

IPA를 활용한 다국어 음성 인식에 관한 연구

김석동^{1*}, 김우성¹, 우인성¹
¹호서대학교 컴퓨터공학과

A Study on the Multilingual Speech Recognition using International Phonetic Language

Suk-Dong Kim^{1*}, Woo-Sung Kim¹ and In-Sung Woo¹

¹Dept of Computer Science, Hoseo University

요 약 최근 다양한 모바일 기기의 사용자 환경과 다양한 음성인식 소프트웨어의 영향으로 음성인식 기술역시 빠르게 발전되고 있다. 그러나 다국어를 대상으로 하는 음성인식의 경우 다국어 혼합음성에 대한 이해 부족과 시스템 성능의 한계로 인하여 원활한 인식율의 개선은 이루어지지 않고 있다. 여러 나라의 혼합 언어로 표현된 음성의 경우 하나의(단일) 음성모델로 구현하는 것이 쉽지 않고, 또한 여러 개의 음성모델을 사용한 시스템의 경우 음성인식 성능의 저하라는 문제점이 있다. 이에 따라 다양한 언어로 구성되어 있는 음성을 하나의 음성모델로 표현할 수 있는 다국어 음성인식 모바일 시스템의 개발 필요성이 증가되고 이에 대한 연구가 필요하다.

본 논문에서는 모바일 시스템에서 다국어 혼합 음성모델을 사용하기 위한 기본연구로써 한국어와 영어 음성을 국제 음성기호(IPA)로 인식하는 통합음성모델 시스템 구축을 연구하였고, 한국어와 영어 음소를 동시에 만족하는 IPA모델을 찾는데 중점을 두어 실험한 결과 우리말 음성은 94.8%, 영어 음성은 95.36%라는 인식률을 얻을 수 있었다.

Abstract Recently, speech recognition technology has dramatically developed, with the increase in the user environment of various mobile devices and influence of a variety of speech recognition software. However, for speech recognition for multi-language, lack of understanding of multi-language lexical model and limited capacity of systems interfere with the improvement of the recognition rate. It is not easy to embody speech expressed with multi-language into a single acoustic model and systems using several acoustic models lower speech recognition rate. In this regard, it is necessary to research and develop a multi-language speech recognition system in order to embody speech comprised of various languages into a single acoustic model.

This paper studied a system that can recognize Korean and English as International Phonetic Language (IPA), based on the research for using a multi-language acoustic model in mobile devices. Focusing on finding an IPA model which satisfies both Korean and English phonemes, we get 94.8% of the voice recognition rate in Korean and 95.36% in English.

Key Words : Multilingual Speech Recognition, IPA

1. 서론

현재 전 세계는 스마트폰(Smart phone) 열풍으로 다양한 모바일 시스템이 서로 연결되어 교류하는 정보량이 증가하는 추세이다. 이런 모바일 시스템의 경우 다양한 OS와 응용S/W로 이루어져 있고, 여러 나라에서 다양하

게 사용되고 있다. 전 세계에서 통용되는 언어는 중국어와 인도어가 압도적으로 많은 사용자를 가지고 있으나, 공식적으로 사용하는 경우는 영어가 제일 많다. 그렇다 보니 자국어 사용자들 역시 자국어+영어의 사용이 일상화되어 있고, 이는 우리나라 역시 마찬가지다. 이렇듯 하나의 음성에 자국어와 영어가 공존하는 경우 두 나라의

본 논문은 2008년도 호서대학교의 재원으로 학술연구비지원을 받아 수행된 연구임(과제번호 20080059).

*교신저자 : 김석동(sdskim@hoseo.edu)

접수일 11년 05월 19일

수정일 11년 07월 06일

게재확정일 11년 07월 07일

음성을 동시에 인식 할 수 있는 시스템은 반드시 필요하다.

