자동통역 기술 동향 및 응용

The Trends and Application of Automatic Speech Translation Technology

모바일 소프트웨어 기술 동향 특집

김승희 (S.H. Kim) 자동통역연구팀 선임연구원
조훈영 (H.Y. Cho) 자동통역연구팀 선임연구원
윤 승 (S. Yun) 자동통역연구팀 UST연구생
김창현 (C.H. Kim) 언어처리연구팀 선임연구원
김상훈 (S.H. Kim) 자동통역연구팀 책임연구원

목 차

Ⅰ. 서론
Ⅱ. 자동통역 기술 동향
Ⅲ. 자동통역 기술 응용
Ⅳ. 결론

근래에 국가간 인적, 물적 교류가 활발해지면서 언어 장벽으로 인한 문제를 해결하기 위한 자동통역 기술의 중요성이 부각되고 있다. 세계 각국에서는 1990년대부터 시작된 정부차원의 연구개발 단계를 거쳐 시범 서비스 및 실용화를 위한 연구개발에 박차를 가하고 있으며, 현재 이라크 내 미군에 의한 군사 목적, 미국 내 병원 진료, 여행시 통역 등의 목적에 자동통역 기술이 활용되고 있다. 본 고에서는 자동통역 기술 및 핵심 요소기술에 대해 설명하고, 최근 자동통역 기술의 개발 동향 및 응용 사례에 대해 기술한다.

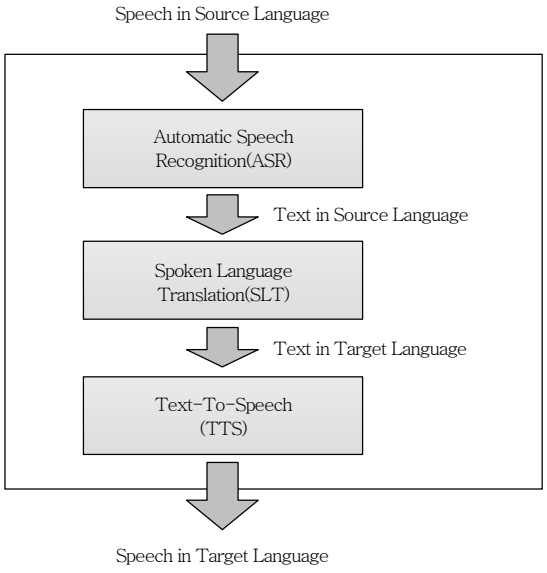
Ⅰ. 서론

자동통역(automatic speech translation) 기술은 서로 다른 언어를 사용하는 사람 간의 대화를 자동으로 통역하여 언어 장벽에 의한 의사소통 문제를 해결해 주는 기술이다.

한 언어의 말소리를 다른 언어의 말소리로 변환하기 위해 자동통역 기술은 다음과 같은 요소 기술들로 구성된다. 우선 말소리를 인식하여 해당 언어의 문자 언어로 변환해 주는 음성인식 기술이 있으며, 한 언어의 문자를 다른 언어의 문자로 변환해 주는 자동번역 기술이 있고, 해당 언어의 문자를 말소리로 변환해 주는 음성합성 기술이 있다. (그림 1)은 단방향 통역 시스템의 일반적인 구조를 나타내고 있다. (그림 1)과는 달리 음성인식과 자동번역을 한꺼번에 수행하는 구조에 관한 연구도 보고되고 있다.

자동통역 기술은, 최근 미군이 이라크에서 군사 및 민간 목적으로 활용하고 있으며, 미국 내 병원에서 의료진과 환자 간의 의사소통을 위해서도 사용되고 있다. 이 외에도 방송뉴스 통역, 강의 통역을 위한 기술들도 활발하게 연구되고 있다.

본 논문에서는 최근 중요성이 급증하고 있는 자동통역 기술 및 핵심 요소 기술의 개발 동향과 그 응용 사례에 대해 알아본다.



(그림 1) 단방향 자동통역 시스템의 구조

Ⅱ. 자동통역 기술 동향

1. 개요

자동통역 기술은 크게 음성인식(automatic speech recognition), 자동번역(spoken language translation) 및 음성합성(text-to-speech synthesis)의 세 가지 요소기술로 구성된다. 이들 각각의 요소분야는 그 자체적으로도 오랜 역사를 가진 독자적인 기술분야이지만, 자동통역 기술은 이 세 가지를 큰 틀에서 아우르는 음성언어정보 기술분야의 궁극적인 목표가 되는 기술이라고 할 수 있다.

자동통역 기술의 상용화를 고려할 경우에는 이 세 가지 요소기술 외에도 실제 사용환경에 존재하는 잡음에 대한 고려와, PDA와 같이 컴퓨팅 자원이 제한적인 휴대단말기 상에서도 위 세 가지 요소기술의 동작이 가능하게 하는 단말 최적화 기술이 추가로 필요하다. 본 장에서는 이들 각각의 분야에 대해 자동통역 기술의 관점에서 최신 연구동향을 기술하기로 한다.

2. 음성인식

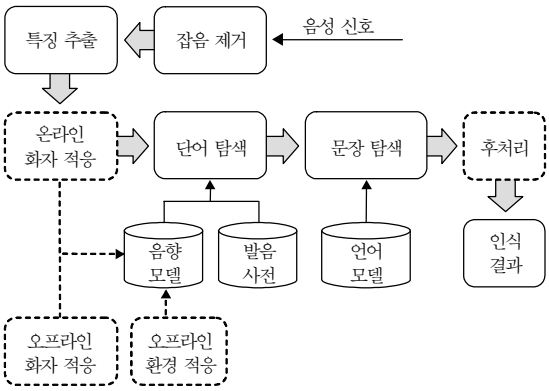
음성인식은 자동통역 시스템에서 가장 처음 단계의 시스템으로서 사용자가 발성한 음성을 텍스트 형태로 변환하는 역할을 수행한다. 음성인식 시스템은 단어단위의 음성을 인식하는 고립단어인식으로부터 연속적인 단어열을 인식하는 연결단어인식, 연속적으로 발성된 음성에서 시스템에 등록된 특정 단어들만을 검출하여 인식하는 핵심어 검출 및 문장형태의 발화를 인식하는 연속음성인식으로 구분할 수 있다. 연속음성인식은 방송뉴스의 진행자가 발성하는 방식처럼 분명하게 발성하는 낭독체 음성인식과 일반인들이 생활 속에서 자연스럽게 발성하는 형태의 대화체 음성인식으로 구분할 수 있다. 낭독체 발성에

