음소변동규칙의 발견빈도에 기반한 음성인식 발음사전 구성*

나민수(서울대), 정민화(서울대)

<차 례>

<Abstract>

Generating Pronunciation Lexicon for Continuous Speech Recognition Based on Observation Frequencies of Phonetic Rules

Minsoo Na, Minhwa Chung

The pronunciation lexicon of a continuous speech recognition system should contain enough pronunciation variations to be used for building a search space large enough to contain a correct path, whereas the size of the pronunciation lexicon needs to be constrained for effective decoding and lower perplexities. This paper describes a procedure for selecting pronunciation variations to be included in the lexicon based on the frequencies of the corresponding phonetic rules observed in the training corpus. Likelihood of a phonetic rule's application is estimated using the observation frequency of the rule and is used to control the construction of a pronunciation lexicon. Experiments with various pronunciation lexica show that the proposed method is helpful to improve the speech recognition performance.

* Keywords: Pronunciation lexicon, Phonetic rules, Continuous speech recognition.

* 본 연구는 서울대학교 신임교수연구정착금 연구과제 지원을 받아 수행되었음.

1. 서 론

연속음성인식에서 인식의 대상이 되는 음성은 연속된 하나 이상의 단어들의 음성적 실현이며 각 단어의 음성은 선행, 후행 단어에 의해 영향을 받아 다양한 발음변이를 가지게 된다[1]. 음성인식을 위한 발음사전을 구성할 때 이러한 발음변이를 모델링하는 방법은 사용하는 정보의 종류에 따라 지식기반 접근방식과 학습기반 접근방식으로 크게 구분할 수 있다[2][3]. 지식기반 접근방식은 명시할 수 있는 정보를 근거로 하향식으로 발음변이를 정의하는 방식이다[3][4]. 이는 체계적으로 정리된 언어학적 지식을 사용할 수 있다는 장점을 가지지만 실제 발화에서 나타나는 다양한 음운변화현상의 특징을 반영하기 어렵다는 단점을 가진다[5]. 학습기반 접근방식은 음성 신호에 담긴 정보를 근거로 상향식으로 단어에 대한 발음변이를 학습하는 모델링 방법이다[6][7]. 이 방식에는 실제 발화에서 발생한 음운변화현상을 기준으로 발음변이를 모델링할 수 있지만 학습 데이터에서 발생하지 않은 발음변이가 테스트 상황에서 발견될 수 있다는 문제점이 있다[7].

본 논문에서는 제한된 크기의 발음사전에 음운변화현상을 효과적으로 표현하는 발음열을 선택하기 위해서 지식기반 발음열 생성 방식으로 실제 발화에서 발생할 수 있는 일반적인 발음열을 우선 생성한 후 학습 발화에서 관찰된 음소변동규칙의 발견빈도를 기준으로 발음사전에 포함될 발음열을 선택한다. 또한 학습 데이터에서 고빈도로 관찰된 음소변동규칙에 의해 생성된 발음열이 우선적으로 발음사전에 포함되도록 사전을 구성하고 인식 실험을 통해서 본 논문에서 제안한 발음열 선택 방식으로 음성인식성능이 향상될 수 있음을 보인다.

본 논문의 구성은 다음과 같다. 2장에서는 자소 문맥과 형태소 및 음절 경계조건 정보를 바탕으로 음소변동규칙을 정의하고 이 규칙들의 발견빈도를 관찰하기 위한 강제인식과 자소-음소의 정렬과정을 기술한다. 3장에서는 각 조건의 상대적 발견빈도를 고려하여 음소변동규칙의 적합도를 조정하는 방법을 제시한다. 4장에서는 HMM 기반의 연속음성인식 실험을 실시하여 지식기반 발음열 생성 시스템과 성능을 비교하고 5장에서 결론을 내린다.

2. 음소변동규칙과 발견빈도 측정

2.1 음소변동규칙의 정의와 발음사전 구성

음성인식 시스템의 어휘모델을 담당하는 발음사전 구성과정에서 음운변화현상의 반영은 필수적이다. 본 논문에서는 형태음운론적 지식에 기반하여 자소문맥, 형태소의 종류, 음절의 경계조건 등에 따라 발생 가능한 음운변화현상을 음소변동

규칙으로 표현하고[4][5][7] 각각의 음소변동규칙을 자소문맥 단위로 적용하여 발음열을 생성한다.

음소변동규칙의 정의는 다음과 같다.

$$r(L,R,m,s): /L,\ R/ \rightarrow [L',\ R'] \text{ with } P_r$$

음소변동규칙 r은 L, R, m, s 등의 입력부와 L′, R′의 출력부로 구성되며, P_r은 이 음소변동규칙 r의 적합도를 의미한다. L과 R은 발음열을 생성하고자 하는 자소들에 대한 정보로 L은 좌측 음절의 종성 자소, R은 L과 인접한 우측 음절의 초성 자소이며, L′과 R′은 각각 L과 R에 대응하는 출력 음소이다. 이때 표제어의 철자를 나타내는 자소는 기호 / /로 표기하고 음소는 기호 [ ]로 표기한다. m과 s는 각각이 입력의 형태소 종류와 음절의 경계조건을 나타낸다.