모바일 음성인식 시스템에서의 음성처리는 모바일 분야에서 핵심이 되는 기술로써 국가별로 사용 할 음성인식에 대한 요구 역시 증가하고 있다[1]. 다국적 통신문화와 인터넷+모바일 환경의 특성상 모바일 환경에서 다국어를 대상으로 하는 음성인식에 대한 요구가 증가하고 있다. 때문에 여러 나라의 서로 다른 언어를 하나의 음성 모델로 표현하는 다국어 시스템의 개발은 매우 중요하다[2]. 기존의 음성인식의 경우 그 동안 자국어만으로 인식 시스템이 개발되고, 사용했다. 본 논문에서는 기존 음성인식 시스템을 확장해서 공식적으로 가장 많이 사용되는 영어와 자국어인 우리말을 동시에 인식될 수 있는 혼합 음성 인식기에 대한 기초적 연구를 수행하였고, 다양한 모바일 시스템 또는 인터넷 통신상에서 다국어 동시인식 시스템의 개발을 이룰 수 있는 기본기술을 연구하였다. 이러한 다국어 혼합인식 시스템 개발이 필요한 이유로는 첫째로 음성모델과 파라미터를 공유함으로써 음성인식 시스템의 복잡도를 줄일 수 있고, 둘째로 통합된 하나의 언어를 하나의 음성모델만으로도 식별이 가능하기 때문이다[3]. 음성인식에서 가장 많이 사용되는 연속음성인식(HMM)의 근본적인 문제점은 여러 나라의 언어들을 어떻게 효율적으로 다룰 것인가에 대한 것이다[4]. 현재 영어로 개발한 음성인식 시스템은 여러 나라의 언어에 성공적으로 이식되어 사용되고 있다. 대표적인 음성인식 시스템으로 IBM, Dragon, BBN, Cambridge, Philips, MIT, LIMSI 등이 있다[5]. 이러한 시스템들은 영어 음성인식 시스템을 독일어, 일본어, 프랑스어, 중국어 등에 적용하여 음성기술을 언어에 독립적으로 일반화를 시켰으며 비슷한 모델링을 통해 여러 언어에 대해 적용하였다[6]. 영어 음성인식 시스템을 기본으로 하고 있는 음성인식의 경우 자국어 이외의 언어에 대한 언어 인식 시스템을 도입할 때 많은 문제점들이 나타난다[7].

본 논문을 통하여 그 동안 각각의 나라별로 개발되는 음성인식 시스템의 확장을 통해서 영어와 한국어가 동시에 인식될 수 있는 혼합 음성인식기에 대한 최근의 연구 성과를 확인할 수 있고, 차후 웹 또는 모바일 통신상에서 기존의 한국어 또는 영어만의 단일어 인식시스템이 아닌 다국어 동시인식 시스템의 개발을 이룰 수 있는 원천적 기술의 습득과 이를 이용한 다국어 ASR(Automatic Speech Recognition)에 대한 현실적인 접근을 위한 중요한 요소에 초점을 맞추었다.

2. 국제음성기호(IPA)

2.1 국제음성기호 (International Phonetic Alphabet)

본 연구는 이 국제음성기호(IPA)를 기본으로 하여 여러 나라 언어의 음성모델을 통합하는 기술을 연구하는데 그 목적이 있다. 이를 위해서 우리말과 영어를 하나의 국제음성기호(IPA)에 결합하여 인식실험을 하였다. 오늘날에는 여러 언어에 대한 대량의 자료를 이용해서 잘 알려진 언어로써 확장되고 있으며 이 자료의 인식을 위해서는 상당히 많은 양의 음성자료가 필요하고 이에 대한 문장들 또한 필요하다. 그러나 현실적으로 많은 음성을 수집하는 것은 매우 어렵다. 첫째로 많은 양의 음성 수집에는 많은 시간과 자원이 필요하다. 둘째, 전 세계에는 4,000여 종류의 언어가 존재하며, 그중에서 약 10%의 언어가 적어도 100,000명 정도의 언어 사용자를 보유하고 있다. 이러한 음성을 모두 수집하기란 현실적으로 불가능하다. 마지막으로 외국인어 발음하는 음성 인식 분야는 여러 가지로 활용이 늘어가고 있지만 그에 대한 많은 양의 자료를 구할 수 없다. 문제에 대한 해결책으로 국제음성기호(IPA) 표기 방법을 이용하였다.