대비되는 대화체 발화의 특성은 간투어, 반복, 수정, 긴 묵음구간, 발음변이, 발화 오류, 발성속도 변이 등을 예로 들 수 있으며 이들을 발성의 비유창성(disfluency)으로 요약할 수 있다.

1990년대 초반에 발표된 JANUS과 같은 초기의 자동통역 시스템들은 자동통역의 요소기술들의 기술적 수준이 매우 낮은 상태였음에도 불구하고 자동통역 기술의 가능성을 보여줄 수 있었다. 그러나, 이 시스템들은 사용자들 간의 자유로운 대화체 발화를 허용할 수 없었고, 제한된 단어들과 구문들에 한하여 미리 정해진 문장들만을 낭독체 형태로 발성해야 했다[1]. 1990년대를 지나 2000년대 중반에 이르기까지 음성인식기술은 수 천 어휘급의 고립단어 인식에서 수 만 내지 수십 만 어휘 규모의 대어휘 낭독체 연속음성인식을 지나 최근에는 사용자의 발성 방식에 제약이 없는 대화체 연속음성인식 기술이 활발히 연구되고 있다.

이에 따라 자동통역에서의 음성인식기술 연구도 기술적 난이도가 높은 대화체 연속음성인식을 목표로 진행되고 있는 추세이다. (그림 2)는 일반적인 연속음성인식 시스템의 구성도 및 음성인식 절차를 나타낸다.

자동통역기 사용자는 비교적 조용한 건물 내부에서 통역기를 사용할 수도 있지만, 주행중인 자동차 내부 또는 사람들이 많은 거리와 같이 다양한 크기와 종류의 잡음이 존재하는 상황에서 시스템을 사용하게 된다. 따라서, 음성인식시스템은 먼저 적절한



(그림 2) 대화체 연속음성인식 시스템의 구성도

잡음처리 과정을 통해 음성신호로부터 잡음을 제거한 후, 음성인식기 입력을 위한 특징파라미터를 추출하게 된다. 잡음처리기법으로는 단일 마이크를 사용하는 방법과 마이크배열에 기반한 방법이 있다. 특징추출기법으로 LPC, PLP를 비롯하여 수많은 방법이 연구되어 왔으나, 최근에는 대부분의 인식시스템이 MFCC를 사용하며, 보다 변별력이 높은 특징파라미터 추출을 위하여 LDA, HLDA, fMPE 등을 특징추출의 후처리 과정으로 도입하기도 한다[2],[3].

높은 음성인식률을 얻기 위해 대부분의 음성인식 시스템에서는 음소의 전후 음소정보를 동시에 모델링하는 문맥종속적 음향모델을 사용하며, 대부분 삼음소(triphone) 모델을 사용하나, 더 많은 문맥정보를 활용하기 위해 때로는 quinphone 등의 단위를 사용하기도 한다. 다양한 잡음 환경과 수많은 사용자의 발성변이에 강한 음향모델을 생성하기 위해 실제 환경에서 수집한 여러 가지 잡음신호를 음향모델의 학습 데이터에 추가하여 음향모델을 생성하는 MST 학습 방식이나, 음향모델들 상호 간에 식별력을 극대화하는 MMI, MCE, MWE, MPE 학습방법 등의 변별학습을 적용하는 추세이다.

IBM의 MASTOR 시스템은 MPE 학습과 더불어 추가적 성능향상을 위해 남성과 여성에 대한 별도의 HMM 음향모델을 사용하며, 이러한 음향모델들은 총 발성시간이 400시간이 넘는 양의 음성코퍼스를 이용하여 생성되었다[3]. BBN의 Byblos 시스템의 경우, DARPA의 TransTac 평가에서 110시간의 음성데이터로 문맥종속 음향모델을 학습하였으며, MPE 학습을 적용하여 22.5%의 오류감소율을 얻을 수 있었고, 영어에 대한 음성인식 성능으로 약 23.1%, 이라크어에 대해 31.9%의 단어오류율을 얻었다. 또한, 여기에 화자 적응을 적용하여 약 23%의 상대적인 성능이득을 얻을 수 있었다.

음성인식의 언어모델은 방대한 용량의 텍스트 코퍼스를 필요로 한다. 낭독체 음성인식의 경우에는 텍스트 코퍼스의 수집이 비교적 용이하지만, 자유발화 대화체에 대해 이처럼 방대한 데이터를 수집하기

는 매우 어렵다. 따라서 제한된 학습 데이터를 이용해서 단어열의 확률분포를 구하는 방법에 대한 지속적인 연구가 필요하다. IBM의 경우, 웹을 통해 자동적으로 학습 데이터를 추가 수집하여 언어모델링 성능을 향상시키고 있으며, 기존에 널리 쓰이는 단어 N-gram의 커버리지 문제를 해결하기 위해 단어의 클래스 N-gram을 적용하기도 한다.

이 외에도 영역에 특화된 언어모델을 다른 언어모델과 조합하는 interpolation 방법 및 제한된 데이터에 대해 분포 추정의 정확도를 높이는 방법에 대한 연구가 필요하다.

