

## 자동통역기술, 서비스 및 기업 동향

Trends in Automatic Speech Translation Technology, Service and Business

김승희 (S.H. Kim)      자동통역연구실 책임연구원  
박준 (J. Park)         자동통역연구실 책임연구원  
김상훈 (S.H. Kim)     자동통역연구실 실장

### 소프트웨어 기술동향 특집

- I. 서론
- II. 자동통역기술 개요
- III. 자동통역기술 발전의 역사
- IV. 대국민 자동통역 시범서비스 지니톡
- V. 국내 특허동향 및 관련 기업동향
- VI. 결론

최근 음성인식, 자동통역, 인공지능 관련 분야에서 글로벌 기업들의 경쟁이 치열하다. 서로 다른 언어를 사용하는 사람들 간의 교류가 급속히 확장되고, 관련 서비스들이 속속 제공됨에 따라 자동통역에 대한 일반인들의 관심도 급증하고 있다. 이에 자동통역 시장경쟁이 본격화되고 있다. 본고에서는 자동통역 요소기술들의 개념 및 최근 동향에 대해 기술하며, 대국민 자동통역 시범 서비스인 지니톡에 대해 기술한다.

## I. 서론

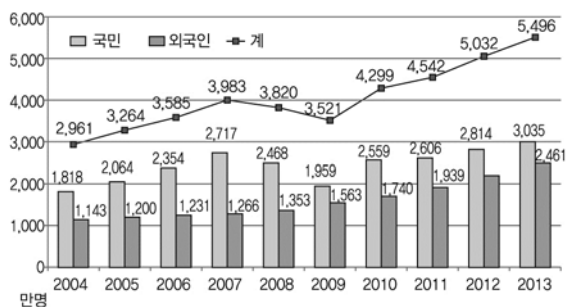
교통, 통신수단이 발달함에 따라 각국 간의 인적, 물적 교류가 활발해져 왔다. 최근의 교류 규모는 예전에 비해 훨씬 커졌다. 법무부는 2013년 국내 출입국자가 5천 4백만명을 넘었다고 발표하였는데, 이는 최근 10년 만에 약 2배 증가한 수치이다[1](그림 1) 참조).

서로 다른 언어를 사용하는 개인이나 집단 간에 교류가 일어날 경우 가장 먼저 부딪히게 되는 문제 중 하나가 언어소통에 대한 것이다. 이 때문에 전문번역사나 통역사에 대한 수요는 오래 전부터 존재해 왔다. 고려시대, 조선시대에는 역관이라는 관직이 있어서 번역, 통역 등의 업무를 수행하였다. 현대에는 각종 주요 국제회의나 정상회의 등에서 전문통역사들이 활발히 활동하는 모습을 볼 수 있다. 회의 규모에 따라서는 수십 명 이상의 전문통역사들이 투입되기도 한다.

하지만, 수요에 비해 전문가의 숫자는 부족하고, 또한 통/번역에 소요되는 경제적 부담도 크다. 그리고 일반인의 경우에는 이들의 도움을 받는 것 자체가 쉽지 않다.

2013년 국내 입국 외국인 중 중국인과 일본인의 숫자가 가장 많은데[1], 이들이 한국에서 쇼핑할 때 겪는 불편한 점 중에 언어소통 문제가 가장 컸다(〈표 1〉 참조)[2].

통역에 대한 수요에 비해, 자동통역 연구가 본격적으로 시작된 것은 오래되지 않았다. 개별 요소 기술들 자체가 가지고 있는 난이도 때문이라 하겠다.



(그림 1) 최근 10년간 출입국자 증감 추이(승무원 포함)[1]

〈표 1〉 한국에서 쇼핑 시 불편사항(단위:%)[2]

	중국인 관광객	일본인 관광객
1위	언어소통 불편(57.3)	상품구입 강요(29.3)
2위	안내표지판 부족(34.0)	언어소통 불편(22.7)
3위	불편한 교통(21.3)	안내표지판 부족(21.3)
4위	비싼 가격(17.3)	종업원 불친절(16.7)
5위	종업원 불친절(12.7)	비싼 가격(10.0)

최근 국내에서 자동통역에 대한 일반인들의 관심이 부쩍 커졌는데, 이는 자동통역에 대한 필요성이 커진 것과 함께 구글(Google)의 자동통역 서비스나 ETRI의 지니톡(GenieTalk)에 기인한 바가 크다고 판단된다.

