOW(continuous bag of words) 모델과 SKIP-gram 모델의 네트워크 구조	62

그림 25 음성인식 성능 비교	66
그림 26 지니톡 기반 산업생태계 구축 개념도	69
그림 27 한-스페인어, 한-프랑스어 영역특화 시나리오	71
그림 28 평창 동계올림픽 테스트 이벤트 자동통역 시범서비스 모습	72
그림 29 자동통역 시범 서비스 흐름도	73
그림 30 자동통역 시범서비스 물리적 구성도	74
그림 31 자동통역 시범서비스 전체 적용 흐름도	75
그림 32 로그 데이터 모니터링 시스템 스크린샷	75
그림 33 MWC 2016 시연 모습	77
그림 34 스페인 현지에서의 실제 사용 장면	78
그림 35 추진체계	85



## 1. 중요성 및 문제점

### 가. 연구개발과제의 중요성

- 세계화의 가속화로 국가 간 인적, 물적 교류가 활발해지면서 언어 간 장벽을 허무는 자동통번역 기술의 확보가 국가 글로벌 경쟁력과 직결됨
- 최근 Google, Apple 등 다국적 기업의 모바일용 OS 플랫폼 개발 사례와 같이 다국어 음성언어정보 기술 또한 외국에 종속될 위기상황에서 국가적 차원의 지속적 투자 필요
- 2012년 Google은 다국어 자동통역서비스 추진 발표(2010.2, 타임지). IBM은 5년 내 상용화 기술 중 자동통역을 파급효과가 가장 큰 기술로 선정 (2007년)
- 2014년 일본정부는 2020년 동경올림픽에서 일본 IT의 부활을 천명하고, 선두 기술로 자동통역서비스를 실시하기로 하고, NTT Docomo, NICT 등이 조인트 벤처기업을 설립, 매년 500억 원의 개발비를 투자하기로 함. 2020년 일본내 자동통번역 시장 10조원 예측 (출처: 일본 UFJ종합연구소, 2006)
- IT 분야는 물론 교육, 관광, 문화, 의료 등 타 산업으로의 파급이 큰 기술

### 나. 연구개발과제 수행의 제약요인

- 단기과제, 단위기술 개발 위주
  - 단기목표 중심 기술개발 후 기술이전으로 사업이 종료함에 따라 지속가능한 음성언어 기술개발이 어려움
  - 최근 SW에 공격적으로 투자를 하고 있는 Google, Apple, IBM 등 선진외국의 다국적 기업이 OS, DBMS를 독점하는 사례와 같이 음성언어처리 기술 또한 종속될 위기 상황에서 국가적인 차원의 장기적인 투자가 필요
- 사용자 데이터 수집을 통한 선순환 기술향상 체계 부재
  - 현 R&D 체계로는 실 서비스 환경 사용자 데이터 수집 불가능. 이에 따라 실 서비스를 바탕으로 자동통번역의 성능을 개선할 수 있는 조직 체계로의 전환이 시급. 또한 사용자 데이터 DB 정제 및 플랫폼 서비스를 위한 상시 인력 필요

- Google은 대용량 다국어 사용자 로그 DB 확보 및 엔진반영의 선순환 구조를 기반으로 다국어 음성인식 및 자동통번역 기술의 강자로 부상하고 있음
- 한국시장에 국한된 기술 개발 위주
- 한국어 위주의 기술 개발로 음성언어 SW의 글로벌 산업경쟁력 확보 어려움. 아시아권 언어, 유럽어, 아랍어 등으로 확대하기 위해 장기간 다국어를 지원하는 상시의 학제간 협력 및 인력 육성 필요
  - 국내 글로벌 업체(현대자동차, 삼성전자 등)에서는 다국어 음성인식 기술이 필요한 상황이지만, 다국어를 지원하는 뉘앙스 등 외산에 종속되어 있음
- 국민 삶의 질 향상이란 국가적 미션 수행 필요
- 국가 글로벌 SW 경쟁력 강화 및 ‘12년 여수 엑스포, ‘14 인천 아시안게임, ‘18 평창 동계올림픽에 자동통번역 대국민 서비스를 통한 국격 제고 필요
  - 급속한 개방과 세계화로 외국어 정보와 지식이 경쟁력이 되는 사회에서 외국어 소외 계층에게 정보접근 통로를 제공하여 외국어 격차(Language divide) 해소

#### 다. 연구개발과제 수행결과 기대효과

- ‘12 여수 엑스포, ‘14 인천 아시안게임, ‘18 평창 동계올림픽 등의 국제행사 대상 자동통번역 서비스 지원을 통한 국격 제고
- 대국민 자동통번역 서비스 실시를 통해 언어장벽 없는 국가 실현 및 국격 제고



그림 1 자동 통번역 對국민 서비스 로드맵 (2012~2018)

- 자동 통번역 산업생태계 조성 및 신규 서비스 창출
  - ETRI의 통번역 SW 플랫폼을 중심으로 응용 SW 개발자, 주요업체(단말기, 포털사, 주력산업 제조업체 등)간의 산업생태계 조성
  - 산/학/연 공동연구 활성화를 통해 ETRI 기술 인프라 공동 활용이 가능하므로 인프라가 부족한 기업의 자생력 강화 및 산업 활성화 가능
  - 자동통번역 산업뿐만 아니라, 휴대폰, 자동차 등 제조업 및 교육, 관광, 문화, 의료 등 서비스 분야와 융합한 新서비스 창출·발전
  
- 국내 SW 경쟁력 제고
  - 공공재적 성격의 융합기술 경쟁력을 확보하고 있는 ETRI와의 기술협력 및 지원을 통해 국내 SW기업의 기술경쟁력 제고
  - 출연(연)이 상대적으로 취약한 기술사업화 부문에서 출연(연) R&D성과 확산 → 국내기업 R&D지원 강화 → 자동통번역 산업생태계 구축 등 상생협력 선순환 고리 정착

- 사교육비 경감 및 창의인재 양성
  - 외국어 사교육비 부담(10조원) 대폭 경감
  - 다학제간 공동 연구 등을 통한 창의 idea 확산 및 창의인재 양성

## 2. 현황 및 접근방법

### 가. 국내·외 현황

#### 1) 국외 기술현황

- 미국방위고등연구계획국 (DARPA: Defense Advanced Research Projects Agency)는 IBM, SRI (Stanford Research Institute), BBN에 대한 지원을 통하여 GALE(Global Autonomous Language Exploitation) 프로젝트를 2006년부터 진행해 오고 있음. 영어, 중국어, 아랍어가 주 대상이며, 실제 사용할 수 있는 실용화 기술개발을 목표로 다양한 언어로 쓰인 대량의 자료를 빠른 시간 내에 통역하는 것을 목적으로 함
- IBM은 5년 내 상용화 가능 기술로서 파급효과가 큰 자동통역을 최우선 순위로 선정. DARPA의 지원 하에 MASTOR (Multilingual Automatic Speech-to-Speech Translator)라는 영/중 자동통역기를 개발. 영어-중국어 양방향 통역이 가능하고 약 3만개의 단어 인식. 여행, 긴급 의료 진단, 군의 자기 방어, 보안 상황을 대상으로 하고 있으며 노트북이나 PDA에서도 사용이 가능함. 현재 이라크에 파견된 미군을 대상으로 시험서비스 중동연구소에서 개발한 TALES (Translingual Automatic Language Exploitation System)라는 번역 엔진을 사용
- 미국 카네기멜론대학교(CMU)에서는 2003년에 음성합성 전문기업인 셉스트럴(Cepstral), 멀티모달 테크놀로지(Multimodal Technologies), 모바일 테크놀로지(Mobile Technologies) 등과 함께 양방향 자동통역시스템인 Speechalator를 개발. PDA에서 작동하고, 의료 정보를 영어와 아랍어로 상호 번역하며 번역 정확도는 약 80%임. 번역기는 두 언어의 매개체로 중간언어(Interlingua)를 이용함. CMU대학에서 창업하여 Facebook에 인수된 미국 지비고(Jibbig)사는 스마



그림 2 Speechalator

트폰 탑재형 영어 기반 일어, 중국어, 스페인어, 아랍어, 양방향 통역기 등 다국어 통역기 개발.

- Google은 2012년 14개 다국어를 대상으로 원격 서버기반 자동통역서비스 실시 계획을 천명하고, 현재 ‘구글 트랜스레이터’로 본격적인 자동통역서비스를 제공. 구글은 200억 단어 이상 규모의 병렬 말뭉치를 이용한 통계 기반의 자동번역 기술을 개발하고 있으며 현존하는 자동통역기 중 가장 많은 언어인 90개 이상 언어 간 번역을 지원
- 미국 SRI (Stanford Research Institute)는 DARPA의 지원 하에 IraqComm이라는 자동통역기를 개발. 영어 및 아랍어에 대해 수만 단어를 처리하며, 자기방어, 보안, 기초 의료 서비스 상황을 통역 대상으로 함. 2006년 미군에 의해 채택
- EU는 의회에서 행해지는 모든 연설을 포함하여, 11개 공식 언어로 문서화하는 작업에 연간 5.5억 유로를 지출하고 있으며 회원국 간의 교류에 있어서 언어장벽이 가장 심각한 문제로 대두됨. 이에 따라, EU 제6차 연구지원사업 프레임워크에서는 의회 연설문의 자동전사 및 통역/번역기술 개발을 목표로 하는 TC-STAR (Technology and Corpora for Speech to Speech Translation) 프로젝트 (2004.4 ~ 2007.3)를 지원. EU 의회에서 통역/번역 서비스의 가능성에 대한 시험 차원에서 평가를 수행함. IBM, IRST, LIMSI, UKA, UPC, RWTH 등에서 참여
- 독일 국책연구소 DFKI는 1991년부터 대화체 자동통역 시스템의 개발을 위해 휴대용 독일/독일 자동통역 시스템 VerbMobil의 개발을 주도하여 비즈니스 상담을 보조하는 구문단위 통역기술을 개발

- 일본 NEC는 2002년에 PDA에서 동작하는 일/영 여행자용 자동통역 시스템을 개발하였으며, 한국어를 포함하여 총 11개 국어를 지원하는 다국어 자동번역 기술을 개발. Bestiland라는 번역 포털 서비스를 실시함. 또, 일/중 통역 시스템도 추가 개발하여 2005년부터 나리타공항에서 시범 서비스를 실시, 이 중 일영 단방향 통역시스템은 2007년 12월 휴대폰 단말기에서 실시간으로 동작함
- 일본 NICT 연구소는 1986년부터 자동통역 기술의 실제 실용화를 목표로 연간 200억 원의 연구비를 투입하여, 일어-영어, 일어-중국어간 양방향 대화체 자동통역 시스템을 개발하고 있음. 2007년도 말부터 휴대폰 단말기에 네트워크 기반으로 일/영 통역 서비스를 제공하고 있음
- 일본 NTT, KDD, NHK, 히타치 등은 산학연간 기술 교류의 활성화를 통해 여행안내, 강의/연설, 비즈니스 회의 등 다양한 분야에서의 실용화를 시도
- 2014년 일본정부는 2020년 동경올림픽에서 일본 IT의 부활을 천명. 선두기술로 자동통역서비스를 올림픽에서 본격 실시하기로 하고, NTT Docomo, NICT 등이 조인트 벤처기업을 설립, 매년 500억 원의 개발비를 투자하기로 함
- 보이스트라(VoiceTra)는 일본 정부 주도로 “언어의 벽을 뛰어넘는 음성 커뮤니케이션 기술의 실현” 프로젝트의 일환으로 개발되었으며 2010년 6월에 최초



그림 3 일본 NICT에서 개발한 VoiceTra

발표. 음성인식의 경우, 일본어, 영어, 중국어, 베트남어, 인도네시아어, 말레이시아어를 지원하고, 자동번역의 경우 이들 언어를 포함하여 대만어, 독일어, 프랑스어, 덴마크어, 네덜란드어, 이탈리아어, 스페인어, 포르투갈어, 포르투갈어, 러시아어, 아랍어, 힌디어, 태국어, 타갈로그어, 한국어 등의 21가지 언어를 지원

하는 등 본격적인 다국어 자동통역기로서의 면모를 보임. 자체적으로 밝힌 바에 따르면 번역 성능은 토익 600점급에 해당한다고 함

- 2014년 5월 마이크로소프트(MS)는 화상통화서비스인 스카이프 통화내용을 통역해주는 '스카이프 트랜스레이터(Skype Translator)' 서비스를 공개. 통화 내용을 실시간 통역해 음성과 자막으로 제공하는 통번역 채팅서비스로, 공개 시연에서는 영어와 독일어 사용자가 대화를 나누어도 전혀 어려움이 없도록 실시간으로 언어소통이 됨. 현재 약 40 가지 언어가 텍스트 형태로 자동번역이 됨

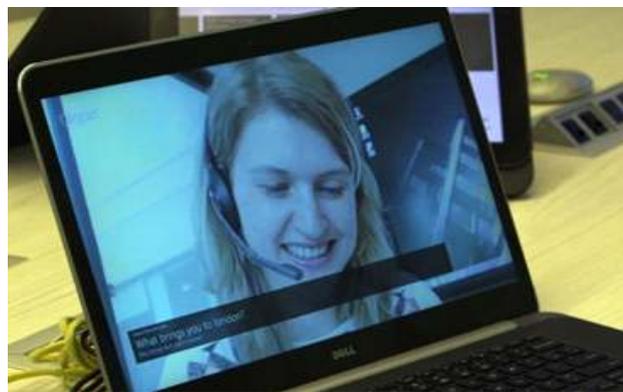


그림 4 MS-스카이프 트랜스레이터

## 2) 국내 기술현황

- 국내 핵심기술 연구는 ETRI 등 국책연구기관을 통하여 이루어지고 있으며, 한국어, 영어, 일본어, 중국어에 대해서는 경쟁력을 유지하나 다국어는 열세임

비교항목	국 내	국 외	사용자 요구수준
상용제품	ETRI 지니톡	미국 Google, 일본 NICT	-
통역 성공률	80%-85%	75%-80%	90%
통역 가능한 언어 수	3개 언어 (한, 영, 일, 중)	15개 언어 (한, 영, 일, 중, 스페인, 독일어, 프랑스 등)	20개 언어
통역 대상 어휘 수	30만 어휘	100만 어휘 이내	100만 어휘 이상
통역 대상 범위	일상, 여행 분야	무제한 분야	무제한 분야

표 1 자동통번역 국내외 기술 수준

- 1999년 한국전자통신연구원은 국제간 자동통역 공동협력연구협의체인 C-STAR (Consortium for Speech Translation Advanced Research)의 핵심회원 기관으로 가입하여 한/영/일/프랑스 4개국 간 실시간 음성언어번역 국제시연을 실시하였음. 고객이 여행사 직원과 여행계획을 상담하는 대화를 대상으로 자동통역 기술의 실현 가능성을 입증하였음. 2012년 한국어와 영어간 여행분야에서 언어소통이 가능한 자동통역 앱 ‘지니톡(GenieTalk)’을 개발함. 2013년에 한국어와 일본어, 한국어와 중국어로 언어를 확장해 현재 총 4개 언어에 대해 자동통역이 가능함
- 2001년 삼성과 히다찌 연구소가 한/일 월드컵 국제행사를 통해 1000 여개의 고정 문장에 대하여 휴대전화로 이용할 수 있는 한/일 자동통역 시범서비스 실시한 바 있음
- 2002년 한/일 번역 업체를 중심으로 PDA 탑재 번역 제품(창신소프트의 「이지토키2002」, 시스메타의 「포켓트랜스위즈」등)이 출시되었으나, 실용적인 면에서 초보적인 수준이며, 영어 및 중국어 등에 대한 다국어 번역 서비스는 제공되고 있지 않음
- 삼성은 2007년 휴대전화에 기본 예문 검색 기능을 탑재함. 12개의 세부 영역 중 하나를 설정하고 버튼 또는 음성으로 예문을 선택하면 대응되는 외국어 문장을 음성으로 들려줌. 미리 정해진 수백 개의 문장 외에는 처리하지 못하여 활용이 제한적임

### 3) 국내외 표준화 현황(또는 향후 기술 발전 추세)

- 2007년 3월 APT ASTAP 산하 표준 전문가그룹인 A-STAR 주관 SNLPEG (Speech and Natural Language Processing Expert Group)결성. 자동통역 모듈 연동 규격 표준 작성
- (국내) ETRI, 삼성전자, LG 등의 순으로 자동통역 기술에 대한 특허를 보유. 한/미/일/유럽을 포함한 주요국에 대비한 출원규모는 약 16.7%를 차지. 주요 출원분야는 통역서비스 향상 방법, 고속 연속음성인식, 통계기반 번역, 합성음의 발화속도 변환, 인식오류 정정방법 등이며 요소기술을 중심으로 분포

