

# 한국인을 위한 영어 발음 교정 시스템에 대한 성능 평가

김 무 중, 김 효 숙, 김 병 기  
(주) 언어과학 부설 언어공학연구소

## Performance Evaluation of English word Pronunciation Correction system

Mujung Kim, Hyosook Kim, Byunggi Kim  
Center for Linguistic Engineering, Eoneo Inc.  
E-mail : {donaldos, hoyssook, killm2}@eoneo.co.kr

### Abstract

In this paper, we present some of experimental results developed in computer-based English Pronunciation Correction System for Korean speakers. The aim of the system is to detect incorrectly pronounced phonemes in spoken words and to give correction comment to users.

Speech data were collected from 254 native speakers and 411 Koreans, then used for phoneme modeling and test. We built two types of acoustic phoneme models: native speaker model and Korean speaker model. We also built language models to reflect Koreans' commonly occurred mispronunciations. The detection rate was over 90% in insertion/deletion/replacement of phonemes, but we got under 75% detection rate in diphthong split and accents.

### I. 서론

영어 학습에서 소리를 통한 학습은 학습의 효율성 및 언어 습득의 과정 측면에서 중요한 과정 중의 하나이며, 이러한 관점에서 영어 발음 학습에 음성인식, 합성을 접목시킨 다양한 영어학습 제품이 출시되고 있다. 그러나 기존의 영어 발음 학습기는 일반 영어 음

성 인식기를 기반으로 하여, 다양한 발음에 대한 유사도를 추출후 백분율로 나타내어 발음의 성취도를 평가하는 수준이다. 이는 잘못 발음된 음소 및 교정 정보를 학습자에게 제공할 수 없다.[1]

본 논문에서 제안하는 한국인을 위한 영어 발음 교정기는 표준 영어 음소와 영어 음소와 음가는 유사하나 음성 자질이 틀린 한국어 음소를 선정하여, 음소 빈도가 일정한 교육용 단어 위주의 녹음 단어셋을 선정, 음성DB를 구축 후 음향모델링을 생성했으며, 한국인 영어 발음들에 대한 오류 유형을 분석 후, 단어별 음소기반 인식 네트워크를 구성하였다. 시스템의 성능을 분석하기 위하여, 오류유형인 음소대체, 삽입, 삭제, 이중모음 분리, 강세오류를 지닌 음성 데이터에 대해 인식실험을 통하여 성능을 분석하였다.

본 논문의 2장에서는 영어 음소와 혼동 가능한 한국어 음소셋 선정 그리고 음성 DB구축, 3장에서는 발음 교정 시스템을 위한 음향 모델링 방법, 4장에서는 오류유형 검출을 위한 언어네트워크 구성, 5장에서는 한국인 발음 데이터 인식 및 검출 성능 그리고 마지막 6장에서는 결론을 맺었다.

### II. 음성 DB 수집 및 음소 선정

한국인의 오류 유형 형태 분석과 혼동 가능한 한국어 음소 등을 반영하기 위하여 통합 음소셋을 선정하여, 녹음 단어들을 만들어 영어 원어민 화자 및 한국

인 화자에게 녹음을 하여 음성 DB를 수집하였다.

## 2.1 영어 음소 기준 및 한국어 음소 설정기준

영어 음소는 CMU 음소셋을 기준으로 하여, 기본자음 24개, 모음 28개, 변이음 16개 음소들과 영어 음소와 음가가 비슷하지만 음성 자질이 다른 한국어 음소인 /ㄱ ㄷ ㅂ ㅅ ㅈ 으/ 6개 음소를 추가적으로 선정하여, 총 74개 음소셋을 선정하였다. 미국인 음성 데이터에는 한국어 음소들을 제외한 음소 기호로 전사되었으며, 한국인 음성 데이터는 한국어 음소 기호가 포함되어 전사되었다.

## 2.2 영어 음성 및 한국어 음성 DB 수집

음성DB를 수집하기 위하여 한국인 및 미국인 남/녀 학생들을 대상으로 한국인은 1인당 2회에 걸쳐 163단어를 미국인은 1인당 10회에 걸쳐 2145단어를 동부 표준 발음을 청취 후, 핸드마이크와 헤드셋을 이용해 2채널로 16KHz 샘플링률과 16비트 샘플당 비트수의 Windows PCM 포맷으로 사무실 환경에서 녹음하였다. 녹음 후 앞뒤 300ms 정도의 묵음 존재, 최대 진폭 15000 정도의 값을 기본으로 양질의 데이터를 선별하는 작업을 하여 녹음 데이터의 약 93%정도의 데이터를 선별하였다. 표 1은 미국인 및 한국인에 대한 남녀 성비로의 녹음 발화자의 구성비를 나타낸다.[2].

