

## 한국어 음성인식 응용서비스 및 연구 과제

구 명 완

### Abstract

In this paper, we present the definition of speech recognition technology and its application services currently working over telephone network. And some research issues especially related to Korean are also shown among basic problems of speech recognition technology. The current statue of the speech recognition technology which will eventually change our future life style is still in the early stage. We show three kinds of services developed by Korea Telecom such as voice dialing, railroad information and name dialing services. And we also describe some research issues obtained throughout our services.

주요어 : 음성인식(Speech Recognition), 한국어 음성인식(Korean Speech Recognition)  
전화 서비스(Telephone Service)

### I. 서론

음성인식 기술이 국내에 소개된 것이 1980년대였으나 이 때는 주로 논문연구 및 단순한 시험시스템 개발에 머물렀다. 1980년대 말부터 음성인식 기술의 상용화가 세계적으로 시작되었고, 신경망 및 퍼지 기술의 개발로 인해 국내에서도 본격적인 연구가 시도되었다. 그러나 국내의 일부 연구소 및 기업에서 개발한 시제품 및 시험서비스는 큰 호응을 얻지 못하였다. 한편 1990년 말부터는 음성인식 기술을 전문으로 하는 중소기업체가 생겨나고 해외의 유명 회사가 국내에 진출하면서 음성인식 기술에 대한 일반인의 관심이 증대되기 시작하였다. 특

히 2000년도에서는 벤처업계의 활성화에 힘입어 음성처리 관련회사가 100여 개로 증가되고 다양한 제품 및 서비스가 제공됨으로써 음성처리 관련산업이 하나의 산업영역으로 대두되기 시작하였다(1)(2).

음성인식 기술의 비전은 “2001: A Space Odyssey”라는 영화에서 나오는 HAL9000 컴퓨터의 음성인식 능력을 최종 목표로 한다. 즉 “사람과 자유롭게 대화하고 장기 및 노래 뿐 만 아니라 문제까지 풀어 줄 수 있는 컴퓨터”이다. 혹은 최근의 영화인 스타 워즈(Star Wars)에 나오는 C3PO라는 로봇의 음성인식 능력이다. 그는 수많은 언어를 이해하고 통역할 수 있는 능력을 가졌다. 그러나 현재의 음성인식 기술은 좁은 영역에서만 인식이 가능하고 음성통역도 특정 영역에서만 가능하다(3)(4). 한편 미국 MIT 대학에서는 음성인식 기술을 21세기를 이끌어 갈 10대 기술의 하나로 간주하였으며, 2001년 1월 스위스 다보스에서 열린 제 31차 경제 포럼에서는 과학 기술자들이 2010년까지 우리 일상생활을 바꿔 놓을 7대 신기술로 음성인식을 선정하였다.

본 고에서는 음성인식 기술의 기본 정의와 최근의 한국어 음성인식 기술을 이용한 대표적인 응용 서비스에 대해서 알아본다. 먼저 II장에서는 음성인식 기술의 기본 정의에 대해 알아보고 III장에서는 최근의 음성인식 응용 서비스에 대해서 기술한다. 특히 한국통신이 최근에 개발한 전화망 응용서비스를 중심으로 소개한다. 그리고 IV장에서는 음성언어학 측면의 연구과제에 대해서 알아본다. 마지막으로 V장에서 결론을 맺는다.

