
남성과 여성의 음성 특징 비교 및 성별 음성인식에 의한 인식 성능의 향상

이창영*

Comparison of Male/Female Speech Features and Improvement of Recognition Performance by Gender-Specific Speech Recognition

Chang-young Lee*

요 약

음성인식에서의 인식을 향상을 위한 노력의 일환으로서, 본 논문에서는 성별을 구분하지 않는 일반적 화자 독립 음성인식과 성별에 따른 음성인식의 성능을 비교하는 연구를 수행하였다. 실험을 위해 남녀 각 20명의 화자로 하여금 각각 300단어를 발성하게 하고, 그 음성 데이터를 여성/남성/혼성A/혼성B의 네 그룹으로 나누었다. 우선, 성별 음성인식에 대한 근거의 타당성을 파악하기 위하여 음성 신호의 주파수 분석 및 MFCC 특징벡터들의 성별 차이를 조사하였다. 그 결과, 성별 음성인식의 동기를 뒷받침할 정도의 두드러진 성별 차이가 확인되었다. 음성인식을 수행한 결과, 성을 구분하지 않는 일반적인 화자독립의 경우에 비해 성별 음성인식에서의 오류율이 절반 이하로 떨어지는 것으로 나타났다. 이로부터, 성 인식과 성별 음성인식을 계층적으로 수행함으로써 화자독립의 인식률을 높일 수 있을 것으로 사료된다.

ABSTRACT

In an effort to improve the speech recognition rate, we investigated performance comparison between speaker-independent and gender-specific speech recognitions. For this purpose, 20 male and 20 female speakers each pronounced 300 isolated Korean words and the speeches were divided into 4 groups: female, male, and two mixed genders. To examine the validity for the gender-specific speech recognition, Fourier spectrum and MFCC feature vectors averaged over male and female speakers separately were examined. The result showed distinction between the two genders, which supports the motivation for the gender-specific speech recognition. In experiments of speech recognition rate, the error rate for the gender-specific case was shown to be less than 50% compared to that of the speaker-independent case. From the obtained results, it might be suggested that hierarchical recognition of gender and speech recognition might yield better performance over the current method of speech recognition.

키워드

음성인식, 성별 음향학적 차이, 성별 음성인식, HMM

1. 서론

정보 전달 수단으로서 음성은 매우 효과적이어서 속

런된 타자수의 기록보다 2배 정도 빠르다[1]. 사람과 기계의 인터페이스로서 음성인식을 구현하려는 노력이 1952년에 처음으로 숫자 인식이 발표된 이래[2] 오늘날

* 동서대학교 정보시스템공학부(seewhy@dongseo.ac.kr)

접수일자 : 2010. 10. 01

심사(수정)일자 : 2010. 10. 30

게재확정일자 : 2010. 12. 10

까지 간단없이 경주되어 오고 있다. 그 결과 지금은 많은 실용화가 이루어져, 이동 통신 기기와 같은 장치를 통하여 우리 생활 주변에 등장하는 단계에 이르렀다[3].

음성인식 시스템의 실용화 여부를 가름하는 어느 인식을 문턱값이 존재하는 바, 만약 인식기의 성능이 그 값 이하로 저조할 때, 사용자는 인식기를 신뢰하지 않고 ‘불편해서 사용불가’라는 판단을 내리게 된다. 반면, 인식률이 그 문턱값 이상의 성능을 발휘하면, 드물게 일어나는 인식의 오류는 사용자로 하여금 자신의 비협조적인 발성에 문제가 있었노라 생각하게 만든다. 이러한 판단은 주관적인 문제이기에 개인별 차이가 존재하겠지만, 실용과 무용의 경계가 되는 인식률의 값은 대략 92% 정도로 여겨지고 있다[4].

인식률이 어느 문턱값을 넘어 실용화 되었다면, 실제 인식률이 (예를 들어) 94%인가 또는 96%인가는 크게 중요하지 않다. 사용자들의 반응은 공히 ‘쓸모있다’로 표현되는 것이다. 그럼에도 불구하고 보다 높은 인식률을 추구해야 하는 중요한 이유가 있다.

