

# 의료분야를 위한 영어 발음열 생성 시스템

김아름\*, 정경석, 박혁로

\*전남대학교 전산학과

{kal7112\*, ksjung}@dal.chonnam.ac.kr

hyukro@chonnam.ac.kr

## A Generation System of English Pronunciation for the medical domain

A-Lum Kim\*, Kyung Seok Jeong, Hyuk Ro Park

\*Dept. of Computer Science, Chonnam National University

### 요 약

본 논문은 의료분야의 음성 인식 시스템의 발음모델의 성능 향상에 필요한 올바른 영어 단어 발음열을 얻고자 한다. 본 시스템의 텍스트는 의료 전문 용어인 영어와 한글의 조합으로 되어있어, 한국어 G2P 성능뿐만 아니라 영어 G2P의 성능도 중요한 문제가 되고 있다. 또한 본 시스템의 의료 음성 데이터가 한국 화자로서, 표기열이 영어식 발음 폰셋으로 변환되면 효율적이지 못하다. 이를 위해, 영어 G2P의 결과를 한국 화자에 적합하게 변환해주는 방법론이 필요하게 된다. 따라서 본 논문에서 제안하는 방법은 음성 텍스트에서 영어만 추출한 후, 영어 G2P 프로그램(addtpt, NIST)을 이용해 발음열을 구한다. 그리고 한국 화자의 실제 음성을 통해 얻은 정답 발음열을 구하여 서로 비교한다. 비교를 위해 각 발음열의 한 폰씩 정렬을 수행한 후, 삽입, 삭제, 대치 에러가 이러나는 쌍과 좌우 바이그램 정보를 추출한다. 마지막으로, 좌우 바이그램 정보에서 best1의 에러 패턴을 통해 모든 단어에 적용한다. 이 때, 최종적으로 실보다 득이 되는 에러패턴만을 추출, 적용한다. 실험에서는 26여개의 에러 패턴을 찾을 수 있어, 8%의 올바른 발음열을 추가적으로 얻는데 성공하였다.

### 1. 서론

음성은 인간의 의사수단 중 가장 자연스럽고 효과적인 정보 교류의 수단이라 할 수 있다. 또 음성에 의해 표현되는 발음은 인간과 인간사이의 의사소통의 수단 일 뿐만 아니라 논리적으로 사물을 생각하는 경우에 있어서도 중요한 역할을 한다. 이러한 음성을 컴퓨터가 인식하는 분야가 음성인식 분야이다. 현재 이러한 음성인식 분야는 상당한 발전을 거듭하여 여러 응용분야에 적용되어왔다.

그러나 의료분야를 적용하다보니 의료 전문 용어인 영어가 많이 나오게 되어, 한국어 G2P(Grapheme to Phoneme) 성능뿐만 아니라 영어 G2P의 성능도 중요한 문제가 된다.

또한 본 시스템의 의료 음성 데이터가 한국 화자로서, 한국인의 고질적인 발음상의 문제를 가지고 있다. 한국인은 실제 원어민의 발음과는 차이가 있

음으로, 이러한 차이를 최소화할 수 있는 방법이 필요하게 된다. 이러한 문제를 해결하기 위해, 94.5%의 음소의 정확률을 가지고 있는 영어 G2P 프로그램[1]을 이용한다.

본 논문에서는 의료분야 텍스트의 영어 단어 발음열 생성에 있어서 오류를 수정하기 위한 방법을 제시하였다. 영어 G2P 결과와 정답발음열을 비교하여 에러패턴을 추출한 후, 이 패턴을 평가해서 유용한 패턴만을 추출한다.

본 논문의 구성은 2장에서는 관련 연구와 제 3장에서는 한국인의 영어 발음 차용방식을 소개한다. 그리고, 제 4장에서는 에러 패턴 학습을 이용한 영어 발음 생성에 대해 설명하고, 제 5장에서는 실험 및 평가를 기술하고 마지막으로 제6장에서는 결론을 맺었다.