2.2 IPA표기 방법의 장점

다국어를 처리하면서 한 나라를 대상으로 하는 음성인식기를 만드는 목적은 첫째, 여러 언어를 대상으로 한 음성인식물 중에서 가장 좋은 결과와 비교하기 위함이고 둘째, 새로운 언어에 대한 적응을 하기 위해서이다. 이러한 목적을 위해 우리는 다국어 연속 음성 인식에 대한 다양한 연구를 수행해 왔다. 특히 여러 나라 언어를 다루기 위해 우리는 언어의 기본 표현방법 중 국제음성표기(IPA) 방법을 이용한다. 언어에 독립적인 음성모델들은 다국어 시스템의 파라미터나 음성인식의 복잡도를 줄이기 위해 음성모델과 자료를 공유한다. 또한 이러한 음성모델은 새로운 언어에 대한 기본 모델로 사용된다. 음성과 언어 모델링에 사용한 통계적인 방법은 많은 분량의 음성과 문장들이 필요할 뿐만 아니라 발음사전과 대량의 문장들이 필요하다.

본 연구에서는 음성모델링에 주요 초점을 맞추고 있기 때문에 다른 자료들의 경우 주어진 언어에서 다룬다고 가정한다. 언어 독립 모델링의 목적은 모든 언어를 포함시키는 음성모델을 만드는데 있는 반면 언어 적응 모델링의 목적은 기존의 모델을 새로운 언어로 최적화하여 인식하도록 적응시키는데 있다. 새로운 언어에 대한 정보가 소량인 경우에도 가능하다. 자료가 소량인 경우 두 가

지 문제가 있다. 하나는 새로운 언어에 대한 초기 음성모델을 위한 적당한 기본 모델을 결정하는 것과 둘째로 문맥 의존(Context Dependent) 음성모델을 만들기 위해 음소 문맥으로 확장할 때 기존 언어와 새로운 언어 사이에 부합이 안 되는 음소가 많을 경우이다. 이와는 반대로 많은 양의 음소 문맥을 사용하면 기존의 음성인식과는 다르게 새로운 음성의 인식률이 크게 증가한다.

3. 음성인식 자료와 학습

3.1 음성모델 비교

음성모델(Acoustic model)의 목적은 주어진 단어에 대하여 임의의 벡터에 대하여 비슷한 정도를 계산하는 것이다. 원칙적으로 확률분포는 각각의 단어와 이에 대한 벡터열로 계산할 수 있으나 대용량의 어휘시스템에는 불가능하므로 대신에 단어열을 음소(Phonetics)라 불리는 기본음성으로 나누어 처리한다. 본 연구에선 우리말과 영어의 음성모델을 BW(Baum-Welch)알고리즘을 이용하여 HMM 모델을 학습하여 새로운 음성모델을 만들고, 두 음성모델을 병합하여 하나의 음성모델을 만들어 사용하였으며 한국어와 영어의 개별인식과 혼합인식실험을 통해 최적의 단일 음성모델을 구축하는데 초점을 두었다.

본 연구에서 사용한 한국어의 경우 한국어 음성모델(총36개)을 자체적으로 개발해 사용하였으며 영어의 경우 CMU(Carnegie Mellon University)에서 개발한 음성인식기인 Sphinx 시스템에서 사용하고 있는 영어 음성모델(총 39개)을 적용하였다. 실험에 사용한 음성모델은 아래 표 1에 정의되어 있다.

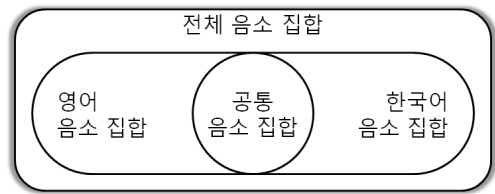
[표 1] 영어와 한국어 음성모델 정의의 예
[Table 1] The Examples of Definitions of Acoustic models in Korean and English

영어 음성모델(39개)					
음소	Exam.	음소	Exam.	음소	Exam.
AA	odd	F	fee	P	pee
AE	at	G	green	R	read
AH	hut	HH	he	S	sea
AO	ought	IH	it	SH	she
AW	cow	IY	eat	T	tea
AY	hide	JH	gee	TH	theta
B	be	K	key	UH	hood
CH	cheese	L	lee	UW	two
D	dee	M	me	V	vee
DH	thee	N	knee	W	we
EH	Ed	NG	ping	Y	yield