우리의 일상생활에서는 상대방의 말에 되묻는 일이 거의 없다. 설령 대화의 일부를 명확하게 듣지 못했더라도, 전후관계를 고려하여 추론할 수 있는 능력 덕택에, 상대방이 전달하려는 정보를 파악할 수 있는 것이다. 하지만 음성을 통한 기계와의 통신에서는 “이것이 맞습니까?”의 확인(confirmation) 과정을 필요로 하는 경우가 많다. 예를 들어 상대방의 은행 계좌번호와 송금액 숫자들을 전화기에 발성하여 이체하는 응용을 지향한 연구가 시도된 바 있지만[5], 그 상용화는 아직 보편적이지 못한 실정이다. 그 이유는, 인식된 단일 후보자에 대한 완벽한 확신이 없는 한, 매 단계마다 자동응답기는 “말씀하신 숫자가 0000이 맞습니까?”라는 확인을 해야 하고, 그로 인하여 음성을 통한 빠른 정보전달의 장점이 상실되기 때문이다. 키워드 검색에서처럼 인식된 후보자를 다수로 제출하는 응용은 적지 않은 상용화가 이루어진 데 반해, 인식 결과가 단일 후보자로 제공되어야 하는 응용이 아직 미흡한 단계에 머물러 있는 것은 이러한 이유에서이다. 음성인식의 궁극적인 목적은, 사람들 사이의 대화가 그러하듯이, 확인을 위한 되물음이 개입되지 않을 정도의 완벽한 인식률로 단일 후보자를 제출하는 수준에 이르는 것이라 할 수 있다.

화자독립 음성인식은 남녀노소를 불문하고 모든 사람을 대상으로 하며, 인식의 많은 오류는 화자들의 다

양한 변화에 기인한다. 예를 들어 훈련 과정과 실제 응용 단계에서 성이나 나이 등의 큰 불일치는 인식에 좋지 않은 영향을 줄 수밖에 없다[6]. 만약 화자들을 작은 그룹들로 나누어 그룹별 인식을 수행한다면, 패턴의 다양함은 줄어들게 되고, 그로 인한 인식률 향상을 기대할 수 있을 것이다. 소그룹들의 음향학적 구분이 선명할수록 그 결과는 더욱 효과적일 것이다.

특정 그룹에 대한 음성인식을 수행하는 사례로서 Wilpon 등은 소아와 노인에 대한 음성 인식을 따로 수행해야 하는 필요성을 제기하였다[7]. 그 근거의 일례로서 남성은 15세 근처의 사춘기를 경계로 피치(Pitch) 주파수의 뚜렷한 차이를 보임을 밝히고, 인식대상을 15세 미만과 이상으로 분리하여 그룹별 인식을 수행할 필요성을 논문은 언급하고 있다. 이외에도, 화자의 나이에 따른 음향학적 정보 및 화자의 나이가 음성인식에 큰 영향을 미친다는 연구 결과들이 다수 보고된 바 있다[8][9].

사람이 발성하는 음성 신호에는 의미를 가진 언어학적 정보 이외에도 화자의 성(性, Gender)이나 출신 국가 또는 지역과 같은 비언어학적 특징이 포함되어 있다[10]. 이러한 특징들을 인식의 한 요소로서 포함시켜 인식률을 높이려는 자연스런 노력이 음성인식에 더러 접목되는데, 예를 들어 성과 억양(악센트) 정보를 MFCC와 같은 다른 특징들과 함께 묶어 구분함으로써 화자인식의 성능을 향상시키기도 한다[11].

화자들을 어떤 소그룹으로 나눌 것인가 하는 일반적인 물음에 대한 답은 쉽지 않지만, 지역과 나이를 불문하고 명확하게 나눌 수 있는 그룹이 있는 바, 바로 남녀의 성에 따른 구분이다. 굳이 주파수 분석을 동원하지 않더라도, 음성학적으로 남녀는 큰 차이를 보이며, 여성은 남성에 비해 고주파쪽 Formant를 갖고 있다는 것은 상식적인 내용이다. 그림 1은 테너와 소프라노의 음역을 보여준다[12].