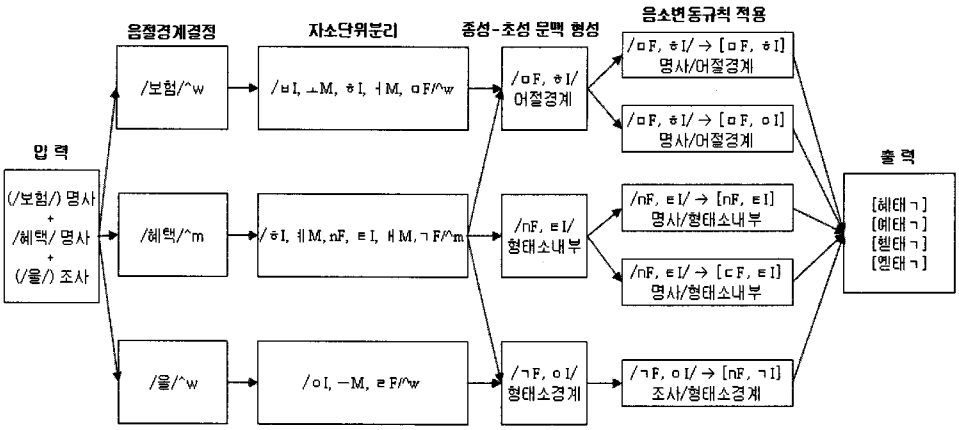
예를 들면 다음과 같다. /음소/에서 /소/의 초성 /ㅅ/에 대한 발음열을 생성하고자 할 때 형태소의 종류는 /음소/의 형태소 종류가 명사이므로 m은 '명사부'이고 음절이 /음소/라는 형태소내부에 위치하므로 s는 '형태소내부'로 정의한다. 이때 음소변동규칙은 /음/의 종성 /ㅁ/과 /소/의 초성 /ㅅ/을 한 쌍의 자소입력으로 받고 입력에 해당하는 음소열을 출력한다. 즉 음소변동규칙은 입력자소열 /L, R/ = /ㅁ, ㅅ/에 대한 출력음소열 [L′, R′] = [ㅁ, ㅅ]을 찾는 역할을 하는 것이다.

한국어의 음운변화현상은 자소의 문맥뿐 아니라 형태소의 종류에 영향을 받고 자소문맥 사이의 음절경계가 어절 또는 형태소의 경계인가, 형태소의 내부인가에 따라 달라진다[3][8]. 따라서 음운변화현상의 반영에서 그와 같은 변인을 고려하기 위해 자소문맥, 형태소 종류, 음절의 경계조건 등 3종류의 정보에 따라서 해당 음소변동규칙의 적용조건을 정의한다. 이때 형태소의 종류는 명사부, 동사부, 어미부, 조사부 등 4종류로 구분하고 음절의 경계조건은 어절경계, 형태소경계, 형태소내부 등 3종류로 구분하여 처리한다.

P_r은 음소변동규칙 r의 적합도를 나타낸다. 이는 발음열 생성 시 해당 음소변동규칙의 적용 가능성을 나타내는 확률적 표현이다. 음소변동규칙은 필수 음소변동규칙과 수의적 음소변동규칙으로 분류하여 필수 음소변동규칙은 0.8에서 1 사이의 적합도 값을 부여하고 수의적 음소변동규칙은 0.7에서 0.9 사이의 적합도 값을 부여한다.

본 논문에서는 발견빈도에 기반하여 음소변동규칙간의 적합도를 상대적으로 조정하는 데 목적이 있고 자음에 대한 음운변화현상만을 조정의 대상으로 삼는다. 따라서 자음에 대한 음운변화현상만을 관찰하여 적합도를 조정하고 모음에 대한 음운변화현상은 기존에 정의된[4] 음소변동규칙의 적합도에 따라 자소 /ㅇ/을 제외한 초성에 인접한 모음의 변화(/ㅖ/→[ㅐ], /ㅒ/→[ㅐ], /ㅙ/→[ㅚ], /ㅞ/→[ㅚ], /ㅢ/→[ㅣ])를 반영하여 처리한다.

발음사전에서 발음열 표기의 단위를 표제어라 한다. 발음열 생성 시 입력은 문장단위의 텍스트를 형태소 단위로 분석하여 형태소의 종류 정보를 부착한 형태이다. 형태소 단위의 입력은 <그림1>과 같이 음절경계결정, 자소단위분리, 종성-초성 문맥 형성, 음소변동규칙 적용 등의 과정을 거쳐 발음열로 출력된다.



<그림 1> 표제어 /혜택/의 발음열 생성 과정

<그림1>에서는 "보험 혜택을"이라는 두 어절로 구성된 입력 텍스트에서 표제어 /혜택/의 발음열을 생성하는 과정을 예로 보여준다. 연속음성인식에서 표제 /혜택/은 인접한 표제어 /보험/의 종성(/ㅁF/)과 /을/의 초성(/ㅇI/) 등과 조음되어 그 발음이 변하므로 인접한 표제어도 입력에 포함하여 그림1에 표시하였다.

음절경계결정 과정에서 인접한 표제어가 띄어쓰기로 분리되는 경우는 '어절경계'로 간주하고 ^w 기호를 부착한다. 표제어가 띄어쓰기로 분리되지 않은 경우는 '형태소경계'로 간주하고 ^m 기호를 부착한다.

자소단위분리 과정에서는 표제어의 자소를 초성, 중성, 종성 등으로 분리하고 자소의 위치에 따라 초성의 자소는 'I', 중성은 'M', 종성은 'F'를 부착한다. 이때 음소변동규칙의 입력형식에 맞추기 위해서 종성이 없는 경우는 그 자리에 'nF'의 자소를 추가하여 표기한다. 따라서 /혜택/은 /ㅎI, ㅖM, nF, ㅌI, ㅐM,ㄱF/로 표기된다.

종성-초성문맥 형성과정에서는 음소변동규칙의 입력 형식에 따라 음절경계에서 좌측 음절의 종성과 우측 음절의 초성을 한 쌍으로 묶는다. 또한 음절사이 경계의 종류가 어절경계(^w)일 경우 '어절경계'를 부착하고 형태소 경계(^m)인 경우 '형태소경계'를 부착한다. 형태소 단위인 표제어 내부에서 생성된 음절경계인 경우 '형태소내부'를 부착한다. 따라서 그림1에서는 /보험/의 종성 /ㅁF/와 그와 인접한 /혜택/의 초성 /ㅎI/가 /ㅁF, ㅎI/로 묶이며, '어절경계' 정보가 부착된다.

