

음성 언어 정보 처리 연구의 동향

한국과학기술원 오영환*

1. 서론

음성은 인간이 사용하는 가장 기본적인 의사소통을 위한 수단이며, 편리함과 경제성의 측면에서 다른 방법에 비해서 우수한 특성을 갖는다. 그동안 음성과 관련된 연구들은 각기 개별적으로 이루어져 왔으며, 그러한 분야들은 언어학, 음성학, 음운학, 생리학, 해부학 등 다양한 학문적인 배경 하에서 진행되어 왔다. 그러한 결과들이 신호처리기술, 기억 장치의 대용량화, 그리고 고속의 정보 처리 기술의 발달 등 급격한 기술의 발전으로 인해서 단순히 실험적인 결과가 아닌 실용적인 측면에서 그러한 결과들을 활용하는 연구가 활발히 진행되어 왔으며, 한편으로는 계산 이론적인 측면에서 음성 처리와 관련된 다양한 연구들이 이루어지게 되었다.

본 고에서는 컴퓨터와 인간간의 의사소통을 위한 차세대 인터페이스로 떠오르고 있는 음성 관련 기술에 대해서 현재의 기술 수준 및 발전 동향, 그리고 분야별 과제 및 실제 수행되고 있는 연구들을 중심으로 살펴보고자 한다. 음성 관련 기술로는 신호처리기술, 음성인식, 음성합성, 음성부호화, 대화처리, 음성변환 등 실로 다양한 분야를 들 수 있으나, 여기서는 음성인식, 합성 및 부호화 기술을 중심으로 현재 연구상황과 관련된 대표적인 기술 및 응용 예를 중심으로 기술한다.

2. 음성인식

음성인식 기술은 컴퓨터를 비롯한 각종 기기에 명령어 입력, 받아쓰기, 자동 통역 등 여러 분야에서 응용 될 수 있기 때문에 많은 사람들의 관심의 대상이 되고 있다. 1970년대 음성인식에 대한 연구가 본격적으로 시작된 이래, 음성인식 형태는 구분 발성된 단어를 인식하는 고립단어 인식과 연속 발성 음성을 인식하는 연속음성인식, 더 나아가 대화체 음성을 대상으로 하는 음성 이해의 형태로 발전하여 왔다. 이 장에서는 최근의 음성인식의 연구 결과들과 동향에 대해 간략히 알아보려고 한다.

2.1 연속 음성인식

미국은 DARPA(Defence Advanced Research Project Agency)의 주도로 1970년대부터 음성이해 연구를 수행해 왔으며 현재는 회화체 연속 음성과 잡음 음성인식에 대한 연구 결과들을 발표하고 있다. 1996년 실시한 DARPA의 Broadcast News 자료에 대한 성능 평가(표 1)에는 BBN Systems and Technology (BBN), Carnegie Mellon대학(CMU), Cambridge 대학(CU-CON, CU-HTK), IBM Watson Lab.(IBM), 프랑스의 LIMSI 그룹, New York 대학(NYU), SRI International(SRI), Rutgers 대학(RU1, RU2) 등이 참가하였다. Broadcast News 테스트에서는 5만 단어 이상의 대어회 연속 음성인식에 대한 실험뿐 아니라 대화체 음성, 배경 잡음 및 음질이 저하된 음성 등에 대한 다양한 평가 실험을 수행하였다(표 2). 이 실험의 결과는 아직 많은 개선의 여지가 남아 있으나, 그동안 실제 환경과 관련된 연속 음성인식 기술이 크게 발전해 왔음을

*종신회원

표 1 DARPA CSR 1996 Broadcast News Hub-4 Benchmark Test(word error rate (%))

System	Complete Test	F10	F1	F2	F3	F4	F5	FX
BBN	30.2	21.6	29.5	32.7	23.3	38.4	31.8	49.9
CMU	34.9	25.8	32.1	38.6	36.6	43.7	36.5	55.8
CU-CON	34.7	25.8	33.5	40.4	33.4	39.3	40.5	53.1
CU-HTK	27.5	18.7	26.5	33.1	23.6	29.1	21.7	51.0
IBM	32.2	21.6	30.4	38.9	28.0	42.2	30.8	54.2
LIMSI	27.1	20.8	26.0	27.1	20.3	33.3	27.8	46.1
NYU	33.0	26.0	32.5	32.6	34.2	38.4	31.1	48.1
RU1	56.1	43.0	51.7	74.6	50.0	81.6	54.8	72.1
RU2	53.8	42.7	51.9	72.9	50.0	59.2	54.8	71.9
SRI	33.3	26.4	33.0	31.7	34.7	38.5	34.4	48.3

