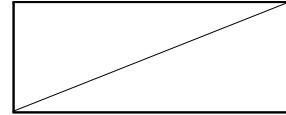


2013년 12월

13ZS1100-01-7040P



언어장벽 없는 국가 구현을 위한 자동통번역 산업 경쟁력 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 확대하는 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.

2013년 12월

한국전자통신연구원 원장 김 홍 남

제 출 문

본 연구보고서는 주요사업인 "언어장벽 없는 국가 구현을 위한 자동통번역 산업 경쟁력 강화 사업"의 결과로서, 본 과제에 참여한 아래의 연구팀이 작성한 것입니다.

2013년 12월

연구책임자 : 책임연구원 김상훈 (자동통역인공지능연구센터)
연구참여자 : 책임연구원 황승구 (소프트웨어 연구부문)
책임연구원 박상규 (자동통역인공지능연구센터)
책임연구원 박 준 (자동통역인공지능연구센터)
책임연구원 이수중 (자동통역인공지능연구센터)
책임연구원 김정세 (자동통역인공지능연구센터)
선임연구원 김승희 (자동통역인공지능연구센터)
선임연구원 왕지현 (자동통역인공지능연구센터)
선임연구원 김영익 (자동통역인공지능연구센터)
선임연구원 김동현 (자동통역인공지능연구센터)
선임연구원 윤 승 (자동통역인공지능연구센터)
선임연구원 최무열 (자동통역인공지능연구센터)
연구원 김기현 (자동통역인공지능연구센터)
책임연구원 김영길 (자동통역인공지능연구센터)
책임연구원 최승권 (자동통역인공지능연구센터)
선임연구원 김창현 (자동통역인공지능연구센터)
선임연구원 서영애 (자동통역인공지능연구센터)
연구원 나승훈 (자동통역인공지능연구센터)
책임연구원 최미란 (자동통역인공지능연구센터)
책임연구원 김기호 (자동통역인공지능연구센터)
연구원 이철순 (자동통역인공지능연구센터)
연구원 김윤영 (자동통역인공지능연구센터)
연구원 이민규 (자동통역인공지능연구센터)
연구원 이담허 (자동통역인공지능연구센터)
연구원 나다니엘 (자동통역인공지능연구센터)
연구원 안재형 (자동통역인공지능연구센터)
연구원 장지수 (자동통역인공지능연구센터)

요 약 문

I. 제 목

언어장벽 없는 국가 구현을 위한 자동통번역 산업 경쟁력 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 2014년까지 한, 중, 영, 일 4개국 통역서비스 제공을 목표로 하고 있으며, 본 서비스는 2014년 아시안게임, 2018년 평창 동계올림픽 지원에도 활용될 예정이다. 특히 대국민 자동통역 서비스를 통해 축적한 사용자 로그데이터는 자동통역 기술의 성능 개선에 직접 활용될 예정이고, 이로부터 최근 구글 등 다국적 기업에 의한 국내 산업을 보호하고, 산업 경쟁력을 강화하고자 한다. 또한 본 사업을 통해 고품질의 자동통역 기술 기반 플랫폼 및 Open API를 국내 업체, 연구소, 대학 등에 지원하여 생태계 구축도 추진하고 있다.

III. 연구내용 및 범위

- 대국민 한/일 자동통역 시범서비스 실시
- 자동통번역 플랫폼 기반 Open API 구현 (1차)
- 사용자 로그데이터 기반 한/일 자동통역 성능 개선 (통역률 ERR 20% 개선)
- 중소기업 시제품 제작 지원 (2차)
- 다국어 음성언어 연구기반 조성 (1차)

- 2014년 인천 아시안게임 대비 한,중,영,일 4개국 자동통역 시범서비스 구현

IV. 연구결과

- 한/일, 한/중 자동통역 대국민 자동통역 시범서비스를 실시함 (창조경제 국무총리상 수상, 안전행정부 공공앱 최우수상 수상)
- 사용자 로그DB 정제를 통한 한/일 자동통역 엔진 성능 개선 (ERR 20% 달성)
- Open API 기반 자동통번역 플랫폼을 구축하여 국내 업체를 대상으로 연구성과 확산. 금년에 9개 업체 지원을 완료함
- 중소기업 시제품 제작지원 사업의 일환으로 금년에 총 3개 업체를 선정 지원하여 자동통역 비즈니스 모델 발굴
- 업체, 대학과 Open R&D를 수행. 국내외 특허 10건 (국제 5건, 국내 5건) 출원 완료
- 업체와 공동으로 순천 정원박람회 등 공공목적용 자동통역서비스 실시 지원 (2013.4~10)

V. 기대성과 및 건의

- '12년 여수 엑스포, '14 인천 아시안게임, '18 평창 동계올림픽에 자동통번역서비스 실시를 통해 언어장벽 없는 국가 실현 및 국격 제고
- ETRI의 자동통번역 플랫폼 및 Open API를 적극 활용하여 새로운 BM 개발 기회가 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 것임

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 till 2014, and support the 2014 Incheon Asian Games as well as 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 the Korean-Japanese automatic speech/text translation trial public service
- Implementation of the automatic speech/text translation platform and Open API (1st phase)
- Improvement of Korean-Japanese translation technology performance based on user log data (20% error reduction rate)

- Support business model creation and application development of small and medium enterprises (2nd phase)
- Set-up of the multilingual spoken language research infrastructure(1st phase)
- Implementation of the four major language (Korean, Chinese, English, Japanese) speech translation system prototype for 2014 Incheon Asian Game

IV. RESULTS

- Opened Korean-Japanese and Korean-Chinese automatic speech translation trial service (Creative Economy Prime Minister Award, The best public App. Award/Ministry of Security and Public Administration)
- Performance improvement of Korean-Japanese translation technology performance based on user log data (20% error reduction rate)
- Set up the automatic speech translation service platform, and supported its open API to 9 enterprises.
- Supported 3 enterprises for building new business model based on automatic speech translation as the speech translation service prototype development supporting project
- Performed Open R&D with industry and university. Applied 10 patents (5 domestic; 5 international)
- Supported the automatic speech translation public service at Suncheon Garden Expo in cooperation with the industry (2013.4~10)

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 Introduction	19
SECTION 1 Importance of the project	19
SECTION 2 Necessity of the project	20
SECTION 3 Expected effects	21
SECTION 4 Technology trend	23
SECTION 5 Key components and approaches	27
CHAPTER 2 Targets and Scope of the project	30
SECTION 1 Final target	30
SECTION 2 Targets and scope per each year	31
SECTION 3 Outcomes of the project	34
SECTION 4 Roles of participating parties and time table	37
CHAPTER 3 Current term targets and results	40
SECTION 1 Current term targets	40
SECTION 2 Development results	42
CHAPTER 4 Industrialization plan and expected effects	62
SECTION 1 Industrialization plan	62
SECTION 2 Expected effects	67
[Appendix 1] Evaluation method for translation success rate	70
[Appendix 2] Evaluation measure for speech recognition	71
[Appendix 3] Acronym table	72

Table List

Table 1 Competitiveness of automatic translation technology	23
Table 2 Project target	30
Table 3 Annual project targets	31
Table 4 Evaluation measures	34
Table 5 Major outcomes (Summary)	42
Table 6 Automatic translation technology evaluation	62
Table 7 Potential application of automatic translation technology	62
Table 8 Market trend and competition status of automatic translation	64
Table 9 Commercialization plan	65
Table 10 Current status of trades	65
Table 11 Major customer	66
Table 12 Market volume estimation	69

Figure List

Figure 1 Automatic translation service road map (2012~2018)	22
Figure 2 Automatic translation system	25
Figure 3 Implementation plan of automatic translation industry eco-system	28
Figure 4 Long term project road-map (2012~2018)	30
Figure 5 Roles of participating parties of the current term	37
Figure 6 GenieTalk article on Yonhab News	49
Figure 7 GenieTalk article on Digital Times	50
Figure 8 GenieTalk article on YTN	51
Figure 9 Japanese Homepage of the Korea Tourist Service, Inc.	52
Figure 10 New Japanese version of Korean-Japanese GenieTalk	53
Figure 11 New Korean version of Korean-Japanese GenieTalk	54
Figure 12 GenieTalk promotion at JATA	55
Figure 13 Web page for automatic speech translation Open API approval/management	56
Figure 14 Demonstration of the Updated Korean-English GenieTalk	56
Figure 15 App. of the GenieTalk-based English learning service for Korean (Confitech, Inc.)	57
Figure 16 App. of the destination-specialized GenieTalk for oversees travelers (Scoopmedia, Inc.)	57
Figure 17 App. of the KTX domain-tailored GenieTalk (Digiquitous, Inc.)	58
Figure 18 Newly developed app. of Korean-Chinese GenieTalk	58

목 차

제 1 장 서론	19
제 1 절 사업 중요성	19
제 2 절 사업추진 필요성	20
제 3 절 사업결과 기대효과	21
제 4 절 국내·외 관련 기술의 현황	23
제 5 절 핵심요소 및 접근방법	27
제 2 장 연구개발 목표 및 내용	30
제 1 절 최종목표	30
제 2 절 연차별 연구개발 목표 및 내용	31
제 3 절 성과목표	34
제 4 절 추진체계 및 일정	37
제 3 장 당해연도 목표 및 수행결과	40
제 1 절 당해연도 목표	40
제 2 절 사업수행 결과	42
제 4 장 사업화 계획 및 기대 효과	62
제 1 절 활용방안 및 사업화 계획	62
제 2 절 기대효과 및 파급효과	67
[부록 1] 번역이해도/통역성공률 평가방안	70
[부록 2] 음성인식률 산출 방법	71
[부록 3] 약어표	72

표 목차

표 1 자동통역 기술의 국내외 수준 비교	23
표 2 사업목표	30
표 3 연차별 연구목표	31
표 4 성과지표	34
표 5 주요 추진 실적 (요약)	42
표 6 자동통역 기술 평가	62
표 7 자동통역 기술의 활용가능성	62
표 8 자동통역 기술의 시장 및 경쟁현황	64
표 9 상용화 계획	65
표 10 국내외 수출입 현황	65
표 11 주요 수요처 현황	66
표 12 국내외 시장규모 예상	69

