호텔예약을 위한 음성번역시스템

A Speech Translation System for Hotel Reservation

구 명 완*, 김 재 인*, 박 상 규*, 김 우 성*, 장 두 성*,
홍 영 국*, 장 경 애*, 김 응 인**, 강 용 범*
(Myoung-Wan Koo*, Jae-In Kim*, Sahng-Gyu Park*, Woosung Kim*, Du-Seong Chang*,
Youngkuk Hong*, Kyung-Ae Jang*, Eung-In Kim**, Yong-Bum Kang*)

요 약

이 논문에서는 호텔예약을 위한 음성번역시스템(KT-STS: Korea Telecom Speech Translation System)에 대해 기술한다. KT-STS는 한국손님이 일본의 호텔을 예약하고자 할 때 사용할 수 있는 시스템으로 한국어 음성을 인식하여 일본어로 번역을 해주는 시스템이다. 이 시스템은 한국어 음성인식부, 한일 기계번역부, 그리고 한국어 음성합성부로 구성되어 있다. 한국어 음성인식부는 HMM(Hidden Markov Model)에 근거한 화자독립, 300 단어급 연속음성인식시스템이다. 언어모델은 바이그램(bigram)을 전향 언어모델로, 의존문법을 후향 언어모델로 사용한다. 기계번역부에서는 의존문법과 직접 번역 방식을 사용하였다. 음성합성부에서 합성단위로 반음소를 사용하며 합성방식은 주기파형분해 및 재배치 방식을 이용한다. KT-STS는 TMS320C30 DSP 보드를 장착한 SPARC20 워크스테이션 상에서 거의 실시간으로 동작한다. 음성인식 실험결과 94.68%의 단어인식률과 82.42%의 문장인식률을 얻었으며, 한일 번역기만의 번역 성공률은 100%였다. 우리는 이 시스템과 일본 KDD에서 개발한 시스템을 전용선으로 연결하여 한일간 자동통역 국제시연을 가진 바 있다.

ABSTRACT

In this paper, we present a speech translation system for hotel reservation, KT-STS(Korea Telecom Speech Translation System). KT-STS is a speech-to-speech translation system which translates a spoken utterance in Korean into one in Japanese. The system has been designed around the task of hotel reservation(dialogues between a Korean customer and a hotel reservation desk in Japan). It consists of a Korean speech recognition system, a Korean-to-Japanese machine translation system and a Korean speech synthesis system. The Korean speech recognition system is an HMM(Hidden Markov model)-based speaker-independent, continuous speech recognizer which can recognize about 300 word vocabularies. Bigram language model is used as a forward language model and dependency grammar is used for a backward language model. For machine translation, we use dependency grammar and direct transfer method. And Korean speech synthesizer uses the demiphones as a synthesis unit and the method of periodic waveform analysis and reallocation. KT-STS runs in nearly real time on the SPARC20 workstation with one TMS320C30 DSP board. We have achieved the word recognition rate of 94.68% and the sentence recognition rate of 82.42% after the speech recognition tests. On Korean-to-Japanese translation tests, we achieved translation success rate of 100%. We had an international joint experiment in which our system was connected with another system developed by KDD in Japan using the leased line.

I. 서 론

최근 음성언어 처리 기술의 발전은 음성 번역의 가능성을 제시하고 있다. ATR의 자동통역 연구소는 미국 CMU (Carnegie Mellon University)와 독일의 Siemens Corporation/Karlsruhe 대학과 공동으로 자동통역 시연회를 가진 바 있다[1,2]. 또 NEC도 화자독립으로 일본어와 영어를 인식하여 영어, 일어 뿐 아니라 다른 언어로도 번역하여 음성합성을 통해 출력해 주는 자동통역 시스템을 만들었다[3]. 독일에서는 Vermobil이라는 대화 통역 과제가 1991년부터 시작되어 2001년까지 진행될 예정이다[4]. 그 밖에 C-STAR II라고 하여 전세계적인 음성번역 과제가 수행중에 있는데, 여기에는 미국의 CMU, 일본의 ATR,

* 한국통신 연구개발본부 멀티미디어연구소
Multimedia Technology Research Laboratory, Korea Telecom Research & Development Group
** 용인공업전문대학
Yong-In Techincal Cotlege
접수일자 : 1996년 3월 29일

독일의 Karlsruhe 대학, Siemens, 한국의 전자통신연구소 등이 주축이 되어 각기 언어들과 어느 언어이든 상대언어로의 번역을 할 수 있도록 하는 목표하에 연구가 진행되고 있으며 이 연구는 1999년 경에 완료될 예정이다.

이 논문에서는 음성번역 시스템에 대해 기술하였다. 2장에서는 시스템 개요를 기술하였으며 3장에서는 한국어 음성인식부, 4장에서는 한일 기계번역부에 대해 설명하였다. 5장에서는 음성합성부에 대해 기술하였으며 6장에서는 성능 평가 결과를 보여주고, 7장에서 결론을 맺는다.

Ⅱ. 시스템 개요

본 시스템에서 다루고 있는 문제 영역은 호텔 예약으로 한정하였다. 이는 음성번역이라는 기술 자체가 아직까지 실용화가 멀었기 때문에 그 영역을 한정하지 않고는 개발하기가 어렵기 때문이며, 또한 서비스의 관점에서도 호텔 예약이라는 영역이 매우 유용할 것이라 판단되었기 때문이다. 이 한정된 영역 내에서 우리는 호텔예약을 위해 실제로 사용되고 있는 자료들을 수집, 분석하여 이를 바탕으로 이 시스템에서 처리할 수 있는 말들에 대한 문법을 작성하였다[5]. 이 문법에서 사용되고 있는 전체 단어수는 약 300단어이다. 또 이 문법으로 생성 가능한 전체 문장수는 100만 문장 이상이다.