표 2 실험 환경

Condition	Dialect	Mode	Fidelity	Background
Baseline Broadcast (F0)	native	Planned	High	Clean
Spontaneous Speech (F1)	native	Spontaneous	High	Clean
Reduced Bandwidth (F2)	native	(any Mode)	Med/Low	Clean
Background Music (F3)	native	(any Mode)	High	Music
Degraded Acoustics (F4)	native	(any Mode)	High	Speech/Other Noise
Nonnative Speakers (F5)	nonnative	Planned	High	Clean
All Other Combinations (FX)	--	--	--	--

보였다. 상용화되어 현재 시판되고 있는 IBM의 ViaVoice, Dragon Systems의 Naturally Speaking 같은 제품은, 2만 단어 이상의 연속 음성인식이 가능할 뿐 아니라 화자독립/화자적응 인식을 지원하고 있어 연속 음성인식 기술이 실용화 단계에 바짝 다가섰음을 보여주고 있다.

이와 더불어 최근에는 자연스러운 대화 형태의 음성을 이해한다는 관점에서 음성 이해 시스템(speech understanding system)의 연구가 활발하다. 음성 이해는 발성된 음성의 완전한 인식 대신에 발성자가 의도한 의미의 정확한 파악을 목적으로 하고 있다. 특히 대화음성 이해의 경우 자유발화음성(spontaneous speech)을 대상으로 하기 때문에 생기는 불필요한 말, 비문법적인 발성 등의 문제를 해결하기 위해서는 고차적인 언어 정보인 구문, 의미, 운율, 대화 정보 등을 처리할 수 있는 새로운 언어 해석 기술 및 대화 처리 기술 등이 필요하다. 대표적인 연구과제로는 미국 DARPA의 ATIS (Air Travel Information System) 프로젝트와

JANUS 프로젝트가 있다. 특히 JANUS 프로젝트는 다국어의 대화체 음성 번역을 목적으로 미국, 독일, 일본, 스페인, 한국 등이 참가하고 있고, 최근 영어, 독일어, 일어에 대한 시스템이 일부 시범을 보인 바 있다.

국내에서는 1980년대 일부 대학을 중심으로 음성인식에 대한 연구가 시작되어 현재 많은 대학과 연구소에서 활발한 연구가 수행되고 있다. 최근 발표된 연구 결과에 따르면 한국과학기술원에서는 PC에서 실시간으로 인식이 가능한 3,000여개의 어휘로 구성된 연속 음성인식 시스템을 개발하였으며[3], 전자통신연구원을 중심으로 수행되고 있는 자동통역 전화 시스템 개발인 CSTAR과제의 연구 결과들과 대규모 음성 데이터베이스 구축 연구 등이 진행되고 있다.

2.2 잡음에 강인한 음성인식

현재의 음성인식 기술은 주변 잡음에 대한 인식기의 성능 저하로 인하여 실생활에 이용하기에는 여러 가지 어려움이 있다. 따라서 잡음

에 의한 음성인식기의 성능 저하를 방지하기 위해 잡음에 강인한 음성인식에 관한 여러 가지 방법들이 연구되고 있다. 잡음처리 수법은, 음성인식기의 앞단에서 처리하는 방법과 음성인식 모델과 결합하여 인식단계에서 잡음을 제거하는 방법으로 나누어 볼 수 있다.

전자에 속하는 방법으로, 캡스트럼 사영척도 [4]과 SMC 방법을 [5]을 들 수 있다. 캡스트럼 사영척도는 백색잡음이 음성신호에 첨가되면 캡스트럼 벡터의 크기는 감소되지만 벡터간 각도는 변화가 적다는 현상을 이용한 잡음에 강인한 거리척도이며, SMC 방법은 인접한 두 개의 분석구간에서 구한 자기상관 수열로부터 선형계수를 구함으로써 잡음에 의한 영향을 감소시킨다. 관별분석을 이용하여 특징파라미터의 분별능력을 향상시키는 연구와 인간의 청각특성을 이용하여 잡음에 강인한 특징파라미터를 추출하는 방법들에 대한 연구도 활발하다. 스펙트럼 차감법은 가산잡음을 제거하는 가장 대표적인 방법으로 음성이 존재하지 않는 구간에서 추정된 잡음을 이용하여 잡음음성에서 잡음을 제거하는 방법이며 [6], 효과적인 잡음의 추정을 위한 비선형적 스펙트럼 차감법과 칼만 필터링과 위너 필터링을 이용한 잡음의 제거도 활발히 연구되고 있다. 이밖에도 잡음음성과 잡음이 섞이지 않은 음성사이의 관계를 신경망이나 선형회귀분석을 사용하여 추정한 후, 변환에 의해서 잡음을 제거하는 방법도 있다.

