

한국어 음성인식을 위한 효율적인 사전 구성에 관한 연구

이 상 복, 최 대 림, 김 종 교
전북대학교 전자정보공학부

Study on Efficient Generation of Dictionary for Korean Vocabulary Recognition

Sang-Bok Lee, Dae-Lim Choi, Chong-Kyo Kim
Dept. of Electronics and Information Eng., Chonbuk National University
E-mail : sang960924@hotmail.com

Abstract

This paper is related to the enhancement of speech recognition rate using enhanced pronunciation dictionary. Modern large vocabulary, continuous speech recognition systems have pronunciation dictionaries. A pronunciation dictionary provides pronunciation information for each word in the vocabulary in phonemic units, which are modeled in detail by the acoustic models. But in most speech recognition system based on Hidden Markov Model, actual pronunciation variations are disregarded. Without the pronunciation variations in the speech recognition system, the phonetic transcriptions in the dictionary do not match the actual occurrences in the database.

In this paper, we proposed the unvoiced rule of semivowel in allophone rules to pronunciation dictionary. Experimental results on speech recognition system give higher performance than existing pronunciation dictionaries.

1. 서 론

인간이 가장 자연스럽게 의사소통에 사용하는 도구는

음성언어(spoken language)이며, 음성언어의 사용은 인간이 다른 동물과 구별되는 가장 큰 특징이다. 그러나 아직 컴퓨터는 이러한 언어능력을 충분히 갖추지 못하고 있다. 여러 이유가 있겠지만 그 중에서도 언어학, 특히 음운론적 이해가 부족하여 음성언어의 근본적 구조 및 복잡다기한 현상은 도외시한 채 단순한 공학적인 접근 방법만을 사용하기 때문이다. 이로 인해 정확한 음성인식 시스템을 구현하는 것이 불가능하다. 이를 위해 발화의 흐름이 음운단위(phonological unit)내에서 조직됨을 강조하는 이론인 운율음운론(prosodic phonology)에 입각한 음운단위를 설정해야 한다[1].

한국어 단어의 철자는 음운 변화 과정을 통해 발음열로 바뀌게 된다. 음운 변화 과정은 음운 변화가 일어나는 위치에 따라 음소 변동 규칙과 변이음 규칙을 사용해서 설명할 수 있다[2]. 해당 음소 문맥에 의해 하나의 음소가 다른 음소로 바뀌거나, 탈락하거나, 첨가되는 양상을 규칙화한 것을 음소 변동 규칙이라 하고 하나의 음소가 발화 상에서 여러 변이음으로 실현되는 양상을 규칙화한 것을 변이음 규칙이라 한다. 본 논문은 변이음 규칙 중에서 반모음의 무성음화 법칙을 보다 정밀하게 적용해서 효율적인 발음사전을 구성하였으며, 이로 인한 인식률의 향상을 이루려고 하였다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 기존의 발음사전 구성방법에 관한 내용을

기술하고, 3장에서는 본 논문에서 제안한 방법을 통해 구성된 사전을 설명하고, 4장에서는 새롭게 구성된 발음사전을 통한 인식 실험 및 결과를 검토한 후, 5장에서 결론을 맺는다.

2. 기존 구성 방법

HMM(Hidden Markov Model)을 기반으로 하는 음성 인식 시스템에서 인식 대상 어휘들에 대한 발음열을 정의한 발음사전이 필요하다. 특히 한국어 음성언어에서는 다양한 음소 문맥에 따라 많은 음운변화 현상이 나타나기 때문에, 각 발음열은 이러한 변화 현상을 잘 반영하고 있어야 한다. 정의된 발음사전을 가지고서 HTK에서는 다음과 같은 과정을 통해 음소 레벨 레이블 정보를 얻어낼 수 있다.

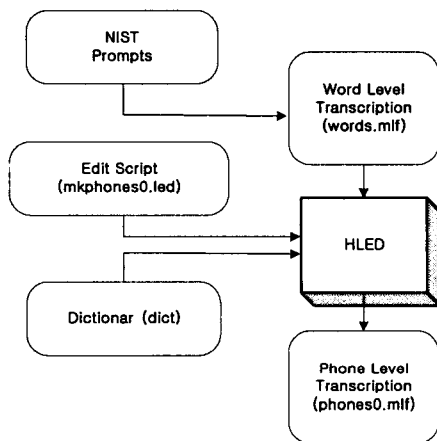


그림 1 Transcription 생성

음소 변동 규칙을 적용시킬 때, 변화되는 음소와 변화에 영향을 주는 음소의 배열을 음소 문맥이라 한다 [3]. 일반적으로 음소 문맥은 연속으로 나오는 두 자음의 순서쌍으로 나타난다. 음소 변동이 일어나는 음절 경계의 앞 음절의 끝 자음과 뒤 음절의 첫 자음의 쌍으로 이루어진다. 기존 사전 구성시에 적용한 음소 변동 규칙의 종류로는 자음 동화, 끝소리 규칙, 자음 축약, 구개음화, 연음 규칙 등이 있다. 그리고 여기에 모음의 발음에 관한 규칙과 장음 처리 규칙을 추가되어 있다. 음소 변동 규칙은 아래 표에 정리되어 있다. 표에 정리되어 있는 규칙들은 앞 음절의 종성, 뒤 음절의 초성(혹은 조사, 접미사), 최종적인 발음 순으로 기

록했다.

