

# 한국어 연속음성인식을 위한 발음사전 구축

이경님<sup>0</sup> 정민화

서강대학교 컴퓨터학과

{knlee@nlpeng.sogang.ac.kr, mchung@ccs.sogang.ac.kr}

## Pronunciation Dictionary For Continuous Speech Recognition

Kyong-Nim Lee<sup>0</sup> Min-Hwa Chung

Dept. of Computer Science, Sogang University

### 요 약

연속음성인식을 수행하기 위해서는 발음사전과 언어모델이 필요하다. 이 둘 사이에는 디코딩 단위가 일치하여야 하므로 발음사전 구축시 디코딩 단위로 표제어 단위를 선정하며 표제어 사이의 음운변화 현상을 반영한 발음사전을 구축하여야 한다.

한국어에 부합하는 음운변화현상을 분석하여 학습용 자동 발음열을 생성하고, 이를 통하여 발음사전을 구축한다. 전처리 단계로 기호, 단위, 숫자 등 전처리 과정 및 형태소 분석 과정을 수행하며, 디코딩 단위인 의사 형태소 단위를 생성하기 위해 규칙을 이용한 태깅 과정을 거친다. 이를 통해 나온 결과를 발음열 생성기 입력으로 하며, 결과는 학습용 발음열 또는 발음사전 구성을 위한 형태로 출력한다. 표제어간 음운변화 현상이 반영된 상태의 표제어 단위이므로 실제 음운변화가 반영되지 않은 상태의 표제어와는 그 형태가 상이하다. 이는 연속 발음시 생기는 현상으로 실제 인식에는 이 음운변화 현상이 반영된 사전이 필요하게 된다.

생성된 발음사전의 효용성을 확인하기 위해 다음과 같은 실험을 통해 성능을 평가하였다. 음향 학습을 위하여 PBS(Phonetically Balanced Sentence) 낭독체 17200문장을 녹음하고 그 전사파일을 사용하여 학습을 수행하였고, 발음사전의 평가를 위하여 이 중 각각 3100문장을 사용하여 다음과 같은 실험을 수행하였다.

형태소 태그정보를 이용하여 표제어간 음운변화 현상을 반영한 최적의 발음사전과 다중 발음사전, 언어학적 기준에 의한 수작업으로 생성한 표준 발음사전, 그리고 표제어간의 음운변화 현상을 고려하지 않고 독립된 단어로 생성한 발음사전과의 비교 실험을 수행하였다. 실험결과 표제어간 음운변화 현상을 반영하지 않은 경우 단어 인식률이 43.21% 인 반면 표제어간 음운변화 현상을 반영한 1-Best 사전의 경우 48.99%, Multi 사전의 경우 50.19%로 인식률이 5~6% 정도 향상되었음을 볼 수 있었고, 수작업에 의한 표준발음사전의 단어 인식률 45.90% 보다도 약 3~4% 좋은 성능을 보였다.

### 1. 서론

대용량 연속음성인식 시스템은 음향모델, 어휘모델, 언어모델의 3가지 주요 구성요소에 대한 지식을 사용하고 있다. 이 중 어휘모델에 사용되는 발음사전은 인식 대상 어휘에 대한 음소 단위의 발음열로 구성된다. 본 논문에서는 인식 시스템의 필수요소인 발음사전 구성을 위하여 한국어 음운변화 규칙을 정리하고, 이에 기반하여 발음열을 출력하는 발음열 자동 생성기를 구축하였다. 또한 발음열 자동 생성기를 이용하여 연속음성인식에 적합한 발음사전의 구성을 비교 실험하였다. 형태소 태그정보를 이용하여 표제어간 음운변화 현상을 반영한 최적의 발음사전과 다중 발음사전, 언어학적 기준에 의한 표준 발음사전, 그리고 표제어간의 음운변화를 고려하지 않고 독립된 단어로 생성한 발음사전과의 비교

실험을 통하여 각각 그 성능평가를 수행하였다.