후자에 속하는 방법들은 음성모델의 특성을 잡음의 제거에 이용하므로, 보다 효과적인 잡음의 처리가 가능하나, 계산량이 증가하는 특성이 있다. 모델기반의 잡음처리 방법으로는 HMM(hidden Markov model)을 일반화하여 음성과 잡음의 두 가지 프로세스를 모델링하고, 음성과 잡음이 결합된 HMM의 상태로 확장하여 탐색공간에서 최적의 경로를 찾아 잡음음성을 인식하는 방법과 HMM의 각 상태에서 정의된 출력확률을 결정하는 평균과 분산행렬을 잡음환경에 적응시키는 방법이 있다 [7].

이밖에도 롬바드 효과라 하여 잡음환경에서 화자는 조용한 환경에서 발성할 때와는 달리 변이된 음성을 발성하므로 이에 대한 처리와, 원격지에서의 정보검색과 예약 등이 가능한 전

화를 이용하기 위해 전화선 채널잡음의 제거에 대한 연구가 활발히 수행되고 있다.

3. 음성합성

가장 발전된 형태의 음성합성기인 문서음성 변환 시스템은 입력 문장을 확실히 이해한 사람이 자연스럽게 문장을 읽었을 때 얻어지는 음성 신호와 가장 가까운 합성음을 생성하는 것을 목표로 한다 [8]. 음성합성시스템은 크게 음성 신호의 합성 방법 및 개발 방법에 따라 formant 합성기 [9], 단위음 연결 합성기 [10], 코퍼스 기반 합성기 [11], 학습 가능한 합성기 [12] 등으로 나뉠 수 있고, 이 중 코퍼스 기반 합성기와 학습 가능한 합성기가 최근에 개발된 방법으로 음성합성에 필요한 정보들을 대용량의 음성 코퍼스로부터 추출한다는 공통점을 갖고 있다. 본 장에서는 위의 합성 방법들 중 코퍼스 기반 합성기와 학습 가능한 합성기에 대해 간략히 기술하고, 국내의 음성 합성 동향에 대해 알아본다.

3.1 코퍼스 기반 합성기

코퍼스에 기반한 합성기는 기존의 음성합성 방법과 달리 신호 처리를 최소화하며, 대량의 음성 코퍼스로부터 단위 음성 신호를 추출하고 이를 연결하여 합성음을 얻는다. 일반적인 음성합성 방법은 운율 생성 모델에서 얻어지는 운율 파라미터에 따라 단위 음성 신호를 시간 축으로 신축한다거나 피치를 조절하게 된다. 하지만 코퍼스 기반 합성기는 주어진 운율 파라미터와 음소의 문맥 등을 키(key)로 하여 가장 가까운 단위 음성 신호를 코퍼스로부터 찾고 이를 신호 처리없이 그대로 연결하게 된다. 이러한 방법의 장점으로서는 합성 신호가 음성 코퍼스로부터 얻어지기 때문에 기존 합성 방법의 문제점인 신호 처리에서 발생하는 음성의 왜곡 정도를 최소화하는 것과, 또한 구현상의 용이점을 들 수 있다. 특히, 구현상의 용이점은 다언어 음성 합성기를 개발함에 있어서 언어에의 의존성을 최소화할 수 있는 장점이 있게 되며, 결과적으로 다언어 음성 합성기 개발 문제를 단위 음성 선택 문제로 바꾸어 해결

한다. 그러나, 이 방법은 대량의 코퍼스를 필요로 한다는 점과 신호 처리 모듈의 부재로 인하여, 원하는 운율이 실린 음성 신호를 얻을 수 없는 경우가 발생하는 단점이 있다.