그림 1. 테너와 소프라노의 음역
Fig. 1 Vocal ranges of tenor and soprano

따라서 남녀를 구분하여 성별로 각각 다른 인식을 수행한다면 화자간 변화가 줄어들게 되고, 그로 인하여 인식률의 향상을 꾀할 수 있을 것으로 기대되는 것이다. 이러한 연유에서 성별(Gender-Specific) 음성 인식이 종종 실행되고 있다[13][14].

본 연구의 목적은 두 가지이다. 하나는 남성과 여성의 음향학적-음성학적 차이를 비교하여, 성별 음성 인식에 대한 근거의 타당성을 파악하는 것이다. 다음은 성별 음성인식을 일반적인 화자독립 음성인식과 비교하여 그 효과에 대해 조사하는 것이다.

한편으로, 화자의 성 인식 또한 패턴인식의 한 분야로서 널리 연구되어 왔다[15-18]. 성의 인식과 성별 음성인식을 연계하여(hierarchically) 수행하면 화자독립 음성인식을 구현하는 것이 가능하다. 본 논문에서는 화자독립(Speaker-Independent, Cross-Gender) 음성인식과 성별(Gender-Specific, Gender-Dependent) 음성인식의 성능을 비교하는 실험을 수행하고, 성 인식과 성별 음성인식을 계층적으로 연계시켜 화자독립을 구현하는 작업은 차후의 연구에 이어질 것이다.

II. 실험

실험을 위해 남녀 각 20명으로 하여금 300 단어를 발성하게 하고, 그 음성신호를 16kHz의 주파수로 샘플링 하였다. 선택된 단어들은 한글의 50여 음소들의 고른 분포를 고려하여(Phone-Balanced) 구성되었다. 얻어진 데이터는 16bit로 양자화 시키고 신호 앞뒤의 묵음 구간을 제거한 후 실험에 사용하였다.

남 또는 여의 성별에 따라 인식을 수행하는 경우와 성별 구분 없이 수행하는 화자독립의 인식률을 비교하기 위하여 40명의 화자를 표 1과 같이 나누었다.

표 1. 화자의 그룹화
Table 1. Grouping of the Speakers

그룹 ID	남성 화자	여성 화자	총 화자
Male	20	0	20
Female	0	20	
Mixed_A	10	10	
Mixed_B	10	10	

또한, 주어진 데이터 범위 내에서 측정 및 조사의 다양화를 위해 300단어를 표 2와 같이 100, 200, 300 단어의 세 집합으로 나누었다. 세 집합 중 두 집합 W_1 과 W_2 는 공유하는 단어가 없으며 ($W_1 \cap W_2 = \phi$), 그 둘의 합집합(Union)은 집합 W_3 와 같다($W_1 \cup W_2 = W_3$).

표 2. 단어의 세 집합
Table 2. Three Sets of the Words

단어 집합 ID	어휘 수
W_1	100
W_2	200
W_3	300

특징벡터 추출은 32ms의 시간 간격에 해당하는 512개의 데이터를 한 프레임으로 하여 수행되었다. 음성신호가 급격하게 변하는 구간에서는 '동시 조음(coarticulation)'[19]이 발생하며, 이러한 정보를 잃지 않기 위해서는 프레임들을 어느 정도 중첩시켜야 한다. 본 연구에서는 인접한 두 프레임을 70% 중첩시켰다. 먼저 각 프레임에 Hanning 윈도우를 적용하여 프레임 가장자리에서의 단절적인 변화를 완화시킨 다음 스펙트럼의 고주파 성분을 강화시키는 pre-emphasis를 거쳐 13차 MFCC를 추출하였으며[20], 시스템 환경에 따른 여러 가지 악영향에 대항하는 강인함을 부여하기 위하여 프레임 베이스로 CMS(Cepstral Mean Subtraction)[21]를 적용하였다.

표 1에 주어진 각 그룹의 20명의 화자는 다시 표 3과 같이 나뉜다. 소그룹 I의 음성은 코드북 생성 및 HMM 파라미터 훈련에 사용되고, 소그룹 II의 음성은 HMM을 훈련시키면서 인식률을 테스트하는 데 쓰인다. 이 그룹에 대해 최고의 인식률을 나타내는 HMM 파라미터가 갱신되며 저장되었다가, 최종적으로 소그룹 III에 대한 인식률 조사에 사용된다.