그림 목차

그림 1 자동 통번역 對국민 서비스 로드맵 (2012~2018)	22
그림 2 통역시스템	25
그림 3 자동 통번역 산업생태계 구축방안	28
그림 4 자동통번역 산업경쟁력 강화 사업의 로드맵 (2012~2018)	30
그림 5 당해연도 사업 추진체계	37
그림 6 연합뉴스 지니톡 기사	49
그림 7 디지털타임즈 지니톡 기사	50
그림 8 YTN 지니톡 기사	51
그림 9 한국관광공사 일본어 홈페이지 홍보 내용	52
그림 10 신규 일본향 한/일 자동통역 ‘지니톡’앱	53
그림 11 업데이트된 국내향 한/일 자동통역 ‘지니톡’ 앱	54
그림 12 일본 세계여행박람회(JATA) ‘지니톡’ 홍보	55
그림 13 자동통역 Open API 신청/승인 관리 화면	56
그림 14 업그레이드 버전 국내향 한/영 자동통역 ‘지니톡’ 앱 시연 장면	56
그림 15 지니톡 기반 한국인 대상 영어 교육용 앱 서비스 구현 화면 (지원업체: 컨피테크)	57
그림 16 해외 여행객을 대상으로 방문도시에 특화가 용이한 자동통 역 서비스 구현 화면 (지원업체: 스크미디어)	57
그림 17 KTX 역내 매표창구에서 언어소통이 가능한 통역 서비스 구현 화면 (지원업체: 디지쿼터스)	58
그림 18 신규 한/중 자동통역 대국민 서비스용 앱	58

제 1 장 서론

제 1 절 사업 중요성

1. 세계화에 따른 언어장벽 문제 해소

- 가. 세계화(Globalization)의 가속화로 국가간 인적, 물적 교류가 활발해지고 있으며, 이에 언어간 장벽을 허무는 자동통역기술의 확보는 무한 경쟁시대에서 국가경쟁력과 직결됨
- 나. 최근 Google, Apple 등 다국적 기업의 모바일용 OS 플랫폼 사례와 같이 다국어 음성언어처리 SW 또한 중속될 위기상황에서 국가적 차원의 지속적인 투자 필요
- 다. Google은 2012년 다국어 자동통역서비스 추진 발표 (2010.2, 타임지). IBM은 5년내 상용화 기술 중 자동통역을 파급효과가 가장 큰 기술로 선정 (2007년). 미국, EU, 일본 등 여러 선진국에서는 자동통역 시장의 중요성을 인식하고 정책과제를 통해 장기적인 연구 지원을 하고 있음

2. 정보통신 산업 블루오션 창출

- 가. 여행자를 위한 모바일 한/영 자동통역 기술은 기존 시장의 대체가 아닌 새로운 서비스와 제품 시장을 창출할 것임
- 나. 글로벌 기업간 비즈니스가 활발해짐에 따라 국적과 언어가 다른 회사들의 의사소통 문제가 크게 부각되고 있음
- 다. 이를 해결해 줄 수 있는 실시간 통번역 기술이 조기 실용화 성공가능성이 가장 큰 기술로 선정됨 (IBM 선정 5대 유망 IT 원천기술, 출처: IDC 뉴스 2007.2.4.)
- 라. 2020년 일본내 자동통번역 시장 10조원 예측 (출처: 일본 UFJ총합연구소, 2006). IT 분야는 물론 교육, 관광, 문화, 의료 등 타 산업으로의 파급이 큰 기술
- 바. 새로운 시장의 모델 정립에 대한 부담으로 산업체가 독자적으로 관련

기술을 개발하기 어려우므로, 정부의 주도로 자동통역 핵심 기술기반을 구축하고, 여러 산업체들이 이를 활용하여 자동통역 시장을 창출, 공략해야 할 것임

3. 자동통역 기술 한계 극복

- 가. 최근 자동통역 기술의 중요성이 부각되고 있으나, 지속적인 연구개발 노력에도 불구하고, 아직 대화체 음성인식 및 대화체 문장번역 등 요소기술의 성능이 사용자의 요구수준을 만족시키지 못하여 실용화가 계속 지연되고 있으며, 기술요소의 융합 및 지식기반 접근방식으로써 이러한 기술적 한계의 극복이 절실히 요구되고 있음
- 나. 자동통역 기술은 대화체 음성인식, 언어번역, 음성합성 등 요소기술이 어우러진 복합기술로서, 이 요소기술들은 지능형 로봇, 텔레매틱스, 디지털홈 등 IT성장동력산업 전 분야에서 HCI (Human Computer Interaction)의 핵심요소로 요구되는 원천 기술임
- 다. 자동통역 기술은 아직 미개척 분야로서 적기에 연구 개발을 추진하여 이에 대한 기술 경쟁력을 확보할 수 있음
- 라. 1999년 미국, 일본, 독일 등 선도 연구기관간의 국제공동 실시간 시연을 통하여 자동통역기술의 실용화 가능성이 확인된 이래, 선진국 중심으로 기술개발이 진행되고 있으며, 거대한 대기 시장 선점을 위하여 지적재산권 확보가 시급한 상태임
- 마. 특히, 대화체 자동통역 기술의 핵심 기술인 비정형 문장의 분석, 생략/조용처리 기술 등은 세계적으로도 실험성이 강한 기반 기술로서 관련 지적재산권의 조기 확보가 필요함

제 2 절 사업추진 필요성

- 가. 단기과제, 단위기술 개발 위주
 - 1) 단기목표 중심 기술개발 후 기술이전으로 사업이 종료함에 따라 지속 가능한 음성언어 기술개발이 어려움
 - 2) 최근 SW에 공격적으로 투자를 하고 있는 Google, Apple, IBM 등 선진

외국의 다국적 기업에 OS, DBMS 사례와 같이 음성언어처리 기술 또한
중속될 위기 상황에서 국가적인 차원의 장기적인 투자가 필요

나. 사용자 데이터 수집을 통한 선순환 기술향상 체계 부재

- 1) 현 R&D 체계로는 실서비스 환경 사용자 데이터 수집 불가능. 이에 따라 실서비스를 바탕으로 자동통번역의 성능을 개선할 수 있는 조직 체계로의 전환이 시급. 또한 사용자 데이터 DB 정제 및 플랫폼 서비스를 위한 상시 인력 필요
- 2) Google은 대용량 다국어 실사용자 로그 DB 수집 및 엔진반영의 선순환 구조를 기반으로 다국어 음성인식 및 자동통번역 기술의 강자로 부상하고 있음

라. 한국시장에 국한된 기술 개발 위주

- 1) 한국어 위주의 기술 개발로 음성언어 SW의 글로벌 산업경쟁력 확보 어려움. 아시아권 언어, 유럽어, 아랍어 등으로 확대하기 위해 장기간 다국어 지원하는 상시의 학제간 협력 및 인력 육성 필요
- 2) 국내 글로벌 업체(현대자동차, 삼성전자 등)에서는 다국어 음성인식 기술이 필요한 상황이지만, 다국어를 지원하는 뉘앙스 등 외산에 종속되어 있음

마. 국민 삶의 질 향상이란 국가적 미션 수행 필요

- 1) 국가 글로벌 SW 경쟁력 강화 및 '12년 여수 엑스포, '14 인천 아시안게임, '18 평창 동계올림픽에 자동통번역 대국민 서비스를 통한 국격 제고 필요
- 2) 급속한 개방과 세계화로 외국어 정보와 지식이 경쟁력이 되는 사회에서 외국어 소외 계층에게 정보 접근 통로를 제공하여 외국어 격차(Language divide) 해소

제 3 절 사업결과 기대효과

- 가. 평창 동계올림픽 등의 국제행사 대상 자동 통번역 서비스 지원을 통한 국격 제고

나. '12년 여수 엑스포, '14 인천 아시안게임, '18 평창 동계올림픽에 자동통번역 서비스 실시를 통해 언어장벽 없는 국가 실현 및 국격 제고



그림 1 자동 통번역 對국민 서비스 로드맵 (2012~2018)

다. 자동 통번역 산업생태계 조성 및 신규 서비스 창출

- 1) ETRI의 통번역 SW 플랫폼을 중심으로 응용 SW 개발자, 수요업체(단말기, 포털사, 주력산업 제조업체 등)간의 산업생태계 조성
- 2) 산/학/연 공동연구 활성화를 통해 ETRI 기술 인프라 공동 활용이 가능하므로 인프라가 부족한 기업의 자생력 강화 및 산업 활성화 가능
- 3) 자동통번역 산업뿐만 아니라, 휴대폰, 자동차 등 제조업 및 교육, 관광, 문화, 의료 등 서비스 분야와 융합한 新서비스 창출·발전

라. 국내 SW 경쟁력 제고

- 1) 공공재적 성격의 융합기술 경쟁력을 확보하고 있는 ETRI와의 기술협력 및 지원을 통해 국내 SW기업의 기술경쟁력 제고
- 2) 출연(연)이 상대적으로 취약한 기술사업화 부문에서 출연(연) R&D성과 확산→ 국내기업 R&D지원 강화→ 자동통번역 산업생태계 구축 등 상생협력 선순환 고리 정착

마. 사교육비 경감 및 창의인재 양성

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

제 4 절 국내·외 관련 기술의 현황

1. 국내 기술동향

- 가. 국내 핵심 기술 연구는 ETRI 등 국책연구기관을 통하여 이루어지고 있으며, 한국어에 대해서는 경쟁력을 유지하나 영어 등 다국어는 열세임
- 나. 현재 자동통번역 국내의 기술 수준

비교항목	국 내	국 외	사용자 요구수준
통역 성공률	70%~75%	75%~80%	90%
통역 가능한 언어 수	3개 언어 (한, 영, 일)	15개 언어 (한, 영, 일, 중, 스페인, 독일어, 프랑스 등)	20개 언어
통역 대상 어휘 수	10만 어휘	100만 어휘 이내	100만 어휘 이상
통역 대상 범위	관광 분야	무제한 분야	무제한 분야

표 1 자동통역 기술의 국내외 수준 비교

- 다. 1999년 한국전자통신연구원(ETRI)은 C-STAR (Consortium for Speech Translation Advanced Research)의 핵심회원 기관으로서 한/영/일/프랑스 4개국 간 실시간 음성언어번역 국제시연을 실시하였음. 고객이 여행사 직원과 여행계획을 상담하는 대화를 대상으로 자동통역 기술의 실현 가능성을 입증하였음
- 라. 2001년 삼성과 히다찌 연구소가 한/일 월드컵 국제행사를 통해 1000 여 개의 고정 문장에 대하여 휴대전화로 이용할 수 있는 한/일 자동통역 시범서비스 실시한 바 있음
- 마. 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

- 바. 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
- 사. CSLI는 구글 음성인식, 중국어 번역의 경우 일본의 자동번역 기술을 도입하여 태블릿PC용으로 한,중,영,일 4개국 통역기 출시. 국내시장 점유 및 해외시장 선점을 위해 기술적 독립성이 요구됨

2. 국외 기술동향

가. 미국

- 1) DARPA(Defense Advanced Research Projects Agency)는 IBM, SRI, 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
- 2) IBM은 5년내 상용화 가능 기술로서 자동통역을 최우선 순위로 선정하여 DARPA의 지원 하에 MASTOR (Multilingual Automatic Speech-to-Speech Translator)라는 자동통역기를 개발. 영어-중국어 양 방향 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 중동 연구소에서 개발한 TALES (Translingual Automatic Language Exploitation System)라는 번역 엔진을 사용
- 3) 카네기멜론대학교(Carnegie Mellon University)에서는 2003년에 음성 합성 전문 기업인 캡스트럴(Cepstral), 멀티모달 테크놀로지(Multimodal Technologies), 모빌 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 중간 언어(Interlingua)를 이용함. CMU대학에서 spin-off한 미국 Jibbigos는 스마트폰 탑재형 영어 기반 일어, 중국어, 스페인어, 아랍어, 양방향 통역기 개발. 한국어 통역은 현재 개발중이며 내년도 출시 예정