본 시스템은 음성인식부, 기계번역부, 음성합성부로 이루어져 있다. 음성인식부는 한국손님이 일본 호텔을 예약하고자 할 때 사용하는 말들을 인식하는 부분으로서 화자독립 연속음성인식을 수행하고 있다. 기계번역부는 인식된 한국어 결과를 받아서 이를 일본어로 번역시켜 주는 부분이다. 음성합성부는 일본에서 전송한 한국어 텍스트를 받아서 음성합성을 통해 한국손님에게 알려주는 부분이다. 이 시스템에서 수행하는 작업은 한국손님이 일본의 호텔을 예약하고자 할 경우에 필요한 부분, 즉 한국어 음성인식, 한→일 기계번역, 한국어 음성합성이다. 또 이에 대응되는 일본 호텔측의 부분은 일본 KDD에서 개발하였으며 이 두 시스템들은 다이얼업(dial-up) 모뎀을 통하여 서로 데이타를 교환할 수 있다. 단 데이타 전송의 안정성을 보장하기 위해 일본 KDD와의 국제 시연에는 전용회선을 사용하여 데이타를 전송하였다. 전체 시스템은 TMS320C30 DSP 보드를 장착하고, 2개의 CPU를 갖는 SPARC 20 workstation 상에서 구현되었다.

그림 1에 KT-STS의 개요가 나타나 있다. 한국손님이 '말하기' 버튼을 누른 후 한국어 문장을 말하면 KT-STS는 자동으로 음성의 끝점을 검출하고 이를 인식하여 N개의 후보 문장을 나타내고, 이 중 사람이 한 문장을 선택하면 그 문장에 대한 일본어 문장으로 번역된다. 한국어 음성합성기는 한국어 문장 텍스트를 받아서 음성으로 출력하는 역할을 한다.

일본 KDD와 공동으로 우리는 KT-STS를 사용하여 일본의 시스템과 1995년 5월 16일, 한일간 자동통역 국제시연회를 가졌다. 한일 양국 각각은 자국어의 음성인식기와 인식된 자국의 언어를 상호간 상대방 언어로 번역시켜 주는 번역기를 개발하였다. 한국통신은 한국손님 측의 역할을 담당하였으며, 일본 KDD는 일본 호텔 예약 담당자와 일본 손님의 역할을 담당하였다. 한국 호텔 예약 담당자의 역할은 전자통신연구소(ETRI)에서 담당하였다[6, 7].

Ⅲ. 음성인식부

본 시스템에서는 음성인식부에 문법과 조음화현상 등 연속음성의 특징을 고려한 N개의 최적 문장을 찾을 수 있는 한국어 연속음성인식 알고리즘을 사용하였다[8]. 그림 2에 한국어 연속음성인식 시스템의 구성을 나타내었다. 음성이 입력되면 음성의 특징이 추출되고 추출된 특징을 이용하여 단어 단위의 비교가 이루어진다. 비교 결과는 후보 단어의 열이된다. 문장 인식은 언어모델을 이용하여 수행되며 결과는 문장이 된다.

1. 특징 추출

음성신호는 8kHz, μ-law 8비트로 샘플링되고 $1-0.95z^{-1}$의 전달 함수를 갖는 필터를 사용하여 pre-emphasize된다. 이 음성은 프레임 단위로 분할되어 처리되는데 각 프레임은 20msec의 길이를 가지며 10msec씩 중첩된다. 매 프레임은 LPC 분석이 수행되고 이 LPC 계수를 이용

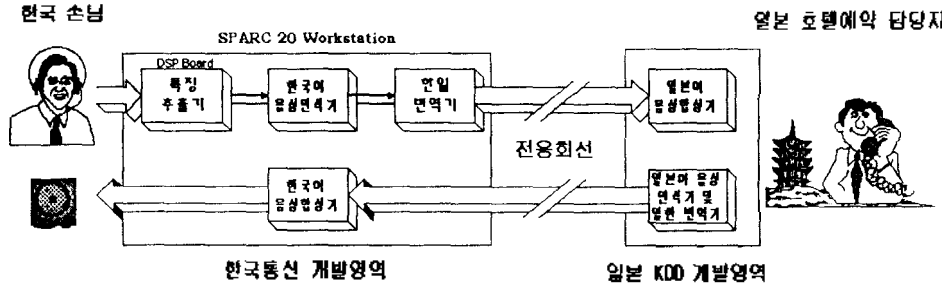


그림 1. 시스템 개요
Fig 1. System overview

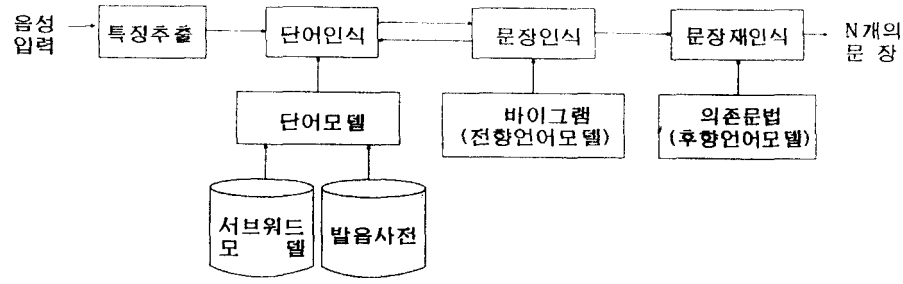


그림 2. 한국어 연속음성인식시스템
Fig 2. Korean continuous speech recognition system

하여 cepstral 계수가 구해진다. 각 프레임에서 구한 LPC 계수는 아래의 가중치 윈도우(weight window) W_c에 의해 가중치가 계산된다.

