한국의 지하철역명을 위한 음소 기반의 음성인식에 관한 연구

A Study on Speech Recognition based on Phoneme for Korean Subway Station Names

김범승[†] · 김순협¹

Beom-Seung Kim · Soon-Hyob Kim

Abstract This paper presented the method about the Implementation of Speech Recognition based on phoneme considering the phonological characteristic for Korean Subway Station Names. The Pronunciation dictionary considering PLU set and phonological variations with four Case in order to select the optimum PLU used for Speech Recognition based on phoneme for Korean Subway Station Names was comprised and the recognition rate was estimated. In the case of the applied PLU, we could know the optimum recognition rate(97.74%) be shown in the triphone model in case of considering the recognition unit division of the initial consonant and final consonant and phonological variations.

Keywords : Subway Station Names, Recognition Unit, Pronunciation Dictionary, Speech Recognition

초 록 본 논문에서는 한국의 지하철역명을 위하여 음운론적 특성을 반영한 음소 기반의 음성인식 구현에 관한 방법을 제시하였다. 한국의 지하철역명의 음소 기반의 음성인식을 위하여 사용되는 최적의 유사음소 단위(PLU: Phoneme-Likely Unit)를 선정하기 위하여 네 가지의 Case 별로 PLU set과 음운 현상을 고려한 발음사전을 구성하여 인식률을 평가하였다. 적용된 유사음소 단위의 경우 초성과 종성 자음의 인식 단위 구분 및 음운 현상을 반영한 경우 트라이폰 모델에서 최적의 인식률(97.74%)을 보임을 알 수 있었다.

주요어 : 지하철역명, 인식단위, 발음사전, 음성인식

1. 서 론

지하철을 이용하기 위한 자동발매기[1], 전화망에 서의 열차시각·노선·운임·소요시간 제공을 위한 지하철역 정보 안내서비스 등에 음성인식 적용시 가장 먼저 고려해야 할 대상이 지하철 역명에 대한 음성인식 DB를 구축하는 것이다. 선정해야 하는 대상어휘 중 지하철역명은 가장 기본적인 항목이다. 현재 코레일, 서울메트로, 인천메트로, 서울도시철도에서 운영하는 수도권을 중심으로 한 광역지하철역은 454개이다[2-5]. 지하철역명을 위한 음성인식 시스템 구축시 역명인식을 모델링하기 위하여 인식 단위를 무엇으로 할 것인가를 고려하는 것은 중요한 문제이다. 음성인식에서 인식 단위는 해당언어의 특징뿐만 아니라 음소문맥에 따른 다양성을 고려하여야 하며 최적의 모델링 단위를 선택하는 것이 무엇보다도 중요하다[6]. 이를 위하여 음소 단위의 트라이폰 모델을 사용하여 음성인식 DB를 구축하였다.

이러한 트라이폰 모델을 구성할 때 너무 작은 단위로 세밀하게 모델링 하게 되면 정교한 형태의 모델을 얻을 수 있지만 반면에 모델링 단위가 커지게 되고 더 많은 양의 음성 DB가 필요하게 되며 HMM(Hidden Markov Model) 파라미터 수를 증가시켜 실제 시스템에서의 복잡도가 한층 높아지고 그 성능이 오히려 저하될 수 있다[7,8]. 이러한 문제

점을 극복하기 위하여 한국의 지하철역명의 음소 문맥에서 음운론적 특성을 고려한 최적의 음소 단위를 선택하기 위하여 각각의 모델과 케이스별로 PLU와 발음사전을 구성하여 인식률 실험을 하였다. 본 논문에서는 실제 인식시스템에서의 이러한 실험 과 평가를 통하여 한국의 지하철역명을 위한 음성학적·음운론적 지식기반의 초성·종성 자음의 인식 단위 구분과 음운 현상 중 경음화·비음화 현상이 반영된 가장 최적의 PLU 및 발음사전을 제시하고자 한다. 2장에서는 한국어의 특징 및 인식 단위에 대한 소개를 3장에서는 이러한 특징을 반영한 최적의 PLU 선정을 위한 분석을 4장에서는 실험 및 평가를 하였고 5장에서 결론을 내렸다.