- 4) 미국 구글은 2012년 다국어 대상 서버기반 통역서비스 실시 계획. 현재 서버형 음성인식/번역 요소기술은 개발되었으나 통역서비스 수준까지는 이르지 못하고 있음
- 5) SRI (Stanford Research Institute)는 DARPA의 지원 하에 IraqComm이라는 자동통역기를 개발. 영어 및 아랍어에 대해 수만 단어를 처리하며, 자기 방어, 보안, 기초 의료 서비스 상황을 통역 대상으로 함. 2006년 미군에 의해 채택
- 6) VoxTec은 DARPA의 지원을 받아 Phraselator라는 PDA 크기의 자동통역기를 개발하였음. 인식률은 70%이상인 것으로 알려져 있음. 음성을 합성하지는 않으며 미리 녹음된 파일을 재생함. DynaSpeak를 인식 엔진으로 사용. 이라크와 아프가니스탄에 파견된 미군부대에서 총 2천대가량이 사용되었으며, 2005년에 미군이 동남아 해일 구조 활동을 할 때 활용되었음. 선박 간의 의사소통을 위해 해군이 성능을 평가하였으며 뉴욕 주의 경찰, 병원 응급실 등에서도 활용을 시도하고 있음
- 7) Fluential은 노트북에서 수행되는 S-MINDS라는 양방향 자동통역기를 개발하였음. 대상 영역을 세분화하고 영역을 선택한 후 사용하는 방법을 채택함. 노트북을 허리에 차고 스피커와 헤드셋을 착용하는 형태임. 표현이 고정되어 있는 군사 훈련시의 교본, 방사선 촬영시의 기본 문답 등을 대상으로 함. 미군에 의한 이라크군의 훈련, 미국내 병원 진료에 사용

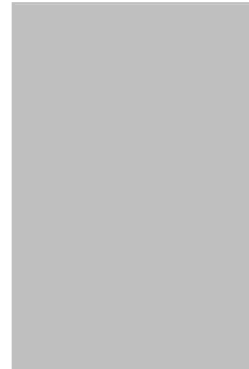


그림 2 통역시스템

나. EU

- 1) EU는 의회에서 행해지는 모든 연설을 포함하여, 11개 공식 언어로 문

서화하는 작업에 연간 5.5억 유로를 지출하고 있으며 회원국 간의 교류에 있어서 언어장벽이 가장 심각한 문제로 대두됨. 이에 따라, EU 제6차 연구지원사업 프레임웍에서는 의회 연설문의 자동전사 및 통역/번역 기술 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

- 2) 독일 국책연구소 DFKI는 1991년부터 대화체 자동통역 시스템의 개발을 위해 휴대용 독일/독일 자동통역 시스템 VerbMobil의 개발을 주도하여 비즈니스 상담을 보조하는 구문단위 통역기술을 개발

다. 일본

- 1) NEC는 2002년에 PDA에서 동작하는 일/영 여행자용 자동통역 시스템을 개발하였으며, 한국어를 포함하여 총 11개국어를 지원하는 다국어 자동번역 기술을 개발하여, Bestiland라는 번역 포털 서비스를 실시함. 또, 일/중 통역 시스템도 추가 개발하여 2005년부터 나리타 공항에서 시범 서비스를 실시, 이 중 일영 단방향 통역 시스템은 2007년 12월 휴대폰 단말기에서 실시간으로 동작함
- 2) ATR 연구소는 1986년부터 자동통역 기술의 실제 실용화를 목표로 연간 200억원의 연구비를 투입하여, 일어-영어, 일어-중국어간 양방향 대화체 자동통역 시스템을 개발하고 있음. 2007년도 말부터 휴대폰 단말기에 네트워크 기반으로 일/영 통역 서비스를 제공하고 있음
- 3) NTT, KDD, NHK, 히다치 등은 산학연간 기술 교류의 활성화를 통해 여행안내, 강의/연설, 비즈니스 회의 등 다양한 분야에서의 실용화를 시도하고 있음

제 5 절 핵심요소 및 접근방법

가. 對국민 자동 통번역서비스 제공

- 1) '12년 여수 엑스포 자동통역 특화를 통한 3개 국어(한/영, 한/일) 자동통역 대국민 서비스 실시
- 2) 2014년 인천 아시안게임용 자동통역 특화를 통한 4개 국어(한/영, 한/중, 한/일) 자동통역 대국민 서비스 실시
- 3) 2018년 평창 동계올림픽용 자동통역 특화를 통한 6개 국어(한/영, 한/중, 한/일, 한/스페인, 한/프랑스) 자동통역 대국민 서비스 실시
- 4) 일상영역 대국민 자동통번역 시범서비스 구축 및 운용

나. 자동 통번역 산업생태계 구축

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



그림 3 자동 통번역 산업생태계 구축방안

다. 국내외 개방형 연구 및 인력양성 추진

- 1) 인문, 공학 등 학제간 융합연구를 통한 창의 연구결과 도출
- 2) 다국어기술 개발을 위한 국제 공동연구 지원
- 3) 기업체, 대학 등 센터로 인력 파견을 통해 공동연구 추진 및 창의인재 육성

라. 혁신적 연구추진 체계 구축

- 1) 기존 자동통번역 개발방법론으로는 한계가 있음. 따라서 국내외 업체에 자동통번역 공통플랫폼 및 Open API를 지원하여 실서비스 환경 대용량 로그데이터 축적을 가능하게 해줌으로서 지속적 성능 개선이 가능한 혁신적인 방법을 추진하고자 함
- 2) 중소기업과 공동으로 자동통번역 BM 개발 및 실용화를 추진함으로써 초기 자동통번역 기술 활용에 따른 위험부담을 줄이며 이에 따라 국내에 자동통번역 산업생태계가 조성되는데 자동통번역센터가 큰 역할

을 할 것으로 보임

- 3) 국내에서는 한국어, 영어, 중국어, 일본어 등 아시아권 주요언어에 치우쳐져 기술개발이 진행되는 한계가 있음. 이에 글로벌 시장에 진출하기 위한 유럽어, 동남아, 중동 등 다국어 음성언어 기반 기술을 개발함으로써 구글, 애플(뉘앙스) 등과 같이 국제적 수준의 기술확보가 가능함

제 2 장 연구개발 목표 및 내용

제 1 절 최종목표

구 분	내 용
최종목표	<ul style="list-style-type: none"> ○ 언어장벽 해소를 위한 대국민 자동통번역 서비스 실시 ○ 자동통번역산업 생태계 구축 ○ '12년 여수엑스포, '14 인천 아시안게임, '18 평창 올림픽 지원을 통한 국격 제고
세부목표	<ul style="list-style-type: none"> ○ '12년 여수엑스포, '14년 아시안게임, '18년 평창올림픽 대국민 자동통번역 서비스 ○ 자동통번역 플랫폼 및 Open API 지원 ○ 실서비스 환경 로그데이터 기반 음성언어자원 구축 및 배포 ○ 다국어 음성언어 연구기반 조성

표 2 사업목표



그림 4 자동통번역 산업경쟁력 강화 사업의 로드맵 (2012~2018)

제 2 절 연차별 연구개발 목표 및 내용

구 분	목 표	내 용
1차년도 (2012)	<ul style="list-style-type: none"> ○ 자동통번역 산업기반 조성(1차) ○ ‘12년 여수엑스포 대국민 자동통번역서비스 실시 	<ul style="list-style-type: none"> ○ 자동통번역 센터 설립 ○ 자동통번역 시범서비스 인프라 구축 ○ 자동통번역 플랫폼 및 Open API 설계 및 조기 구축 ○ 사용자 로그데이터 구축 환경 조성 및 도구설계 ○ 중소기업 BM 발굴 및 응용개발 지원 ○ 사용자 로그데이터 기반 한/영 자동통번역 성능 개선 (통역률 ERR(Error Reduction Rate) 20% 개선) ○ 여수엑스포 한/영 자동통번역 대국민 서비스 실시
2차년도 (2013)	<ul style="list-style-type: none"> ○ 자동통번역 산업기반 조성(2차) 	<ul style="list-style-type: none"> ○ 자동통번역 플랫폼 및 Open API 구현 보완 ○ 사용자 로그데이터 구축 및 반자동 정제 도구 개발 ○ 중소기업 BM 발굴 및 시범서비스 지원 ○ 표준형 음성언어 평가데이터 배포 ○ 자동통번역 플랫폼 기반 app개발 프로모션(1차) ○ 다국어 음성언어 연구기반 조성(1차) ○ 사용자 로그데이터 기반 한/영/일 자동통번역 성능 개선(통역률 ERR 20% 개선) ○ 아시안게임용 한, 중, 영, 일 4개국 자동통번역 서비스 운영
3차년도 (2014)	<ul style="list-style-type: none"> ○ 자동통번역 산업생태계 구축(1차) ○ ‘14년 인천 아시안게임 한, 중, 영, 일 	<ul style="list-style-type: none"> ○ 자동통번역 플랫폼 및 Open API 홍보 및 배포 ○ 사용자 로그데이터 기반 한/중 자동통번역 성능 개선 (통역률 ERR 20% 개선) ○ 중소기업 BM 발굴 및 상용화 지원(1차) ○ 자동통번역 플랫폼 기반 app개발 프로모션(2차) ○ 다학제간 산/학/연 협력연구 지원(1차)

	4개국 자동통번역 대국민 서비스 실시	<ul style="list-style-type: none"> ○ 다국어 음성언어 연구기반 조성(2차) ○ 아시안게임용 한/중 자동통번역 성능 개선 및 한, 중, 영, 일 4개국 대국민 서비스 실시
4차년도 (2015)	○ 자동통번역 산업생태계 구축(2차)	<ul style="list-style-type: none"> ○ 자동통번역 플랫폼 및 Open API 확산 ○ 사용자 로그데이터 기반 한/스페인어 자동통번역 성능 개선 (통역률 ERR 20% 개선) ○ 중소기업 BM 발굴 및 상용화 지원(2차) ○ 다국어 음성언어 데이터 구축(유럽어) ○ 다학제간 산/학/연 협력연구 지원(2차) ○ 다국어 음성언어 연구기반 조성(3차)
5차년도 (2016)	○ 자동통번역 글로벌 경쟁력 확보(1차)	<ul style="list-style-type: none"> ○ 다국어 지원용 자동통번역 플랫폼 및 Open API 설계 ○ 사용자 로그데이터 기반 한/불어 자동통번역 성능 개선 (통역률 ERR 20% 개선) ○ 글로벌시장 대상 BM 발굴 및 상용화 지원(1차) ○ 다국어 음성언어 데이터 구축(유럽어) ○ 다국어 음성언어 연구를 위한 국제협력 지원(1차)
6차년도 (2017)	○ 자동통번역 글로벌 경쟁력 확보(2차)	<ul style="list-style-type: none"> ○ 다국어 지원용 자동통번역 플랫폼 및 Open API 구현 ○ 사용자 로그데이터 기반 한/독어 자동통번역 성능 개선 (통역률 ERR 20% 개선) ○ 글로벌시장 대상 BM 발굴 및 상용화 지원(2차) ○ 다국어 음성언어 데이터 구축(아시아 주요언어) ○ 다국어 음성언어 연구를 위한 국제협력 지원(2차) ○ 다국어 통번역 플랫폼 기반 app개발 프로모션(1차) ○ 평창올림픽용 한, 중, 영, 일, 스, 불, 독 7개국 자동통번역 특화 시범서비스 운용