<표 1>은 2006년에 개최된 제 2차 TC-STAR 평가 워크숍에서 발표된 인식성능 비교평가 결과를 나타낸다. 이 평가에는 독일 IBM, 독일의 RWTH, 이탈리아의 ITC-irst, 프랑스의 LIMSI, 핀란드의 NOKIA, 독일 SONY, 독일 UKA 대학이 참여하였다. 비교평가를 위해서 정의된 태스크로는 세 가지가 있는데, <표 1>에서는 EPPS 태스크에 대한 성능을 기술하고 있다. EPPS 태스크는 EC에서 출간한 의회 논쟁 자료의 오디오 신호와 그에 해당하는 텍스트 자료를 포함하며, 영어와 스페인어로 되어 있다. 세 가지 인식 태스크에 대해서 인식기의 학습조건을 다시 세 종류로 구분하였다. 첫째, 한정된 조건(restricted condition)은 평가 참가자들이 TC-STAR 프로젝트에서 제공된 자료만을 인식기의 학습에 사용해야 하며, EPPS English 자료의 경우 총 166시간, EPPS Spanish의 경우 총 151시간 및 CORTES Spanish 코퍼스는 40시간의 자료로 구성되어 있다. 둘째로 공개 데이터 조건(public data condition)은 LDC 등을 통해 공개된 데이터들을 모두 사용이 가능하다. 셋째로 오픈 조건(open condition)에서는 특정 날짜 이전의 모든 데이터를 사용할 수 있다. 인식기 평가자료로는 EPPS 태스크의 경우, 2005년 9월부터 11월 의회 녹음자료를 사용하고, CORTES의 경우는 같은 해 11월 24일의 자료를 사용하였다. <표 1>에서 각 열은 인식기의 세 가지 학습조건을 의미한다. 평가결과 TC-STAR 시스템이 공개 데이터 조건에서 6.9% 단어 오인식률로 최고의 성능을 나타내었으며, 이 시스템은 ROVER 방법에 의해 여러 인식기의 출력 결과를 조합한 것이다. 스페인어의 경우, 오류율은 10.2%에서 28.4%까지 다양하였으며, 마찬가지로 ROVER로 통합한 결과는 8.1%의 단어 오인식률을 나타내었다[4].

〈표 1〉 English EPPS 연속음성인식 태스크에 대한 기관별 단어오류율

(단위: %)

Site	Open	Public	Restricted
IBM		8.8	
ITC-irst		11.0	
LIMSI		8.2	
NOKIA	18.3		
RWTH			10.2
SONY	37.1		
UKA		14.0	
TC-STAR		6.9	

<자료>: TC-STAR 2006 평가보고서

ETRI의 자동통역연구팀에서 2008년부터 4년간 수행하는 휴대형 한영 자동통역기 개발과제에서는 여행영역에 대해서 발성이 자유롭고 변화가 많은 대화체 음성 DB를 구축하고, 대화체 특성을 반영한 음향 모델링 기술, 대화체 언어 모델링 기술 등에 대한 연구를 진행중에 있으며, 문장 단위의 발성을 실시간으로 처리할 수 있는 고속 탐색 기술, 대화의 흐름에 따른 문맥 지식을 활용하는 문맥지식기반 음성인식 기술, 대화의 대상물과 대화가 이루어지는 상황정보를 활용하는 상황지식기반 음성인식 기술에 대한 연구개발을 점차적으로 수행할 예정이다.

3. 자동번역

일반적으로 자동번역(machine translation) 기술이란 텍스트 원문을 자연어처리 기법을 이용하여 다른 언어의 문장으로 변환하는 기술을 말한다.

2차 대전 후 미국과 옛 소련에 의해 군사적인 목적으로 개발된 자동번역 기술은 1980년대 중반부터 유럽과 일본에 의해 다시 본격적인 연구가 시작되었다. 유럽은 다국어 문화권이지만 유럽연합(EU)

이라는 공동체의 특성상 언어장벽을 해소해야 할 필요성이 증가했으며, 일본은 Toshiba, Fujitsu 등의 기업 주도로 자동번역에 대한 연구 개발이 활발히 추진되었다.

자동번역은 그 자체로 하나의 완전한 기술일 뿐만 아니라 자동통역 기술의 하부기술로써의 요구 등도 존재한다. 본 장에서는 이러한 두 가지 관점에서 자동번역을 살펴보기로 한다.

자동번역 기술은 번역 방법론에 따라 크게 규칙기반 방법과 말뭉치 기반 방법으로 나눌 수 있다. 규칙기반 방법에서는 언어학자, 번역가들이 자동번역에 필요한 지식을 자신의 언어능력을 사용하여 구축하고, 이렇게 구축한 언어학적 규칙(예: 구조분석 규칙, 변환 규칙 등)을 이용해 자동번역이 이루어지는 반면, 말뭉치기반 방법에서는 인간의 주관적 언어능력 대신 말뭉치로부터 객관적 번역지식을 학습함으로써 자동번역이 이루어진다. 말뭉치기반 방법은 크게 예제기반 방법과 통계기반 방법으로 나누어 볼 수 있다. 1980년대까지는 규칙기반 방법이, 1990년대에는 말뭉치기반 방법이, 그리고 2000년대 들어서는 규칙기반과 말뭉치기반 방법이 독립 또는 공존하는 시기로 구분할 수 있다[5].

