

# 청음 음성학적 지식에 기반한 음가분류에 의한 핵심어 검출 시스템 구현

김 학 진<sup>†</sup> · 김 순 협<sup>††</sup>

## 요 약

본 논문은 한국어 대용량 음성인식에 기초가 되는 유사 음소 단위의 분류와 한국어 7종성과 8종성의 유효성 등 두 가지 관점에서 대해 설명한다. 유사 음소는 음소를 음성학적으로 조음 위치와 조음방법에 따라 분류하며, 보통 한국어 음성 인식에서는 50개 전후의 유사 음소 단위(PLU, Phone-likely unit)를 사용하고 있다. 본 연구에서는 유사 음소 단위의 분류에 청음 음성학적 지식을 적용하여 45개의 유사 음소 단위를 제시한다. 모음 /이, 히/를 [ee], /히, 헤/를 [ye], 그리고 /시, 새, 게/를 [we] 등을 유사음소로 분류하였다. 두 번째로는 현재 사용되고 있는 한글 맞춤법 통일안의 7종성 체제와 훈민정음 해례의 8종성가족용에 대해 설명한다. 한국어의 종성에 활용되는 음소 중 /ㄷ/와 /ㅅ/에 대한 음가가 동일한가? 하는 의문은 학계의 오랜 논쟁거리의 하나다. 본 연구에서는 한국어 종성의 변천 과정을 살펴보고, 7종성과 8종성가족용을 음성 인식에 활용하여 그 유효성을 검증한다. 실험은 고립단어 인식과 연속 음성 인식으로 나누어 실험하기 위해 먼저, 고립단어 인식은 PBW452를 이용하고 남·녀 약 50명이 5개조로 나뉘어 각각 50단어를 발성하여 실험하였다. 연속음성 인식은 구현되는 증거거레 시스템에서 활용할 수 있도록 증거거레 71문장의 문장 코퍼스와 이를 발성한 음성 코퍼스를 수집 활용하였으며, 남·녀 각 5명이 문장을 각각 2회씩 발성하여 실험하였다. 실험 결과 종성으로 8종성가족용을 적용했을 때 평균 약 1.45%의 인식 성능이 향상되었으며, 8종성가족용과 청음 음성학을 동시에 적용한 유사 음소 단위를 적용할 때에는 평균 1.5~2.02%의 인식을 향상을 확인하였다. 연속 음성에 대한 인식 실험에서는 기존의 49개 또는 56개 유사 음소 단위를 이용한 경우보다 평균 약 1~2%의 인식 성능 향상을 확인하였다.

## The Design of Keyword Spotting System based on Auditory Phonetical Knowledge-Based Phonetic Value Classification

Hack-Jin Kim<sup>†</sup> · Soon-Hyub Kim<sup>††</sup>

## ABSTRACT

This study outlines two viewpoints the classification of phone likely unit (PLU) which is the foundation of Korean large vocabulary speech recognition, and the effectiveness of *Chiljongseong* (7 Final Consonants) and *Paljongseong* (8 Final Consonants) of the Korean language. The phone likely classifies the phoneme phonetically according to the location of and method of articulation, and about 50 phone-likely units are utilized in Korean speech recognition. In this study auditory phonetical knowledge was applied to the classification of phone likely unit to present 45 phone likely unit. The vowels '이, 히' were classified as phone-likely of [ee]; '히, 헤' as [ye]; and '시, 새, 게' as [we]. Secondly, the *Chiljongseong* System of the draft for unified spelling system which is currently in use and the *Paljongseonggajokkyong* of Korean script haerye were illustrated. The question on whether the phonetic value on 'ㄷ' and 'ㅅ' among the phonemes used in the final consonant of the Korean language is the same has been argued in the academic world for a long time. In this study, the transition stages of Korean consonants were investigated, and *Chiljongseong* and *Paljongseonggajokkyong* were utilized in speech recognition, and its effectiveness was verified. The experiment was divided into isolated word recognition and speech recognition, and in order to conduct the experiment PBW452 was used to test the isolated word recognition. The experiment was conducted on about 50 men and women - divided into 5 groups - and they vocalized 50 words each. As for the continuous speech recognition experiment to be utilized in the materialized stock exchange system, the sentence corpus of 71 stock exchange sentences and speech corpus vocalizing the sentences were collected and used 5 men and women each vocalized a sentence twice. As the result of the experiment, when the *Paljongseonggajokkyong* was used as the consonant, the recognition performance elevated by an average of about 1.45%; and when phone likely unit with *Paljongseonggajokkyong* and auditory phonetic applied simultaneously, was applied, the rate of recognition increased by an average of 1.5% to 2.02%. In the continuous speech recognition experiment, the recognition performance elevated by an average of about 1% to 2% than when the existing 49 or 56 phone likely units were utilized.

키워드: 핵심어검출(Keyword Spotting), 음가분석(phonetic value), KWPLU, 청음음성학(Auditory Phonetics), 음성신호처리(Speech Signals)

\* 본 연구는 광운대학교 2002학년도 교내 학술 연구비의 지원으로 이루어졌습니다.