7차년도 (2018)	<ul style="list-style-type: none"> ○ 자동통번역 글로벌 시장 진출 ○ '18년 평창 동계올림픽 한, 중, 영, 일, 불, 스, 독 7개국 자동통번역 대국민 서비스 실시 	<ul style="list-style-type: none"> ○ 다국어 자동통번역 플랫폼 및 Open API 홍보 및 배포 ○ 글로벌시장 대상 BM 발굴 및 상용화 지원(3차) ○ 다국어 음성언어 데이터 구축(기타 주요언어) ○ 다국어 플랫폼 기반 app개발 프로모션(2차) ○ 평창올림픽 한, 중, 영, 일, 스, 불, 독 7개국 자동통번역 대국민 서비스 실시
----------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

표 3 연차별 연구목표

제 3 절 성과목표

1. 성과지표 선정 및 목표설정

성과목표		목표도출 근거	성과지표	당해 연도 목표 ('12년도)	평가(검증)방법	배점
output	언어장벽 없는 국가구현을 위한 자동통번역 산업경쟁력 강화 사업	자동통번역 분야 기술경쟁력 제고 및 국내 연구중심/허브 역할 제고의 성과로 판단	사용자 로그데이터 기반 한/일 통역성능 개선 (ERR)	20%	3인 이상의 전문 번역가 평가 실시	40
			특허출원 건수 (국내/국제)	5(3/2)	통번역 플랫폼 Open API 활용 기관 수 및 DB배포 건수	10
			중소기업 BM발굴 및 지원 (건)	3	중소기업 BM관련 응용 prototype 구현 건수	20
			소계			70점
outcome/impact		자동통번역 산업 경쟁력 강화를 위한 지표로 판단	공공적/공익적 연구성과 활용 실적	1	국가차원 공공행사에 대 국민 자동통역서비스 지원 여부	30
합계						100점

표 4 성과지표

2. 성과지표별 목표 추세

세부성과목표		성과지표	실적*				3년간 목표				장기목표**
			계	10년	11년	12년	계	13년	14년	15년	
output	언어장벽 없는 국가구현을 위한 자동통번역 산업경쟁력 강화 사업	사용자 로그데이터 기반 한/일 통역성능 개선 (ERR: Error Reduction Rate)	한/영 통역률 ERR 20% 개선	-	-	한/영 통역률 ERR 20% 개선	한, 영, 일, 중 통역률 ERR 20% 개선	한/일 통역률 ERR 20% 개선	한/중 통역률 ERR 20% 개선	한/스 한/불 통역률 ERR 20% 개선	로그데이터 기반 상용수준의 8개국 자동통역엔진 개선
		특허출원 건수 (국내/국제)	1 (1/0)			1 (1/0)	15 (9/6)	5 (3/2)	5 (3/2)	5 (3/2)	국내외 특허출원을 통한 지적재산권 30건 이상 선점
		중소기업 시제품제작 지원 (건)	4			4	9	3	3	3	중소업체 대상 시제품제작 지원사업 10건 추진
outcome/impact		공공적/공익적 연구성과 활용 실적(건)	1	-	-	1	3	1	1	1	국가차원의 공공적 행사에 매년 1건 활용

3. 목표달성을 위한 추진전략 및 실천계획

성과목표	추진전략	로드맵						재원 투입 계획 (2013)	추진일정 (2013)	
		'12	'13	'14	'15	'16	'17			'18
사용자 로그데이터 기반 한/일 통역성능 개선 (%)	<ul style="list-style-type: none"> 충남대, 외대 등 대학내 원어민과 시스템 평가 등 협력 업체보유 자원 협력 사용자 로그데이터 구축을 통한 특화영역 성능 개선 국내 협력연구 추진 	한/일 영 ER R 20 % 개 선	한/일 ER R 20 % 개 선	한/중 ER R 20 % 개 선	한/스 ER R 20 % 개 선	한/독 ER R 20 % 개 선	8개 국 ER R 10 % 개 선	8개 국 ER R 10 % 개 선	3MY / 6억	<ul style="list-style-type: none"> 2월: 로그 데이터 정제 용역 추진 5월: 자동통역 1차 평가 10월: 자동통역 2차 평가
특허출원 건수 (국내/국제)	<ul style="list-style-type: none"> 기존R&D결과를 기반으로 통번역 플랫폼화 추진 Open API 지원 	1 (1/)	5 (3/2)	5 (3/2)	5 (3/2)	5 (3/2)	5 (3/2)	5 (3/2)	1MY / 3억	<ul style="list-style-type: none"> 상반기: 국내외 특허 3건 출원 하반기: 국내외 특허 2건 출원
중소기업 시제품제작 지원 (건)	<ul style="list-style-type: none"> 중소업체 2~3곳 선정하여 BM개발 추진 금년에는 최소 1곳 이상 모범사례를 만드는 데에 주력 	4	3	3	3	4	4	5	3MY / 6억	<ul style="list-style-type: none"> 3월: 중소기업 선정 6월: 1차 시제품 구현 9월: 시제품 구현 완료 10월: 시범서비스 운영
공공적/공익적 연구성과 활용 (건)	<ul style="list-style-type: none"> 업체와 공동으로 인천 아시안 게임에 통역서비스 지원 대국민 시범서비스 운영 	1	1	1	1	1	1	1	1MY / 2억	<ul style="list-style-type: none"> 2월: 한/일 통역서비스 통합 4월: 대국민 한/일 통역시범서비스 제공

제 4 절 추진체계 및 일정

1. 당해연도 사업 추진체계

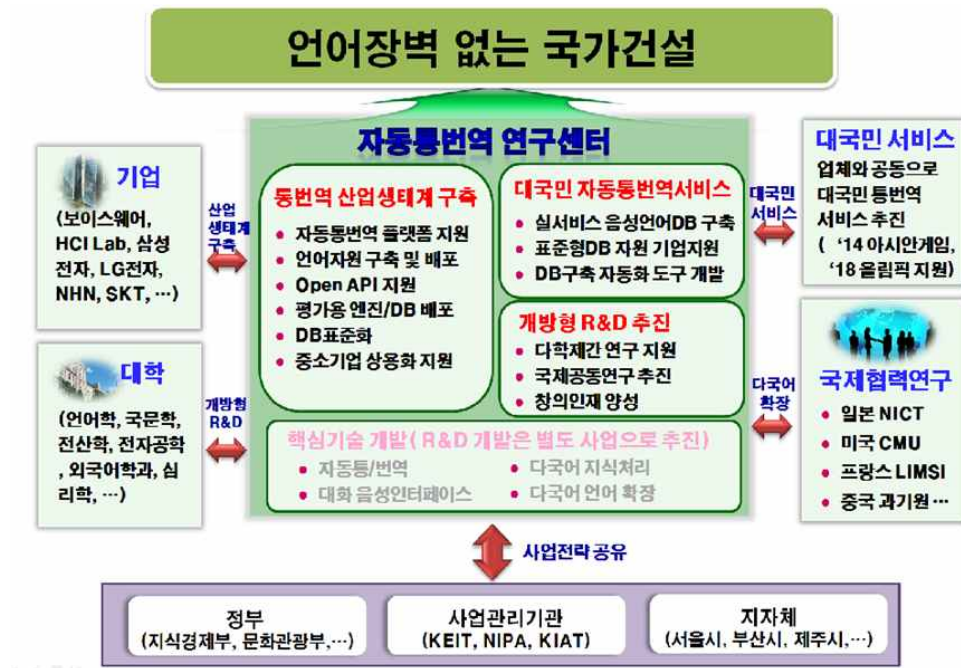


그림 5 당해연도 사업 추진체계

2. 추진전략

- 한/일 자동통역 시범서비스의 경우, 당초 사용자 로그데이터 축적을 위한 목적이 주요하므로 일본어 원어인 데이터를 확보하기 위해 일본에서도 시범서비스를 동시 실시하였으며, 9/12일부터 4일간 일본 동경에서 열리는 세계여행박람회(세계 160여개국, 13만 여명이 참석)에 참석하여 일본인들에게 '지니톡'을 홍보, 더 많은 사용자를 확보하고자 하였음
- 일본 동시 시범서비스 실시가 데이터 축적에 기여가 되는 것으로 판단되면, 향후 중국어, 영어 등도 외국에서 직접 지니톡을 서비스하는 방향으로 추진할 예정임. 특히 중국 현지 IDC센터를 통해 직접 서비스 하는 방안도 고려중임
- 중소기업 시제품제작 지원 사업의 경우, BM 도출을 위해 ETRI-업체와 밀접합

하여 기획안을 도출하고, 지속적인 feedback을 통해 제품의 완성도를 높이고자 함

- 다학제간 협업을 위해 충남대, KAIST 등 주변 대학 및 부산대, 이대 통번역대학원 등과 연계하여 기술 개발 추진할 계획임. 현재 부산대 불문학과와 한/불 대역사전 구축 업무를 진행하고 있고, 이대 통번역대학원에서 자동통역 기술 소개(11/25)를 통한 협력 분야를 파악하였음
- 창의인재 육성을 위해 UST 석박사 과정 학생을 적극 선발하고, 해양대 등 대학과 연구연수생 교류 추진 예정. 현재 UST 석사과정 2명, 박사과정 1명과 함께 연구를 진행중이며, 석사과정 1명이 내년 2월부터 신규로 합류할 예정임
- 자동통역 기술 기반 시장이 형성되도록 Open API지원을 확대할 예정임. 내년 상반기에 본격적인 지원체계를 구축하고 홍보를 통해 자동통역 응용 제품 확산 및 시장 창출에 힘쓸 예정임