규칙기반 방법은 그 분석의 깊이에 따라 직접번역방식, 간접변환방식, 중간언어방식 등으로 세분할 수 있다. 직접번역방식에서는 입력문을 형태소 분석, 태깅(tagging) 등의 과정을 통해 매우 낮은 단계에서 분석을 마친 후, 변환 사전(bilingual transfer dictionary) 등을 참조해 대역문장을 생성해 낸다. 이 기법은 초창기 자동번역 시스템에서 많이 사용되었으며, 최근에도 한국어와 일본어, 스페인어와 이탈리아어 등과 같이 언어학적으로 유사한 언어 쌍에 대해 많이 사용되고 있다. 간접변환방식에서는 형태소 분석을 거쳐 통사구조(syntactic structure), 의미구조(semantic structure)에 대한 분석을 더 거친 후 목표언어로의 변환을 하며, 이 변환된 구조로부터 대역 문장을 생성하게 된다. 이 방식은 비교적 개발이 용이하고, 소수의 규칙만을 구축하더라도 비교적 높은 성능을 낼 수 있으므로, 현재 국내외에서 상용화되어 판매되고 있는 대부분의 자동번역 시스템에 채택되고 있다. 중간언어방식에서는 개별 언어 독립적인 의미표상(language-independent semantic representation)을 도입하고, 입력문을 분석 단계를 거쳐 이 언어 독립적인 의미표상으로 매핑한다. 따라서 다수 개의 변환모듈이 필요한 간접변환방식과는 달리, 중간언어방식은 단지 개별언어로부터 중간언어로의 매핑을 위한 분석모듈, 중간언어로부터 목표언어를 생성하기 위한 생성모듈만이 필요하다. 따라서 이 방식은 다국어 자동번역에 적합하다고 할 수 있다.

예제기반 방법은 유추에 의한 번역(translation by analogy)이라고도 불리며, 수많은 번역 쌍들을 데이터베이스에 저장한 후, 입력문이 들어왔을 때 입력문과 가장 유사한 예문을 찾아, 예문의 번역을 참조하여 번역을 하는 것이다. 이 방법의 장점은 대용량의 대역 코퍼스와 잘 정의된 시소러스가 있으면 어느 언어 쌍에도 비교적 쉽게 적용할 수 있다는 점이다. 그러나 이 방법의 단점은 높은 성능을 내기 위해서는 대용량의 대역코퍼스가 필요한데, 많은 언어 쌍의 경우 이것이 쉽지 않다는 점이다. 또 하나의 문제점은 대역 코퍼스의 도메인에 따라 번역률 차이가 많다는 점이다.

통계기반 자동번역(SMT) 기술은 통계적 분석을 통해 이중언어 말뭉치로부터 모델 파라미터를 학습하여 문장을 번역하는 기술이다. SMT 기술은 1949년 Warren Weaver[6]에 의해 소개된 이후, 1991년 IBM의 Thomas J. Watson 연구소 연구원들에 의해 다시 소개되면서부터 연구가 부활하여[7], 현재 가장 활발하게 연구되는 기계번역 기술이다.

SMT 기술이 활발히 연구되는 배경에는 다음과 같은 요인이 작용하고 있다. 1) 모델 파라미터를 학습할 수 있는 대용량의 가용 말뭉치가 구축되고 있다. 2) 특정 언어 쌍에 제한 받지 않고 모델을 자동으로 학습해 낼 수 있다. 3) 규칙기반/패턴기반 기계번역이 번역지식을 구축하는 데 상당한 비용을 요구하고, 다른 언어들에 일반화시켜 적용하기 어렵다는 문제가 있다. SMT의 기본 요소는 통계적 번역모델

과 언어모델, 이중언어 말뭉치로부터 은닉된 번역지식 파라미터를 찾아내는 학습 알고리듬, 그리고 학습된 번역모델에 기반하여 최적의 번역결과를 탐색하는 디코딩 알고리듬으로 구성된다. SMT 기본 모델인 단어단위 모델(IBM model 1-5)은 문장 길이에 따라 어순 재배열 계산 복잡도가 너무 높기 때문에 계산 복잡도를 낮추기 위한 많은 연구들이 시도되었고, 특히 2003년 어순 재배열에 따른 계산복잡도 감소 및 번역 효율을 고려한 구문단위 번역 모델[8]이 소개되면서 기술이 급격히 발전하여 현재 state-of-the-art를 이루었다. 그리고 최근에는 통사적 언어 구조를 모델에 접합시키기 위한 연구([9],[10]), 미등록어 및 관용적 숙어 표현의 효과적 처리를 위한 paraphrasing 기법 또는 번역 지식 일반화 등에 대한 연구가 진행중이다[11],[12].

자동통역의 하부기술로써의 자동번역(spoken language translation)은 그 핵심 기술에서는 앞서 설명한 자동번역(machine translation)과 크게 다르지 않다. 그러나, 이러한 환경에서 가장 논점이 되고 있는 것은 크게 두 가지이다. 첫째는 음성 인식의 결과를 자동번역 입력으로써 사용하는 문제이고, 둘째는 자동통역의 대상이 되는 번역 도메인 문제이다. 일반적인 자동번역은 올바른 문장을 입력 단위로 가정하고 있다. 그러나, 음성 인식의 결과는 이를 보장할 수 없으며, 이를 해결하기 위한 여러 연구들이 시도되고 있다. 대표적인 방법론들로는 음성인식 결과를 하나의 문장이 아닌 다수 후보 형태로 구성하여 자동번역의 입력으로 사용하는 것이다. 이러한 형태들로는 가장 단순한 형태인 N-best 문장 형태, 이론적으로는 가장 적절한 단어 격자 형태, 자동번역의 탐색 측면에서 가장 적절한 혼합 네트워크(confusion network) 형태 등이 있다. 그러나, 이러한 복잡한 입력 형태에 대해 현재까지 조사된 연구 결과는 만족스럽지 않다. 음성 인식 오류의 자동 보정에 대한 연구와는 별도로 음성 인식 결과를 발화자 자신이 직접 확인하고 오류를 수정하는 형태의 멀티모달 방법은 또 다른 대안이 될 수 있으며, 실제로 상용화를 고려하는 자동통역 제품들 가운데에서는 이러한 시도가 고려되고 있다. 자동통역은 실시간으로 사람들의 발화를 번역하는 것을 요구한다. 특히 사람들의 발화는 생략, 축약, 구어체적 표현, 비문법적인 문장, 반복, 간투사 등의 번역에 적합하지 않은 특성들을 포함하고 있다. 이러한 대화체의 특성은 기존의 자동 번역에서 주로 다루었던 문어체 위주의 번역 도메인과는 확연히 구별된다. 한국어와 같은 교착어의 경우 생략, 축약 등은 형태소 분석 단계에서의 많은 오류를 야기시키기 때문이다. 자동통역의 문제점 분석 결과들에 따르면 의미 오류, 개념의 오류, 개념의 누락 등의 자동번역 모듈 오류가 상당 부분을 차지한다고 보고되고 있다. 이는 대화 도메인 및 대화 문맥을 사용해야만 해결될 수 있는 문제들이다. 이러한 대화체 특성에 따른 문제점 및 문맥 적용에 대한 연구들은 현재 시작 단계에 있으며, 이들의 결과에 따른 자동통역 상용화 시기가 결정될 것으로 보인다.