음소변동규칙 적용과정에서는 종성-초성문맥의 자소문맥(/ㅁF, ㅎI/)과 발음열을 생성하려는 표제어(/혜택/)의 형태소 종류(명사), 자소문맥 사이의 음절의 경계조건(어절경계, ^w)에 따라 적용될 수 있는 음소변동규칙을 찾고 적용하여 음소열을 출력한다. 조건 {(/ㅁF, ㅎI/), 명사, 어절경계}에 적용될 수 있는 음소변동규칙은 자소를 음운변화 없이 음소로 변환하는 규칙(/ㅁF, ㅎI/→[ㅁF, ㅎI])과 ㅎ-탈락규칙(/ㅁF, ㅎI/→[ㅁF, ㅇI])이다. 각 입력조건에 가능한 모든 음소변동규칙을 적용하면 [혜태ㄱ], [예태ㄱ], [혣태ㄱ], [옏태ㄱ] 등의 발음열을 생성할 수 있다.

표제어 단위의 발음열에 대해 발음열 생성 시 적용된 모든 자소단위 음소변동규칙의 적합도를 누적하여 그 값을 저장한다. 누적된 값과 생성된 발음열 중 최대 누적값을 비교해 그 비율이 주어진 컷오프 비율보다 클 때 해당 발음열을 발음사전에 포함한다. 따라서 음소변동규칙의 적합도는 발음사전의 발음열 구성에 중요한 역할을 하고 있고 음소변동규칙의 적합도를 개선함으로써 음성인식의 탐색공간 구성을 향상할 수 있다.

2.2 음소변동규칙의 발견빈도 측정

본 논문에서 베이스라인으로 사용한 발음열 생성 및 발음사전 구축 시스템은 지식기반의 발음열생성기[4][5]를 바탕으로 하고 있으며, 여기에 추가되는 각 음소변동규칙의 적합도를 정의하기 위해서 적합도의 초기값을 설정하고 설정된 적합도를 이용하여 인식 실험을 실시하였다. 인식결과를 직관적으로 분석하여 각 음소변동규칙의 적합도를 재조정하는 과정을 반복하였고 그 결과로 음운현상의 종류별로 0.7에서 1 사이의 적합도 값을 부여하였다. 이러한 실험을 통한 최적화 방법은 성능이 시행횟수에 따라 달라지고 기존의 시행결과를 연장하여 개선하기가 쉽지 않다는 단점을 가진다. 또한 음운변화현상의 종류별로 적합도를 부여하기 때문에 각 자소문맥에서 발생할 수 있는 음운변화현상간의 상대적 발생확률의 차이를 반영할 수 없다. 예를 들면 종성 /ㄱ/과 초성 /ㄱ/의 문맥에서 경음화가 일어나는 확률과 종성 /ㄱ/과 초성 /ㅅ/의 문맥에서 경음화가 일어나는 확률이 다를 수 있는데 음운변화현상의 종류를 단위로 적합도를 부여하면 이러한 확률의 차이를 반영할 수 없다. 본 논문에서는 실제의 음운변화현상의 발생확률을 음소변동규칙의 적합도에 반영하기 위해서 음성 데이터로부터 음소변동규칙의 발견빈도를 관찰하고 자소문맥을 단위로 정규화하여 적합도로 변환하였다. 본 절에서는 음소변동규칙의 발견빈도를 측정하는 과정을 기술한다.

<그림1>의 예에서와 같이 하나의 입력조건에 대해서는 하나 이상의 음소변동규칙이 발생할 수 있지만 각각의 입력과 출력 사상관계는 하나의 음소변동규칙으로 고유하게 정의된다. 2.1절에서 {(/ㅁF, ㅎI/), 명사, 어절경계}의 입력조건에서 발생 가능한 음소변동규칙은 자소를 음운변화 없이 음소로 변환하는 규칙과 ㅎ-탈락

규칙 등 2개이지만 해당 조건에서 실제 관측한 음소열이 [ㅁF, ㅇI]이라고 한다면 입력과 출력의 사상관계는 하나의 음소변동규칙(/ㅁF, ㅎI/→[ㅁF, ㅇI])으로 정의될 수 있다. 고유한 입출력 셋이 하나의 음소변동규칙에 대응되므로 자소열, 형태소, 음절경계 등의 입력정보와 그에 해당하는 출력 음소열을 수집해서 입력정보에 대응하여 어떤 음소변동규칙이 발생하였는지 측정할 수 있고 발견빈도를 계산할 수 있다.

음소변동규칙의 입출력을 관찰하기 위해서 규칙의 출력인 음소열 정보가 필요한데 음소열을 전사하는 방법은 크게 수작업에 의한 방법과 음성인식기를 이용한 자동 전사방법으로 나눌 수 있다[1][3]. 수작업 전사는 음운변화현상에 대한 전문적인 지식이 필요하고 작업에 많은 시간이 들어가며, 일관성을 유지하기가 쉽지 않다. 또한 사람의 음소 범주와 음성인식기의 음소 범주가 정확히 일치하지 않기 때문에 수작업에 의한 정확한 전사가 음향모델 수준에서도 정확한 전사라고는 볼 수 없다. 따라서 본 논문에서는 음성인식기가 인식하는 음소범주를 반영할 수 있을 뿐만 아니라 시간과 비용 면에서 유리한 음성인식기를 사용한 강제인식으로 자소열에 해당되는 음소열을 전사한다.

2.3 자소열-음소열 정렬

강제인식은 일반적인 음성인식과정과 달리 음성에 대한 정답 전사열인 텍스트와 각 단어에 대한 다중 발음열을 포함하는 발음사전을 입력으로 주고 음성 데이터에 대한 인식을 수행하여 해당 음성에 가장 적절한 발음열을 찾아낸다[1][3].

