한국인 영어 학습자의 발음 정확성 자동 측정방법에 대한 연구

윤원희* 정현성** 장태엽***
* 경남대학교 영어학부 ** 한국교원대학교 영어교육 *** 한국외국어대학교 영어과

A Study on Automatic Measurement of Pronunciation Accuracy of English Speech Produced by Korean Learners of English

Weonhee Yun* Hyun-Sung Chung** Tae-yeoub Jang***
* Division of English, Kyungnam Univ. **English Education, Korea National Univ. of Education
*** English Division, Hankuk University of Foreign Studies

whyun@kyungnam.ac.kr, hschung@ knue.ac.kr, tae@hufs.ac.kr

Abstract

The purpose of this project is to develop a device that can automatically measure pronunciation of English speech produced by Korean learners of English. Pronunciation proficiency will be measured largely in two areas; suprasegmental and segmental areas. In suprasegmental area, intonation and word stress will be traced and compared with those of native speakers by way of statistical methods using tilt parameters. Durations of phones are also examined to measure speakers' naturalness of their pronunciations. In doing so, statistical duration modelling from a large speech database using CART will be considered. For segmental measurement of pronunciation, acoustic probability of a phone, which is a byproduct when doing the forced alignment, will be a basis of scoring pronunciation accuracy of a phone. The final score will be a feedback to the learners to improve their pronunciation.

Ⅰ. 서론

국제 공용어로서인 영어의 중요성이 날로 증대되고 있음에 따라 효과적인 영어교육의 방법도 다각적으로 연구, 개발되고 있다. 기초 생활회화, 문법 등에 중점을 두었던 영어교육이 좀 더 고차원적인 활용, 즉, 국제회의나 비즈니스를 위한 영역으로 확대되면서 의사소통 기능뿐만 아니라 정확한 발음 구사의 중요성이 더욱 강조되는 시점이 되었다. 하지만 영어교육의 영역에 비해 발음교육에 대한 체계적인 연구는 상대적으로 부족해왔던 것이 사실이다.

최근에는 의사소통 능력의 관점에서 영어 발음 교육은 이전의 접근 방법과 차이점을 보이고 있다. 즉 분절음(segment)의 정확성뿐만 아니라 발음의 유창성에 수반되는 억양, 음조, 리듬 등의 초분절음적 자질의 학습이 중요시 되고 있다.

일반적으로 영어 발음의 정확성 및 유창성을 판단하는 것은 영어 모국어 화자의 영역에 속한다. 그러나 영어 교육자 혹은 평가자가 영어 모국어 사용자가 아닐 경우, 영어 학습자의 발화가 어느 정도 모국어 화자의 발음에 근접한지를 판단하기란 쉽지 않다. 영어 교육에 숙련된 모국어화자라 할지라도 청취테스트에 의존하는 기존의 평가방법에는 평가자들 간 의견 차이, 평가자내부의 일관성 부족 등으로 정확하고 객관적인 평가결과를 도출하는데 한계가 있다. 또한 영어 학습자의 경우 자신의 발음을 매번 영어 모국어 화자에게 확인 받기가 현실적으로 어려우므로, 발음의 정확도 향상에 즉각적인 효과를 보기가 힘들다.

그러므로 객관적이고 실용적으로 영어 발음의 정확성을 측정할 수 있는 방법의 개발이 요구되며, 특히 평가자의 주관적인 판단에만 의존하지 않고 자동화된 기계의 작동을 통해 발음의 정확성을 측정하여 점수화 함으로서 즉각적인 평가를 할 수 있는 방법이 개발된다면 영어 발음교육에 큰 효과를 나타낼 것이다.