$$W_c(m) = 1 + \frac{\theta}{2}\sin\left(\frac{\pi m}{\theta}\right) \qquad 1 \le m \le Q$$

여기서 Q는 LPC 차수이다. 결국 음성인식에는 weighted LPC cepstral 계수와 그들의 폐기(difference), 이차 폐기(second order difference), 로그 파워의 일차, 이차 폐기 값을 벡터 양자화하여 사용하여 이들을 특징으로 사용한다. 각 파라미터들의 코드워드 수는 각각 256개, 256개, 256개, 그리고 로그 파워의 일차, 이차 폐기 값은 2차원으로 구성되며 12개의 코드워드를 갖는다.

2. 음소 모델

HMM에 근거한 음성인식 시스템은 음성인식의 기본단위가 필요한데 본 시스템은 음소와 유사한 단위(phoneme-like unit)를 사용하였다. 기본 유니트 갯수는 56개를 사용하였으며 조음화현상을 고려하여 문맥 종속 음소를 구하였다. 음소 모델은 7개의 상태와 12개의 전이를 가지며 그 전이들은 세개의 그룹으로 묶을 수 있으며 같은 그룹의 전이는 같은 출력 확률을 갖게 된다[9]. 그리고 HMM의 출력 확률을 계산하기 위하여 discrete 확률 분포를 사용하였다.

본 시스템은 아래와 같이 세 종류의 조음화현상을 고려하여 300개의 문맥 종속 음소를 구하였다. 실제 이론적으로는 더 많은 문맥종속음소수가 존재하지만, 음소수가 많아지면 그만큼 탐색공간이 커지게 되므로 성능이 저하될 수 있다. 또 실제로 우리가 사용하는 언어의 현상을 살펴보면 잘 사용되지 않는 음소들 또는 구태여 구분할 필요가 없는 음소들이 많이 있기 때문에 unit reduction rule에 의해 음소수를 줄였으며, 실제로 음소의 개수를 여러개로 바꿔가며 실험을 해보기도 했지만 300개일 경우에 성능이 가장 우수하였다.

1) 묵음 모델 : 연속음성을 발음할 경우 단어 사이의 묵음은 사람에 따라서 지켜지거나 연이어서 발음을 할 수 있다. 본 시스템에서는 null transition을 만들어서 묵음을 모델링하였으며 탐색영역이 줄도록 매 단어 끝에 묵음을 추가하였다.

2) 단어내 조음화 모델 : 단어내의 조음화 현상을 모델링하기 위하여 트라이폰(triphone)을 사용하였다. 트라이폰 갯수를 결정할 때는 양이 너무 커지지 않게 하기 위하여 unit reduction rule을 사용하였다[10].

3) 단어간 조음화 모델 : 단어간 조음화 현상은 매 단어의 앞과 뒤에서 발생한다. 본 시스템은 훈련시에는 단어간 조음화 현상을 하나의 트라이폰으로 모델링하였으며 인식 단계에서는 가능한 모든 트라이폰으로 모델링하였다. 그림 3에 단어간 조음화 현상을 고려한 트라이폰 구성을 나타내었다.

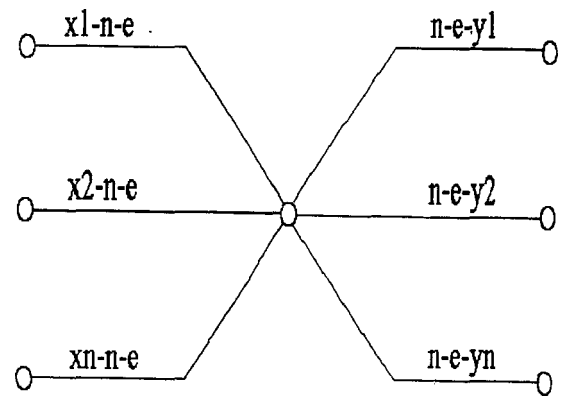
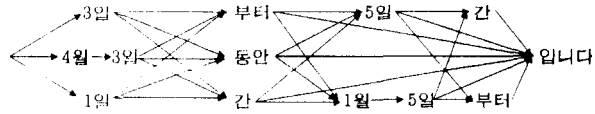


그림 3. "네"의 단어간 조음화 모델
Fig 3. Phonetic transcription of word "ne"

3. 언어모델

연속음성인식에서 언어모델은 신호 처리 방식에 의해 인식된 단어들의 연결된 리스트를 입력으로 하여 이들 단어가 이룰 수 있는 문장들 중 비문법적인 문장을 제거하는 역할을 한다. 본 시스템에서는 전향 언어모델로 바이그램(bigram) 문법을, 후향 언어모델로는 의존문법을