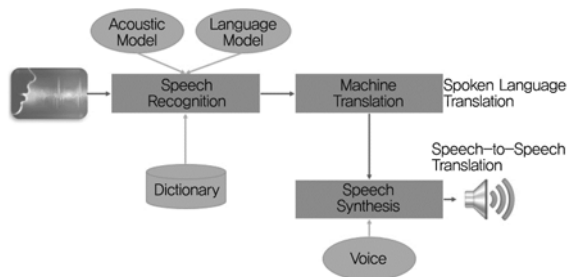
본 논문의 나머지 부분은 다음과 같은 내용들을 설명한다. II장에서는 자동통역 관점에서 각 요소 기술들의 개념 및 최근 동향과 요소 기술의 통합에 대해 설명한다. 그리고, III장에서는 자동통역기술 발전의 역사를 간략히 기술한다. IV장에서는 지니톡에 대해 설명하며, V장에서는 자동통역과 관련된 국내 특허동향 및 기업들의 최신 동향을 기술한다. 그리고 결론을 맺는다.

## II. 자동통역기술 개요

### 1. 개요

자동통역은 어떤 언어 A로 된 사람의 발화를 음성인식, 자동번역 등의 과정을 거쳐서 다른 언어 B로 변환하고, 이를 자막으로 출력하거나 혹은 음성합성 후 스피커를 통해 들려주는 과정 및 기술을 의미한다.

(그림 2)에 자동통역 시스템의 대략적인 구성을 도시하였다. 자동통역 시스템은 사람의 발화를 입력으로 한다. 음성인식 모듈이 입력되는 음성을 처리하는 과정을 담당한다. 일반적으로 음성인식기의 결과는 문자열이다. 언어 A로 된 문자열을 언어 B의 문자열로 변환하는 과정을 담당하는 것이 자동번역기이다. 자동통역 시스템의 출력은 응용분야에 따라 달라진다. 서로 다른 언어를 사용하는 사람들 간의 통역에서는, 번역된 텍스트를



(그림 2) 자동통역 시스템의 구성

음성합성 모듈을 이용하여 음성신호로 합성한 후 스피커를 통해 음성으로 출력한다. 이 경우를 speech-to-speech translation이라고 한다. 반면에, 방송뉴스를 다른 언어로 통역하는 경우에는 음성합성 모듈을 거치지 않고 번역된 결과를 자막으로 내보낼 수도 있다. 이런 시스템은 spoken language translation system이라고 한다[3]. 하지만, spoken language translation을 일반적인 의미의 자동통역으로 사용하기도 한다.

방송뉴스 통역과 같은 경우에는 언어 A에서 언어 B로만 통역이 이루어진다. 이런 형태의 자동통역을 단방향(one-way) 자동통역이라고 한다. 반면에 지니톡과 같은 대면(face-to-face) 형태의 자동통역에서는 A→B, B→A의 방향으로 자동통역이 이루어지게 된다. 이런 형태를 양방향(two-way) 자동통역이라고 한다.

## 2. 음성인식

음성인식은 사람의 입으로부터 나온 음성신호를 자동으로 인식하여 문자열로 변환해 주는 과정 혹은 기술을 의미한다. ASR(Automatic Speech Recognition), voice recognition, 혹은 STT(Speech-to-Text)라고 한다.

현재 대부분의 음성인식은 확률통계 방식에 기반하고 있다. 음성인식 과정에서 사용되는 음향모델(AM: Acoustic Model), 언어모델(LM: Language Model)로 확률통계에 기반한 모델을 쓰고 있다. 그리고, 핵심 알고리즘인 HMM(Hidden Markov Model)도 역시 확률통계에 기반한 것이다.

음향모델로 대표적인 것이 GMM(Gaussian Mixture Model)이며, 언어모델은 N-gram이 대표적이다. 하지만 최근에는 GMM을 대신하여 딥 러닝(deep learning) 아키텍처 중 하나인 DNN(Deep Neural Network)이 확산되고 있다.

음성처리나 영상처리 등의 여러 분야에서 딥 러닝은 기존의 최고 성능보다 훨씬 나은 성능을 보여주고 있다. 가트너(Gartner)는 2014년 주목할만한 기술분야 중 하나로 딥 러닝을 꼽았다. 2017년이 되면 컴퓨터의 10%는 데이터 처리가 아닌 딥 러닝 기반의 학습을 하고 있을 것이며, DNN 알고리즘을 활용하는 음성인식 애플리케이션은 2배로 늘어날 것이라고 예측하였다[4].

구글은 안드로이드 OS 4.1인 Jelly Bean의 음성검색 서비스에 DNN을 적용하였고[5], 마이크로소프트(Microsoft)도 자사의 음성인식 서비스인 Bing 음성검색에 DNN을 적용하였다[6]. 국내에서는 네이버가 음성인식 서비스에 DNN을 적용하였다.

자동통역 관점에서의 음성인식은, 시스템의 구조적인 특성상 다른 응용분야에 비해 어려운 요소들이 있다[7].