2. 관련연구

기존에 음성 인식분야에서 패턴 매칭을 이용하는 방법은 인식된 결과에서 단어의 올바른 유무를 가리는데 후처리로 사용되었다. 그리고 이런 방법론은 크게 규칙을 이용하는 방법과 통계적인 방법으로 나뉜다. 하지만, 규칙을 직관적으로 알기 어렵거나 고려해야 할 도메인이 클수록 통계적인 방법이 선호되고 있다.

Ringger et al.[2]은 음성 인식기가 학습한 환경과 실제 사용되어지는 환경의 차이를 Noisy Channel Model로 해석하였다. 이는 음성 인식의 오류에 대해 각 환경적인 요소를 변수로 하여 체계적으로 모델링한 점에 의미가 있다. 하지만, 이것은 각 단어가 아닌 각 음소간의 언어 모델을 적용하기에는 한 음소에서 또 다른 음소로 갈수 있는 확률은 너무 랜덤하여 언어 모델의 장점을 잘 살리지 못한다.

Satoshi Kaki[3]는 음성 인식의 결과를 이용해 에러 패턴을 모아 저장하고 올바른 corpus를 이용하여 에러 패턴에 해당하면 corpus의 올바른 문장으로 치환하는 방법이다. 이 방법은 응용 도메인이 작을 경우 효율적인 에러 수정을 기대할 수 있으며, n:m 대치 오류를 잘 처리한다. 하지만 이 방법도 각 음소간 나올 수 있는 경우가 랜덤하여 올바른 음소 또한 빈번히 고치게 되는 경우가 자주 발생한다.

김용현[4] 연구는 error pattern을 검출/수정하는 것은 [3]과 유사하나, n-best 후보에 error 패턴과 좌우 bigram 정보를 적용하여 최대의 가능성을 보이는 후보를 선택하는 방법이다. 하지만 이도, 현재 음향 모델의 기술로 음소간의 n-best를 구하는 것은 쉽지 않은 일이므로, 이 방법을 의료분야에 적용하기는 어렵다.

마지막 방법으로 윤용욱[5]의 방법은 어휘의미패턴을 기반한 통계적인 오류 수정이다. 이 또한 음소단위는 의미를 나타낼 수 없으므로 의료분야에 적용하기 어렵다.

3. 한국인의 영어 발음 차용방식

영어와 한국어의 음절구조를 살펴보면, 영어는 C(3)VC(4)로 음절 초 최대 자음3개, 모음, 음절 말 최대 자음4개로 구성되고, 한국어는 C(1)VC(1)로 음절 초 최대 자음1개, 모음, 음절 말 최대 자음1개로 구성된다. 이와 같은 음절, 음향 차이로 인하여 한국인이 발화 시 한국어의 음절 구조를 맞추기 위해서 모음이 삽입되는 등 많은 차용을 보인다.

조사된 영어와 한글의 음절, 음향적 성격의 차이로 인한 차용은 다음 [표 1]과 같다.