† 정 회 원: 신용보증기금

†† 정 회 원: 광운대학교 컴퓨터공학과 교수

논문접수: 2002년 10월 4일, 심사완료: 2003년 3월 3일

### 1. 서 론

음성 인식에 있어 하나의 문장으로 구성된 연속 음성을 인식하기 위해서는 문장 전체를 완벽하게 인식하는 방법이 가장 좋은 방법이나, 문장 중 부분적인 중요 단어만을 인식하는 방법도 사용 용도에 따라 효과적이다[1]. 핵심어 검출은 문장으로 된 연속 음성 중에서 태스크에 필요한 핵심어(중심어, Keyword)만을 인식하는 것이다[1]. 대용량 음성 인식에는 언어모델과 어휘모델 그리고 음향모델 등의 지식 정보가 필요하며 이 세 요소 중 음향모델과 어휘모델의 근간이 되는 음소의 음가 재분류를 통해 핵심어 검출기의 성능 향상을 시도하고자 한다. 이를 위해 먼저, 음가의 분류에 청음 물리 음성학적 지식을 적용하여 유사 음소 단위를 재분류하였으며, 두 번째로는 현재까지도 학자들간의 논란이 되고 있는 훈민정음 해례의 8중성가죽용과 한글맞춤법 통일안에서 제시하는 7중성을 음성 인식에 적용하여 그 유용성을 실증해보고자 한다. 대어휘 연속 음성 시스템의 구현에 적합하며, 현재 가장 많이 사용되고 있는 부단어(subword) 단위의 인식 모델로 유사음소단위가 사용된다[2]. 한국어와 영어의 경우에 대략 50여개로 구성된 유사음소를 사용하며, 대표적인 것으로 광운대 KWPLU(Kwangwoon phone-likely unit) 49개, 서강대 유사음소(SGPLU, Seokang phone-likely unit) 49개 또는 54개[3], 전남대 유사음소(JNPLU, Jeonnam phone-likely unit) 56개 등으로, 이들 유사 음소는 변이음과 받침을 기준으로 한 음소 분류를 채택하고 있다. 본 논문에서는 유사 음소를 분류하기 위해 음운학적 지식과 청음 물리 음성학적 지식을 적용하여 유사 음소를 분류한 KWPLU를 제안하고자 한다. 한국어 맞춤법 표기에서 많이 틀리는 /나, 내, 네/와 /기, 히/, /히, 키/를 청음 물리 음성학적 지식을 기초로 각각 하나의 유사음소로 분류하였다. 두 번째 방법은 음소의 질의어 집합을 구성하기 위하여 한국어의 발음구조와 특성에 따른 음소의 분류를 초성은 자음 18개, 중성은 모음(단모음, 이중모음) 17개와 종성에는 한국어 7중성 외에 'ㄱ'을 추가한 8중성가죽용을 적용하여 구성하였다. 8중성가죽용은 훈민정음 해례 중성해 규정에서 중성에 쓸 수 있는 8개 자음을 정한 것으로 1933년 한글맞춤법 통일안이 제정되기 까지 사용되었던 것이다. 최근 해외 기업의 국내 진입은 차후 증권시장에 까지 진출할 것으로 보여져 외래어로 기업체명을 접하는 것은 어렵지 않다. 또한 본 논문에서 태스크 시스템으로 구현하고자 하는 증권시장에 등록된 기업의 상당수가 외래어로 된 상호를 갖고 있어 이들의 인식 성능 제고의 일환으로 8중성가죽용의 유효성을 검증하였다.

본 논문의 구성은 2장에서 기존의 한국어의 음가 구분과 한국어 종성의 변천 과정을 소개하고 3장에서 청음 음성학에 근거한 음소 분류에 대하여 제안하였으며, 4장에서 그

실험 결과에 대하여 비교 분석하고 5장에서 결론을 맺는다.

### 2. 한국어의 음운학적 음가구분 및 종성 표기법 연구

#### 2.1 음운학적 음가 구분

한국어 음소는 자음과 모음 체계로 자음 19개와 모음 21개로 구성되어 있다[4]. 이 음소의 분류 방법으로는 혀의 위치와 입 모양에 따라 모음을 구분하고, 조음의 위치와 방법에 따라 자음을 구분한다[5-9]. 다시 자음을 조음방법에 따라 파열음, 마찰음, 파찰음, 비음, 유음으로 세분류한다. 파열음의 조음방법은 구강의 한 부분을 막는 것과 막음을 지속하는 것, 구강을 개방하는 단계에 의해 발생된다. 파열음은 조음위치에 따라 양순음, 치조음, 그리고 연구개음으로 나뉘고, 긴장(tense) 및 기식성(aspiration) 유/무에 따라 파열음으로 나뉜다. /ㅂ, ㅃ, ㅍ/는 조음기관의 긴장 없이 약간의 기식성을 동반하는 연음이고, /ㅃ, ㅍ, ㅍ/는 조음기관의 긴장과 강한 기식성을 수반하는 유기경음이고, /ㅃ, ㅃ, ㅍ/는 조음기관의 긴장은 있고, 기식성은 거의 없는 무기경음이다. 우리말의 파찰음은 모두 연구개에서 조음된다. 마찰음은 파열음과 같이 막음 및 지속단계, 개방단계와 개방 후 마찰단계로 발생된다. 비음은 구강을 막고 연구개를 내려 폐로부터 나오는 기류를 비강으로 내보내어 발생되는데, 조음위치에 따라 양순음, 치음, 연구개음으로 나뉜다[7].

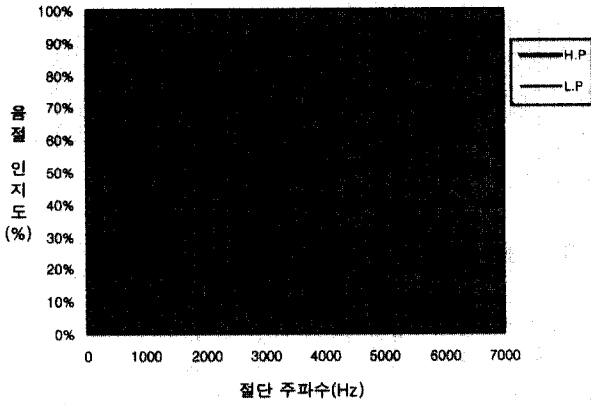
<표 2.1> 한국어 조음 방법과 위치에 따른 자음분류

방법 \ 위치	양순음 (bilabial)	치경음 (alveolar)	경구개음 (hard palate)	연구개음 (velar)	성문음 (glottal)
파열음(plosive)	ㅂ, ㅃ, ㅍ, ㅍ	ㄷ, ㄸ, ㅌ, ㅎ		ㄱ, ㅋ, ㆁ	
마찰음(fricative)		ㅅ, ㅆ			ㅎ
파찰음(affricate)			ㅈ, ㅉ, ㅊ		
비음(nasal)	ㅁ	ㄴ		ㅇ	
설측음(lateral)		ㄹ			