사용하였다. 즉 입력된 음성을 시간대의 전향으로 탐색하는 과정을 전향 언어모델이라고 한다면, 이 과정에서 단어 w_1 다음에 단어 w_2가 올 확률값 $P(w_2/w_1)$를 훈련시 구하고 인식 단계에서는 이 값을 단어 천이 확률값으로 사용한다[11]. 사용된 바이그램 문법은 단어로부터 추출한 44개의 클래스에 의거한 것이다. 또 음성의 입력이 끝난 후 인식된 단어의 격자 구조를 다시 후향 탐색하는 과정을 후향 언어모델이라고 할 수 있으며, 여기서는 의존문법[12]을 후향 언어모델로 구성함으로써 언어모델의 일반성과 제한된 인식 영역에서의 효율성을 동시에 가질 수 있도록 하였다[13]. 의존문법은 형태소간의 결합을 나타내는 구문요소화 규칙과 각 구문요소 간의 의존관계를 나타내는 의존규칙으로 이뤄지며, 한국어와 같은 중심어가 후위에 위치하는 언어에서 후향으로 분석을 할 경우 조기에 문법에 맞지 않는 문장을 제거할 수 있는 특징이 있다. 의미 모델은 사용하지 않는다.



a) 전향 언어모델이 적용된 단어인식의 결과

후보문장1 : 3일부터 5일간입니다.
후보문장2 : 4월 3일부터 5일간입니다.
후보문장3 : 3일간입니다.

후보문장9 : 3일부터입니다

b) 후향 언어모델을 통해 생성된 9개의 후보문장

그림 4. 중간결과
Fig 4. Intermediate results

그림 4는 제안된 언어모델을 사용한 한국어 연속음성 인식시스템의 흐름도이다. 이 시스템은 음성이 입력됨에 따라 음성의 특징을 추출하여 미리 수집된 단어 모델을 사용하여 단어 인식을 한다. 이 과정에서 전향 언어모델인 바이그램이 음성의 매 프레임 단위로 적용되어 각 단어들의 공기 가능성을 제약한다. 결국 그림 4(a)와 같이 인식된 단어의 후보 격자(Lattice)가 생성된다. 이 단어의 격자를 후향 탐색하는 과정에 후향 언어모델인 의존문법을 적용하여 그림 4(b)와 같이 N개의 인식 후보 문장의 리스트를 생성한다.

사용된 의존문법은 기존의 음성 인식시스템에서 사용되었던 문법과는 달리 한국어를 일반적으로 기술하고 있으며, 이러한 특성으로 인하여 인식시스템에서 후향탐색시 의존문법을 이용하여 분석된 의존트리는 기계번역과 같은 음성 인식된 결과를 입력으로 하는 자연 언어 처리 시스템에 쓰일 수 있다. 즉, 기계번역기의 입장에서 보면 구문 분석 이전의 단계를 음성 인식시스템과 공유하게 되는데 이는 인식 후보 문장을 선택할 때 같이 생성된 그 문장의 의존트리를 구문 분석 결과로 사용하고 있기 때문이다. 이 생성된 의존트리는 기존의 구구조문법을 사용한 언어모델에서와 달리 기계번역에서 변환 및 생성을 하기 위한 충분한 정보를 가지고 있으므로, 기계 번역기에서는 단지 변환 및 생성만을 담당할 수 있다. 이러한 구문분석기의 공유는 전체 시스템의 크기를 감소시키며, 기계번역만을 위한 구문 분석 단계가 없기 때문에 전체 통역 시간을 감소시키는 장점이 있다.

4. 탐색 알고리즘

탐색 알고리즘은 Viterbi 알고리즘을 사용하였으며 인식 시간을 향상시키기 위하여 beam 탐색 알고리즘을 사용하였다. 최적 문장의 갯수 N은 5를 사용하였다. N개의 최적 문장을 찾는 알고리즘은 word-dependent 알고리즘을 사용하였는데 단어내의 매 state에서 N보다 작은 n개의 가능한 path를 저장한다[14].

5. 데이터베이스

음성 데이터베이스는 호텔 예약에 많이 쓰이는 약 300개의 단어를 사용하여 20대부터 50대까지의 남성 및 여성으로 구성된 60명이 1인당 약 100문장씩을 발음한 것으로 구성하였다. 화자는 헤드셋트(headset) 마이크를 사용하여 사무실 환경에서 발음하고 발음된 음성은 CS4215 코덱이 내장된 A/D 변환 장치 SAIB를 통해 PCM 방식에 의해 컴퓨터에 저장되었다. 데이터베이스의 구성은 표 1과 같다. 60명의 화자 중 54명의 음성을 훈련에 사용하였으며 6명의 음성을 성능평가에 사용하였다. 또한 한 문장의 평균 단어 수는 8.8개이다.

표 1. 데이터베이스 구성
Table 1. Characteristics of evaluation database

연 령	사 람 수	
	남	여
20대	15	5
30대	15	5
40대	10	5
50대	5	

Ⅳ. 기계번역부

1. 호텔예약을 위한 대화체 기계번역 시스템의 설계

음성번역 시스템은 대화를 번역하기 때문에 기존의 기계번역과는 몇 가지 차이점을 가지고 있다. 대화는 화자와 청자간의 신뢰에 근거하여 이루어지기 때문에 문장의 일부가 생략되거나, 이전 대화에서의 내용을 문제 삼는 경우가 많다. 더욱이 이러한 대화의 번역은 실시간에 이루어져야만 하며, 모든 가능한 번역 형태 중에서 단 하나만의 번역 결과만을 생성해야 한다. 이러한 차이점은 기

존의 문장체를 대상으로 하는 기계번역 방식을 그대로 적용하지 못하게 하는 이유이며, 또한 대화체의 번역을 어렵게 하는 이유이다.

