2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

영어 학습 시의 발성 교정 기술에 관한 연구

김재민,백승권, 한민수,
한국정보통신대학원대학교

Study on the pronunciation correction in English Learning

Jae-Min Kim,Seung-Kwon Beack,Minsoo Hahn,
Information and Communications University
Skbeack@icu.ac.kr

<본 연구는 ㈜한국엑시스의 연구비 지원에 의해서 이루어졌습니다.>

ABSTRACT

In this paper. we implement an elementary system to correct accent, pronunciation, and intonation in English 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 calculate the spectral distance measure for each phoneme between input and reference. For the intonation evaluation. we propose nine pattern of slope to estimate pitch contour. then we grade test sentences by accumulated error obtained by the distance measure and estimated slope. Our result shows that 98 percent of accent and 71 percent of pronunciation evaluation agree with perceptual measure. As the result of the intonation evaluation. system represent the similar order of grade for the four sentences having different intonation patterns compared with perceptual evaluation.

1.서론

영어를 습득하는데 있어서 발성에 대한 교정을 받기 위해서는 전문가의 지도를 받거나 어학용 기기를 통한 반복청취 및 발성을 통해 습득하는 것이 상례이다. 본 연구에서는 영어를 익히고자 하는 초보자에게 본인의 발성 및 청취도에 대한 자기 진단을 하면서 영어를 습득할 수 있도록 하는 시스템을 제공함이 목적이다. 시스템의 평가항목은 영어단어 발성시 액센트와 발음을, 영어문장에서는 억양과 발음을 평가해 줄 수 있도록 구성하였다. 액센트를 평가하기위해서는 에너지와 피치정보를 제안된 알고리즘에 적용하였고, 발음에 대한 평가는 음성인식에서와 유사한 방법으로 spectral 거리 측정을 이용하여 구현하였다.[4] 억양에 대해서는 발성 패턴과 기준패턴의 피치궤적에서 각각의 포락선을 제안된 9개의 linear slope 패턴을 이용하여 모델링하고 여기서 발생하게 될 81 가지 mapping 패턴을 테이블로 만들어 유사도에 따른 가중치를 구하였다. 각 절에서 평가 항목별로 구현된 알고리즘을 설명하였다. 특히 종전의 연구에[6] 추가된 사항과 문장부분에서 억양부분에 대한 평가를 주되게 다루었으며 시스템의 평과결과와 지각적 평가결과의 유사도를 조사하여 결론을 맺었다.

2. 액센트 평가

일반적으로 한국어 단어에서는 무강세,즉 액센트가 없는 실제로 한국어 음성신호에서는 큰 강세 변화를 찾아 보기 힘들다. 한편 영어는 많은 정보가 액센트에 의해 전달 되므로 발음이 정확하다 해도 액센트의 위치가 틀리면 의사소통이 안되는 경우가 많다. 액센트라하면 화자의 억양이나 지방사투리로 간주하여 화자 인식을 하기위한 지표가 되기도 하는데 본 논문에선 음의 세기로 간주였다.[2]