Ⅱ. 연구 방법 및 내용

1. 발음 측정기의 개괄

영어 발음 측정기의 목적은 영어 학습자의 발음을 측정하는 것이다. 따라서 주어진 문장을 영어 학습자가 읽음으로서 미리 준비된 음성 인식기(발음 측정기)는 그 문장의 정확성을 모델링한 통계에 따라 원어민의 발음과 비교하여 학습자의 영어 발음을 점수화 시킨다. 이 점수는 발음에 대한 전체 점수뿐만 아니라 각 모듈 즉 분절음, 억양 및 강세, 길이 등의 세 가지 모듈로 나누어 점수화 할 수 있도록 만들고, 그것을 통하여 영어 학습자가 어느 부분에 더욱 노력을 기울여야 하는지 피드백(feedback) 되도록 한다.

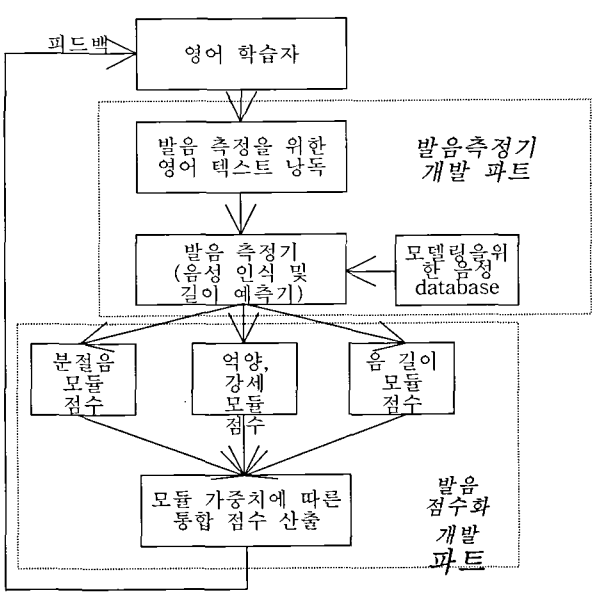

그림 1 영어 학습자와 발음 측정기 사용에 관한 순서도

<그림 1>에 따르면 영어학습자는 주어진 발음 측정용 텍스트를 읽고 그 음성을 마이크를 통하여 발음 측정기에 입력 하게 된다. 발음 측정기는 미리 조성된 음성 데이터베이스를 통하여 모델링 된 음성 인식기 혹은 음 길이 예측기다. 입력된 음성은 음향 파라미터로 변환되어 발음 측정기의 세 가지 모듈로 보내진다. 이 세 모듈은 각각, 분절음 모듈, 억양, 강세 모듈, 음 길이 모듈로 나눈다. 각 모듈에서는 확률 혹은 길이 예측값 등을 이용해 입력된 영어 화자의 발음에 대한 점수를 출력해 낸다. 각 모듈에서 보내진 발음 점수는 모듈별 가중치 혹은 영어 학습자의 점수에 대한 신뢰성을 가질 수 있을 점수로 환산하여 영어 학습자에게 피드백시킨다. 모듈 점수를 통해서 부족한 부분을 다시 훈련할 수 있도록 도와준다.

2. 분절음 측정기 개발

발음 측정기에 있어서 분절음의 발음 정확도를 측정하는 것은 음성 인식 기술을 활용하는 것이다. 따라서 분절음에 관련하여 발음 측정기는 분절음 인식기를 개발하는 것과 같다.

본 연구에 사용될 음성 인식 기술은 기반 연구가 충분히 이루어져있는 HMM을 이용해서 구축하려한다. 또한 HMM을 이용한 음성 인식기의 경우 기존의 HTK(Hidden Markov Toolkit)을 이용하여 복잡한 프로그래밍 없이 간단하고 값싸게 구축할 수 있는 장점이 있다.[1] 분절음 인식기는, 분절음의 특징, 즉 자음과 모음의 조음 방법과 조음 위치에 따라 달리 나타나는 스펙트럼상의 주파수 특징을 모델링하는 것이다.