- (국외) AT&T, IBM, Microsoft, Nuance, NEC, NICT, NTT, Hitachi, Sharp 등을 중심으로 자동통역 및 관련 요소기술에 관한 특허를 다수 보유. Sharp가 109건으로 가장 많고, IBM 85건, Microsoft 74건, NEC 64건, NICT 50건, AT&T 48건 등의 순위로 분포. Fluential, VoxTec, BBN, SRI는 핵심특허를 보유하여 특허권을 행사하는 경우도 있음

#### 4) 동일, 유사내용에 대하여 국내·외 관련자들의 수행내용

- 미국 IBM은 현재 세계최고 수준의 기술력을 보유하고 있으며, 2012년 상용화를 목표로 2007년부터 본격적으로 실용화 개발에 착수하였음. 여행, 의료, 군사 등 영역에서 영어/중국어, 영어/아랍어간 양방향 자동통역 기술을 개발하고 있으며, 영어/아랍어 자동통역기는 이라크에 파견된 미국을 대상으로 시범서비스를 실시하고 있음
- 미국 뉘앙스는 세계 최대의 음성솔루션 업체이고 세계 주요 언어에 대한 음성인식 기술을 보유하고 있어 향후 중요 경쟁자가 될 것임. 2008년 말 애플 iPhone에 1,500 문장 수준의 단방향 일영, 일중 통역 소프트웨어를 출시. 아직은 정해진 문장만을 인식하는 수준임
- 구글은 주요 언어간의 번역 기술을 보유하고 있고 2008년 휴대전화와 연동한 서버 방식의 연속 음성인식 기술을 선보임. 통역분야 향후 주요한 경쟁 상대임
- 일본 국책연구소인 NICT는 1986년부터 자동통역 기술의 실상용화를 위하여 연간 200억원 규모의 연구비를 투입하여 일본어와 영어, 중국어, 독일어간 양방향 자동통역 기술을 개발하고 있음. 2008년부터 네트워크기반의 일/영, 일/중 통역 시범서비스를 실시하고 있음. 일본어와 한국어의 언어적 특성이 유사하므로 한국어 관련 기술 개발에 착수하는 경우 강력한 경쟁상대가 될 것임

경쟁국가	기관명	연구동향	Google 대비 경쟁력 (%)			
			지원언어 수준	음성인식 기술수준	자동번역 기술수준	사용자 편의성
	ETRI	• 2012년 <b>지니독 대국민 시범서비스 실시</b> . 현재 한,중,영,일 4개국 통역서비스 지원 및 스페인어, 불어 등 총 8개국 자동통역서비스 개발 계획	4개국	100	110	110
	NTT	• 일본 총무성, NICT와 NTT Docomo, Systran, FueTrek이 조인트 벤처 설립 <b>2015년부터 2년간 연 500억원 투입</b> 하여 통번역 기술 개발 추진	21개국	85	90	80
	Google	• Google, 2012년 <b>14개 국 다국어</b> 를 대상으로 <b>'Google translator' 제공</b> • Google의 최대 강점인 사용자 기반 빅데이터와 세계 최대 컴퓨팅 파워를 결합하여 더욱더 진보된 기술이 나올 것으로 예상	14개국	100	100	100
	MS	• MS- Skype는 2014년 5월 화상통화서비스인 스카이프 통화내용을 통역해주는 <b>'Skype Translator' 공개</b>	2개국	95	95	90
	IBM	• IBM Watson연구소가 개발중인 MASTOR 자동통역시스템은 영어-아랍어간 통역. <b>이라크파병군민에게 보급</b> 하여 현지인과 통역에 사용	4개국	90	90	90
	Lexifone	• 이스라엘 start-up 업체 Lexifone, 2012년 <b>양방향 실시간 전화 동시통역 상용서비스 실시</b> . 현재 세계 주요언어 8개국에 대한 대화체 통역 서비스 제공	10개국	80	80	100

그림 5 국내외 자동통역 연구동향 및 수준

## 5) 동일, 유사내용과 관련하여 제안자가 이미 수행한 사업 또는 연구개발과제

- 지식경제부 산업원천기술개발 사업의 일환으로 “휴대형 한/영 동시통역 기술 개발(2008~2011)” 사업 수행. 본 사업을 통해 스마트폰 기반 한/영 자동통역 핵심기술 확보
- 지식경제부 산업원천기술개발 사업의 일환으로 “글로벌 소통을 위한 한/영, 한/일 동시통역 응용 SW 개발(2011~2013)” 과제에 참여기관으로 수행
- 미래창조부 산업융합원천기술 개발 사업의 일환으로 “지식학습기반의 다국어 확장이 용이한 관광/국제행사 통역률 90%급 자동통번역 소프트웨어 원천 기술 개발 (2012~2016)” 에 참여

## 나. 핵심요소 및 접근방법

- 對국민 자동 통번역서비스 제공
  - 2012년 여수 세계엑스포 영역특화를 통한 3개 언어(한/영, 한/일) 자동통역 대국민 서비스 실시

- 2014년 인천 아시안게임용 자동통역 특화를 통한 4개 언어 (한/영, 한/중, 한/일) 자동통역 대국민 서비스 실시
- 2018년 평창 동계올림픽용 자동통역 특화를 통한 7개 언어(한/영, 한/중, 한/일, 한/스페인, 한/프랑스, 한/독일어, 한/러시아어) 자동통역 대국민 서비스 실시 계획
- 일상영역 대국민 자동통번역 시범서비스 구축 및 운용

○ 자동 통번역 산업생태계 구축

- 자동통번역 플랫폼 지원: Open API를 이용하여 자동통번역 서비스 적용 및 생태계 창출을 위한 플랫폼 지원
- 언어자원 구축 및 배포: 시범서비스를 통해 축적한 로그 DB를 가공·정제하여 기술개발에 활용할 수 있도록 업체 대상 DB 배포
- 자동통번역 Open API(Application Program Interface) 및 DB 표준화: 국내업체·연구소·대학 등의 공통 활용을 위해 API, DB 표준화 추진
- 중소기업 지원 및 공동개발: 자동통번역 산업진흥을 위한 중소기업 지원 및 국내외 업체와 협업을 통한 新비즈니스 창출



그림 6 자동 통번역 산업생태계 구축방안

- 국내외 개방형 연구 및 인력양성 추진
  - 인문, 공학 등 학제간 융합연구를 통한 창의 연구결과 도출
  - 다국어기술 개발을 위한 국제 공동연구 지원
  - 기업체, 대학 등 센터로 인력 파견을 통해 공동연구 추진 및 창의인재 육성

#### 다. 혁신성과 독창성

- 국내외 업체에 자동통번역 공통플랫폼 및 Open API를 지원하여 실서비스 환경 대용량 로그데이터 축적을 가능하게 해줌으로서 지속적 성능 개선이 가능한 혁신적인 방법을 추진하고자 함
- 중소기업과 공동으로 자동통번역 BM 개발 및 실용화를 추진함으로써 초기 자동통번역 기술 활용에 따른 위험부담을 줄이며 이에 따라 국내에 자동통번역 산업생태계가 조성되는데 자동통번역센터가 큰 역할을 할 것으로 보임
- 국내에서는 한국어, 영어, 중국어, 일본어 등 아시아권 주요언어에 치우쳐 기술 개발이 진행되는 한계가 있음. 이에 글로벌 시장에 진출하기 위한 유럽어, 동남아어, 중동어 등 다국어 음성언어 기반 기술을 개발함으로써 구글, 애플(뉘앙스) 등과 같이 국제적 수준의 기술확보가 가능함

### 3. 사업 개요

#### 가. 사업 정의

대국민 자동통역 시범서비스 실시를 통해 자동통번역 산업생태계를 구축하고, 2018년 평창 동계올림픽 8개 언어 자동통역 공식 서비스 실시를 통해, 세계최고 수준 자동통번역 기술 확보 및 글로벌 시장 진출

#### 나. 기술 개념

“외국인과 모국어로 언어소통을 가능하게 하는 기술”



그림 7 자동통역시스템 구성도

다. 기술개발 로드맵



그림 8 제주~평창까지 자동통역 서비스 실시 계획

## 4. 사업 목표

### 가. 최종 연구목표



그림 9 최종 연구목표

나. 단계별 연구목표



그림 10 단계별 연구목표

다. 연차별 수행내용

구분	목표	내용
1차년도 (2012)	<ul style="list-style-type: none"> <li>○ 자동통번역 산업기반 조성 (1차)</li> <li>○ '12년 여수엑스포 대국민 자동통번역 서비스 실시</li> </ul>	<ul style="list-style-type: none"> <li>○ 자동통번역 센터 설립</li> <li>○ 자동통번역 시범서비스 인프라 구축</li> <li>○ 자동통번역 플랫폼, Open API 설계 및 초기 구축</li> <li>○ 사용자 로그데이터 구축 환경 조성 및 도구설계</li> <li>○ 중소기업 BM 발굴 및 응용개발 지원 (1차)</li> <li>○ 사용자 로그데이터 기반 한/영 자동통번역 성능 개선 (통역률 ERR(Error Reduction Rate) 20% 개선)</li> <li>○ 여수엑스포 한/영 자동통번역 대국민 서비스 실시</li> </ul>

2차년도 (2013)	○ 자동통번역 산업기반 조성(2차)	<ul style="list-style-type: none"> <li>○ 대국민 한/일 자동통역 시범서비스 실시</li> <li>○ 자동통번역 플랫폼 기반 Open API 구현 (1차)</li> <li>○ 사용자 로그데이터 기반 한/일 자동통역 성능 개선 (통역률 ERR 20% 개선)</li> <li>○ 중소기업 시제품 제작 지원 (2차)</li> <li>○ 다국어 음성언어 연구기반 조성 (1차)</li> <li>○ 2014년 인천 아시안게임 대비 한,중,영,일 4개 언어 자동통역 시범서비스 구현</li> </ul>
3차년도 (2014)	<ul style="list-style-type: none"> <li>○ 자동통번역 산업 생태계 구축(1차)</li> <li>○ '14년 인천 아시안 게임 한,영,일,중 4개국 자동통번역 대국민 서비스 실시</li> </ul>	<ul style="list-style-type: none"> <li>○ 대국민 한/중 자동통역 시범서비스 실시</li> <li>○ 자동통번역 플랫폼 기반 Open API 구현 (2차)</li> <li>○ 사용자 로그데이터 기반 한/중 자동통역 성능 개선 (통역률 ERR 20% 개선)</li> <li>○ 중소기업 시제품 제작 지원 (3차)</li> <li>○ 다국어 음성언어 연구기반 조성 (2차)</li> <li>○ 2014년 인천 아시안게임 한,중,영,일 4개 언어 자동통역 시범서비스 실시</li> </ul>
4차년도 (2015)	<ul style="list-style-type: none"> <li>○ 자동통번역 산업생태계 구축(2차)</li> <li>○ 자동통번역 글로벌 경쟁력 확보(1차)</li> </ul>	<ul style="list-style-type: none"> <li>○ 한/스, 한/불 자동통역 프로토타입 서비스 구현</li> <li>○ 자유발화 일상대화형 동시통역 원천기술 개발(1차)</li> <li>○ 자동통번역 플랫폼 기반 Open API 구현 (3차)</li> <li>○ 중소기업 시제품 제작 지원 (4차)</li> <li>○ 대학, 산업체와 협업을 통한 다국어 음성언어 연구기반 조성 (3차)</li> <li>○ 2015년 광주 유니버시아드 한,중,영,일 4개 언어 자동통역 서비스 지원</li> </ul>

5차년도 (2016)	○ 자동통번역 글로벌 경쟁력 확보(2차)	<ul style="list-style-type: none"> <li>○ 한/스, 한/불 자동통역 성능 개선 (통역률 ERR 10% 개선)</li> <li>○ 핸드프리 양방향 자동통역 원천기술 개발(2차)</li> <li>○ 다국어 음성인식 기술 개발(독일어, 러시아어)</li> <li>○ 자동통번역 플랫폼 기반 Open API 구현 (4차)</li> <li>○ 중소기업 시제품 제작 지원 (5차)</li> <li>○ 다국어 음성언어 연구기반 조성 (4차)</li> <li>○ 2016년 평창 올림픽 테스트 이벤트용 5개 언어 자동통역 구현</li> </ul>
6차년도 (2017)	○ 자동통번역 글로벌 경쟁력 확보(3차)	<ul style="list-style-type: none"> <li>○ 한/독, 한/러 자동통역 프로토타입 서비스 시스템 구현 (통역률 ERR 10% 개선)</li> <li>○ 핸드프리 양방향 자동통역 원천기술 개발(3차)</li> <li>○ 다국어 음성인식 성능 개선(8개 언어 대상)</li> <li>○ 자동통번역 플랫폼 기반 Open API 구현 (5차)</li> <li>○ 중소기업 시제품 제작 지원 (6차)</li> <li>○ 대학, 산업체와 협업을 통한 다국어 음성언어 연구기반 조성 (5차)</li> <li>○ 2018년 평창 동계올림픽 대비 한,중,영,일,불,스,독,러 7개 언어 자동통역 서비스 구현</li> </ul>
7차년도 (2018)	<ul style="list-style-type: none"> <li>○ 자동통번역 글로벌 시장 진출</li> <li>○ '18년 평창 동계 올림픽 한,중,영,일,불,스,독 7개 언어 자동통번역 대국민 서비스 실시</li> </ul>	<ul style="list-style-type: none"> <li>○ 8개국 일상대화형 동시통역 시범서비스 실시(평창:지니톡SUM)</li> <li>○ 자유발화 일상대화형 동시통역 원천기술 개발(4차)</li> <li>○ 자동통번역 플랫폼 기반 Open API 구현 (6차)</li> <li>○ 중소기업 시제품 제작 지원 (7차)</li> <li>○ 다국어 음성언어 연구기반 조성 (6차)</li> <li>○ 2018년 평창 동계올림픽 한,중,영,일,불,스,독,러 7개 언어 자동통역 서비스 실시</li> </ul>

라. 주요 성과목표 및 계획

○ 성과목표 달성실적

성과지표 (주요성능 Spec)	세계최고 수준	기술개발 목표치( '16)	달성 실적	목표치 산출근거
한/스페인어 통역성능 (통역률)	통역률=60 %( Google)	통역률=69% (ETRI)	76.1% (BMT 결과 구글은 72.1%로 구글 대비 4% 우수함)	자동통역 성능개선 률 측정을 통해 정 량적 평가가 가능 하고, 특히 세계 수준(Google) 대 비 벤치마킹을 통 해 상대적 우수성 입증이 가능하여 목표로 설정함
한/불어 통역성능 (통역률)	통역률=60 %( Google)	통역률=68% (ETRI)	61.8% (BMT 결과 구글은 56.4%로 구글 대비 5.4% 우수함)	

표 2 성과목표 달성 실적

- ※ 통역율(%)=음성인식율(%) \* 자동번역률(%)
- ※ 음성인식률(%)=20명 이상이 발성한 1,000발화 대상 단어인식률 평가
- ※ 자동번역률(%)=번역 전문가 3인 이상 대상 MOS 5점 기준 주관적 평가 실시
- ※ 당해년도 새로 만든 평창올림픽용 평가셋의 난이도가 전년도 평가셋보다 높아 당초 목표성능에는 미치지 못했으나 구글과 BMT를 통해 ETRI시스템 성능이 한/스 4%, 한/불 5.4% 각각 우수한 것으로 평가되어 목표를 달성함

○ 정량적 실적

공통지표(필수제시)			자율지표(자율제시)					
지표명			목표	실적	지표명		목표	실적
SCI 논문(건)			1	1 (제출)	과학적 성과	표준화된 IF 상위 20% SCI 논문(건)	1	0
특허(건)	국내	출원	7	8	기술적 성과	특허활용률 (기술이전건수 / 특허등록보유건수)	20% (10건/50건)	3건
		등록	4	0		국제표준특허(건)	2	2건 (반영)
	국제	출원	5	7		국제표준승인 표준기고서(건)	1	1건 (제출)
		등록	2	2		3급 특허(건)	1	0
기술이전(건)			2	3	경제적 성과	연구비 대비 기술료 수입(%)	8%	4.3%
기술료(억원)			1	1.26 (부가세포함)				