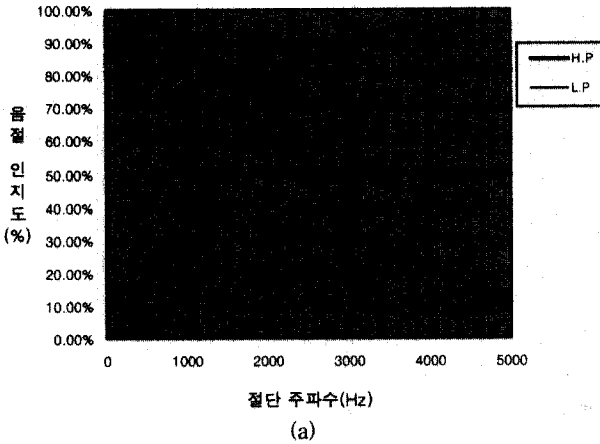
<표 2.1>은 자음을 조음의 위치와 방법에 따라 분류한 자음 음가 분류 표이다[7,9-11]. 혀의 위치에 따라 분류되는 모음(vowel)은 성대의 진동으로 작은 공기가 성도에서 공명을 일으켜 만들어지는데, 이때 성도가 어떠한 모양을 하고 있는가에 따라서 소리의 질이 결정된다. 모음의 조음에서 가장 중요한 변수는 혀의 고저 및 전후, 입술 모양 등 세가지 조음현상이 가장 중요한 변수로 작용한다. 혀의 고저와 함께 입을 적게 벌리고 내는 소리는 폐모음(close), 혹은 고모음(high), 입을 크게 벌리고 내는 소리를 개모음(open), 저모음(low), 고모음과 저모음의 사이에서 나는 소리를 중모음(Mid)이라고 한다. 혀의 전후 위치가 앞쪽에 있으면 전설모음(Front), 뒤쪽에 있으면 후설모음(Back), 그 중간에 있는 중설 모음(Central)이라고 한다. 또한 모음의 조음에서



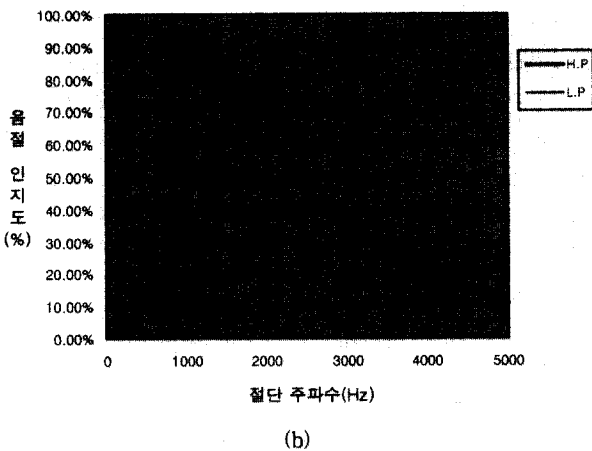
각 평균 약 72%, 71%의 인식률이 가능한 것으로 나타나고 있다.



(그림 3.1) 필터링 후 음성 인식률(HP : 고주파 LP : 저주파)



(a)



(b)

(그림 3.2) 음성 'ㄱ(a), ㄴ(b)'의 필터링 후 음성 인식률(HP : 고주파 LP : 저주파)

또한, 다른 방법으로 두 모음의 통합과정을 알아보기 위해 'ㄱ'과 'ㄴ'의 음향 분석 결과를 통해 이 두 모음의 조음 과정이 시대에 따라 변했음을 알 수 있었다. <표 3.1>은 1950년대부터 2000년대까지의 방송 자료에서 'ㄱ'과 'ㄴ' 모

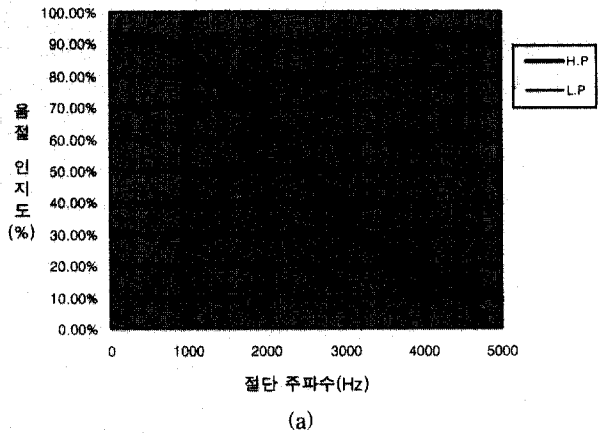
음의 포먼트 값을 나타낸 것이다.

<표 3.1> 모음 'ㄱ'과 'ㄴ'의 음가 변화

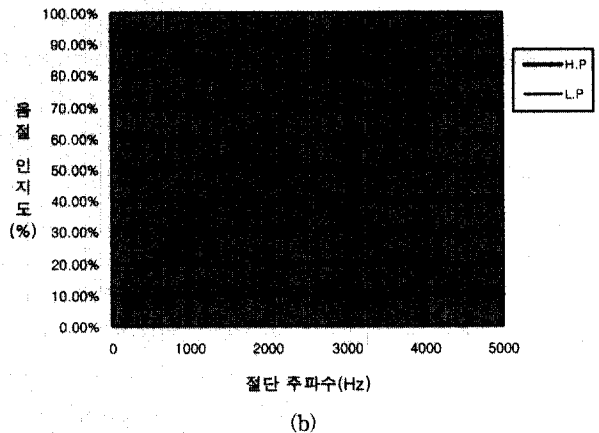
구분	/ㄱ/		/ㄴ/	
	F1	F2	F1	F2
1950년대	571 Hz	2033Hz	421Hz	2090Hz
1960년대	517Hz	1914Hz	452Hz	1937Hz
1970년대	547Hz	1880Hz	404Hz	1991Hz
1980년대	419Hz	1985Hz	374Hz	2070Hz
1990년대	415Hz	1707Hz	381Hz	1797Hz
2000년대	496Hz	1785Hz	449Hz	1817Hz

<표 3.1>에서 보는 바와 같이 1950년대부터 1970년대까지는 'ㄱ'과 'ㄴ' 모음의 차이가 뚜렷하게 나타나고 있으나, 1980년대 이후부터는 두 모음의 차이가 현저히 줄어드는 것을 알 수 있다. 이와 같은 사실은 두 모음의 혼동이 1980년대 이후에 급격히 확산되어 갔음을 알 수 있다[13].