ER	hurt	OW	oat	Z	zee
EY	ate	OY	toy	ZH	seizure
한국어 음성모델(36개)					
AA	까닭은	L	음식물	R	비바람
AE	빠내	M	아침	S	따라서
AO	따라서	N	아니고	T	비슷한
B	어업은	NG	일등상	UW	부도
CH	엄청난	OW	일곱개	W	꽤
D	영탁군	P	발표에	Y	뿔혀
EH	영향에	KK	이끌	TC	어젯밤에
G	여쭙보고	PP	어젯밤	EU	어느
HH	열흘	TT	받았던	KC	역사의
IY	열린	AN	복상하고	PC	유입
JH	위주의	EN	쟁긴	SS	앞서서
K	육해공군	ON	충청	JJ	이쪽

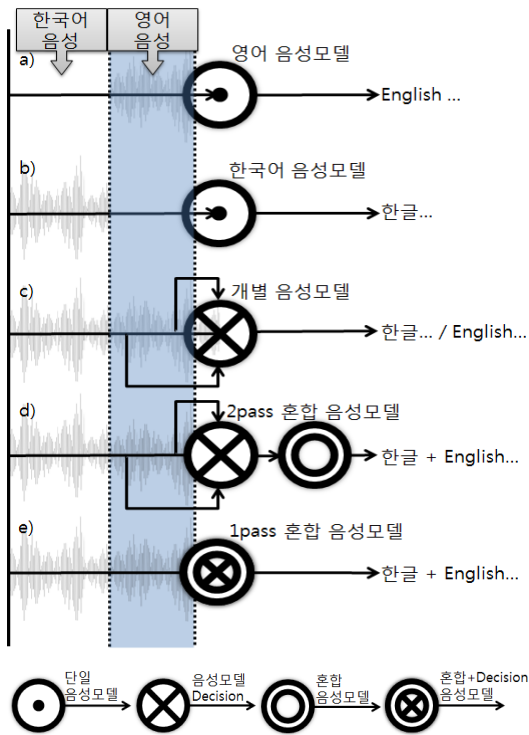
3.2 음성학습 및 인식

일반적으로 음성학습을 위하여 각 나라별 음성을 가지고 각 나라별 음성모델을 만든다. 이때 우리말과 영어의 발음현상과 규칙이 다르므로 각각에 적당한 음성모델이 필요하다. 따라서 그림 1 같이 각각의 언어가 공통으로 사용하는 음소집합과 우리말 또는 영어 고유음소집합을 결정해야 한다.



[그림 1] 공통음소집합
[Fig. 1] A set of common phonemes.

본 연구에서는 그림 2에서와 같이 영어음성 [a]와 한국어음성 [b]를 가지고 개별 음성모델을 만들었지만 음성학습은 영어음성을 가지고 학습한다. 개별 음성모델을 이용하여 한국어와 영어의 혼합된 음성을 인식하기 위해서는 [c]와 같이 한국어와 영어를 개별적으로 인식한 후 각각 구분하는 음성모델 Decision이 필요하다. 혼합 음성모델을 생성하기 위해서는 [d,e]같이 수행한다. 우리말과 영어를 동시에 학습 자료로 사용하여 음성모델을 생성한다.



[그림 2] 음성학습 및 인식 실험 방법
 [Fig. 2] The Experimental methods for Voice Learning and Recognition

본 연구에서는 공통음소집합의 크기를 늘려, 혼합음성 모델의 시스템 효율을 높이는데 목적이 있다. 예를 들면 한국어 고유의 음소 집합의 경우 모음과 자음별로 “-”와 같은 이중모음, “ㄱ”, “ㄷ” 등의 자음 등의 음소들이 있고, 영어 고유 음소 집합의 경우에는 “AX”, “AXR” 등이 있을 수 있다. 따라서 고유한 음소와 별개로 공통분모인 공통 음소를 최대한 많이 찾아 비슷한 음소를 최대한 공유하여 음성모델의 복잡도를 줄여 시스템의 효율을 높이 고자 한다.

3.3 음성모델 생성시 국제음성기호(IPA) 활용

실험에서 다국어의 연속 음성 인식을 수행하기 위해서는 음소의 음성학적인 표현 방법을 우선 고려해야 하며 이는 음성학적 표현방법이 나라별로 관계없이 서로 비슷하다는 가정 하에 나라에 따라 고유한 음소를 찾고 이를 표기하기 위해 국제음성기호(IPA)를 이용한다.