표 3 정량적 실적

○ 정성적 실적

- 세계최초 Zero-effort UI/UX 통역용 어어셋 개발 성공 (2017.8월 ISO 국제표준 예정)
- 세계수준 8개 언어 다국어 음성인식 기술 확보 (국내유일 다국어 기술 보유)
- 세계최초 발음사전 없는 음성인식 기술 및 topic LM 디코더 개발
- 실시간 방송음성 정제를 통한 대용량 학습용 DB 자동확보 기술 구현에 성공 (한국어, 독일어, 불어, 스페인어에 적용)
- 한컴인터프리 (ETRI연구소 기업 및 한컴그룹 자회사) ‘말랑말랑 지니톡’ 30만 앱 다운로드 기록 및 MWC2016 출품 (한컴은 단말통역기, 통역박스 등 다양한 제품으로 사업 확대 추진 중)
- 평창동계올림픽 테스트이벤트에 5개 언어 자동통역 시범서비스 적용 (2016.11~2017.3)

○ 사업화 실적

사업화 기업체	사업화 내용	사업화 제품 이미지
한컴인터프리	<ul style="list-style-type: none"> <li>• 2018 평창동계올림픽 8개 언어 자동통번역 서비스 공식공급사로 선정 (지니톡)</li> <li>• 韓-5개 언어 자동통역서비스 사업화 추진</li> <li>• 韓-3개 언어 단말탑재형 자동통역 사업화 추진</li> <li>• 세계 최초 자동통역용 이어셋 글로벌 사업화 추진</li> <li>• 중국, 인도, 중동 등 글로벌 진출 계획 수립</li> </ul>	
시스템인터내셔널	<ul style="list-style-type: none"> <li>• 수원, 제주, 인천공항 등 지자체 대상 통역서비스 사업화</li> <li>• 중국, 일본 대상 통역서비스 사업화 추진</li> </ul>	

표 4 사업화 실적

## 5. 세부 추진실적

### 가. 정성적 추진실적

#### 1) 한/스, 한/불 자동통역서비스시스템 구현

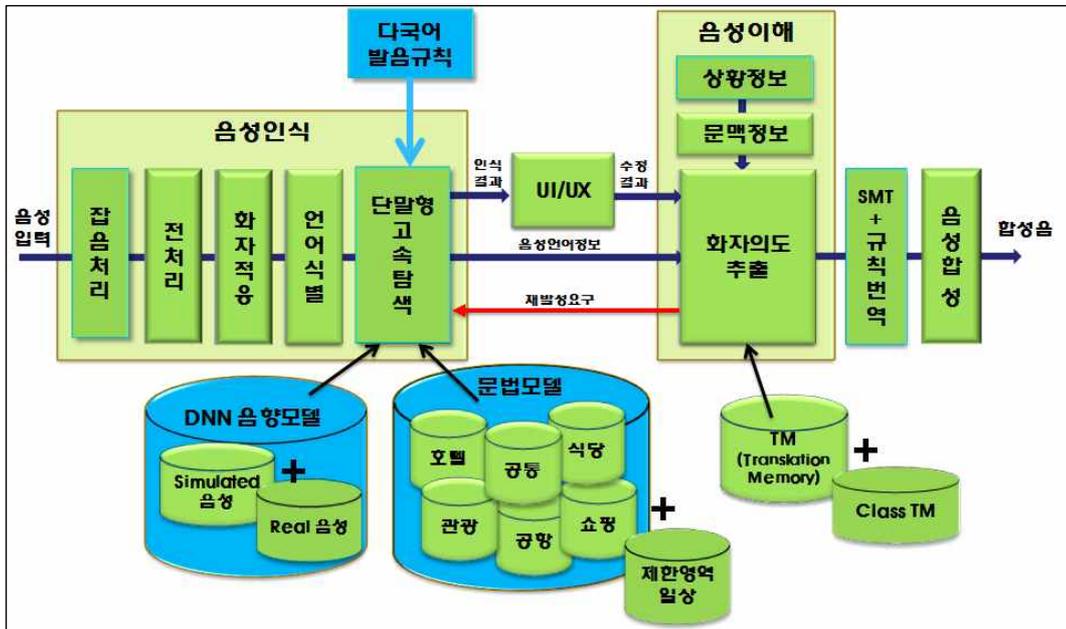


그림 11 자동통역시스템 구조

#### ○ 한/스, 한/불 지니특 자동통역 시스템 개선

- 문장부호복원 기술은 음성인식결과에 번역을 위해 필요한 문장부호를 복원하는 것으로, 평서문(.)과 의문문(?)을 구별하여 부호를 붙인다. 이 때 평서문과 의문문을 구분하는 규칙들이 존재하나, 이는 품사태깅, 구문분석 등의 과정을 거쳐야만 정확하게 구분이 가능하다. 본 시스템에서는 앞서 언급한 복잡한 과정을 생략하기 위해 키워드 분포 정보를 이용하여 평서문/의문문을 구별하였다. 최근에는 이보다 더욱 진보해 word2vec과 LSTM (Long Short Term Memory)을 적용한 방법을 도입하여 문장부호 복원 성능을 대폭 향상시켰다.

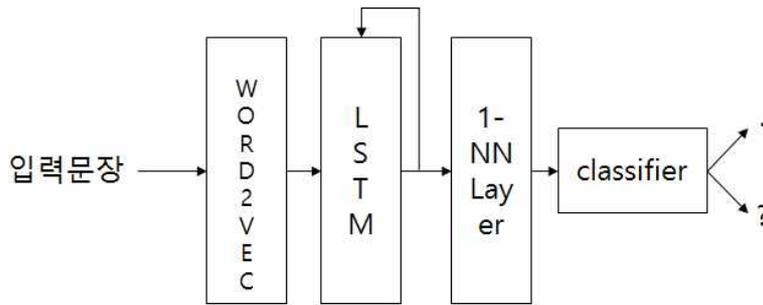


그림 12 딥러닝 기반 문장부호복원 구조도

- 키워드 기반의 문장부호복원이란, 평서문과 의문문으로 구분된 문장들로부터 문장의 앞과, 뒤에서 n개의 단어들로 구성된 리스트들을 추출하고, 평서문과 의문문에서의 분포를 계산하여 의문문에 특정 확률 이상이면 의문문 키워드로 선택하고, 아니면 평서문의 키워드로 선택하게 하는 방법이다. 참고로 키워드 기반의 문장부호복원에서 한국어는 스페이스 단위의 키워드를 사용하기에는 어절의 수가 너무 많아 음절을 사용하였다.
- 새롭게 제안한 방법은 평서문/의문문 문장부호 복원을 위하여 RNN-LSTM 딥러닝 방식을 적용하였다. 이 때 훈련 데이터로는 프랑스어 자막 1억 문장(평서문 8천만/의문문 2천만)을 Gensim word2vec의 입력 데이터로 사용하였다.
- 이렇게 제안된 방식을 프랑스어에 적용했을 때, 프랑스어 문장부호 복원 성능은 <표 5>에서와 같다. 평가셋으로는 텍스트 평서문 5만3천 문장과 의문문 3만6천 문장을 사용하였다.

	평서문(예측)		의문문(예측)		Recall	
	Keyword (기존방식)	LSTM (제안방식)	Keyword (기존방식)	LSTM (제안방식)	Keyword (기존방식)	LSTM (제안방식)
평서문(정답)	51,708	50,160	1,643	3,191	<b>0.969</b>	<b>0.940</b>
의문문(정답)	6,741	2,607	29,255	33,389	<b>0.813</b>	<b>0.927</b>
Precision	0.884	0.951	0.94	0.912	Accuracy	
					<b>0.906</b>	<b>0.935</b>

표 5 프랑스어 문장부호 복원 성능

- 문장부호 복원 기술을 다국어에 적용한 결과는 아래 <표 6>에서와 같다.

언어	키워드기반 (평서문/의문문)	RNN-LSTM (평서문/의문문)
영어	95.5% (97.9%/87.7%)	96.8% (97.5%/94.7%)
한국어	88.3% (93.2%/73.4%)	93.8% (93.2%/95.6%)
스페인어	83.9% (95.7%/64.7%)	90.9% (95.8%/82.7%)
불어	90.6% (96.9%/81.3%)	95.9% (95.9%/95.7%)
독어	90.8% (95.8%/82.3%)	96.8% (97.4%/95.6%)
러시아어	84.1% (95.5%/66.7%)	91.5% (95.5%/85.5%)

표 6 다국어 문장부호 복원 기술의 성능 평가 결과

- 결과에 있어 스페인어의 문장부호 복원 성능 평가 결과가 다른 언어에 비해 낮은 이유는 다음과 같다. 먼저 스페인어의 어순이 비교적 자유로운 편이므로 평서문에 의문부호만 붙여도 의문문이 되므로 대부분의 경우 화자의 억양에 의존할 수밖에 없다는 점이다. 또한 동사변형만으로도 주어를 알 수 있는 경우가 많기 때문에, 구어체에서는 대부분 주어를 생략하는 경우가 많으므로, 주어가 생략된 평서문이 의문문처럼 보일 수도 있는 위험이 있다. 이러한 요소를 해결하기 위해서는 향후 음향 정보도 함께 문장부호 복원에 활용하는 것을 검토할 수 있다.
- 한-스, 한-불 자동통역 시스템은 한컴인터프리의 말랑말랑 지니톡에 <그림 13>과 같이 통합을 완료하였으며 현재 국내 시범서비스를 진행하고

있다.

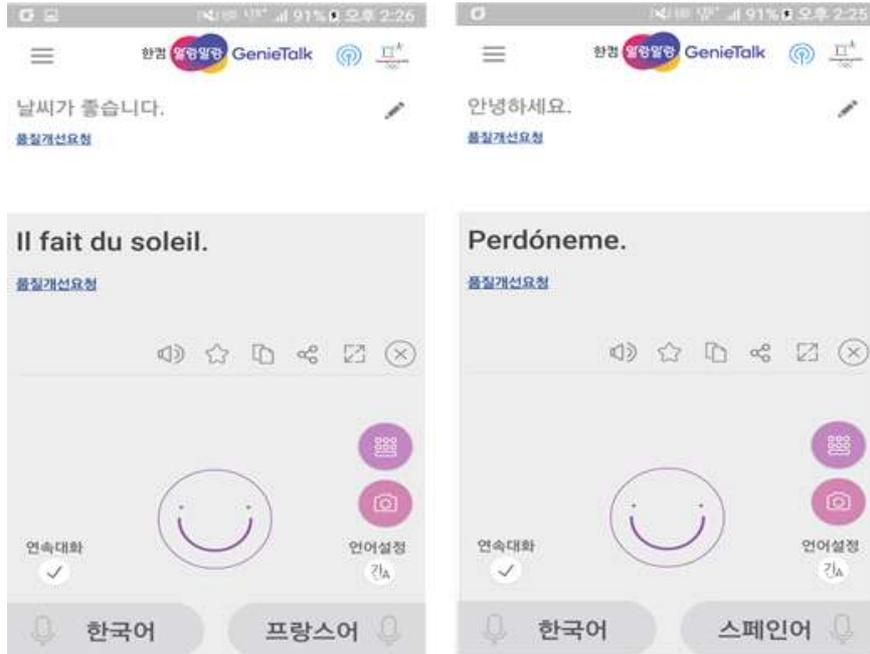


그림 13 한/스, 한/불 자동통역 화면

○ 한/스, 한/불 자동통역 성능 개선 (통역률 ERR 10% 달성)

- 자동통역 성능 개선을 위해 유사문장 검색 기능을 도입하였다. 유사문장 검색이란 한글과 스페인어, 한글과 불어, 각각의 쌍으로 구축된 TM(Translation Memory)을 기반으로 입력된 문장과 유사한 문장을 키워드 및 클래스 정보를 활용하여 대용량의 TM에서 검색하여 보여주는 기능이다. 유사문장 검색 기능은 아래와 같은 흐름을 가진다.

- . 한국어->스페인어와 한국어->불어의 경우, 한국어 입력을 검색하기 위해 형태소 품사태깅 과정을 거쳐 조사와 어미를 제외한 키워드 추출
- . 인명, 지명 등을 하나의 클래스로 적용하여 클래스 변환
- . 키워드를 기준으로 키워드 매칭 개수에 따라 순위를 정해 TM매칭
- . 상위 n개를 추출하고, 같은 키워드로 구성된 TM은 제거 (이때 같은 키워드라 하더라도 문형과 시체가 다르면 제거하지 않음)

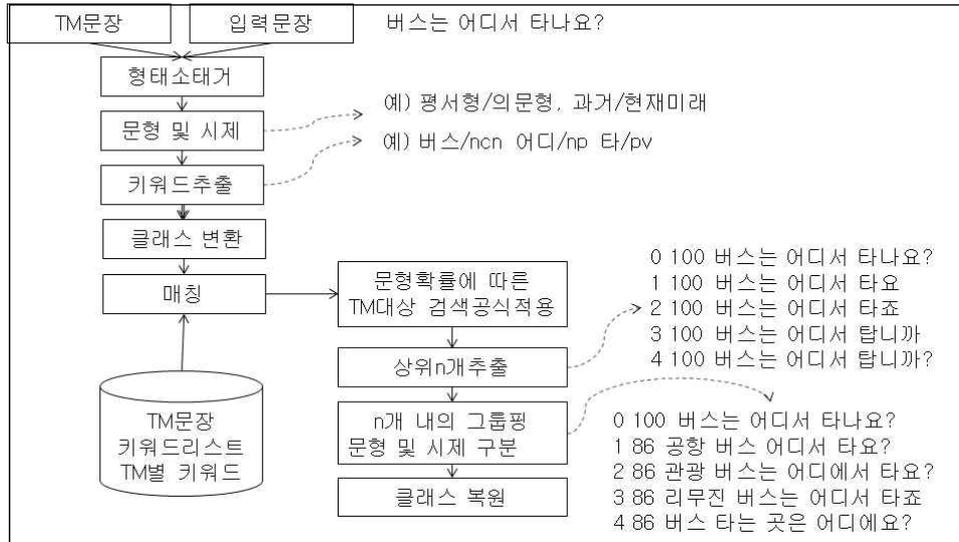


그림 14 유사문장 검색기 구조

	구조화 유사문	평가셋	번역률 개선
한국어->스페인어	760,568문장	500문장	11.6%
스페인어->한국어	673,490문장	500문장	0.3%
한국어->불어	605,094문장	500문장	1.9%
불어->한국어	538,765문장	500문장	0.7%

표 7 한국어/스페인어, 한/불어 유사문장 검색 기술의 성능평가 결과

○ 대용량 사용자 로그데이터 및 미디어 데이터 반자동 구축 및 DB화

- 딥러닝(Deep Learning)을 이용한 머신 러닝 기술은 다른 알고리즘보다 데이터베이스의 보강이 성능에 큰 영향을 미친다. 음성인식 성능향상을 위해 딥 러닝 기술을 적용하기 위해서는 대용량의 음성 데이터베이스가 요구되는데, 대용량 DB를 수동 구축할 경우, 데이터를 정제하는데 많은 인력과 비용이 소모되므로 자동화의 필요성이 매우 크게 요청된다. 이에 실시간 방송 콘텐츠 및 웹 데이터를 정제하여 음성언어DB 구축 자동화 기술 개발을 추진하였다.

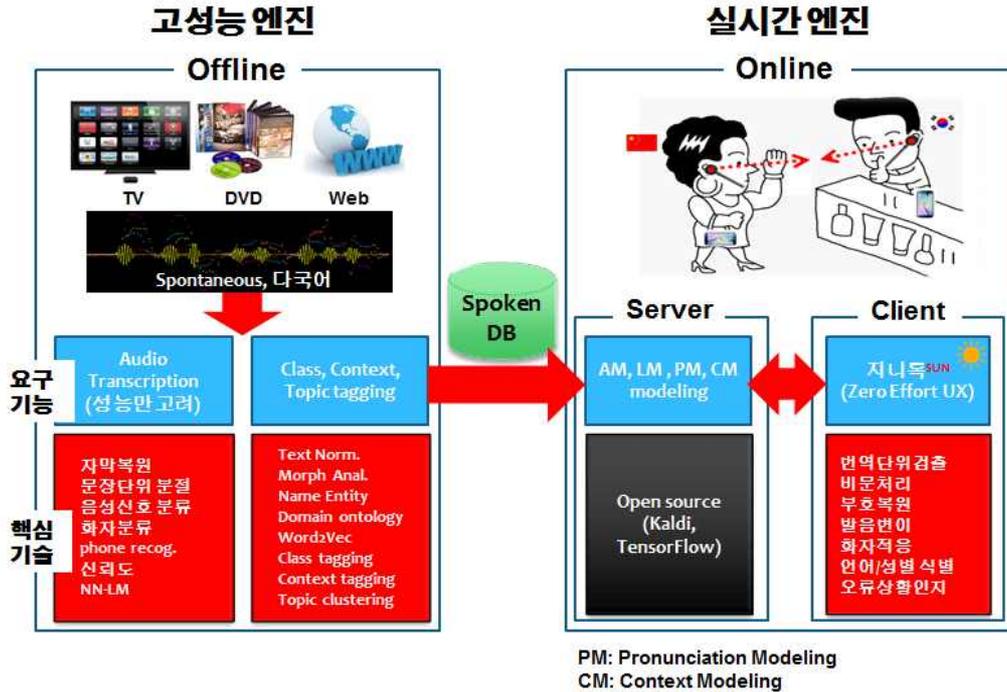


그림 15 방송 미디어데이터 기반 자동 DB구축 개념도

- 먼저 자막 방송 데이터를 수집하고, 수집된 자막과 방송의 음성을 시간적으로 일치 시켜 학습용 음성DB로 활용하였다. 제안된 방식은 다음의 절차를 따르며, 실험결과 자막과 음성이 100% 일치되는 경우만 가정했을 때, 전체 방송음성 분량의 약 25% 정도 음성을 DB화 가능하였다. 자동 DB 구축은 다음과 같은 순서로 이루어진다.

