

《主 題》

# 음성 자동통역 기술의 현황

구 명 완

(한국통신 연구개발원 소프트웨어연구소)

□ 차 례 □

I. 서 론

II. 음성 자동통역 시스템의 개발 현황

III. 결 론

## I. 서 론

음성 자동통역 기술이란 어떤 언어를 사용하고 있는 사람과 다른 언어를 사용하고 있는 사람이 각자의 언어를 사용하더라도 언어의 불편을 느끼지 않고 의사소통이 가능하도록 해주는 꿈의 기술이다. 이 기술은 음성을 인식하는 음성인식 기술, 인식된 음성을 다른 언어로 변경시켜 주는 기계번역 기술, 그리고 변경된 언어를 음성으로 출력시켜 주는 음성합성 기술로 나눌 수 있다. 이러한 음성인식, 기계번역 및 음성합성 기술은 음성 자동통역 기술이 대두되기 이전부터 독자적으로 각 분야의 전문가에 의해 연구가 진행되어 왔다[1]. 이로 인해 음성 자동통역 기술이 별도로 존재할 수가 있는지 그 실체에 대해 회의를 품는 경우가 있으나 음성 자동통역 기술을 위한 음성인식, 기계번역 및 음성합성 기술은 기존의 기술과는 다음과 같은 차이점이 있다.

● 음성인식 분야

- 기존의 음성인식 기술은 인간 대 기계 사이의 대화(human-computer interaction)에 중점을 두었으나 음성 자동통역 기술은 인간 대 인간 사이의 대화(human-human interaction)에 발생하는 대화 음성을 인식해야 하는 특징이 있다.

- 특히 음성 자동통역시스템은 인간 대 기계 대 인간 사이의 대화(human-machine-human interaction) 중

기계의 역할을 수행하는 것으로 보는 견해도 있으므로 이에 관한 연구가 진행되고 있다[2].

● 기계번역 분야

- 기존의 기계번역 시스템의 입력은 문법에 맞는 문어체 문장이었으나 음성 자동통역 기술에서는 음성인식 시스템의 출력이다. 음성인식 시스템의 출력은 문법에 맞지 않는 경우가 발생할 수 있으며 문장의 끝을 표시해 주기 어렵다.

● 음성합성 분야

- 기존의 음성합성 기술의 상당 부분이 그대로 이용될 수 있으나 음성 자동통역 시스템에서의 음성합성은 번역된 언어를 발화자의 음색에 맞게끔 합성되어야 하는 특징이 있다.

● 시스템 분야

- 음성 자동통역 기술은 음성인식, 기계번역 및 음성합성 등의 단위 기술이 서로 호환성이 있도록 결합되어야 하며 뿐만아니라 각 단위 기술내에 있는 유사한 작업들은 독립적인 반복수행이 아니라 한번의 수행으로 시스템 효율을 높힐 수 있도록 결합시켜야 하는 시스템 기술이다.

현재 음성 자동통역 분야에서 연구되고 있는 주요 분야는 다음과 같다.

● 대화체 음성 연구

- 대화체 음성은 발음이 분명하지 않고 발음이 틀리는 경우가 많으며 반복과 생략현상이 빈번하다. 그

리고 기침소리와 같은 소음 및 머뭇거림이 많으며 문법에 맞지 않은 문장이 많다. 이러한 현상에 대한 연구가 진행되고 있으며 대화체 음성인식시스템 개발이 이루어지고 있다[2][3].

- 인간 대 인간 사이의 대화에 나타나는 담화 연구 및 데이터 수집방법

인간 대 인간사이의 대화방법, 대화내용 등에 대한 연구가 진행되고 있으며[4], 데이터 수집을 위한 실제 상황 연출 및 음성 자동통역 시스템의 기능에 대한 연구도 진행되고 있다[5]. 최근의 자동통역 시연의 결과에 따르면 향후 음성 자동통역 시스템은 음성뿐만 아니라 화상, 문자 및 그림의 교환이 필요할 것이라는 연구결과가 제시되었다[4].

- 무제한 단어인식 연구

- 일반적인 대화 과정에서 나타나는 단어의 수는 무제한이기 때문에 무제한 단어인식에 대한 연구가 진행되어야 한다. 그러나 현재는 1000-3000 단어 수준의 대화 영역에 국한하고 있으며 인식 대상에 해당되지 않는 단어가 대화과정에 나타날 경우의 처리방법에 대한 연구만 주로 진행되고 있다[6].

본 논문에서는 현재 개발되고 있는 음성 자동 통역 시스템을 소개하고자 한다. 먼저 미국 MIT(Massachusetts Institute of Technology)에서 개발한 CCLINC와 미국 CMU(Carnegie Mellon University)에서 개발한 JANUS 시스템을 소개하고 일본 ATR(Advanced Telecommunication Research Institute)에서 개발한 ASURA에 대해서 기술하겠다. 또한 한국통신이 최근에 개발한 KT-STTS에 대해 소개하며 마지막으로 결론을 맺는다.