표 2.1 구개 음화

규칙 1	'ㄷ, ㅌ' 과 'ㅣ' → [ㅈ, ㅊ] 예) 곧이듣다[고지드따], 굳이[구지]
규칙 2	'ㄷ' 과 '히' → [치] 예) 굳히다[구치다], 닫히다[다치다]

표 2.2 자음 동화

규칙 3	'ㄱ, ㄷ, ㅂ' 과 'ㄴ' → [ㅇ, ㄴ, ㅁ] 예) 밥물[밤물], 깎는[깁는]
규칙 4	'ㄴ, ㅇ' 과 'ㄹ' → [ㄴ] 예) 담력[담녁], 강릉[강능]
규칙 5	'ㅂ, ㄷ, ㄱ' 과 'ㄹ' → [ㅁ, ㄴ, ㅇ] 예) 막론[망논], 협력[협녁]
규칙 6	'ㄴ' 과 'ㄹ' 앞 뒤 → [ㄴ] 예) 난로[날로], 물난리[물랄리]
규칙 7	'ㄱ, ㄷ, ㅂ' 뒤 'ㄱ, ㄷ, ㅂ, ㅅ, ㅈ' → 된소리 예) 국밥[국빱], 곱돌[곱똥]
규칙 8	'ㄴ, ㄴ' 과 'ㄱ, ㄷ, ㅅ, ㅈ' → 된소리 예) 신고[싣꼬], 삼고[삼꼬]

표 2.3 자음 축약

규칙 9	'ㄱ, ㄷ, ㅂ, ㅈ' 과 'ㅎ' → [ㅋ, ㅌ, ㅍ, ㅊ] 예) 좋다[조타], 맞히다[마치다]
------	--

표 2.4 끝소리 규칙

규칙 10	'ㅃ-' 과 자음 → [ㅃ] 예) 밥다[밥따], 밥소[밥쏘]
규칙 11	'ㄹ, ㄹ, ㄹ' 과 어말 및 자음 → [ㄱ, ㄴ, ㅂ] 예) 닭[닥] 삶[삼]
규칙 12	'ㄱ, ㅋ', 'ㅅ, ㅆ, ㅈ, ㅊ, ㅌ', 'ㅍ' 과 어말 및 자음 → 대표음[ㄱ, ㄷ, ㅂ] 예) 닭다[닥따], 앞[압]
규칙 13	'ㄱ', 'ㄴ', 'ㄹ, ㄹ, ㄹ', 'ㅃ' 과 어말 및 자음 → [ㄱ, ㄴ, ㄹ, ㅂ] 예) 녀[녁], 할다[할따]
규칙 14	'ㄹ' 과 'ㄱ' → [ㄴ] 예) 맑게[말게], 물고[물꼬]
규칙 15	'ㅎ, ㄴ, ㄹ' 과 'ㄱ, ㄷ, ㅈ' → [ㅋ, ㅌ, ㅊ] 예) 농고[노꼬], 닭지[달치]
규칙 16	'ㅎ, ㄴ, ㄹ' 과 'ㅅ' → [ㅆ] 예) 당소[다쏘], 많소[만쏘]
규칙 17	'ㅎ' 과 'ㄴ' → [ㄴ] 예) 놓는[논는], 쌓네[싼네]

규칙 18	'ㄴ, ㄹ' 과 'ㄴ' → [ㅎ] 발음안함 예) 앉네[안네], 앉는[안는]
규칙 19	'ㅎ, ㄴ, ㄹ' 어미 및 접미사 → [ㅎ] 발음안함 예) 낳은[나은], 닳아[다라]

앞에서 표로 정리된 4가지의 기본 규칙에 연음 규칙이 추가된다. 이러한 연음 규칙에는 세부 규칙으로 3가지가 있는데 홀받침, 쌍받침, 겹받침이 뒷음절의 조사나 어미와 결합될 때 연음되는 현상들이 기술되어 있다. 그리고 모음의 발음에서는 용언의 활용형에 나타나는 '저, 쩌, 처' 가 '저, 찌, 처' 로 발음되는 것과 자음을 첫소리를 갖는 '의' 가 '이' 로 발음되는 것 두 가지만을 고려하였다. 또한 복합어, 파생어에 나타나는 장음처리 부분이 포함되어 있다. 발음열 표기시에 사용한 표기 기호의 리스트는 아래 표 2.5와 같다.