3.2 학습 가능한 합성기

학습 가능한 합성기는 코퍼스 기반 합성기와 함께 음성 코퍼스로부터 음성 합성에 필요한 모든 정보를 얻는다는 공통점을 가진다. 그러나 학습 가능한 합성기는 코퍼스 기반 합성기에 비해 합성기 개발 방법에 더욱 자동화를 가한 것으로, 문장 집합과 그 문장들을 발성한 음성 코퍼스만 얻어지면, 합성 시스템을 구축할 수 있다는 장점을 갖고 있다.

이 방법은 우선, 음성 코퍼스를 HMM과 같은 확률 모델을 이용하여 음소 단위로 자동 분할하고, 확률 모델의 음성 신호 발생 확률 등을 이용하여 단위 음성 후보들 중 최적의 후보들을 얻게 된다. 또한 운율 모델을 이용하여 운율 템플릿을 수집하고, 음성 합성시에는 얻어진 운율 템플릿과 미리 구축된 단위 음성 신호를 연결하여 합성음을 얻게 된다. 이 방법과 코퍼스 기반 합성기와의 차이점은 코퍼스 기반 합성기는 운율 파라미터와 단위음이 결합된 상태의 음성 신호 자체를 이용하는데 반하여, 학습 가능한 합성기는 단위 음성과 운율 모델을 따로 작성하는데 있다. 그러나, 학습 가능한 합성기를 구현하기 위해서는 우선적으로 HMM과 같은 확률 모델과 이를 이용한 높은 정확률의 자동 음소 분할 프로그램을 필요로 하는 단점이 있다. 이 방법을 이용하는 합성기로는 마이크로소프트에서 개발한 Whistler[12]가 있다.

한편, HMM을 합성기 개발에 이용하는 방법에 있어서 HMM의 모델 파라미터를 합성에 그대로 이용하는 HMM 기반 합성기[13]도 최근에 연구되고 있다. HMM 기반 합성기는 HMM을 최대 우도 방법에 기반하여 학습한다는 점을 이용하여, 합성시에는 모델에서 가장 높은 확률로 발생하는 모델 파라미터를 얻은 후, 이를 이용하여 합성음을 생성하게 된다.

3.3 국내의 음성 합성 연구

국내에서 개발된 대표적인 합성기로는 LG

전자기술원에서 개발한 LG-TTS[14], 삼성 종합기술원에서 개발한 MagicVoice[15], 한국과학기술원에서 개발한 KAISTalk[16] 등을 들 수 있다.

LG 전자기술원에서 개발한 LG-TTS는 합성 단위음으로 수정된 음절 7,270개를 사용하며, 합성 방식으로는 성도의 스펙트럼 포락선(PSE; Power Spectrum Envelope)의 시간 영역에서의 임펄스 응답을 이용한다. 운율 생성을 위해서는 총 33개의 품사군으로 어절을 분석하고, 관측 어절, 이전 어절, 다음 어절을 특징 파라미터화하여 피치 패턴을 얻고, 휴지구간의 결정은 어절의 분석 결과에 따른 휴지구간 테이블을 참조한다. 음의 지속 시간 결정 방법은 어절 단위, 음절 단위, 음소 단위별로 지속 시간 변화 파라미터를 이용하는데, 예를 들어, 어절에 존재하는 음절의 개수가 커질수록 표준 음소 길이보다 더욱 짧아지도록 하고, 각 음소별로 좌우 음소의 영향을 고려한다.

삼성종합기술원의 MagicVoice는 전문 성우가 발성한 약 2,000문장의 음성 코퍼스로부터 운율 모델을 학습시켰다. 휴지구간의 결정은 문장의 의존 트리로부터 얻어지는 구문 깊이와 사람의 음성 발화에 필요한 폐활량 등을 이용하고, 음소의 지속 시간과 에너지 등은 선형 회귀 모델을 이용한다. 피치 패턴은 발화구의 유형, 발화구의 마지막 단어의 품사에 따라 결정하게 된다.

한국과학기술원에서 개발한 KAISTalk는 총 3,300개의 VCV(Vowel-Consonant-Vowel) 체인을 기본 합성 단위로 하며, PSOLA 방법을 이용하여 신호를 합성한다. 언어처리를 위해 통계적 품사 태깅 방법과 통계적 구문 분석 방법을 이용하며, 운율 생성을 위해서는 트리에 기반한 패턴 인식 방법을 이용한다. 휴지구간의 결정은 결정 트리를 이용하고, 휴지구간 및 음소의 지속 시간 결정은 회귀 트리를 이용한다. 한편, 피치 패턴의 결정을 위해 의존 변수를 벡터로 확장한 회귀트리를 이용한다.