## II. 음성 자동통역 시스템의 개발 현황

### 2.1 미국 MIT의 CCLINC

MIT에서는 제한된 영역에서 음성 자동통역을 할 수 있는 CCLINC를 개발하였다[7]. 현재 주관심이 되고 있는 응용 분야는 다양한 언어를 사용하고 있는 연합군내에서 서로간의 군사작전에 관련된 통신을 원활히 하기 위한 분야이다. CCLINC는 영어를 프링스어 혹은 한국어로 음성 자동통역시켜 주는 화자 독립 시스템이며 현재 약 700단어를 통역시켜 준다.

그림 1에는 CCLINC의 시스템 구조가 그려져 있다. CCLINC는 여러개의 모듈로 구성되어 있으며 다국어에 맞도록 설계되어 있고 음성인식, 언어이해, 언어생성 및 음성합성으로 구성되어 있다. 음성이 인식되어 언어의 뜻이 이해되면 의미정보는 시맨틱 프레임(semantic frame)으로 표현되어 CCL(common coalition language)로 번역되어 저장된다. CCL은 군사용 중간언어(interlingua)로서 특정분야의 통신내용을 공통 형태로 표현해 주는 방식이다. CCL로 표현된 내용은 CCL망을 통해 여러 나라의 언어로 번역될 수 있다.

#### 2.1.1 음성인식부

CCLINC 시스템의 음성인식기는 대용량 화자독립 연속 음성인식 시스템(링컨 음성인식 시스템)을 사용하였으며[8], 언어이해 모델은 TINA를 사용하였다[9]. 음성인식부와 자연어 모델과의 결합 방식은 CCLINC 개발의 중요한 연구 목표 중의 하나였다. TINA 언어모델을 음성인식부내의 검색과정에 직접

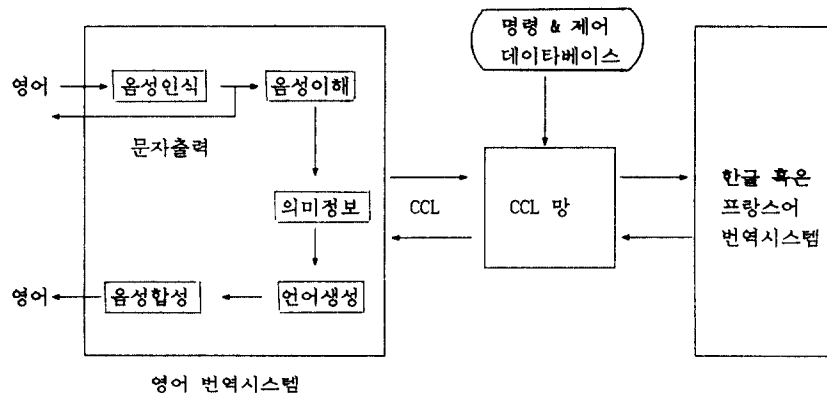


그림 1. CCLINC의 시스템 구조

결합시키는 밀접 결합(tightly-coupled) 방식과 음성 인식 결과의 출력을 언어 이해 모델로 연결시켜 주는 느슨한 결합(decoupled) 방식으로 나누어서 시스템을 구성한 후 서로 비교하였다. 밀접 결합 방식은 TINA 언어 모델을 직접 사용하기 때문에 인식 과정에서 특별히 다른 언어 모델을 사용하지 않았으며 느슨한 결합 방식으로 구성된 시스템은 인식 과정에서 바이그램(bigram), 트라이그램(trigram) 등과 같은 통계적 언어 모델을 별도로 사용하였다. 순수한 음성인식 결과만 고려하면 느슨한 결합 방식이 약간 좋았으나 번역 결과까지 포함하면 밀접 결합 방식이 좋은 것으로 나타났다[7].

### 2.1.2 기계번역부

CCLINC 시스템은 언어이해를 위해 TINA를 사용하였고 상대편 언어생성을 위해 GENESIS 시스템을 사용하였으며[10] 이 두 모듈을 결합하여 기계번역 과정을 수행한다. TINA는 구분 및 의미분석 능력을 지닌 문맥 자유 문법(context-free grammar)을 기본으로 한다. 실제로 기계번역이 되는 과정을 설명하면 다음과 같은 순서로 진행된다.

- ① 형태소 분석기가 수행된다.
- ② 형태소 분석기의 결과를 이용하여 파서가 수행되고 그 결과는 파서 구조 형태로 표현된다.
- ③ 파서 구조 형태는 의미 프레임(semantic frame)으로 표현되어 언어 생성 모듈인 GENESIS의 시작점이 된다.
- ④ 의미 프레임의 내용에 따라 사전을 참조하여 상대 언어로 변경시켜 주며 또한 문법에 따라 순서를 정한다.
- ⑤ 상대 언어의 문법에 따라 관사 등이 바르게 수정 보완된다.

위의 순서중 ①, ②, ③ 과정은 TINA 모듈에서 수행되며 ④, ⑤ 과정은 GENESIS 모듈의 수행내용이다.

### 2.1.3 음성합성부

CCLINC에서 고려하고 있는 음성합성기는 한국어, 프랑스어 및 영어를 위한 것이다. 한국어는 디지털에서 개발한 가라사대를 구입하여 사용하고 있으며 영어는 Eloquent Technology 회사 제품을 사용하고 있다.