표 1. 발화자 남/녀, 국적 구성수

구 분	미국인			한국인			계
	남	녀	계	남	녀	계	
인원	121	133	254	163	248	411	665

## III. 영어 및 혼동 가능 한국어 음소 기반의 음향 모델링 기법

음향모델은 선정된 음소를 기준으로 한국인 화자에 의해 발화된 데이터와 미국인 화자에 의해 발화된 데이터를 구분하여 각각 음소 단위 HMM모델을 생성 후 통합시키는 방식을 사용하였다. 미국인과 한국인 음성 데이터에 대해서 pre-emphasis와 10ms 마다 25ms 크기의 Hamming window를 적용하여 특징벡터를 추출하였다. 특징 벡터는 사람의 청각기관의 특징을 고려

한 MFCC를 사용하여 12차 MFCC와 log energy 그리고 각각의 delta 및 delta-delta를 구한 39차 특징 벡터를 사용하였다. 음향모델은 left-to-right 형식의 3개의 상태를 기반으로 하였다. 한국인 음성 데이터에 대한 음소셋과 외국인 음성 데이터에 대한 음소셋이 상이기 때문에, 다른 방식으로 2 종류의 음향모델을 생성하였다.

### 3.1 영어 음소 모델링 기법

미국인 발음을 반영한 음향 모델은 전체 음성 데이터에 대한 global mean과 global variance로 프로토타입을 구성 후, Baum-Welch re-estimation 알고리즘을 이용하여 68개 음소 단위 1차 음향모델을 생성하였다. 생성된 음향 모델을 기반으로 음소 경계 시간 정보를 추출 하여, K-means 알고리즘을 통해 음소 단위 초기 음향모델을 생성한 후, Baum-Welch re-estimation 알고리즘을 통해 2차 최종 모델을 선정하였다.

### 3.2 한국인 발화데이터 음소 모델링 기법

한국인이 발음한 음성데이터는 오류 발음 유형이 포함되어 있기 때문에 혼동 가능한 한국어 음소를 반영한 모델을 만들기 위해서는 3단계를 거쳐 최종 모델링을 하였다.

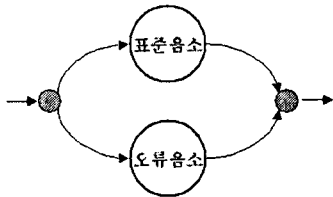
첫 번째 단계는 생성된 영어 음소 단위 음향모델을 기반으로 자동 세그멘테이션 작업 후, 한국인 영어 음성DB의 30%의 음성 데이터에 대해 음성학 전문가들의 수작업을 통해 오류 유형 분석 및 전사 그리고 시간 정보 수정등의 매뉴얼 세그멘테이션 작업을 하였다. 두 번째 단계는 수작업된 30%의 음성데이터에 대해 K-means 알고리즘을 이용하여 74개 음소 단위 HMM의 초기 음향 모델을 생성 후, Baum-Welch re-estimation 알고리즘을 이용하여 2차 음향 모델을 생성한다. 세 번째 단계는 생성된 모델을 이용하여 나머지 70%의 음성 데이터의 자동 세그멘테이션 후, 30% 매뉴얼 세그멘테이션 작업된 음성 데이터들과 통합하여 K-means 알고리즘을 이용하여 초기 음향 모델을 생성 한 후, Baum-Welch re-estimation 알고리즘을 이용하여 최종 모델을 생성한다.[3][4]

미국 원어민 음성 데이터와 한국인 발화 음성 데이터로 생성된 데이터는 Test data로 인식 실험에서 적절한 mixture 개수를 찾은 결과, 미국 원어민의 데이터에 대해서는 mixture 7로 한국인의 데이터에 대해서는 mixture 10으로 최종 모델을 생성하였다.[4]