4. 음성부호화

음성에 관한 연구는 음성을 전송 또는 저장

하거나 음성 정보를 가공하고 음성 자체를 생성하는 일로 나눌 수 있으며, 음성 신호는 보통 디지털로 표현된다. 디지털 신호는 아날로그 신호에 비해서 처리하기가 용이하며 상대적으로 왜곡에 강하다고 알려져 있다. 디지털 신호의 중요한 속성 중 하나는 전송률이다. 전송률은 1초 분량의 음성 신호를 표현하는데 얼마만큼의 비트가 필요한지를 나타내는 값이다. 디지털 음성 신호를 전송하는 경우 전송선의 대역폭은 전송률에 의해서 결정된다. 디지털 음성 신호를 저장하는 경우에도 마찬가지이다. 음성부호화는 디지털화된 음성 신호의 전송률을 줄이는 기법을 의미하며, 음성부호화기는 부호기와 복호기로 구성된다. 부호기는 디지털 음성 신호를 입력으로 받아서 저전송률의 비트열을 생성하며, 복호기는 이러한 비트열을 입력으로 받아서 원음 음성 신호에 가까운 음성 신호를 생성해 낸다.

일반적인 아날로그/디지털(A/D) 변환기는 대부분 PCM(pulse-code-modulated) 신호를 출력한다. 전화선을 통해 전송되는 음성(300-3400Hz 대역폭)의 음질을 제대로 표현하기 위해서는 8kHz 샘플링에 한 샘플당 16bit를 사용한 선형 PCM이 필요하다고 알려져 있다. 이 경우 전송률은 128kb/s가 된다. 1980년대 이전에는 음성부호화 방법의 비효율성 및 구현상의 어려움으로 인해 음성부호화기가 거의 사용되지 않았으나, 디지털 신호를 처리하는 하드웨어와 음성부호화 방법의 급격한 발전으로 최근에는 많은 분야에서 사용되고 있다.

4.1 음성부호화 기술

1980년대에 들어서는 하드웨어 기술의 발전으로 A/D 변환기와 D/A 변환기가 실제의 음성 부호화 시스템과는 분리되었다. 이러한 시스템의 경우 음성 신호는 먼저 128kb/s PCM 신호로 변환되며 이 신호는 음성부호화 알고리즘에 의해서 더 낮은 전송률을 갖는 비트열로 다시 변환된다. 이러한 방식을 취한 첫번째 응용 프로그램으로는 아날로그 방식의 전화 통신에 보안성을 추가한 시스템이 있다. 이 시스템에서는 먼저 음성 신호를 디지털 비트열로 변환하고 암호화한 후 모뎀을 통해서 전송하게

된다. 그 당시 모뎀 기술을 고려하여 음성부호화기로 2.4kb/s에서의 미 연방 표준 부호화기였던 LPC-10e 부호화기[17]가 사용되었다. 이러한 보안 통신 시스템은 국방 등에서 매우 중요한 위치를 차지하고 있었기 때문에 낮은 합성음 음질과 비싼 비용에도 불구하고 많이 사용되었다. LPC-10e 표준 부호화기는 산업 표준 부호화기로 지정됨으로써 통신 장비간의 호환성을 보장하게 되었다. LPC-10e가 제안된 지 10년간, 하드웨어의 가격대 성능비가 비약적으로 상승했다. 음성부호화기는 하나의 칩(DSP(digital signal processing) chip)에서 쉽게 구현될 수 있게 되었다. 결과적으로 음성부호화 기술은 많은 통신 시스템에서 중추적인 역할을 하게 되었으며 이에 대한 많은 연구가 이루어지게 되었다.