### 2.1.4 평가

500문장을 이용하여 훈련을 시키고 190 문장에 대한 음성인식률과 번역 성공률을 표 1에 나타내었다. 음성인식 실험 결과 단어 및 문장의 오인식률은 각각

69.0%, 54.7%였다. 표 2에서는 음성 자동통역시스템의 음성번역률을 나타낸 것이다. 언어처리가 음성인식과 결합되면 정확한 음성인식이 안되더라도 의미적인 파악으로서 상대어로 번역이 가능하기 때문에 번역률은 문장인식률보다 높아질 수 있다. CCLINC는 프랑스어에 대해서는 58.9%의 번역을 나타내었으며 한국어에 대해서는 66.3%로 나타났다. 여기서 번역률은 번역 대상을 사용하는 사람이 적당한 번역이라고 생각되는 것을 점수화 한 결과를 나타낸 것이다.

표 1. CCLINC의 음성 오인식률

단어 오인식률	문장 오인식률
69.0%	54.7%

표 2. CCLINC의 번역률

파서 성공률(%)	번역률	
	프랑스어	한국어
76.8%	58.9%	66.3%

## 2.2 CMU의 JANUS

CMU에서 개발한 JANUS는 대화음성을 번역시켜 주는 음성 자동통역 시스템이며 JANUS-1은 500단어로 구성된 낭독체 음성(reading speech)을 통역시켜 주며[11], JANUS-2는 2000단어 정도가 사용되는 예약 업무에서 대화체 음성을 자동통역시켜 줄 수 있다[12]. 현재는 약속체 회의부(appointment scheduling)에서 영어와 독일어의 음성을 인식하여 영어, 독일어, 스페인어, 일어 및 한국어로 자동통역 시킬 수 있다. 또한 한국어 및 스페인어를 인식할 수 있도록 작업이 진행되고 있다[13]. 그림 2에는 전체적인 시스템 구성이 나타나 있다. 전체적인 시스템은 음성인식부, 파싱부, 담화(discourse)처리부, 생성부 및 음성합성부로 구성되어 있으며 각 부분은 언어에 독립적으로 설계되어 있는 것이 특징이다.

음성이 입력되면 음성인식부에서는 인식과정을 수행한 후 N개의 문장이 출력될 수 있는 구조로 변형시켜 파싱부로 입력이 되도록 한다. 파싱부에서는 입력된 인식구조를 근거로 파싱작업이 진행되고 언어에 독립적인 의미 형태로 인식 구조를 변형 시킨다. 파싱후엔 담화처리부에 입력이 되어 현재 입력의 파싱 결과가 애매할 때 과거의 대화내용에 따라 애매성을 해소시켜 준다.

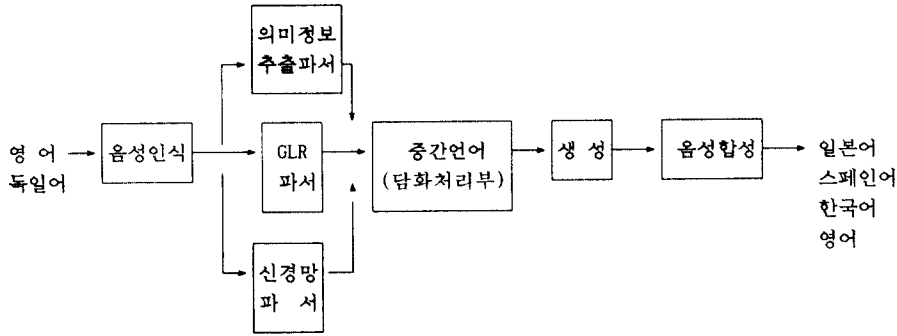


그림 2. JANUS의 시스템 구조

2.2.1 음성인식부

음성인식부는 특징추출, 음소모델 및 검색 알고리즘으로 크게 나누어지는데 각 부분의 특징은 다음과 같다.

가. 특징추출

주파수 영역이 멜 스케일 형태로 되도록 하였으며 파워를 사용하였다.

나. 음소모델

LVQ-2(learning vector quantization)나 반 연속 HMM(semi-continuous hidden Markov model)을 사용하여 트라이폰(triphone) 모델을 사용하였으며 소음 제거를 위해 소음모델을 독립적으로 사용하였다.

다. 검색 알고리즘

기본적인 검색 알고리즘은 비터비(Viterbi) 검색 알고리즘을 사용하였으며 N개의 문장을 출력할 수 있도록 단어의존(word-dependent) 검색방식을 채택하였다. 그리고 문법은 바이그램을 사용하였다.