두 번째로 발성 음성 중 '예, 애'를 이용하여 2000Hz 이상의 고·저주파수를 제거하고 테스트하였다. (그림 3.3)에서 볼 수 있듯이 고·저주파수를 제거해도 각각 평균 약 65%, 63%의 인식률이 가능한 것으로 나타나고 있다.



(a)



(b)

(그림 3.3) 음성 'ㄱ(a), ㄴ(b)'의 필터링 후 음성 인식률(HP : 고주파 LP : 저주파)

세 번째로는 발성 음성 중 '의, 왜, 왜'를 이용하여 2000Hz 이상의 고·저주파수를 제거하고 테스트하였다. 고·저주파수를 제거한 태스크에서 각각 평균 약 62%, 65%, 64%의 인식률이 가능한 것으로 나타났다. 따라서, 본 논문에서는 이를 근거로 단모음으로 구분하는 /니/는 청음 음성학적 지식을 적용하여 이중모음(diphthong)인 유사음 /내, 게/와 같은 그룹으로 묶었고, 모음의 /H/도 청음 정보를 가지고 /게/와 묶었으며, /H/와 /키/도 같은 그룹으로 묶는 등 이들 7개의 음소를 각각 하나의 음가로 그룹화하여 아래 <표 3.2>과 같은 New KWPLU를 제안하였다.

<표 3.2> KWPLU 집합의 구성요소

자 음	PLU	설 명	모 음	PLU	설 명
ㄱ	gs	초성ㄱ	ㅏ	aa	
	ge	종성ㄱ	ㅑ	av	
ㄴ	ns	초성ㄴ	ㅓ	oo	
	ne	종성ㄴ	ㅕ	uu	
ㄷ	ds	초성ㄷ	ㅡ	ww	
	de	종성ㄷ	ㅣ	ii	
ㄹ	ls	초성ㄹ	ㅞ, ㅟ	ee	청음음성학적
	le	종성ㄹ	ㅚ, ㅜ, ㅠ	we	청음음성학적
ㅁ	ms	초성ㅁ	ㅜ	ui	
	me	종성ㅁ	ㅟ	ya	
ㅂ	bs	초성ㅂ	ㅛ	yv	
	be	종성ㅂ	ㅝ	yo	
ㅅ	sb	초성ㅅ	ㅠ	yu	
	se	종성ㅅ	ㅞ, ㅟ	ye	청음음성학적
ㅇ	ng	종성ㅇ	ㅑ	wa	
ㅈ	js	초성ㅈ	ㅑ	wv	
ㅊ	cs	초성ㅊ	ㅑ	wi	
ㅋ	ks	초성ㅋ			
ㅌ	ts	초성ㅌ	단복음	sp	short term
ㅍ	ps	초성ㅍ	장복음	sil	long term
ㅎ	hs	초성ㅎ			
ㄱ	gg	초성ㄱ			
ㄴ	dd	초성ㄴ			
ㅁ	bb	초성ㅁ			
ㅂ	ss	초성ㅂ			
ㅅ	jj	초성ㅅ			

3.2 종성 'ㅅ'의 실증적 검증[18]

“ㅅ'과 'ㄷ'의 음가를 같이 볼 것인가, 다르게 볼 것인가?” 하는 문제는 학계의 오랜 논쟁거리의 하나로 학자들마다 입장 차이를 보이고 있다. 동일하다고 보는 입장에는 허웅[18], 이기문[19], 이근수[20], 이은정[21], 이익섭[22] 등이 있으며, 두 개의 음가가 달랐다는 주장은 이기문[23], 허웅[24], 안병희[25] 등이 있다. 15세기 중세국어에서 받침 'ㄷ'과 'ㅅ'의 소리가 변별될 수 있을까? 하는 문제는 훈민

정음의 8종성에서 이들이 구별되어 있다는 점에서 음소적이던 형태적이던 변별되어 사용했던 것으로 추측되고 있다. 두 받침의 음가가 서로 달랐다는 편에 있는 이들의 주장은 먼저, 훈민정음에 구분하여 들어가 있으며, 둘째는 중세국어의 모습을 보여주는 「조선관역어」를 보면 「훈민정음」에 종성 'ㅅ'으로 나타나는 모든 예는 '사(思)'로 표기되어 있음에 반하여, 종성 'ㄷ'으로 나타나는 예들은 엄격히 구별되어 있어 이들이 변별되어 사용했음을 알 수 있다. 세 번째로는 15세기 국어에 있어서 당시 문헌에 'ㄷ'에 표기된 음절말 자음은 /ㄴ/위에서 동화되어 /ㄴ/으로 변한 예를 볼 수 있으나, 'ㅅ'으로 표기된 음절말 자음은 이런 동화를 보여주지 않는다[26]. 네 번째로 15세기 국어의 끝소리는 '간'(아내), '깁'(기둥)과 '깃'(닭의 등우리) 등이 구별되어 사용되었는데 이는 이시기에 이 낱말들이 각각 다르게 발음되었기 때문이다[27]. 이러한 근거를 들어 훈민정음 창제 무렵의 받침 'ㄷ'과 'ㅅ'은 표기에서는 물론 실제 발음에서 구별되었다고 주장하였다. 그러나, 이와 반대편에 있는 이들의 견해는 국어의 대표음규칙은 음절 말의 소리가 /k, p, t, l, m, n, ŋ/ 들 중 하나로 끝난다. 그리고 현대 국어에서 음절 말 'ㅅ'과 'ㄷ'은 /t/로 중화되며, 음절 초에서는 그 중화 현상이 일어나지 않는다.