## II. 음성인식기술의 정의

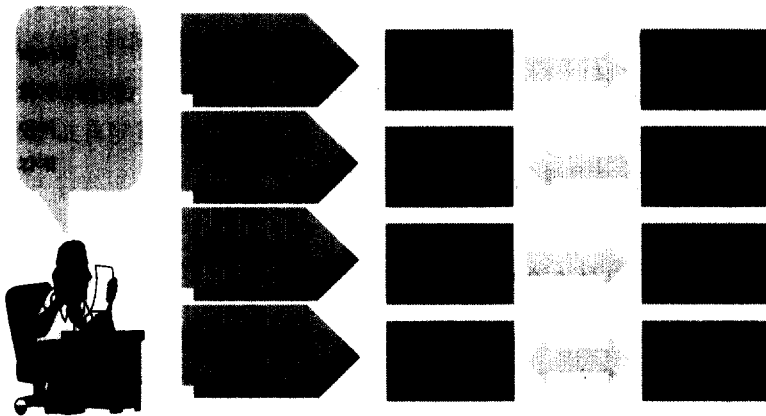
음성인식 기술의 정의는 <그림 1>에 나타나 있듯이 음성으로부터 의미를 추출하여 문자로 변환시켜주는 기술이다. 음성은 사람마다 다르고 지역마다 다르기 때문에 사용하는 사람의 목소리만 인식할 수 있는 화자 종속 시스템과 남녀 노소의 모든 음성을 인식하는 화자 독립 시스템으로 나누어진다. 또한 인식할 수 있는 단어 량에 따라 시스템이 구분이 된다.

음성합성 기술은 음성인식과 반대로 문자로부터 음성을 변환시켜 주는 기술

이다. 즉 문자가 입력이 되면 음성으로 변환이 되는데 남자, 여자, 어린이 등의 목소리로 변환시켜주는 기술이다. 이 기술의 핵심은 사람과 같은 자연스러운 음질을 만드는 것이다. 최근에는 한국어 음성합성 기술의 발달로 상당히 이해도가 높은 수준의 합성기가 개발되어 있다.

화자인식 기술은 음성의 개인성 정보를 이용하는 것이다. 즉 사람마다 목소리가 다르다는 점을 이용하여 음성으로부터 사람을 확인하는 기술이다. 이 기술은 음성으로부터 사람을 찾는 화자확인(speaker identification) 기술과 목소리를 이용하여 문을 열거나 하는 보안에 사용될 수 있는 화자 인증(speaker verification) 기술로 나누어진다.

음성번역은 한국어 음성을 영어 혹은 일어로 통역해주는 기술이다. 이 기술은 음성인식, 음성합성 및 기계번역 기술이 결합되어 있는 기술로 아직 연구단계에 있다. 최근의 일본어와 영어사이의 음성통역 연구결과에 따르면 토익 550 정도의 수준의 기술이 개발되었다고 한다.



<그림 1> 음성인식 기술의 정의

### Ⅲ. 음성인식 응용서비스

#### 1. 한국통신의 전화망 응용서비스

국내에서 가장 많이 사용되고 있는 전화망 응용서비스는 음성인식 증권 서비스이다. 현재 많은 증권정보 안내 사이트에 음성인식 기술이 적용되어 있다. 최근에는 미국 스피치워orks(SpeechWorks) 제품을 사용하여 증권 매매까지도 가능한 서비스가 출현하였다. 또한 항공기, 기차표 예매에도 음성인식 기술이 사용되고 있다.

본 절에서는 통신사업자인 한국통신이 자체 연구를 통해 개발한 서비스들에 대하여 소개하고자 한다(5). 적용되는 기술의 상세한 설명보다는 현재 종합 통신사업자인 한국통신에서 음성언어 연구의 결과들을 적용한 응용사례를 살펴봄으로써 국내의 음성 언어 연구의 맥을 짚어보고자 한다.