2.2.2 기계번역부

기계번역부는 언어에 독립적인 구조로서 약속계획 업무에 필요한 정보를 근거로 하여 개발하였기 때문에 대상 언어마다 구조를 변경시킬 필요가 없다[14]. 이 기계번역부의 특징은 구문분석(syntactic analysis)을 수행하지 않고 직접 의미 정보로 표현되어 번역된다는 것이다. 그러므로 비 문법적인 표현이 많으며 음성인식 결과가 낮은 대화체 음성의 기계번역에 적합하다. 기본적인 구조는 강한 GLR(robust generalized left-right) 파서, 의미정보 추출 파서 및 신경망 파서

로 이루어지는 세 종류의 파서를 동시에 수행하여 의미정보가 잘 표현이 되는 중간 언어(interlingua) 형태로 변형시킨다는 것이 특징이다.

가. 강한 GLR 파서

강한 GLR 파서는 입력 문장이 파싱이 되지 않을 경우에도 입력단어를 생략하여 파싱이 성공할 수 있도록 하는 기능이 있다. 파싱의 결과는 의미를 나타내는 중간언어 형태로 변형된다.

나. 의미정보 추출 파서

특정 업무가 주어지면 그 업무를 표현해 주기 위한 정보는 언어에 관계없이 일정하므로 특정 업무를 의미정보로 표현해 줄 수 있도록 의미정보 추출 파서를 개발하였다. 이 파서는 피닉스 음성언어 시스템에서 사용되는 피닉스 파서와 유사하며[15], 입력을 의미 프레임으로 표현해 준다. 특히 구문분석을 수행하지 않기 때문에 입력이 문법에 정확하지 않아도 괜찮다.

다. 신경망 파서(connectionist parser)

신경망 파서 PARSEC는 강한 GLR 파서가 입력 문장 등을 파싱 못할 경우에 사용된다[16]. 이 파서의 특징은 훈련 과정에서 파서에 필요한 규칙을 자동적으로 구할 수 있으며 비 문법적인 요소가 강한 대화체 문장에 적합하며 억양(intonation)과 같이 다른 정보를 이용하기 위해서 쉽게 결합이 될 수 있다는 것이다[17].

이러한 세가지 파서에 의해 의미정보가 표현되면 그 의미정보에 해당되는 번역대상 단어가 채택되고 번역 문법에 의해 언어의 특성에 맞게끔 문장이 만들

어진다.

### 2.2.3 음성합성부

음성합성은 상용화된 제품을 사용하였다.

### 2.2.4 성능 평가

JANUS 시스템의 음성인식 결과는 사람과 사람사이의 대화체 음성을 사용하였기 때문에 성능이 낮다. 표 3에는 영어, 독일어 및 스페인어에 대한 음성인식률이 나타나 있으며 표 4에는 파싱 및 번역률에 대한 실험 결과가 나타나 있다[14].

표 3. JANUS의 대화체 음성 인식률

	영 어	독일어	스페인어
단어 인식률	66%	72%	61%

표 4. JANUS의 파싱 및 번역률

	파싱 성공률(%)	번역률(%)
영어	90.0	90.2
독일어	89.6	87.3
스페인어	58.3	82.2

## 2.3 일본 ATR의 ASURA

일본 ATR이 개발한 ASURA는 일본어를 영어 혹은 독일어로 번역하여 주는 음성 자동통역 시스템이다. 이 시스템은 음성인식부, 기계번역부 및 음성합성부로 크게 나누어져 있으며, 이전에 개발한 SL-TRANS라는 음성 자동통역 시스템의 성능을 향상시킨 것이다[18]. 그림 3에는 ASURA의 구조가 나타나 있다[19].

ASURA는 1993년 1월 28일 미국 CMU와 독일 Siemens 회사와의 동시 음성 자동통역 국제세미나회에서 일본 측에 의해 사용된 시스템이다.

### 2.3.1 음성인식부

ASURA에서 사용되는 음성인식기는 HMM-LR에

기반을 두었다[20]. HMM-LR이란 LR 과정을 사용하여 다음에 가능한 음소를 예측하는 음소기반 화자종속 HMM시스템이다. 또한 화자적응과정을 개발하여 화자독립시스템이 되도록 하였다[21].

### 가. 특징추출

음성특징추출을 위하여 LPC(linear predictive coding)를 이용한 켈스트럼(cepstrum)분석 방식을 사용하였으며, 16차 켈스트럼, 텐타 켈스트럼, 파워어 및 파워어를 특징으로 이용하였다.

### 나. 음소모델

음소모델을 위해서 HMnet(hidden Markov network)을 개발하였다[22]. HMnet는 기본 유니트로 이음(allophone)을 사용하였다. 그러나 이음은 문맥종속 유니트이므로 많은 종류가 발생한다. 그러므로 유사한 이음을 합치고 또한 출력확률이 비슷한 상태(state)를 다른 이음과 동시에 사용하여 음소모델의 안정성을 도모하도록 SSS(successively state splitting) 알고리즘이 개발되어 있다.

### 다. 화자적응

ASURA시스템이 채택한 화자적응 알고리즘은 VFS(vector field smoothing)이다. 이 알고리즘은 새로운 화자가 10단어를 말할 경우 그 음성으로부터 새로운 화자의 특성을 파악하여 기존의 HMnet을 변경시켜 준다[23].