3. 추진일정 및 예상결과물

과제내용	추진 일정												활동 책임자	연구개발비 (천원)	참여 인력 (M/Y)
	1	2	3	4	5	6	7	8	9	10	11	12			
<ul style="list-style-type: none"> ○ 대국민 한/일 자동통역 시범서비스 실시 • 한/일 대국민 자동통역 시범서비스 구현 • 일본형 스마트폰 지원 지니톡 앱 구현 • 한/일 자동통역 시스템 성능 및 안정성 평가 													김상훈	200,000	3.5
<ul style="list-style-type: none"> ○ 자동통역 플랫폼 기반 Open API 구현 (1차) • 한/일 자동통역 공통플랫폼 설계 보완 • 업체 요구사항에 따른 자동통역 Open API 구현 보완 													김상훈	800,000	1.5
<ul style="list-style-type: none"> ○ 사용자 로그데이터 기반 한/일 자동통역 성능 개선 (통역률 ERR 20% 개선) • 사용자 로그데이터 반자동 정제 도구 구현 • 사용자 로그데이터 음성언어 데이터 자동정제 기술 개발 • 대용량 사용자 로그데이터 구축 및 DB화 • 사용자 로그데이터 기반 다중 발음모델링 기술 개발 • 사용자 로그데이터 기반 언어모델링 기술 개발 • 실환경 최적 음향모델링 기술 개발 • 특화영역별 자동통역 튜닝 도구 개발 													박준	400,000	5.0
<ul style="list-style-type: none"> ○ 중소기업 시제품 제작 지원 (2차) • 한/영/일 자동통역 기반 중소기업 시제품 제작 지원 추진 • 한/영/일 자동통역 BM 상용화 기술 지원 													김상훈	200,000	1.16
<ul style="list-style-type: none"> ○ 다국어 음성언어 연구기반 조성 (2차) • 다국어 음성언어 처리 요구사항 도출 • 통계 기반 한/일 구어체 번역 성능 개선 기술 개발 • 한/불어 대역사전 구축 및 불어 발음변환 기술 개발 • 한국어, 영어 표준형 음성언어 평가DB 구축 													김영길	200,000	1
<ul style="list-style-type: none"> ○ 2014년 인천 아시안게임 대비 한,중,영,일 4개국 자동통역 시범서비스 구현 • 아시안게임을 한,영,일 자동통역 영역 특화 • 클라우드 기반 자동통역 서비스 시스템 구현 및 운용 • 아시안게임을 한/영/일 자동통역 시범서비스 시스템 운영 													김상훈	200,000	2
<p style="text-align: center;">주요 Milestone <u>완성점에서의</u> 수행결과</p>	<ul style="list-style-type: none"> ○ 로그데이터 1차 정제 ○ 한/일 자동통역 대국민 시범서비스 실시 						<ul style="list-style-type: none"> ○ 로그데이터 2차 정제 ○ 중소기업 시제품 모범 사례 1건 창출 							2,000,000	14.16

제 3 장 당해연도 목표 및 수행결과

제 1 절 당해연도 목표

가. 개발목표: 자동통번역 산업기반 조성 (2차)

1. 대국민 한/일 자동통역 시범서비스 실시
 - 가) 한/일 대국민 자동통역 시범서비스 구현
 - 나) 일본향 스마트폰 지원 지니톡 앱 구현
 - 다) 한/일 자동통역 시스템 성능 및 안정성 평가

2. 자동통번역 플랫폼 기반 Open API 구현 (1차)
 - 가) 한/일 자동통번역 공통플랫폼 설계 보완
 - 나) 업체 요구사항에 따른 자동통번역 Open API 구현 보완

3. 사용자 로그데이터 기반 한/일 자동통역 성능 개선 (통역률 ERR 20% 개선)
 - 가) 사용자 로그데이터 반자동 정제 도구 구현
 - 나) 사용자 로그데이터 음성언어 데이터 자동정제 기술 개발
 - 다) 대용량 사용자 로그데이터 구축 및 DB화
 - 라) 사용자 로그데이터 기반 다중 발음모델링 기술 개발
 - 마) 사용자 로그데이터 기반 언어모델링 기술 개발
 - 바) 실환경 최적 음향모델링 기술 개발
 - 사) 특화영역별 자동통역 튜닝 도구 개발




4. 중소기업 시제품 제작 지원 (2차)
 - 가) 한/영/일 자동통번역 기반 중소기업 시제품제작 지원 추진
 - 나) 한/영/일 자동통번역 BM 상용화 기술 지원

5. 다국어 음성언어 연구기반 조성 (1차)
 - 가) 다국어 음성언어 처리 요구사항 도출
 - 나) 통계 기반 한/일 구어체 번역 성능 개선 기술 개발
 - 다) 한/불어 대역사전 구축 및 불어 발음변환 기술 개발
 - 라) 한국어, 영어 표준형 음성언어 평가DB 구축

6. 2014년 인천 아시안게임 대비 한,중,영,일 4개국 자동통번역 시범서비스 구현
 - 가) 아시안게임용 한,영,일 자동통번역 영역 특화
 - 나) 클라우드 기반 자동통번역 서비스 시스템 구현 및 운용
 - 다) 아시안게임용 한/영/일 자동통번역 시범서비스 운영

제 2 절 사업수행 결과

가. 주요 추진실적 (요약)

	성과목표	추진실적	달성도
대표성과	한/일 자동통역 시범서비스 실시 	<ul style="list-style-type: none"> • 5/8일 한국, 일본 동시 실시 완료 • 한/일 통역률 86.3% 달성, 일/한 통역률 88.8% 달성 (Google 대비 5~10% 우위) • 2013년 11월말 현재 158만 다운로드 기록 • 일본 세계여행박람회에 한/일 자동통역 ‘지니톡’ 전시 (동경, 9/12~9/15) • 창조경제 국무총리상 수상 (12/12) • 안전행정부 공공앱 최우수상 수상(12/10) 	100%
	한/중 자동통역 시범서비스 실시 	<ul style="list-style-type: none"> • 한/중 대국민 시범서비스 실시 (12/12) • 한/중 통역률 77% 달성, 중/한 통역률 80% 달성 (Google 대비 15~20% 우위) • 중국어 인식엔진 확보 (여행분야에서 Google 대비 동등 수준) • N-best 제시를 통한 앱 사용성 개선 완료 	100%
	순천정원박람회 자동통역서비스 지원 	<ul style="list-style-type: none"> • 기술이전 업체(CSLI)에서 순천정원박람회, 충주국제조정대회 자동통역서비스 제공 • ETRI는 업체에 순천정원박람회 자동통역(한국어, 영어) 특화 지원 완료 	100%

성과 지표	<p>사용자 로그데이터 기반 한/일 통역성능 개선 (목표 ERR 20% 개선)</p>	<ul style="list-style-type: none"> 5월 시범서비스 실시 때 한/일 번역률이 83.0%에서 11월 버전은 86.3%로 3.3% 개선 (ERR 19.4% 개선) 한국어 언어모델용 로그데이터 1,000만 문장 반영 완료 (ERR 12% 개선) 잡음에 강인한 특징추출로 한국어, 영어, 일본어 음향모델링 개선 (ERR 10% 개선) 한국어 로그데이터 40만 문장 정제 완료 일본어 로그데이터 5만 문장 정제 완료 	100%
	<p>특허출원 건수 (목표 5건)</p>	<ul style="list-style-type: none"> 10건 (국제 5건, 국내 5건) 출원 완료 국제특허 1건 (특허명: 핸드프리 자동통역)은 ITU-T 국제표준으로 추진 중 	100%
	<p>중소기업 BM 발굴 지원 (목표 3건)</p>	<ul style="list-style-type: none"> KTX 통역용(지원업체: 디지쿼터스), 영어 교육용(지원업체: 컨피테크), 해외도시별 특화용(지원업체: 스크미디어) 3건 추진 완료 시제품 개발 기획(안) 작성 및 설계 완료 시제품 구현 완료. 사용성평가 추진 	100%
	<p>공익적/공공적 연구성과 활용 실적 (목표 1건)</p>	<ul style="list-style-type: none"> 순천정원박람회, 충주국제조정경기대회 등 국제행사용 자동통역 특화 지원 2건 완료 부산시티투어, 인천공항공사 등 자동통역 사업화를 위한 업체 상용화 현장지원 	100%

표 5 주요 추진 실적 (요약)

나. 세부 추진실적 (정성적)

추진계획	추진실적
<p>대국민 한/일 자동통역 시범서비스 실시</p>	<ul style="list-style-type: none"> • 5/8일 한/일 자동통역 시범서비스 한국, 일본 동시 실시 완료 (그림 6~9: 언론보도자료 참조) • 2013년 11월말 현재 158만 다운로드 기록 <ul style="list-style-type: none"> ※ 미국, 캐나다, 호주, 홍콩, 일본 등 세계 10여개국에서 사용 • 일본향 지니톡 앱 구현 완료 (그림 10) <ul style="list-style-type: none"> ※ 한국관광공사 일본어 사이트에서 일본향 지니톡 홍보 ※ 일본향 앱은 일본에서만 다운로드 받을 수 있음 • 국내향 지니톡 앱 구현 완료 (그림 11) • 일본어 인식엔진 구현 완료 (14만 단어급 90% 인식률) • 한/일 SMT 자동번역 기술 구현 완료 (11월 평가지, 한/일 번역률 86.3%, 일/한 번역률 88.8%) <ul style="list-style-type: none"> ☞ 5월 평가 대비 한/일 번역률 3.3% 개선 (ERR 19.4% 개선) ☞ 개선요인: 부정문/의문문 후처리 반영, 한/일 대역 코퍼스 보강, 디코더 성능 개선 ☞ 현재 구축진행 중인 지니톡 로그데이터 기반 한/일 대역코퍼스 20만 문장을 반영하면 성능개선 효과가 기대됨 • 한/일 유사문장 제시 기능 구현 완료 (hit율: 18.4%) <ul style="list-style-type: none"> ☞ 한/일 번역률 약 1~2% 개선 효과가 있음 (ERR 5.8% 개선 효과) • 일본어 발음에 대한 로마나이즈 기능 구현 완료 (변환성 공률: 97.4%)

	<ul style="list-style-type: none"> • 지니톡 사용자 요청에 따라 자동통역 결과를 타 응용에 사용할 수 있도록 copy&paste 기능 추가 • 동시접속자 확대. 서버당 기존 20명에서 40명 동시접속자 수용 구조로 개선 완료 • 한국관광공사 일본어 홈페이지에 지니톡 기술 소개 • 9/12~9/15일 일본에서 열리는 세계여행박람회 (JATA)에 한/일 지니톡 자동통역 기술 전시 (그림 12) • JATA 전시회 참여성과 <ul style="list-style-type: none"> ☞ 여행전문 언론매체 등 15 개 대상 홍보 ☞ 일본내 온라인 여행업체 2 개 대상 홍보. 업체 일본어 게시판에 홍보하기로 협의 함 ☞ 일본인 대상 700~800 명 정도 개별 홍보: 홍보리플릿 전달 및 시연 ☞ 서울관광공사 일본어 게시판에 지니톡 홍보하기로 협의함 ☞ 코레일 공항철도공사와 KTX 영역 자동통역 서비스 시범적용을 검토하기로 함 ☞ 일본인 대상 지니톡 기반 교육용 앱 개발 필요성 충분 ☞ 행사에 출전한 각 지자체/관광공사에 ‘지니톡’ 홍보
<p>자동통번역 플랫폼 기반 Open API 구현 (1차)</p>	<ul style="list-style-type: none"> • 국내업체 대상 자동통역 Open API 구현 완료 (그림 13) • 업체 요청시 승인과정을 거쳐 3개월 사용이 가능한 Open API 지원 • 업체는 자동통역 Open API를 이용하여 자동통역 사업화 가능성을 리스크 없이 파악할 수 있고, 이후에 기술이전 단계로 발전하는 절차 수립 완료 • 현재 SKC&C, 디오텍 등 9개 업체 대상 지원 완료
<p>사용자 로그데이터 기반 한/일</p>	<ul style="list-style-type: none"> • 한/영 자동통역 기술 성능 개선 완료 (그림 14) • 한국어 텍스트 로그데이터 1,000만건에 대한 자동정제 실