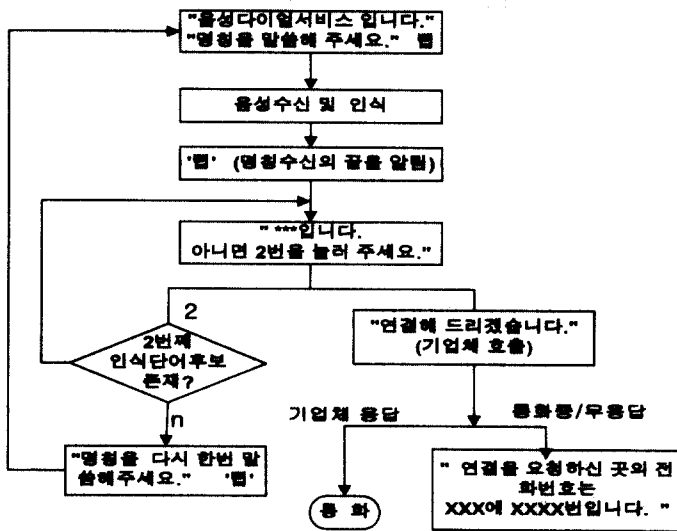
현재 한국통신에서 자체 개발한 음성인식 엔진을 이용하여 서비스중인 것으로 DSP(Digital Signal Processor)를 이용하는 기업체음성 다이얼링(1577), 음성인식 증권정보 서비스(700-3399), 그리고 음성인식 열차예매 서비스(700-1188)가 있다(6)(7). 그리고 최근에는Windows NT 상에서 구동되는 음성인식 무인 자동교환서비스도 개발한 바 있다(8). 다음에는 현재 운영되고 있는 음성인식 서비스 중 일부에 대하여 자세히 기술한다.

##### 1.1. 기업체 음성 다이얼 서비스

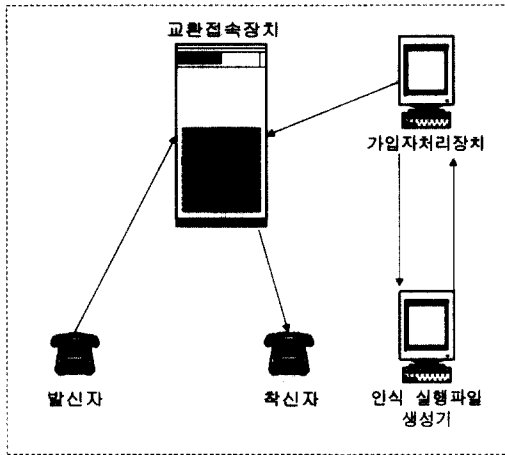
기업체 음성 다이얼링 서비스는 기업체의 이름이나 상품명을 음성으로 입력하면 해당되는 전화번호로 전화를 걸어주는 서비스이며 만일 해당업체가 통화중일 경우 전화번호를 안내해 준다(6). 이것의 기본 서비스 시나리오는 <그림 2>와 같으며 <그림 3>은 시스템 구성도이다. 가입자가 전화망을 통하여 서비스 시스템의 접속번호(1577(서울), 02-1577(지방))를 누르면 시스템에 연결되고 시스템은 통화하고자 하는 기업체의 명칭을 음성으로 입력할 것을 요구한다. 가입자가 안내방송에 따라 인식명칭을 입력하면 시스템은 음성인식 후 인식결과

에 대해 가입자에게 확인을 요구하고 맞으면 기업체 데이터베이스에서 착신전화번호를 추출하여 자동으로 다이얼링하여 연결을 시켜준다. 이때 착신 기업체가 통화 중 또는 부재중이어서 통화가 완료되지 못하면 해당 전화번호를 음성으로 안내하여 준다. 현재 서비스가 가능한 인식대상 회사명은 웹 사이트(<http://voice.kotel.co.kr>)에 저장되어 있다.

본 서비스는 한국통신에서 상용화 진행중인 지능망(AIN-IP)의 서비스로서 추가될 예정에 있으며, 시스템은 호 처리를 위해 교환기를 기반으로 하였으며 음성인식 기능을 수행하는 하드웨어를 설계하여 구현하였다.