4. 음성합성

자동통역의 마지막 단계인 음성합성은 목적 언어(target language)로 번역된 텍스트를 사람들이 들을 수 있는 말의 형태로 제공하는 기술이다. 합성음을 생성하는 가장 간단한 방법으로 사용자에게 들려줄 안내 멘트를 미리 녹음하여 가지고 있다가 필요한 시점에서 이를 재생하여 들려주는 방법이 있다. 이 방법은 매번 정해진 내용만 반복해서 들려주므로 한계가 있기 때문에 문장의 기본 틀은 고정시켜 둔 상태에서 문장의 일부분만을 변경해서 합성음을 생성하는 편집합성 방식이 있다. 그러나 이 방법은 기본적으로 무제한 음성합성이 어렵다는 한계로 인하여 주로 단문 형식의 안내 멘트 합성용으로 사용된다. 음편조합 방식을 이용한 음성합성 기술은 일반적으로 단어보다 훨씬 작은 소리 단위를 조합하여 합성음을 생성하며 편집합성처럼 녹음된 음편을 그대로 결합하여 합성음을 생성하므로 음성신호 생성을 위한 음성 신호 조작을 거의 이용하지 않아 신호처리에 의한 왜곡이 없고 빠른 속도의 음성합성을 수행할 수 있다. 그러나, 이 방식은 조음효과로 인하여 음질열화가 발생하며, 음성

신호에 대한 조작이 어려워 합성음의 운율조절이나 음색조절이 어렵다는 단점과 통상적으로 수백 MB에서 수 GB급의 저장공간을 필요로 하여 상대적으로 저장공간의 크기가 제한된 분야에서는 고품질 합성음을 생성하기가 어려운 점이 있다.

음성합성 기술은 시스템의 크기에 따라 크게 서버용 음성합성 기술과 단말기 내장형 음성합성으로 구분할 수 있다. 기존 서버형 음성합성 기술을 응용한 시장의 성장은 비교적 더딘 반면에, 최근 임베디드 단말기에 내장되어 사용되는 내장형 음성합성 수요는 폭발적으로 증가하고 있다. 이에 따라 소용량 및 다국어 음성합성기술과 관련된 연구가 활발히 진행되고 있다.

본 연구팀에서는 임베디드 단말기에서 사용할 수 있는 내장형 소용량 음성합성 기능을 가지며, 대화체에 특화되어 있고, 음색변환에 용이한 HMM 기반 내장형 소용량 대화체 음성합성 기술을 개발하고 있다. 이 기술은 음성신호의 스펙트럼 정보, 피치 정보, 지속시간 정보를 각각의 독립된 Gaussian 확률분포를 가지는 HMM 모델로 훈련하여 합성용 보이스폰트를 생성한다. 합성 시에는 훈련된 HMM 모델 파라미터로부터 합성음 생성을 위한 음성 특징 파라미터를 생성하고, 이를 적당한 방법으로 보간하여 합성음 궤적을 생성한다. 기본적인 프로토타입 버전은 PC 환경에서 초벌을 개발하고, 개발된 HMM 기반 합성엔진을 저사양 프로세스와 스피커를 가진 임베디드 환경에서 문제점 및 성능개선 부분 등을 고찰하기 위하여 60MHz clocks/second 처리 속도의 ARM 720T 프로세서, 16MB NAND 메모리, 32MB SDRAM으로 구성된 ITS 단말용 OBE 보드에 정수형 버전 및 일부 모듈은 연산의 고속화를 위해 ARM 어셈블리어로 구현하였다(<표 2> 참조)[13].

최근 뉘앙스 사(Nuance Inc.)의 경우에는 모바일 플랫폼에서 14개 이상의 언어를 지원 가능하고, 엔진의 크기가 음질에 따라 2MB 또는 5MB로 가변적인 소용량 음성합성 제품을 출시하고 있다[14].

자동통역기를 위해서는 음성인식과 동일하게 대화체 음성합성 기술이 요구되고 있다. 대화체 음성합성이란 뉴스읽기와 같은 단조로운 낭독체 음성합성이 아닌 전달하고자 하는 메시지의 내용에 따라, 사용자의 의도에 따라 합성음의 분위기가 다르게 표현되는 기술을 의미한다. ETRI의 자동통역연구팀에서도 휴대형 한영 자동통역기 개발과제에서 통계적 음성 모델링 기법의 일종인 HMM 기반의 내장형 소용량 음성합성 기술을 개발함으로써 기존에 음성인식 분야에서 널리 연구되어 온 다양한 화자적응기법들을 자연스런 대화체 합성음 생성에 적용 가능하도록 할 예정이다.

〈표 2〉 내장형 음성합성 시스템의 메모리 사용량

(단위: MB)

	NAND	SDRAM
언어처리사전	1.7	2.1
보이스폰트	0.93	1.3
합성엔진	0.47	2.2
총합	3.19	5.6

Ⅲ. 자동통역 기술 응용

자동통역 기술의 응용사례는 인식 어휘 규모, 발화 스타일, 대상 영역, 출시 연대, 플랫폼 등에 따라 다양하게 분류할 수 있으나 여기에서는 크게 플랫폼을 기준으로 두 가지로 나누어 살펴보기로 한다. 이는 목표로 하는 자동통역 대상에 따라 자동통역 시스템이 탑재되는 플랫폼이 결정되며 결정된 플랫폼에 따라 채용될 수 있는 기술들이 나누어지기 때문이다. 따라서 이 곳에서는 PC 기반 플랫폼과 핸드헬드 기반 플랫폼의 두 항목으로 나누어 대표 응용사례와 기술적인 특성을 소개하도록 하겠다.