이 가정에 따라 두 나라 언어의 특정한 음소를 IPA의 음소 집합중 하나로 놓게 되고, 이 경우 어느 음소가 어떤 기호에 부합되는지 찾는 것이 문제가 된다. 이러한 문

제들을 실험을 통해 학습시키고 오류를 분석하여 다시 실험하는 과정이 반드시 필요하다. 현재 나라와 독립적으로 연속 음성 인식을 수행하기 위해서는 음소의 음성학적인 표현 방법을 우선 고려해야한다. 본 연구에서는 음성학적 표현방법이 나라별로 관계없이 서로 비슷하다는 가정에 입각해 수행한다. 이에 따라 우리는 나라에 따라 고유한 음소를 표기하기 위해 국제음성기호(IPA)와 부합시켜 공통점을 찾았다.

본 연구에서는 다양한 실험에 의해 가장 적합한 IPA 음소를 찾고자 하였다. 특히 이중 모음의 경우 하나의 음소로 표현하는 방법(IPA(1))과 CD 모델같이 두 개의 음소로 표현하는 방법(IPA(2))으로 나누어 실험을 하였다. 예를 들어 ‘에’ 인 경우 하나의 IPA음소인 e와 두 개 음소의 결합인 ‘에+에’로 나타낼 수 있다. 우리는 실험을 통해 단일 음소와 두 개의 음소 결합 중 인식이 좋은 IPA음소를 찾았다.

4. 실험 및 결과

본 논문에서 사용된 음성모델들의 생성을 위해 영어와 한국어 음성의 음소인식을 모두 하였고, 각각의 음소인식을 위한 학습자료의 경우 한국어의 경우 한글 음성 학습 자료는 남성과 여성을 포함한 총 162명의 음성을 21시간에 걸친 음성자료를 이용하여 학습하였고, 영어 음성 학습 자료는 2,997개의 단어로 구성된 미국 신문 자료를 원어민이 발음한 음성을 학습하여 본 실험에 이용하였다. 언어모델의 경우 생성에 사용된 문서자료는 인터넷을 통해 한국의 신문사 및 방송사의 자료를 수집하여 그 중에서 공개된 문서자료만을 사용하였다.

4.1 실험에 사용한 학습자료

본 논문은 2,997개의 단어, 68명의 화자, 136번의 발음 자료를 이용하였다. 처음실험에서 우리의 기본 설계안은 30.72%의 초기 오류율을 보였다. 이는 비용을 들이지 않았다는 점(시간과 비용)을 고려해볼 때 상당히 낙관적인 시작이었다. 최소한의 전문 미세 조정으로 오류율을 20.07%로 낮추었다.

모든 자료를 수동으로 만들 때의 결과보다 낮은 성공률을 보이긴 하지만 일단 생성되면 거의 노력 없이 극적인 개선을 이루어 낼 수 있음을 간접적으로 보여주었다.

4.1.1 학습 방법

먼저 우리의 방법을 한국어에 적용하였다. 두 언어 모

델에 사용된 텍스트 자료와 음성 데이터 자료는 온라인 사이트에서 공개적으로 수집하였다. KSC완성형 코드에서 얻어진 텍스트는 유니코드(UCS-2)로 전환되었고 문장의 발음에 따라 구분되었다. 발음이 음성화 된 후 녹음된 스크립트에서 음성적으로 풍부한 하나의 세트를 뽑아내기 위해 최소한의 보존 설계를 하였다. 발음 사전은 자동 생성 프로그램에 의해 만들었다. 이렇게 만든 발음사전은 음성모델(Acoustic model)의 훈련과 인식에 모두 사용되었다.

4.1.2 음성모델

표 2는 본 연구에 사용된 IPA(1),IPA(2)의 음성모델이다. 음성인식에서 여러 언어에 걸친 ASR시스템 비교는 어려운 것이 현실이다.

[표 2] IPA(1,2) 음성모델 정의

[Table 2] The Definitions of Acoustic models in IPA(1,2)

No	TiMit	Korea IPA	MIX	English IPA	ASCII
1	AA		a		aa
2	AE		æ		ae
3	AH			ɑ	ah
4	AO	ɔ		ʌ	ao
5	AW			au	aw
6	AX			ə	ax
7	AXR			ə	axr
8	AY			aɪ	ay
9	b		b		b
10	BD			bd	bd
11	CH		tʃ		ch
12	d		d		d
13	DD			d	dd
14	DH			ð	dh
15	DX			dx	dx
16	EH		ɛ		eh
17	ER			ɜ	er
18	EO	ɔ*			er
19	EY			eɪ	ey
20	f			f	f
21	G	G			g
22	GD			gd	gd
23	GG	GG			gg
24	HH		h		hh
25	IH			ɪ	ih
26	IX			-	ix
27	IY		i		iy
28	JH		dʒ		jh
29	k		k		k
30	KD			kd	kd
31	l		l		l