〈그림 2〉 기업체 음성다이얼 서비스 기본 시나리오

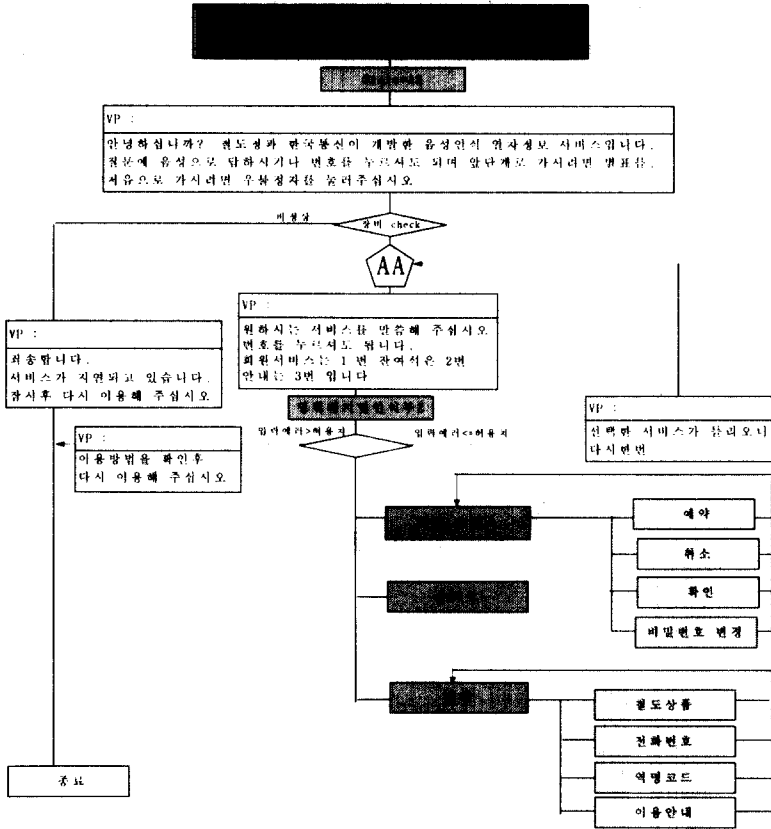


〈그림 3〉 기업체 음성다이얼 서비스 시스템 구성도

## 1.2. 음성인식 철도 안내서비스

본 서비스는 철도청 열차 예매 시스템과의 연동을 통하여 잔여석 정보 등을 실시간으로 제공하며, 철도회원의 경우에는 예매까지 가능한 서비스이다[8]. 이 서비스는 전국 동일 전화번호 700-1188로 접속되며 전자식 버튼 혹은 음성을 인식하여 철도정보를 제공하여 준다. 〈그림 4〉는 음성인식 철도정보 안내 시스템 초기 주 메뉴를 나타내고 있다. 현재 음성을 입력할 수 있는 단계는 서비스 메뉴, 열차 역명, 열차 종류 명(새마을호, 무궁화호, 통일호), 객실종류 등이다. 인식 단어 수는 총 580개의 명칭이다. 현재 회원 카드번호를 음성으로 인식하는 연구를 수행 중에 있다. 이 서비스는 전국 어디서나 정보 이용료 없이 시내전화 요금으로만 제공하여 주고 있다.