우선, 음성검색과 같은 분야에 비해 음성인식의 정확도가 더 높아야 한다. 음성인식 모듈의 출력이 자동번역기의 입력이 되는데, 자동번역 모듈이 음성인식 오류에 대해 검색 등의 언어처리 모듈보다 더 민감하기 때문이다. 게다가, 자동통역이 처리대상으로 하는 음성들이 대체로 대화체의 음성이라는 점은 음성인식의 난이도를 높이는 요소이다.

그리고, 음성인식 모듈은 단순히 입력된 음성에 대응하는 전사(transcription)뿐만 아니라 문장 경계에 대한 정보도 제공해야 한다. 문장 단위로 구분되지 않은 텍스트를 처리할 경우 번역성능이 떨어지기 때문이다. 하지만, 방송뉴스나 강연, 강의 등의 연속된 음성입력에 대해 자동으로, 더군다나 지연시간을 최소화하면서 문장 경계를 찾는 것은 쉬운 일이 아니다.

또한, 음성인식의 성능을 높이기 위해서는 양질의 음

향모델, 언어모델을 학습해야 하고, 이를 위해서는 다량의 학습 DB가 필요하다. 그러나, 자동통역이 대상으로 하는 대화체의 음성이나 언어 DB를 확보하기가 쉽지 않다.

### 3. 자동번역

자동번역은 언어 A로 쓰여진 문장을 다른 언어 B의 문장으로 변환해 주는 기술이다. 대표적인 방법론으로 규칙에 기반한 방법과 말뭉치에 기반한 방법이 있다.

규칙 기반 방법은 분석 깊이에 따라 다시 직접 번역방식이나 간접 변환방식, 중간 언어방식으로 나눌 수 있다 [8].

직접 번역방식은 처음 자동번역 시스템을 개발할 당시에 많이 쓰였다. 주로 한국어와 일본어처럼 문장구조가 비슷한 동종 언어 쌍 간에 적절한 방식이다. 하지만 영어나 한국어처럼 문장구조가 다른 이종 언어 간에는 적용하기 어렵다는 단점이 있다.

간접 번역방식은 구문이나 의미구조까지 분석한 다음, 번역 문장을 만들어내는 점에서 직접 번역방식과 차이를 나타낸다.

간접 변환방식은 다국어 자동번역을 위해 다수의 변환모듈을 필요로 한다. 중간 언어방식은, 문장을 분석하여 특정 언어에 의존적이지 않은 의미 표현 형태의 중간 언어로 바꾼 다음, 최종 목적 언어로 번역하는 형태다 ((그림 3) 참조).

언어 쌍에 따른 변환모듈을 필요로 하지 않는 장점이 있으나[8], 의미분석모듈을 수작업으로 개발해야 하고, 모든 언어에 공통되는 의미표상을 설계하기가 힘들다는

단점을 가지고 있다[7].

말뭉치에 기반한 방법론으로는 예제 기반 방법[9]과 통계기반 방법이 있다.

통계 기반 자동번역(SMT: Stochastic Machine Translation)기술은 통계적 분석을 통해 이중언어 말뭉치로부터 모델 파라미터를 학습하여 문장을 번역하는 기술이다. 문법이나 의미표상을 개발할 때 수작업으로 하지 않고 번역하고자 하는 언어 쌍에 대한 말뭉치로부터 번역에 필요한 모델을 만든다. 그래서, 말뭉치만 확보할 수 있다면 비교적 용이하게 언어확장을 할 수 있다. 단점은, 대규모의 이중언어 말뭉치가 필요하고, 다수의 언어들을 연결하는 공통된 의미표상이 없다는 것이다[7].

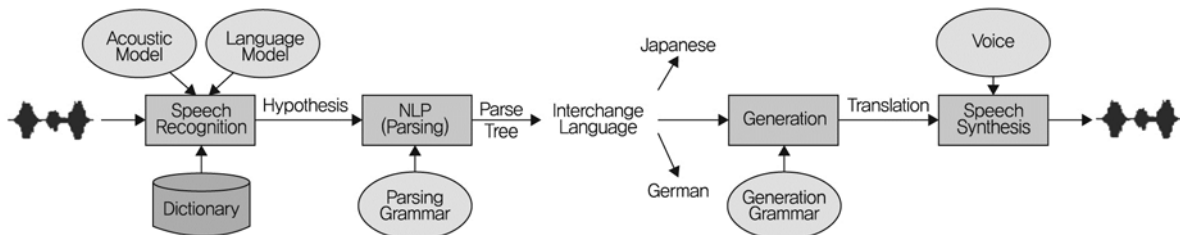
제한된 도메인에서는 중간 언어의 설계가 가능하고 유용할 수 있으나, 도메인 제한이 없는 경우에는 일반적으로 SMT를 선호해 왔다[7].