실험을 위해 수집한 DB(DataBase)는 미국인이 발성한 5 개단어를 선정하였으며 각각의 DB 는 액센트의 위치와 음절내의 음소특성, 액센트 받는 음절 내의 유무성음 포함여부에 따라 3 음절 이상의 단어들로 구성하여 기준 패턴으로 삼았다. 액센트위치를 평가하기 위해서는 정확한 액센트위치를 찾아야 하고 찾은 액센트위치가 어느 음절에 포함되어 있는가를 판별할 수 있어야 액센트의 옳고 그름을 판단해 줄 수 있다. 따라서 정확한 액센트위치를 찾을 수 있는 방법과 발성된 입력패턴과 기준패턴사에 대한 세그먼트 정보가 필요하다. 전자의 방법은 그림 1 과 식 2.1 을 적용하였다. 이는 액센트위치를 추정하는데 에너지와 피치정보를 이용한 것으로 에너지 궤적의 최대 변곡점 2 개를 후보자로 삼고 피치궤적의 최대값과 거리를 측정하여 최소거리를 갖는 후보자를 액센트 위치로 판별하였다. 세그먼트정보는 DTW (Dynamic Time Wapping)를 응용하여 입력패턴의 세그먼트정보를 얻었다. 이는 두패턴간의 정합경로를 찾은 후 기준패턴의 세그먼트정보로부터 대응되는 좌표로부터 입력패턴의 세그먼트 정보를 얻을 수 있다. 표 1 은 시스템이 내린 액센트 평가에 대한 결과 표이다. 여기서 에너지 정보만을 이용할때와 피치정보만을 이용할 때 부정확한 액센트 위치를 찾게 된다. 피치의 최대치가 강세를 받는 음절을 결정하는데 중요한 요소임은 사실이다.[1] 그러나 피치의 최대치가 천이구간에 나타날 경우 액센트 음절을 판별하는데 오류가 발생될 수 있다. 따라서 제안된 알고리즘에 의한 방법을 적용할 때 가장 정확한 액센트 위치를 찾아내고 평가내릴 수 있다.

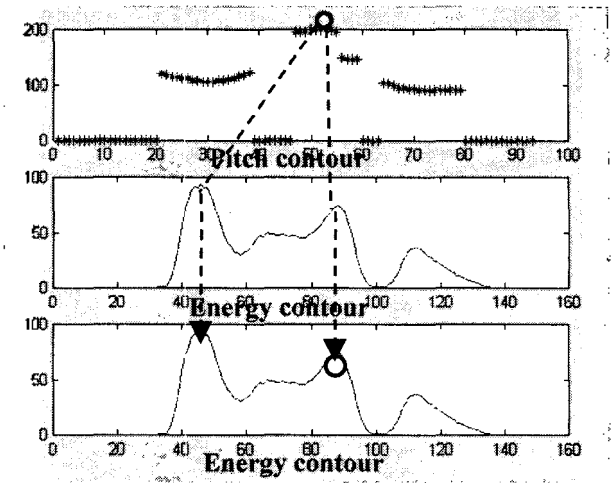

그림 1. 'however'에서 에너지 정보와 피치 정보를 이용한 액센트 위치 판별 예

표 1 액센트 평가

	Energy		Pitch		Energy, Pitch,			
	A	B	A	B	A	B	A	B
1	42/67	3	12/67	1	48/67	3	42/67	3
2	43/76	1	43/76	3	47/76	3	43/76	1
3	3/51	1	3/51	1	10/51	2	3/51	1
4	35/60	2	10/60	1	39/60	3	35/60	2
5	8/73	2	3/73	1	10/73	2	8/73	2

A: 액센트위치/전체지속시간, B:액센트 음절

1:arizona. 2:avocation , 3:obvious, 4:However 5:reluctantly

$$Sp = \min(dist(MaxE\_i, MaxP)), \quad \text{식(2.1)}$$

$dist : |A - B|, Sp$: Selectedframe.

$MaxE\_i$: Maximum energy frame 1,2

$MaxP$: Maximum pitch point

3. 발음 평가

발음 평가를 하는 방법은 음성인식 알고리즘과 유사한 과정을 거치게 되는데 차이점이 있다면 발성되어질 단어와 문장이 정해져 있다는 것이다. 따라서 정해진 후보자를 기준패턴으로 삼고 발성된 입력패턴이 기준패턴과 얼마나 일치하는가를 판별하게 되는 것이다. 그림 2 를 보면 2 절에서와 마찬가지로 기준패턴의 세그먼트 정보로부터 DTW 를 이용하여 입력패턴의 세

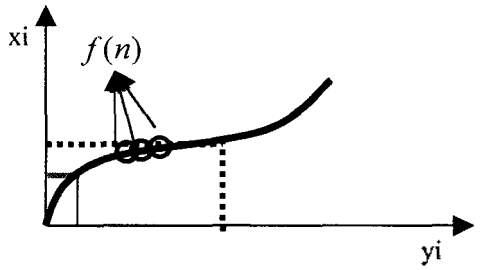

그림 2 동적 시간정합 알고리즘을 이용한 세그먼트 정보와 발음 평가

$$f(n)=\frac{1}{3}\sum_{i=0}^{3}\sum_{p=1}^{11}\left(x_i(p)-y_i(p)\right)^2 \quad (식\ 3.1)$$