- . 오디오 분할: 길이가 긴 오디오 파일을 자막의 시간 정보를 이용하여 자막 별 오디오 파일로 분할
- . 텍스트 토큰화: 자막 텍스트를 음성인식기 출력과 동일한 단위로 변환
- . 음성인식: 분할된 오디오 파일에서 음성인식기를 이용하여 각 단어의 시간 정보를 가지는 가설 텍스트 추출
- . 텍스트 정렬: 자막 및 가설 텍스트 간의 공통 문자열을 탐색하고 이를 정제된 음성 데이터로 추출

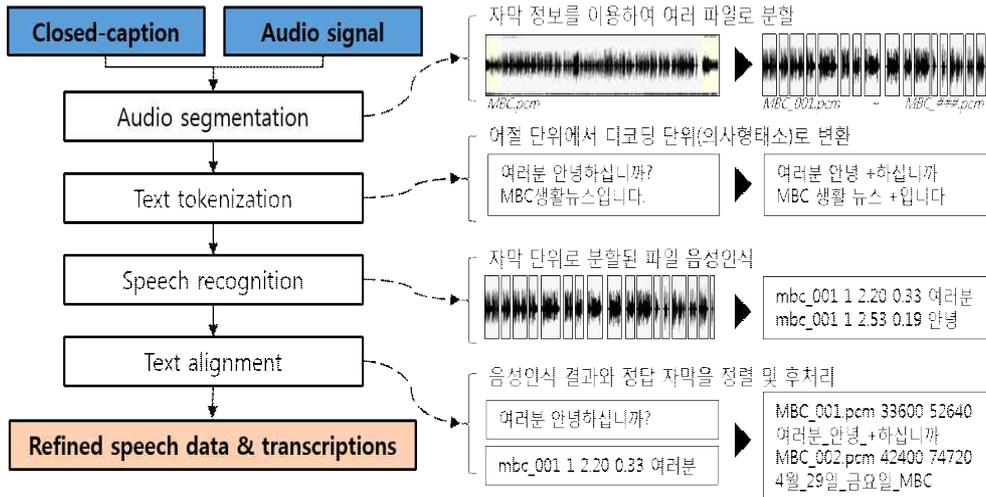


그림 16 음성DB구축 자동화 방법

- <표 8>은 방송음성 DB 중에서 일부를 이용하여 자동정제가 인식성능에 미치는 영향을 초기 실험한 것이다. 방송음성 DB 약 490시간과 자막정보를 이용하여 음향모델을 구축한 경우와 자동정제 과정을 거친 후 120시간에 대한 데이터만을 이용하여 음향모델을 구축한 경우의 음성인식 성능을 비교하였다. 음향모델에 사용된 시간은 적으나 자동정제된 음성데이터를 사용한 경우 음성인식 성능이 월등히 향상 되는 것을 확인할 수 있었다. 이 때 드라마의 경우 인식성능이 뉴스 인식성능에 비해 비교적 높지 않거나 이는 드라마의 음성이 사람의 자유발화 특성을 다분히 포함하고 있기 때문에 음성인식에 있어서 난이도가 상당히 높은 태스크이기 때문이다.

인식모델	인식평가(방송용)	
	뉴스	드라마
방송음성 DB 정제전 (490 시간)	69.7%	15.2%
정제후 (120 시간)	85.5%	57.6%

표 8 방송음성 DB의 자동 정제 전/후 성능 비교 (WA%)

## 2) 핸즈프리 양방향 자동통역 원천기술 개발 (2차)

### ○ 자유발화 대화체 음향모델링 및 초별 성능 평가

- 자유 발화 대화체 음향모델링은 자유 발화를 대량으로 포함하고 있는 방송음성DB를 이용하여 이루어졌다. 이를 위해 자동정제된 방송음성 DB로 음향모델을 학습하였고, 언어 모델의 경우 역시 대량의 방송자막을 이용하여 학습하였다. 이 때 음향모델은 총 5개월 동안 취득된 7개 채널의 방송음성을 사용하였다. 방송음성의 장르로는 뉴스, 다큐, 드라마, 예능 등 다양한 장르의 자유발화 대화체 방송이 포함되어 있다. 총 6000여 시간의 방송음성을 수집하였으며 자동정제 과정을 통해 최종적으로 1,400시간 분량의 데이터만을 훈련에 사용하였다.
- 평가용 데이터로는 훈련에 사용되지 않은 방송음성 데이터 중 뉴스와 드라마 각각 1시간 분량의 음성을 사용하였다. 뉴스와 드라마 음성을 수동으로 문장 단위의 발화구간을 나누고 전사 작업을 거쳐 평가용 데이터로 구축하였다. 또한 기존의 음성인식 성능 비교를 위해 자동통역에 사용되는 평가용 데이터에 대해서도 단어인식률(%)을 측정하였다.

인식모델	자유발화 인식평가		정형발화 인식평가				
	뉴스	드라마	사무실 1	식당	가정(TV)	지하철	사무실 2
방송음성 DB (1,400 시간)	89.6%	67.3%	95.7%	94.8%	94.5%	93.2%	94.3%

표 9 자유발화 방송음성 DB를 이용한 음성인식 성능비교 결과

- <표 9>는 자동정제된 음성 1,400시간을 사용하여 만든 음성인식 모델에 대해 실험한 결과로 <표 8>의 방송용 음성 인식 평가 결과에 비해 뉴스와 드라마 각각 28.5% 22.7%의 오류 감소율을 나타내었다. 특히 방송음성 DB 만으로 훈련한 인식모델을 이용하여 기존의 자동 통역용 음성 인식 평가를 수행한 결과 매우 높은 인식성능을 보임을 확인 할 수 있었다. 또한 가정환경이나 식당 등 잡음환경에서도 기존 모델보다 개선된 성능을 확보함으로써 방송음성을 이용한 음향모델링이 잡음에 매우 강인함을 알 수 있었다. 이로써 방송음성 DB의 자동정제과정을 통해 기존의 음성인식 데이터 구축 방식의 패러다임을 획기적으로 향상시킬 수 있는 방안을 마

련했다고 할 수 있다.

○ 동시통역용 웨어러블 디바이스 구조 연구 및 국제표준화 추진 (지니톡<sup>SUN</sup>)

- 기존 스마트폰 기반 자동 통역 과정에서의 답답함과 불편함을 해소하기 위하여 동시 통역용 웨어러블 디바이스를 개발하였다. 웨어러블 디바이스로 핸드프리어용 이어셋을 채택하였으며 이는 귓속마이크와 외부마이크를 이용한 2채널 음성인식 기능을 수행하는 것으로, 귓속마이크는 화자의 음성발화 여부와 음성끝점검출에 활용하고, 외부마이크로 들어오는 음성은 음성인식에 활용한다. 또한 합성음을 출력하는 동안에도 발화가 가능한 Barge-In 기능 구현을 완료하여 자유발화 기반의 실시간 통역이 가능한 세계최초 zero-effort UI/UX 도입이 가능하게 되었다.

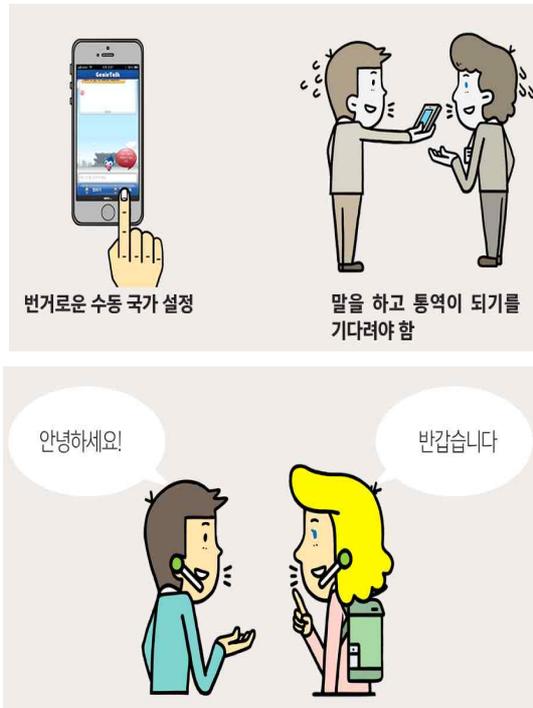


그림 17 기존 스마트폰 통역대비 제안된 핸드프리 통역의 사용자편의성 향상

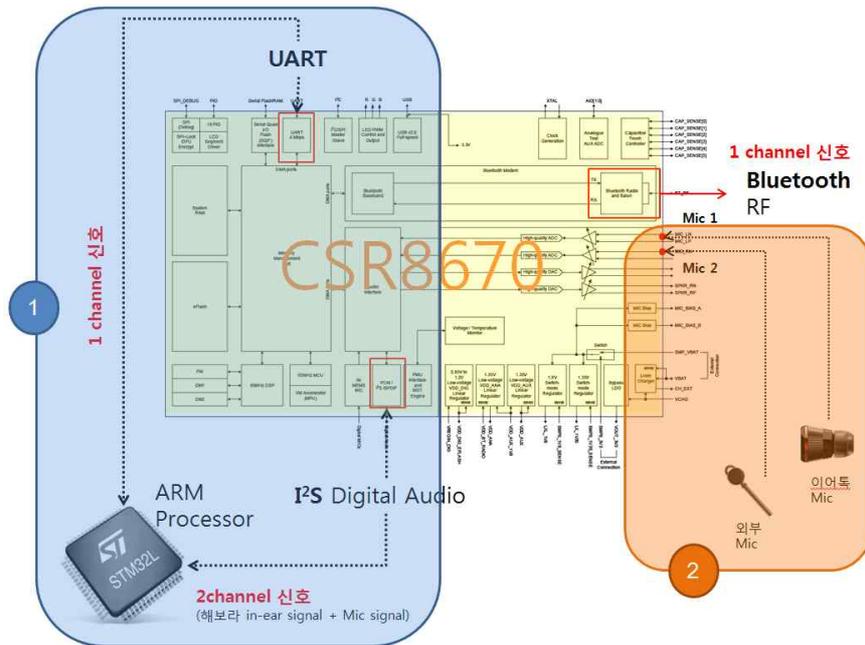


그림 18 핸즈프리 통역용 이어셋 2채널 음성신호 처리 구조



그림 19 핸드프리 통역용 이어셋 샘플

- 핸드프리용 이어셋(지니톡<sup>sun</sup>)을 이용한 자동통역은 현재 국제표준화를 추진 중에 있으며, 관련 국제특허를 표준특허로 다수 반영 예정이다.
  - . 2016.2: CD (Committee Draft) voting 결정 (로마)
  - . 2016.8: DIS (Draft International Standard) voting 결정 (서울)
  - . 2017.2: FIS (Final International Standard) 결정 (베를린)
  - . 2017.8: IS (International Standard) 결정 (제네바)
  
- 현재 연구소기업인 한컴인터프리는 2018년 평창동계올림픽에 지니톡<sup>sun</sup>을 적용할 계획이며 내년 스페인 바르셀로나에서 열리는 MWC2017에 출품하여 핸드프리 통역용 이어셋 기술을 국제적으로 홍보할 계획이다.

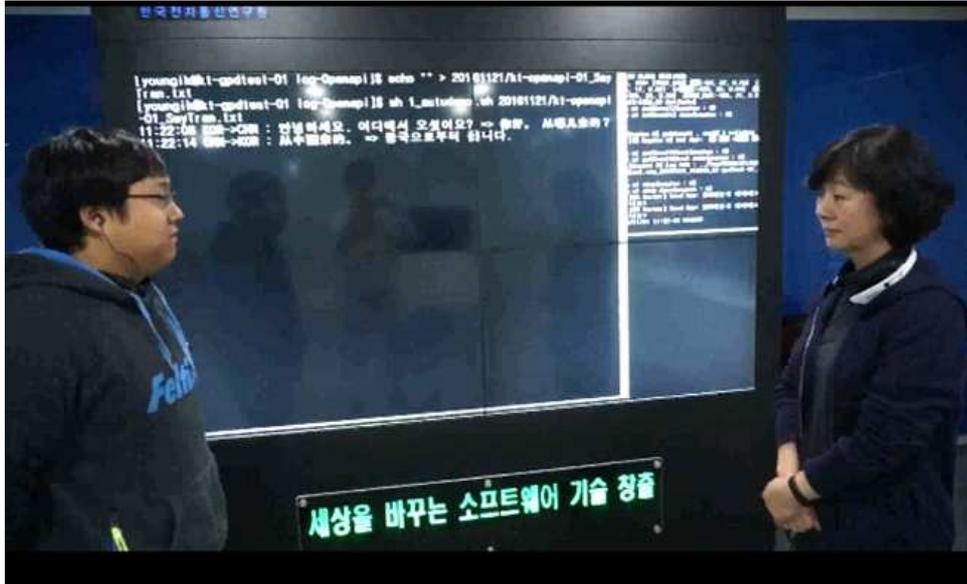


그림 20 실제 핸즈프리 양방향 자동통역 시연 장면

### 3) 다국어 음성인식 기술 개발

#### ○ 독일어, 러시아어 전처리, 발음사전 및 발음 변환기 구현

- 독일어는 발음이 규칙적이긴 하나, 예외발음이 많아 23만 엔트리에 대해 발음사전을 구축하였다. 이 때 발음사전에 없는 단어들에 대해서는 발음 변환 규칙을 적용하였다. 발음 표기에는 총 47개 폰셋을 사용하였다. 구현된 발음변환기 성능을 측정한 결과 14.8만 발성목록에서 98% 사전 기반으로 매칭이 이루어졌고, 사전에 없는 단어의 발음을 규칙으로 생성할 경우에는 91.5% 성능을 보였다.
- 러시아어의 경우 단어의 악센트를 기준으로 발음이 규칙적으로 변화하는 특징을 보인다. 이러한 점을 반영하여 발음변환기를 구현하였으며 또한 러시아어를 표기할 때는 단어의 악센트를 거의 표시하지 않는 점을 감안하여 100만 여개의 러시아어 악센트 사전을 별도로 구축하였다. 러시아어의 경우에는 발음표기에 43개 폰셋을 사용하였다. 구현된 발음 변환기의 성능을 측정한 결과 5천 발성목록에서 95.7%가 사전에 매칭이 이루어졌고, 사전에 없는 것을 규칙으로 생성할 경우에는 90.6%의 성능을 나타냈다.

#### ○ 클라우드 소싱 플랫폼을 이용한 음성DB 및 다국어 발음사전 정교화

- 음성인식 성능 개선을 위해서는 대용량의 DB 수집 및 발음사전 보강이 필수적이다. 이를 위해 클라우드 소싱 플랫폼에서 번역 문장을 수집하는 과정에서 추가로 다국어 음성을 녹음하였다. 이러한 과정을 통해 대용량의 음성 DB를 구축할 수 있었다. 또한 클라우드 소싱 기법을 도입함으로써 화자 및 녹음 환경에서의 다양성을 확보하여 강인한 음성인식기를 개발할 수 있도록 하였다.(프랑스어, 스페인어, 독일어, 러시아어 각 20만, 전체 80만 발화 구축 완료)
- 또한 발음사전을 규칙적으로 생성하기 어렵고 정제된 발음 사전을 확보하기 어려운 언어인 영어, 러시아어, 독일어, 프랑스어, 중국어, 일본어를 대상으로 평균 각 10만 엔트리에 대해 발음을 언어 전문가에 의해 수작업으로 기술하였다. 이를 통해 발음 사전의 강건성을 확보하는 동시에 외연을 확대할 수 있었다.

