

## 영어 단어 학습시의 발성 교정 기술에 관한 연구

Study on the pronunciation correction in English words

백 승 권, 최 정 규, 한 민 수

Seung Kwon Beack, Jungkyu Choi, Minsoo Hahn

### ABSTRACT

In this paper, we implement an elementary system to correct accents and pronunciations in English words spoken by non-native English speakers. In case of the accent evaluation, energy and pitch information are used to find stressed syllables, and then we extract the segment information of input patterns using a dynamic time warping method to discriminate and evaluate accent position. For the pronunciation evaluation, we utilize the segment information using the same algorithm as in accent evaluation, and perform the spectral distance measure for each phoneme between input patterns and reference patterns. Based on these spectral distances, we decide whether to recommend the pronunciation correction or not. Our results show that 98 percent of accent and 71 percent of pronunciation evaluation agree with the perceptual measure.

### 1. 서 론

외국어를 습득하는데 있어서 액센트 및 발음을 교정하기 위해서는 전문가의 지도를 받거나 어학용 기기를 통한 반복 청취 및 발성을 통해 습득하는 것이 상례이다. 본 연구에서는 외국어를 익히고자 하는 초보자에게 좀 더 빠른 시일 내에 본인의 발성 및 성취도에 대한 자기 진단을 하면서 외국어를 습득할 수 있도록 발음 및 액센트를 고립단어 범위에서 평가하고 교정할 수 있는 초벌 시스템을 구현하였다. 이 시스템은 학습자에게 잘못 발성된 액센트 및 발음부분에 대해서 지적을 해주고 이를 인지한 학습자는 잘못된 발성 부분에 대해 지속적인 반복과 학습을 통해 좀 더 빠른 시일 내에 교정을 받을 수 있도록 한다. 본 시스템을 구현하기 위해서 음성 신호의 기본적인 여러 특징들과 인식 알고리즘을 적용하였으며 단어발성에 대해 평가 및 교정이 가능한 초벌 시스템을 구현하였다.

## 2. 실험 과정

### 2.1 Data base 구축

발성교정 시스템을 구축하기 위해서 화자가 발성한 음성신호와 비교하여 이를 평가해 줄 수 있는 기준 패턴(Reference Pattern)이 필요하다. 이를 기준으로 화자가 발성한 단어의 액센트와 발음을 교정하게 되므로 기준 패턴은 발음이 정확한 영어 원어민이 발성한 단어음성이 필요하다. 따라서 본 논문에서 제안하는 시스템에서는 영어 원어민의 발성 DB를 얻기 위해 간접적으로 어학용 테이프를 통해 5개 단어를 수집하였다. 단어의 선정 조건은 다음 사항을 고려하여 수집하였다.

- 액센트 위치에 따른 단어 선정.
- 액센트 받은 음절 내의 유성, 무성 포함 여부에 따른 선정.
- 다양한 음소를 고려한 단어 선정.

DB 목록은 표 1과 같다.

### 2.2 액센트 교정 및 평가

일반적으로 한국어 단어에는 무강세, 즉 액센트가 없는 언어라고 하며 실제로 한국어 음성신호에서 큰 강세 변화를 찾아보기 힘들다. 한편, 영어는 많은 정보가 액센트에 의해 전달되므로 발음이 정확하다 해도 액센트의 위치가 틀러지면 의사소통이 안되는 경우가 많다. 따라서 한국인이 영어를 배울 때 가장 어려움을 느끼고 틀리기 쉬운 것 중의 하나가 단어 내에서의 액센트 패턴이다.