2. 한국어의 특징 및 인식 단위

2.1 한국어의 특징 및 인식 단위

음성 언어 인식의 모델링을 위한 기본단위로 단어, 음절, 음소, PLU 등을 사용할 수 있다. 음성 언어 인식 측면에서 보면 기본 발음단위가 하나의 인식 단위로 사용 가능하다. 따라서 한국어는 독립해서 발음할 수 있는 최소 단위인 음절을 기본 발음단위로 생각 할 수 있으며, 또한 형태소적으로 음소 단위에서 결합이 가능하여 음소를 기본 발음단위로 볼 수도 있다. 하지만 한국어의 경우 음절의 종류와 수가 많고 초성, 중성, 종성의 결합이 다양하여 음절을 인식 단위로 사용하기는 거의 불가능하다. 이러한 인식 단위 외에 외국에서 많이 쓰이는 있는 다이폰이나 트라이폰이 있지만, 기

[†]교신저자 : 한국철도공사 정보기술단

E-mail : bluedav@korrail.com

¹광운대학교 컴퓨터공학과

Fig. 1 Triphone model

본 음소의 조합으로 이루어져 그 개수가 음소 개수의 제곱, 세제곱에 이르므로 한국어의 음성인식 시스템에 적용되기 위해서는 매우 많은 학습용 음성 데이터가 필요하다[6-8]. 이중 트라이폰(Triphone)의 유사음소 단위의 문맥 종속 단위가 우수한 인식 단위로 알려져 있다[9,10]. 트라이폰의 유사음소 단위의 상태별 구성은 Fig. 1과 같이 나타낼 수 있다.

음소(Phoneme)는 단어에서 뜻을 구분 할 수 있는 음운의 최소단위로서 한국어의 경우 자음 19개, 모음 20개로 구성된다. 음소는 환경에 따라 다른 음소로 바뀌거나 없어지는 등 변동을 겪는데 이러한 음소의 변동을 음운 현상이라고 하며 변화 양상에 따라 크게 한 음소가 다른 음소로 바뀌는 “대치(비음화, 유음화, 경음화 등)”, 한 음소가 없어지는 “탈락(자음군 단순화, 후음 탈락, 유음 탈락 등)”, 없던 음소가 새로 끼어드는 “첨가(ㄴ-첨가, 활음 첨가)”, 두 음소가 합쳐져 다른 음소로 바뀌는 “축약(유기음화, 모음축약)”, 두 음소가 서로 자리를 바꾸는 “도치(자음 도치, 모음 도치)”의 다섯 가지로 나눌 수 있다. 한국어에 존재하는 이러한 다섯 가지 유형의 음운 현상에 의해서 하나의 형태소가 여러 발음을 가질 수 있고, 여러 형태소가 하나의 발음이 될 수 있다[11]. 이러한 형태소와 발음과의 사상(Mapping)에는 단순한 사상(Mapping) 이상의 정보가 필요하다[12]. 따라서 한국어 문서의 형태소 분석에 사용되는 형태소 사전을 음성인식에 그대로 적용할 수 없고 음소의 변동을 표시할 수 있는 사전구조가 필요하다[13-15]. 이러한 문제를 해결하기 위하여 인식실험을 통한 적절한 PLU의 선택 및 음운 현상을 고려한 발음사전의 구성이 필요하다.