4.2 음성부호화 기술의 전망

1980년대 이후부터 음성부호화 분야에서는 놀랄만한 기술의 발전이 있었다. 특히, 음성 신호의 구조를 더 잘 이해할 수 있게 되었을 뿐만 아니라 인간의 청각 시스템에 대해서도 많은 연구 결과가 발표되었다. 고성능 양자화 기법이 제안되었으며 신호 처리용 하드웨어의 성능도 놀랍게 향상되었다. 이러한 경향은 앞으로도 계속될 것으로 전망되며, 이에 따라 여러 음성부호화 연구자들 그룹의 관심사도 점차 변화할 것으로 보인다. 1985년 이후, 선형 분석(linear prediction)에 기반한 분석에 의한 합성 방법(analysis-by-synthesis)을 사용한 LPAS(Linear Prediction Analysis-by-Synthesis) 방식의 부호화기들이 음성부호화 방법의 주류를 이루어왔다. 이러한 방식의 대표적인 부호화기로 MP-LPC(MultiPulse LPC) 부호화기[18], DoD-CELP(DoD Codebook Excited LP) 부호화기[19], VSELP(Vector Sum Excited LP) 부호화기[20]를 들 수 있다. 현재 DoD-CELP 부호화기는 미 국방성의 4.8kbps 음성부호화 표준안으로, VSELP는 유럽 지역 TDMA 무선 이동 통신 표준안으로 각각 사용되고 있다. 앞으로도 향후 몇 년간은 LPAS는 4kb/s~16kb/s 사이의 부호화기에서 가장 많이 사용되는 방법으로 남을 것으로 보인다. 하

지만 4kb/s 이하의 전송률에서는 다른 방법들에 비해서 LPAS가 갖는 장점이 거의 없다. 이러한 사실은 2.4kb/s의 전송률을 갖는 부호화기에서는 LPAS를 거의 사용하지 않는다는 데서 확인할 수 있다[21]. 음성부호화기의 전송률을 더욱 줄이기 위해서는 LPAS가 아닌 다른 방법이 새로 제안되어야 할 것으로 보인다. 이러한 목적으로 최근에 제안된 부호화 방법으로 MBE(Multi-Band Excitation) 부호화기[22], WI(Waveform Interpolation) 부호화기[23] 등이 있다. 최근에는 새로운 음성 부호화 방법에 대한 연구뿐만 아니라, 잡음 환경 하에서의 음성부호화 방법, 가변 전송률을 갖는 부호화 방법, 1kbps 이하의 극저전송률 부호화기 등에 대한 연구가 활발하게 이루어지고 있다.

5. 결 론

본 고에서는 음성 정보 처리의 현황에 대해 음성인식, 음성합성, 음성부호화 분야를 중심으로 분야별 이슈 및 실제적으로 수행되고 있는 연구들을 기술하였다. 음성인식 기술은 현재 연속음성인식이 가능한 제품의 상용화가 활발히 이루어지고 있으며, 미국, 일본 및 유럽국가들이 국가주도로 자동 통역 등 실제 환경에 근접한 연구가 매우 활발하게 이루어지고 있고 국내에서도 자동통역 전화 연구를 위해 외국의 다른 연구 그룹들과 공동 연구가 진행중에 있다. 음성합성은 외국의 경우 코퍼스 기반, 학습 가능한 합성기를 이용한 다국어 합성기에 대한 연구가 활발하며 국내의 경우 최근 비교적 양질의 합성기들이 발표되어 다양한 응용분야에서 사용되고 있는 추세다. 음성부호화 분야에서는 무선 전화, 인터넷 폰 등에서의 수요 증가로 인하여 양질의 부호화기와 저전송률 부호화기를 중심으로 지속적인 연구가 수행되고 있다. 이제 음성 정보 처리 기술은 차세대 컴퓨터와 인간간의 의사 소통을 위한 인터페이스로서 더 이상 연구 분야에 머물러 있지 않고 일상 생활에 깊숙이 파고들고 있는 분야이다.

보다 많은 연구자들에 의한 국어 음성 연구가 유기적으로 이루어지고, 선진 외국의 경우

처럼 이를 주도할 범국가적인 추진 주체가 조속히 구성되고 국가 지원 하에 장기 연구 프로젝트가 가시화되기를 기대한다.

참고문헌

- [1] David S. Pallett, Jonathan G. Fiscus, Mark A. Przybochi, "1996 Preliminary Broadcast News Benchmark Tests," 1997 DARPA Speech Recognition Workshop proceeding.
- [2] John S. Garofolo, Jonathan G. Fiscus, William M. Fisher, "Design and Preparation of the 1996 Hub-4 Broadcast News Benchmark Test Corpora," 1997 DARPA Speech Recognition Workshop proceeding.
- [3] Seong Jin Yun, Yung Hwan Oh, Gyung Chul Shin, "Improved Lexicon Modeling for Continuous Speech Recognition," International Conference on Acoustics, Speech, and Signal Processing, pp. 1, 827~1,830, Munich, Germany, Apr. 1997.
- [4] D. Mansour and B. J. Juang, "A family of distortion measures based upon projection operation for robust speech recognition," IEEE Trans. on ASSP, Vol. 37, No. 11, pp. 1,659~1,671, 1989.
- [5] D. Mansour and B. J. Juang, "The short-time modified coherence representation and its application for noisy speech recognition," IEEE Trans. on ASSP, Vol. 37, No. 6, pp. 795~804, 1989.
- [6] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. on ASSP, Vol. 27, No. 2, pp. 113~120, 1979.
- [7] M. J. F. Gales, "Model-based techniques for noise robust speech recognition," Ph. d. thesis, Cambridge university, september, 1995.