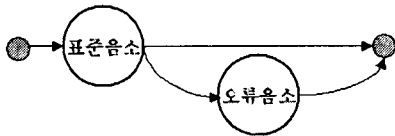
## IV. 언어 네트워크 구성

### 4.1 오류 유형을 검출 할 수 있는 음소기반 인식 네트워크

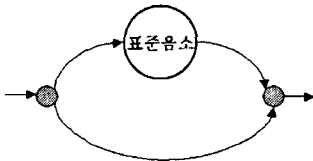
한국인의 영어 발음에서 오류 유형은 크게 음소 또는 변이음의 대치, 이중모음의 분리와 음소의 삽입, 삭제에 있다. 전자는 한국어와 영어의 음운 체계, 특히 음소목록이 차이가 나기 때문에 생기는 유형이며 후자는 한국어와 영어의 음절구조가 다르기 때문에 생기는 유형이다. 한국인의 영어 단어 발음시 발생하는 오류 유형은 영어의 유성 파열음인 /g,d,b/는 한국어 음소인 /ㄱ,ㄷ,ㅂ/로 음소를 대치하는 경우, 3 음절 이상의 자음연쇄시 [으] 또는 음절꼬리의 자음 다음에는 [이]가 삽입되는 경우, /er//kw/ 발음시 [r][w] 발음 등이 삭제되며 경우와, 모음의 경우는 이중모음/au//oy/등의 발음을 단음소로 분리한 경향과 강세에 대한 오류 유형을 보인다. 다음 그림 1은 오류 유형을 검출하기 위한 음소기반 인식네트워크를 설명한 것이다.[5]



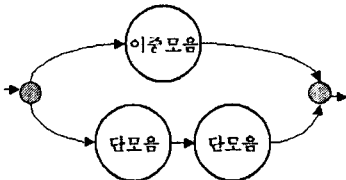
a) 음소대치 및 모음 강세 오류 검출 네트워크



b) 음소 삽입 오류유형 검출 네트워크



c) 음소 삭제 오류유형 검출 네트워크



d) 이중모음 분리 오류유형 검출 네트워크

그림 1. 오류 발음 형태를 검출하기 위한 음소 기반 인식 네트워크

## V. 발음형태에 따른 인식률 분석

본 논문에서는 한국인의 영어 단어 발음에서 오류 유형 검출을 위한 시스템을 구축에 대하여 설명하였다. 분석하는 관점을 대치,삽입,삭제,이중모음 분리,강세의 5개 오류유형에 대하여 정확한 미국 원어민 발음과 부정확한 한국인 발음을 입력 데이터로 구성하여 인식결과를 구하고, 이를 실험음성학 전문가들의 검토를 통하여 정검출과 오검출율로 성능을 테스트 하였다.

### 5.1 오류 발음 형태별 음소 검출률

음소의 삽입,삭제,대치,이중모음분리,강세등의 오류 유형별로 대표되는 단어 10개를 선정 한 후, 각 단어별로 1차로 실험음성학 전문가들이 선정한 정확한 발음 10개, 부정확한 발음 10개의 데이터를 준비하여, 오류 유형별 정확한 발음 100개, 부정확한 발음 100개의 테스트 데이터를 구성하였다. 4장에서 설명한 음소기반 인식네트워크를 구성하여 인식실험을 실시하였다.

#### (1) 대치

한국인은 /g,d,b/음소로 시작하는 단어 발음시 한국어 /ㄱ,ㄷ,ㅂ/로 발음하는 경향이 많다. 이를 검출하기 위해 /g,d,b/로 시작하는 단어인 bus, dish, gold 등의 단어를 선정 후, 정확한 발음 입력시는 /g,d,b/등의 음소가 인식이 되어 정검출이 되는 것이며, 부정확한 발음 입력시는 /ㄱ,ㄷ,ㅂ/가 출력이 되어 부정확한 발음의 정확한 검출(정검출)이 되는 것이다.

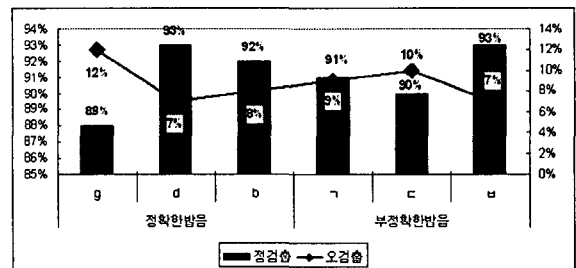


그림 2. 음소 대치 오류 인식시 검출률

#### (2) 삽입

한국인은 자음 세 개의 연속 또는 마지막 음절의 끝이 자음으로 끝날 경우 /으//이/음소를 삽입하는 유형이 있다. 이러한 오류유형 단어는 strike, church등의 단어가 있으며, 정확한 발음 인식시에는 [으][이]의 음소가 검출되지 않아야 하며, 부정확한 발음 인식시에는 [으] 또는 [이]가 출력되어 정검출이 되는 것이

다.

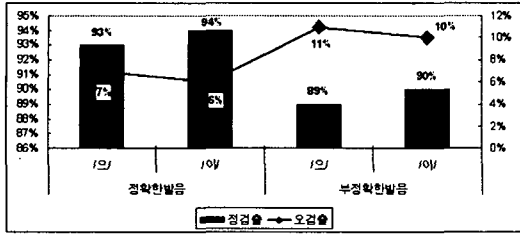


그림 3. /w//i/삽입 오류 인식시 검출률

(3) 삭제

한국인은 /er/ /kw/ 발음에서 [r] [w] 발음을 각각 삭제하는 경향이 있으며, 이 오류유형을 검출하기 위하여 teacher, quiz, wood 등의 단어를 선정하였다. 각각의 단어 발음시 정확한 발음은 [r] [w] 음소가 인식되어야 하며, 부정확한 발음 입력에는 [r] [w] 음소가 인식되지 않아야 한다.