2.2 한국의 지하철역명의 특징 및 인식 단위

역명에는 여러 종류가 있으며 소단위로 지하철 역명이 있고, 중간 단위로 국철 역명이 있고, 가장 큰 단위로 고속철도 역명이 있다. 우리는 전통적으로 역명을 제정하는데 속지주의 원칙에 따른 지명을 취해 역명을 지어왔다[16]. 열차를 가장 많이 타고 내리는 승객이 그 역이 위치한 지역의 거주민이 때문에 역이 속한 지명을 역명으로 사용하게 된다. 그리고 타 지역의 사람들도 여러 가지 목적으로 이곳을 왕래하기 위하여 타고 내리기 때문에 이곳 지명을 역명으로 사용해야 손쉽게 찾아 올 수 있는 장점이 있기도 하다[16]. 광역지하철역명의 경우 그 정착장이 기존 역의 지하일 경우에는 서울역, 청량리역, 용산역 등과 같이 기존 역명을 그대로 쓰고 있다. 그렇지 않을 경우에는 대개 종로3가역, 을

지로3가역, 충무로역 등과 같이 소(小) 지명을 택한다. 국철 역명의 경우 먼 단위 이상의 중(中)·대(大) 지명을 쓰며 고속철도의 경우 광역지명(大)을 쓰게 된다[16,17]. 즉 사실상 역명은 모두 지명인 고유명사이다. 한국의 광역지하철역명의 경우 위 사례에서와 같이 대부분이 소단위 지명이며 2자(음절)가 63%, 3자가 22%, 4자가 8%, 5자가 5%, 6자 이상이 2%를 차지하고 있다. 또한 초성이 자음으로 시작되는 역명이 83%이고 초성이 모음으로 시작하는 역명이 17%를 차지하고 있다. 그리고 몇 개의 음소가 모여서 이루어지는 음절의 구조는 다양하지만, 광역지하철역명의 경우 음절구성을 위한 음소배열 유형을 살펴보면 CVC가 49.8%, CV가 33.7%, VC가 11.7%, V가 4.8%를 나타내고 있다. 일반적으로 대부분의 언어는 하나의 자음에 하나의 모음으로 구성된 것이 가장 자연스럽다. 따라서 개별언어인 한국어도 CV가 가장 자연스러운 형태이다. 15세기 당시 표음문자인 훈민정음을 표기하는데 있어 연철 방식을 택한 것을 보아도 CV가 가장 안정적이고 자연스러운 형태임이 분명하다[18]. 하지만 영어와 달리 한국어는 /야탐/, /오류동/, /이수/ 등과 같이 모음 하나만을 가지고 음절을 이룰 수 있기 때문에 두 번째로 자연스러운 음절은 V이며 폐음절을 만드는(C)VC가 그 다음이다. 따라서 선호도가 가장 낮은 C₁VC₂에서 C₂에 오는 미파음은 음향학적 자질 보다는 어휘부에 저장된 어휘와 음운규칙에 의거해 인지하게 되는 것이다[18]. 하지만 위에서 살펴 보았듯이 수도권 지하철 역명의 경우의 선호도는 CVC > CV > VC > V이다. 이것은 역명이 지명인 고유명사인 이유도 있긴 하지만 첫 음절의 종성에 위치하는 자음과 그 다음 음절에 오는 초성이 자음인 경우가 많음을 의미하며 음향학적인 자질 보다는 음운론적 자질의 영향을 받을 수 있음을 알 수 있다. 따라서 지하철역명의 발음시 발생하는 자음의 초성, 종성 구분 및 음운 현상을 고려하여 음소 기반의 인식 단위를 선정할 필요가 있다.

3. 최적의 PLU 선정을 위한 분석

3.1 음성·음운론적 특성을 고려한 유사음소 단위(PLU) 선정

한국어 자음의 경우 19개의 초성과 27개의 종성으로 구성되나 음절구조 제약에 의해 종성의 자리에 /ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅇ/만(음절의 끝소리 규칙)이 올 수 있으며 인접 음소에 따라 초성과 같은 음자로 공유된다[8,11]. 자음은 공기가 입 안에서 장애를 받아 만들어지는 소리이며 발음할 때에는 마찰이나 폐쇄 등 다양한 장애가 일어난다. 또한 홀로 발음되지 못하고 모음에 의지해서 발음되는 닿소리이며 이것은 스스로 음절을 이룰 수 없음을 의미한다[11]. 본 논문에서는 초성·종성 자음의 인식 단위를 달리하는 부분에 대해서는 /ㄱ, ㄴ, ㅂ, ㅅ/을 고려하였다. Fig. 2, 3, 4는 초성·종성 자음의 인식 단위를 달리하는 /교간-옥수/, /구로-구일/, /보라매-압구정/의 파형과 스펙트로그램이다.