### 4. 음성합성

음성합성은 글자로 된 말을 소리로 바꿔주는 기술이다. TTS(Text-to-Speech) 혹은 voice synthesis라고 한다.

대표적으로 쓰이는 기술이 음편조합방식이다.

먼저 문장을 분석한다. 분석된 결과에 맞는 음편들을, 미리 만들어 놓은 대규모의 음편 DB에서 가져와서 이어 붙인다. 분석된 결과에 대응되는 음편들이 다수 존재하고, 이들을 다양한 방식으로 붙이게 되면 합성음으로 출력할 후보들이 많아진다. 이들 중 운율, 연결했을 때의



(그림 3) 중간 언어방식의 자동통역 시스템 구조[7]

매끄러움 등의 기준으로 가장 적합한 것을 고른다. 그리고, 이것을 후처리한 후 스피커를 통해 출력하게 된다.

자동통역 관점에서는 발화자의 음성과 최대한 가깝도록 합성음을 생성하는 방식이 관심을 끌고 있다. 발화자의 감정을 인지하여 합성음에 감정을 싣거나, 발화 스타일이나 음색을 반영하는 방법들이 연구되고 있다[7].

### 5. 요소 기술의 통합

자동통역의 난이도는 다음의 원인들 때문에 일반적인 자동번역보다 높다.

우선 번역기의 입력에 오류가 포함된다. 일반적인 자동번역에서는 입력되는 텍스트에 오류가 없다고 볼 수 있다. 하지만 자동통역에서는 음성을 텍스트로 변환하는 과정에서 이미 상당한 오류가 발생한다.

또한, 자동통역은 사람의 말을 취급한다. 그런데, 사람의 말은 글로 쓰여진 문장과는 아주 다르며 상당히 비문법적인 형태를 띠고 있다. 이 때문에 자동통역을 위한 자동번역은 난이도가 높다.

그리고, 텍스트에는 포함되어 있는 구문론적, 의미론적 단서들, 예를 들어 문단의 구분이나, 구두점, 문장부호, 대소문자 정보 등이 음성인식 결과에는 포함되어 있지 않다. 즉, 텍스트 분석 과정에서 유용하게 사용될 많은 정보들이 음성인식 결과에는 없다[10][11].

이 때문에 음성인식과 자동번역을 통합하여 성능을 개선하고자 하는 연구들이 진행되어 왔다[11]-[17].

가장 단순한 구조의 자동통역기는 음성인식기와 자동번역기를 순차적으로 배치하고, 단일의 음성인식 결과를 번역기의 입력으로 전달하는 것이다.

하지만, 음성인식 결과에 오류가 포함되기 때문에 가급적 많은 인식 후보들이나 관련 정보를 번역기의 입력으로 전달하는 방식을 생각할 수 있다. 예를 들어 N-best 인식결과[14], confusion network[15], word lattice[13][16][17] 등을 전달하는 것이다. 또한, 참고문

〈표 2〉 EPPS\* 스페인어-영어 통역과제에서의 통역결과: 각 인터페이스의 비교(단위: %)[10]

INTERFACE	WER*	PER*	BLEU*
VERBATIM(NO ASR ERROS)	41.0	31.2	48.0
SINGLE BEST	50.0	39.2	37.6
10 BEST	49.5	38.7	38.6
100BEST	49.1	38.6	38.9
CONFUSION NETWORK	49.5	38.6	39.2

\* EPPS(European Parliament Plenary Session)

\* PER(Phone Error Rate)

\* WER(Word Error Rate)

\* BLEU(Bilingual Evaluation UnderStudy)

헌 [18]에서는 기존의 단어 기반 정보전달 대신에 소리에 해당하는 음소 기반의 정보를 전달하는 방식을 제시하고 있다.

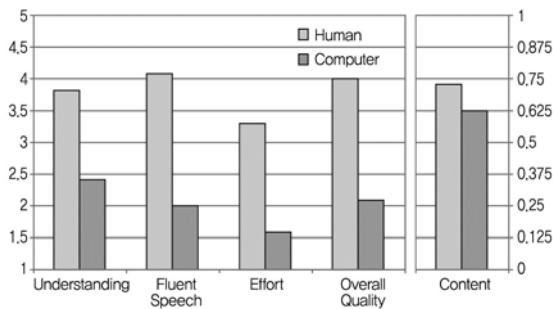
〈표 2〉에서 각 인터페이스별 성능의 차이를 볼 수 있다.

SMT가 보편화되면서 음성인식과 통합된 확률통계 방식을 적용하고자 하는 연구들이 진행되어 왔다[11][16]. 즉, 성능향상을 위해 음성인식과 자동번역의 과정을 결합함으로써 자동통역을 하나의 단일 작업으로 보는 것이다.