<표 1> 강제인식 예

텍스트:
(보험) 혜택 (을)
자소열:
(1) (ㅁ) ㅎ ㅖ ㅌ ㅐ ㄱ (ㅇ)
강제인식:
(1) (M) JE TH EH K (WW L)
(2) (ㅁF) 예태ㄱ (ㅡM ㄹF)

<표 1>은 '(보험) 혜택 (을)'이라는 텍스트가 음성 데이터에서 어떤 음소열로 실현되었는지 관찰하기 위해 강제인식을 실시한 결과를 보여준다. 자소에 대한 음소를 찾기 위해서 텍스트를 자소단위로 분리한다. 분리결과는 <표 1>의 자소열(1)이다. 그리고 '(보험) 혜택 (을)'에 대한 음성 데이터와 텍스트로 강제인식을 수행하여 얻은 음소 단위의 전사열이 <표 1>의 강제인식(1)이다. 강제인식(1)은 인식기의 음소출력을 PLU(Phone-like unit) 단위로 표현한 형태이고 강제인식(2)는 PLU 단

위 음소열을 자소로 변환하여 음절단위로 조합한 결과이다. 이에 따르면 '(보험) 혜택 (을)'이라는 텍스트에서 표제어 '혜택'에 대한 음소는 'JE TH EH K'로 실현되었고 음성 데이터에서 '혜택을'은 [예태글]로 발음되었음을 알 수 있다.

강제인식으로 얻은 음성데이터의 음소열은 입력 자소열의 탈락, 종성의 복합자음 및 연음 등에 의한 음소 이동, 음가변동 등의 음운변화현상이 실현된 결과이므로 자소열과 일대일 대응이 되지 않는다. 예를 들면, <표 1>에서 자소열은 '(ㅁ) ㅎ ㅖ ㅌ ㅐ ㄱ (ㅇ)'이어서 로 자소의 개수는 5개이고 강제인식을 통해 구한 음소열은 '(M) JE TH EH K (WW L)'이고 음소의 개수는 4개라서 자소와 음소가 서로 맞대응되지 않는다. 따라서 특정 자소와 음소를 입출력으로 갖는 음소변동규칙을 찾기 위해서 먼저 자소와 음소를 초성, 중성, 종성 순서에 맞도록 정렬한다.

강제인식으로 얻은 음소열에 음절내의 삼성 위치와 비교해서 탈락, 종성의 복합자음 및 연음 등에 의한 음소 이동, 인식단위에 따르는 자소 이동 등을 고려한 정렬 과정을 적용하여 그 결과에서 각 자소에 해당하는 음소의 배열을 찾는다. 자소열은 종성이 없는 경우에 기호 'n'을 삽입하여 항상 종성을 갖추도록 한다.

<표 1>과 <표 2>에서 현재 정렬대상은 '혜택'이고 좌우 문맥은 '보험'과 '을'이다. 강제인식 결과는 <표 1>과 같이 PLU(Phone-like unit) 단위로 출력되고 자소열과 음소열을 정렬하면 <표 2>와 같다. 이 때 PLU는 초성과 종성의 위치에 따라 동일한 자소를 다른 기호로 표시한다. 예를 들면 음소 'K'는 초성 자소 'ㄱ'에 해당하고 음소 'KK'는 종성 자소 'ㄱ'에 해당한다.

<표 2> 자소-음소 정렬 예

자소열:
(ㅁ) ㅎ ㅖ n ㅌ ㅐ ㄱ (ㅇ)
음소열:
(M) n JE n TH EH n K (WW L)
자소-음소 정렬:
ㅁ+ㅎ → ㅁ+ㅇ (M+n)
n+ㅌ → n+ㅌ (n+TH)
ㄱ+ㅇ → n+ㄱ (n+K)

<표 2>와 같이 삼성 위치에 맞춰 자소열과 음소열을 정렬하고 각 위치의 자소와 음소에 상응하는 음소변동규칙을 찾기 위해서 이전 음절의 종성과 다음 음절의 초성을 단위로 자소열과 음소열을 그룹화한다.

정렬결과에서 자소 'ㅁ+ㅎ'은 음소 'ㅁ+ㅇ'으로 실현되었음을 알 수 있다. 자소 'ㅁ'은 '보험'의 종성자소이므로 명사 태그를 가지고 자소 'ㅎ'은 '혜택'의 초성자소이므로 '명사' 태그를 가진다. '보험'과 '혜택'은 서로 다른 어절로 구분되므로

'ㅁ'과 'ㅎ'의 음절 경계 종류는 '어절경계'이다. 자소열과 음소열, 형태소 태그, 음절 경계에 대응되는 음소변동규칙은 자소열 'ㅁ+ㅎ'에 대한 ㅎ-탈락이다. 이와 같은 방식으로 입출력 조건에 상응하는 음소변동규칙을 찾고 검출된 규칙의 발견빈도를 측정한다.

2.4 자소열-음소열 정렬 결과분석

자소에 대한 음소열의 실현양상을 관찰하기 위해서 다양한 음운변화현상을 포함하도록 설계된 대용량의 삼성 PBS(Phonetically Balanced Sentence) 낭독체 연속음성 코퍼스를 사용한다. 이 코퍼스는 음운균형 문장 셋으로 구성되어 있으므로 음운변화현상의 관찰에 적합하다. 음성데이터베이스의 통계는 <표 3>과 같다.

<표 3> 실험에 사용된 삼성 PBS 낭독체 연속음성 코퍼스 통계

단위		개수	평균
문장		43,000	9.4 어절
어절		405,629	2.1 형태소
형태소		844,121	1.6 음절
음절		1,328,492	2.4 자소
음절 경계	어절경계	405,629	-
	형태소경계	438,493	-
	형태소내부	484,370	-
	계	1,328,492	-
자소		3,141,484	-

강제인식은 43,000 문장의 음성 데이터를 대상으로 하고 HTK(Hidden Markov Toolkit)의 HVite 함수를 이용한다[9].

음소변동규칙은 음운변화현상의 발생을 형태음운론적 지식에 의해 규칙적으로 예측할 수 있는지 여부에 따라 필수 음소변동규칙과 수의적 음소변동규칙으로 분류될 수 있다. 필수 음소변동규칙은 발생여부를 규칙적으로 예측할 수 있는 음운변화현상에 대한 음소변동규칙이고 수의적 음소변동규칙은 발생여부를 규칙적으로 규정할 수 없으나 특정 조건에서 발생할 수 있다고 정의된 음운변화현상에 대한 규칙이다. 따라서 발음열 생성 시 음소변동규칙의 적용에 있어서 필수 음소변동규칙과 수의적 음소변동규칙을 분류할 필요가 있고 두 종류의 규칙간의 적합도 범위를 차별화함으로써 분류하였다[4][5]. 필수 음소변동규칙에는 0.8에서 1 사이의 적합도 값을 부여하고 수의적 음소변동규칙에는 0.7에서 0.9 사이의 적합도 값을 부여하였다.