하지만, 대화는 항상 가정된 일정한 상황에서만 일어난다는 예측 가능성이 있으므로, 대화가 이루어지는 영역을 가정할 수가 있다. 그러므로, 일반적인 목적의 기계번역 시스템을 개발하기 전에 일정한 영역에서의 대화체 기계번역 시스템은 개발하는 것이 가능하다. 본 연구에서는 호텔예약 상황에서 이루어질 수 있는 대화체를 대상으로 기계번역 시스템을 구현하였으며, 이 과정을 통해 대화체 기계번역에 대한 기술을 습득하였다. 또한, 구현된 시스템은 전화망을 통한 음성번역 시스템의 일부로 쓰였다.

2. 호텔예약을 위한 한국어 해석기

호텔예약을 위한 한국어 해석기는 구문분석기와 사전, 의존문법으로 구성되어 있으며 현재 한국어 형태소 해석기를 사용하지 않는다. 그 이유는 입력의 단위가 형태소 혹은 어절로 구성되어 있는 음성인식기의 결과로서, 굳이 형태소 분석을 거치는 것이 불필요했기 때문이다. 음성번역기는 번역의 정확도와 더불어 실시간 처리도 중요한 성능 평가의 요소가 된다. 그러므로 인식기의 결과를 형태소 단위의 사전을 통해 매번 분석하는 것보다는 미리 분석된 형태소 분석 결과를 선택하는 방식을 사용하는 것이 유리하였기 때문에, 형태소 분석 과정을 생략하고 대신 단어 단위의 사전과 구문요소화 과정을 통해 구문 정보를 얻는 방법을 사용하였다.

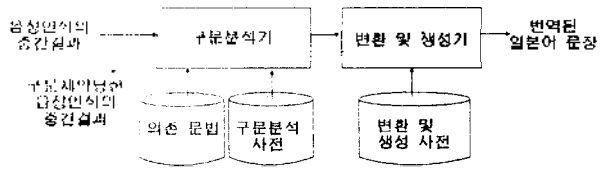


그림 5. 호텔예약을 위한 대화체 기계 번역 시스템
Fig 5. Dialogue machine translation system for hotel reservation

구문분석기는 음성인식기의 음성 탐색 과정에서 바이그램을 통해 1차 선택된 단어들의 격자를 입력으로 받아 구문 분석을 시도한다. 단어들의 격자를 입력으로 받는 과정에서 형태소 혹은 어절로 되어 있는 입력을 선형 시간 내에 구문 분석의 단위인 구문 요소로 바꾸어 주고 사전에서 구문 정보를 입수하는 구문요소화 과정을 거친다. 구문분석기가 사용하는 파싱 알고리즘은 챠트파싱이며, 입력 문장의 후위에서 문두쪽으로 분석해 나간다. 또한 사용하는 문법은 의존문법의 형태를 띄며, 이 의존문법을 사용하여 한국어를 우→좌로 분석해 나가는 방법은 챠트파싱에서 적은 파싱 결과를 생성할 뿐 아니라, 비문법적인 문장을 조기에 발견하는 이점을 가진다[15]. 이러한 이점은 한국어 해석기가 음성인식기의 후향 언어모델로 사용될 때 음성인식기의 성능을 높이는 데 큰 역할을 한다.

한국어 해석기에서 사용되는 사전은 호텔예약 영역에서의 사용 가능한 단어들과 이 단어들의 구문 정보를 저장하고 있으며, 구문 정보는 단어가 문장의 일부로서 담당하는 기능의 정보인 기능 범주 정보와 문장 전체에서 그 단어에 해당하는 의미를 나타내는 의미 범주 정보로 구성되어 있다. 사전의 검색어는 음성인식기에서 사용하는 인식 대상 단어와 동일한 형태, 즉 형태소 혹은 어절이며, 검색어가 어절일 경우 검색어의 구문 정보에는 기능 범주 정보와 의미 범주 정보가 있고 형태소일 경우 형태소의 종류에 따라 그 중 하나의 정보만이 저장되어 있다.

3. 한→일 변환 및 일본어 생성기

이 시스템에서의 번역 방식은 분석된 한국어의 의존트리에서 일본어에 맞는 구조의 변환을 일부 거친 후 일본어를 생성해 내는 방식을 사용하고 있다. 한국어와 일본어는 그 문장의 구조가 비슷하며, 특히 호텔예약 영역은 단문 중심의 대화체이기 때문에 그 문장의 구조 변환은 최소한으로 그치고 있으며, 대신 분석된 문장 구조는 일본어를 생성하는 과정에서 일본어 형태소들의 변환에 많은 도움을 준다. 변환 및 일본어 생성 과정은 변환 및 생성 사전을 참조하여 이루어지며, 이 사전은 구조 변환 및 일본어 생성에 관한 정보를 수록하고 있다.

생성 사전에서의 일본어는 그 구현 및 관리의 효율을 위하여 KSC5601안에 할당되어 있는 일본어 코드로 구성되어 있으며, 이는 실제 한·일 음성번역 시스템에서 일본어 합성기의 입력으로 쓰이지 못한다. 그러므로, KSC5601에서 JIS0212로 변환하는 루틴을 매핑테이블로 구성하여 코드 변환을 위해 사용하였다.