음소와 음소가 만나면 충돌이 일어난다. 서로 다른 특성을 지닌 음소들이 만남으로 충돌이 일어나는 것은 당연하다.

Fig. 2 Wave and Spectrogram of /고잔-옥수/

Fig. 3 Wave and Spectrogram of /구로-구일/

Fig. 4 Wave and Spectrogram of /보라매-압구정/

Table 1 PLU and Pronunciation Dictionary Example

자음	표기	예 제	설명
ㄱ	g	강남 g a n k n a m, 강동 g a n k d o n k, 고대 g j o d e, 제기동 z e g i d o n k	초성 ㄱ
ㄲ	gq	간석 g a n s y g q, 녹천 n o g c y v n, 성북 s y n k b u g q, 안국 a n g u g q	중성 ㄱ
ㄴ	r	계동 g e r o n k, 구로 g u r o, 미아섬거리 n i a s a n g v r i, 운문로 c u n k m u r o	초성 ㄴ
ㄷ	l	구발 g u l i, 구파발 g u p a b a l, 제물포 z e m u l p o, 하여울 h a g y u l	중성 ㄴ
ㅂ	b	강변 g a n b j v n, 계동 g e b o n k, 대방 d e b a n k, 방배 b a n k b e	초성 ㅂ
ㅃ	bq	신당 s i n d a b q, 아랍 j a t a b q, 용답 j o n d a b q, 주입 z u j v b q	중성 ㅂ
ㄸ	gg	경북궁 g j v n x b g g u n k, 압구정 a b q g u z v n k, 역곡 j v g u g g o q	ㄱ→ㄸ(경음)
ㄹ	dd	길동길 g i l d o n k, 목동 m o g d d o n k, 숙대입구 s u g d d e i b q g u	ㄷ→ㄹ(경음)
ㅅ	bb	녹면 n o g b b v n, 독바위 d o g b b a v i	ㅂ→ㅅ(경음)
ㅆ	ss	가락시장 g a r a g s s i z a n k, 독산 d o g s s a n, 수락산 s u r a g s s a n	ㅂ→ㅅㅅ(경음)
ㅈ	zz	무학동 m u a g z z e, 을지로7가 y z l i r o s e a n k, 합정 h a b z z v n k	ㅈ→ㅈㅈ(경음)
ㄴ	n	목달문 d o n k n i m a u n, 청량리 c y v n k n a n k n i, 창신리 w a n x s i n n i	ㄴ→ㄴ(비음)

Fig. 5 Wave and Spectrogram of /독립문/

중 음소 문맥에서 발생할 수 있는 경음화, 비음화를 고려하였고 그 예제와 내용은 Table 1과 같다.

Fig. 5의 좌측은 /독/, /립/, /문/을 각각 발음하여 /독립문/을 합성한 것이고 우측은 /독립문/을 발음한 파형이다. 이 경우 인위적으로 합성한 경우 음소간의 충돌을 무마하지 못하여 /독이븐/(4명), /독이플/(3명), /독이뿔/(3명)로 들렸다(10명의 20·30대 남녀에게 들려 주었을 때의 인지 실험). 하지만 정상적으로 발음한 경우 /독립문/→/동닙문/으로 비음화 되어 들리는 것을 알 수 있다. 이는 음소 문맥 사이의 음소의 변동을 반영하는 것이 음운·음향학적 상관 관계를 명쾌하게 설명하기는 힘들지만 음성을 인식함에 있어 영향을 줄 수 있음을 알 수 있다.