<표 4>는 각 음절경계조건에서 발견된 필수 음소변동규칙의 비율을 나타낸다.

어절의 경계와 형태소 내부에서는 경음화규칙이 가장 많이 발견되었고 형태소의 경계에서는 연음규칙이 가장 많이 발견되었다. 세 경계조건에서 발생한 대부분의 음운변화현상은 연음과 경음화이고 장애음의 비음화, 격음화, 유음의 비음화 등의 순서로 발견되었다.

<표 4> 음절경계조건 별 필수 음소변동규칙의 발견비율

음운현상	어절경계 (%)	형태소경계 (%)	형태소내부 (%)
음절말중화	3.426	0.207	0.979
자음군 단음화	0.169	0.506	0.168
격음화	0.880	2.882	5.813
연음	30.499	69.535	31.709
유음화	4.797	0.244	6.549
장애음의 비음화	2.848	3.464	5.577
유음의 비음화	0.056	0.135	9.152
구개음화	0.000	0.103	0.652
경음화	50.084	22.401	39.063
ㅎ-탈락	0.000	0.523	0.337
ㄴ-첨가	7.241	0.000	0.000
계	100.000	100.000	100.000

<표 5> 수의적 음소변동규칙의 발견비율

음운현상	수의적 규칙 비율 (%)
경음화	7.589
ㄴ-첨가	10.032
자음탈락	0.112
중복자음화	25.737
변자음화	27.473
초성 ㅎ-탈락	29.057
계	100.000

<표 5>는 각 음절경계조건에서 발견된 수의적 음소변동규칙의 발견비율을 나타낸다. 필수 음소변동규칙의 총 발견빈도는 989,977회인데 반해서 수의적 음소변동규칙의 전체 발견빈도는 42,981회였다. 수의적 음소변동규칙은 필수 음소변동규칙에 비해 빈도수가 적으므로 발견빈도의 측정과 적합도 조정에서 음절경계의 종류에 따라 분리하지 않았다. 분석 결과 가장 많이 발견된 수의적 음소변동규칙은 초성 ㅎ-탈락이고 변자음화, 중복자음화 등의 순서로 발견되었다.

<표 4>와 <표 5>에 나타난 음소변동규칙의 발견빈도 순서는 동일한 음성 데이터 영역에 대해서 지식기반 방법으로 발음사전을 생성했을 때의 음소변동규칙의

적용 양상에 대한 연구결과[8]와 일치한다. 본 논문에서 추가로 구한 음소변동규칙의 발견빈도는 발음열 생성 시 음소변동규칙의 적합도를 조정하는 기준이 된다.

3. 음소변동규칙의 적합도 조정

자소열과 음소열을 정렬하여 음소변동규칙의 발견빈도를 측정한 결과를 보면 어절경계의 종성/ㄱ/과 초성/ㄱ/의 문맥에서 경음화된 경우는 전체 문맥조건의 93%이고 /ㄱ/, /ㅅ/의 문맥에서 경음화 된 경우는 전체 문맥조건에서 67%이다. 이는 같은 종류의 음운변화현상(경음화)이라도 자소 문맥에 따라 실현되는 확률에 차이가 있을 수 있음을 보여준다. 본 장에서는 음성 데이터에서 상대적으로 더 높은 빈도로 발견된 음소변동규칙이 발음열 생성 시에도 높은 적합도를 가지도록 조정하는 과정을 보인다. 발음열 생성 시스템은 생성 대상 표제어에 대해 자소단위로 음소변동규칙을 적용하고 각 음소변동규칙에 할당된 적합도를 누적하여 결과 발음열의 적합도를 계산한다. 각 대상 표제어마다 최대 적합도를 가지는 하나의 발음열이 발음사전에 필수적으로 탑재되며, 해당 발음열의 적합도가 정해진 컷오프 비율 이상으로 클 때 2순위 이상의 발음열이 추가로 발음사전에 포함된다.

3.1 음소변동규칙의 적합도

2장의 자소-음소 정렬 결과를 사용해서 각 입출력 조건에 대한 음소변동규칙의 발견빈도를 조사하고 빈도를 기준으로 적합도를 조정한다. 동일한 입력조건(/L, R/)에서 k개의 종류의 음소변동규칙이 적용될 수 있다고 가정할 때 각 음소변동규칙(r_{ni})의 적합도(P_{ni})는 다음과 같이 정의된다.

$r_{n1}(L,R,m,s)$: /L, R/ $\rightarrow$ [L_1', R'_1] with P_{n1}
$r_{n2}(L,R,m,s)$: /L, R/ $\rightarrow$ [L_2', R'_2] with P_{n2}
...
$r_{nk}(L,R,m,s)$: /L, R/ $\rightarrow$ [L_k', R'_k] with P_{nk}

$\Rightarrow r_{ni}(L,R,m,s)$: /L, R/ $\rightarrow$ [L'_i, R'_i], where $0 \le rate_{ni} \le 1$ and $1 \le i \le k$