1. PC 기반 플랫폼

자동통역 기술 개발 초기에는 주로 워크스테이션급의 플랫폼에서 자동통역이 이루어졌다. 1990년대 중반까지가 이 시기에 해당하며 대표적인 응용사례로는 C-STAR I을 꼽을 수 있다. 그러나 이 시기의 자동통역 시스템은 인식 어휘나 발화 스타일, 자동

통역 대상 영역에 제약이 많이 존재하였다는 한계를 가진다.

이후 1990년대 중반부터 현재에 이르러서는 PC 기반의 플랫폼에서 자동통역이 주로 이루어지고 있다. PC 기반 플랫폼은 채용된 기술의 특성에 따라 몇 가지로 나누어 살펴볼 수 있다. 먼저 무제한 연속 음성 인식 기술을 채택하고 있으나 대상 영역에는 제한이 있는 경우이다. 이들의 대표적인 응용 사례로는 C-STAR[15], Verbmobil[16], Nespole[17], TransTac[18], A-STAR 등을 꼽을 수 있다. C-STAR의 경우 ETRI, ATR, CMU, CLIPS 등이 참여한 국제 공동 컨소시엄을 통해 연구되었으며 여행 계획 영역을 자동통역 대상으로 삼았다. 중간언어방식을 채택하였으며 네트워크를 통해 상대방 시스템과 통신한다는 점이 특징이다. Verbmobil의 경우 독일의 연구소, 대학, 기업들이 공동 참여해 개발하였으며 일정 수립, 여행 계획, 호텔 예약 등의 상황에서 독일-영국, 독일-일본 자동통역이 가능하다. 기본적으로 면대면 상황을 가정하였으며 휴대폰을 통해 자동통역 서버에 접속하여 자동통역이 가능한 것이 특징이다. Nespole는 비디오 콜센터에 자동통역을 채용하는 것에 관해 연구하였다. 은행, 소비자 서비스, 여행, 전자상거래 등의 상황에서 소비자가 비디오 콜센터에 접속한 후 상담원과 소비자 간에 자동통역을 통한 상담이 이루어지도록 하였다. Nespole의 경우 멀티모달 입력을 보조 수단으로 활용할 수 있도록 한 것이 특징이다. TransTac의 경우 DARPA의 지원 아래 이루어진 프로그램으로써 1단계로 군사 목적 등의 전술적인 상황에서 쓰이는 영어-아랍어간 자동통역 시스템을 개발한 데 이어 현재는 새로운 언어와 대상 영역에 빠르게 적응할 수 있는 기술을 개발중이다. 그리고 최근에는 자동통역 기술이 선진국뿐만 아니라 비교적 주변 국가로까지 확대되어 연구되는 경향을 나타내기도 한다. 관련하여 아시아권에서는 2009년 7월 아시아 지역의 자동통역 연구를 위하여 조직된 컨소시엄인 A-STAR 주관으로 한국어, 중국어, 일본어, 타이어, 인도네시아어, 말레이시아어, 베트남어, 힌디어 등을 대상으로 네트워크 기반의 여행 영역 아시아어권 자동통역 국제 시연을 실시하기도 하였다[19]. ETRI도 본 시연에 참여하였으며 시연에 참여한 시점에서 ETRI에서 개발중인 자동통역 시스템의 경우 2만 단어급 83.8%의 한국어 인식 성능과 83.4%의 영어 인식 성능, 그리고 한국어로부터의 영어의 경우 81%, 영어로부터 한국어의 경우 83%의 자동통역 성공률을 나타내었다[20].

지금까지 소개한 자동통역 응용 사례들은 주로 사용자들 간의 대화를 자동통역하는 사례에 해당된다. 그러나 최근에는 이러한 연구 동향과 별도로 단방향의 자동통역 기술이 연구되고 있다. 이들은 주로 뉴스, 연설, 강의 등을 대상으로 한 것으로 GALE[21], TC-STAR[22], Lecture Translator[23] 등이 대표적인 사례이다. 이들은 자동통역 대상 영역이 제한되어 있지 않다는 점을 가장 큰 특징으로 가진다. GALE의 경우, DARPA의 지원 아래 IBM, SRI, BBN 등이 참여해 연구중인 것으로 뉴스, 토크쇼 및 문어 자료 등을 대상으로 대량의 자료를 단시간 내에 이해 가능한 형태로 번역하는 것을 목표로 하고 있다. TC-STAR의 경우에는 공식 언어가 20여 가지 이상인 EU의 언어 장벽을 해소하기 위하여 연구된 것으로 IBM, ITC-irst, LIMSI, UKA, UPC, RWTH, NOKIA, SONY 등이 참여해 의회 연설문 등의 자동전사 및 통역/번역 서비스 제공 가능성을 시험하였다. Lecture Translator는 세미나 및 강의를 통역하는 것을 목표로 개발된 시스템이다. 이전에 언급한 시스템에 비해 발화 스타일이나 통역 대상이 좀 더 다양하기 때문에 난이도가 높아 아직은 실험적인 수준에 그치고 있다. Lecture Translator의 경우 영어로부터 스페인어, 독일어, 아랍어 번역이 가능하며 부가적으로 통역 결과를 지향성 스피커를 이용해 재생하거나 안경 또는 별도의 스크린에 표시할 수 있다는 특징을 가지고 있다.

2. 핸드헬드 기반 플랫폼

최근 가장 활발하게 자동통역 기술이 응용되고 있는 사례가 핸드헬드 기반의 자동통역 시스템이다.