수도권 1~9호선의 광역지하철역명의 음소에 대한 빈도수 및 그 분포를 분석해 본 결과는 아래 Fig. 6과 Table 2와 같다. n/ㄴ/의 경우가 빈도수가 제일 높으며 전체 중 17.33%를 차지하고 있고 wi/ㄱ/가 0.09%로 가장 작다. 이중 발성시 초성·중성 자음의 인식 단위 구분에 영향을 줄 수 있는 초성 g/ㄱ/, 중성 gg/ㄱ/, 초성 r/ㄹ/, 중성 l/ㄹ/, 초성 b/ㅂ/, 중성 bq/ㅂ/의 경우 14.9% 이고 경음화·비음화 현상을 발생시킬 수 있는 gg/ㄱ/, dd/ㄷ/, bb/ㅃ/, ss/ㅆ/, zz/ㅈ/, n/ㄴ/이 차지하는 비율이 2.7%를 나타내고 있다. 이런 고려에 대한 예제는 Table 1과 같고 이는 초성·중성 자음의 인식 단위 구

지하철역명의 음소에 대한 빈도수

Fig. 6 The frequency of phoneme

Table 2 The frequency of phoneme

음소	g	gq	n	d	r	l	n	b	bq	s
빈도수	127	54	390	95	43	84	89	55	22	137
분포	5.64%	2.40%	17.83%	4.22%	1.91%	1.51%	8.95%	2.44%	0.98%	6.09%
음소	nx	z	c	t	p	h	gg	dd	bb	ss
빈도수	209	61	46	11	19	43	19	8	3	17
분포	9.28%	2.71%	2.04%	0.49%	0.84%	1.91%	0.84%	0.36%	0.13%	0.76%
음소	zz	a	ja	v	jv	o	jo	u	ju	y
빈도수	7	182	10	113	36	147	9	99	4	25
분포	0.31%	8.09%	0.44%	5.02%	1.60%	6.53%	0.40%	0.18%	0.18%	1.11%
음소	i	e	je	wa	we	wi	wv	yi		
빈도수	109	72	7	25	4	2	21	6		
분포	4.84%	3.20%	0.31%	1.11%	0.18%	0.09%	0.93%	0.27%		

분과 경음화·비음화 현상이 인식률에 영향을 미치는 정도가 17.6% 정도의 영향도가 있음을 알 수 있다. 위 사항을 고려한 실험을 거쳐 최적의 인식률을 나타내는 기본 유사음소 단위인 PLU set은 Table 6과 같다.

3.2 음소 단위의 모델링

본 논문에서는 객관적인 실험을 하기 위하여 널리 사용되고 있는 영국의 CAM(University of Cambridge)에서 공개하고 있는 HTK(Hidden Markov Model Tool Kit)를 사용하여 인식 실험을 수행하였다[19,20]. HTK는 전 세계적으로 널리 그 인식 성능이 검증된 신뢰가 높은 인식 툴이다. 인식기를 구축하기 위하여 학습에 필요한 20대 남성 30명과 여성 30명이 322개 역명[1,20]을 녹음한 여취를 사용하였다. 음성 파일로부터 특징추출을 하기 위하여 39 MFCC(Mel Frequency Cepstral Coefficients)를 사용하였으며 CHMM(Continuous Hidden Markov Model)을 이용한 음향모델을 구축하였다[1,21,22]. 인식 단위를 음소로 할 경우 단어나 음절에 비하여 그 수가 적고 학습에 필요한 충분한 자료를 모으기가 용이한 반면에 음소 앞뒤에 위치하는 음소에 영향을 많이 받으므로 이를 고려하여 세분화된 문맥중속 음소모델(sub-word unit modeling)을 구성하여야 한다. 이전 연구들[8,14,15]에서 문맥독립 음소는 문맥중속 음소에 비하여 많은 변이를 포함하고 있으므로 모델링이 어려워지고 인식률도 낮게 나타났다. 따라서 문맥독립 음소모델을 사용할 경우 인식 단위에 대한 정확한 모델링 뿐만 아니라 분별학습,