표 3. 그룹 내 20명 화자의 구분
Table 3. Division of 20 Speakers within a Group

소그룹 ID	화자 수	역할
I	12	훈련
II	4	훈련 중 테스트
III	4	최종 테스트

표 3의 소그룹 I에 대해 Linde-Buzo-Gray 클러스터링 알고리즘을 적용하여 클러스터 수 512의 코드북을 생성하였다. 음성신호는 512개의 클러스터 센트로이드와 비교되고, 그 중 가장 가까운 두 개의 클러스터 인덱스와 멤버십 값을 계산하는 퍼지 벡터 양자화(FVQ)를 수행하였다. 이 결과는 HMM(Hidden Markov Model)에 전달되어 인식 과정이 진행된다.

음성인식을 위한 HMM으로서 후방의 상태로 전이가 금지되는 Bakis 모델을 채용하였다. 상태 수는 각 단어의 평균 프레임 수에 비례하도록 단어별로 따로 설정하였다[22]. 훈련이 진행됨에 따른 파라미터의 재평가는 Baum-Welch 공식에 의해 이루어지고, 사건 관측확률은 하한선을 설정함으로써 0이 되지 않도록 하였다.

HMM 파라미터를 반복 훈련하면서 세 가지 항목이 자동으로 모니터링되도록 하였다. 주어진 HMM 파라미터에서의 총 사건확률, 표 3의 그룹 II에 대한 인식률, 그리고 사건관측확률 파라미터들의 변화가 그것이다. 매 반복 순환 훈련이 이루어지고 나서, 그룹 II에 대한 인식률을 조사하고, 그 최고값이 경신될 때마다 HMM 파라미터를 따로 저장하였다. 훈련의 수렴이 충분하여 이들 변화가 어느 문턱값 이하가 되면 훈련을 종료시키고 표 3의 그룹 III에 대해 인식률을 조사하였다. 이 때 사용되는 HMM 파라미터는 그룹 II에 대해 최고의 인식률을 기록한 것이다. 이 처방에 의해, 인식기가 훈련에 동원되는 화자들에 지나치게 특성화되어 그 이외의 일반적인 화자에 대한 적응이 떨어지는 것을 막는 효과를 얻을 수 있다[23].

III. 결과 및 토론

그림 2는 남녀 각 20명으로부터 발생된 모음 /a/의

주파수 분석을 수행하여 남성과 여성에 대해 따로 평균값을 취한 것이다. 음향준위의 효과를 배제하기 위하여 개인별 스펙트럼의 최고값을 1로 정규화 시켰다. 예상할 수 있었던 바와 같이, 여성(Female) 그룹에 대한 스펙트럼 전체가 남성(Male) 그룹에 비해 고주파쪽으로 편향되어 있음을 볼 수 있다.

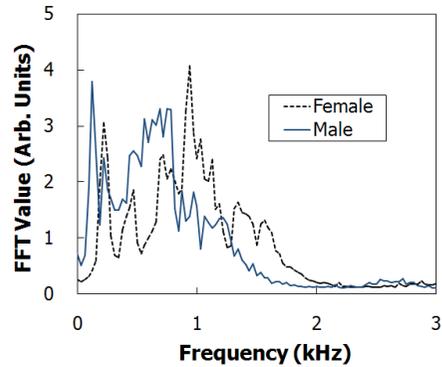


그림 2. 모음 /a/에 대한 남성과 여성의 스펙트럼 평균
Fig. 2 Spectrum of Phoneme /a/ Averaged over 20 Male and 20 Female Speakers

그림 3은 남녀 각 20명으로부터 발생된 모음 /a/, 즉 그림 2에 사용되었던 음성 토큰에 대한 Formant 주파수 분포를 나타낸다. 남성과 여성의 Formant 주파수 분포가, 최소한 본 연구에서 수행된 대상들에 대해서는 확연히 분리됨을 볼 수 있다. 남녀가 공유하는 주파수 대역은 700-800Hz 뿐이다. 이 그림으로부터 남성과 여성의 음향학적 차이가 작지 않음을 알 수 있다.