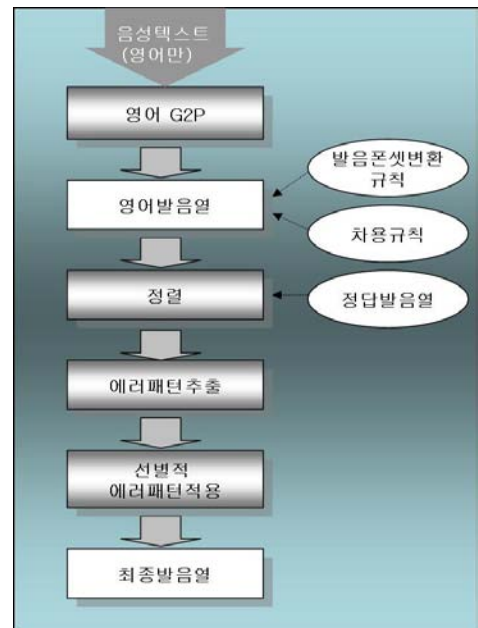
[표 1] 영어 발음 차용방식

종류	변이현상	예
삽입 모음	양순음뒤 [u] 삽입	brown→[puraun]
	구개음뒤 [i] 삽입	bench→[bentʃi]
	나머지환경 [I ] 삽입	christmas→[khIrisImasI]
Stop음	무성폐쇄음은 격음 대체	pie→[phai]
	단어초에 경음화	boy→[poi]
마찰음	뒤에 모음 오거나	sing→[Sain]
	단어말 올때 [S] 대체	bus→[bəSI]
	뒤에 자음 올때 [s] 대체	star→[sItha]
[r], [l]음	음절말 [r]은 발음 안됨	car→[kha]
	음절초 [l]은 [r]로 대체	love→[rəbI]
	모음사이의 [l]은 [ll]로대체	olive→[ollibI]

4. 에러 패턴 학습을 이용한 영어 발음 생성

4.1 영어 발음 모델 구조도

영어 발음 모델 구조도는 [그림 1]과 같이, 영어로 된 음성텍스트가 들어오면 영어 G2P프로그램을 통해, 영어 발음열을 얻게 되는데, 여기에서 영어 발음열은 발음 폰셋 변환규칙과 차용규칙을 적용한다. 그 다음단계로 영어발음열과 정답발음열을 비교하여 에러패턴을 추출한다. 그러나 추출된 에러패턴을 각 단어에 적용함으로써, 틀린 폰을 고쳐 줄수도 있지만 맞는 부분도 틀리게 할 수 있으므로 실보다 득이 되는 에러패턴만을 선별하여 최종발음열을 얻는다.



[그림 1] 영어발음모델 구조도

4.2 휴리스틱 규칙 적용

기존에 조사 된 [표 1]의 규칙들 중, 본 도메인과 특정 화자에게 자주 발견된 규칙들만을 적용한다. 본 실험에서는 빈번히 발견된 삽입 모음 [I]와 [I]이 [r]로 차용됨을 고려하였다. 또한, 향후 비 영어를 포함한 음소 셋을 고려해야함으로 발음 폰 셋 변환 규칙을 적용하였다.

4.3 정렬

영어 G2P의 발음열과 정답 발음열을 비교하여 에러 패턴 쌍을 추출하기 위해 정렬 알고리즘을 [표2]에서 정의하였다. 본 정렬 알고리즘은 [4]의 정렬 알고리즘과 비슷하나 적용할 대상이 음소임을 고려하여 수정 설계 하였다.

[표 2] 정렬 알고리즘

알 고 리 즘	1. 각 음소 열에 대해 인덱스 정보에 -1을 초기값으로 설정한다. 2. 정답 발음열과 영어 G2P 발음열 중 짧은 발음열을 기준으로 왼쪽 첫 음소부터 비교한다. 2-1. 만약 음소가 일치하면 그 음소의 위치 정보를 인덱스 정보에 입력한다. 2-2. 만약 일치하지 않으면 그 다음 번째음소와 비교한다. (이는 3회이상이면 그 인덱스는 -1로 그 대로 둔다) 3. 인덱스 정보를 이용하여 에러 패턴을 구별한다. 3-1. 인덱스가 -1이면, 삽입 에러 3-2. 양수가 2이상 증가하면 삭제 에러 3-3. 둘 다 인덱스가 -1이면, 대치 에러
예	입력 단어 : bleeding 정답 발음열 : b I R I d I Q 1-1 2 3 4 5 -1 영어 G2P 발음열 : b r I d I n 1 3 4 5 6 -1 에러 패턴(정답:영어 G2P) 쌍 : <b>I:-</b> (삭제), <b>Q:n</b>

4.4 에러 패턴 추출

정렬을 통해 얻어진 에러 패턴을 통해 영어 G2P의 좌우 음소의 bigram 정보를 함께 저장한다. 그리고 좌우 음소를 기준으로 해서 1-best 에러 패턴만을 다시 추출한다. 만약, 빈도가 같거나 비슷한 경우, 두 가지를 모두 선택하고, 차후 선별적 에러 패턴 과정에서 탈락되는 에러 패턴을 제거하고, 아니면 임의적으로 하나의 패턴을 제거한다. 다음[표 3]는 추출된 에러 패턴을 몇 가지 예를 보인다.