### 2.3.2 기계번역부

음성인식부의 결과로서 N개의 문장이 기계번역부에 입력이 되면 구문 및 의미분석이 수행되어 분석이 가능한 한개의 문장이 선택된다. 기계번역부의 기본 알고리즘은 의도번역방식(intention translation method)에 근거를 두고 있다. 구문의 의미분석에 의해 의도정도가 표현이되면 변환과정에서 대상언어의 표현이

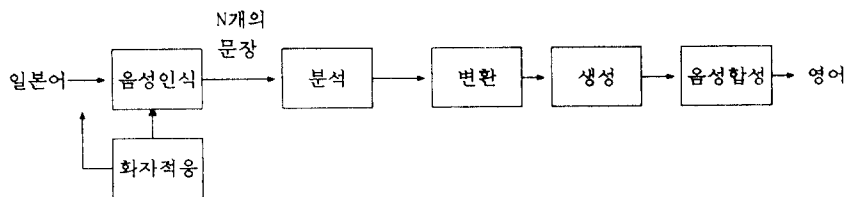


그림 3. ASURA의 시스템 구조

가능하도록 의도특징정보가 변환된다. 마지막으로 생성과정에서는 의도특징정보에 대응하는 구문트리 (syntax tree)가 구성되고, 해당되는 단어가 선정되어 번역이 완료된다.

2.3.3 음성합성부

ASURA는 ATR에서 개발한 음성합성기를 사용하며 이 음성합성기는 HP9000/730에서 수행된다.

2.3.4 성능평가

ASURA는 국제학술대회에 관한 문의를 할 때 사용되는 일본어 음성을 영어 혹은 독일어로 음성 자동통역시켜 주는 시스템이다. 약 700단어로 이루어진 문장에 대한 음성인식률과 번역성공률을 표 5에 나타내었다.

표 5에서 700단어로 이루어진 연속음성 인식시스템의 단어인식률은 top 1(첫 후보만을 고려했을 경우)으로서 86.7%, top 3(세번째 고려했을 경우)로서 92.5% 이었다. 음성 자동통역 시스템의 전체를 고려하면 90.3%의 번역 성공률을 얻을 수 있다. 그러나 한

문장이 번역되는 시간은 약 25초가 소요되었다[24].

표 5. ASURA의 음성인식률 및 번역성공률

	음성인식률(%)		번역 성공률(%)
	Top 1	Top 3	
일어→영어	86.7	92.5	90.3

2.4 한국통신의 KT-STs

한국통신이 개발한 KT-STs는 한국어를 일본어로 통역시켜주는 음성 자동통역 시스템이다[25]. 이 시스템은 한국 손님이 일본 호텔을 예약하고자 할 경우 사용되는 한국어 연속음성을 인식하여 일본어 문자로 표현하여 준다. 일본어 음성인식, 음성합성 및 일한 기계번역 시스템 개발을 위해 KDD와 공동연구를 수행하였으며 1995년 5월 16일 국제 시연회를 개최하였다. 그림 4에는 일본 KDD와 분담된 개발영역을 나타내었으며 그림 5에는 KT-STs의 구성도가 나타나 있다.

KT-STs는 약 300단어로 구성된 한국어 연속음성

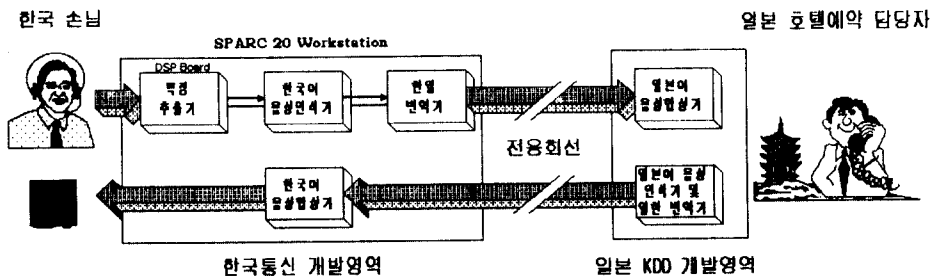


그림 4. 일본 KDD와의 개발영역

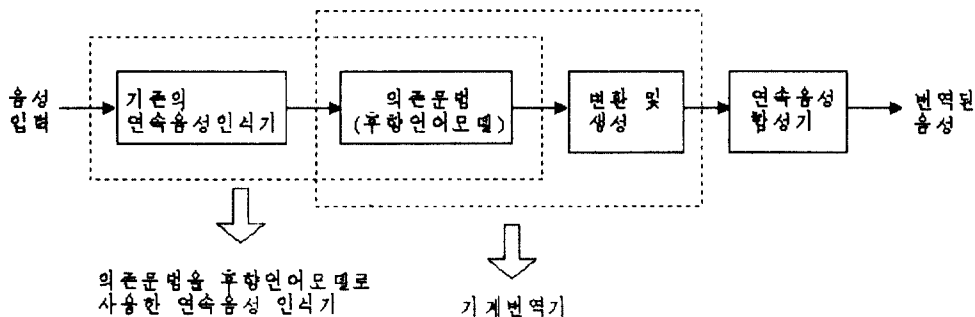


그림 5. KT-STs의 구성도

을 인식하는 화자독립 음성인식 시스템과 한-일 기계 번역 및 음성합성 시스템으로 구성되어 있다.