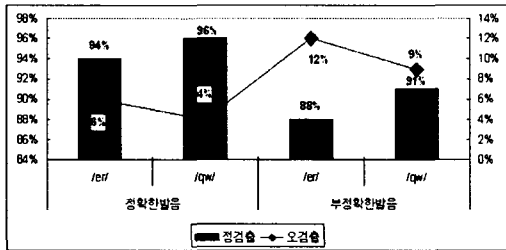


그림 4. /er/ /qw/에 대한 음소 삭제 오류 검출률

(4) 이중음소 분리 및 강세

/ai,au,ou,oi/ 등 이중모음 발음시 분리하여 발음하는 경우로 fine, town, toy, slow 등의 10개 단어를 선정하였다. 부정확하게 발음한 경우에는 /a/i/ 등 이중모음이 분리되어 인식된다. 강세는 음소셋에 제 1강세와 미강세 모음을 선정하여 모델을 생성하였으므로 강세 모음 대신 미강세 모음이 인식될 경우 잘못된 검출(오검출)로 본다.

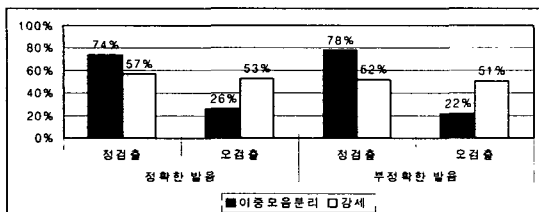


그림 5. 이중모음 분리 및 강세 오류 검출률

한국인 영어 발음시 단어내 나타날 수 있는 오류유형인 삭제, 대치, 삽입에 대해 부정확하게 발음한 데이

터 및 정확하게 발음한 데이터는 90% 이상 인식률을 보이며 바람직한 검출 결과가 나타났으나, 모음에 대한 이중분리 및 강세 오류는 75%이하로 낮게 나타났다. 이는 한국인 음성 DB구축시 모음에 대한 manual segmentation 작업의 부족과 training data에 대한 에너지 불균형에 문제로 추측되어 진다.

VI. 결론

본 논문에서 제안한 영어 음소 및 혼동 가능한 한국어 음소 단위 음향 모델과 오류 발음 분석을 통한 음소 단위 인식네트워크를 기반으로 한 한국인의 영어 오류 발음 검출 시스템에서 오류 형태인 대치, 삽입, 삭제 형태는 인식을 통한 검출률이 90% 이상으로 높았으나, 모음에 대한 이중모음 분리 및 강세 오류의 인식을 통한 검출률은 75%이하로 낮았다.

낮은 검출률을 보완하기 위하여 이중 모음 분리는 해당 모음에 대한 duration 검출과 표준 음소의 duration의 비교방식을 통한 방법을 연구 중이며, 강세는 모음에 대한 피치 및 에너지를 통한 새로운 접근방식이 상기 결과를 보완하는 방법으로 연구 진행 중이다.

추후 리듬적인 요소 및 문장 내 발음 교정 기술이 연구가 영어 발음학습에 적용에 있어서 단어 단위가 아닌 구 및 문장단위의 발음 학습 기법이 연구가 되어야한다.

참고문헌

- [1] 백승권, 최정규, 한민수, "영어단어발성시의 오류교정 기술에 관한 연구," 음성과학회 논문집 No.8, pp.83-90 2000년.
- [2] L. Neumeyer, H. Franco, M.Weintraub and P. Price, "Automatic Text-Independent Pronunciation Scoring fo Foreign Language Student Speech," Proc. of ICSLP 96, pp.1457-1460. 1996.
- [3] 김무중, 권철홍, "HMM 및 보정 알고리즘을 이용한 자동 음성 분할 시스템," 음성과학 제 7권 제 4호, pp.256-274, 2002.
- [4] Y.J Chung and C.K. Un, "Use of different numbers of mixtures in continuous density hidden Markov Models," IEEE Lett., Vol.29, No.9, pp.824-825, 1993.
- [5] 김효숙, "한국인을 위한 영어발음교정 시스템 'Dr. Speaking' 소개," 대한음성학회 창립 25주년 기념 학술대회 발표 논문집, pp.47-50, 2002.