발성에서의 액센트라 하면 음의 세기나 강세를 말하거나 발성한 화자의 억양이나 지방 사투리를 뜻하기도 하는데, 이는 음성인식이나 화자인식을 하는데 중요하게 활용되기도 한다.[1, 2] 본 논문에서의 액센트는 전자를 뜻하며 이때 액센트는 음의 세기로 간주하여 에너지 함수에 의해 구하도록 한다.[3]

표 2는 액센트 위치를 에너지 궤적의 최대치로 추정하였을 때 찾은 입력 단어의 액센트 위치와 실제, 즉 기준 패턴의 액센트 위치를 비교한 결과이다. 이 표에서 무성음을 포함하고 있는 음절에 강세가 올 경우나 상대적으로 긴 유성음 구간을 갖고 있는 단어의 경우 액센트 위치를 찾는 데 오류가 발생함을 알 수 있다.

표 1. 기준 패턴 DB

단어 \ 특성	음절수	액센트 위치	액센트에서의 음소 특성
Obvious	3음절	1음절	V
Avocation	4음절	3음절	U+V
Reluctantly	4음절	2음절	V+U
Arizona	4음절	3음절	U+V
However	3음절	2음절	V

V: voiced . U: unvoiced

일반적으로 3음절 이상의 단어에서의 액센트는 강세와 약세가 존재하게 된다. 따라서 에너지 궤적에서 두 개의 최대값을 찾아 후보자로 두고 이들 중 피치궤적의 최대값과 거리 측정을 하여 최소거리를 갖는 후보자를 강세 위치로 판별하였다.

표 2. 에너지 궤적의 최대치를 이용한 액센트 평가

구분 단어	시스템이 찾은 액센트 위치	실제 액센트 위치	기준패턴과 비교 후 내린 최종 평가결과
Obvious	1음절	1음절	Correct
Avocation	3음절	3음절	Incorrect
Reluctantly	2음절	2음절	Incorrect
Arizona	1음절	3음절	Incorrect
However	1음절	2음절	Incorrect