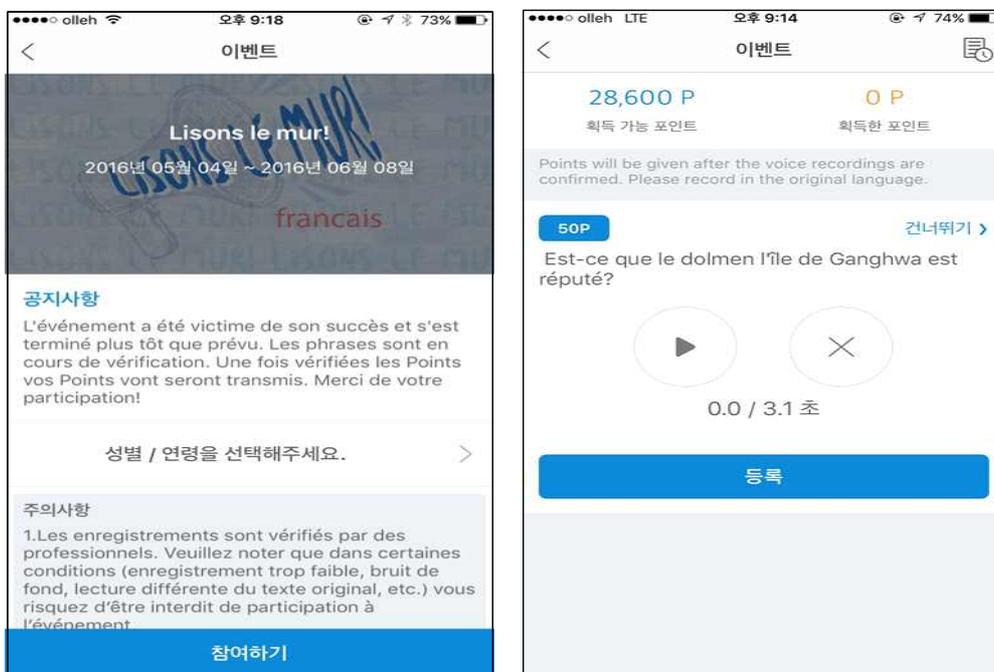


그림 21 클라우드 소싱 플랫폼 앱 실행 화면

○ 다국어 음향/언어모델링

- 스페인어의 경우 음향모델링을 위해 총 57만 발화에 해당하는 571시간의 음성데이터를 수집하였다. 수집된 데이터들은 기본적으로 16kHz의 대화

체, 낭독체 중심의 데이터이며 절대적인 음성데이터 양을 보충하고자 채널이 두 가지 이상 존재하는 DB에 대해서는 같은 발화에 대해 PC와 스마트폰 채널의 음성데이터를 모두 활용하여 훈련하였다. 이렇게 구축된 DB의 양은 총 933시간이다(<표 10>). 또 잡음에 대한 강인성을 높이기 위해 다양한 환경으로부터 수집된 잡음을 음성데이터에 섞어 최종적으로 1,866시간의 데이터를 음향모델링을 위해 사용하였다. 최종적으로 <표 11>에서 알 수 있듯이 평균 92.6%의 단어 인식률을 나타내었다.

DB명	수집채널	발화개수/화자수	발성시간
ESP_ASR001	PC	4만 발화/200화자	36 시간
Nation_DB2013	PC,PHONE	10만 발화/700 화자	92 시간
HumanMedia2014	PC,PHONE	8만 발화/400 화자	86 시간
Nation2014	PC,PHONE	6만 발화/400 화자	61 시간
Flitto_ESP_2015	PHONE	10만 발화/2,200화자	70 시간
King-ASR-142	PHONE	9만3천 발화/310화자	132 시간
Flitto_ESP_2016	PHONE	9만7천 발화/390화자	94 시간
계		57만 발화	571 시간

표 10 스페인어 음성데이터 현황

Image	Beam	Lm Weight	단어인식률
ESP.AM.DNN.5bvr6r3-20161108_ES.SP L2+NEWS_90.600p.g11.e10-1s	16	3.0	92.6%

표 11 스페인어 음성인식 평가결과

- 프랑스어의 경우에는 음향모델링을 위해 총 43만 발화에 해당하는 493시간 분량의 음성데이터를 수집하였다. 프랑스어 역시 수집된 데이터들은 기본적으로 16kHz의 대화체, 낭독체 중심의 데이터이다. 이 때 스페인어 훈련에서 같은 발화에 대해 채널별 DB를 모두 통합하여 훈련한 효과가 최초 기대보다는 낮게 나타나 프랑스어에서는 단일 채널의 DB로 훈련하였다. 그리고 특기할만한 사항으로는 AudioBook2015 DB의 경우 텍스트북을 낭독한 형태의 DB이기 때문에 화자의 수가 다양하지 않아 전체 344시간 중 일부만 훈련에 사용하였다.

DB명	수집채널	발화개수 및 화자수	발성시간
FRF_ASR003	PC	1만 발화/100 화자	26.3 시간
HumanMedia2013	PC,PHONE	6만 발화/300 화자	48 시간
HumanMedia2014	PC,PHONE	4만 발화/200 화자	43.3 시간
Nation2014	PC,PHONE	12만 발화/800 화자	122시간
AudioBook2015	PC		70(344) 시간
Flitto_FRA_2015	PHONE	11만 발화/1,600화자	99 시간
Flitto_FRA_2016	PHONE	9만7천 발화/1,031화자	85 시간
계		43만 발화	493.6 시간

표 12 프랑스어 음성데이터 현황

- 프랑스어의 경우에도 잡음에 대한 강인성을 높이기 위하여 잡음을 섞은 음성데이터를 생성하여 통합하여 음향모델링을 위해 사용하였으며 이렇게 훈련한 결과 <표 13>에서와 같이 90.6%의 단어 인식률을 나타내었다.

Image	Beam	Lm Weight	단어인식률
FRA.AM.DNN.5bv6r5-20161111_ FRA.LM.PTYPE5-1	16	3.0	90.6%

표 13 프랑스어 음성인식 평가결과

○ 다국어 음성인식 성능 개선

- 영어의 경우 언어모델을 3-gram에서 4-gram으로 확장하여 영어 음성 인식 성능 개선을 완료하였다. 인식 어휘 수 역시 15만 단어급(2016년1월)에서 84만단어급(2016년12월)으로 확장되었다.

구분		2016.01	2016.12
Order		3-gram	4-gram
단어 수		15만 5천 단어	84만 단어
ppl	Office1	50.6	34.8
	Office2	90.9	61.8
WER(%)	Office1	5.42	3.43
	Restaurant	6.09	4.36
	HomeTV	10.01	7.84
※ Word Error Rate	Subway	12.79	9.97
	Office2	7.89	5.9

표 14 영어 음성인식 성능 개선

- 발음변환기 제거를 위한 grapheme 인식기 구현

영어 인식기의 경우 발음 변환기 없이 훈련하는 grapheme 인식기를 시범적으로 개발하였다. 통상 특정언어를 음성인식하기 위해서는 소리인 음성 신호를 문자와 연결할 수 있는 연결 단위가 필요하다. 이 때 주로 신호와 단어를 연결하기 위해 음소라고 하는 기본 요소를 전문가의 지식으로 정하고, 데이터를 이용하여 통계적 값을 모델링 하게 된다. 이 과정에서 언어나 소리의 특성이 달라서 기본 음소의 개수와 종류가 다르게 되며 지금까지는 이를 위해 전문가의 통계적 분석이 요구되었다. 특히 단어를 음소열로 나타내는 발음사전은 더더욱 전문가의 손길이 필요하였는데, 이는 발음 규칙에 따라 음소가 변하는 현상을 반영하여 G2P(grapheme to phoneme) 프로그램을 생성해야 하는 작업이 음성인식을 위해서 반드시 필요하기 때문이다. 여기에 새로운 단어가 추가될 때 마다 새로운 발음의 음소열을 결정하는 것은 지속적인 전문가의 손길이 요구되는 작업이며 이 과정에서 실제 발음과의 차이에 따른 오류도 증가하는 경향이 있다.

이런 문제를 해소하기 위하여 character 단위의 grapheme 정보를 발음사전의 음소로 활용해 G2P 없이도 음향모델을 생성하는 방법을 시도하였다. 기본적인 아이디어는 character의 좌우 context를 증가시켜 음성의 소리와 유일하게 연결되는 유사 음소가 설정되도록 하는 방법이다. 그리

고 DNN의 senone을 증가시켜 기계학습과 데이터에 의한 분류가 이루어져 음성인식이 이루어지도록 하는 것이다.

이러한 아이디어를 검증하기 위해 음소 43개에 senone 8K를 기본으로 갖는 영어 음성인식기를 대상으로, 기본 성능이 <표 15>에서 처럼 96.19%의 성능을 보이는 음향모델의 환경에 character 26개를 기본 음소처럼 사용하는 grapheme 기법을 적용해 보았다.

	with GtP (발음변환기 포함)	without GtP (발음변환기 제거)		
		G_C3S8	G_C5S13	G_C5S13Ab
단어인식률	<b>96.2%</b>	92.3%	<b>94.3%</b>	<b>95.2%</b>
문장인식률	81.0%	66.2%	71.5%	75.9%

표 15 Phone 음성인식 vs. grapheme 음성인식

최초의 시도(G\_C3S8)는 character의 좌우 context를 1개씩 보고, senone 개수를 8K로 했을 때이고, 그 이후(G\_C5S13)에는 좌우 context를 2개로 늘리고 senone 개수도 13K로 증가시켰다. 그리고 영어의 특성상 대문자를 사용하는 약어 발음의 데이터가 많은 비중을 차지하고 있어서, 마지막 실험(G\_C5S13Ab)에서는 아래 예제와 같이 대문자 발음을 소문자 character의 자음과 모음으로 할당하도록 다음과 같이 고정하여 처리하였다.

Ex) A → e i, B → b i, C → c i, D → d i, E → i, F → e f, G → g i, ...

위와 같은 실험을 통해 기존의 G2P 프로그램이 없이도 기존 성능과 1% 차이에 불과한 성능을 확보할 수 있음을 알 수 있었다. 이는 향후 G2P 구현과 유지 보수가 어려운 다국어 음성 인식기 개발에 활용될 수 있을 것으로 보인다.

- word2vec 기반 Topic LM (Language Model) 구현

Topic LM의 도입은 대화의 주제 정보를 이용해 음성인식 성능을 향상시키는 것을 목적으로 한다. 즉, 인식된 발화에서 주제를 인식하고 이를 통해

다음 발화의 음성인식 후보 단어들의 발생 가능성을 조정하는 것이 목적이다. 이 때 사용한 단어 임베딩 알고리즘이 Word2Vec 알고리즘이다. 이는 단어의 의미는 주변단어들과의 관계로부터 결정된다는 가정으로부터 알고리즘이 시작한다. 즉 Word2Vec에서, 단어  $w_k$ 가 주어졌을 때 같은 문장의 일정 범위(window)에서  $w_j$ 가 나타날 확률  $y_j$ 는 <그림 22>와 같은 신경망을 이용하여 아래 수식과 같이 계산된다.

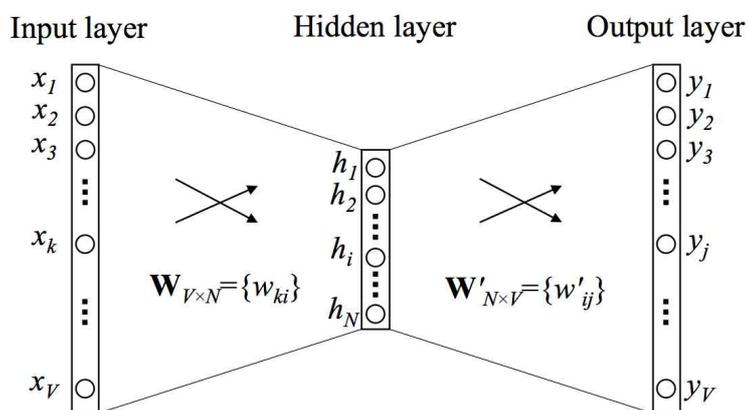


그림 22 Word2Vec 기본 모델

$$p(w_j|w_I) = \frac{\exp(\mathbf{v}'_{w_j}{}^T \mathbf{v}_{w_I})}{\sum_{j'=1}^V \exp(\mathbf{v}'_{w_{j'}}{}^T \mathbf{v}_{w_I})}$$

그림 23 말뭉치 특성

Topic LM의 적용을 위한 Word2Vec 훈련에 사용된 말뭉치의 규모는 아래와 같다.

말뭉치	문장 수	고유문장 수	문장 평균길이	어절 수	단어 수
위키 말뭉치	3,339,405	2,858,475	18.8	56,064,190	1,604,750
나무위키 말뭉치	8,644,779	7,991,345	22.8	184,867,680	3,925,329
자막 말뭉치	24,075,126	8,513,410	6.5	122,897,922	812,777
블로그 말뭉치	21,765,365	18,007,107	8.4	161,402,416	1,204,378

표 16 실험에 사용된 말뭉치

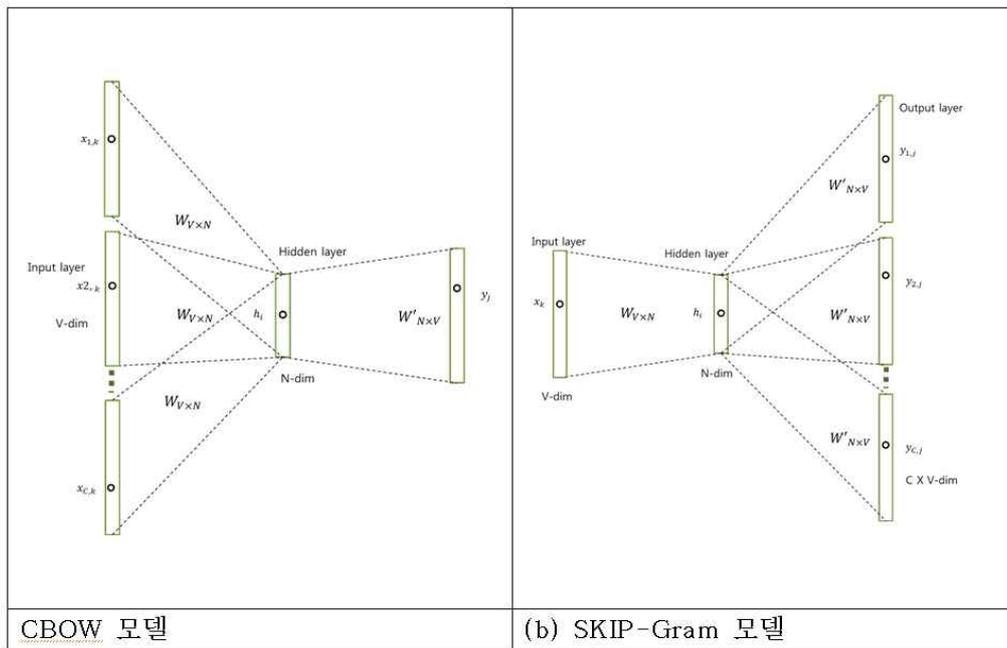


그림 24 CBOW(continuous bag of words) 모델과 SKIP-gram 모델의 네트워크 구조

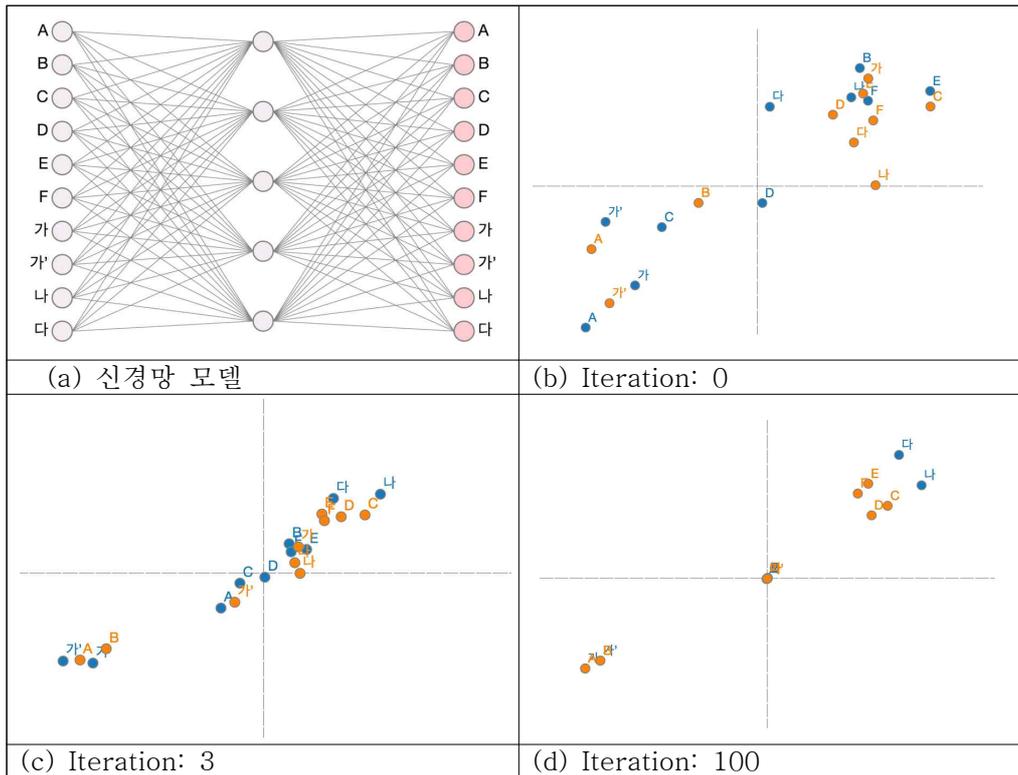


표 17 Word2Vec 학습 예 (파란색 점: 입력벡터, 주황색 점: 출력벡터)

본 연구에서는 기존의 CBOW, SKIP-Gram 단어 임베딩 방법과는 달리, Topic LM의 특성을 감안해 단어 임베딩에서 일반적으로 사용되는 입력 매트릭스와 함께, 출력 매트릭스도 단어 임베딩으로 사용해 보는 아이디어를 도출하였다. 이는 학습 과정에서 대상 단어의 입력 벡터와 문맥 단어의 출력 벡터가 서로 가까워지는 과정을 적용해보고자 하기 때문이다. 이를 통해 대상 단어와 가까이 있는 단어가 그 단어와 같이 나올 수 있는 단어들로 조정되기를 기대하는 것이며 이러한 모델을 Dual Embedding 모델이라고 한다.