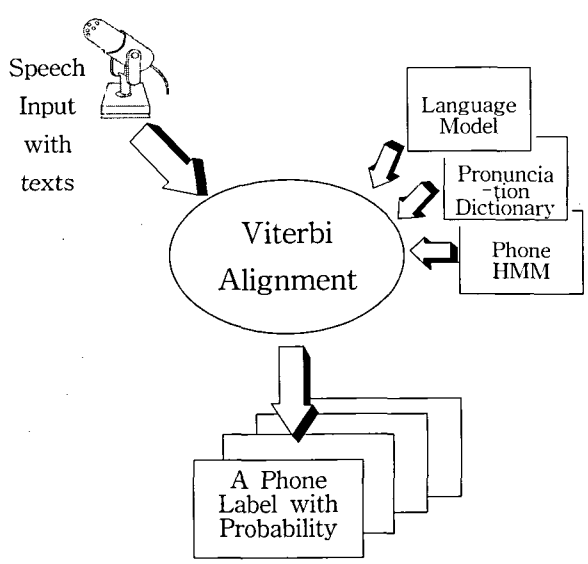

그림 2 Viterbi 알고리듬을 이용한 음성 분할의 개괄도

본 연구에서 개발하려고 하는 발음 측정기의 경우 음성 인식 기술을 이용하여 분절음의 발음 정확도를 확률적으로 계산해 낼 수 있다. 전술하였듯이, 발음 측정기는 학습하지 않은 문장을 입력대상으로 삼지 않는다. 분절음 인식기는 일반적인 음성 인식기 구축과 같은 방법으로 이루어지지만, 발음 측정기로써 입력

대상은 음성뿐만 아니라 그 음성에 해당하는 텍스트도 포함한다. 이 경우, 문장에 대한 학습자의 음성과 그 문장 텍스트를 분절음 인식기에 입력하면 각 분절음 위치에 따른 음성 분할이 이루어지도록 인식기를 조정할 수 있다. 이러한 과정을 forced alignment라 부르고 이 과정을 <그림2>에서 보여주고 있다.

위 과정을 통하여 얻어진 음성 분할은 결국 분절음 단위의 레이블링을 자동적으로 얻어내는 결과를 가져온다. 분절음 단위의 레이블이란, 입력 음성과 문장에서 문장 속에 들어간 모든 분절음의 위치를 음성 파형 위에 표시할 수 있도록 그 음의 시작과 끝에 해당하는 시간 정보를 나타내는 것이다. 자동 음성 인식기를 통하여 음성 분할을 했을 경우 해당 분절음의 시간 정보와 더불어 각 분절음의 확률을 얻어낼 수 있다. 이 확률로부터 원어민의 발음과 실제 영어 학습자의 발음의 차이점을 유추해 낼 수 있다.

3. 강세, 억양 등의 음향 자질 추출

강세나 억양의 음향단서로써 F0의 중요성이 많은 연구들을 통해서 밝혀져 있다.[2][3][4][5] 또한 intensity와 duration도 무시할 수 없는 단서로 인식되고 있으므로 이들의 고려도 필요하다[6][7].

구체적으로, 독립어휘 강세를 정확히 발음할 수 있는 가에 대한 모델로서 2음절 이상의 목표 단어들을 추출하고 이 단어들에 대한 원어민 화자의 발화데이터를 확보하고 분석하여 강세음절과 비강세 음절에서 나타나는 F0와 intensity를 측정하여 그 수치를 통계적으로 모델링한다. reference model을 만들 때에는 다수의 화자로부터, 일정하게 운율구조가 고정되도록 만들어진 상태에서 발화된 음성을 녹음하게 하여 화자 및 기타 환경의 변이를 최대한 고정시키는 효과를 거둔다.

억양의 정확성을 측정하기 위한 방법으로 억양경계톤(intonational boundary tone)의 모델링을 통해서 비교해 볼 수 있다. intonation meaning의 가장 핵심적인 정보라고 알려져 있는 억양의 오른쪽 경계에 나타나는 F0의 양상을 모델링 할 수 있다면 원어민 발화와 영어학습자 발화에 있어서 운율적인 음성적 구현의 차이가 무엇인지를 파악할 수 있다.