32	m		m		m
33	n		n		n
34	NG		ŋ		ng
35	OW		oʊ		ow
36	IX				ix
37	OY			ɔɪ	oy
38	p		p		p
39	PD			pd	pd
40	r		r		r
41	s		s		s
42	SH			ʃ	sh
43	t		t		t
44	TH			θ	th
45	TS			ts:	ts
46	UH			ʊ	uh
47	UW		u		uw
48	v			v	v
49	w		w		w
50	y		j		j
51	z			z	z
52	ZH			ʒ	zh
53	KK	k			kk
54	TT	t			tt
55	PP	p			pp
56	SS	s			ss
57	JJ	tʃ			jj
58	KC	k*			kc
59	TC	t*			tc
60	PC	p*			pc
61	EU	u			eu
62	YA	YA(야)	Y AA		ya
63	YEO	ja(쟤)	Y EO		yeo
64	YO	jo(요)	Y OW		yow
65	YU	ju(유)	Y UW		yuw
66	YAE	jae(쟤)	YAE		yae
67	YE	je(쟤)	Y EH		je
68	WA	wa(와)	W AA		wa
69	WAE	wae(와)	W AE		wae
70	WO	wɔ(와)	W EO		wo
71	WE	we(와)	W EH		we
72	UI	ij(이)	EU IY		ui
73	WI	wi(y)(위)	W IY		wi
74	OE	wɛ(외)(위)	W EH		we

그럼에도 불구하고 다양한 노력의 결과로 한국어 음성 모델이 영어 음성모델과 유사하게 작동하는 것을 실험으로 확인하고 실험에 사용된 한국어 음성모델이 PLI(Phoneme Language Identity)파일을 이용하여 자동적으로 생성되는 발음사전(Lexicon)을 이용하여 조절하였다. 기존에 사용되던 수동 생성 발음사전을 활용한 실험 결과와 PLI를 사용한 자동 생성 발음사전을 활용한 실험 결과가 유사한 패턴을 보이는 것을 미루어 볼 때 이는 PLI로 만드는 것이 수동 생성 발음사전의 실용적인 대안

이라는 주장을 뒷받침할 수 있었다.

이에 본 논문의 실험에는 기존의 음성모델을 확장하여 IPA로 혼합음성모델을 만들어 PLI를 이용해 실험하여 기존에 사용되던 각각의 음성모델을 하나의 IPA 음성모델로 작성하여 시스템의 복잡도를 줄일 수 있는 방법을 추구한다.

4.2 IPA 분리

우리는 실험에서 기존의 IPA를 IPA(1), IPA(2)로 구분하여 실험하였다. IPA(1)의 경우 우리말의 이중모음을 배제한 음소만(61개)을 이용하여 실험하였고, 표2에서 음영으로 표시한 IPA(2)의 경우 이중모음을 추가하여(61개 + 13개) 실험한 결과이다.

4.2.1 음성학습에 의한 IPA(1)방식 인식

IPA(1)의 경우 한글에서 초성 /ㄱ/을 사용 시점에 따라 두 가지로 나누어 처음 시작하는 /ㄱ/은 G로 글자 사이에 있는 초성 /ㄱ/은 GG로 나누어 처리하였으며, 영어와 한글의 /어/발음은 ‘ɔ’와 ‘ʌ’로 분리해서 사용하였다. 이렇게 분리한 음소까지 포함하여 한글 고유 음소 13개, 영어 고유 음소 26개, 그리고 공통의 음소 22개까지 총 61개의 음소를 이용하여 실험하였다. 여기에 사용된 음소는 표 3과 같다.