Table 3 PLU Set of Case 1

구분	31 PLU Set				
자음	g/ㄱ/	n/ㄴ/	d/ㄷ/	r/ㄹ/	m/ㅁ/
	b/ㅂ/	s/ㅅ/	nx/ㅇ/	z/ㅈ/	c/ㅊ/
	t/ㅌ/	p/ㅍ/	h/ㅎ/		
모음	a/ㅏ/	ja/ㅑ/	v/ㅓ/	jv/ㅕ/	o/ㅗ/
	jo/ㅛ/	u/ㅜ/	ju/ㅠ/	-/ㅛ/	i/ㅣ/
	e/ㅔ/·ㅖ/	je/ㅖ/	wa/ㅘ/	ve/ㅙ/	vi/ㅚ/
	wv/ㅝ/	yi/ㅣ/			
목음	Sil				

Table 4 PLU Set of Case 2

구분	34 PLU Set				
자음	g/ㄱ/	n/ㄴ/	d/ㄷ/	r/ㄹ/	m/ㅁ/
	b/ㅂ/	s/ㅅ/	nx/ㅇ/	z/ㅈ/	c/ㅊ/
	t/ㅌ/	p/ㅍ/	h/ㅎ/	gg/ㄲ/	dd/ㅌ/
	ss/ㅆ/				
모음	a/ㅏ/	ja/ㅑ/	v/ㅓ/	jv/ㅕ/	o/ㅗ/
	jo/ㅛ/	u/ㅜ/	ju/ㅠ/	-/ㅛ/	i/ㅣ/
	e/ㅔ/·ㅖ/	je/ㅖ/	wa/ㅘ/	ve/ㅙ/	vi/ㅚ/
	wv/ㅝ/	yi/ㅣ/			
목음	Sil				

Table 5 PLU Set of Case 3

구분	36 PLU Set				
자음	g/ㄱ/	n/ㄴ/	d/ㄷ/	r/ㄹ/	m/ㅁ/
	b/ㅂ/	s/ㅅ/	nx/ㅇ/	z/ㅈ/	c/ㅊ/
	t/ㅌ/	p/ㅍ/	h/ㅎ/	bb/ㅃ/	ss/ㅆ/
모음	a/ㅏ/	ja/ㅑ/	v/ㅓ/	jv/ㅕ/	o/ㅗ/
	jo/ㅛ/	u/ㅜ/	ju/ㅠ/	-/ㅛ/	i/ㅣ/
	e/ㅔ/·ㅖ/	je/ㅖ/	wa/ㅘ/	ve/ㅙ/	vi/ㅚ/
	wv/ㅝ/	yi/ㅣ/			
종성 자음	gq/ㄱ/	l/ㄹ/	bq/ㅂ/		
목음	Sil				

Table 6 PLU Set Case 4

구분	39 PLU Set				
자음	g/ㄱ/	n/ㄴ/	d/ㄷ/	r/ㄹ/	m/ㅁ/
	b/ㅂ/	s/ㅅ/	nx/ㅇ/	z/ㅈ/	c/ㅊ/
	t/ㅌ/	p/ㅍ/	h/ㅎ/	gg/ㄲ/	dd/ㅌ/
	bb/ㅃ/	ss/ㅆ/	zz/ㅈ/		
모음	a/ㅏ/	ja/ㅑ/	v/ㅓ/	jv/ㅕ/	o/ㅗ/
	jo/ㅛ/	u/ㅜ/	ju/ㅠ/	-/ㅛ/	i/ㅣ/
	e/ㅔ/·ㅖ/	je/ㅖ/	wa/ㅘ/	ve/ㅙ/	vi/ㅚ/
	wv/ㅝ/	yi/ㅣ/			
종성 자음	gq/ㄱ/	l/ㄹ/	bq/ㅂ/		
목음	Sil				

후처리 등의 충분한 뒷받침 없이는 높은 인식률을 기대하기 어렵다[14,15]. 반면 문맥중속 음소모델은 문맥독립 음소모델에 비해 음향의 가지수는 많지만 음소에 의한 변이음을 고려한 모델로서 강건한 음향모델을 생성하는 방법으로 많이 사용되고 있다[8,23]. 유사음소 단위는 최소 인식단위로 많

이 사용되며 기본적인 음소에 변이음을 포함하고 있는 음소이다. 음성학적 및 음향학적 유사성이 큰 경우에는 음소와 유사음소 단위는 동일하게 취급될 수 있지만 그렇지 않은 경우에는 큰 차이가 있다. 본 논문에서는 최적의 인식단위의 PLU 구성 실험을 위해 유사음소 단위 집합을 Table 3,4,5,6과 같이 4가지의 경우로 구분하여 음향모델을 작성하였다.