V. 음성합성부

음성합성부는 음성학적 전처리부, 운율발생부 그리고 합성부의 중요한 세 부분으로 나누어진다. 음성학적 전처리부에서는 파일이나 키보드 입력에 의한 문장을 성분분석하여 음운 변동을 수행하고 운율 정보 발생을 위한 기본적인 구문 분석을 수행한다. 다음 운율발생부에서는 전처리의 결과를 받아 한국어에 적합한 억양, 길이, 세기 등의 운율을 발생시킨다. 합성부에서는 합성의 기본단위를 가져와 연결시킴으로써 문장을 만들어 내는데, 운율 구현 및 음소간의 인터폴레이션(interpolation)을 동시에 수행한다[16].

1. 합성단위

본 시스템에서는 새롭게 제안된 반음소(demiphone)를 합성의 기본단위로 사용한다. 반음소는 음소를 그것의

정상 상태 시점인 중심을 기준으로 해서 다시 양분함으로써 얻어진다. 음소를 양분하여 얻어진 두개의 반음소 중에서 먼저 것을 전반 음소(initial demiphone), 나중 것을 후반 음소(final demiphone)라고 한다. 반음소의 경계는 음소 및 다이폰의 경계와 일치한다. 따라서 전반 음소와 후반 음소들을 적당히 결합함에 따라 음소를 만들 수도 있고 다이폰을 만들 수도 있다. 이와 같은 성질 때문에 반음소는 음소와 다이폰의 장점을 동시에 가지게 된다. 다시 말하자면 음소와 마찬가지로 다루기 쉽고 메모리 양을 적게 필요로 하며, 다이폰과 마찬가지로 합성시 얻어지는 합성 음성의 음질이 좋게 되는 장점이 있다[17].

2. 음성학적 전처리

무제한 음성합성 시스템에 있어서 음성학적 전처리 단계는 문자열 정형화부(text preformatting block), 문장 구조 추출부(parsing block), 음운 변동 처리부(phonetic recoding block)로 세분될 수 있다.

문자열 정형화부는 문자열이 입력되면 숫자, 알파벳, 많이 쓰이는 영어 단어, 약자 등을 입력한 사전을 참조하여 문자열 속에 있는 모든 약어, 숫자, 특수 기호 및 수식, 수사 처리를 하여, 발음 가능한 문자열과 제어 문자 구독기호(punctuative symbol)들로만 구성된 정형화된 문자열을 생성하며, 1300여 단어의 발음 예외 사전의 탐색 및 치환 과정을 수행한다.

문장 구조 추출 과정은 조사, 어미, 선어말어미 등으로 구성된 형식 형태소 사전(약 400단어)과 관형사, 부사, 불완전명사들로 구성된 실질 형태소 사전(약 230단어)을 참조하여 정형화된 문자열에 대해 구문 해석을 함으로써 수식, 피수식 관계 및 구, 절의 경계점을 검출하고, 구 및 절의 통사적 기능을 결정한다. 음운 변동 처리부는 발음 예외 사전 탐색 결과 형태소 분석 결과 문장 구조 추출 결과 및 음운 변동 알고리듬을 이용하여 자소 단위의 변환을 수행한다[18,19].

3. 운율 생성

운율조절부에서는 운율의 기본 요소인 각 음소의 길이, 억양 그리고 세기를 조절한다. 첫째로, 길이 조절을 위하여서는, 음소의 길이가 음성학적 요인, 구문론적요인 그리고 발음 속도에 의하여 변하는 것으로 모델링을 하여 길이 조절 규칙을 제안하였다. 즉, 임의의 음소의 길이는 전후의 음소의 음성학적 성질에 따라 적절히 줄어드는 것으로 모델링을 하였으며, 문장의 끝이나 억양구의 끝에서는 길이가 늘어나도록 하였다. 둘째로 억양에 있어서는 어절 단위, 억양구 단위 그리고 절단위의 억양 조절 규칙을 사용하였는데, 이 조절규칙은 기본적으로 어절 단위의 피치 패턴 규칙과 baseline resetting규칙으로 이루어져 있다. 다음 세번째로 각 음소의 세기 조절 규칙에서는 먼저 음소별 기준 세기를 정한 후에 부여된 피치의 크기에 따라서 선형적으로 증감하는 방법으로 사용하였다[20].

4. 합성방식

원음성(original speech) 중의 유성음 구간의 신호를 각 성문 펄스(glottal pulse)에 의해 만들어지는 한 주기분의 음성파형에 해당하는 단위파형(unit waveform 또는 wavelet)들로 분해하는 주기파형분해 방식과 저장된 단위파형 들 중 배치시키고자 하는 위치에 가장 가까운 단위파형을 선택하여 그것들을 서로 중첩시킴으로써 원음성의 음질을 그대로 가지면서도 음성 단편의 지속 시간(duration)과 피치 주파수(pitch)를 임의대로 조절할 수 있게 하는 시간 왜곡식 단위파형 재배치 방식을 합성방식으로 사용한다.

주기적 음성을 그것의 스펙트럼 포락 함수의 시간 영역 함수인 임펄스 응답과, 주기적 음성과 주기가 같고 평탄한 스펙트럼 포락을 가진 주기적 피치펄스열 신호로 디콘벌루션(deconvolution)한 다음 이포크 검출 알고리즘(epoch detection algorithm)[21]과 같은 시간 영역에서의 피치 검출 알고리즘을 이용하여 주기적 피치펄스열 신호나 시간 영역의 음성파형으로부터 피치 펄스들의 위치를 구한 후 피치 펄스가 한 주기구간당 하나씩 포함되도록 피치펄스열 신호를 주기적으로 분할하고, 유효 지속 시간에 따라 파라미터 연장과 영샘플 추가를 한 후, 이 피치펄스 신호들을 그 주기 구간 동안의 임펄스 응답과 다시 콘벌루션시키면 단위파형이 구해진다.