- [8] Mark Y. Liberman and Kenneth W. Church, "Text Analysis and Word Pronunciation in Text-to-Speech Synthesis," *Advances in Speech Signal Processing*, edited by S. Furui and M.M. Sondhi, Eds., Marcel Dekker, New York, 1992.
- [9] Allen J., Hunnicutt S. and Klatt D., *From text to speech: the MITalk system*, MIT Press, Cambridge, Massachusetts, 1987.
- [10] Sagisaka Y., Kaiki N., Iwahashi N. and Mimura K., "ATR v-Talk speech synthesis system," *ICSLP, Banff, Canada*, 1992, pp. 483~486.
- [11] Andrew J. Hunt and Alan W. Black, "Unit Selection in a Concatenative Speech Synthesis System using a Large Speech Database," *ICASSP*, 1996. pp. 373~376.
- [12] Huang X., Acero A., Hong H., Ju Y., Liu J., Meredith S. and Plumpe M., "Recent Improvement on Microsoft's Trainable Text-to-Speech System-Whistler," *ICASSP, Munich*, 1997, pp. 959~962.
- [13] Tokuda K., Kobayashi T. and Imai S., "Speech Parameter Generation from HMM using Dynamic Features," *ICASSP, Detroit*, 1995, pp. 660~663.
- [14] 이준우, 김세린, 김상수, 이종석, 김민성, "수정된 음절을 이용한 한국어 문장-음성 변환 시스템", 제13회 음성통신 및 신호처리 워크샵, 1996, pp. 237~240.
- [15] 김정수, 이혜정, "언어정보 및 통계 데이터 활용을 이용한 한국어 운율 생성", 제 13회 음성통신 및 신호처리 워크샵, 1996, pp. 227~231.
- [16] 오영환, 김연준, 이상호, 변효진, 김은경, "KAISTalk-I", *지능기술 엑스포 '97 발표자료집*, 1997, pp. 3~12.
- [17] T. E. Tremain, "The government standard linear predictive coding algorithm," *Speech Technology*, pp. 40~49, April 1982.
- [18] P. Kroon, E. Deprettere, and R. Sluyter, "Regular-pulse excitation: A novel approach to effective and efficient multipulse coding of speech," *IEEE Trans. on ASSP*, October 1986, 1,054~1,063.
- [19] M. Schroeder and B. Atal, "Code excited linear prediction: High quality speech at low bit rates," *Proc. of ICASSP*, pp. 937~940, 1985.
- [20] I. Gerson and M. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 4.8kbps," *Proc. of Int. Mobile Satellite Conf.*, pp. 678~683, Ottawa, 1990.
- [21] M. A. Kohler and L. M. Supplee, "Progress towards a new government standard 2400 bps voice coder," in *Proc. Int. Conf. Acoust. Speech Sign. Process.*, (Detroit), pp. 448~451, IEEE, 1995.
- [22] D. W. Griffin and J. S. Lim, "Multiband excitation vocoder," *IEEE Trans. Acoust. Speech Sign. Process.*, vol. ASSP-36, no. 8, pp. 1,223~1,235, 1988.
- [23] W. B. Kleijn, "Continuous representations in linear predictive coding," in *Proc. Int. Conf. Acoust. Speech Sign. Process.*, (Toronto), pp. 201~204, 1991.

오 영 환



1972 서울대학교 공과대학 전자공학과
 1974 서울대학교 교육대학원 공업교육학과(석사)
 1980 Tokyo Institute of Technology 정보공학전공(박사)
 1981 충북대학교 공과대학 전산학과 조교수
 1983~1984 University of California(Davis) 연구교수

1985~현재 한국과학기술원 전산학과 교수로 재직중
 관심분야: 음성인식, 음성합성, 음성코딩, 화자인식, 대화판리, 신경회로망, 전문가 시스템