최근에는 음성입력뿐만 아니라 영상입력까지 함께 받아들여 처리하는 multi-modal 방식의 자동통역에 대한 연구도 이루어지고 있다[19]. 얼굴의 표정을 인식하여 화자의 감정상태를 파악하고, 그 감정상태를 음성합성 과정에 반영하여 감정표현을 시도하는 것이다.

### III. 자동통역기술 발전의 역사

자동통역은 인류의 오랜 꿈 중의 하나이다. 그러나, 자동통역 연구가 본격적으로 시작된 것은 불과 20여 년 전이다. 요소 기술들의 발전에 기초하여 1980년대 말과 1990년대 초의 기간 동안에 자동통역 연구가 본격적으로 시작되었다[7]. 그 이후 20년 동안 많은 연구개발 프



(그림 4) 통역전문가 vs. 자동통역 품질 비교[7]

로젝트들이 수행되었고, 시스템들이 개발되었다. <표 3>과 <표 4>에 주요 프로젝트들 및 시스템들에 대해 요약하였다.

대략의 발전방향을 요약하면, 데스크톱/랩톱 등의 플랫폼에서 모바일/클라우드로 진화하였고, 통역 대상 영역이 넓어지고 어휘수가 확장되었다. 또한 딱딱한 어투의 짧은 문장에서 대화체의 발화를 처리하는 방향으로 발전하였다.

(그림 4)는 TC-STAR에서 자동통역기의 품질을 평가한 내용을 요약한 것으로, 전문통역사에 의한 통역품질과 컴퓨터에 의한 통역품질을 비교한 것이다.

피험자들을 대상으로 통역결과에 대한 적합성을 판단하도록 하였다. 통역결과에 담긴 메시지를 이해할 수 있는가, 통역결과가 유창한가, 듣기 쉬운가, 전체적인 품

<표 3> 자동통역 프로젝트 및 평가활동들[11]

PROJECT/CAMPAIGN	Active period	SCOPE, SCENARIOS AND PLATFORMS
C-STAR	1992-2004	One/Two-way Limited domain spontaneous speech
VERBMOBIL	1993-2000	One-way Limited domain conversational speech
DARPA BABYLON	2001-2003	Two-way Limited domain spontaneous speech; laptop/Handheld
DARPA TRANSTAC	2005-2010	Two-way Small domain spontaneous dialog; Mobile/Smartphone
IWSLT	2004-	One-way; Limited domain dialog and unlimited free-style talk
TC-STAR	2004-2007	One-way; Broadcast speech, political speech; Server-based
DARPA GALE	2006-2011	One-way; Broadcast speech, conversational speech Server-based
DARPA BOLT	2011	One/two-way Conversational speech; Server-based

<표 4> 시스템 차원의 자동통역 성능향상 요약[7]

	YEARS	VOCABULARY	SPEAKING STYLE	DOMAIN	SPEED	PLATFORM	EXAMPLE SYSTEMS
First Dialog Demonstration Systems	1989-1993	Restricted	Constrained	Limited	2-10×RT	Workstation	C-Star-I
One way Phrasebooks	1997-Present	Restricted, Modifiable	Constrained	Limited	1-3×RT	Handheld	Phraselator, Ectaco
Spontaneous Two-way Systems	1993-Present	Unrestricted	Spontaneous	Limited	1-5×RT	PC/ Handheld Devices	C-Star, Verbmobil, Nespole, Babylon, Transtac
Translation of Broadcast News, Political Speeches	2003-Present	Unrestricted	Read/Prepared Speech	Open	Offline	PCS, PC-Clusters	NSF-Strudust, EC TC-Star, Darpa Gale,
Simultaneous Translation of lectures	2005-Present	Unrestricted	Spontaneous	Open	Realtime	PC, Laptop	Lecture Translator

질은 어떠한가 등의 관점에서 평가한 결과이다. 점수는 1(매우 나쁨)에서 5(매우 좋음)까지이다. <표 4>에서 다섯번째 결과는 정확도(%)에 대한 것인데, 통역결과의 내용과 관련된 질문에 피험자가 대답할 수 있는가에 대한 것이다.

자동통역의 품질은 전문가의 통역결과에 비해 현저히 떨어지지만, 전문가에 의한 통역결과와 비교했을 때 사용할 만하고 이해할 만한 수준에 도달한 것을 볼 수 있다[17]. 다만, 전문 통역사들은 실시간 통역을 하고, 또한 중요한 정보를 선택해야만 한다는 점과, 이에 비해 자동통역기들은 모든 정보를 처리하려고 하는 경향이 있다는 점을 고려해야 한다[20].