[표 3] IPA(1)에서 사용한 한글,영어 혼합 음소
[Table 3] the mixed phonemes in Korean and English, used in IPA(1)

한글 음소(13개)	영어 음소(26개)	공통 음소(22개)
ɔ, ɔ̃, G, GG, k, t, p, s, ʃ, k*, t*, p*, u	a, ʌ, au, ə, ə̃, aɪ, bɔ̃, d, ɔ̃, dɔ̃, ɜ, eɪ, f, gɔ̃, ɪ, -, kɔ̃, ɔɪ, pɔ̃, ʃ, θ, ʃ:, ɔ̃, v, z, ʒ	a, æ, b, ʃ, d, e, h, I, ɔ̃, k, l, m, n, ɪ, oɔ̃, p, r, s, t, u, w, j

4.2.2 음성학습에 의한 IPA(2)방식 인식

IPA(1)에서는 한글중 이중모음에 해당하는 문자들의 경우 없는 글자들로 인하여 인식이 낮아지게 되었고 이에 기존의 IPA(1)의 61개의 음소에 이중모음 Y, W에 새로운 음소 13개(YA(야), jA(쟤), jo(요), ju(유), jæ(혜), je(헤), wa(와), wæ(왜), wA(워), we(웨), ij(-), wi(y)(기), e(φ)(시))를 추가하여 총 74개의 음소를 이용하여 실험하였고 여기에 사용된 음소는 표 4에 정리해 놓았다.

[표 4] IPA(2)에서 사용한 한글,영어 혼합 음소
[Table 4] The Mixed phonemes in Korean and English, used in IPA(2)

한글 음소(26개)	영어 음소(26개)	공통 음소(22개)
ɔ, ɔ̃, G, GG, k, t, p, s, ʃ, k*, t*, p*, u, YA, jA, jo, ju, jæ, je, wa, wæ, wA, we, ij, wi(y), e(φ) (기본+이중모음)	a, ʌ, au, ə, ə̃, aɪ, bɔ̃, d, ɔ̃, dɔ̃, ɜ, eɪ, f, gɔ̃, ɪ, -, kɔ̃, ɔɪ, pɔ̃, ʃ, θ, ʃ:, ɔ̃, v, z, ʒ	a, æ, b, ʃ, d, e, h, I, ɔ̃, k, l, m, n, ɪ, oɔ̃, p, r, s, t, u, w, j

4.3 인식실험 결과

IPA(1)에서 사용된 총 61개의 음소를 이용한 실험의 결과는 한글의 경우 90.62%, 영문의 경우 91.71%의 인식률을 얻었다. 그리고 IPA(1)의 61개의 음소에 한글음소의 이중모음 13개를 추가하여 사용한 IPA(2)의 실험의 경우 한글의 경우 94.8%, 영문의 경우 95.36%로 IPA(1) 대비 높은 인식률을 얻었다. 인식률이 증가한 원인으로는 첫 째로 기존에 음성에서 인식하지 못했던 음소들을 이중모음으로 다시 정의한 것과 또 하나 한글음소인 이중모음의 추가로 영문음소의 수에는 변화가 없었으나 불특정했던 한글의 이중모음에 해당하는 한글음소가 인식됨으로 이를 영어 음소 안에서 다시 찾아 인식실패의 경우를 줄임으로 영어 인식률의 동반상승 효과를 얻을 수 있음을 알게되었다.

[표 5] 영어와 한국어 혼합인식 결과
[Table 5] The Results of Mixed recognition in Korean and English.

	Recognition Rates (%)	
	한글	영문
IPA(1)	90.62	91.71
IPA(2)	94.80	95.36

5. 결론

인터넷의 발전과 함께 사용자간에 교류되는 정보량이 폭발적으로 늘어나고 있다. 인터넷을 통해 국경의 벽을 넘어 다양한 나라의 정보들이 넘쳐나고 있으며, VoIP와 다양한 음성검색 S/W들의 활발한 사용으로 음성정보 역시 넘치고 있다. 모바일기기의 혁신적인 발전에 힘입어 다양한 음성인식 방법들이 제시되고 다양한 콘텐츠들이 웹서비스를 통해 교류되고 있다. 이렇게 서로 주고받는

다양한 음성정보의 경우 하나의 음성에 우리말과 영어가 혼합되어 나오는 경우가 많아지고 있다. 이러한 경우 우리말과 영어 음성을 정확히 인식하는 시스템의 필요성이 증가하고 있고, 단일어 음성정보 기반의 현재 시스템의 문제점들이 제기되면서 단일어가 아닌 다국어 음성정보의 인식에 대한 중요성 역시 증대되고 있다.