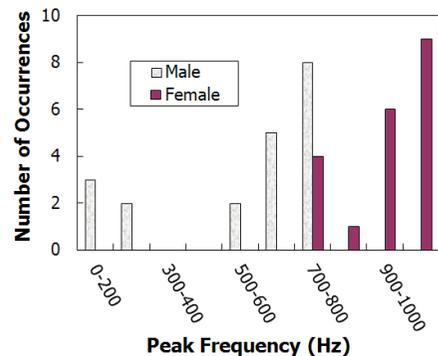


그림 3. 모음 /a/에 대한 Formant 주파수 분포
Fig. 3 Distribution of Formant Frequencies of Male and Female for Vowel /a/

그림 4는 남녀 각 20명에 대해 추출된 MFCC 특징 벡터들의 성별 평균을 보여준다. 가로축은 MFCC 차수, 세로축은 MFCC 값을 나타낸다.

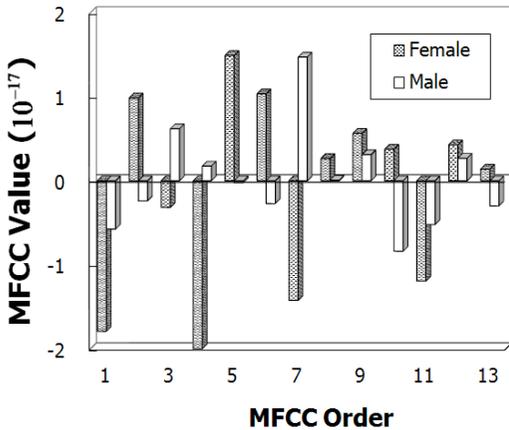


그림 4. 남녀 성별 20명 화자들에 대한 MFCC 특징벡터들의 평균

Fig. 4 MFCC Feature Vector Averaged over 20 Male and 20 Female Speakers

그림으로부터 13개의 성분들 중 7개의 성분(2, 3, 4, 6, 7, 10, 13)이 남성과 여성에 대해 극성이 반대임을 볼 수 있다. 또한 2개의 성분 5와 8도 그 구분이 비교적 뚜렷하다. 나머지 네 개의 성분(1, 9, 11, 12)만이 남성과 여성에 대해 비교적 분리가 어려운 분포를 보이고 있다. 따라서 MFCC 특징벡터들이 분포하는 13차원 공간에서 남녀를 따로 분리하여 패턴을 분리하는 것이 모든 대상을 한꺼번에 다루는 것보다 훨씬 용이하리라는 것이 기대된다.

표 4는 표 1에 주어진 네 그룹에 대한 인식 오류율을 나타낸다. 세 어휘 경우 모두 성별 (Male 또는 Female) 인식에서의 오류율의 값이 혼성(Mixed_A와 Mixed_B)을 대상으로 경우에 비해 절반 정도 또는 그 이하로 낮아짐을 볼 수 있다.

표 4. 네 그룹에 대한 인식 오류율
Table 4. Recognition Error Rate for Four Groups

어휘 수	그룹 ID	인식 오류율(%)	평균 인식 오류율(%)
100	Female	0.5	0.5
	Male	0.5	
	Mixed_A	2	1.9
	Mixed_B	1.75	
200	Female	1.75	1.6
	Male	1.5	
	Mixed_A	4.25	3.9
	Mixed_B	3.5	
300	Female	3.08	2.6
	Male	2.17	
	Mixed_A	5.67	5.1
	Mixed_B	4.58	

그림 5는 표 4의 데이터를 그래프로 나타낸 것이다. 오른쪽 두 그룹으로 표시된 혼성의 인식 오류율에 비해 왼쪽 두 그룹으로 표시된 성별 인식 오류율이 두드러지게 작음을 확인할 수 있다.