#### IV. 대국민 자동통역 시범서비스 지니톡

##### 1. ETRI에서의 자동통역 연구 역사

ETRI에서는 90년대 초부터 자동통역을 목표로 관련 분야의 연구, 개발을 수행하였다. KT와 함께 일본의 KDD와 다국 간 자동통역 전화 개발에 착수, 시연하였는데 우리나라 통역기술의 첫 응용사례라고 할 수 있다. 1995년에는 미국의 CMU(Carnegie Mellon University), 일본의 ATR, 이탈리아의 IRST, 프랑스의 CLIPS 등과 함께 C-STAR II를 결성하여 국제 공동연구를 진행하였다. 그리고, 1999년 여행분야에 한정하여 6개국 언어 간 국제 실시간 자동통역을 성공적으로 시연하였다[8].

하지만 이후 자동통역에 대한 연구는 정체기를 맞게 되었다. 요소 기술들인 음성인식, 자동번역, 음성합성 등에 대한 투자 및 연구가 집중되었고, 자동통역 자체에 대한 연구는 미미하였다.

##### 2. 지니톡 개발과정 및 현황

본격적인 자동통역 연구는 2008년에 다시 시작되었다. 2008년부터 4년간 휴대형 한/영 자동통역기술 개발

을 시작하였다. 2008년에 노트북에서 수행되는 1만 단어급, 2009년에는 MID(Mobile Internet Device)에서 수행되는 2만 단어급, 2010년에는 안드로이드 스마트폰에서 수행되는 5만 단어급의 자동통역기를 개발하였고, 2012년에는 안드로이드/아이폰에서 동작하는 27만 단어급의 서버기반 자동통역기를 개발하였다.

연구결과물을 바탕으로 하여 2011년 12월에 제주지역에서 한/영 자동통역 시범서비스를 실시하였다. 2012년 5월에는 여수 엑스포에 자동통역 서비스를 지원하였고, 2013년 순천정원 박람회, 충주국제조정경기대회를 지원하였다.

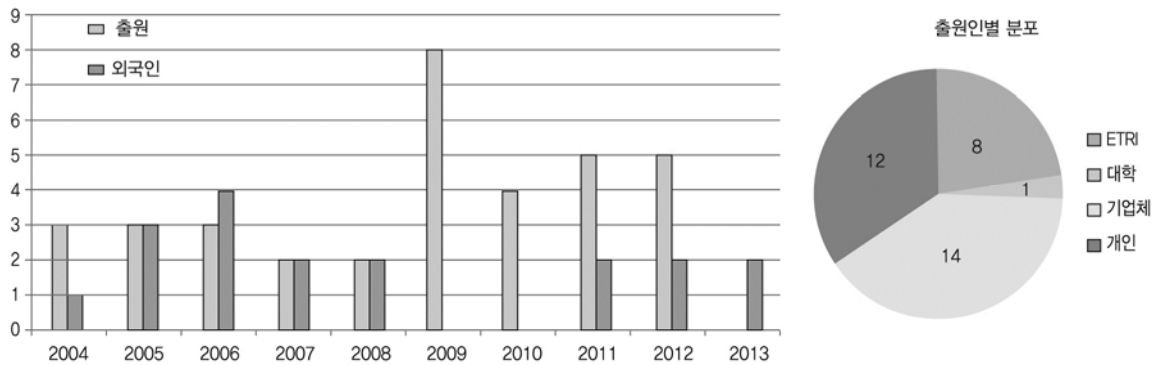
그리고, 2012년 10월에 대국민 한/영 자동통역 시범 서비스인 지니톡을 실시하였다. 2013년 5월에는 한/일 자동통역 시범서비스를 한국과 일본에 동시에 실시하였다. 또한 2013년 12월에 한/중 자동통역을 추가하였다.

지니톡의 다운로드 수는 최초 서비스 실시 후 7일만에 72만을 넘었고, 2012년 12월에 100만을 기록하였다. 2014년 5월 현재 180만건의 다운로드를 기록하였다. 하루 평균 15만건의 데이터가 누적되고 있고, 하루 3만명 이상 사용하는 것으로 추정된다.

#### V. 국내 특허동향 및 관련 기업동향

##### 1. 국내 특허동향

2004년부터 현재까지 발명 제목에 '통역'이 들어간 특허를 검색하여 분석하였다. 이 중에서, 수화통역, 동시통역 등 자동통역과 관련 없는 특허를 제외한 바, 공개 특허 35건, 등록 특허 18건이 검색되었다. (그림 5)에서 볼 수 있듯이 2005~2008년에는 출원이 미미하였고, 2009년 이후 증가 추세를 보이고 있다. 이는 ETRI에서 실용화를 위한 자동통역 과제 수행 및 지니톡 서비스 실시, 구글의 통역서비스 실시와 맞물려 있음을 시사하고 있다.