$P_{ni} = 0.8 + 0.2 * rate_{ni}$

$rate_{ni} = c_{ni} / (c_{n1}+c_{n2}+...+c_{nk})$

c_{ni} = observation frequency of r_{ni}

여기서 r_{ni}는 n 번째 입력조건 /L, R/에서 적용될 수 있는 i 번째 음소변동규칙을 가리킨다. 예를 들면 '명사' '어절경계'의 '종성 /ㅈ/'과 '초성 /ㅇ/' 문맥에서는 /ㅈF, ㅇI/→[nF, ㅈI]와 /ㅈF, ㅇI/→[nF, ㅅI]의 두 종류의 연음규칙과 음절말 중화(/ㅈF, ㅇI/→[ㄷF, ㅇI])규칙이 적용될 수 있다. 이 때 첫 번째 연음규칙을 r_{n1}, 두 번째 연음규칙을 r_{n2}, 음절말 중화규칙을 r_{n3}라 한다. 2장의 자소-음소 정렬 결과에 따라 구한 각 음소변동규칙 r_{ni}에 할당된 발견빈도를 c_{ni}라 할 때 입력조건 /L, R/에서 발견된 모든 음소변동규칙의 발견빈도($c_{n1}+c_{n2}+c_{n3}$)에 대한 비율이 $rate_{ni}$이다. 데이터에서 발견되지 않았으나 실험 시에는 발생할 수 있는 음운변화현상이 있으므로 정의된 모든 음소변동규칙의 적합도가 0.8 이상이 되도록 상수항을 더하여 최종적으로 음소변동규칙 r_{ni}에 대한 적합도 P_{ni}를 정의한다. 이 때 최소 적합도 값 0.8은 실험에 의해서 선정한 값으로 이보다 작을 경우 발음열 생성에서 최소 적합도를 가지는 음소변동규칙이 발음열 생성에 반영되지 않았다.

베이스라인 시스템에서는 적합도를 음소변동규칙에 할당할 때 음소변동규칙별로 고정된 값을 실험에 의해 찾아서 할당하였다. 앞의 예에서는 명사 어절경계의 종성 /ㅈ/과 초성 /ㅇ/ 문맥에 대한 적합도는 각각 0.9, 0.9, 1.0이었다. 음소변동규칙의 통계에 따르면 동일조건의 문맥이 발견된 횟수는 총 16번이고 이 중 첫 번째 연음규칙(/ㅈF, ㅇI/→[nF, ㅈI])이 8번(c_{n1}), 음절말 중화규칙(/ㅈF, ㅇI/→[ㄷF, ㅇI])이 8번(c_{n3}) 발견되었다. 이에 따라 $rate_{ni}$를 각각 0.5, 0, 0.5로 계산하고 각 음소변동규칙의 적합도 P_{ni}를 0.9, 0.8, 0.9로 조정한다. 따라서 발음사전 생성 시 /ㅈF, ㅇI/에 대한 두 번째 연음규칙 r_{n2}가 적용된 발음열은 r_{n1} 또는 r_{n3}가 적용된 발음열에 비해 발음사전에 포함되기 어려워질 것이다.

3.2 조정 적합도 분석

3.1절의 과정에 의해서 정의된 음소변동규칙의 적합도와 베이스라인 시스템의 적합도를 비교하여 적합도가 조정된 문맥조건과 음운변화현상 별 음소변동규칙의 비율을 분석하였다.

어절경계에서는 전체 420개의 문맥조건 중 234개의 문맥조건에서 2개 이상의 음소변동규칙이 발생할 수 있고 이중 126개의 조건에 해당하는 음소변동규칙들의 적합도가 조정되었다. 형태소경계에서는 각각 420개, 73개, 73개였고 형태소내부에서는 420개 52개, 48개였다. 문맥조건 별로 적용 가능한 음소변동규칙들의 적합도가 발생빈도에 따라 상대적으로 조정되었으므로 발음열 생성 시에 각 문맥조건에서 고빈도로 발견된 음소변동규칙을 포함하는 발음열이 발음사전에 포함될 가능성이 커진다.

각 음소변동규칙 별 적합도의 조정양상은 <표 6>과 같다. 베이스라인 시스템의 적합도와 비교하여 높게 조정된 음소변동규칙의 비율은 '상향'으로 표시하고 낮게

조정된 음소변동규칙의 비율은 '하향', 같은 비율은 '유지'로 표시한다.

어절의 경계에서 가장 고빈도로 발견되었던 경음화의 경우 기존의 적합도에 비해서 유지 또는 상향 조정된 음소변동규칙의 비율이 높았다. 이는 고빈도로 발견된 경음화의 음소변동규칙에 대한 적합도가 상향조정되었음을 알 수 있는 결과이다. 그에 비해 형태소의 경계에서 고빈도로 발견되었던 연음의 경우 기존의 적합도와 비교할 때 유지 또는 하향 조정된 음소변동규칙의 비율이 높았다. 적합도가 그대로 유지된 음소변동규칙의 경우의 대부분은 /ㄱF, ㅇI/, /ㄴF, ㅇI/, /ㄹF, ㅇI/ 등과 같이 해당문맥조건에서 동시에 발생할 수 있는 음소변동규칙이 없는 경우이기 때문에 높은 발견빈도를 가졌지만 그 적합도는 유지되었다. 또한 /ㅅF, ㅇI/, /ㅈF, ㅇI/, /ㅊF, ㅇI/ 등과 같은 문맥조건에서는 동일한 조건에서 발생 가능한 음소변동규칙이 모두 연음규칙이기 때문에 문맥에서의 발견비율에 따라 적합도가 분배되어 하향 조정된 규칙들의 비율이 높았다.

<표 6> 적합도가 조정된 음소변동규칙의 비율

	어절경계 (%)			형태소경계 (%)			형태소내부 (%)		
음운현상	하향	상향	유지	하향	상향	유지	하향	상향	유지
음절말 중화	4.444	20.556	75.000	1.905	90.476	7.619	8.163	0.000	91.837
자음군 단순화	0.000	12.810	87.190	2.959	97.041	0.000	20.714	0.000	79.286
격음화	5.263	21.053	73.684	15.000	80.000	5.000	5.263	5.263	89.474
연음	25.000	34.375	40.625	41.026	33.333	25.641	45.455	13.636	40.909
유음화	11.111	22.222	66.667	40.000	50.000	10.000	37.500	0.000	62.500
장애음의 비음화	9.091	45.455	45.455	24.324	64.865	10.811	18.750	0.000	81.250
유음의 비음화	10.526	36.842	52.632	19.048	71.429	9.524	16.667	0.000	83.333
구개음화	0.000	0.000	100.000	0.000	0.000	100.000	0.000	0.000	100.000
경음화	10.084	41.176	48.739	28.889	60.741	10.370	24.742	1.031	74.227
ㅎ-탈락	0.000	0.000	100.000	0.000	100.000	0.000	0.000	0.000	100.000
ㄴ-첨가	0.000	100.000	0.000	0.000	0.000	100.000	0.000	0.000	100.000

음성데이터에서 빈번하게 발견되었던 장애음의 비음화, 격음화, 유음의 비음화 등의 적합도도 상향조정 되었고 음성데이터에서 절대적 발견빈도는 낮았지만 해당문맥조건에서 높은 비율로 발생한 형태소 경계의 음절말 중화, 자음군 단순화 등의 적합도는 상향조정 되었다. 형태소내부에서는 각 문맥조건에서 음소변동규칙이 문맥조건에서 여러 규칙이 동시에 발생할 수 없는 경우가 많기 때문에 유지의 비율이 높다.