### 음성인식 ARS 시나리오



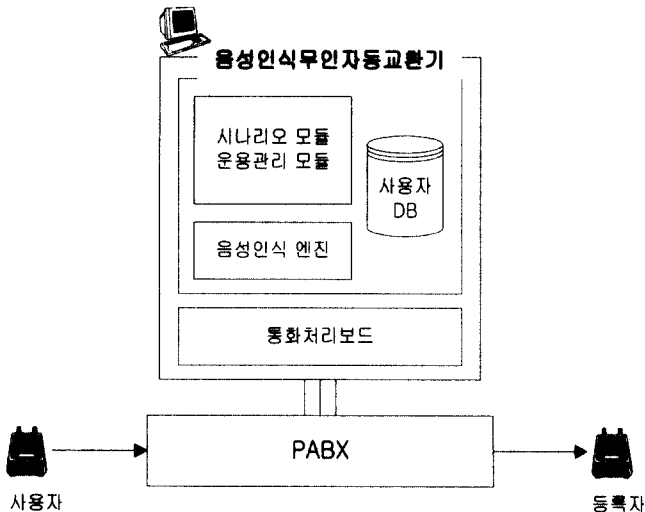
〈그림 4〉 음성인식 철도안내 서비스 초기 주 메뉴

### 1.3. 음성인식 무인자동 교환 서비스

음성인식 무인자동교환서비스는 일반 사업체나, 학교 등에 설치하여 통화를

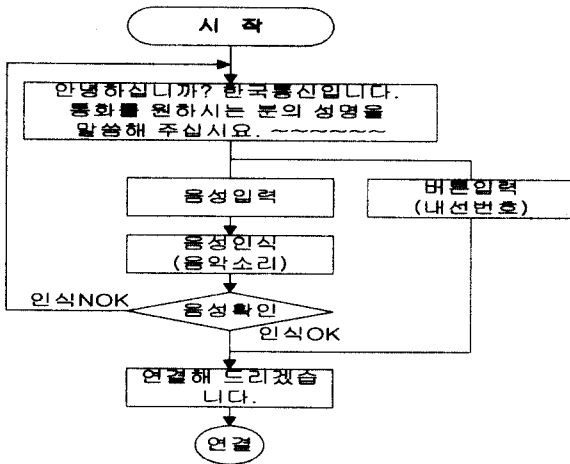
하고자 하는 사람의 이름이나 부서 명을 말하면 이를 인식하여 자동으로 연결해주는 서비스이다. 이 서비스는 전화번호를 일일이 기억할 필요 없이 통화를하고자 하는 사람의 이름이나 부서 명만으로 충분히 전화를 할 수 있으며, 안내원을 이용하는 기업체인 경우 안내원과 통화가 가능한 시간적 제약을 극복할 수 있고, 인적자원의 절약 또한 가져올 수 있다. <그림 5>는 무인자동 교환 시스템의 구성도이며, 서비스 시나리오는 <그림 6>과 같다. 본 서비스는 현재 한국통신 연구개발본부 건물내의 직원에 대해 526-5114번으로 서비스되고 있으며, 한국통신 본사는 700-5114번으로 시험 운영 중에 있다[9].

음성인식 무인자동교환시스템은 PC기반으로 구성되어 시스템 크기가 작고, 음성인식엔진을 라이브러리 형태(DLL)로 개발하였기 때문에 다양한 음성인식 시스템 개발에 응용할 수 있는 장점을 갖고 있다.



<그림 5> 무인자동 교환 시스템 구성도





〈그림 6〉 무인자동 시스템 서비스 시나리오

## 2. 컴퓨터 응용서비스

### 2.1. 음성 워드프로세서

음성을 이용하여 컴퓨터에게 명령을 내리는 음성명령기가 상용화되어 있으며 음성으로 워드프로세서를 작동할 수 있는 음성 워드프로세서 소프트웨어도 개발되어 있다. 현재 미국에서는 L&H가 Dragon 회사를 인수하여 이 분야의 시장을 80% 점유하고 있다고 한다. L&H의 음성워드 프로세서인 "Voice Xpress"는 64,000단어를 실시간 인식할 수 있다. IBM에서는 "ViaVoice"라는 소프트웨어를 출시하여 판매하고 있다. 〈표 1〉은 음성워드프로세서의 장단점을 비교하였다[10].

〈표 1〉 음성워드프로세서의 비교

	Dragon Naturally Speaking Refereed4	FreeSpeed 2000	L&H VoiceXpress Professional4	IBM ViaVoice Millennium Edition
가격	\$160	\$150	\$150	\$180
인식단어/최대단어	160K/250K	60K/670K	34K/64K	64K/2000K
Outlook, PowerPoint 등 MS s/w와 호환 여부	가능	가능	가능	가능

## 2.2. 멀티미디어 응용서비스

최근에는 음성인식 기술이 향상되고 멀티미디어 서비스가 제공됨에 따라 음성인식 기술을 이용한 멀티미디어 서비스가 선보이고 있다. 대표적인 서비스로는 음성인식 기술을 이용하여 영상을 검색하는 기술이다[11]. 예를 들면 뉴스의 동화상이 인터넷에 있을 경우 뉴스 내용 중 특정단어가 있는 부분의 뉴스 검색에 음성인식 기술이 사용된다. 즉 뉴스 아나운서의 특정단어 음성을 인식하여 그 부분을 검색하여 동영상을 보여 준다. 현재 이 분야의 연구만을 전문으로 하여 사업하고 있는 업체도 있다.