이를 위해서 사용될 모델은 Tilt intonation 이론이다.[8] 언어학적인 차원에서 운율정보의 표기를 위해 가장 널리 쓰이는 ToBI(Tone and Break Indices)가 있으나[9] 이 모델은 억양경계톤의 모델링을 자동적인 방법으로 구현하는데 그 한계가 있다. 보다 실용성이 있고 대용량데이터를 통한 모델링에 용이한 Tilt 모델을 사용할 것이다. 이 모델의 핵심은 원어민이 발화하는 대표억양을 가진 문장들의 boundary tone의 모양을 모델링하여 이를 한국인 영어학습자의 것과 비교하여 그 차이를 보여주는 것이다. Tilt 모델의 파라미터 추출을 위한 요소는 <그림 3>에서 볼 수 있다.

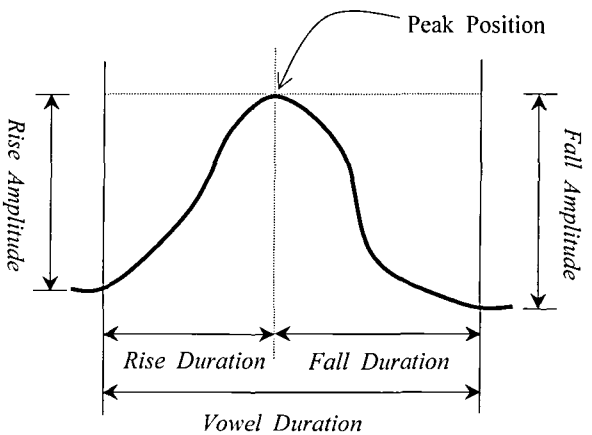

그림 3 Tilt 파라미터 계산을 위한 요소
([10]에서 추출)

추출된 각 파라미터들을 조합하여 최종적으로 하나의 수치를 산출해내는데 이 수치가 Tilt 값이며, 아래와 같은 조합을 사용한다.

(1) Tilt 파라미터를 이용한 Tilt 값 추출

$$tilt = \frac{1}{2}\left(\frac{|A_{rise}| - |A_{fall}|}{|A_{rise}| + |A_{fall}|} + \frac{D_{rise} - D_{fall}}{D_{rise} + D_{fall}}\right)$$

A: Amplitude

D: Duration

이러한 방식으로 지정된 문장에 대해 원어민이 발화한 각 억양경계톤의 Tilt값을 모델링하여 보관하고 있는 상태에서 한국인 학습자가 동일한 문장을 발화하게 하여 그 값들을 비교해 보면 운율에 대한 발음정확도를 측정하는 하나의 지표로 사용할 수 있다.

4. 음의 길이에 대한 특징 추출

화자변이로 인한 음 길이의 편차를 극복하기 위해서 음성 데이터베이스의 길이 정보 추출 시, 길이의 절대값을 측정 기준치로 사용하지 않고, 개별 분절음의 길이를 모두 z-score로 전환하여 사용한다. 각 분절음의 길이를 우선 log값으로 전환 후, 평균과 표준편차를 이용하여 각 분절음의 z-score를 계산한다.

(2) z-score = (분절음의 길이 - 개별 분절음의 평균값) / 표준편차

개별분절음의 z-score에는 이미 운율자질에 따른 길이의 변화가 반영되어 있다. 따라서, 별도의 길이 추출 모델이 필요치 않다. 발음 측정을 위한 텍스트와 모델링을 위한 음성 데이터베이스의 문장이 동일하기 때문에 동일 환경의 동일 분절음의 길이에 대한 z-score를 영어학습자가 낭독한 영어텍스트의 그것과 비교 측정함으로써, 영어학습자의 발음의 정확성을 측정할 수 있다. 동일한 분절음에 대해 영어학습자의 발음이 얼마나 정확성을 보이는가를 측정하기 위해, 본 연구에서는 영어학습자의 영어발음에서 길이의 정확성의 기준을 음성합성 및 음성인식 등에서 흔히 쓰이는 평균제곱근오류(RMSE: root mean squared prediction error)와 상관계수(correlation coefficient)를 사용하여 구하고자 한다. 기존의 합성 및 인식을 위한 연구결과에서는 평균제곱근오류 20 밀리세컨드 전후, 상관계수 0.8 이상을 우수한 성능으로 판단하고 있다.