어절 또는 형태소 경계에서 상향조정된 음소변동규칙들이 형태소 내부의 경우와 비교할 때 더 많았다. 이것은 형태소 내부에서보다 어절 또는 형태소 경계에서 발생 가능한 음소변동규칙의 적합도가 더 상향조정되었다는 의미이고 따라서 경

계에서의 음소변동규칙이 강화되어 발음열 생성과정에 반영될 것이다.

4. 실험 및 결과

3장에서는 자소-음소열의 정렬결과로 음소변동규칙의 발견빈도를 측정하고 그에 따라 적합도를 조정하였다. 본 장에서는 적합도의 조정으로 발음열을 선택하여 구성한 발음사전의 인식성능을 평가하기 위해서 기존의 적합도로 발음열을 선택한 발음사전과의 비교실험을 실시한다.

4.1 실험환경

실험에서 사용한 음성데이터는 45,000 문장으로 구성된 삼성 PBS 낭독체 연속음성 코퍼스이다. 이 중 43,000 문장을 음향모델의 학습 데이터로 사용하고 학습에 사용하지 않은 문장 중 600문장을 테스트 데이터로 사용한다. 음성은 39차의 MFCC(Mel-frequency cepstral coefficient)로 코딩하고 음향모델은 triphone 기반의 HMM을 12개의 Gaussian mixture로 확장하여 학습한다. 발음사전은 발견빈도가 높은 상위 24,000개의 형태소 단위 표제어를 대상으로 생성하고 언어모델은 backoff-bigram을 사용한다.

4.2 실험결과 및 분석

적합도 조정에 의한 발음사전의 인식성능을 평가하기 위해서 발음사전을 평균변이음의 수에 따라 6종류로 분류한다. 평균변이음의 수는 각 표제어 당 평균 1.3에서 0.2개 단위로 증가시켜 평균 2.3개까지 총 6종류이다.

베이스라인 발음사전은 기존[4][5]의 음소변동규칙을 사용하여 생성한 발음사전이며, 음소변동규칙별로 실험에 의해 고정된 값을 찾아서 적합도를 할당하였다. 제안한 발음사전은 음소변동규칙의 통계에 기반하여 적합도를 조정하고 발음열을 생성한 발음사전이다. 발음사전의 각 표제어는 최대 15개의 발음열을 가질 수 있다. 각 표제어에 대한 발음열 후보의 적합도가 해당 표제어의 후보 발음열 중 가장 높은 적합도에 비해 정해진 컷오프 비율보다 높다면 해당 후보의 발음열을 발음사전에 기록한다. 발음사전의 평균변이음 개수는 사전전체의 발음열 개수를 표제어 개수로 나눈 값이다.

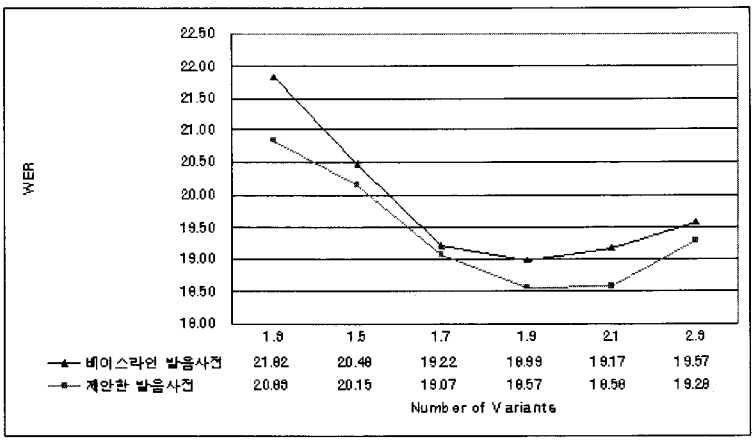
<표 7>은 각 평균변이음 별 베이스라인 발음사전으로 탐색네트워크를 구성했을 때의 음성인식성능을 나타내고 <표 8>은 제안한 발음사전을 사용했을 때의 성능을 나타낸다. <그림 2>에서 두 종류의 발음사전에 대한 인식성능을 비교하였다.

<표 7> 베이스라인 시스템의 발음사전 평가

	평균변이음 개수	대치오류 (%)	삭제오류 (%)	삽입오류 (%)	WER (%)
베이스라인 발음사전	1.3	14.66	4.62	2.54	21.82
	1.5	13.84	3.78	2.86	20.48
	1.7	12.85	3.71	2.66	19.22
	1.9	12.74	3.67	2.58	18.99
	2.1	12.83	3.70	2.64	19.17
	2.3	13.11	3.71	2.74	19.57

<표 8> 제안한 시스템의 발음사전 평가

	평균변이음 개수	대치오류 (%)	삭제오류 (%)	삽입오류 (%)	WER (%)
제안한 발음사전	1.3	14.09	3.96	2.77	20.83
	1.5	13.46	3.86	2.83	20.15
	1.7	12.56	3.70	2.62	19.07
	1.9	12.48	3.52	2.57	18.57
	2.1	12.51	3.47	2.60	18.58
	2.3	13.09	3.43	2.75	19.28



<그림 2> WER 기준의 인식 성능 비교

두 방식의 발음사전 모두 평균변이음의 개수가 1.9 이하에서는 발음사전의 크기가 커짐에 따라 인식 성능(Word Error Rate 기준)이 향상됐고 1.9를 기준으로 발음사전의 크기가 증가함에 따라 인식 성능이 하락함을 볼 수 있다. 이것은 평균변이음의 개수가 1.9에서 탐색네트워크의 크기가 최적화되었음을 알 수 있는 결과로, 평균변이음이 1.3~1.9 사이일 때는 발음열 정보의 추가가 인식성능에 도움이 되지만 1.9 이상에서는 정보의 추가가 네트워크의 혼잡도를 높여 인식성능을 하락시킨 것으로 볼 수 있다.