또 다른 응용 서비스로 음성 요약(gisiting) 서비스가 있다. 이것은 음성을 요약하여 정리 해주는 서비스이다. 예를 들면 5분 뉴스를 요약 정리하여 1분 뉴스로 만들어 주는 서비스이다. 현재는 중요한 단어, 혹은 문장을 찾아내는 방식으로 진행하지만 미래에는 요약 정리된 새로운 문장을 만들 수 있을 것이다.

## IV. 음성언어학 측면의 연구과제

### 1. 음성의 기본단위

음성의 기본 단위는 음소라고 알려져 있다. 하지만 한국어 음성인식 기술에

서 기본 단위로 음소를 사용하는지에 대한 검증이 제대로 이루어져 있지 않다. 영어인 경우에는 인식 엔진을 개발하는 연구소마다 사용하는 기본단위는 거의 표준으로 되어 있지만 한국어인 경우에는 체계적인 연구가 되어 있지 않다. 어떤 학자에 따르면 한국어도 중국어와 같이 음절 언어이므로 음절이 기본 단위가 되어야 한다는 의견이 제시되고 있지만 음성 공학자들 사이엔 음소를 기본 단위로 사용하는 경우가 더 많은 것 같다.

음소 혹은 음절을 기본단위로 사용한다 하더라도 한국어의 고유 특징중의 하나인 종성자음의 음성학적 특징연구가 실용화 수준까지 다다르지 못한 것 같다. 예를 들면 “우”와 “욱” 같이 “ㄱ”이 종성에 위치하면 폐쇄자음(unreleased stop)이 되는데 이것에 대한 연구가 부족해서 음성인식 시스템에 적용시 오인식의 원인이 되곤 한다.

다음으로는 기본 단위의 사용 개수이다. 음소를 기본 단위로 하였을 경우 적당한 음소 개수에 대한 의견도 통일되어 있지 못하다. 특히 문제가 되고 있는 음소는 이중모음과 “에”, “애”의 구분 등이다. 더구나 이음현상을 고려해야 할 경우에 더욱 고려해야 할 점이 많다.

## 2. 발음사전의 기본단위

음성인식 기술에서 발음사전이란 인식기의 기본 인식 단위를 말하는 것이면 또한 음성 인식기의 문법의 기본 단위가 되기도 한다. 음성 인식기의 성능 및 인식률에 중요한 역할을 한다. 한국어인 경우에 기본 단위를 “단어”, “어절” 및 “형태소” 중 어느 것을 선택하는 것이 적절한가에 대한 체계적인 연구가 진행되어 있지 않다. 영어인 경우에는 단어를 기본단위로 사용하고 있으며 일본어인 경우엔 어절을 사용하고 있다. 현재 한국어인 경우에는 단어 혹은 어절을 주로 사용하고 있으며 형태소를 변형한 유사형태소를 기본단위로 사용하는 경우도 있다. 그렇지만 한국어의 특성과 음성 인식기의 성능을 고려한 최적 기본단위에 대한 공통적인 결과가 도출되는 것이 바람직하다.

### 3. 의미정보를 위한 기본단위

음성 인식기에서 의미정보를 위한 기본 단위란 한국어의 문장에서 의미전달에 필수적으로 존재해야 되는 단위를 말한다. 음성을 이해하는데 있어서 모든 것을 인식하는 것보다 의미적인 기본단위로 인식하는 것이 더 쉽기 때문에 이에 대한 연구가 필요하다.