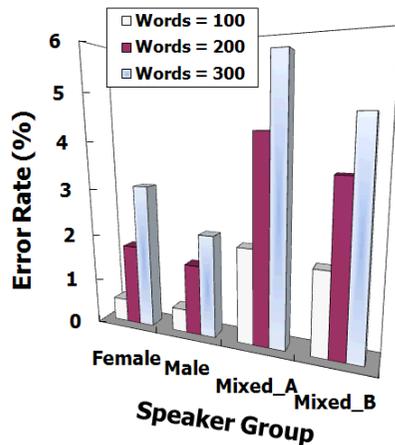


그림 5. 표 4의 데이터를 나타낸 그래프
Fig. 5 Graph for the Data Given in Table 4

성별 음성인식에서의 오류율이 혼성의 경우에 비해 절반 이하로 줄어든다는 것을 확인함으로써, 현재의 오류율을 줄일 수 있는 새로운 방법이 모색될 수 있다. 음성신호가 입력되면 우선 성을 인식한다. 이어서, 그 결과에 따라, 성별 음성인식을 수행하는 것이다. 성별 음성인식의 오류율이 50% 정도 줄어든다면, 성

의 인식 단계에서의 오류율이 90%에 불과하다 할지라도 총 오류율은 현재의 혼성에 대한 음성인식 방식에 비해 45% 정도 줄어들 것이 예상되는 것이다. 물론 이에 대해서는 실험적 검증이 이루어져야 할 것이다. 그림 6은 계층적 음성인식을 통하여 화자독립 음성인식을 수행하는 과정을 나타낸 것이다. 이의 구현에 대한 연구는 차후에 진행될 것이다.

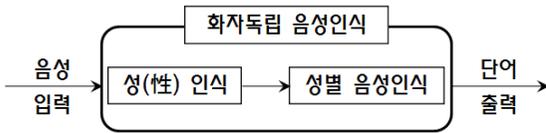


그림 6. 계층적 인식을 통해 성별 인식으로부터 화자독립 음성인식을 구현하는 과정

Fig. 6 Block Diagram for Hierarchical Implementation of Gender Recognition and Gender-Specific Speech Recognition

다른 가능성으로서, 사용자는 자신이 남성인지 여성인지 우선 음성인식 기기에 성별을 등록할 수 있다. 이후에 음성인식 시스템은 성별로 따로 훈련된 파라미터에 의해 동작한다.

IV. 결론

본 논문에서는 남녀 각 20명의 화자를 대상으로 성별(Gender-Specific) 음성인식과 성을 구분하지 않는 일반적인 화자독립(Speaker-Independent) 음성인식을 비교하는 연구를 수행하였다. 우선 그러한 음성인식에 대한 근거의 타당성을 조사하기 위하여 성별 주파수 스펙트럼을 비교하였다. 여성의 평균 스펙트럼은 남성에 비해 고주파쪽으로 편향되어 있음이 확인되었고, 남성과 여성의 Formant 주파수는 거의 확연히 분리되는 정도였다. 남성과 여성에 대한 MFCC 특징벡터들의 평균값을 분석한 결과, 13개의 성분들 중 7개 성분의 극성이 반대로 나타났다. 이는 13차원의 특징벡터 공간에서 남성과 여성의 특징벡터들이 비교적 분리가 쉽다는 것을 의미한다. HMM 인식기를 통한 음성인식 테스트 결과, 성별 음성인식의 경우 인식 오류율이 혼성의 경우에 비해 50% 이하로 낮아짐이 확인되었다. 이로부터, 성의 인식과 성별 인식을 계층적으

로 결합하여 화자독립 음성인식을 구현함으로써 인식을 향상할 수 있을 것으로 사료된다.