표 2 기준 패턴 헤더 파일 정보

Word Header	Phoneme segment information	
	Phoneme lable	
	Accent position information	
Sentence Header	Word segment Information	Phoneme segment Information
	Word lable	Phoneme lable

그먼트 정보를 얻게 되고 각 구간별로 spectral 거리를 측정하게 된다. spectral 거리 측정은 음소구간중 안정구간의 3 개의 frame 에 대해서 수행하게 된다. 이는 천이구간에 대한 평가를 제외 시키기 위함이다.

또한 지속시간에 따른 발성오류를 지적해 주기 위해 가중치 α 를 사용하였다.[6]

시스템의 발음 평가는 단어에서는 음절의 음소별로, 문장에서는 문장내의 단어의 음소별로 지적을 해 줄 수 있도록 구현하였다. 따라서 기준패턴에 대한 헤더 파일 정보를 표 2 와 같이 구성하였다. Spectral 거리 측정방법은 식 3.1 과 같이 12 차 LPC(linear predictive coefficient) cepstrum 을 사용하였고 0 차 계수는 제외하였다.[4] 기준 패턴이 되는 문장 DB 는 미국인이 발성한 5 개 문장을 사용하였다.

4.억양평가

억양은 발성문장의 음의 높낮이로 문장의 종류에 따라 패턴을 달리한다. 이는 피치궤적으로부터 얻어지게 되는데 같은 문장이라 할지라도 억양에 따라 의미를 달리하게 된다. 억양는 한국어에서도 중요한 운율정보이지만 영어에서는 더욱 그러하다. 한국어와 비교해 볼 때 영어의 억양 패턴은 언어 특성상 많은 운율의 변화를 보이며 보다 변화가 심한 형태의 피치궤적을 그리게 된다. 따라서 억양패턴을 비교하기 위해서는 피치궤적의 높낮이 뿐만아니라 궤적의 변화 모습도 모델링되어 져야 한다. 영어문장에서 억양 특성을 음절구간으로 한정하여 관찰할 때 크게 네가지 형태로 표현되어 지는데 이는 'a rise', 'a fall', 'a rise-fall', 'a fall-rise'로 구분되어 진다. [1]

본 논문에서는 문장의 억양 패턴을 비교하기 위해 음절별로 나누어 피치궤적의 모습을 두개의 linear slope 을 이용하여 모델링하였다. 모델링 하는 방법은 우선 음절내의 피치구간을 찾아내고 구간의 시작점과 끝점 사이의 slope 을 연결시켜줄 중간지점을 찾아 내게 된다. 이때 중간지점은 피치구간내의 최대치, 또는 최소치를 갖는 지점을 선택하게 된다 중간지점의 선택기준은 피치구간내의 평균치보다 큰 편차를 보이는 최대치 또는 최소치를 선택하게 된다. 그 결과 피치구간을 모델링 하는데 대표되는 9 가지 패턴을 생성하게 되었고 입력패턴과 기준패턴사이의 81 가지의 mapping Table 을

표 3. 9 개의 피치궤적을 모델링한 linear slope pattern

Slope pattern					
P1		P4		P7	
P2		P5		P8	
P3		P6		P9	

표 4. Slope 비교시 가중치 table

Tst / Ref	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0.2	0.8	1.5	2	1.5	1.5	1.5	1.5	0.8
P2	0.8	0.2	0.8	2	1.5	2	0.8	0.8	1.5
P3	1.5	0.8	0.2	1.5	1.5	2	0.8	0.8	1.5
P4	2	1.5	1.5	0.2	0.8	1.5	1.5	0.8	1.5
P5	1.5	2	1.5	0.8	0.2	0.8	0.8	0.8	1.5
P6	1.5	1.5	2	1.5	0.8	0.2	1.5	1.5	0.8
P7	1.5	0.8	1.5	1.5	0.8	1.5	0.2	0.8	0.8
P8	2	1.5	0.8	0.8	0.8	1.5	0.8	0.2	2
P9	0.8	1.5	1.5	2	1.5	0.8	0.8	2	0.2