<표 18>은 블로그 말뭉치를 통해 학습한 SKIP-Gram 모델에서 **결혼**과 유사도가 높은 단어를 추출한 결과이다. 표에서 IN-IN은 **결혼**의 단어 벡터를 모델의 입력 매트릭스로부터 가져오고 비교 대상 단어들의 단어 벡터도 입력 매트릭스의 값을 사용한 결과이다. IN-OUT은 **결혼**의 단어 벡터는

입력 매트릭스로부터, 비교 대상 단어들의 단어 벡터는 출력 매트릭스의 값을 이용한 결과이다.

<표18>의 결과를 살펴보면, IN-IN에서는 **결혼**과 유사하게 인생의 특별한 이벤트들이 **결혼**과 가까운 단어로 나타났다. IN-OUT에서는 결혼 기념일, 결혼을 앞둔, 결혼 예물과 같이 주로 결혼과 같이 한 문장에서 나타날 수 있는 단어들이 **결혼**과 가깝게 나타났다.

	IN-IN	IN-OUT
1	이혼(0.771)	결혼(0.373)
2	약혼(0.768)	신혼(0.215)
3	퇴사(0.762)	신부(0.208)
4	출산(0.752)	기념일(0.205)
5	연애(0.723)	앞두고(0.203)
6	이직(0.719)	예비(0.2)
7	결혼식(0.718)	앞둔(0.19)
8	군입대(0.706)	연애(0.17)
9	취직(0.702)	혼수(0.162)
10	상견례(0.701)	예물(0.161)

표 18 '결혼'의 dual embedding 비교

<표19>에서는 태도와 가까운 단어들을 추출하였다. IN-IN에서는 앞서 <표 18>과 유사하게 행태, 말투, 언행 등 태도와 문법적으로 유사하고 대체 가능한 단어들이 주로 나타난 반면에 IN-OUT에서는 진중한 태도, 무심한 태도와 같이 태도와 같이 많이 쓰이는 단어 들이 등장하는 것을 볼 수 있다.

	IN-IN	IN-OUT
1	행태(0.775)	태도(0.278)
2	행동거지(0.773)	무심(0.261)
3	고압적(0.762)	자신(0.211)
4	뒹뒹이(0.745)	행동(0.203)
5	말투(0.741)	그들(0.195)
6	정의감(0.736)	진중(0.185)

7	대하던(0.735)	타인(0.183)
8	언행(0.734)	의식(0.182)
9	행동(0.734)	대하는(0.175)
10	인류애(0.73)	말하는(0.166)

표 19 ‘태도’의 dual embedding 비교

기존의 단어 임베딩은 단어와 가까운 단어를 찾을 때, 주로 자신과 문법적으로, 또는 의미적으로 대체 가능한 단어들이 가깝게 나오는 경향이 있다. 하지만 주제를 이용한 단어 선택에는 유사단어를 찾는 것보다 대상 단어와 같은 문장에서 쓰일 수 있는 단어를 찾는 능력이 중요하므로 본 연구에서는 이러한 효과를 나타내는 단어 임베딩을 찾기 위해 여러 시도를 해보았다. 그 결과 일반적으로 단어 임베딩 학습 후 버려지는 임베딩 학습 모델의 출력 가중치 매트릭스의 활용에 대해 주목하고 그 가능성에 대해 살펴 보았으며 실험 결과 출력 가중치를 함께 사용하는 방법이 본 연구에서 추구하는 단어 임베딩의 특성을 만족함을 알 수 있었다. 또한, 이러한 단어 임베딩을 적용한 Topic LM을 실제 음성 인식기에 구현하였고 다양한 방법으로 이의 성능에 대해 추가 실험할 예정이다.

○ 원어민 대상 성능 평가 (구글 음성인식엔진과 8개 언어 대상 비교평가)

- 지니톡 음성인식기의 성능평가를 위해 동일한 평가셋에 대해 구글의 음성 인식기와 비교 평가를 실시하였다. 그 결과 한국어 및 중국어의 경우 ETRI 음성인식기의 성능이 뛰어나며 영어의 경우에는 구글이 뛰어난 것을 알 수 있었다. 기타 언어에 대해서는 약 1% 내외의 성능 차이로 언어별 편차를 나타내었다. 평가를 통하여 개발된 음성 인식기의 성능이 세계 최고 수준에 근접하다는 것을 다시 한 번 확인할 수 있었다.

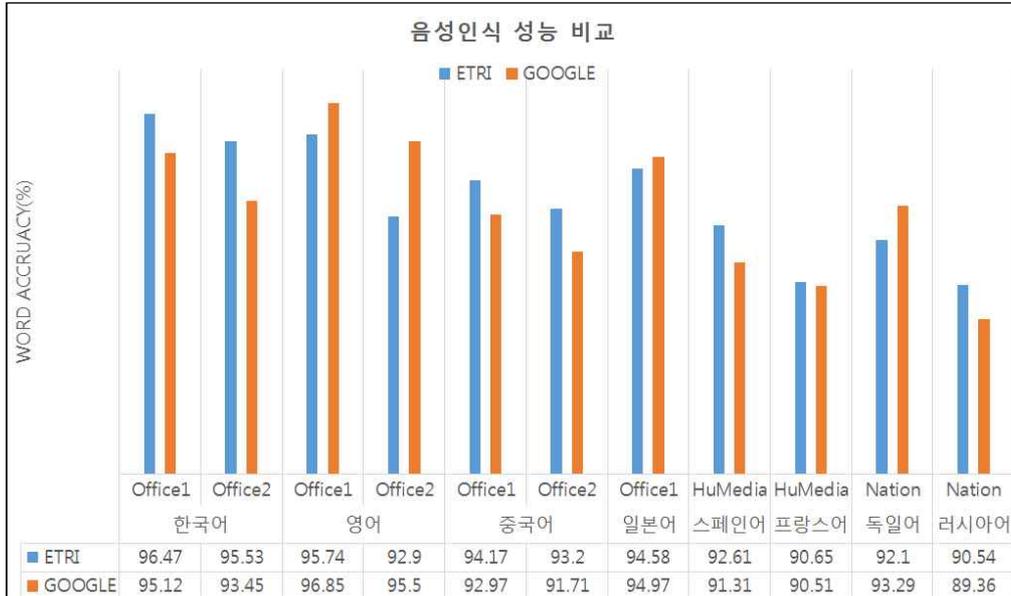


그림 25 음성인식 성능 비교

#### 4) 자동통역 플랫폼 기반 Open API 개선 (4차)

- 한/스, 한/불 통합을 위한 자동통역 플랫폼 보완 설계 완료
  - 한/스, 한/불 통합을 위하여 기존에 설계되어 있던 한-영, 중, 일 통역 플랫폼에 대상 언어인 스페인어, 프랑스어를 추가하였다. 이는 한컴인터프리의 말랑말랑 지니톡 플랫폼을 대상으로 이루어졌으며 통합 후 시범 서비스를 통해 프랑스어, 스페인어 자동 통번역 기능이 원활히 동작함을 확인할 수 있었다.
  
- 스페인어, 불어 통번역 기능 제공을 위한 지니톡 플랫폼내 Open API 추가 완료
  - 동일한 방법으로 한/스, 한/불 통합을 지니톡 플랫폼에도 실시하였으며 현재 프랑스어, 스페인어 음성 인식 기능을 Open API서버를 통해 업체에 지원이 가능한 상황이다.

#### 5) 중소기업 시제품 제작 지원 (5차)

- 중소기업 대상 자동통역 BM 발굴 및 시제품제작 지원
  - 2016년도에는 기존에 지원해오던 업체 외에 신규 업체로 해보라와 (주)소프트모터 2개사에 대해 자동통역 BM 발굴 및 시제품 제작을 지원하였다. 지원 내용을 표로 정리하면 아래와 같다.

지원업체	지원 내용
해보라	<ul style="list-style-type: none"> <li>• 해보라는 이어셋 제조업체로 음성인식 및 통역서비스에 자사 이어셋 적용을 통해 시장을 확장하고자 함</li> <li>• ETRI는 해보라 이어셋을 음성인식 및 통역성능 평가를 지원하고, 통역용 이어셋 음질 개선을 위한 시제품 제작 지원을 추진함</li> </ul>
(주)소프트모터	<ul style="list-style-type: none"> <li>• (주)소프트모터는 영상콘텐츠를 대상으로 음성인식 및 번역 결과를 부가한 UCC 서비스를 신규사업으로 추진</li> <li>• 영어 음성인식엔진 성능 개선방안을 지원하고, 향후 통번역 open API 지원을 통한 비즈니스 모델 검증을 추진하기로 함</li> </ul>

표 20 중소기업 대상 자동통역 BM발굴 및 시제품 제작 지원 내용

- 자동통역 사업화를 위한 전단계로 시범서비스 실시 지원
  - 한컴인터프리 대상으로 평창동계올림픽 테스트 이벤트에 시범적으로 적용

할 수 있도록 한/스, 한/불 자동통역 및 독일어, 러시아어 음성인식 기술을 지원하였으며 현재 시범 서비스 중이다.

#### 6) 대학, 산업체와 협업을 통한 다국어 음성언어 연구기반 조성 (4차)

○ 다국어 음성언어정보처리 기반 기술 연구

- 본 사업을 통해 아래 <표 21>에 정리한 것처럼 3개 대학과 다국어 음성언어 기반 기술 연구를 수행하였다.

대학협력	협력 내용
충남대학교	<ul style="list-style-type: none"> <li>• Word2vec 방식을 응용한 Topic LM 실효성 검증을 공동으로 추진</li> <li>• 지니톡 음성인식 디코더에 Topic LM을 적용하여 성능개선 실험을 추진 중임</li> </ul>
충북대학교	<ul style="list-style-type: none"> <li>• 음성데이터와 자막을 align하여 딥러닝 학습용 음성DB를 자동으로 구축하는 기술을 공동으로 개발</li> <li>• 한국어에 대해 1차 실험 결과, 금년 7~8개월 동안 구축한 음향모델이 지난 10년간 구축한 음성DB를 기반으로 하는 음향모델과 대등한 성능을 확보</li> </ul>
한국외국어대학교	<ul style="list-style-type: none"> <li>• 다국어 음성인식기 성능 개선을 공동으로 추진</li> <li>• 평창동계올림픽 자동통역 성능 평가에 우선 협력하기로 함</li> </ul>

표 21 ETRI-대학 협력 현황

- 업체의 경우, 연구소기업인 한컴인터프리 및 시스트란과 기반기술 연구를 수행하였다. 특히 한컴인터프리의 경우, 지니톡 자동통역 플랫폼을 여행, 교육, SNS 등 다양한 분야에 응용할 예정이며, 이에따른 다국어 음성언어 기반기술의 성능개선을 위한 협의체를 구성하여 정기적 미팅을 추진하였

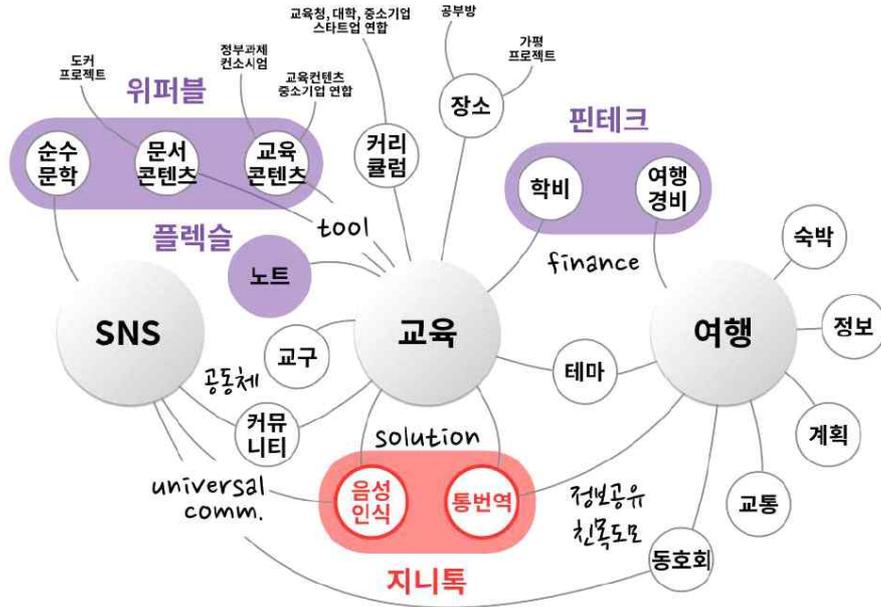


그림 26 지니특 기반 산업생태계 구축 개념도

- 다국어 음향모델 및 언어모델용 음성언어 DB 확장 구축
  - 음성인식 성능 향상을 위해 기본 언어모델링 및 Topic 언어모델링에 필요한 대규모 자막 DB를 구축하였으며 각 언어별 DB구축 규모는 다음과 같다.

언어	문장(만줄)	크기(MB)
영어	30,862	9,658
중국어	11,205	2,204
일본어	2,851	868
프랑스어	6,776	2,182
스페인어	3,025	1,251
독일어	2,534	799
러시아어	10,266	5,096
한국어	28,394	9,910

표 22 다국어 자막 데이터 수집 현황

- 또한 자막 이외에 여행, 맛집 등 통역영역에 활용빈도가 높은 텍스트 대상으로 대용량 블로그 데이터를 수집하였다. 여행, 일상, 맛집 등을 주제로 삼는 블로그를 대상으로 한국어 1,400만 페이지, 일본어 900만 페이지 분량의 원시 데이터를 확보하였으며 이를 정제하여 DB화 하였다. 뿐만 아니라 독일어, 러시아어도 초기 수집 단계에서 20만 페이지 분량의 데이터를 확보하였으며 프랑스어, 스페인어, 러시아어로 확대 중에 있다.
- 그리고 딥러닝용 음향모델링에 소요되는 DB구축을 위해 다국어 방송음성을 대상으로 정제를 통해 대규모 DB 구축을 추진하였다. 자막방송을 실시하는 한국어 방송 7개 채널을 대상으로 2만 시간의 원시 음성 데이터를 수집하였고, 본 연구에서 개발된 음성과 자막 정렬 도구를 이용한다면 약 5천 시간의 순수한 음성 DB를 확보하게 된다. 이러한 경험을 다국어로 확장하여 현재 독일어 2,500 시간, 스페인어 1,700 시간, 프랑스어 2,000 시간, 영어 300시간 분량의 원시 음성 데이터를 수집 완료하였다.