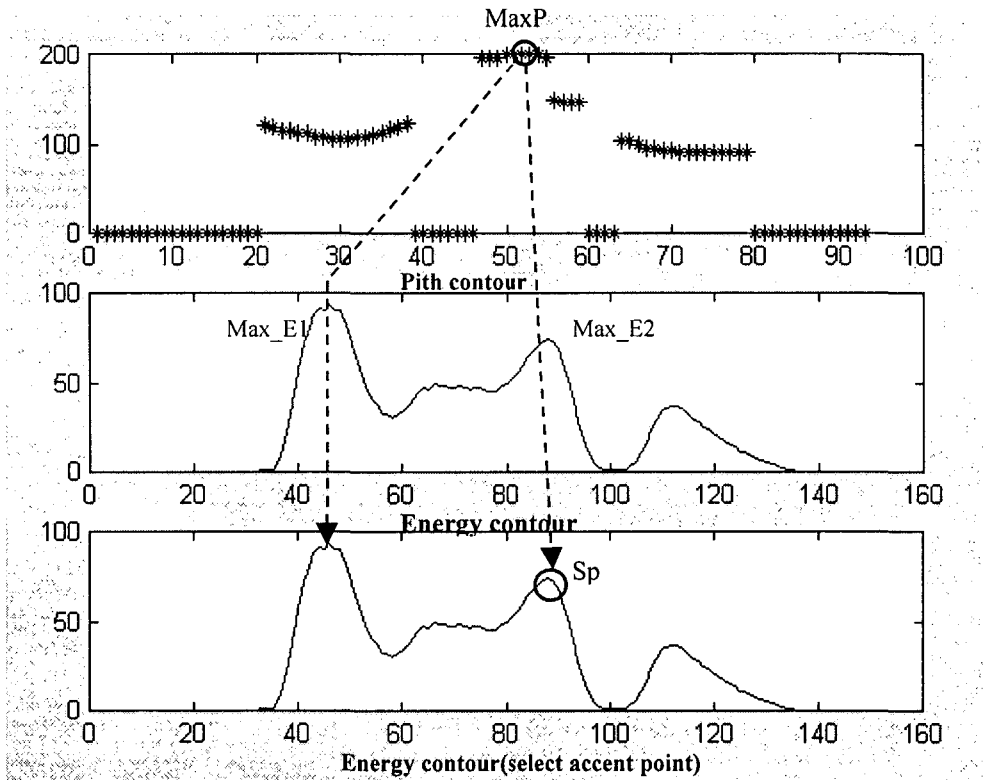


그림 1. 'however'에서 에너지 정보와 피치 정보를 이용한 액센트 위치 판별 예

표 3. 에너지와 피치정보를 이용한 액센트 평가

구분 단어	시스템이 찾은 액센트 위치	실제 액센트 위치	기준패턴과 비교 후 내린 시스템의 최종 평가
Obvious	1음절	1음절	Correct
Avocation	3음절	3음절	Incorrect
Reluctantly	2음절	2음절	Incorrect
Arizona	3음절	3음절	Incorrect
However	2음절	2음절	Incorrect

$$Sp = \min(\text{dist}(\text{MaxE}_i, \text{MaxP})), \quad (2.2.1)$$

$\text{dist} : |A - B|, Sp$  : Selected frame,  $\text{MaxE}_i$  : Maximum energy frame 1,2

$\text{MaxP}$  : Maximum pitch point

그림 1은 본 연구에서 제시한 방법으로 액센트 위치를 판별하는 예를 나타내고 있으며 식 (2.2.1)을 적용하였다. 표 3은 제시한 알고리즘에 의해 찾은 액센트 위치를 가지고 평가를 내린 결과 표이다.

표 2와 표 3을 살펴보면 시스템이 찾은 액센트 위치와 실제 액센트 위치가 일치하더라도 기준패턴과 비교 후 내린 시스템의 최종평가는 부정확한 것으로 평가를 내렸음을 확인할 수 있다. 이는 액센트 위치를 정확히 찾았다하더라도 발성 패턴과 기준패턴을 비교하여 평가를 내릴 경우 두 패턴의 음절 간 지속시간 차이로 인해 발성패턴의 액센트가 속한 음절의 위치가 기준패턴의 어느 음절에 속하는지 판단하기 어렵기 때문이다. 이러한 현상은 강세를 받는 음절에서 에너지의 궤적이 커지는 것에 비례하여 지속시간도 길어지기 때문이다.[6]

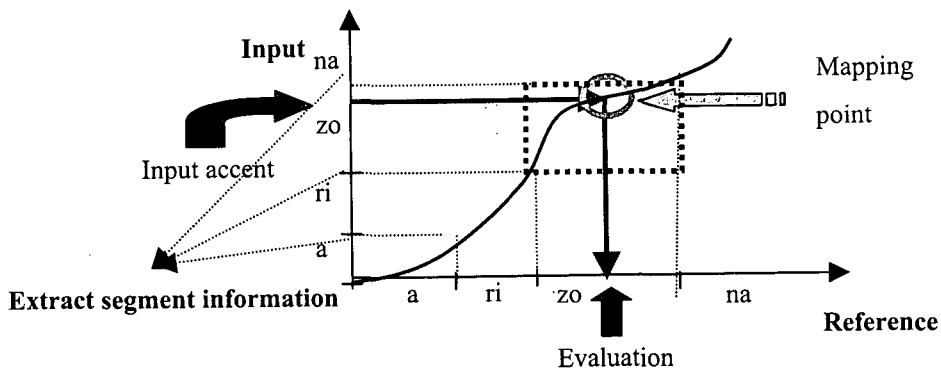


그림 2. 동적 시간정합 알고리즘을 이용한 세그먼트 정보와 액센트 평가

따라서 정확한 평가를 하기 위해서는 각 패턴에 대한 세그먼트정보를 필요로 하게 된다. 본 연구에서는 기준 패턴에 대해서 수작업에 의한 세그먼트를 수행하였고 입력 패턴에 대해서는 동

적시간정합 알고리즘을 응용하여 세그먼트정보를 추출하고 이를 근거로 기준패턴과 비교하여 평가를 내렸다. 최종평가 결과는 표4에 나타내었다.