영어학습자의 리듬구현의 오류분석을 위해서는 운율정보에 따른 길이 모델링이 필요하다. 길이 모델링을 통해, 영어학습자가 중대한 오류를 범했을 경우, 그 개선방향을 제시할 수 있다. 길이 모델링을 위해서는 문장의 운율경계 정보를 추출하고, 그에 따른 음의 환경을 매개변인으로 하여 길이를 예측한다. 비선형적 운율구조 형성을 위해서는 [11]을 사용할 것이다.

5. 발음의 점수화

발음 점수화 개발 파트는 각 모듈, 즉 분절음, 억양과 강세, 그리고 음 길이의 모듈에서 원어민과 학습자의 발음을 비교하여 그 결과물을 어떻게 점수화 할 것인가를 연구하는 부분이다. 따라서 각 모듈별 점수를 어떻게 계산할 지에 대한 연구를 포함하여, 세 모듈의 점수를 동일한 가중치로 처리할지 혹은 모듈마다 다른 가중치를 적용할지에 대해서 연구한다.

Ⅲ. 결론

본 연구는 한국인 영어 학습자의 영어 발음을 음성공학 기술을 활용하여 자동으로 측정할 수 있는 방법을 연구하는 것이다. 이러한 연구는 학문적으로 음성학과 음성 공학이라는 두 학제 사이의 결과물로써 그 학문적 발전에 기여할 것이다. 이는 인문학 내에서 뿐만 아니라 다른 학문 제 분야의 공동 연구를 촉진하는데 긍정적 영향을 미칠 것으로 기대된다.

본 연구의 실용적 결과물은 영어 발음 측정기이다. 현재 영어 학습과 교육에 높은 사회적 비용을 지불하고 있는 것이 사실이다. 영어 발음 측정기는 이러한 사회적 비용을 낮출 수 있는 효과적인 교육용 기자재로서 쓰일 수 있을 것이다.

참고문헌

[1] Young, S., J. Jansen, J. Ollason & P. Woodland, *HTK Book*, Entropic. 1996.

[2] McClean, M. D. & W. R. Tiffany. "The acoustic parameters of stress in relation to syllable position, speech loudness and rate", *Language and speech* 16. 283-291. 1973.

[3] Lieberman, P., "Some acoustic correlates of word stress in American English", *Journal of Acoustical Society of America* 32, 451, 1960.

[4] Bolinger, D. L, "A theory of pitch accent in English", *Word* 14, 109, 1958.

[5] Morton, J. & W., "Jassem, Acoustic correlates of stress", *Language and speech 8, 159, 1965.*

[6] Fry, D. B. "Duration and intensity as physical correlates of linguistic stress", *Journal of the Acoustical Society of America* 27, 1141, 1955.

[7] Nakatani, L. & C. Aston. "Perceiving stress patterns of words in sentences", *Journal of the Acoustical Society of America* 63, S55, 1978.

[8] Taylor, P., "Analysis and synthesis of intonation using the tilt model", *The Journal of the Acoustical Society of America*, 107(3), 1697-1714, 2000.

[9] Beckman, M. E. & J. Hirschberg, *The ToBI annotation conventions*. MS. and accompanying speech materials, Ohio State University, 1994.

[10] Jang. T. Y., *Phonetics of segmental F0 and machine recognition of Korean speech.* Ph.D. Dissertation. University of Edinburgh, *2000.*

[11] Huckvale, M., "Representation and processing of linguistic structures for an all-prosodic synthesis system using XML", *Proceedings of Eurospeech '99*, 4, 1847-1850, 1999.