두 방식을 비교한 결과 제안한 시스템의 발음사전으로 음성인식 시스템을 구성하였을 때 인식오류(WER)가 18.99%에서 18.57%로 최대 0.42% 감소하였다. 각 평균변이음의 경우에서 제안된 방식의 발음사전의 사용으로 일관적인 성능향상이 관찰되었다. 따라서 적합도 조정이 인식성능 향상을 위한 발음사전 구성방법으로 의미가 있는 것으로 판단된다.

음소변동규칙 적용의 적합도 조정으로 실제 음성데이터에서 통계적으로 빈번하게 발생한 규칙이 발음열 생성에 더 비중이 크게 작용하였고, WER의 향상은 적합도가 큰 음소변동규칙에 의해 생성된 발음열이 발음사전에 포함되어 탐색 네트워크의 구성이 개선된 결과로 해석된다.

5. 결 론

본 논문에서는 제한된 발음사전의 크기 내에서 음성데이터의 음운변화현상을 효과적으로 표현할 수 있는 발음열을 선택하기 위해서 고빈도로 발견되는 음소변동규칙을 기준으로 발음사전을 구성하였다. 음소변동규칙의 발견빈도를 측정하기 위해서 음성데이터에 강제인식을 수행하여 자소열에 대한 음소열을 얻었고 자소-음소 정렬결과에 따라 음소변동규칙의 적합도를 조정했다. 어절의 경계에서는 경음화 규칙의 발견비율이 상대적으로 높았기 때문에 경음화의 적합도를 상향조정하였다. 형태소의 경계에서는 연음 규칙의 절대적 발견빈도는 높지만 상대적 비율은 낮았기 때문에 적합도를 하향조정하고 절대적 발견빈도가 낮지만 상대적 비율이 높은 음절말 중화, 자음군 단순화, 격음화 등의 적합도는 상향조정하였다. 조정된 적합도를 기준으로 발음열을 선택하고 발음사전을 구성한 실험을 통해서 제안한 방식이 연속음성인식에 더 효과적임을 확인하였다.

실험에서 음성데이터에 대한 음소변동규칙의 발견빈도 측정 결과는 규칙의 입력조건 별로 발생 가능한 음소변동규칙 간의 상대적 차이를 적합도에 반영하는데 사용되었다. 이에 추가로 음절경계조건에서의 음운변화현상 별 발생비율 차이를 반영하여 적합도 조정을 개선하는 연구가 필요하다.

음소변동규칙의 발견빈도를 측정하기 위하여 강제인식을 수행한 결과 입력조건과 출력음소열이 맞대응 되지 않는 경우, 즉 기존에 정의된 음소변동규칙에 사상되지 않는 음운변화현상이 발견되었다. 이는 실제 음성데이터에 기존에 정의되지 않은 음운변화현상이 존재하는 것을 나타낸다. 이와 같은 미정의 음소변동규칙 중 데이터 잡음이 아닌 유의미한 규칙을 선별하여 학습할 수 있는 방법의 도입이 필요하다.

감사의 글

본 실험에 사용한 삼성종합기술원의 PBS 음성 데이터베이스 사용허가에 감사드립니다.

참 고 문 헌

[1] H. Strik, C. Cucchiarini, "Modeling pronunciation variation for ASR: A survey of the literature", *Speech Communication*, Vol. 29, Nos. 2-4, pp. 225-246, 1999.

[2] H. Strik, "Pronunciation adaptation at the lexical level", *Proc. 4th ISCA Tutorial & Research Workshop on Adaptation Methods for Speech Recognition*, pp. 123 - 131, 2001.

[3] 정민화, 이경님, "한국어 연속음성인식 시스템 구현을 위한 형태소 단위의 발음 변화 모델링", *말소리*, 제49호, pp. 107-121, 2004.

[4] 이경님, 전재훈, 정민화, "한국어 연속음성 인식을 위한 발음열 자동 생성", *한국음향학회지*, 제20권, 제2호, pp. 35-43, 2001.

[5] K. Lee, M. Chung, "Morphome-based modeling of pronunciation variation for large vocabulary continuous speech recognition in Korean", *IEICE Transactions on Information and Systems*, Vol. E90-D, No. 7, pp. 1063-1072, 2007.

[6] J. Kessens, C. Cucchiarini, H. Strik, "A data-driven method for modeling pronunciation variation", *Speech Communication*, Vol. 40, No. 4, pp. 517-534, 2003.

[7] J. Jeon, M. Chung, "Automatic generation of domain-dependent pronunciation lexicon with data-driven rules and rule adaptation", *Proc. Interspeech*, pp. 1337-1340, 2005.

[8] 이경님, 정민화, "발음열 자동 생성기를 이용한 한국어 음운 변화 현상의 통계적 분석", *한국음향학회지*, 제21권, 제7호, pp. 656-664, 2002.

[9] S. Young, et al., *The HTK Book (for HTK Version 3.2)*, Entropic Cambridge Research Laboratory, 2002.

접수일자: 2007년 11월 17일
게재결정: 2007년 12월 19일

▶ 나민수(Minsoo Na)
주소: 서울시 관악구 관악로 599번지 서울대학교 대학원 협동과정 인지과학전공
소속: 서울대학교 대학원 협동과정 인지과학 전공
전화: 02) 880-9039
E-mail: dix39@snu.ac.kr

▶ 정민화(Minhwa Chung) : 교신저자
주소: 서울시 관악구 관악로 599번지 서울대학교 언어학과
소속: 서울대학교 언어학과
전화: 02) 880-9195
E-mail: mchung@snu.ac.kr