### 2.4.1 음성인식부

KT-STs의 음성인식부는 입력된 음성을 A/D하고 실시간으로 특징을 추출하는 특징추출과정과 특징에 따라 음성을 인식하는 검색과정으로 나뉘어진다. 특징추출과정은 DSP(digital signal processor) 보드 내에서 수행되며 검색과정은 SPARC20 컴퓨터에서 이루어진다. DSP 칩으로는 TI(Texas Instrument)회사에서 개발한 TMS320C30이 사용되었다. 음성이 입력되는 동시에 특징 추출 및 검색과정이 DSP 보드 및 SPARC20 컴퓨터에서 수행되므로 실시간으로 작동된다.

음성인식부의 기본 알고리즘은 VQ(vector quantization)을 이용한 비연속 HMM 알고리즘이다. 각 부분별의 상세 내용은 다음과 같다.

#### 가. 특징 추출

음성신호는 8 kHz,  $\mu$ -law, 8 비트로 샘플링되고,  $1-0.95z^{-1}$ 의 전달함수를 갖는 필터를 사용하여 전처리(pre-emphasize)된다. 이 음성은 프레임 단위로 분할되어 처리되는데 각 프레임은 20 msec의 길이를 가지며 10 msec씩 중첩된다. 매 프레임은 LPC 분석이 수행되고 이 LPC 계수를 이용하여 켈스트럼(cepstrum) 계수가 구해진다. 각 프레임에서 구한 LPC 계수는 아래의 가중치 윈도우(weight window)  $W_c$ 에 의해 가중치가 계산된다.

$$W_{c(m)} = 1 + \frac{Q}{2} \sin\left(\frac{\pi m}{Q}\right), 1 \leq m \leq Q$$

여기서 Q는 LPC 차수이다. 음성인식에는 가중치 LPC 켈스트럼(weighted LPC cepstrum) 계수와 그들의 빼기(difference), 이차 빼기(second order difference), 로그파워의 일차, 이차 빼기값을 벡터 양자화하여 사용한다.

#### 나. 음소 모델

HMM에 근거한 음성인식 시스템은 음성인식의 기본단위가 필요한데 본 시스템은 음소와 유사한 단위(phoneme-like unit)를 사용하였다. 기본 유니트 갯수는 60개를 선택하였으며 조음화현상을 고려하여 300개의 문맥종속 음소를 확장하였다. 음소모델은 7개의 상태와 12개의 천이(transition)를 가지며 그 천이들은 3개의 그룹으로 묶을 수 있으며 그룹의 천이는 같은

출력확률을 갖게 된다[26].

음성인식부는 아래와 같이 세 종류의 조음화현상을 고려하였다[27].

① 묵음 모델: 연속음성을 발음할 경우 단어사이의 묵음은 사람에 따라서 지켜지거나 연이어서 발음할 수 있다. 본 시스템에서는 널 천이(null transition)을 만들어서 묵음을 모델링하였다.

② 단어내 조음화 모델: 단어내의 조음화 현상을 모델링하기 위하여 트라이폰(triphone)을 사용하였다. 트라이폰 갯수를 결정할 때는 양이 너무 커지지 않게 하기 위하여 유니트 축소규칙(unit reduction rule)을 사용하였다[28].

③ 단어간 조음화 모델: 단어간 조음화 현상은 매 단어의 앞과 뒤에서 발생한다. 본 시스템은 훈련시에는 단어간 조음화 현상을 하나의 트라이폰으로 모델링하였으며 인식단계에서는 가능한 모든 트라이폰으로 모델링하였다[29].

#### 다. 언어처리 알고리즘

KT-STs에는 두 종류의 언어처리모델이 사용되었다. 첫번째 언어처리 모델은 바이그램이며 음성인식부에서만 사용된다. 두번째 언어처리 모델은 의존 문법을 사용한 차트 파서(chart parser)이며 음성인식부와 기계번역부에서 동시에 사용된다. 언어모델의 복잡도(perplexity)는 약 30이다. 한편 사용된 바이그램은 클래스에 근거하였으며 32개의 클래스가 사용되었다.

#### 라. 탐색 알고리즘

음성언어 시스템에서 음성과 언어처리의 효율적인 인터페이스를 위해 N개의 최적문장을 찾는 알고리즘을 사용하였다. 이 알고리즘은 시간 동기 비터비 빔(Viterbi beam)탐색 알고리즘으로서, 주어진 값(beam threshold)내에서 가장 적합한 N개의 문장을 찾아주는 알고리즘이다[30]. 최적문장의 갯수 N은 편의상 5로 지정하였다. 가능한 여러개의 문장을 찾아내는 방법은 여러가지가 있다. 현재는 문장종속 알고리즘과 격자 알고리즘의 중간 단계인 단어종속 N-best 알고리즘을 사용하고 있다.