Ⅵ. 성능 평가

본 시스템은 호텔예약 과정에서 사용될 수 있는 문장들에 대해 실패 없이 음성번역이 가능하도록 구현되었다. 대화체 기계번역 시스템은 본 시스템에서 사용된 모든 문장을 1μsec이내에 번역할 수 있도록 구현되었으며, 음성합성 시스템 또한 사용된 모든 문장을 사람이 큰 주의를 기울이지 않고도 쉽게 알아들을 수 있는 수준의 음성을 실시간에 합성하도록 구현되었다.

연속음성인식 시스템의 인식률은 표 2에 나타나 있다. 실험에 사용한 데이터 베이스는 앞서 설명한 바와 같다. 표에서 Top1은 첫번째 후보만을 고려한 인식률이며, Top5는 5번째 후보까지 포함하였을 경우의 인식률이다. 이 시스템은 94.68%의 단어인식률과 82.42%의 문장인식률을 보였다.

표 2. 한국어 연속음성 인식시스템의 성능평가표
Table 2. Evaluation results of Korean continuous speech recognizer

단어 인식률 (%)		문장 인식률 (%)	
Top1	Top5	Top1	Top5
94.68	98.27	82.42	95.07

한국어 음성 인식된 결과를 갖고 실험한 한일 기계번역시스템만의 번역 성공률은 100%이었다. 이는 본 시스

템이 호텔예약이라는 한정된 분야에서만 동작되기 때문이다. 즉 음성 인식된 한국어 결과가 이미 우리가 만든 특정한 문법 범위내에서만 존재하며 따라서 이를 입력으로 하여 생성될 수 있는 일본어 번역 결과도 매우 제한된 것이기 때문이다. 단 여기서 말하는 번역 성공률은 음성인식된 결과를 기준으로 한 것이기 때문에 음성인식 자체의 오류는 제외된 것이며, 음성인식시스템의 오류를 포함한 전체 음성번역률은 음성인식률과 동일한 결과, 즉 82.42%가 된다. 전체 시스템이 한국어 음성 입력으로부터 일본어 문장을 생성하기까지 거의 실시간에 동작한다.

Ⅵ. 결 론

이 논문은 한국손님이 한국어 만을 사용하여 일본 호텔을 예약할 수 있도록 해 주는 호텔예약을 위한 음성번역 시스템에 관한 것이다. 음성번역 기술은 음성인식, 합성, 기계번역이 결합된 첨단 기술로서 본 연구에서는 한국어 음성인식, 한→일 기계번역, 한국어 음성합성 시스템이 결합된 음성번역 시스템을 구현하였다. 일본 호텔측의 영역은 일본 KDD에서 개발하였고, 우리가 개발한 시스템과 다이얼업 모뎀 또는 전용선을 통해 통신할 수 있다. 음성인식 시스템은 HMM을 이용하는 300 단어급 화자독립 연속음성인식 시스템으로서 전향언어모델로 바이그램, 후향언어모델로 의존문법을 사용하여 N-best 문장을 생성해 낸다. 실험결과 94.68%의 단어 인식률, 82.42%의 문장 인식률을 얻었다. 기계번역은 직접 번역 방식을 채택하였고, 음성합성은 반음소를 기본 단위로 주기파형분해 및 재배치 방식을 사용하였다. 이 시스템을 사용하여 일본 KDD 시연 시스템과의 국제 시연에 성공하였으며 이를 통해 음성번역의 가능성을 제시하였다고 할 수 있다.