자동통역 성능 개선
(통역률 ERR 20% 개선)

시 후 언어모델(LM: Language Model)에 반영 완료



※ 본 언어모델은 기 한/영 자동통역 기술을 이전 받은 업체에 업데이트 버전으로 기술이전 완료함

- 텍스트 로그데이터 자동정제를 위한 띄어쓰기, 철자오류 수정 기술 개발

기술명	성능	평가방법
텍스트 자동정제	98.6%	지니톡 로그 1,000문장 대상 평가

- 음성 로그 자동정제를 위한 신뢰도 측정 기술 개발
- 구글 n-gram 기반 언어모델 interpolation 기술 개발
- 배경잡음에 강인한 음향모델링을 위한 잡음과 음성 믹싱 기술 개발
- 기존 39차 특징벡터 대신 배경잡음에 강인한 53차 음성 특징추출 기술 적용 (ERR 10% 개선)

※ ERR(Error Reduction Rates): 오류 감소율

- 문장부호 복원기술 개발 (복원률: 91.9%)

	<ul style="list-style-type: none"> • 한국어 클래스 태거 기술 개발 (태깅률: 52.8%) • 한/영 유사문장 검색 기술 개선 (hit율: 20.8%) • 한국어 음성 로그데이터에 대해 약 40만 건 수동 정제 진행 완료 • 한/일 자동통역 성능 개선을 위한 사용자 로그데이터 정제 5만건 완료 • 지니톡 로그데이터로부터 한/일 대역문장 20만건 구축 진행 완료
<p>중소기업 시제품 제작 지원 (2차)</p>	<ul style="list-style-type: none"> • 3개 업체를 대상으로 시제품제작 지원 추진 (지원대상업체: 컨피테크, 스크미디어, 디지쿼터스) • 각 업체별로 자동통역 기반 신규 BM 발굴 완료 • 컨피테크는 영어 외국어 교육용 앱 개발 진행. 3차 기획작업 완료 (그림 15) • 스크미디어는 해외 여행객을 대상으로 방문도시에 특화가능한 자동통역 기술 적용. 2차 기획작업 완료 (그림16) • 디지쿼터스는 향후 국내 외국인 방문이 증가함에 따라 KTX 역내 매표창구에서 언어소통이 가능한 통역 서비스 개발 진행. 1차 기획 작업 완료 (그림 17) • 3개 시제품 구현 완료. 사용성 평가 후 성능 개선 추진
<p>다국어 음성언어 연구기반 조성 (1차)</p>	<ul style="list-style-type: none"> • 다국어 언어처리 관련 위탁과제 2건 추진 완료 • 부산대(윤애선교수)는 프랑스어 발음변환용 사전 구축 및 대역 사전 구축 완료 (약 7만 엔트리 구축) • 경인여대(여상화교수)는 영/프랑스, 영/스페인 SMT 번역 기술 프로토타입 구축 완료

	<p>※ SMT (Statistical Machine Translation): 통계기반 번역</p>
<p>2014년 인천 아시안게임 대비 한,영,일,중 4개국 자동통역 시범서비스 구현</p>	<ul style="list-style-type: none"> • 분산형 언어모델 기술을 사용한 N-best 제시 기능을 중국어에 최초 적용 • 중소기업 지원사업으로 진행중인 스콧미디어를 통해, 인천아시안게임용 자동통역 앱 프로토타입을 구현 진행 • 인천 아시안게임 대비 중국어 언어모델용 코퍼스 구축 용역 추진 완료 (총 200만 문장 수집 완료) • 미국 DB배포기관인 LDC 회원으로 가입하여 중국어 음성DB 120시간분 확보 <p>※ LDC(Linguistic Data Consortium): 미국 음성언어DB 배포기관</p> <ul style="list-style-type: none"> • 인천 아시안게임 준비를 위한 업체와 역할 분담 및 협력 방안 수립 • 광주 유니버시아드 조직위 및 정보시스템 구축업체 (SKC&C)와 자동통역 서비스 지원 방안 협의

ETRI, 지니톡 한일 통역 버전 출시

(대전=연합뉴스) 박주영 기자 = ETRI(한국전자통신연구원)는 한·영 자동 통역이 가능한 모바일 단말용 애플리케이션 '지니톡(GenieTalk)'의 일본어 통역서비스를 실시한다고 7일 밝혔다.

서비스는 한국과 일본에서 동시에 시행되며, 여행·관광·어학 등 분야의 일본어 대화체 13만 단어를 음성 인식할 수 있다.

ETRI는 내년 인천 아시안게임은 물론 2018년에 열릴 평창 동계올림픽까지 한국어, 영어, 중국어, 일어, 스페인어, 불어, 독일어, 러시아어 등 총 8개 국어의 자동통역 서비스를 완료한다는 계획이다.

지난해 10월 출시한 지니톡의 다운로드 수는 7개월 만에 130만건을 넘어섰으며, 연구 성과의 우수성을 인정받아 문화체육관광부 시행 '2012 스마트콘텐츠 어워드'와 지식경제부 시행 '오픈기술상'을 수상했다.

jyoung@yna.co.kr

(끝)

<저 작 권 자(c)연 합 뉴 스. 무 단 전 재-재 배 포 금 지.>

[2013-05-07 09:53 송고]

그림 6 연합뉴스 지니톡 기사

기사 주소: http://www.dt.co.kr/contents.html?article_no=2013050802011376731002

자동통역 '지니톡' 일본어도 서비스

이준기 기자 bongchu@dt.co.kr | 입력: 2013-05-07 20:03

한국전자통신연구원(ETRI)은 자동통역 앱 '지니톡'의 일본어 통역 시범 서비스를 한국과 일본에서 동시에 실시한다고 7일 밝혔다.

이 서비스는 한·영 자동통역 앱인 지니톡(화면)에 일본어를 추가로 제공함에 따라 이뤄지는 것으로, 초기 화면의 메뉴 버튼을 클릭한 후 환경설정에서 통역 언어를 일본어로 변경하면 사용할 수 있다.

한·일 통역 서비스는 일본어 대화체 13만 단어의 음성 인식을 제공해 여행, 관광, 어학 등의 분야에서 활용할 수 있다. ETRI는 오는 2014년 아시안게임은 물론 2018년 평창 동계올림픽까지 한국어, 영어, 중국어, 일본어, 스페인어, 프랑스어, 독일어, 러시아어 등 8개 국어의 자동통역 서비스를 완료한다는 계획이다.

박상규 자동통역인공지능연구센터장은 "내년 인천 아시안게임에는 관련 기술을 이전받아 사업화를 진행하고 있는 국내 자동통역 서비스 업체와 함께 자동통역 서비스를 제공해 언어장벽 없는 세상을 실현하는데 더욱 노력하겠다"고 말했다.

지난해 10월 출시한 지니톡 다운로드 건수는 130만건을 넘어섰으며, 구글 플레이스토어 또는 애플의 앱스토어에서 '지니톡'이나 'GenieTalk'로 검색한 후 내려받아 사용할 수 있다.

대전=이준기기자 bongchu@



그림 7 디지털타임즈 지니톡 기사

[대전·대덕] ETRI, 한-일 자동통역 '지니톡' 시범서비스

2013-05-08 00:00



한국전자통신연구원 ETRI(에트리)는 세계 최고 수준의 순수 국산 자동통역 애플리케이션 '지니톡'이 한-영 자동통역에 이어 한-일 자동통역 시범서비스에 들어갔습니다.

한국과 일본에서 동시에 실시되는 한-일 '지니톡' 자동통역 시범서비스는 많은 사용자가 한꺼번에 서비스를 이용할 경우 일부 제약을 받을 수 있지만 기술이전업체에서 제공하는 상용서비스는 모든 기능을 원활하게 사용할 수 있습니다.

ETRI는 오는 2014년까지 아시안게임은 물론 2018년 평창 동계올림픽까지 한국어와 영어, 중국어, 일어, 스페인어, 불어, 독어, 러시아어 등 모두 8개 국어의 자동통역 서비스를 실시한다는 계획입니다.

이정우 [ljwwow@ytn.co.kr]

그림 8 YTN 지니톡 기사



写真左はAndroid（日⇄韓・韓⇄英）、右はiPhone（日⇄韓）

スマートフォンやタブレット端末に向かって日本語で話すと同語に訳して読み上げてくれる、無料の音声通訳アプリが登場しました。韓国電子通信研究院（ETRI）が開発したもので、旅行、観光、語学分野の通訳ができる、旅行者のためのアプリです。韓国語の単語は27万語、日本語は13万語が認識できます。

日本語⇄韓国語の音声通訳の他、テキスト入力での翻訳もでき、音声は男性、女性の声を選べます。また、日常会話検索機能もあり、例えば「ホテル」と音声入力すると、ホテルに関する会話例が表示され、韓国語で読み上げてくれます。

通訳処理はサーバーで行われるため、インターネットに接続する必要があります（データ利用料金が発生することがあります）、韓国内ではあちこちにあるWi-Fi（ワイファイ）ゾーンで利用されると便利です。

☞ 韓国で使えるWi-Fi（ワイファイ）ゾーン総まとめ

旅行のお供に便利な日韓音声通訳アプリ「Genietalk（ジーニートーク）」は、Apple App StoreもしくはGoogle Playでダウンロードできます。

more info

日韓音声通訳アプリ「Genietalk（ジーニートーク）」

☞ ダウンロード：iPhone ⇒ Apple App Store、Android ⇒ Google Play

☞ 言語：日本語⇄韓国語と韓国語⇄英語のアプリがあります。

日本語⇄韓国語は「ジーニートーク」とカタカナで検索、韓国語⇄英語は「Genietalk」と英語で検索。

※日本以外で契約しているアンドロイド利用者の場合は、「Genietalk」で検索して、設定部分で日本語⇄韓国語に変更してください。

☞ 問い合わせ：韓国電子通信研究院（ETRI）

ホームページ <http://genietalk.etri.re.kr>（韓国語） / Eメール p-genie@etri.re.kr（日本語可）

出所：韓国観光公社

그림 9 한국관광공사 일본어 홈페이지 홍보 내용



그림 10 신규 일본향 한/일 자동통역 ‘지니톡’ 앱



그림 11 업데이트된 국내향 한/일 자동통역 ‘지니톡’ 앱



(한국관 장면)