## 7) 평창동계올림픽 테스트 이벤트용 5개 언어 자동통역서비스 구현

### ○ 평창동계올림픽용 자동통역 영역 특화

- 평창동계올림픽 테스트이벤트 시범서비스에 적용하기 위해 한국관광공사에서 제공한 관광공사 외국어 용례표기사전과 평창조직위에서 제공한 올림픽 관련 용어를 대상으로 한국어, 영어 대상 영역특화 단어를 모델에 반영 완료하였다.

**한-스페인어 대관령 양떼목장 시나리오**

스페인어 : Puede recomendar un lugar en que pueda sentir el medio ambiente en vez del centro? ( 시내 말고 자연환경을 많이 느낄 수 있는 곳을 추천 해 주시겠어요?)

한국어 : @대관령 양떼목장@@이라는 곳이 있는데 혹시 알고 계신가요? (Hay un lugar llamado @@대관령 양떼목장@@. ¿Sabe usted?)

스페인어 : Cuando pregunté a otras personas sobre el lugar de turismo antes, todos me recomendaron allí (이전에 다른 사람들한테 관광지를 물어보니까 모두들 이야기하더군요)

한국어 : 자연환경이 좋아서 추천하는 곳이에요. (Todos lo recomiendan porque el medio ambiente es bueno)

스페인어 : ¿Se puede reservarlo aquí? (이곳에서 예약할 수 있나요?)

한국어 : 예. 가능합니다. (Si, es posible)

**한-프랑스어 음식 시나리오**

프랑스어 quelle ville était la meilleure pour vous? (가보신 곳 중에 어디가 좋았습니까?:)

한국어 : 저는 개인적으로 춘천이 좋았어요(Personnellement, j'aime @@Chuncheon@@.)

프랑스어 : Quel plat est célèbre de la ville? (특별히 맛볼만한 음식이 있나요?)

한국어 : 닭갈비가 있어요. 막국수도 유명하고요. (Il y en a Dalkgalbi. Des mak nouilles sont aussi fameuses. )

프랑스어 :Je voudrais les essayer. (먹어보고싶네요)

한국어 : 거리가 좀 멀긴 하지만, 꼭 가보세요. (C'est un peu loin d'ici, mais vous devez visiter la ville )

그림 27 한-스페인어, 한-프랑스어 영역특화 시나리오

- 번역 성능개선을 위한 대용량 대역코퍼스 구축
  - 번역 성능개선을 위해서는 클라우드 소싱(Crowd Sourcing)을 이용하여 다국어 대역코퍼스를 저비용으로 대용량 구축하였다, 구축대상 언어 및 규모는 영어 -> 프랑스어, 스페인어, 독일어, 러시아어 대화체 대역문장 각 20만 (총 80만) 문장과, 한국어 -> 영어, 중국어 각 15만, 일본어 10만 문장으로 총 40만 문장의 대역코퍼스를 구축하였다
- 평창 동계올림픽 테스트이벤트 자동통역 시범서비스 및 평가

- 평창 동계올림픽 테스트 이벤트로 열리는 FIS 스노보드 빅에어 월드컵 (2016.11.23.~26)에서 자동통역 시범서비스를 제공하고 정성평가를 실시하였다. FIS 스노보드 빅에어 월드컵에는 23개국 100여명의 선수를 비롯하여 협회 임원, 관계자, 지원 STAFF 등이 참여했으며 이들 올림픽 패밀리 및 경기 관람 관중들을 대상으로 자동통역 시범 서비스를 제공하였다. 또한 원어민 또는 외국어를 모국어 수준으로 구사하는 이중 언어 화자를 대상으로 정성 평가를 별도로 실시하였다. 이 때 테스트 이벤트에서 활용된 서비스 로그를 위치 정보와 함께 별도 수집하여 이용 양태 분석에 활용하고 있다.



그림 28 평창 동계올림픽 테스트 이벤트 자동통역 시범서비스 모습

- 평창 동계올림픽 테스트 이벤트에 적용된 자동통역 시범서비스는 다음 <그림 29>과 같은 흐름으로 제공되었다.

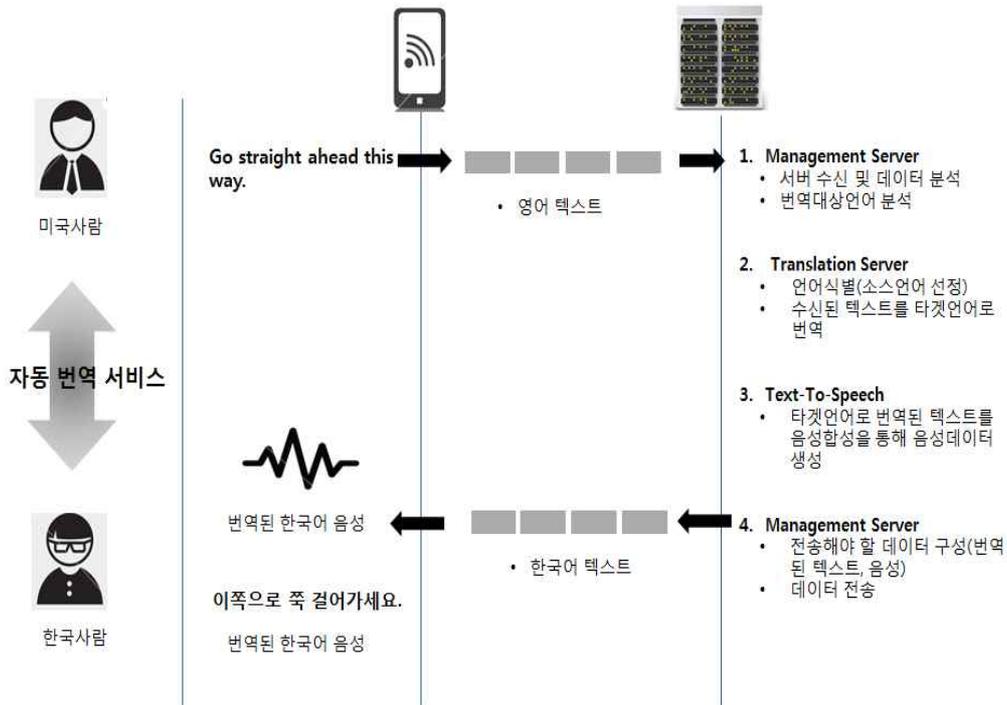


그림 29 자동통역 시범 서비스 흐름도

- 또한 자동통역 서버의 물리적 구성을 그림으로 표현하면 <그림 30>와 같다.

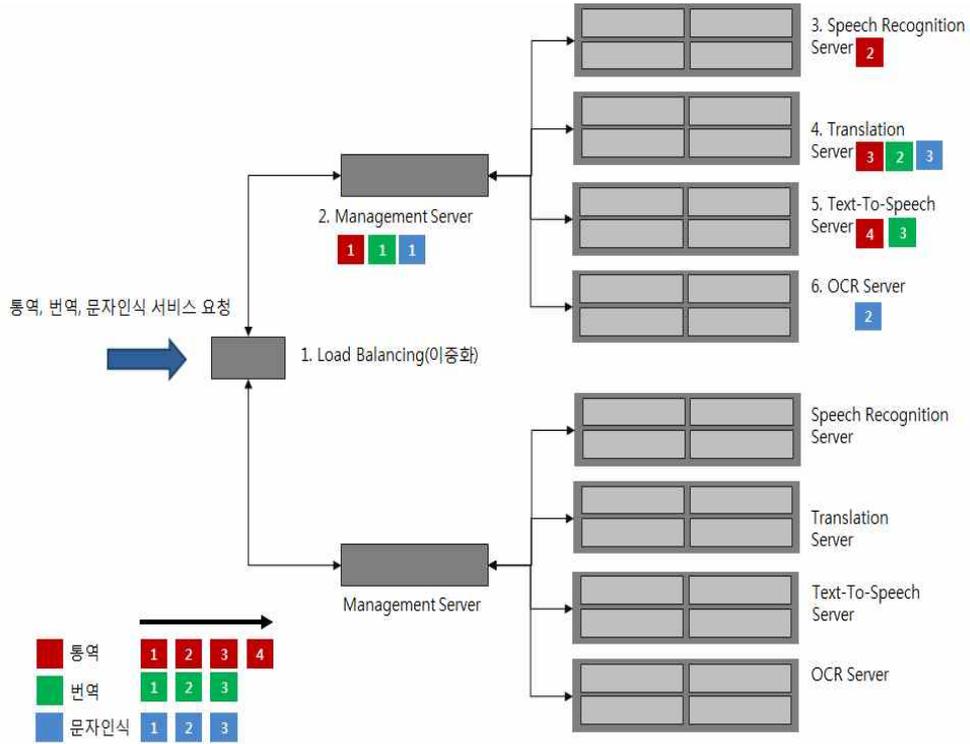


그림 30 자동통역 시범서비스 물리적 구성도

- 이렇게 구성된 자동통역 서비스는 단순히 서비스에 그치는 것이 아니라 로그 데이터를 통하여 언어별, 터치포인트별 필터링을 통해 양상을 분석하고 서비스 제공 결과를 보고할 수 있다.

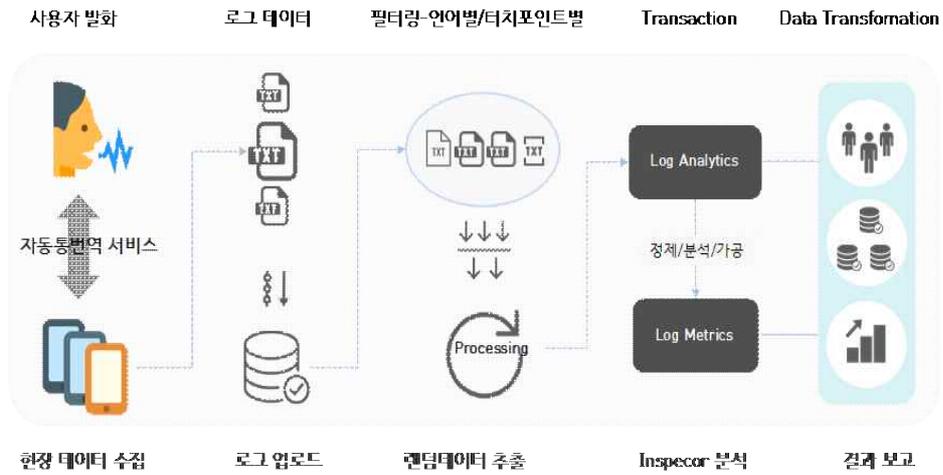


그림 31 자동통역 시범서비스 전체 적용 흐름도

- 다음 <그림 32>에 자동통역 서비스 현황 분석을 위한 트랜잭션 로그 데이터 모니터링 시스템에서 시간별 리퀘스트 현황을 질의한 결과를 예로 보였다.



그림 32 로그 데이터 모니터링 시스템 스크린샷

- 테스트 이벤트 시범 서비스에 대한 자동통역 성능 평가는 크게 두 가지로 이루어진다. 첫 번째로는 테스트 이벤트 기간에 원어민 패널을 동원하여 정성 평가를 실시하는 것이다. 정성 평가로는 터치 포인트별 샘플 예문을 제시한 후 터치 포인트 검증, 서비스 품질(정확도, 사용성, 안정성, 속도, 만족도), 서비스 개선 의견의 3가지 분야별 세부 평가를 실시하였다. 두 번째로는 터치포인트와 일반 관광 예문으로 구성된 평가 예문 DB와 수집된 서비스 로그를 이용하여 정량 평가를 실시할 예정이다(12월). 정량 평가를 위해서 앞서 한국어를 기반으로 터치포인트 4,000 문장, 일반 관광 1,000 문장의 평가 예문을 작성하고 이를 7개 언어로 번역하여 총 4만 문장의 평가 예문 후보 DB를 구축한 바 있다. 이렇게 구축한 평가 예문 4만 문장에서 랜덤 선정한 문장 80% 및 테스트이벤트 기간 중 수집한 실사용 통번역 문장 20% 비율로 평가 데이터를 구성하여 정량 평가를 실시할 예정이다. 정량평가는 언어쌍별 통번역 전문가 3인이 참여하여 LREC (Language Resource Evaluation Conference) 평가 방법에 기반하여 5 단계 수동 평가를 실시하게 된다.(평가 방법과 관련된 자세한 내용은 부록으로 별도 첨부하였다.)
- 그리고 평창 동계 올림픽에 참여하는 외국인들이 원활하게 서비스를 이용하기 위해서는 미리 앱을 설치하고 한국에 도착해야 한다는 점을 고려하여 자동통역 시범 서비스의 확산 및 외국에서의 사용성 평가를 위하여 지난 2월 스페인 바로셀로나에서 열린 Mobile World Congress 2016에 참가하였다. ETRI 지니톡 기술을 이전받은 한컴인터프리에서 부스를 개설하였고 일평균 150여명의 방문객이 부스에 방문하여 관심을 보였으며 이들은 일상회화에 있어 스페인어 음성 인식률이 높다는 반응을 보였다.



그림 33 MWC 2016 시연 모습

- 또한 실제 현지에서 사용성 평가를 진행해 본 결과 전반적으로 일상 회화에서 의사 소통을 하는데 큰 무리 없는 모습을 보였다. 다만 이 과정에서 주변의 소음에 민감한 부분에 대해 조절이 필요하다는 결론을 얻어 UI에 대한 개선을 실시하였다. 그리고 서버형으로 원격 서비스가 됨에 따라 로밍으로 인한 통역 속도를 공항, 버스 안, 지하철 행사장 등 여러 장소에서 테스트를 실시해본 결과 음성 입력 종료 뒤 3~4초 후에 결과가 나오는 등 국내 속도 대비 다소 느린 감이 있었으나 테스트가 3G 로밍 상태에서 이루어진 점을 감안한다면 실제 서비스에는 별다른 지장이 없을 것이라는 결론을 얻었다.



그림 34 스페인 현지에서의 실제 사용 장면

나. 정량적 추진실적

1) 논문/특허

○ 논문

번호	명 칭	게재일 (또는 제출일)	게재지명
1	(SCI) Incremental layer Transfer adaptation for large vocabulary speech recognition using deep neural networks(제출)	2016.04.28.	IEEE Signal Processing Letter
2	(연구재단 등재지) 합성곱 신경망을 이용한 On-Line 주제 분리	2016.10.13.	정보처리학회논문지(소프트웨어 및 데이터 공학)
3	(국내 학술대회) 정제된 방송 음성 데이터를 이용한 음성인식기 성능 향상	2016.11.18	한국음성학회 2016 가을 학술대회 논문집

표 23 논문리스트

○ 특허 (빨간색은 표준특허로 반영)

번호	명 칭	등록일 (또는 출원일)	국명
1	자동 번역 및 통역 장치 및 그 방법(등록)	2016.03.22	미국
2	다중음향모델을 이용하여 음성을 인식하기 위한 장치 및 그 방법(등록)	2016.06.28	미국
3		2016.02.25	미국
4		2016.05.05.	미국
5		2016.06.21.	미국
6		2016.06.24.	미국
7		2016.06.27.	미국
8		2016.07.06.	미국
9		2016.07.19.	미국
10		2016.01.04.	대한민국
11		2016.01.05.	대한민국
12		2016.01.05.	대한민국
13		2016.01.05.	대한민국
14		2016.05.25.	대한민국
15		2016.06.20.	대한민국

16		2016.09.09.	대한민국
17		2016.10.07.	대한민국

표 24 특허리스트

## 2) 표준화 성과

번호	표준기고서 제목	표준화 기구
1	face to face speech translation system	ISO/IEC JTC1 SC35 User Interface

## 3) 기술이전

번호	기술명	이전일자	이전업체	기술료(천원)
1	영어 음성인식 탐색 네트워크 생성 기술	2016.11.07	네오스마트링크	110,000
2	한국어 공통음성 DB 중 음성인식용 문장(대화체) DB	2016.02.25.	삼성SDS	6,072
3	한국어 공통음성 DB중 음성인식용 DB 2종	2015.12.11.	사운드잇	10,560
				1.26억원 (부가세 포함)