참 고 문 헌

1. T. Morimoto, et al., "ATR's speech translation system: ASURA", *Proc. 3rd European Conf. on Speech Communication and Technology*, pp. 1291-1294, Sep. 1993.
2. A. Waibel, et al., "JANUS-a speech-to-speech translation system using connectionist and symbolic processing strategies," *Proc. 1991 International Conf. on Acoustics, Speech, and Signal Processing*, pp. 793-796, May 1991.
3. K. Hatazaki, et al., "INTERTALKER: an experimental automatic interpretation system using conceptual representation," *Proc. 1992 International Conf. on Spoken Lang. Processing*, pp. 393-396, Oct. 1992.
4. W. Wahlster, et al., "Verbmobil, translation of face-to-face dialogs," *Proc. 3rd European Conf. on Speech Communication and Technology*, pp. 29-38, Sep. 1993.
5. 김 우성, 구 명완, "호텔 예약을 위한 한국어 음성 데이터 베이스에 관한 연구", *한국 정보과학회 '92 추계 학술발표회 논문집*, 1992.
6. Myoung-Wan Koo, Il-Hyun Sohn, Woo-Sung Kim, Du-Seong Chang, "KT-STS: A Speech Translation System for Hotel Reservation and a Continuous Speech Recognition System for Speech Translation," *Proc. European Conference on Speech Communication and Technology*, pp. 1227-1230, Madrid, Sep. 1995.
7. 구 명완, 김 응인, 김 재인, 도 삼주, 강 용범, 박 상규, 손 일현, 김 우성, 장 두성, 이 종락, 김 진영, "호텔예약을 위한 자동통역 시스템", *음성통신 및 신호처리 워크샵 논문집*, pp. 105-108, 1995년 6월.
8. 구 명완, "N개의 최적문장을 찾을 수 있는 한국어 연속음성인식 시스템", *음성통신 및 신호처리 워크샵 논문집*, pp. 48-51, 1994년 10월.
9. K. F. Lee, *Automatic speech recognition: the development of the SPHINX system*. Kluwer Academic Publisher, Norwell, Mass., 1989.
10. C. H. Lee et al., "Acoustic modeling for large vocabulary speech recognition," *Computer Speech and Language*, vol. 4 pp. 127-165, 1990.
11. F. Jelinek, "The development of an experimental discrete dictation recognizer," *Proc. IEEE*, pp. 1616-1624, 1985.
12. I. A. Mel'cuk, *Dependency Syntax: Theory and Practice*, The State Univ. of New York Press, 1988.
13. Du-Seong Chang, Myoung-Wan Koo, "A Korean continuous speech recognition system using dependency grammar as a backward language model", *Proc. of the Natural Language Processing Pacific Rim Symposium '95*, pp. 646-651, 1995.
14. R. Schwartz and S. Austin, "Efficient, high-performance algorithms for N-best search," *Proc. of the DARPA speech and natural language workshop*, pp. 6-11, 1990.
15. C. H. Kim et al., "A Right-to-Left Chart Parser for Dependency Grammar using Headable Paths," *Proc. of the 1994 International Conference on Computer Processing of Oriental Languages*, pp. 175-179, 1994.
16. 김 응인, 김 재인, "한소리: 무제한 음성 합성 시스템," *음성통신 및 신호처리워크샵 논문집*, pp. 342-345, 1994년 10월.
17. 이 종락, "반음소: 새로운 음성합성 및 인식단위," *음성통신 및 신호처리워크샵 논문집*, 1993년 8월.
18. 강 용범, 안 치홍, "무제한 음성합성 시스템을 위한 문장구조 추출에 관한 연구," *음성통신 및 신호처리 워크샵논문집*, 1993년 8월.
19. 강 용범, 김 진영, "무제한 음성합성 시스템을 위한 전처리 과정," *음성통신 및 신호처리워크샵 논문집*, pp. 334-337, 1994년 10월.
20. 김 진영, 성 굉모, "한국어의 억양에 관한 연구," *Korean-Japan Joint Symposium on Acoustics*, pp. 292-297, 1991.
21. C.d'Alessandro, J.S. Lienard, "Decomposition of the Speech Signal into Short-Time Waveform Using Spectral Segmentation," *IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1988.

▲구 명 완(Myoung-Wan Koo)

1982년 : 연세대학교 전자공학과 졸업(학사)
1985년 : 한국과학기술원 전기 및 전자공학과 졸업(석사)
1991년 : 한국과학기술원 전기 및 전자공학과 졸업(박사)
1985년~현재 : 한국통신 멀티미디어연구소 음성언어연구팀 팀장
※주관심분야 : 음성번역시스템, 음성인식, 합성, 신경망

▲김 응 인(Eung-In Kim)
1984년 : 경북대학교 전자공학과 졸업(학사)
1986년 : 경북대학교 대학원 전자공학과 졸업(석사)
1986년~1996년 : 한국통신 멀티미디어연구소 음성언어연구팀 선임연구원
1996년~현재 : 용인공업전문대학 전임강사
1994년~현재 : 한국과학기술원 박사과정
※주관심분야 : 음성신호처리, 디지털 신호처리, 이동통신

▲김 재 인(Jae-In Kim) : 14권 1호 참조
현재 : 한국통신 멀티미디어연구소 음성언어연구팀 선임연구원

▲강 용 범(Yong-Bum Kang)
1991년~1995년 : 한국통신 멀티미디어연구소 음성언어연구팀 전임연구원

▲박 상 규(Sahng-Gyu Park)
1989년 : 경북대학교 전자공학과 졸업(학사)
1991년 : 한국과학기술원 전기 및 전자공학과 졸업(석사)
1991년~현재 : 한국통신 멀티미디어연구소 음성언어연구팀 전임연구원
※주관심분야 : 음성인식, 합성, 음성신호처리

▲김 우 성(Woosung Kim)
1990년 : 한국과학기술원 과학기술대학 전산학과 졸업(학사)
1992년 : 포항공과대학교 대학원 전자계산학과 졸업(석사)
1992년~현재 : 한국통신 멀티미디어연구소 음성언어연구팀 전임연구원
※주관심분야 : 음성인식, 언어처리, 신경망

▲장 두 성(Du-Seong Chang)
1990년 : 전남대학교 전산통계학과 졸업(학사)
1993년 : 한국과학기술원 전산학과 졸업(석사)
1993년~현재 : 한국통신 멀티미디어연구소 음성언어연구팀 전임연구원
※주관심분야 : 자연언어처리, 음성언어처리

▲홍 영 국(Youngkuk Hong)
1992년 : 포항공과대학교 전자계산학과 졸업(학사)
1994년 : 포항공과대학교 대학원 전자계산학과 졸업(석사)
1994년~현재 : 한국통신 멀티미디어연구소 음성언어연구팀 전임연구원
※주관심분야 : 자연언어처리, 음성언어처리, 한국어 정보처리

▲장 경 애(Kyung-Ae Jang)
1990년 : 경북대학교 자연대학 유전공학과 졸업(학사)
1991년~1995년 : 한국통신 경주전화국 근무
1995년~현재 : 한국통신 멀티미디어연구소 음성언어연구팀 전임연구원
※주관심분야 : 음성인식, 합성, 음성정보서비스 시스템