tst: test ref: reference

만들어 내었다. Mapping Table 의 수치는 가중치 factor 로 사용하였고 피치궤적의 유사도에 따라 그 값을 결정하였다. 그림 4 는 9 개의 linear slope 을 나타내며 표 3 은 mapping 테이블을 나타내고 있다. 실험결과는 구간별 궤적간의 거리측정에 대한 누적에러를 이용한 방법, 모델링된 linear slope 을 이용한 가중치 factor 의 누적치를 구한방법, 거리측정 누적에러에 가중치 factor 를 적용한 방법 3 가지로 하였다.

지각적인 평가방법은 성인 남성 10 인에게 기준패턴 문장을 들려 주고 각기 다른 억양패턴을 갖는 4 개의 테스트 문장에 대하여 6 단계의 등급을 메기도록 청취 테스트를 하였다.[3] 그 결과를 표 5 에 나타내었다.

그 결과 제안된 알고리즘에 의한 서열이 지각적 서열과 가장 유사함을 보였고 명확한 구분을 지을
수 있도록 서로간에 큰 편차를 보였다. 전체결과는 각 측정방법의 누적에러값에 따라 등급을 매겨 지각적평가방법과 얼마만큼의 유사도를 갖는지를 조사하여 표 6 에 나타내었다.

5.결론

본 연구에서는 영어를 익히고자 하는 초보자에게 자기 진단할 수 있도록 액센트, 발음, 억양에 대해 평가 및 교정을 해 줄 수 있는 초벌시스템을 구현하였다. 각각의 시스템 평가결과는 지각적인 평가방법과 비교했을 때 액센트 및 발음부분에서 98%, 71%일치하였다.[6] 억양 평가부분에서는 4 개의 각기 다른 억양패턴을 갖는 입력패턴과 기준패턴의 유사도에 따라 6 등급으로 나누어 지각적인 평가를 수행하였고 이에 대한 결과와 본 논문에서 제안한 알고리즘에 의해 수행된 평가결과를 비교 했을 때, 지각적 평가의 선호도에 대한 서열과 제안된 알고리즘의 누적 에러에 의한 서열이 유사함을 보였다.

표 5 지각적 억양 평가결과

	A	B	C	D	E	F	Total Score
Tst1		1		3	4	2	-15
Tst2		1	2	3	4		-7
Tst3	1	1	3	1	3	1	-3
Tst4	5	4		1			20

A: best F: worst

Grade score : A= 3, B=2, C=1, D=-1, E=-2, F=-3

표 6

	Perceptual Measure		Amplitude Distance(AD)	
	Grade	Score	Grade	Error
Tst1	4	-15	2	68.4
Tst2	3	-7	4	126.6
Tst3	2	-3	2	68.2
Tst4	1	20	1	60
	Slope Distance(SD)		AD + LD	
	Grade	Error	Grade	Error
Tst1	4	7	4	129.4
Tst2	3	5.3	2	95.41
Tst3	2	4	3	111.2
Tst4	1	1.2	1	64.6

참고문헌

[1] Colin W and Mori Ostendorf, "Automatic recognition of intonational features", in proc. of IEEE Int.Conf.ASSP,pp I-221.

[2] Teixeira,C.and Trancoso,I."Accent Identification",Spoken Language,1996.ICSLP 96.

[3] S.M.F.smyth, J.V.McCanny and P.Challenr."An independent evaluation of the performance of the CCITT G.722 wideband coding recommendation using music sound",proc of IEEE Int Conf,Sep 1988.

[4] 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,최정규,한민수, " 영어단어발성시의 오류교정 기술에 관한 연구",음성과학회,음성과학회 논문집 No.8,pp83-90,2000 년 4 월.