본 연구는 이 IPA(International Phonetic Alphabet)를 기본으로 하여 여러 나라 언어의 음성모델을 통합하는 기술을 연구하는데 그 목적이 있다. 그래서 그동안 개별적으로 개발되는 단일어 음성인식 시스템을 확장해서 영어와 우리말을 동시에 인식될 수 있는 혼합 음성인식에 대한 기초적 연구를 수행하였고, 또한 이를 이용하여 다양한 모바일 통신상에서 한국어 또는 영어만의 단일어 인식시스템의 확장이 아닌 다국어 동시인식 시스템의 개발을 이룰 수 있는 효과적인 기술의 범주로 응용될 수 있다. 활용여부에 따라 어휘모델의 생성을 통한 다국어인식과 함께 각 언어에서 지역적 특색인 사투리 및 서로 다른 발음의 단어들을 인식 할 수 있다.

본 연구를 활용하면 여러 언어를 대상으로 어휘모델을 생성 할 수 있으며, 기본적으로 혼합 음성인식기를 이용하여 하나의 시스템으로 서로 다른 언어의 음성에 대하여 인식이 가능하기 때문에 다른 언어 처리에 사용될 시스템 구성 시 비용절감에 도움이 된다. 국제음성기호(IPA)를 음성인식기술에 효과적으로 이용하기위해선 더욱 효율적인 음성모델을 찾고 다양하면서도 음성자원의 효율성을 높일 수 있는 연구가 필요하며, 향후 실시간 시스템 처리를 위한 효율적인 인식방법의 연구가 이루어져야 한다.

References

[1] Stefanie Tomko, and Roni Rosenfeld. " A Speech-and Language-based Information Management Environment". In Proc. IEEE Int.l Conference on Acoustics, Speech and Signal Processing, Toulouse, France, May 2006.

[2] Z. Al Bawab, B. Raj, and R. M. Stern, "Analysis-by-synthesis features for speech recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing, April 2008, Las Vegas, Nevada.

[3] H.-M. Park and R. M. Stern, "Missing-feature speech recognition using dereverberation and echo suppression in reverberant environments," IEEE International Conference on Acoustics, Speech, and Signal

Processing, April 2007, Honolulu, Hawaii.

[4] Thomas K. Harris, Arthur Toth, James Sanders, Alexander Rudnicky. "Towards Efficient Human Machine Speech Communication". ACM Transactions on Speech and Language Processing, February 2005..

[5] Jahanzeb Sherwani et el " Towards Speech-based Access by Semi-literate Users". In Proc. Speech in Mobile and Pervasive Environments, Singapore, September 2007.

[6] John S. Garofolo, Jonathan G. Fiscus,William M. Fisher "Design and prtparation of the 1996 HUB-4 Broadcast News Benchmark Test Corpora." DARPA Speech Recognition Workshop, Feb. 1997, pp. 15 - 21.

[7] A. G. Hauptmann, et el. "Multi-Lingual Broadcast News Retrieval", TRECVID'06 TREC, NIST Gaithersburg, MD., November 2006.

김 석 동(Suk-Dong Kim)

[정회원]



- 1984년 2월 : 아주대학교대학원 전자공학과 (공학석사)
- 1993년 2월 : 아주대학교대학원 전자공학과 (공학박사)
- 1984년 3월 ~ 현재 : 호서대학교 컴퓨터공학과 교수

<관심분야>

Speech Recognition, Ubiquitous Technology

김 우 성(Woo-Sung Kim)

[정회원]



- 1983년 5월 : Texas A&M 대학원 컴퓨터공학과 (공학석사)
- 1993년 8월 : 서강대학교대학원 컴퓨터공학과 (공학박사)
- 1987년 3월 ~ 현재 : 호서대학교 컴퓨터공학과 교수

<관심분야>

지능정보 시스템, USN, 장비제어, 영상처리

우 인 성(In-Sung Woo)

[정회원]



- 2003년 8월 : 호서대학교대학원 정보보호학과 (공학석사)
- 2008년 8월 : 호서대학교대학원 컴퓨터공학과 (공학박사)
- 2008년 10월 ~ 2009년 10월 : 美SOU Post-doc.
- 2003년 9월 ~ 현재 : 호서대학교 컴퓨터공학과 강사

<관심분야>

Microprocessor, Ubiquitous, Voice XML, HMI