### 2.4.2 기계번역부

기계번역부에서는 한국어 구문분석, 한일변환, 일본어 생성의 과정을 거친다. 한국어의 문법구조가 일본어와 유사하고, 호텔예약에 필요한 대화로 그 범위

가 한정되어 있기 때문에 한일 기계 번역은 비교적 간단히 구현될 수 있다. 음성인식기가 형태소 분석의 역할도 일부 수행하기 때문에 형태소 분석기는 사용하지 않았으며, 의존문법에 근거한 차트파서(chart parser)를 사용하였다. 한국어의 특성상 단어의 어순이 비교적 자유롭기 때문에, 고정된 어순의 특성을 갖는 언어에 적합한 구구조(phrase structure) 문법을 그대로 적용하기는 어렵다. 의존문법은 수식어와 피수식어간의 관계에 중점을 두기 때문에 이런 문제점을 해결해 준다[31]. 그리고 한국어의 경우에 술어가 문장의 맨끝에 오기 때문에 일반적으로 사용되는 방법과 달리 문장의 후미에서 전향으로 구문분석하는 방법을 사용하였다. 변환단계에서는 직접변환 방식이 사용되고, 생성단계에서는 간단한 생성규칙이 적용되었다.

2.4.3 음성합성부

음성합성부는 음성학적 전처리부, 운율발생부 그리고 합성부의 중요한 세 부분으로 나누어진다. 음성학적 전처리부에서는 입력된 문장을 성분분석하여 음운변동을 수행하고 운율정보 발생을 위한 기본적인 구문분석을 수행한다. 다음 운율발생부에서는 전처리의 결과를 받아 한국어에 적합한 억양, 길이, 세기 등의 운율을 발생시킨다. 합성부에서는 합성의 기본단위를 가져와 연결시킴으로써 문장을 만들어내는 데, 운율구현 및 음소간의 인터플레이션(interpolation)을 동시에 수행한다. 음성합성의 기본단위로 반음소(demiphone)를 사용하였다[32].

2.4.4 성능평가

KT-STSTS의 성능결과가 표 6에 나타나 있다. 연속음성인식의 성능은 top 1일 경우 94.5%, top 5일 경우 98.5%이며 전 시스템을 고려한 번역 성공률은 94.7%로 나타났다.

표 6. KT-STSTS의 음성인식률 및 번역성공률

인식률 (%)		번역성공률 (%)
Top 1	Top 5	
94.53	98.5	94.7

III. 결 론

본 고에서는 음성 자동통역 기술의 현황을 최근에

개발하였거나 현재 개발이 진행중인 음성 자동통역 시스템의 개발 내용을 중심으로 기술하였다. 미국은 MIT, CMU 등의 대학을 중심으로 영어, 일어, 독일어, 프랑스어 및 한국어 사이의 음성 자동통역 시스템을 만들고 있으며 특히 음성인식부와 기계번역을 위한 언어처리부 사이의 통합을 꾀하고 있다. 또한 낭독체 음성이 아니라 대화체 음성을 통역하는 분야에 연구가 집중되고 있다. 일본은 1993년 자동통역 연구소를 번역통신 연구소로 개편하여 음성 자동통역 시스템 개발에 박차를 가하고 있다. 국내에서는 한국통신과 한국전자통신연구소가 자동통역 시스템 개발에 참여하고 있다.

현재의 기술수준으로 볼 때 동시 통역자들이 통역할 수 있는 수준의 시스템을 개발한다는 것은 당분간 불가능하지만 업무영역을 잘 설정한다면 사용자에게 따라 제한적으로 사용가능한 시스템이 다양하게 개발될 수 있다. 그러나 자동통역 시스템의 실용화를 위해서는 단위기술인 음성인식, 합성 및 기계번역 기술의 실용화가 전제되어야만 한다. 최근에는 광대역 전송이 가능한 교환기가 개발됨에 따라 음성, 음악, 화상 및 문자 등의 멀티미디어 데이터의 전송이 가능해지게 되었고 음성 자동통역도 하나의 미디어 변환으로 고려될 수 있기 때문에 큰 의미에서 멀티미디어의 한 분야로 볼 수 있다. 그러므로 향후 음성 자동통역 분야의 연구방향은 음성번역만을 고려하는 경우와 다양한 미디어로 활용하여 음성번역을 시도하는 방향으로 연구가 진행될 것이다.

참 고 문 헌

1. S. Furui, M. M. Sondy, *Advances in speech signal processing*, Marcel Dekker, Inc. 1992
2. M. Okada, H. Ostuka, "Incremental elaboration in generation spontaneous speech.," *Proc. of Int., Symposium on Spoken Dialogue*, Tokyo, Japan, pp. 49-pp. 52 Nov. 1993
3. S. Seto, et al., "Spontaneous speech dialogue system TOSBURG II and its evaluation," *Proc. of Int. Symposium on Spoken Dialogue*, Tokyo, Japan, pp. 41-44, Nov. 1993
4. F. Yato, et al., "Important issues for automatic interpreting telephone technology," *Proc. of International Symposium on Spoken Dialogue*, pp. 235-238, Nov. 1993