### 2. 관련연구

음성인식에서 사용되는 학습데이터는 실제 발음형태의 발음처리가 사용되어진다. 수작업으로 발음열을 작성하기 위해서는 작성자가 한국어 음운변화에 대한 전문적인 지식이 필요할 뿐만 아니라 작성에 있어서도 많은 시간 및 노력과 일관성 유지가 필요하다. 따라서 문자열을 자동으로 발음열 변환하는 시스템 구성이 요구된다. 일반적으로 변환 규칙[3]을 이용하거나, 미리 구축된 발음사전을 확장하는 형식, 또는 학습을 통하여 발음열을 생성하는 경우가 일반적 방식이다[1]. 또한 인식에 이용되는 발음사전 자동생성도 음운변환 규칙 및 학습을 이용하여 생성하는 방식으로 구성되고 있다[2][4].

### 3. 발음열 생성기

본 논문에서는 발음열 자동생성을 위해 규칙기반의 발음열 자동 생성 시스템[2]을 안정화하여 낭독체 음성인식을 위한 발음사전 생성에 유용하게 사용할 수 있게 하였다.

한국어의 음운 변화 현상은 단일 형태소 내부와 여러 형태소가 결합하여 한 어절을 이루는 경우, 그 형태소 경계, 그리고 어절 경계에서 서로 다른 음운 규칙이 적용된다. 따라서 언절이나 문장 등의 입력을 음소열로 변환하기 위해서는 형태소 분석, 태깅 작업이 반드시 수행되어야 올바른 발음열을 유도할 수 있다. 제안한 시스템은 한국어의 형태음운현상을 반영하기 위해 형태소 분석을 선행한 후, 한국어에서 빈번하게 발생하는 음운 변화 현상의 분석을 통해 정의된 음소 변동 규칙과 변이음 규칙을 선택적으로 적용하여 형태소, 어절, 언절 또는 문장 등의 다양한 형태의 입력에 대해 발음열을 생성한다.

#### 3.1 전처리기

전처리 단계로 기호, 단위명사, 영문 및 숫자를 포함하는 고유명사, 일반 아라비안 숫자 처리를 위한 전처리 과정 및 형태소 분석 과정을 수행하며, 의사형태소 단위[6]로 생성하기 위해 규칙을 이용한 태깅 과정을 거친다. 이렇게 얻어진 결과는 어휘모델의 표제어 단위로 언어모델의 경우 디코딩 단위가 된다.

#### 3.2 음소 변동 규칙

한국어의 음운 변화 과정[7]을 형태음운론적 분석에 기반하여 17개의 음소변동 규칙으로 재정의하였다. 음소변동 규칙은 특정 음소 문맥에 따라 하나의 음소가 다른 음소로 변환되는 규칙으로 본 논문에서 실제 시스템에 사용된 규칙 수는 총 725개이다.

규칙 번호	규칙의 종류	세부 규칙수
1	음절말 중화	117
2	자음군 단순화	256
3	격음화(기식음화)	21
4	연음규칙	42
5	유음화	10
6	장애음의 비음화	34
7	유음의 비음화	19
8	구개음화	3
9	경음화	136
10	중성 ㅎ-탈락	1
11	ㄴ-첨가	30
12	동일 조음위치 자음탈락	7
13	중복 자음화	6
14	변자음화	17
15	초성 ㅎ-탈락	5
16	자음 첫소리 '의' 모음화	18
17	용언의 활용형 '저, 쯤, 처' 의 모음화	3