표 25 기술이전리스트

## 6. 활용(산업화) 방안

### 가. 기술평가

구 분	구체적인 내용 (과학적/기술적 원리 및 응용과정과 관련된 내용)
기준(선행)기술과 비교하여 유리한 점	<ul style="list-style-type: none"> <li>○ 국내외 업체에 자동통번역 공통플랫폼 및 Open API를 지원하여 현재 ETRI 조직체제로 해결하지 못하는 실서비스 환경 대응량 로그데이터 축적을 가능하게 해줌으로서 기술적 성능 개선이 가능함</li> <li>○ 글로벌 시장에 진출하기 위한 다국어 음성언어 기반 기술을 개발함으로서 구글, 애플(뉴앙스) 등과 같이 국제적 수준의 기술 확보가 가능함</li> </ul>
기준(선행)기술과 비교하여 불리한 점	없음

### 나. 활용(상용화) 가능성

구 분	구체적인 단계 및 내용
기술의 응용분야 및 활용방법	<ul style="list-style-type: none"> <li>○ 자동통번역 기술은 최근 세계화에 따른 언어소통 불편을 해소하는 기술로, 스마트폰 통역, 통역전화기, 통역 로봇, 교육용 SW, 비즈니스용 통역 등 응용분야가 매우 다양함</li> <li>○ 특히 자동통역기술은 음성인식, 합성, 자동번역 등 핵심 HCI 기술로 구성되어 있어 최근 이슈가 되고 있는 애플 siri, 스마트 TV 등 대화형 음성인터페이스에도 직접 활용할 수 있는 기술임</li> </ul>
적용상의 애로점과극복(개선) 방안	<ul style="list-style-type: none"> <li>○ 잡음환경에서의 자동통번역기 성능 저하는 오류를 빈번히 발생시켜 사용상의 불편을 가져올 수 있음</li> <li>○ 자동통번역기의 오류를 사용자가 수정할 수 있는 멀티모달 기반의 UI/UX 기술이 보완되어야 타 기술과 차별성을 가질 수 있을 것임</li> </ul>
제품/서비스의 예상 수요자(층)	<ul style="list-style-type: none"> <li>○ 예상수요자               <ul style="list-style-type: none"> <li>• 모바일 기기 동시통역 SW 및 내외국인 입/출국자를 대상으로 한국어를 중심으로 2015년 영어, 중국어 일본어에 대한 통역 세계시장 631억원(점유율 1%) 점유, 다국어 확장에 의해 2020년 7조원(점유율 45%) 점유로 세계 1위 가능</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>○ 수요업체 <ul style="list-style-type: none"> <li>• 스마트폰 제조업체: 삼성전자, LG전자, HTC(대만), 도시바(일본) 등</li> <li>• 포털업체: 다음, NHN, 파란 등</li> <li>• 통신사업자: SKT, LG U+, KT 등</li> <li>• 중소기업: 파인디지털, 아이리버, 샤프(일본) 등 업체</li> </ul> </li> </ul>
<p>제품/서비스화하여 시장에 도입되기까지 요구되는 시간의 정도</p>	<ul style="list-style-type: none"> <li>○ 자동통번역 기술은 기술개발 단계로는 발아기(Technology Trigger)이나 2015년 초에 자동통역기술 실용화가 활발하게 이루어 질 것으로 예상함</li> <li>○ 현재 자동통번역 시장은 세계적으로 기술개발 후 시장이 형성되고 있는 초기단계에 해당</li> </ul>

다. 시장 및 경쟁

구분	구체적인 내용
○내수시장 창출효과	<ul style="list-style-type: none"> <li>○ 해당 기술의 국내 시장(생산기준)은 2015년 약 720억 원 수준에서 2020년에는 약 1,328억 원 정도가 될 것으로 전망, 연평균 21.2%의 성장추세</li> <li>○ 국내 휴대형 한/영 자동통번역 산업의 경우, 전체의 경제적 파급효과는 '13~ '20년 누적 1.5조원의 생산유발, 6,740억 원의 부가가치유발, 5,995명의 고용유발 효과 기대</li> <li>○ 다국어로 확장할 경우, 한/영 자동통역 시장의 10배 이상 시장창출 효과가 있을 것으로 예상</li> </ul>
○수출효과/ 수입대체 효과	<ul style="list-style-type: none"> <li>○ 자동통번역 세계시장은 2010년 5.8억불 수준에서 2020년에는 44.6억불 정도가 될 것으로 전망, 연평균 22.7%의 성장추세</li> </ul>
○잠재적/ 현재적 경쟁자와 그들이 갖고있는 기술/제품 특성	<ul style="list-style-type: none"> <li>○ 현재 경쟁자는 구글, 애플, 닐앙스임</li> <li>○ 구글의 경우, 40개 언어 이상 다국어 자동통역 기술 확보하였음. 대용량 서비스 로그데이터를 기반으로 점진적으로 성능을 개선하고 있음</li> <li>○ 애플의 경우, Siri라는 영어기반 대화형 인터페이스 기술을 확보하였고, 조만간 다국어로 확장중임</li> <li>○ 닐앙스는 20여개 언어 대상 음성인식 기술을 확보하여 음성인식 시장을 독점하고 있음</li> </ul>

라. 이전가능한 기술목록

기술이전 목록	주요 핵심기술내용	이전여부 및 시기	이전형태
다국어 음성인식 기술	한국어, 영어, 일본어, 중국어, 스페인어, 불어 등 다국어 음성인식 기술	2015.06	SW
다국어 자동통역 기술	한/영, 한/일, 한/중 자동통역 기술	2015.09	SW
음성언어 DB	음향모델용 음성DB 및 언어모델용 텍스트DB	2016.11	SW

※ 이전형태는 특허, 설계도, 회로도, 공정도, S/W, 시제품 등을 지칭함

마. 자동통역 응용서비스의 상용화 계획

구분	구체적인 내용
○ 형태/규모	<ul style="list-style-type: none"> <li>○ 본 과제의 결과물을 상용화하기 위한 상세 사업계획을 수립하고, 상용 제품 개발 인력을 별도 투입하여 전담 마케팅 인력 투입과 별도의 마케팅 비용을 책정할 계획</li> <li>○ 실 서비스 사업 방안 도출은 과제 진행 중 전문 인력을 통해 수행 예정</li> </ul>
○ 상용화 능력 및 자원보유	<ul style="list-style-type: none"> <li>○ 시스템인터내셔널, 한컴인터프리, 디오텍 등 국내업체는 자동통번역 소프트웨어 개발, 판매 및 서비스사업을 수행해 왔으며, 현재 다음과 같은 상용화 전문인력 및 관련자원을 확보하고 있음</li> </ul>
○ 상용화 계획 및 일정	<ul style="list-style-type: none"> <li>○ 2016년                             <ul style="list-style-type: none"> <li>• 통역 응용 앱 시장성 조사</li> <li>• 플랫폼, API 구현</li> <li>• 중소기업 BM 프로토타입 서비스 구현</li> </ul> </li> <li>○ 2017년                             <ul style="list-style-type: none"> <li>• 통역 응용 앱 개발을 통한 통역시범 운용 후 사업성 평가</li> </ul> </li> </ul>

## 7. 추진체계 및 전략

### 가. 추진체계



그림 35 추진체계

### 나. 추진

- 미래부, 문체부, 조직위, 업체와 공동협력하여 평창 동계올림픽 자동통역서비스 성공적 실시 준비
- 평창 동계올림픽 조직위에서 추진 중인 테스트이벤트에 참가하여 자동통역서비스의 품질을 제고하고 시행착오를 줄여 성공적 자동통역 서비스에 기여
- 대용량 다국어 음성언어데이터 자동정제 기술 개발을 통해 자동통역 엔진의 성능 개선 추진
- 다국어 사용자 로그데이터 확보를 위해 시스템, 한컴인터프리 등 서비스업체

와 협력

- 중소기업 시제품 지원을 통해 개발한 자동통역 시제품을 사업화하고, 모범적인 상용화 사례를 구축하여 IT분야 신시장 창출
- 지니톡 Open API 기반 자동통역 플랫폼을 국내업체에 홍보하여 산업생태계 구축 추진
- ‘핸즈프리 자동통역’ 기술에 대해 ITU-T에서 일본 NICT와 공동협력하여 국제표준으로 시도
- UST 석박사 과정을 사업에 적극 참여시켜 등 연구원과 협업을 통한 사업의 효과적인 추진
- 다학제간 협업을 위해 서울대, KAIST, 경북대, 부산대 등 대학과 연계하여 기술 개발 추진

## 8. 기대성과

### 가. 기술적 측면

- 자동통번역 기술은 대화체 음성인식, 언어번역, 음성합성 등 요소기술이 어우러진 복합기술로서, 이 요소기술들은 지능형 로봇, 텔레매틱스, 디지털 홈 등 IT성장동력산업 전 분야에서 **HCI(Human Computer Interaction)의 핵심요소**로 요구되는 원천 기술임
- 자동통번역 세계시장 선점을 위한 국가간 기술개발 경쟁 치열한 상황임. 다국어 언어처리 기반 기술 개발을 통해 향후 유럽어, 아시아권 언어 등으로 자동통번역 및 음성인식기술을 확장할 수 있는 기반 기술이 확보됨

### 나. 경제 산업적 측면

- 현재 자동통번역 관련 중소기업이 4~5개 정도임. 본 사업을 통해 매년 2~3개

- 중소기업을 지원하여 2018년에는 10여개 강소형 중소기업이 육성되고, 기존 자동통번역 업체는 글로벌 경쟁력을 가질 수 있는 중견기업으로 육성될 것으로 기대
- 본 사업을 통해 구축한 리소스의 공동활용으로 자동통번역 기술개발 기간 단축 및 DB구축에 소요되는 비용을 절감하여 자동통번역산업의 국제경쟁력이 강화될 것으로 기대
  - 2012년의 경우, 우리나라 출국자 2,400만 명, 입국외국인 670만 명으로 추정됨. 따라서 해외출입국자 3천만 명의 20%가 자동통역 서비스(이용료 1만원)를 이용할 경우, 연 600억원의 서비스 시장이 창출됨
  - 2020년 전세계 자동통역 시장규모는 자동통역 단말기 6.3조원, 비즈니스용 자동통역 2.5조원, 교육용 자동통역 1.2조원 등 총 10조원으로 예측 (출처: 일본 UFJ총합연구소, 2006)
  - 경제적 파급효과는 중립적 시나리오 하에, '13~' 20년 누적 약 2,497억 원의 생산유발, 1,110억 원의 부가가치유발, 987명 수준의 고용유발 효과가 예상
  - - ※ 국내 휴대형 한/영 자동통번역 산업 전체의 경제적 파급효과는 '13~ '20년 누적 1.5조원의 생산유발, 6,740억 원의 부가가치유발, 5,995명의 고용유발 효과 기대  
(단위: 억원/명)

구분		2013	2014	2015	2016	2017	2018	2019	2020	합계
생산유발	낙관적	163	277	381	531	703	513	278	150	2,997
	중립적	136	231	318	443	586	427	232	125	2,497
	비관적	108	185	254	354	469	342	185	100	1,998
부가가치유발	낙관적	72	123	170	236	313	228	124	67	1,332
	중립적	60	103	141	197	261	190	103	56	1,110
	비관적	48	82	113	157	208	152	82	44	888
고용유발	낙관적	64	109	151	210	278	203	110	59	1,185
	중립적	54	91	126	175	232	169	92	49	987
	비관적	43	73	101	140	185	135	73	40	790

\* 출처: 한국은행(2008) 산업연관표(서비스업-통신 및 방송)를 활용하여 산출  
 \* 생산유발계수 : 1.883, 부가가치유발계수 : 0.837, 고용유발계수 : 7.444

## 9. 향후계획

- 세계최고 중국어 음성인식기술 보유 업체인 iFlytek 대비 동등수준 중국어 인식 엔진 성능 확보. 국내업체가 중국어 음성인식 시장에 진출할 경우, 경제적으로 산업적으로 큰 파급효과가 있을 것으로 예상됨
- 시스트란에 단말탑재형 자동통역기 기술이전을 통해 연말 상용화 추진
- 업체 기술지원 및 사업화 (현재 진행중인 업체: 시스트란, 헬로챗, 한글과컴퓨터, 스마트비투엠, 에버트란, SK텔링크, 국과수, LexiFone(이스라엘 통역서비스 업체) 등
- 차년도 이후에는 국내업체의 글로벌 경쟁력 확보를 위해 자유발화 일상대화형 동시통역이 가능한 세계최초 원천기술 개발에 도전할 예정이며, 2018년 평창동계 올림픽에서 시험적용하도록 차년도 수행사업계획서(2017~2018)에 반영예정임

## 10. 부록

### 가. 번역이해도/통역성공률 평가방안

#### 1. 평가방법

- 가. 평가 문장 풀(Pool)로부터 평가 대상 분야의 평균 문장길이를 고려하여 원문 선정
- 나. 외부용역 기관 5~7명의 전문번역가에 의한 객관적인 평가: 엔진 버전별 결과 혼합, blind 평가
- 다. 평가기준에 따른 각 문장별 0~4점 스코어링
- 라. 각 문장별로 최고, 최저 점수를 제외한 평균점으로 합산

#### 2. 산출방법

- 가. 번역이해도/통역성공률 (%) = 3점이상문장수/평가문장수 x 100
- 나. 예: 80/100 x 100 = 80% (평가 대상 100문장 중 3점 이상인 문장 수가 80문장인 경우)

#### 3. 평가 점수 부여 기준

점수	평가 기준
4.0	원어문의 의미가 그대로 전달된 경우
3.5	복문에서, 문장의 동사구가 정확히 전달되어 문장의 전체적인 의미의 골격이 전달되지만 동사를 제외한 1-2단어의 대역어가 잘못된 경우
3.0	문장의 동사구가 정확히 전달되어 문장의 전체적인 의미의 골격이 전달되는 경우
2.5	하나의 동사절이라도 정확히 번역되어 부분적으로 문장의 의미를 전달할 경우
2.0	하나 이상의 구가 정확히 번역되지만 전체적인 문장의 의미를 파악하기 어려운 경우
1.0	문장 중에 하나의 단어 또는 구라도 정확히 번역된 경우
0.0	번역문 출력이 안 된 경우

4. 출처:

- DARPA(Defense Advanced Research Projects Agency) 1994 adequacy test(Doyon, Taylor, and White, 1996)
- Workshop at the LREC 2002 Conference: Machine Translation Evaluation: Human Evaluators Meet Automated Metric

## 나. 음성인식률 산출 방법

### 1. 평가방법

- 가. 음향모델이나 언어모델 훈련에 사용되지 않은 음성신호 데이터를 1,500문장 이상 선정
- 나. 발화단위로 음성인식엔진에 입력하고, 인식 결과를 파일로 저장
- 다. 평가데이터 원문과 인식결과를 대응시켜 아래와 같이 음성인식률 산출

$$\text{인식률} = \frac{N - (S + D + I)}{N}$$

N: 전체 평가데이터에 포함된 단어 수

S: 원문의 단어와 달리 다른 단어로 인식된 단어의 수

D: 원문에는 존재하나 인식되지 않고 건너 뛴 단어의 수

I: 원문에는 존재하지 않으나, 인식결과에 추가로 포함된 단어의 수

### 2. 산출방법

- 가. 단어인식률의 경우, 레퍼런스와 인식결과간 단어대 단어간 비교하여 일치하는 개수를 카운트하여 전체 단어 대비 맞는 단어개수의 비율을 백분율로 측정
- 나. 문장인식률의 경우, 문장을 구성하는 단어가 모두 정인식 되었을 때 정인식 문장 개수를 카운트하여 전체 문장개수 대비 맞는 문장 개수의 비율을 백분율로 측정

다. 약어표

약어	의미
ERR	Error Reduction Rate
Open API	Open Application Program Interface
BM	Business Model
DARPA	Defense Advanced Research Projects Agency
ASP	Application Service Provider
GALE	Global Autonomous Language Exploitation
HCI	Human Computer Interaction
C-STAR	Consortium for Speech Translation Advanced Research
HMM	Hidden Markov Model
MID	Mobile Internet Device
MASTOR	Multilingual Automatic Speech-to-Speech Translator
PDA	Personal Digital Assistant
PMP	Portable Media Player
SNLP	Speech and Natural Language Processing
SRI	Stanford Research Institute
TALES	Translingual Automatic Language Exploitation System
TC-STAR	Technology and Corpora for Speech to Speech Translation
UMPC	Ultra Mobile Personal Computer
BMT	Bench Mark Test
UI	User Interface
OTG	On-The-Go
SMT	Statistical Machine Translation
LM	Language Model
TM	Translation Memory
DNN	Deep Neural Network
EPD	End-Point Detection

## 주 의

1. 이 연구보고서는 한국전자통신연구원의 주요사업으로 수행한 연구 결과입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 한국전자통신연구원에서 수행한 주요사업 결과임을 밝혀야 합니다.