(그림 5) 국내 자동통역 관련 특허현황

출원인 분포에서는 개인 출원이 전체의 1/3 이상으로 많이 차지하고 있으며, 대학에서는 포항공대가 유일하게 1건, 기업에서는 삼성전자 3건, 모토로라 1건 등 차지하고 있다. '요약'으로 살펴본 발명의 내용은 주로 통역방법에 관한 것으로서 전체 시스템이나 서비스 구성에 대한 것이 많았다. 그리고, 스마트폰이 확산됨에 따라 이를 활용하는 발명과 통역 기능에 기반한 외국어 학습, 의료분야 특화 관련 특허도 눈에 띄었다. 이번 조사는 발명 제목에 '통역' 포함 여부로 검색하였으나, 음성인식, 번역 등 요소에서 자동통역을 다루는 특허도 다수 존재하여 이에 대한 추가 작업이 필요하다. 2012년 출원 특허가 현재 공개 중에 있으며, 2013년 이후 출원은 아직 미공개 상태라 통계에서 제외되었다.

## 2. 최신 기업동향

주요 거대 기업들 사이에 딥 러닝 기술확보를 위한 경쟁이 치열하다. 구글은 2013년 3월에, 자사의 DNN 성능을 개선하기 위하여 캐나다 토론토 대학의 DNN 전문가인 Hinton 교수가 세운 DNNresearch라는 신생기업을 인수하였다[21]. 또한, 2014년 초에는 인공지능 관련 신생기업인 DeepMind를 인수하였는데, 이 회사는 딥 러닝에 대한 연구를 진행해 왔다[22]. 페이스북(Facebook), IBM, 야후(Yahoo) 등도 딥 러닝 관련 경쟁력 확보에 열을 올리고 있다[23].

마이크로소프트가 무료 인터넷 화상전화 서비스인 Skype의 통화를 동시통역하여 제공하는 Skype Translator를 2014년 5월에 발표하였다. CEO인 Satya Nadella는 음성인식, 자동번역, 음성합성이 통합된 자동통역 시스템에 DNN 기술을 적용하였다고 밝혔다[24]. 페이스북은 2013년에 모바일 자동통역 앱인 지비고(Jibigo)의 개발사인 Mobile Technologies를 인수하였다[25]. 페이스북은 DeepMind를 인수하기 위해 구글과 경쟁하기도 하였다[26].

국내의 자동번역 솔루션 기업인 씨에스엘아이(CSLi)가 2014년 5월에 글로벌 유명 업체인 시스트란(Sys-tran)을 인수하였다. 시스트란은 프랑스 파리에 본사를 둔 세계 최대의 자동번역 솔루션 개발기업이다[27].

## VI. 결론

본고에서는 자동통역의 개념, 역사, 요소 기술들 및 최근 동향, 그리고 대국민 자동통역 서비스인 지니톡에 대해 알아보았다. 음성, 영상 등 여러 분야에서 최근 화두가 되고 있는 딥 러닝, DNN에 글로벌 기업들의 관심이 집중되고 있으며, 인력 및 기술확보 경쟁이 치열하다. 또한, 관련 분야의 주요 기업들이 발전된 요소 기술들에 기반하여 자동통역 서비스를 제공하고 있다. 자동통역분야의 경쟁이 본격화되고 있는 상황이다. 기술,



DB 확보 및 서비스 경쟁에서 뒤떨어진다면 다시 경쟁에 뛰어드는 것이 쉽지 않을 수 있을 중요한 시기이다.

#### 용어해설

**의미표상** 표상은 실세계와 대응되거나 그 대응을 가능하게 만들어주는 표현으로써 의미표상은 의미의 언어학적 표현을 말한다.

**이중언어 말뭉치** 같은 뜻을 가진 용례가 두 언어로 되어 있는 일정한 규모 이상의 크기를 갖추고 내용적으로 다양성과 균형성이 확보된 자료의 집합체

#### 약어 정리

AM	Acoustic Model
ASR	Automatic Speech Recognition
BLEU	Bilingual Evaluation UnderStudy
CMU	Carnegie Mellon University
DNN	Deep Neural Network
EPPS	European Parliament Plenary Session
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
LM	Language Model
MID	Mobile Internet Device
PER	Phone Error Rate
SMT	Stochastic Machine Translation
STT	Speech-to-Text
TTS	Text-to-Speech
WER	Word Error Rate

#### 참고문헌

[1] 법무부, “작년 출입국자 5천4백96만 명으로 사상 최대,” 2014. 1. 13.

[2] 대한상공회의소, “한국 방문 중일 관광객 쇼핑현황 실태 조사,” 2014. 5. 7.

[3] X. He and L. Deng, “Speech-Centric Information Processing: An Optimization-Oriented Approach,” *IEEE*, vol. 101, no. 5, May 2013.