(지니톡 홍보 부스 장면)



(미디어데이 언론사에 홍보 장면)



(한국관내 부스에 방문한 일본인)



그림 12 일본 세계여행박람회(JATA) '지니톡' 홍보



그림 13 자동통역 Open API 신청/승인 관리 화면

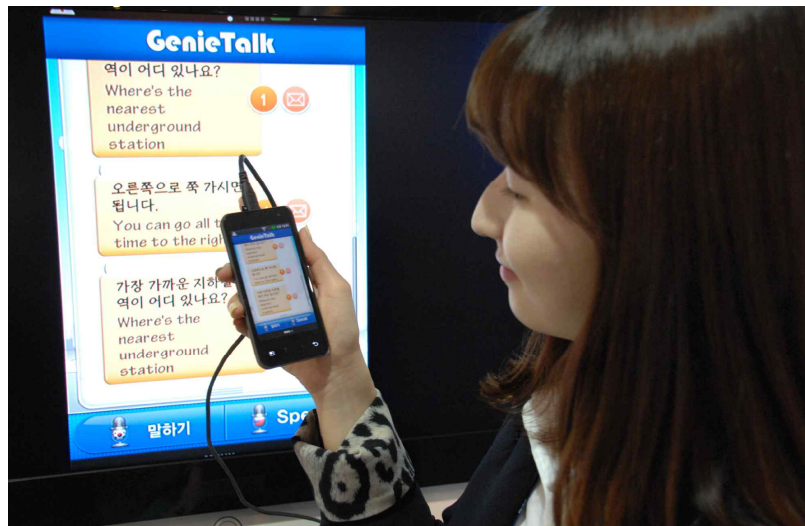


그림 14 업그레이드 버전 국내향 한/영 자동통역 ‘지니톡’ 앱 시연 장면



그림 15 지니톡 기반 한국인 대상 영어 교육용 앱 서비스 구현 화면
(지원업체: 컨피테크)



그림 16 해외 여행객을 대상으로 방문도시에 특화가 용이한 자동통역 서비스 구현 화면 (지원업체: 스킵미디어)

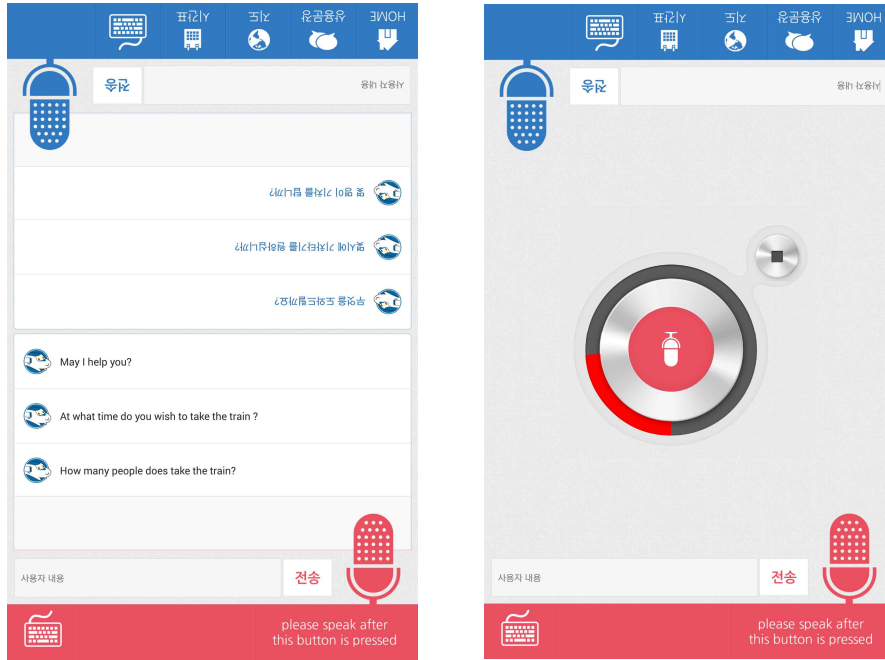


그림 17 KTX 역내 매표창구에서 언어소통이 가능한 통역 서비스 구현 화면 (지원업체: 디지쿼터스)



그림 18 신규 한/중 자동통역 대국민 서비스용 앱

다. 세부 추진실적 (정량적)

1) 특허/논문 성과

(단위 : 건수)

구분	과제원 총원 (M/Y)	특허				논문					
		국제		국내		국제			국내		
		출원	등록	출원	등록	SCI, SSCI	학술	기타	국내 저널	학술	기타
1차년도 (2013)	14.16	5	0	5	0	(1)	0	0	1	1	0
총 계	14.16	5	0	5	0	(1)	0	0	1	1	0

※ () 심사중

- ‘핸즈프리 자동통역’ 기술은 미국, 일본, 중국에 국제특허 출원 및 국제표준화 추진
- 특허리스트

번호	종류	명칭	출원 일	등록 일	국명	등록번호	발생차 수
1	국내	핸즈프리 자동통역	20130821		한국	출원중	
2	"	디바이스 의존누적적용을 이용한 음성인식 시스템 및 그 방법	20130917		한국	2013-0111918	
3	"	멀티센서 데이터의 모델분류를 이용한 음성인식시스템 및 그 방법	20130917		한국	2013-0111673	
4	"	한일 자동통역 언어의 발성음을 상대국 언어로 표기하는 방법	20130729		한국	2013-0089649	
5	"	특화영역 교체형 음성인식 시스템 및 서비스	201311		한국	출원중	

		스	128				
6	국 제	Handsfree speech translation apparatus	20130 821		미국, 일본, 중국	출원중	
7	"	Speech recognition system and method using model classification of multi-sensor data	20130 528		미국	출원중	
8	"	Speech recognition system and method using incremental device-based model adaptation	20130 514		미국	출원중	
9	"	Additional transcription of the pronunciation into the other party language in Korean-Japanese automatic speech translation processes	20130 510		미국	출원중	
10	"	Automatic speech recognition system and service based on replaceable specific domain search network	20131 128		미국	출원중	

○ SCI 논문

- (ETRI 저널) Handling Negative and Interrogative Expressions in Korean-to-Japanese Statistical Machine Translation (submitted)

2) 표준화 성과

- ‘핸즈프리 자동통역’ 기술에 대해 ITU-T SG2에 Preliminary Work Item으로 제출
- 기고서 작성하여 ITU-T SG2에 제안 진행 중임
- 일본 NICT 표준화전문가 Dr. Hori와 회의를 통해 상기 표준화에 공동협력하기로 함

3) 기술이전

구 분	제 목	주 요 내 용	비 고
S/W	실서비스 환경에 강인한 한국어 대화체 음성인식 용 언어모델 기술	지니톡 로그데이터 기반 성 능개선된 언어모델 기술전수 (전수업체: CSLI)	착수료(500만원)를 낮게하고, 매출액 대비 로열티(매출액 2.5%)를 높게 책정하여 계약함
S/W	한국어 공통음성 DB	영어 대화체 음성DB (엔에이 치엔)	착수료 1,000만원

제 4 장 사업화 계획 및 기대 효과

제 1 절 활용방안 및 사업화 계획

1. 활용 방안

가. 기술평가

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

표 6 자동통역 기술 평가

나. 활용(상용화) 가능성

구 분	구체적인 단계 및 내용
○ 기술의 응용 분야 및 활용 방법	<ul style="list-style-type: none"> ○ 자동통번역 기술은 최근 세계화에 따른 언어소통 불편을 해소하는 기술로, 스마트폰 통역, 통역전화기, 통역 로봇, 교육용 SW, 비즈니스용 통역 등 응용분야가 매우 다양함 ○ 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
○ 적용상의 애로점과 극복(개선) 방안	○ 잡음환경에서의 자동통번역기 성능 저하는 오류를 빈번히 발생시켜 사용상의 불편을 가져올 수 있음 ○ 자동통번역기의 오류를 사용자가 직접 수정할 수 있는 멀티모달 기반의 UI/UX 기술이 보완되어야 타 기술과 차별성을 가질 수 있을 것임
○ 제품/서비스의 예상 수요자(층)	○ 예상수요자 <ul style="list-style-type: none"> • 모바일 기기 동시통역 SW 및 내외국인 입/출국자를 대상으로 한국어를 중심으로 2013년 영어, 중국어 일본어에 대한 통역 세계시장 631억원(점유율 1%) 점유, 다국어 확장에 의해 2017년 7조원(점유율 45%) 점유로 세계 1위 가능 ○ 수요업체 <ul style="list-style-type: none"> • 스마트폰 제조업체: 삼성전자, LG전자, HTC(대만), 도시바(일본) 등 • 포털업체: 다음, NHN, 파란 등 • 통신사업자: SKT, LG U+, KT 등 • 중소기업: 파인디지털, 아이리버, 샤프(일본) 등 어학 학습기 업체
○ 제품/서비스화하여 시장에 도입되기 까지 요구되는 시간의 정도	○ 자동통번역 기술은 기술개발 단계로는 발아기(Technology Trigger)이나 2015년 초에 자동통역기술 실용화가 활발하게 이루어 질 것으로 예상함 ○ 현재 자동통번역 시장은 세계적으로 기술개발 후 시장이 형성되고 있는 초기단계에 해당

표 7 자동통역 기술의 활용가능성

2. 사업화 계획

가. 시장 및 경쟁

구분	구체적인 내용
○ 내수 시장 창출효과	<ul style="list-style-type: none"> ○ 해당 기술의 국내 시장(생산기준)은 2013년 약 720억 원 수준에서 2020년에는 약 1,328억 원 정도가 될 것으로 전망, 연평균 21.2%의 성장추세 ○ 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 ○ 다국어로 확장할 경우, 한/영 자동통역 시장의 10배 이상 시장창출 효과가 있을 것으로 예상
○ 수출효과/ 수입대체효과	<ul style="list-style-type: none"> ○ 자동통번역 세계시장은 2010년 5.8억불 수준에서 2020년에는 44.6억불 정도가 될 것으로 전망, 연평균 22.7%의 성장추세
○ 잠재적/현재 적 경쟁자와 그들이 갖고 있는 기술/ 제품 특성	<ul style="list-style-type: none"> ○ 현재 경쟁자는 구글, 애플, 뉘앙스임 ○ 구글의 경우, 14개국 이상 다국어 자동통역 기술 확보하였음. 대용량 서비스 로그데이터를 기반으로 점진적으로 성능을 개선하고 있음 ○ 애플의 경우, Siri라는 영어기반 대화형 인터페이스 기술을 확보하였고, 조만간 다국어로 확장중임 ○ 뉘앙스는 20여개국 대상 음성인식 기술을 확보하여 음성인식 시장을 독점하고 있음