PC 기반 자동통역 시스템의 경우 휴대가 불가능해 그 활용 용도가 제한적인데 반해 핸드헬드 기반 시스템의 경우 다양한 상황에서 응용이 가능하다는 특징을 지닌다. 다만 컴퓨팅 파워가 높지 않은 까닭에 일반적으로는 발화스타일이나 자동통역 대상 영역에 제약을 가진 경우가 많다. 응용 사례 중에 대표적인 예로 Phraselator를 들 수 있다. Phraselator는 DARPA의 지원 아래 Voxtec에서 개발하였다. 단방향 통역만이 가능하며 음성 인식 기능 또한 미리 기억된 문장만을 음성으로 선택할 수 있는 수준이어서 비교적 제약이 많은 편이다. 그럼에도 불구하고 다양한 언어를 지원하며 의사 전달을 위주로 하는 제한된 용도에서는 활용 범위가 넓어 이라크와 아프카니스탄에 파병된 미군에 보급된 실적이 있고 뉴욕주의 경찰 및 병원 응급실에서도 사용되고 있다. 또한 2005년 미군의 동남아 해일 구조에서도 사용된 바 있다((그림 3) 참조).

<자료>: http://www.voxtec.com

(그림 3) Phraselator

IBM에서도 MASTOR[24]라는 휴대형 통역기를 개발하였다. MASTOR는 DARPA의 지원으로 IBM Watson 연구소에서 개발하였으며 약 3만 단어급의 영어-중국어 양방향 통역 기능을 지원한다. 노트북, 핸드헬드 양쪽 모두에서 실행이 가능하며 여행, 긴급 의료 진단, 군의 자기 방어, 보안 상황 등을 통역 대상으로 삼고 있다. MASTOR를 이용하여 이라크에 파병된 미군에서 시험 서비스를 진행한 바 있다.

Ectaco에서 개발한 Speech Guard는 대중적으로 판매가 많이 이루어진 제품이다. Speech Guard는 군용, 의료용, 경찰용 등 다양한 응용 영역에서 활용할 수 있도록 개발되었으며 30여 가지 이상의 다양한 언어를 지원하는 제품군이 있다. 통역 목적에 따라 단방향 및 양방향 자동통역을 수행할 수 있으나 음성 인식의 경우 Phraselator와 같이 음성 인식 문장 검색 기능이 탑재된 수준이다. Ectaco에서는 여행용 자동통역 시스템으로 iTRAVL도 판매하고 있다((그림 4) 참조).

<자료>: http://www.ectaco.com

(그림 4) Speech Guard

전용 단말기 외에 범용 PDA 기반의 자동통역 시스템들도 상당수 존재한다. 대표적인 응용 사례로 CMU에서 개발한 Speechlator[25]와 PanDoRa[26]를 들 수 있다. Speechlator의 경우 HP iPAQ PDA에서 동작하는 제품으로 양방향 통역이 가능하며 영어와 아랍어로 의료 정보를 통역할 수 있다. PanDoRa 역시 CMU 주도로 개발된 휴대형 자동통역기로 영어-아랍어, 영어-중국어, 영어-일본어를 대상으로 자동통역을 지원하며 여행, 의료, 자기 방어 등을 통역 대상으로 삼는다. SMT 기반으로 번역을 수행한다는 것이 특징이다((그림 5) 참조).

NEC에서도 PDA에서 동작하는 여행자 영역의 1만 단어급 영어-일본어 자동통역 시스템을 개발하

<자료>: http://www.mobytrans.com

(그림 5) PanDoRa

였다. NEC에서 개발한 자동통역기는 일본 나리타 공항에서 실시한 e-Airport 시범 사업을 통해 대중에 선보인 바 있다((그림 6) 참조).

(그림 6) 일본 나리타 공항 e-Airport 시범 서비스

이러한 움직임은 최근 확대되어 일본 NICT의 경우 일본 최대 여행업체인 JTBGMT와 함께 2010년 1월에서 2월에 걸쳐 일본 각지의 숙박시설과 관광시설을 대상으로 자동통역 시범 서비스를 실시하고 있다. 이는 한국어, 일본어, 중국어, 영어를 대상으로 하고 있으며 비교적 대규모로 이루어지고 있어 자동 통역 실용화가 머지 않았음을 알리고 있다.

그리고 최근에는 스마트폰의 대중화와 함께 스마트폰에 자동통역기술을 탑재하려는 움직임이 확산되고 있다. PDA 시장이 하향세인 것과 비교해 스마트폰의 보급이 급속도로 이루어지고 있는 것을 고려한다면 스마트폰 기반의 자동통역 기술이 일반에 대중화될 가능성은 매우 높다고 할 수 있다. 구글의 경우에도 구글이 개발한 안드로이드 OS 기반의 스마트폰에 탑재될 수 있는 자동통역 애플리케이션을 개발하기 위한 연구를 진행하고 있으며 실제 아이폰의 경우 이미 영어-스페인어를 대상으로 여행과 의료 영역에서 4만 단어급의 통역 성능을 보이는 Jibbigo와 같은 애플리케이션이 출시되어 사용자로부터 좋은 반응을 얻고 있다((그림 7) 참조).

<자료>: http://www.jibbigo.com

(그림 7) Jibbigo

Ⅳ. 결론

자동통역 기술은 날로 가속화되고 있는 세계화의 시대에서 언어장벽 문제를 해결할 수 있는 중요한 기술로 부각되고 있다. IBM은 개발 완료시 효과가 큰 실용화 대상 기술로 자동통역을 1위로 선정했으며(2007년 2월 4일), DARPA 50년 역사의 5대 발

명품 중의 하나로 자동통역 기술이 선정되기도 했다(2008년 5월 15일). 이미 미국, EU, 일본 등 선진 각국에서는 정부 차원의 대규모 지원으로 10여 년간의 연구 개발 단계를 거쳐 제한된 영역에 대한 시범 서비스를 시도하고 있다. 미국은 주로 군사적인 목적으로, EU는 11개의 공식 언어를 지원하기 위해, 일본은 관광 등 민간 분야를 대상으로 실용화 연구개발을 추진하고 있다. 자동통역 기술은 현재 발아기로서 기술의 확보 여부는 국가 경쟁력과 직결된다고 할 수 있다.