4. 실험 및 평가

4.1 실험도구 · 환경

- 1) 실험에 사용한 PC : Window Vista 환경의 Notebook LG Xnote E200,
- 2) 녹음용 Microphone : Sony ECM-221 Condenser Microphone & Xnote E200 내장 마이크,
- 3) 음성파형 분석 툴 : Praat[24],
- 4) 분석대상 어휘 : [1]에서의 대상 역명(314)에 8개 역명(개룡,국수,녹양,덕계,덕정,보산,양주,지행)을 추가한 322개 어휘,
- 5) 음성분석 조건 : Sampling Rate 8000Hz, 양자화 레벨 16 bit.

4.2 PLU 구성 별 실험

본 논문에서는 최적의 유사음소 단위를 선택하기 위하여 PLU set을 Table 7과 같이 4가지 Case 별로 구성 하였고 모노폰, 트라이폰 단위의 학습용 발음열과 발음사전을 구성하여 인식을 실험을 진행 하되 모노폰, 트라이폰 기준으로 실험을 진행하였고 인식을 평가를 위해 학습에 참여하지 않은 20명의 화자가 녹음한 음성파일을 사용하였다.

그 수행 결과는 아래 Fig 7, Table 8과 같다. 제안된 4가

Table 7 PLU set with four Case

구 분	내 용	비 고
Case1(C1)	초성 종성의 자음 구분이 없고 음운 현상을 고려하지 않은 경우의 31 PLU set	Table 3
Case2(C2)	초성 종성의 자음 구분이 없고 음운 현상이 반영된 경우의 34 PLU set	Table 4
Case3(C3)	초성 종성의 자음 구분이 있고 음운 현상이 적용되지 않은 경우의 36 PLU set	Table 5
Case4(C4)	초성 종성 자음의 구분과 음운 현상을 모두 고려한 39 PLU set	Table 6

Fig. 7 Recognition Rate with four Case

Table 8 Recognition Rate with four Case

Recog Units	C1	C2	C3	C4
Tri.Rate	93.48%	93.86%	94.79%	97.74%
Mono.Rate	50.00%	52.48%	55.28%	55.59%

지 Case의 경우 모노폰 구성 C1과 C2는 2%정도의 차이가 있으나 C3와 C4의 경우에는 미미한 차이를 보였다. 하지만 트라이폰 구성의 경우 C3와 C4의 경우 3% 이상의 차이를 보였다. 이는 초성·종성 자음의 인식 단위 구분과 음운 현상을 고려한 Case 4가 트라이폰 단위에서 그 인식률이 가장 높음을 알 수 있다.

5. 결 론

본 논문에서는 한국의 지하철역명의 음소기반의 음성인식을 위하여 사용되는 최적의 유사음소 단위를 선택하기 위하여 네 가지의 Case 별로 PLU set을 구성하였고 Case 별 인식 성능을 평가하고 최적의 (초성·종성 자음의 인식 단위 구분과 음운 현상이 고려된) 인식 단위인 PLU를 선별하였다. 적용된 유사음소 단위의 경우 초성·종성 자음의 인식 단위 구분 및 음운 현상을 반영한 경우 트라이폰 모델에서 최상의 인식률을 보임을 알 수 있었다. 향후 전국 기차역 및 광역시별 지하철역의 음성인식 DB를 구축하여 다양한 방식으로 음소문맥의 경계에 존재하는 음운 현상을 반영한 음성 인식 시스템을 구축하기 위하여 연구할 계획이다.

후 기

본 연구는 광운대학교 2011학년도 교내 학술 연구비 지원으로 이루어졌습니다.