그림 2는 동적 시간정합 알고리즘을 이용하여 세그먼트 정보를 얻는 방법을 나타내고 있다. 입력 패턴이 들어오면 기준패턴으로부터 세그먼트 정보를 얻게 되고 동적 시간정합 알고리즘을 통하여 정합경로를 찾게 된다. 정합경로로부터 입력패턴의 세그먼트정보를 얻게 됨으로써 정확한 액센트 음절의 위치를 파악할 수 있다. 따라서 같은 음절로 판별된 위치에 액센트의 존재 여부를 판단하여 강세의 옳고 그름을 평가 내리게 된다.

### 2.3 발음 교정 및 평가

발음 교정 및 평가는 2.2절에서와 마찬가지로 동적 시간정합 알고리즘을 이용하게 된다. 동적 시간정합 알고리즘은 보통 고립단어 인식기에 적용되는 알고리즘으로 입력된 신호와 기준 패턴으로 저장된 자료 간의 특징벡터를 추출하고 이를 이용하여 두 패턴 간의 비선형의 최적 정합경로를 찾아 서로 다른 길이의 패턴을 비교하는 방법이다.[4] 본 논문에서는 이러한 동적 시간정합 알고리즘을 응용하여 2.2절에서와 마찬가지로 정합경로를 찾아 입력 패턴에 대한 세그먼트 정보를 얻었다. 이로부터 각 음절, 주요 음소별로 특징 벡터들 간의 정합거리를 구하게 된다. 사용된 특징벡터는 12차 LPC cepstrum 계수를 이용하였고 정합거리를 구하는 방법으로는 Euclidian distance measure를 사용하였다. [ 5 ]

일반적인 고립단어 인식에 쓰이는 알고리즘과는 달리 본 논문에서는 정합거리를 찾을 때 정해진 기준패턴 하나 만을 대상으로 한다. 이렇게 구해진 정합거리로부터 입력신호의 세그먼트정보를 가지고 식(2.3.1)과 같이 각 구간마다 특징벡터 간의 정합거리를 구하게 된다.

$$f(n) = \frac{1}{N_k} \sum_i^{N_k} \sum_j^P (x_i(j) - y_i(j))^2 \tag{2.3.1}$$

$N_k$  : Number of frame at k-th phoneme,  $P$  : LPC cepstrum order

$i$  : frame index,  $j$  : LPC cepstrum coefficient index

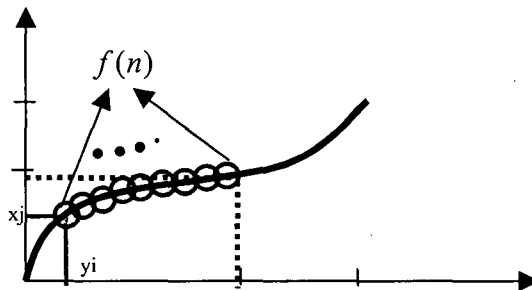


그림 3. 동적 시간정합 알고리즘을 이용한 세그먼트 정보와 발음 평가

각 음절 및 주요 음소구간의 발음평가는 정합거리로부터 얻어진 누적 에러와 가중치 벡터  $\alpha$ 를 이용하여 식(2.3.2)의 최종비용 함수로부터 구하게 된다

$$f'(n) = f(n) \times \alpha, \quad \alpha = \frac{\text{input\_segment\_duration} / \text{ref\_segment\_duration}}{\text{input\_duration} / \text{ref\_duration}} \quad (2.3.2)$$

가중치 벡터  $\alpha$ 는 발생된 음절의 지속시간을 반영하여 발생단어 내에서 기준 패턴과 비교하여 상대적으로 길게 발생된 부분에 대해서 교정해 주기 위함이다. 이는 해당 음절에 액센트가 올 경우 다른 음절에 비해 지속시간이 상대적으로 길어지게 되므로 같은 음소로 이루어진 음절이라 할지라도 액센트가 위치할 경우 발생이 틀려지게 된다. 실험적으로  $\alpha$ 의 범위는 0.8~1.2를 벗어난 범위에서 가중치를 할당하게 하였다.