### 4. 대화체 문법연구

대화체는 문어체에 비해서 문법이 제대로 지켜지지 않으며 의미 없는 단어의 반복, 혹은 동일한 문장의 반복 및 생략 등 여러 특징을 가지고 있다. 이러한 특징을 잘 표현할 수 있는 문법에 대한 연구가 진행되어야 한다. 현재는 확률적 통계방식에 의한 바이그램(bigram) 혹은 트라이그램(trigram)이 사용되고 있다.

## V. 결론

본 고에서는 음성인식 기술의 정의와 한국통신이 개발한 음성인식 응용서비스에 대해서 소개하였다. 또한 음성언어학 측면에서 심층적으로 연구되어야 할 연구과제를 소개하였다. 최근의 음성인식 기술의 발달로 음성인식 응용서비스가 시장에 출시되고 있으나 아직까지는 음성인식 성능의 한계 때문에 폭발적인 확산에 장애가 되고 있다. 이러한 원인중의 하나가 한국어에 대한 심층적인 연구부족에 기인한다고 할 수 있겠다. 현재 음성인식 기술의 성능향상에 도움을 줄 수 있는 대표적인 연구과제를 소개하였다. 이러한 연구는 단기적으론 음성 인식기의 성능 향상에 도움을 줄 것이고 장기적으론 음성인식의 비전을 푸는 열쇠가 될 것이다.

최근에는 음성인식 기술의 중요성을 정부도 공감하면서 정보통신부, 산업자원부가 음성처리 기술에 과감한 투자를 하고 있어 음성언어 관련 연구자들이 고무되어 있다. 반면 한국시장의 잠재성 때문에 최근에는 세계 굴지의 음성인식기술 업체가 한국에 진출하고 있다. 한국인의 자존심인 "우리 말"에 관련된 기술을

우리 손으로 개발하고 장기적으로 외국어 음성인식기술 까지 개발하기 위해선 다양한 분야 연구자들의 공동연구가 한국어 음성인식기술의 생존에 필요조건이 되고 있다.

## 참고문헌

- (1) 구명완. 1993. "음성인식기술의 현황과 전망." 「대한전자공학회 학회지」 제20권 제5호.
- (2) 양재우. 2001. "음성정보처리 기술동향 및 전망." 음성정보처리 산업협의회 창립총회 및 세미나.
- (3) C.H. Lee. 2000. "Voice user interface(VUI) for new services over internet and wireless communications." Proc. of 9-th Korea Telecom International Symposium.
- (4) S. Yamamoto. 2000. "Corpus-based spoken language translation technologies." *Korea Telecom Journal*.
- (5) 박용기. 2000. "한국통신의 음성언어 연구현황." 「제9회 한국음성과학회 학술발표대회 논문집」.
- (6) 류창선 · 전호현 · 김재인. 1999. "한국통신 음성인식 전화정보 시스템의 음성 인식기 운용 현황." *COMSW99*.
- (7) 전호현 · 장경애 · 류창선. 1999. "대용량 음성인식 증권정보 시스템의 개발 및 상용 서비스 운용 현황." 차세대 통신소프트웨어 학술대회 (NCS'99).
- (8) 류창선 · 전호현 · 구명완. 2000. "음성인식 철도정보안내서비스의 시범 운용." 한국음향학회 제17회 음성통신 및 신호처리 학술대회.
- (9) 김희경 · 김문식. 2000. "음성인식 무인 자동교환서비스의 시범운용." 한국음향학회 제17회 음성통신 및 신호처리 학술대회.
- (10) PC Magazine. 1999.
- (11) Y. Wang. et. al. 2000. "Multimedia content analysis." *IEEE Signal Processing magazine*.

14 한국어학 13

구명완(Koo, Myoung-wan)  
한국통신 멀티미디어연구소  
137-792 서울시 서초구 우면동 17번지  
전화 : 02-525-5090  
전자우편: mwkoo@kt.co.kr

접수일 : 2001. 4. 1.  
게재결정일: 2001. 5. 26.