참고문헌

- [1] B.S. Kim, S.H. Kim (2009) A Study on the Speech Recognition for Commands of Ticketing Machine using CHMM, *Journal of the Korean Society for Railway*, 12(2), pp. 285-290.
- [2] Korail, <http://www.korail.com>, accessed on 20 January 2011.
- [3] Seoul Metro, <http://www.seoulmetro.co.kr>, accessed on 20 January 2011.
- [4] Incheon Metro, <http://www.incheonmetro.co.kr>, accessed on 20 January 2011.
- [5] Seoul Metropolitan Transit Corp, <http://www.smrt.co.kr>, accessed on 20 January 2011.
- [6] K.N. Lee, M.H. Chung (2002) Discriminative Allophone Rules for Optimizing Pronunciation Dictionary in Korean LVCSR, *KSCSP, Acoustical Society of Korea*, 19(1).
- [7] H.J. Kim, M.H. Chung (2002) Korean Continuous Recognition Using Phonological Context and Crossword Phonological Variations, *KSCSP, Acoustical Society of Korea*, 19(1).
- [8] G.H. Shen, H.J. Seo, S.J. Hahm, J.G. Kim, et al. (2004) A

- Study on Phone-Like Units for Korean Continuous Speech Recognition in Noisy Environments, *2004 Autumn Conf. Acoustical Society of Korea*, 23(2).
- [9] R.D.R. Fagundes, J.S. Correa, P. Dumouchel (2002) A New Phonetic model for continuous speech recognition systems, *Proc. ICSP*, pp. 572-575.
- [10] K.N. Lee (2006) Morpho-Phonological Modeling of Pronunciation Variation for Korean Large Vocabulary Continuous Speech Recognition, PhD Thesis, Sogang University.
- [11] J.H. Lee (2009) *Korean Phonology Lecture*, SAMGYENG-MUNHWASA.
- [12] G.S. Lee, S.I. Yang, Y.H. Gwon (2001) *Speech Recognition*, Hanyang University Press.
- [13] K.H. Kim, G.B. Lee, J.H. Lee (1995) Integration of Phoneme Recognition with Morphological Analysis for Spoken Korean Processing, *The Korean Institute of Information Scientists and Engineers*, 22(10), pp. 1488-1498.
- [14] J.H. Jeon, S.H. Cha, M.H. Chung (1997) Korean Pronunciation Generation Based on Morphonological Analysis, *Autumn Conf. The Korean Institute of Information Scientists and Engineers*, 24(2), pp. 247-250.
- [15] M. Suzuki, S. Makino, A. Ito, H. Aso, H. Shimodaira (1995) A New Hmnet Construction Algorithm Requiring No Contextual Factors, *IEICE transaction on Information and systems*, E78-D(6), pp. 662-668.
- [16] S.H. Do (2010) *New Research for The Place Name of Korea*, JNC.
- [17] D.Y. Park (2007) A Study on the Subway Station Names of Incheon Region, *The Place Name Society of Korea*, 13, pp. 147-177.
- [18] S.H. Jeong (2007) The Study on Phonetical Information for Speech, *The Association of The Research On Korean Language And Literature*, 49, pp. 135-160.
- [19] HTK, <http://htk.eng.cam.ac.uk>, accessed on 20 January 2011.
- [20] S. Young, G. Evermmana, M. Gales, T. Hain, et al. (2006) The HTK Book for HTK Version 3.4.
- [21] L.R. Rabiner (1989) A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition, *Proc. IEEE*, 77(2), pp. 257-286.
- [22] D. Jurafsky and J. H. Martin (2008) *Speech and Language Processing*, Prentice Hall(2nd).
- [23] L. Gu and K. Rose (2000) Sub-state tying in tied mixture hidden Markov models, *Proc. IEEE, Acoustics, Speech, and Signal Processing*, pp. 1062-1065.
- [24] Praat, <http://www.praat.org>, accessed on 20 January 2011.
- 접수일(2011년 1월 31일), 수정일(2011년 6월 2일),
게재확정일(2011년 6월 11일)