### 3. 실험 결과

표 4와 표 5는 액센트, 발음 교정 및 평가에 대한 최종 결과이다. 평가방법은 본 논문에서 구현한 시스템의 평가방법이 주관적인 평가 방법과 얼마만큼 일치하는지에 초점을 맞추어 실험 하였다.

표 4는 20대 중반 남성이 각 단어에 대해 의도적으로 각각의 음절에 액센트를 주어 발생했을 때 본 시스템의 평가와 98% 일치함을 식 (3.1)과 같이 구하였다.

표 4에서 화자의 의도와 다르게 평가를 내린 부분은 'Arizona'에서 1회 발생하였는데 4음절에 액센트가 있는 것으로 평가하였다. 이는 'Arizona'의 '/zo/'와 '/na/' 사이의 천이구간에서 비음화되어 입력패턴에 대한 세그먼트 정보가 다소 부정확했기 때문인 것으로 조사되었다.

표 4. 액센트 교정 및 평가 결과

단어 \ 음절	1음절	2음절	3음절	4음절
Obvious	5/5	5/5	5/5	
Avocation	5/5	5/5	5/5	5/5
Reluctantly	5/5	5/5	5/5	5/5
Arizona	5/5	5/5	5/5	4/5
However	5/5	5/5	5/5	

A/B : 정확한 액센트 평가 횟수/발성 횟수

표 5. 발음교정 및 평가 결과

단어 \ 평가	Arizona				Obvious			Reluctantly				Avocation				However		
	A	R	Z	N	O	V	O	R	L	T	T	A	V	C	T	H	E	V
주관적 평가	-	2	4	-	1	-	1	-	4	-	4	-	2	-	1	-	-	4
시스템 평가	-	-	I	-	-	-	I	-	I	I	I	-	I	I	I	-	I	I

'I' : 부정확한 발성으로 판정한 음소 부분

$$\text{Accent\_evaluation}(\%) = \frac{\text{count\_of\_correction}}{\text{total\_count}} \times 100 \quad (3.1)$$

$$\text{Pr onunciation\_evaluation}(\%) = \frac{\text{Cr} - \text{iCr}}{\text{subjective\_decision\_count}} \times 100 \quad (3.2)$$

Cr : System correct evaluation , iCr : System different evaluation

발음평가는 20대 중반 남성 3인과 30대 초반 남성 1인, 40대 초반 남성 1인 등 5인의 남성에게서 해당 단어를 3회씩 반복 청취하게 한 후, 주관적인 입장에서 부정확하다고 느끼는 발성 부분을 지적하도록 하고 이에 대한 통계를 구하여 표 5에 나타내었다. 시스템에 의한 평가는 각 단어를 시스템에 입력하였을 때 평가한 것으로 해당 단어를 반복 평가하는데 있어서 동일한 입력에 의해 이루어져 같은 결과를 출력하게 되므로 1회 실험한 결과이다. 따라서 부정확하다고 평가한 발성 부분에 대해서 'I'로 표기하고 식(3.2)에 의해 주관적인 평가와 얼마만큼 일치하는지를 구하였다. 그 결과 71% 정도가 일치함을 보였다. 표 5의 평가 결과를 살펴보면 'Arizona'의 '/zo/', 'Reluctantly'의 '/luc/', 'Avocation'의 '/vo/', 'However'의 '/ver/' 등 한국인이 구별하기에 힘든 '/z/', '/v/', '/l/'에서 부정확한 발성을 보이는 것으로 평가되었다.

#### 4. 결 론

본 연구에서는 외국어를 익히고자 하는 초보자에게 좀더 빠른 시일 내에 자기 진단을 하면서 외국어를 습득할 수 있도록 발음 및 액센트를 고립단어 범위에서 평가해주고 교정할 수 있는 초별 시스템을 구현하였다.