- a. 발 + - 에서 → 발에서 /patesə~pasesə/
- 발 + - 이 → 발이 /pa.čhi~pasi/
- b. 꽃 + - 에서 → 꽃에서 /k'o.čhesə~k'osesə/
- 꽃 + - 이 → 꽃이 /k'o.čhi~k'osi/
- c. 숲 + - 에서 → 숲에서 /suphesə~subesə/
- 숲 + - 이 → 숲이 /suphi~subi/

그런데 위에서 '발'과 '꽃'과 '숲'은 각각 /pat/과 /k'ot/과 /sup/으로 발음되지만 이들이 모음으로 시작되는 형식 형태소와 연결될 때에는 원래의 음가를 잃고 /t<sup>h</sup>/와 /č<sup>h</sup>/가 /s/로, /p<sup>h</sup>/가 /b/로 발음되는 경향이 있다. 특히 /t<sup>h</sup>/가 /s/로 발음되는 현상에 유의해야 한다. 원래 모음으로 시작하는 형식형태소와 연결될 때 기저의 음소가 모두 발음되던 것이 변화되어 대표음이 발음되게 되었다고 해도 어말에서 /t<sup>h</sup>/의 대표음은 /s/가 아니라 /t/이기 때문에 문제가 된다. 이 문제는 '8종성 표기'와 '7종성 표기'에서 'ㄷ'과 'ㅅ'의 관계를 규명하는 문제와 관련이 있다. 또한 「훈민정음」, '예의'의 '종성해'에서는 “여빛곳위이화 열의갓위호피 이사자가이통용 고지용ㅅ자”라 하여 ‘ㅅ, ㅊ, ㅌ’ 받침을 ‘ㅅ’으로 쓴다고 하였는데 만약 종성 ‘ㅅ’의 음가를 /s/로 본다면 ‘ㅅ, ㅊ, ㅌ’이 종성에서 /s/로 중화된다는 설명이 되어 현대국어에서 종성 ‘ㅅ, ㅊ, ㅌ’이 /t/로 중화되는 현상과 거리가 멀다. 물론 중세국어 당시에는 ‘ㅅ, ㅊ, ㅌ’이 ‘ㅅ’과 같은 치음 계열이었기 때문에 현대국어에서보다는 중세국어 시기에 이들의 관계가 더 가까웠으리라는 짐작은 할 수 있다.

그러나 이들이 가지고 있는 자질을 통하여 중화 현상을 생각해 볼 때 ‘ㅈ, ㅊ’이 /s/로 중화되었을 가능성을 찾아 보기 힘들다. 중세국어의 ㅈ계 각자병서와 선행음절 중성의 ‘ㅈ’과 후행음절 초성의 ‘ㄱ, ㄷ, ㅂ, ㅅ’이 만난 형태(예, 샓기 → 새끼)가 현대국어에서 된소리로 발음되는 현상도 이 ‘ㅈ’의 음가가 여말에서 /s/가 아닌 /t/일 때에 가능하다. 그렇다면 음절말에서 변별을 잃고 중화되는 ‘ㅈ’과 ‘ㄷ’을 무엇 때문에 구별하여 썼을까 하는 문제가 발생하는데 지춘수[28, 29], 이인자[30]에서는 ‘ㄷ’은 설음을 대표하고 ‘ㅈ’은 치음을 대표하기 때문에 구별하지 않은 것이라고 주장하였다. 즉 사성칠음의 체계를 살리기 위해 같은 치음 계열인 ‘ㅈ, ㅊ, ㅊ, ㅈ’의 음절말 소리를 설음 계열인 ‘ㄷ’이 아니라 치음 계열인 ‘ㅈ’이 담당하게 했다는 것이다. 그런데 17세기에는 ‘ㅈ’을 ‘ㄷ’으로 대체하는 경향이 보이더니 18세기에는 반대로 ‘ㄷ’을 ‘ㅈ’으로 대체하는 경향이 두드러졌다.

- a. ㅈ: 중세국어의 기저형 ‘ㅈ’ /pat/ → 근대국어의 기저형 ‘ㅈ’ /pas/
- b. ㅊ: 중세국어의 기저형 ‘ㅊ’ /kot/ → 근대국어의 기저형 ‘ㅊ’ /kos/
- c. ㅊ: 중세국어의 기저형 ‘ㅊ’ /mit/ → 근대국어의 기저형 ‘ㅊ’ /mis/
- d. ㅈ: 중세국어의 기저형 ‘ㅈ’ /pat/ → 근대국어의 기저형 ‘ㅈ’ /pas/

중세국어에서 ‘ㄷ’ 받침이던 것이 근대국어에서 ‘ㅈ’ 받침으로 표기된 것은 위에서 보듯이 기저의 종성 /t/가 종성 /s/로 바뀌었다. 음절 말에서는 기저가 /s/로 바뀌었다더라도 불파에 의해 /t/로 실현되기 때문에 이 교체가 감지될 수 없지만 음절 초에서는 바뀐 기저음 /s/로 실현되기 때문에 그 기저음을 살려 음절 초의 위치에서만 아니라 음절 말의 위치에서도 기저형을 반영하여 ‘ㅈ’으로 표기하게 된 것이다. 실제로 현대 국어에서 기저 발음으로 ‘ㅈ’을 갖고 있는 ‘ㅈ, ㅊ, ㅊ’ 등은 모두 중세 시대에는 ‘ㄷ’ 종성으로 표기되었었는데 ‘7종성’ 표기에서부터 ‘ㅈ’으로 표기된 이래 그것이 계속 현대국어에까지 이어지는 것이다. 이익섭[22]은 이들의 종성 음가가 언제 /t/에서 /s/로 바뀌었는지 알 수 없다고 했지만 이들의 종성이 ‘ㅈ’으로 바뀌었기 때문에 표기에서도 ‘ㄷ’을 ‘ㅈ’으로 바꾼 것으로 생각된다.