본 고에서는 자동통역 기술 및 핵심 요소 기술의 개발 동향을 살펴보고, 응용 사례에 대해 살펴보았다. 자동통역 기술은 대화체 음성인식, 자동번역, 음성 합성 등 요소 기술이 어우러진 복합 기술이며, 아직 미개척 분야로써 적기에 연구 개발을 추진하면 이에 대한 기술 경쟁력을 확보할 수 있다.

● 용 어 해 설 ●

HMM: 관측된 음성신호의 통계적 특성 및 해당 음성신호의 숨겨진 통계적 상태를 모델링하는 2차 통계모델

언어모델: 선행 단어에 대한 후속 단어의 관계(언어적 가능성)를 정의한 것을 의미

약어 정리

A-STAR	Asian Speech Translation Advanced Research
C-STAR	Consortium for Speech Translation Advanced Research
DARPA	Defense Advanced Research Agency
EPPS	European Parliament Plenary Sessions
GALE	Global Autonomous Language Exploitation
fMPE	feature Minimum Phone Error
HLDA	Heteroscedastic Linear Discriminative Analysis
HMM	Hidden Markov Model
LDA	Linear Discriminative Analysis
LPC	Linear Prediction Coefficient
MCE	Minimum Classification Error
MFCC	Mel-Frequency Cepstral Coefficient
MMI	Maximum Mutual Information
MPE	Minimum Phone Error
MST	Multi-Style Training
MWE	Minimum Word Error
PLP	Perceptual Linear Prediction
ROVER	Recognizer Output Voting Error Reduction
SMT	Statistical Machine Translation
TC-STAR	Technology and Corpora for Speech to Speech Translation

참 고 문 헌

[1] Alex Waibel, “Speech Translation: Past, and Future,” *In Proc. of INTERSPEECH*, 2004, pp.353-356.

[2] David Stallard et al., “Recent Improvements and Performance Analysis of ASR and MT in a Speech-to-Speech Translation System,” *In Proc. of ICASSP*, 2008, pp.4973-4976.

[3] Xiaodong Cui et al., “Developing High Performance ASR in the IBM Multilingual Speech-to-Speech Translation System,” *In Proc. of ICASSP*, 2008, pp.5121-5124.

[4] D. Mostefa, M.-N. Garcia, O. Hamon, and N. Moreau, “Evaluation Report,” TC-STAR, http://www.tcstar.org, 2006.

[5] 최승권, 홍문표, 박상규, “다국어 자동번역 기술,” 전자통신동향분석, 제20권 제5호, 2005. 10., pp. 16-27.

[6] Machine Translation of Languages, MIT Press, Cambridge, MA.

[7] P. Brown, S. Della Pietra, V. Della Pietra, and R. Mercer, “The Mathematics of Statistical Machine Translation: Parameter Estimation,” *Computational Linguistics*, Vol.19, No.2, 1991, pp. 263-311.

[8] P. Koehn, F.J. Och, and D. Marcu, “Statistical Phrase Based Translation,” *In Proc. of the HLT/NAACL*, 2003.

[9] Y.S. Hwang, A. Finch, and Y. Sasaki, “Improving Statistical Machine Translation Using Shallow Linguistic Knowledge,” *Computer Speech and Language*, Vol.21, No.2, 2007.

[10] D. Chiang, "A Hierarchical Phrase-based Model for Statistical Machine Translation," *In Proc. of ACL'05*, 2005.
[11] C. Bannard and C.B. Callison, "Paraphrasing with Bilingual Parallel Corpora," *In Proc. of ACL'05*, 2005.
[12] Y.S. Hwang, Y.K. Kim, and S.K. Park, "Paraphrasing Depending on Bilingual Context toward Generalization of Translation Knowledge," *In Proc. of the Third Int'l Joint Conf. on Natural Language Proc.*, 2008.
[13] 김종진, 김정세, 김상훈, 박준, "내장형 음성합성 기술 동향 및 사례," 전자통신동향분석, 제23권 제1호, 2008. 2., pp.77-88.
[14] http://www.nuance.com/realspeak/mobile
[15] C-STAR Project, http://www.c-star.org
[16] Verbmobil Project, http://verbmobil.dfki.de
[17] NESPOLE! Project, http://nespole.itc.it
[18] http://www.darpa.mil/IPTO/programs/transtac/transtac.asp
[19] Sakriani Sakti et al., "The Asian Network-based Speech-to-Speech Translation System," *In Proc. of ASRU*, 2009, pp.507-512.
[20] Ilbin Lee et al., "An Overview of Korean-English Speech-to-Speech Translation System" In Proc. of TCAST Workshop, Singapore, 2009, pp.6-9.
[21] http://www.darpa.mil/IPTO/programs/gale/gale_concept.asp
[22] TC-STAR Project, http://www.tcstar.org
[23] Christian F¨ugen et al., "Open Domain Speech Translation: From Seminars and Speeches to Lectures," In Proc. of TC-STAR Workshop Speech-to-Speech Translation, Barcelona, Spain, Sep. 2006, pp.81-86.
[24] MASTOR, http://domino.watson.ibm.com/comm/research.nsf/pages/r.uit.innovation.html
[25] Alex Waibel et al., "Speechalator: Two-way Speech-to-Speech Translation on a Consumer PDA," *In Proc. of EUROSPEECH 2003*, Geneva, Switzerland, Sep. 2003, pp.369-372.
[26] Ying Zhang and Stephan Vogel, "PanDoRA: a Large-scale Two-way Statistical Machine Translation System for Hand-held Devices," In Proc. of MT SUMMIT XI, Copenhagen, Denmark, Sep. 2007, pp.543-550.