[표 1] 음소변동 규칙

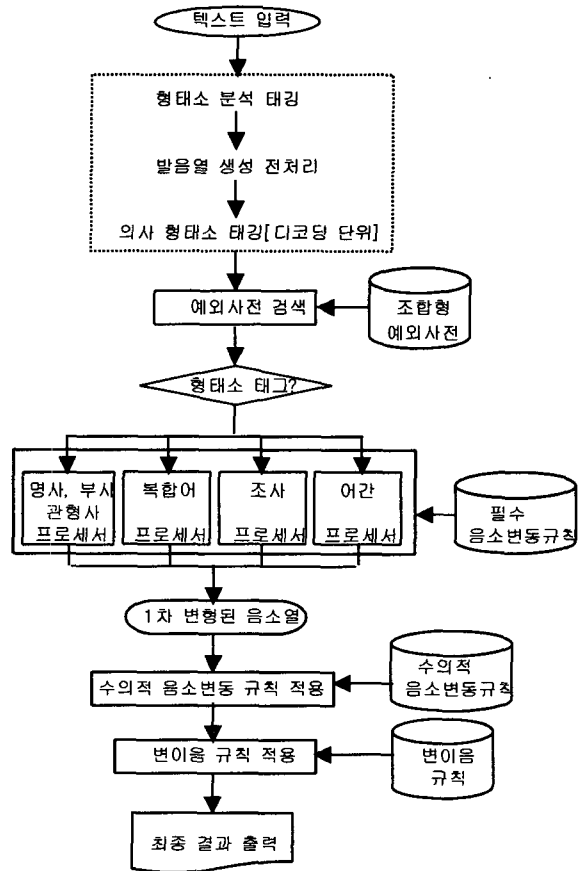
위의 규칙 중 1-11번까지는 필수적 음운변동 규칙이고, 12-15번까지는 수의적 음운변동 규칙이다. 이러한 음소변동 규칙은 음절경계의 앞 음절 끝자음과 뒤 음절의 첫자음의 쌍으로 이루어진다. 위에 정의한 규칙들은 인접한 음소간의 적용여부가 앞 뒤 음절의 중성까지 보아야 하는 경우에는 허용범위가 넘어가 적용하기 어려워진다. 16-17번의 규칙이 이런 경우에 해당하며, 적용범위가 앞 뒤 자소에는 영향을 미치지 않으므로 다른 규칙을 수행하기 전에 이 규칙을 먼저 적용하여 처리하고 나머지 규칙을 적용하는 식으로 구성하였다.

#### 3.3 변이음 규칙

변이음 규칙은 각 음소가 실제 발음되는 양상에 관한 규칙으로 한국어의 변이음 규칙 중 발화 파형 분석을 통해 변이음의 특징이 뚜렷이 구분되는 규칙을 분류하여 음성 인식의 기본이 되는 유사 음소(PLU: Phoneme Like Unit)로 정의하였다. 적용된 변이음 규칙은 무파화 규칙과 구개음화 규칙으로 silence 포함하여 총 48개의 PLU를 사용하였다.

#### 3.4 시스템 구성도

발음열 생성기 시스템 구성도는 다음과 같다.



[그림] 전처리 및 발음열 생성 알고리즘

4. 발음사전 구성

표제어간 음운변화 현상을 반영하기 위해 앞서 구축된 발음열 생성기를 이용하여 발음사전을 구축하였다. 사전 구성 형식에 따라 발음열 생성기의 출력인 발음열 리스트와 사전 표제어를 정렬하여 사전을 구성하였다. 인식실험에 사용할 발음사전은 다음과 같은 특성을 지닌 사전형태로 구축하였다.

구분	사전 특성
TagVarBest	입력 언절내의 표제어간 음운변화 현상을 태그정보를 이용하여 얻은 최적의 발음열로부터 구축된 발음사전
TagVarMulti	입력 언절내의 표제어간 음운변화 현상을 태그정보를 이용하여 얻은 다중 발음열로부터 구축된 발음사전
Standard	전문가에 의한 수작업으로 구성된 1-Best 표준발음사전
WordBest	입력 언절내의 표제어간 음운변화를 고려하지 않고 독립된 단어 입력으로 생성된 발음사전

[표 2] 구축된 발음사전 특성

5. 실험 및 평가

구축된 발음사전의 평가를 위하여 인식 실험을 통하여 그 성능을 비교하였다. 사용된 베이스라인 음성인식기는 HMM 기반의 HTK[5] 이다.

음향모델을 위해서는 20 대 남성화자 172 명의 낭독체 17200문장의 음성을 이용하였고, 학습에는 발음열 전사(Auditory Transcription) 결과를 이용하였다. 이때 사용된 Triphone 수는 총 18967 개였다.