본 논문에서는 이들의 오랜 논쟁인 ‘ㅈ’과 ‘ㄷ’에 대한 음가 동일성 여부 및 유효성 여부를 실제 음성 실험을 통해 검증하였다. 이를 위해 먼저, 음소의 분류를 초성 18개, 종성 17개로 하고 종성에는 한국어 7종성을 적용한 실험을 실시하고, 두 번째로는 ‘ㅈ’을 추가한 8종성가속음을 적용하여 실험하였다. 8종성가속음은 훈민정음 해례에서 종성에 쓸 수 있는 8개 자음을 정한 것으로 1933년 한글맞춤법 통일안이 제정되기 까지 사용되었던 것이다. 또한 한글맞춤법

통일안에는 7종성 발음만을 인정하도록 제정하였으나, 외래어 표기법[15]에는 ‘ㅈ’을 추가한 8종성을 적용토록 하고 있어 이를 살펴보고자 한다. 특히 최근 해외 기업의 국내 진입은 차후 증권시장에 까지 진출할 것으로 보여져 외래어 기업체명을 접하고 사용하게 될 것이다. 따라서, 외국 기업의 국내 증권시장에 진출을 앞두고 한글과 외래어 표기법에서 사용되는 7종성 중 ‘ㅈ’과 ‘ㄷ’을 서로 바꾸어 사용되고 있는 것에 대한 음가의 동일성을 검증하였다.

#### 4. 인식 실험 및 고찰

본 논문에서는 고립 단어 인식과 연속 음성 인식을 동시에 수행할 수 있는 증권거래 시스템을 구현하기 위해 상호, 가격, 수량, 목적어, 필러 모델로 핵심어 278단어와 이들로 구성된 71문장을 DB로 구축하였다.

##### 4.1 음성 데이터 구축 및 실험 환경

본 논문에서 사용하는 음성 데이터베이스는 국내 증권시장에서 통용되는 71 문장을 선정하고, 이를 남자 20명 여자 15명의 전화음성을 DB로 구축하였으며, 음성보드는 Dialogic의 D/120 JCT-LS Analog 12 Channel CSP(Continuous Speech Processing)음성보드를 사용하였다. 전화 음성은 mulaw로 하였으며, G.711의 규약에 의거하여 mulaw to linear PCM을 16bit로 변환한 음성 DB를 이용하였다. 다음 <표 4.1>은 입력 음성의 조건을 나타내고 있다.

<표 4.1> 입력 음성 데이터의 조건

설정 내용	설정 값
Sampling Rate	8000 Hz
Channel	Mono
Quantization	16 bit PCM (8bit alaw to linear)
음성 데이터 수	증권거래 71문장
화자 수	남성화자 20명, 여성화자 15명
전화녹음	A-law, Analog Line
환경	조용한 사무실
음성보드	D/120JCT-LS (Dialogic)

본 논문에서 사용된 단어와 문장의 일부를 <표 4.2>와 <표 4.3>에서 보여주고 있다. 이들을 통해 핵심어 검출 실험을 수행하였다.

<표 4.2> 핵심어 검출 실험을 핵심어 목록

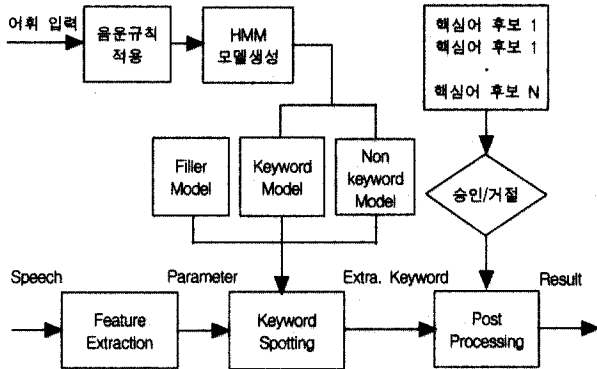
핵심어 유형	구성 어휘
상 호(Firm)	강원산업우    경남모직우    경수종금    골드금고    남성정밀
수 량(Amount)	구백오십주    구백주    백사십주    백주
가 격(Cost)	매도호가    매수호가    상한가    하한가
결 정(Decision)	매    도    매    수    얼    마    도    움    말

<표 4.3> 문장 데이터베이스의 일부 예

- 1) 하한가에 모아택을 매도하겠습니다
- 2) 상한가에 백주
- 3) 강원산업우 매도호가
- 4) 경수종금을 칠백주 매도하겠습니다
- 5) 골드금고를 매수하겠습니다

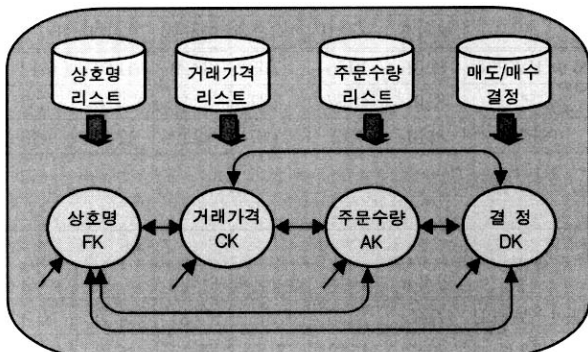
4.2 시스템 구성

인식 실험을 위해 증권거래 시스템을 구현하였으며, 시스템의 구성은 12차 LPC-based mel-cepstrum과 delta mel-cepstrum으로 이루어진 전체 24차의 특징 벡터를 사용하였다. 각 단어의 HMM 상태수는 단어의 평균 프레임 수에 비례하게 할당하였다. 핵심어 인식의 연결단어인식 알고리즘은 프레임 동기 one-pass viterbi search를 이용하였으며, 사용된 문법은 전이가 가능한 단어사이에서만 단어의 전이를 허용하는 word-pair 문법을 사용하였다. 전체적인 구성은 (그림 4.1)과 같다.



(그림 4.1) 핵심어 검출 시스템 구성도

(그림 4.2)는 구현한 증권거래 시스템을 실험하기 위한 발성 중 핵심어의 천이 과정을 나타내고 있다. 4개의 핵심어는 발성 순서에 상관없이 핵심어 그룹에 해당하는 것만 있으면 주식 거래가 가능하며, 여러 번에 나누어 발성할 수 있도록 하였다.