표 8 자동통역 기술의 시장 및 경쟁현황

나. 자동통역 응용서비스의 상용화 계획

구분	구체적인 내용
○ 형태/ 규모	○ 본 과제에의 결과물을 상용화하기 위한 상세 사업계획을 수립하고, 상용 제품 개발 인력을 별도 투입하여 전담 마케팅 인력 투입과 별도의 마케팅 비용을 책정할 계획 ○ 실 서비스 사업 방안 도출은 과제 진행 중 전문 인력을 통해 수행 예정
○ 상용화 능력 및 자원보유	○ 엘앤아이소프트, CSLI, 디오텍 등 국내업체는 자동통역 소프트웨어 개발, 판매 및 서비스사업을 수행해 왔으며, 현재 다음과 같은 상용화 전문인력 및 관련자원을 확보하고 있음
○ 상용화 계획 및 일정	○ 2013년 <ul style="list-style-type: none"> • 통역 응용 앱 시장성 조사 • 플랫폼, API 구현 • 중소기업 BM 프로토타입 서비스 구현 ○ 2014년 이후 <ul style="list-style-type: none"> • 통역응용 앱 개발을 통한 통역시범 운용후 사업성 평가

표 9 상용화 계획

다. 국내외 시장 규모 및 수출·입 현황

(단위 : 백만원)

년도	(2012년)	(2015년)	(2018년)
	현재년도	개발 종료후 1년	개발 종료후 3년
세계 시장 규모	-	27	61
한국 시장 규모	-	63	142
년도	(2010년)	(2011년)	(2012년)
	2년 전	1년 전	현재년도
수출 규모	-	-	-
수입 규모	-	-	-

표 10 국내외 수출입 현황

- 1) 자동통역 기술은 발아단계의 기술로서, 2012년 시장이 활성화될 것으로 예상(Gartner 기술전망 Hype-cycle, 2006)
- 2) 본격적인 시장이 형성되기 전인 2011년까지 기존 초동단계의 제품에 과제에서 개발하는 요소기술을 지원하여 차별화된 제품 매출을 유도함
- 3) 상용화가 예상되는 2012년 기준으로, 전체 출입국자 3천만명의 2%가 자동통역서비스(이용료 만원) 이용할 경우 60억 규모의 시장 창출을 시작으로 매년 50% 신장 예측 (출입국관리현황자료, 2007.8). 출입국자 국적 기준 비율(7:3)을 적용하여 한국시장 및 세계시장을 산정

라. 국내·외 주요 수요처 현황

수요처	국명	수요량	관련제품
단말기 제조사, 통신/포털업체 (삼성전자, LG전자, NHN, KT)	한국, 미국, 일본	-	MID, UMPC, 스마트폰

표 11 주요 수요처 현황

제 2 절 기대효과 및 파급효과

1. 기대효과

- 가. '12년 여수 엑스포, '14 인천 아시안게임, '18 평창 동계올림픽에 자동통번역 서비스 실시를 통해 언어장벽 없는 국가 실현 및 국격 제고
- 나. ETRI의 자동통번역 플랫폼 및 Open API를 적극 활용하여 새로운 BM 개발 기회가 확대되고 이에 따라 대학, 중소기업, 응용SW 개발자, 수요업체(단말기, 포털사, 주력산업 제조업체 등)간의 산업생태계 조성
- 다. ETRI와의 기술협력 및 향후 다국어 처리 기술지원을 통해 향후 3~4년내 국내 SW기업의 기술력이 글로벌 경쟁력을 갖추게 될 것임
- 라. 중소기업과 자동통번역 플랫폼 및 Open API 기반 신규 BM을 개발하여 시범서비스 실시 및 지속적 성능 개선을 통한 상용화를 시도할 예정임. 이로부터 모범사례를 만들어 추후 대상업체를 확장해나갈 계획임
- 마. 공공재적 성격의 기술개발, 융합기술 관련 경쟁력을 기 확보하고 있는 ETRI와의 유기적 협업체계를 통하여 국내기업의 기술경쟁력 보장 및 기술개발 리스크 축소
- 바. 출연(연)이 상대적으로 취약한 기술사업화 부문에서 출연(연) R&D성과 확산→ 국내기업 R&D지원 강화→ 자동통번역 산업생태계 구축 등 상생협력 선순환 고리 정착
- 사. 자동통번역 공통플랫폼 지원, 공통 언어자원 구축 및 배포, 자동통번역 Open API(Application Program Interface) 및 DB 표준화를 통한 산업생태계 구축. 이에 따라 정부의 중점 현안인 일자리 창출이 가능하고 단순창업이 아닌 경쟁력을 갖춘 기술창업이 활성화되는 계기를 마련함으로써 기술집약형 글로벌 기업으로 성장할 수 있는 초석으로 작용
- 아. 다학제간 연구, 다국어기술 개발을 위한 국제공동연구 등 개방형 R&D 추진을 통한 창의 idea 확산 및 글로벌 창의인재 확보
- 자. 산/학/연 공동연구 활성화를 통해 ETRI 기술 인프라 공동 활용이 가능하므로 인프라가 부족한 기업의 자생력 강화 및 산업 활성화 가능

2. 기술적, 경제적, 사회적 파급효과

- 가. 세계화 시대에 요구되는 궁극적인 정보통신기술인 자동통역 기술 실현에 있어서 현재 가장 걸림돌이 되고 있는 대화체 음성인식 및 대화체 문장번역에 대한 한계를 극복함으로써, 자동통역 실용화 기반을 구축하여 새로운 정보통신산업 블루오션을 창출할 것임
- 나. 우리나라가 선도하고 있으나 언어장벽으로 인하여 세계시장 공략에 애로를 겪고 있는 디지털콘텐츠, 온라인게임, 전자상거래 등 정보통신서비스 및 문화콘텐츠 산업에 대한 글로벌 비즈니스 인프라를 구축함으로써 관련 산업의 세계화를 통한 시장 확대 및 활성화에 기여함
- 다. 자동통역의 구성 기술요소인 음성인터페이스 및 언어번역 기술은 정부의 성장동력산업 중 지능형로봇, 디지털방송, 홈네트워크, 차세대 PC 등 산업에 사용자 인터페이스 및 서비스 구현에 직접 활용할 수 있어, 관련 산업의 육성에 커다란 기여가 기대됨
- 라. 2000년대 초 벤처열풍을 타고 설립된 많은 음성언어정보 산업체들이 현재 기술력 부족으로 시장이 축소, 정체되면서 어려움을 겪고 있음. 본 과제에서 기술개발을 통하여 자동통역이라는 킬러 어플리케이션(killer application)을 창출함으로써 관련 산업을 크게 부흥시킬 수 있을 것임
- 마. 자동통역 기술은 언어소통에 불편을 겪고 있는 잠재적 수요자의 수가 헤아릴 수 없이 많으므로 사용자의 요구수준을 만족시킴으로써 엄청난 규모의 시장창출이 기대됨
- 바. 모바일 환경이 도래하면서 대부분의 사람들이 휴대폰을 비롯하여 UMPC, MID, PDA, DMB폰, 게임기 등 다양한 모바일 기기를 보유하고 있으며 자동통역 S/W는 추가적인 장비 구입비용 없이 S/W를 다운로드하여 사용할 수 있음. 국내 휴대폰 보급률이 78%를 넘었으며(한국전자거래진흥원 2006년 3월 월간보고 자료), 기타 모바일 기기를 포함하면 최소 1인 1대 모바일 기기가 보급됨. 이에 전체 기기의 3%의 사용자가 자동통역 소프트웨어(패키지 당 3만원 예상)를 사용할 경우 450억원 규모의 단말용 자동통역 소프트웨어 시장을 창출할 것임
- 사. 모바일 기기 동시통역 SW 및 내외국인 입/출국자를 대상으로 한국어를

중심으로 2013년 영어, 중국어 일본어에 대한 통역 세계시장 631억원(점유율 1%) 점유, 다국어 확장에 의해 2017년 약 7조원(점유율 45%) 점유로 세계 1위 가능

	'13년	'14년	'15년	'16년	'17년	'18년
매출액(억원)	63,100	79,632	100,496	126,826	160,054	201,988
수출액(억원)	631	2,389	5,025	53,267	72,024	96,954
세계시장점유율(%)	1 (예상순위 5위 미만, 1위는구글)	3 (예상순위 5위 미만, 1위는구글)	5 (예상순위 5위, 1위는 구글)	42 (예상순위 2위)	45 (예상순위 1위)	48 (예상순위 1위)
고용창출(명)	131	496	1,043	11,053	14,945	20,118

표 12 국내외 시장규모 예상

- ※ 1단계(2011~2013) 개발 및 사업화를 통해 세계시장 2013년 1%, 2015년 5% 산정, 2단계(2014~2016) 다국어 확장을 통한 세계시장 40% 이상으로 확대. 매출액 및 수출액은 2012년 시장 규모(50억불, 연 23.1% 성장) 대비 점유율을 근거로 산정
- ※ SW산업(컴퓨터관련서비스)의 취업유발계수(명/10억당) 16.5명, DB 구축의 경우 25명 적용(한국은행, 2005년기준 산업연관표 부속 고용표), 자동통번역 SW의 경우 개발 및 DB구축 비중이 50%씩 산정
- 아. 2020년 전세계 자동통역 시장규모는 자동통역서비스 단말기 6조 3,000억원, 비즈니스용 자동통역기 2조 5,000억원, 교육용 자동통역기 1조 2,000억 원으로 총 10조원으로 예측 (출처: 일본 UFJ총합연구소 마케팅 조사보고서, 2006)
- 자. 세계화(Globalization)가 가속화 되고 있는 상황에서 자동통역 기술의 개발은 언어장벽의 해소를 통하여 보다 다양한 분야의 사람들이 국제협력 등 활동을 할 수 있게 되며, 한류의 확산, 외국인의 국내활동 기반 강화, 국제 무역의 확대, 관광산업의 진흥 등 국제화를 통하여 국제사회에서의 우리나라의 위상을 높이고 국가 경쟁력을 강화시킴
- 차. 자동통역 기술이 적용된 외국어 교육도구 개발로 외국어 교육비 부담을 대폭 경감할 수 있음. 2006년 어학연수와 유학비용으로 해외에 지출된 돈이 4조 4천억 원에 이르고, 영어 사교육비만 10조원이 훨씬 넘는 등 외국어 교육비 부담이 상당함. 이러한 문제에 대하여 자동통역/번역 기술을 기반으로 교육도구 개발 등으로 교육비 부담 경감이 가능함

[부록 1] 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

[부록 2] 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

[부록 3] 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
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

본 문서에서 음영 처리된 부분은 () 정보공개법 제9조의 비공개대상정보와 저작권법 및 그 밖의 다른 법령에서 보호하고 있는 제3자의 권리가 포함된 저작물로 공개대상에서 제외되었습니다.

주 의

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