[표 3] 추출된 에러패턴의 예

입력 단어	영어 G2P 발음열	정답 발음열	에러패턴
orifice	<b>E</b> r i p a s I	o r i p i s I	-/E/r //o
The	<b>T</b> E	d E	-/T/E //d
overlapping	o b E r r e p i n	o b E r r e p i Q	i/n/- //Q
ureter	i u R r i t E R	u R r i t E R	-/i/u// -
large	r a R z I	r a R z i	z/I/-// i

여기서 에러 패턴을 표현하는 방법은 영어 G2P 발음열의 좌측 음소/에러 음소/우측 음소//정답 음소가 된다. 그리고, '-'는 음소가 null임을 표시한다.

4.5 선별적 에러 패턴 적용

에러 패턴을 각 단어에 적용함으로써, 틀린 폰을 고쳐줄 수도 있지만, 맞는 부분을 틀리게 할 경우도 있게 된다. 이를 해결하기 위해, 각 단어의 정답 발음열과 에러 패턴이 적용된 발음열을 비교함으로써 실보다 득이 되는 에러패턴만을 선별한다.

본 실험의 선별하는 조건식은 다음[표 4]와 같다. .

[표 4] 선별적 에러패턴 조건식

```

TER = EER + λ2 * CER;
if(ECR >= λ1 && TER < ECR){
    최종 에러 패턴 선택;
}
ECR : 에러가 맞게 될 확률
TER : 모든 에러 확률
EER : 에러가 또 틀릴 확률
CER : 맞은 것을 틀리게 할 확률
    
```

본 실험에서는 λ<sub>1</sub> = 0.85, λ<sub>2</sub> = 3.5일 때 가장 좋은 교정을 보였고 총 1-best 에러패턴 450개 중, 최종적으로 26개의 에러패턴을 구하였다. 여기서 교정이란 좌우 음소에 근거하여 에러 패턴이 정답 패턴으로 교정되는 것으로, 모든 발음열이 다 맞아야 하는 것만이 아니라 부분적으로 교정된 경우도 이에 해당된다. 이는 부분 교정도 전체 인식 성능의 향상에 기여하기 때문이다.

5. 실험 및 평가

5.1 실험

본 실험은 의료 영역 음성에 부합되는 음성 텍스트(2003년 5월~12월)중, 소화기계, 비뇨생식기계 부분만을 고려한다.

이 텍스트에 대한 정보는 다음[표 5]와 같다.

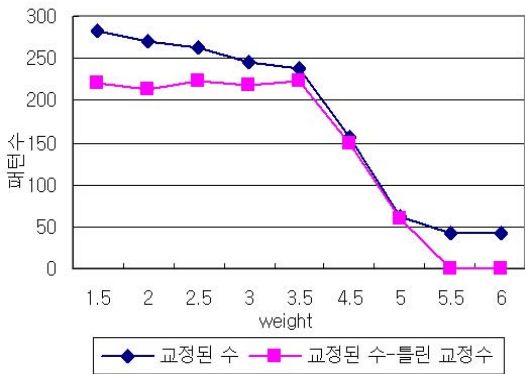
[표 5] 텍스트 정보

문장의 수	55939
총 어절의 수	360982
총 형태소의 수	511787
유일한 형태소의 수	4607
유일한 영어의 수	3898

위의 표와 같이 의료 텍스트는 영어가 84%가 넘는다. 결국은 발음 사전을 잘 구축하기 위해 영어 발음 생성이 얼마나 중요한 영향을 끼치는 줄 알 수 있다.