5. C. Boitet, et al., "Human-machine-human interactions in interpreting telecommunications," *Proc. of International Symposium on Spoken Dialogue*, pp. 247-250, Nov. 1993
6. D. S. Pallet, et al., "1994 benchmark tests for the ARPA spoken language program," *Proc. of Spoken Language System Technology Workshop*, pp. 5-36, Jan. 1995.
7. D. Tummala, et al. "CCLINC: System architecture and concept demonstration of speech to speech translation for limited-domain multilingual," *Proc of 1995 ARPA Spoken Language Systems Technology Workshop*, pp. 227-232, Jan. 1995
8. D. B. Paul, et al., "The Lincoln large vocabulary stack-decoder HMM CSR," *Proc of ICASSP'93*, pp. 660-663, April, 1993
9. S. Seneff, "TINA: A natural language system for spoken language applications," *Computational Linguistics*, vol. 18, no. 1, pp. 61-86, 1992
10. J. Glass. et al., "Multilingual language generation across multiple domains," *Proc. ICSLP'94*, pp. 983-986, Sep. 1994
11. L. Osterhdtz, et al., "Testing generality in JANUS: a multi-lingual speech to speech translation system," *Proc. of ICASSP'92*, pp. 209-212, 1992
12. N. Woszczyna, et. al., "Recent advances in JANUS: a speech translation system," *Proc. of Human Language Technology*, pp. 211-216, Mar. 1993
13. B. Suhm, et al., "JANUS: toward multilingual spoken language translation," *Proc. of Spoken Language Systems Technology Workshop*, pp. 221-226, Jan. 1995
14. L. Mayfield, et al., "Concept based speech translation," *Proc. of ICASSP'95*, pp. 97-100, 1995
15. W. Ward, "Understanding spontaneous speech: the phoenix system," *Proc. of ICSSP'91*, pp. 365-367, 1991
16. A. J. Jain et al., "PARSEC: a structured connectionist parsing system for spoken language," *Proc. of ICASSP'92*, pp. 205-208, 1992
17. A. Waibel, et al., "Learning complex output representations in connectionist parsing of spoken language," *Proc. of ICASSP'94*, pp. 365-368, 1994
18. T. Morimoto, et al., "Enhancement of ATR'S spoken language translation system: SL-TRANS," *Proc. of International Conf. on Spoken Lang. Processing*, pp. 397-400, 1992
19. T. Morimoto, et al., "ATR' speech translation system: ASURA" *Proc. of EUROSPEECH'93*, pp 1291-1294, 1993
20. K. Kita, et al., "HMM continuous speech recognition using predictive LR parsing," *Proc. of ICASSP'89*, pp. 703-706, 1989
21. K. Ohkura, et al., "Speaker adaptation based on transfer vector field smoothing with continuous mixture density HMMs," *Proc. of Interaction Conf. on Spoken Lang. Processing*, pp. 369-372, 1992
22. J. Takami, et al., "A successive state splitting algorithm for efficient allophone modeling," *Proc. of ICASSP'92*, pp. 573-576, 1992.
23. H. Hattori, et al., "Vector field smoothing principle for speaker adaptation," *Proc. of Int. Conf. on Spoken Lang. Processing*, pp. 381-392, 1992
24. S. Satayama, et al., "ATREUS: a speech recognition front-end for a speech translation system," *Proc. of EUROSPEECH'93* pp. 1287-1290
25. M.W. Koo, et al., "KT-STTS: a speech translation system for hotel reservation and a continuous speech recognition system for speech translation," *Proc. of EUROSPEECH'95*, To be appeared, Sep., 1995
26. K.-F. Lee, *Automatic speech recognition: the development of the SPHINX system*. Kluwer Academic Publisher, Norwell, Mass., 1989
27. 구명완, "N개의 최적문장을 찾을 수 있는 한국어 연속음성인식 시스템," *음성통신 및 신호처리 워크샵 논문집*, pp.48-51, 1994년 10월.
28. C. H. Lee, et al., "Acoustic modeling of subword units for speech recognition," *Proc. 1990 IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 721-724, April, 1990
29. E. P. Giachin, et al., "On the use of inter-word context-dependent units for word juncture modeling," *Computer Speech and Language*, vol. 6, pp. 197-213, 1992
30. Y. Chow, et al., "The N-best algorithm: an efficient procedure for finding top N sentence hypotheses speech recognition performance," *Proc. Speech and Natural language*, pp. 147-149, Oct., 1989

31. I. A. Mel'cuk, *Dependency Syntax: Theory and Practice*, The State Univ. of New York Press, 1988
32. 김용인, 김재인, "한소리: 무제한 음성 합성 시스템," 음성통신 및 신호처리워크샵 논문집, pp. 342-345, 1994년 10월



구 명 완

- 1982년 : 연세대학교 전자공학과(학사)
  - 1985년 : 한국과학기술원 전기및 전자공학과(석사)
  - 1991년 : 한국과학기술원 전기및 전자공학과(박사)
  - 1985년 ~ 현재 : 한국통신 소프트웨어연구소 선임연구원, 음성언어연구팀장
- ※ 관심분야: 자동통역전화시스템 개발, 음성인식, 신경망, 음성합성