구현된 시스템을 평가한 결과 액센트의 경우는 화자가 의도적으로 발성한 액센트 위치와 시스템이 평가한 액센트 위치가 98% 일치함을 보였고, 발음평가의 경우 한국인 성인 남성 5인에 의해 주관적인 평가를 수행하여 시스템의 평가와 비교한 결과 71%가 일치함을 보였다.

액센트의 경우는 음의 세기와 높낮이를 나타내는 에너지와 피치에 의해 결정되기 때문에 주관적인 평가와 시스템의 평가가 거의 일치함을 보였다. 그에 반해 발음 부분을 평가하는 데 있어서 주관적인 평가는 개인마다 성향이 다르고 모국어가 아닌 관계로 평가기준이 다소 상이함을 보였고 시스템의 평가에서도 에너지나 피치에 비해 정보량이 많은 특징 벡터(Linear Prediction Coefficient)를 이용하게 되므로 화자가 동일한 발성을 반복하였다고 느낄지라도 평가 결과는 다르게 나타날 수 있었다. 그러나 주관적인 평가와 시스템이 평가한 항목 중 71%가 일치함을 보였고 주관적인 평가에서 다수가 인정하는 부분에 대해서는 시스템의 평가도 일치함을 보였다. 발음 교정 및 평가 결과, 주관적인 평가와 시스템에 의한 평가 모두가 /z/, /v/, /l/ 등에서의 발음이 잘못된 발성이라고 평가한 경우가 많았는데 이는 영어가 모국어가 아닌 한국인이 제대로 발음하고 분별하기 힘든 음소이기 때문으로 판단된다.

향후 계획으로는 영어 원어민의 주관적 평가에 기반하여 좀더 객관적인 평가지표를 마

련하고 이를 통하여 시스템 성능의 신뢰도를 높이려 한다. 또한 본 논문에서 제시한 알고리즘을 일반화 하기 위하여 등록 단어 수를 늘리고 이에 대한 시스템 성능 평가를 할 예정이다며 문장단위 입력음성에 대해서도 발음 및 억양에 대해서 평가와 교정해 줄 수 있는 시스템으로 확장할 예정이다.

### 참 고 문 헌

- [1] Liu Wai Kat; Fung, P. "Fast accent identification and accented speech recognition" Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on vol. 1, pp. 221-224, 1999.
- [2] Teixeira, C. and Trancoso, I. "Accent Identification", Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on Volume: 3, pp. 1784 -1787, vol. 3, 1996.
- [3] L.R. Rabiner and R.W. Schafer, "Digital processing of speech signals", pp. 117-126, 1979.
- [4] H. Sakoe and . Chiba, "Dynamic programming optimization for spoken word recognition", IEEE Trans. Accoustics, Speech, Signal Proc., ASSP-26 (1):43-49, Feb. 1978.
- [5] Itakura, F., "Line spectrum representation of linear predictor coefficients of speech signal.", Trans. Committee on Speech Research, Acoust. Soc. Jap., S75-34, 1975.
- [6] 백승권, 김원철, 한민수, "구문분석을 이용한 한국어 음성합성의 운율생성 연구", 음향학회, 음향학회 논문집 Vol. 18, No. 2, pp. 37-40, 1999년, 11월.

접수일자: 2000. 4. 07.

게재결정: 2000. 5. 28.

#### ▲ 백승권

대전광역시 유성구 화암동 58-4  
한국정보통신대학원대학교 통신공학부 (우: 305-348)  
Tel : (042)866-6196, H/P : 011-9957-0118  
e-mail : skbeack@icu.ac.kr

#### ▲ 최정규

대전광역시 유성구 화암동 58-4  
한국정보통신대학원대학교 통신공학부 (우: 305-348)  
Tel : (042) 866-6183, H/P : 016-466-8927  
e-mail : kyuro@icu.ac.kr



▲ 한민수

대전광역시 유성구 화암동 58-4

한국정보통신대학원대학교 통신공학부(우: 305-348)

Tel : (042)866-6123, H/P : 011-897-2559

e-mail : mshahn@icu.ac.kr