(그림 4.2) N Group 핵심어 발성 천이과정

5.3 인식 실험

5.3.1 '스'추가 8종성가족용 인식 결과

기존에 광운대학에서 사용하던 49개 음소로 구성된 KWPLU에 한글맞춤법 통일안의 7종성과 훈민정음 해례의 8종성가족용을 각각 반영하여 인식 성능을 실험하였다. 인식에 활용된 DB는 국어공학센터의 PBW 452를 이용하였으며, 참여한 50명이 5개조로 나뉘어 각 조별 고립단어를 각각 50개씩 발성하여 실험하였다. 인식 실험 결과는 아래 <표 5.1>에서 보여주고 있다.

<표 5.1> Old KWPLU에 '스'추가 8종성 인식 결과

조 별	인식률	7종성법 적용 인식률	8종성가족용 적용 인식률 (7종성법 + /스/종성)
MHJY0		89.33%	88.63%
MGSY0		87.52%	88.92%
MGSC0		92.73%	93.52%
MCYJ0		87.84%	91.25%
MBTG0		90.34%	92.70%
평 균		89.55%	90.21%

기존의 7종성을 8종성 가족용으로 분류하여 실험한 결과, 대체로 인식률이 평균 약 1.45% 상승하는 것으로 나타나고 있다. 실험 대상 중 4개조의 평균 인식률은 상승하였으나, 1개조는 약 0.7% 하락하는 것으로 나타났다.

5.3.2 청음 음성학 지식 기반 실험 결과

<표 5.2> /#,#,#/, /#,#/, /#,#,#/ 구분 인식 결과

(a)

조 별	인식률	Old KWPLU (7종성법)	New KWPLU (7종성법 + /스/종성)
MHJY0		89.33%	88.37%
MGSY0		87.52%	89.73%
MGSC0		92.73%	94.38%
MCYJ0		87.84%	89.86%
MBTG0		90.34%	92.93%
평 균		89.55%	91.05%

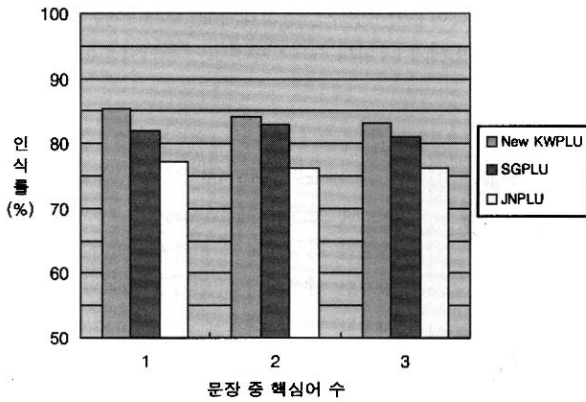
(b)

조 별	인식률	Old KWPLU (7종성법)	New KWPLU (7종성법 + /스/종성)
MHJY0		89.21%	92.11%
MGSY0		86.32%	88.56%
MGSC0		91.22%	94.38%
MCYJ0		90.22%	89.86%
MBTG0		90.22%	92.93%
평 균		89.55%	91.57%

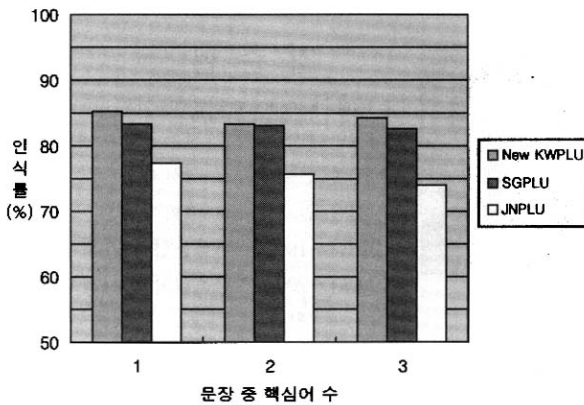
본 실험도 기존에 광운대학에서 사용하던 49개 음소로 구성된 KWPLU에 한글맞춤법 통일안의 7종성을 적용한 Old







(a) 남성 화자



(b) 여성 화자

(그림 5.1) 화자의 성별 PLU별 인식률

(그림 5.1)은 본 논문에서 제안한 New KWPLU와 타 유사 음소 단위와의 비교 실험 결과를 남·녀 화자로 나누어 보여주고 있다. 실험 결과 본 논문에서 제안한 유사 음소를 적용했을 때 남녀 모두 타 유사 음소 단위 대비 약 1~2% 정도의 인식률이 향상된 것을 알 수 있다.

### 7. 결 론

본 논문에서는 한국어 7종성과 8종성가족용에 속해 있는 자음의 음가에 대한 동일성 여부와 음성 인식에 기초가 되는 유사 음소 단위의 분류에 청음 음성학 도입 가능성에 대해 논하였다. 기존 유사 음소 49개의 KWPLU에 8종성가족용을 적용하여 고립단어 인식 실험하였으며, 청음 음성학의 도입은 스펙트로그램 상의 특정한 음향적 특성이 우리의 귀에 전달되는 단서를 찾기 위해 음성을 일부씩 분리하여 그 영향력을 알아 보았다. 이를 위해 본 연구에서는 음성의 일부분을 변경하거나 제거한 뒤 본래의 음성과 비교하여, 청자의 이해도를 알아보는 방법을 이용하였다. 음성을 일부 변경하거나 제거하기 위해 고·저주파 필터를 사용하였다. 8종성가족용 및 청음 음성학 지식 기반 유사 음

소 단위의 유효성을 검증하기 위해 증권거래 시스템을 구현하고 실험하였다.