5.1.1 영어 발음 생성 성능

영어 발음 생성 성능으로는 영어 G2P 결과에 휴리스틱 규칙을 적용한 후, 정렬을 수행하여 에러 패턴을 추출한 결과는 다음과 같다. 총 에러패턴의 수는 606개, 1-best 에러 패턴의 수는 450개이다. 이를 기반으로 다음은 선별적 에러 패턴을 적용하기 위한 실험을 수행한 결과는 다음 [그림 2]와 [표 5]와 같다.



[그림 2] 가중치에 따른 패턴수

[표 5]  $\lambda_1$ 이 85%일 때,  $\lambda_2$ 에 따른 결과

$\lambda_2$	1.5	2	2.5	3	3.5	4.5	5	5.5	6
에러패턴의 수	43	40	36	32	26	23	17	16	16
교정된 수(A)	282	271	264	245	237	155	61	41	41
틀린 교정수(B)	62	57	42	27	15	6	2	10	10
(A)-(B)	220	214	222	218	222	149	59	31	31
음소의 정확률	92.5	92.7	93.1	93.6	94.2	89.2	87.9	87.2	87.2
단어의 정확률	0.62	0.62	0.63	0.64	0.66	0.66	0.66	0.67	0.67

[그림2]와 [표 5]를 비교할 때 변수  $\lambda_1 = 85\%$ 으로 하였을 때,  $\lambda_2 = 3.5$ 일 때, 최적의 변수임을 알 수 있었다. 26개의 패턴을 분석해 3898개에 적용한 결과, 단어의 정확률이 60.5%에서 65.8%로, 음소의 정

확률이 86.1%에서 94.2%로 향상 되었고 과반수이상 대부분이 빈도수가 1~3정도인 패턴들로 판명되었다.

5.2 평가

본 실험에서 사용된 영어 G2P 프로그램이 실제 의료 전문 영어에 최적화된 시스템은 아니기 때문에 기대 이하의 정확률을 보인다. 하지만 본 실험은 한국 화자로 인한 실제 발음열과 영어 G2P의 발음열의 차이를 줄이는 에러패턴을 추출함으로써, 이러한 문제를 해결하는데 공헌한다.

6. 결론

본 논문에서는 의료분야 텍스트의 영어 단어 발음열 생성에 있어서 오류를 수정하기 위한 방법을 제시하였다. 영어 G2P 결과와 정답발음열을 비교하여 에러패턴을 추출한 후, 이 패턴을 평가해서 유용한 패턴만을 추출한다. 이 패턴을 실제 영어 텍스트에서 적용해본 결과 영어 G2P를 그대로 사용한 결과 보다 8%의 성능향상을 보였다. 그러므로 에러패턴을 활용하는 방법은 유용하다.

향후 연구로는 음향 정보를 이용한 전 처리로 에러 패턴 적용을 한정하여, 무분별한 적용을 줄임으로써 보다 많은 에러 패턴을 추가적으로 찾을 수 있는 방법을 연구할 계획이다.

참고문헌

[1] W.M. Fisher., "A Statistical Text-to-Phone Function using Ngrams and Rules," IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume 2, pp. 649-652, March 1999.  
 [2] E.K.Ringger et al., "A Fertility Model for Post Correction of Continuous Speech Recognition," ICSLP, pp897-900, 1996.  
 [3] Satoshi Kaki et al., "A Method for Correcting Speech Recognition Using the Statistical features of Chracter Co-occurrence," COLING-ACL, pp.653-657, 1998.  
 [4] 김용현, 정민화, "에러패턴 학습과 후처리 모듈을 이용한 연속 음성 인식의 성능향상," 한국정보과학회 봄 학술발표논문집, pp.441-443, 2000.  
 [5] 윤용욱, 정한민, 이근배, "어휘의미패턴을이용한 음성인식오류검출 및 수정," 한글 및 한국어 정보처리 학술대회 논문집 pp62-68, 2002.