[4] Gartner, “Gartner Reveals Top Predictions for IT Organizations and Users for 2014 and Beyond,” Oct. 2013. <http://www.gartner.com/newsroom/id/2603215>

[5] Google Research Blog, “Speech Recognition and Deep Learning,” Aug. 2012. <http://googleresearch.blogspot.kr/2012/08/speech-recognition-and-deep-learning.html>

[6] Bing blogs, “Bing Makes Voice Recognition on Windows Phone More Accurate and Twice as Fast,” June 2013. [http://www.bing.com/blogs/site\\_blogs/b/search/archive/2013/06/17/dnn.aspx](http://www.bing.com/blogs/site_blogs/b/search/archive/2013/06/17/dnn.aspx)

[7] A. Waibel and C. Fügen, “Spoken Language Translation,” *IEEE Signal Process. Mag.*, May 2008, pp. 70–79.

[8] 김승희 외, “흔히 보이는 음성언어기술,” 전자신문사, 2009.

[9] W. Wahlster, “Verbmobil: Foundations of Speech-to-Speech Translation,” *Springer*, Sept. 2000.

[10] F. Casacuberta et al., “Recent Efforts in Spoken Language Translation,” *IEEE Signal Proc. Mag.*, May 2008, pp. 80–88.

[11] B. Zhou, “Statistical Machine Translation for Speech: A Perspective on Structures, Learning, and Decoding,” *IEEE*, vol. 101, no. 5, May 2013, pp. 1180–1202.

[12] H. J. Ney, “Speech Translation: Coupling of Recognition and Translation,” *ICASSP*, vol. 1, 1999, pp. 517–520.

[13] E. Matusov, S. Kanthak and H. Ney, “Integrating Speech Recognition and Machine Translation: Where Do We Stand?,” *ICASSP*, 2006.

[14] R. Zhang et al., “A Unified Approach in Speech-to-Speech Translation: Integrating Features of Speech Recognition and Machine Translation,” *Proc. Int. Conf. Comput. Linguist*, 2004.

[15] N. Bertoldi et al., “Efficient speech translation through confusion network decoding,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 8, Nov. 2008, pp. 1696–1705.

[16] L. Mathias and W. Byrne, “Statistical Phrase-Based Speech Translation,” *Proc. ICASSP*, 2006, pp. 561–564.

[17] B. Zhou, L. Besacier, and Y. Gao, “On Efficient Coupling of ASR and SMT for Speech Translation,” *Proc. ICASSP*, 2007, pp. IV-101–IV-104.

[18] J. Jiang et al., “Phonetic Representation-Based Speech Translation,” *Proc. 13th Machine Translation Summit*, 2011.

- [19] Z. Ahmed et al., "A System for Facial Expression-based Affective Speech Translation," *Proc. Inter. Conf. Intelligent User Interfaces*, 2013.
- [20] O. Hamon, D. Mostefa and K. Choukri, "End-to-End Evaluation of A Speech-to-Speech Translation System in TC-STAR," *Proc. MT-Summit*, 2007, pp. 223-230.
- [21] Business Insider, "Google Has Bought A Startup To Help It Recognize Voices And Objects," Mar. 2013. <http://www.businessinsider.com/googlebuysdmresearch-2013-3>
- [22] The Science Times, "구글, 딥마인드 인수...인공지능(AI) 기술 선점," 2014. 5. 29.
- [23] TechCrunch, "Google Acquires Artificial Intelligence Startup DeepMind For More Than \$500M," Jan. 2014. <http://techcrunch.com/2014/01/26/google-deepmind/>
- [24] The Official Microsoft Blog, "Microsoft demos breakthrough in real-time translated conversations," May 27th, 2014. [http://blogs.technet.com/b/microsoft\\_blog/archive/2014/05/27/microsoft-demos-breakthrough-in-real-time-translated-conversations.aspx](http://blogs.technet.com/b/microsoft_blog/archive/2014/05/27/microsoft-demos-breakthrough-in-real-time-translated-conversations.aspx)
- [25] TechCrunch, "Facebook Acquires Mobile Technologies, Developer of Speech Translation App Jibigo," Aug. 12th, 2013. <http://techcrunch.com/2013/08/12/facebook-acquiresmobile-technologies-speech-recognition-and-jibigo-app-developer>
- [26] The Information, "Google Beat Facebook for DeepMind, Creates Ethics Board," Jan. 27th, 2014. <https://www.theinformation.com/Google-beat-Facebook-For-DeepMind-Creates-Ethics-Board>
- [27] 디지털데일리, "국내 자동번역 시장의 유일한 생존자, 프랑스 업체 인수 '눈길'," 2014. 5. 27.