실험은 고립단어와 연속 음성 인식으로 나누어 실시하였으며, 고립단어는 남·녀 50명이 10명씩 5개조로 나누어 각 조별 고립단어를 각각 50개씩 발성하여 실험하였다. 연속 음성 인식은 핵심어가 1~3개씩 포함하고 있는 71문장을 각각 2회씩 발성하여 실험하였다. 실험 결과 PBW452를 이용한 고립단어 인식에서는 종성으로 8종성가족용을 적용했을 때 평균 약 1.45%의 인식 성능이 향상되었으며, 8종성가족용과 청음 음성학을 적용한 유사 음소 단위를 적용할 때에는 평균 1.5~2.02%의 인식률 향상을 확인하였다. 연속 음성에 대한 인식 실험에서는 기존의 49개 또는 56개 유사 음소 단위를 이용한 경우보다 평균 약 1~2%의 인식 성능 향상을 확인하였다.

따라서, 8종성가족용에 있는 자음 /ㄷ/와 /ㅌ/에 대해 유사 음소 단위를 분리했을 때 약간의 인식률이 상승하는 결과로 미루어 이들의 음가가 완전히 동일하지는 않는 것으로 보이며, 청음 음성학을 적용한 모음 /ㅞ/, /ㅟ/, /ㅠ/, /ㅡ/, /ㅢ/, /ㅣ/, /ㅤ/는 실험 결과로 볼 때 음성 인식을 위하여 음성이 갖고 있는 모든 고·저주파수의 성분 음들을 필요로 하지는 않으며, 좁은 주파수대에서도 인식이 가능하다는 것을 실증하였다. 앞으로 청음 음성학을 좀 더 깊이 연구할 계획이다.

### 참 고 문 헌

- [1] Jay G. Wilpon, Lawrence R. Rabiner, Chin-Hui Lee, E. R. Goldman, "Automatic Recognition of Keyword in Unconstrained Speech Using Hidden Markov Models," IEEE Trans. Acoust., Speech, Signal Processing, Vol. 38, No.11, pp.1870-1878, Nov., 1990.
- [2] 오영환, "음성언어정보처리", 홍릉과학출판사, 1998.
- [3] 이경남, "의사 형태소 단위의 한국어 연속 음성 인식", 서강대학교 석사학위 논문, 1997.
- [4] 문교부고시 제88-2호, "국어 어문 규정집", 문화체육부, 1988.
- [5] S. J. Young, N. H. Russel, J. H. S Thornton, "Token passing: a simple conceptual model for connected speech recognition systems," Cambridge University Engineering Department, 1989.
- [6] Lawrence Rabiner, Biing-Hwang Juang, "Fundamentals of speech recognition," Prentice-Hall, 1993.
- [7] Daniel Jurafsky, James H. Martin, "Speech and Language Processing," Prentice Hall, 2000.
- [8] 유승덕, 김학진, 김순협, "한국어 자소 음가 분류에 관한 연구", 한국음향학회지, 제20권 제2(s)호, pp.89-92, 2001.
- [9] 윤성희, "음성 언어인식을 위한 사전참조 및 후처리에 관한 연구", 상명대학교, 1999.

- [10] 이활림, "음소 HMM을 이용한 핵심어 검출 시스템의 성능 향상에 관한 연구", 부산대학교 석사학위논문, 1996.
- [11] 신지영, "말소리의 이해 ; 음성학 · 음운론연구의 기초를 위하여", 한국문화사, 2000.
- [12] 서봉수, "가변어휘 음성 인식기 구현 및 탐색기간 단축 알고리즘 비교", 석사학위논문, 전남대학교, 2001.
- [13] 정명숙, "음성 자료에 나타난 국어의 사적 변천", 고려대 민족문화연구원 국어연구소, 2002.
- [14] 여재열, "한글 종성 표기의 변천에 관한 연구", 홍익대학교 석사학위논문, 1993.
- [15] 문교부고시 제85-11호, "외래어 표기법", 문화체육부, 1988
- [16] 이용재, "영어 음성학", 고려대학교, 2000.
- [17] 이동석, "구개음화의 어휘화와 'ㅅ' 종성에 대하여", 한국어학회, 제6권, pp.9-15, 1997.
- [18] 허 용, "국어음운론", 정음사, 1958.
- [19] 이기문, "십육세기 국어의 연구", 문리논총(고려대), 1959
- [20] 이근수, "ㄷ, ㅅ 종성에 대하여", 탑출판사, 1986.
- [21] 이은정, "8종성에서의 '-ㅅ'에 대하여", 한글, 1986.
- [22] 이익섭, "음절말 표기 'ㅅ'과 'ㄷ'의 사적 고찰", 성곡논총, p.18, 1987.
- [23] 이기문, "국어사개설", 민중서관, 1961.
- [24] 허 용, "국어음운론", 정음사, 1965.
- [25] 안병희, "십오세기 국어의 활용어간에 대한 형태론적 고찰", 국어연구, p.7, 1959.
- [26] 이기문, "국어표기법의 역사적 고찰", 한국연구원, 1963.
- [27] 허 용, "국어음운론", 정음사, 1987.
- [28] 지춘수, "ㅅ 종성 재론", 한글, 1971.

- [29] 지춘수, "국어표기사연구", 경희대학교 박사학위논문, 1986.
- [30] 이인자, "15세기 국어의 'ㄷ·ㅅ'종성고", 동학어문논집, p.20, 1985.



**김학진**

e-mail : kimhj@shinbo.co.kr  
 1986년 건국대학교 전자계산학과(학사)  
 1997년 연세대학교 전자공학과(석사)  
 1998년~현재 광운대학교 컴퓨터공학과 박사과정  
 1986년~현재 신용보증기금 근무

2000년~현재 명지전문대 겸임교수  
 관심분야 : 음성신호처리, 음성인식, 대화처리시스템, 자연어처리

**김순협**



e-mail : kimsh@daisy.gwu.ac.kr  
 1974년 울산대학교 전자공학과 학사  
 1976년 연세대학교 석사과정  
 1983년 연세대학교 공학박사  
 1998년~1999년 한국음향학회 회장  
 1999년~2003년 광운대학교 사회교육원 원장

1979년~현재 광운대학교 컴퓨터공학과 교수  
 2000년~현재 한국음향학회 명예회장  
 2003년~현재 광운대학교 정보과학기술대학원 원장  
 관심분야 : 음성신호처리, 음성인식, 자연어처리, 3-D 멀티미디어