참고 문헌

- [1] G. Kaplan, "Words Into Action I," IEEE Spectrum, Vol. 17, pp. 22-26, 1980.
- [2] K. H. Davis, R. Biddulph, and S. Balashek, "Automatic Recognition of Spoken Digits," J. Acoust. Soc. Am., Vol. 24, No. 6, pp. 637-642, 1952.
- [3] B. H. Juang & L. R. Rabiner, "Automatic Speech Recognition - A Brief History of the Technology Development," Encyclopedia of Language and Linguistics, 2nd Ed., Elsevier, 2005.
- [4] L. Rabiner & B. Juang, "Fundamentals of Speech Recognition," Prentice Hall, New Jersey, pp. 485-486, 1993.
- [5] Z. Bo, L. Juan, P. Gang, & W. Wang, "A High Performance Mandarin Digit Recognizer," Fifth International Symposium on Signal Processing and Its Applications, Vol. 2, pp. 629-632, 1999.
- [6] O. Deshmukh, C. Y. Espy-Wilson, & A. Juneja, "Acoustic-Phonetic Speech Parameters for Speaker-Independent Speech Recognition," International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 593-596, 2002.
- [7] J. G. Wilpon & C. N. Jacobsen, "A Study of Speech Recognition for Children and the Elderly," International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 349-352, 1996.
- [8] S. Yildirim & S. S. Narayanan, "An Information-Theoretic Analysis of Developmental Changes in Speech," International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 480-483, 2003.
- [9] I. Kudo, T. Nakama, & T. Watanabe, "An Estimation of Speaker Sampling in Voice Across Japan Database," International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 825-828, 1996.
- [10] P. Dutta & A. Haubold, "Audio-Based Classification of Speaker Characteristics," 2009

- International Conference on Multimedia and Expo (ICME), pp. 422-425, 2009.
- [11] S. Deshpande, S. Chikkerur, & V. Govindaraju, "Accent Classification in Speech," Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pp. 139-143, 2005.
- [12] <http://en.wikipedia.org/wiki/File:Tenor.png> & <http://en.wikipedia.org/wiki/File:Sopran.png>.
- [13] R. Muralishankar & D. O'Shaughnessy, "A Comprehensive Analysis of Noise Robust Speech Features Extracted from All-Pass Based Warping with MFCC in a Noisy Phoneme Recognition," The Third International Conference on Digital Communications (ICDT), pp. 180-185, 2008.
- [14] I. Gavut & C. O. Dumitru, "ASR for Romanian Language," 14th International Workshop on Systems, Signals, and Image Processing (IWSSIP), pp. 300-303, 2007.
- [15] G. Tzanetakis, "Audio-Based Gender Identification Using Bootstrapping," IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp. 432-433, 2005.
- [16] H. Kim, K. Bae, & H. Yoon, "Age and Gender Classification for a Home-Robot Service," The 16th International Symposium on Robot and Human Interactive Communication, pp. 122-126, 2007.
- [17] T. Bocklet, A. Maier, J. Bauer, F. Burkhardt, & E. Noth, "Age and Gender Recognition for Telephone Applications Based on GMM Supervectors and Support Vector Machines," International Conference on Acoustics, Speech, and Signal Processing, pp. 1605-1608, 2008.
- [18] X. Zhao, D. O'Shaughnessy, & N. Minh-Quang, "A Processing Method for Pitch Smoothing Based on Autocorrelation and Cepstral F0 Detection Approaches," International Symposium on Signals, Systems, and Electronics (ISSSE), pp. 59-62, 2007.
- [19] J. R. Deller, J. G. Proakis, & J. H. L. Hansen, "Discrete-Time Processing of Speech Signals," Macmillan, New York, pp. 143-145, 1994.
- [20] J. Wang, J.-F. Wang, & Y. Weng, "Chip Design of MFCC Extraction For Speech Recognition," The VLSI Journal, vol. 32, pp. 111-131, 2002.
- [21] W. Xu, et al., "A Noise Robust Front-End Using Wiener Filter, Probability Model and CMS for ASR," International Conference on Natural Language Processing and Knowledge Engineering, pp. 102-105, 2005.
- [22] M. Dehghan, K. Faez, M. Ahmadi, and M. Shridhar, "Unconstrained Farsi Handwritten Word Recognition Using Fuzzy Vector Quantization and Hidden Markov models," Pattern Recognition Letters, vol. 22, pp. 209-214, 2001.
- [23] L. Fausett, "Fundamentals of Neural Networks," Prentice-Hall, New Jersey, p. 298, 1994.

저자 소개



이창영(Chang-young Lee)

1982년 2월 서울대학교 물리교육
학과 졸업 (이학사)

1984년 2월 한국과학기술원 물리
학과 졸업 (이학석사)

1992년 8월 뉴욕주립대학교 대학원 물리학과 졸업
(이학박사)

동서대학교 정보시스템공학부 교수

※ 관심분야 : 음성인식, 신호처리