표 2.5. 발음열 기호

1	sil	16	h	31	i
2	c	17	a	32	S
3	v	18	n	33	Ui
4	N	19	yo	34	j
5	wa	20	b	35	yu
6	d	21	yv	36	wv
7	E	22	ya	37	r
8	k	23	l	38	b0
9	m	24	o	39	J
10	p	25	ye	40	g0
11	t	26	u	41	D
12	g	27	s	42	B
13	U	28	we	43	G
14	e	29	wE		
15	wi	30	d0		

3. 제안한 구성 방법

하나의 음소는 음성 환경에 따라 서로 다른 음가의 소리들로 실현되는데, 이 소리들을 그 음소의 변이음(allophone)이라 한다. 변이음의 가장 중요한 결정 요소는 동시조음이다[4,5]. 동시조음에 의한 변이음 규칙들은 동시조음의 방향성에 따라 순행, 역행, 순행 및 역행, 좌우행 동시조음의 4가지로 분류된다. 본 논문에서는 순행 동시조음에 해당하는 변이음 규칙 중에서 반모음의 무성음화 규칙을 새롭게 추가하였다.

반모음은 성문마찰음 'ㅎ'나 기(aspiration)자질의 뒤,

혹은 치조 마찰음 'ㅅ, ㅆ' 다음에서 무성음화되어 발음된다. 이어서 성문마찰음이나 기자질 또는 치조 마찰음과 융합(Coalescence)하여 무성의 마찰음으로 발음된다. 이를 규칙화하여 표현하면 표 3.1 과 같다.

표 3.1 반모음의 무성음화

규칙 1	성문마찰음 'ㅎ' 과 /j, w/ → 무성음화 예) 향수, 화방 등
규칙 2	기(aspiration) 와 /j, w/ → 무성음화 예) 편수, 튀김 등
규칙 3	치조마찰음 'ㅅ, ㅆ' 과 /j, w/ → 무성음화 예) 오셔서, 싸 등

세부 규칙 3개를 추가함으로써 발음사전에 발음열 표기로 사용되는 음소셋의 리스트에 표 3.2가 추가된다.

표 3.2 추가된 발음열 기호

44	wA	낙화암, 백화점
45	yO	표적수사, 표면에
46	yU	컴퓨터, 휴대용
47	yV	펼치고, 편집국장

4. 인식실험 및 결과

인식 실험에 이용한 데이터는 음소별로 고르게 분포되도록 설계된 452 단어의 PBW(Phonetically Balanced Word)이다. 모델 훈련은 20명의 남녀 화자 2회씩 발성한 총 18,080 단어를 이용하였고, 10명의 남녀 화자 9,040 단어로 인식 성능을 평가하였다.

실험은 펜티엄 IV 1.7G에서 HTK ver. 2.2를 사용하여 수행하였다. 표 4는 모노폰 모델을 만든 후, mixture 수에 따른 인식 결과를 나타낸다.

표 4 인식률 비교(%)

	기존 사전	제안한 사전
None	76.17	76.74
Mixture 2	85.95	86.11
Mixture 3	91.13	91.42
Mixture 5	94.10	94.37
Mixture 7	96.16	96.27

5. 결 론

본 논문에서는 통계적 모델인 HMM을 이용해서 단어 인식을 수행하였다. 음운학적인 지식을 기반으로 한 발음 사전을 구성하는데, 변이음 규칙중에서 반모임의 무성음화 규칙을 보다 정밀화하여 적용함으로써, 인식률의 향상을 보이려고 하였다. Monophone 기반에서는 7개의 mixture 까지 인식률의 향상을 가져왔다. 하지만 9개 이상의 mixture 에서는 변별력있는 인식률의 향상을 가져오지는 못했다.

이는 보다 많은 규칙과 그에 따른 변이음을 세분히 구분하여 인식함으로써, 인식 시스템을 구성하는데 있어 PLU(Phoneme-Like Unit : 유사음소)의 수를 늘리게 되어 학습시에 데이터 부족을 초래했기 때문인 것으로 보인다. 따라서 향후, 보다 적은 수로 이러한 발음들을 통합할 수 있는 연구가 필요하다고 본다.

참고문헌

- [1] Lori Lamel and Gilles Adda, "On Designing Pronunciation Lexicons for Large Vocabulary, Continuous Speech Recognition," Proc. of International Conference on Spoken Language Processing, 1996.
- [2] 전재훈, 형태음운학적 분석에 기반한 한국어 발음 열 자동 생성, 서강대학교 전자계산학과 석사학위 논문, 1997.
- [3] Peter Beyerlein, Meinhard Ullrich and Patricia Wilcox, "Modeling and Decoding of Crossword Context Dependent Phones in the Phillips Large Vocabulary Continuous Speech Recognition", Proceedings of Eurospeech, vol. 3, pp.1167-1170, 1997.
- [4] 이호영, 국어음성학, 태학사, 1996.
- [5] 이호영, 한국어의 변이음 규칙과 변이음의 결정 요인들, 부산 수산대, 1993.