실제 인식에 사용된 문장은 3100 문장으로 사전 표제어 단위 및 언어모델에 사용된 디코딩 단위는 의사형태소이다. 이때, 사전의 표제어(단어 리스트) 수는 10104 단어이고, 문장당 디코딩 수는 평균 19.36 단어로 구성되었다. 인식시 언어모델로는 Backoff -bigram을 사용하였고, 문장의 Perplexity 는 58.72, Entropy 는 5.88bits였다.

	Word Corr(%)	Word Acc(%)	Dictionary (rate)
TagVarBest	48.99	45.82	1.11
TagVarMulti	50.19	46.64	1.52
Standard	45.90	43.10	1.0
WordBest	43.21	40.99	1.0

[표 3] 발음사전별 인식률 및 사전크기

표제어간에 필수적 음운변화 규칙이 적용된 TagVarBest 사전과 수의적 규칙까지 적용된 TagVarMulti 사전을 비교하여 보면, 다중 발음사전을 사용한 경우 인식률이 약 1% 남짓 증가한 것을 볼 수 있으나, 사전 크기의 증감을 고려해 볼 경우 큰 효과를 예상하기 힘들다. 반면 TagVarBest 사전과 Standard 사전의 사용을 비교하였을 경우 약 3% 정도의 성능 향상을, WordBest 사전보다는 5.5% 정도의 인식성능 향상을 얻을 수 있었다. 따

라서 일반적으로 사용되는 발음사전의 형태로는 연속음성 인식시 생기는 음운변이 현상을 반영할 수 없으므로, 표제어간의 음운변이 현상이 반영된 발음사전을 이용함으로써 인식성능을 높일 수 있음을 확인하였다.

또한 보조용언 등과 같이 단음절 어절간에 빈번히 발생하는 경음화 오류의 경우, 의사형태소 태깅을 함으로써 사전 표제어 내부에서의 음운변이 현상으로 발음사전 내에 반영하였다.

6. 결론 및 향후 연구과제

본 논문에서는 기존의 발음열 생성기[2]에서 가지고 있던 규칙 오류 수정 및 적용되지 않았던 음운변화 현상을 재정리하여 시스템을 안정화 시켰다. 이로써 낭독체 연속음성 인식에서의 학습용 발음열 맞춤치 생성 도구로서의 이용과 인식용 발음사전 생성기로서 유용하게 이용할 수 있도록 하였다.

특히 표제어간의 음운변화를 반영할 수 없었던 기존의 인식 시스템의 성능향상을 위해 발음열 생성 결과를 이용하여 발음사전을 구축함으로써 인식 성능을 향상 시킬 수 있었다.

현재 구축된 발음사전은 발음열 생성 결과만을 사용하고 있지만 앞뒤의 문맥정보에 따른 변이 현상의 반영과 한국어가 가지고 있는 상위 언어지식 및 제약사항을 이용할 수 있는 사전을 구축하고자 한다.

7. 참고 문헌

[1] Helmer Strik, Catia Cucchiarini, " Modeling Pronunciation Variation For ASR: Overview and Comparison of Methods ", Proc. Of ESCA Workshop, pp.137-144, 1998.  
 [2]. Jehun Jeon, Sunhwa Cha, Minhwa Chung, Jun Park, and Kyuwoong Hwang, " Automatic Generation Of Korean Pronunciation Variants By Multistage Applications Of Phonological Rules ", ICSLP98, 1998.  
 [3] Maria-Barbara Wesenik, " Automatic Generation of German Pronunciation Variant s", Proc. Of ICSLP, 1996.  
 [4]. Paul C. Bagshaw, " Phonemic transcription by analogy in text -to-speech synthesis: Novel word pronunciation and lexicon compression ", Computer Speech and Language, pp.119-142, 1998.  
 [5]. Steve Young, *The HTK Book*, 1999.  
 [6]. 이경님, 정민화, " 의사형태소 단위의 연속 음성 인식", 제 15회 음성통신 및 신호처리 워크샵, pp.309-314, 1998.  
 [7]. 이호영, 국어음성학, 태학사, 1996.