말소리와 음성과학 제2권 제2호 (2010)
pp. 69~75

정제 알고리즘을 이용한 한국인 화자의 영어 발화 자동 진단 시스템

Automatic Pronunciation Diagnosis System of Korean Students' English Using Purification Algorithm

양 일 호[1] · 김 민 석[2] · 유 하 진[3] · 한 혜 승[4] · 이 주 경[5]
Yang, IL-Ho · Kim, Min-Seok · Yu, Ha-Jin · Han, Hyeseung · Lee, Joo-Kyeong

ABSTRACT

We propose an automatic pronunciation diagnosis system to evaluate the pronunciation of a foreign language without the uttered text. We recorded English utterances spoken by native and Korean speakers, and utterances spoken by Koreans are evaluated by native speakers based on three criteria: fluency, accuracy of phones and intonation. The system evaluates the utterances of test Korean speakers based on the differences of log-likelihood given two models: one is trained by English speech uttered by native speakers, and the other is trained by English speech uttered by Korean speakers. We also applied purification algorithm to increase class differentiability. The purification can detect and eliminate the non-speech frames such as short pauses, occlusive silences that do not help to discriminate between utterances. As the results, our proposed system has higher correlation with the human scores than the baseline system.

Keywords: English education, automatic pronunciation diagnosis, purification, GMM

1. 서론

영어 교육의 필요성이 증대됨에 따라 컴퓨터를 이용한 다양한 영어 학습 시스템들이 개발되고 있다. 컴퓨터를 이용한 영어 학습 시스템들은 멀티미디어를 활용하여 영어 학습자와 컴퓨터 간 상호작용이 가능한 특성을 지닌다. 몇몇 영어 학습 시스템들의 경우, 학습자의 발화를 녹음하여 들려주는 기능을 지원함으로써 책과 오디오에 의존하는 기존의 영어 학습법에서 실현하기 어려웠던 영어 말하기 교육을 제공한다. 특히 자동으로 영어 발화 수준을 진단해주는 시스템은, 원어민 교사 없이도 개인 학습자에게 적절한 피드백을 제공함으로써 영어 말하기 교육을 보다 널리 보급하는데 이바지할 수 있다.

제 2언어 학습자의 발화를 자동 진단하는 연구는 세계 각지에서 다양한 형태로 수행되고 있다. (Neumeyer, 1996)는 미국인의 불어 발화를 자동 진단하는 시스템에 대해 연구하였다. (Franco, 1997)는 다양한 발화 진단 점수를 혼합하여 성능을 높이는 방법을 제시하였다. (Neumeyer, 1998)는 제 2언어 학습자를 위한 진단 시스템을 구축하고 스페인어 · 불어 · 영어 발화에 대하여 각각 성능을 평가하였다. (Cucchiarini, 2000)는 원어민 / 비원어민의 네덜란드어 발화를 4종류의 척도를 기준으로 진단하였다. (Moustroufas, 2007)는 그리스인의 영어 발화를 자동 진단하였으며, 발성 문장 입력 없이 발화를 진단하는 방법을 제안하였고, 제 2 언어 학습자의 모국어 모델을 구성하여 시스템 성능을 개선하였다.

국내에서도 한국인 화자의 영어 발화를 자동 진단하는 방법에 대한 연구가 수행되었다. (Kim, 2002)은 한국 아동 대상 영어발음교정 시스템을 개발하였다. (Kim, 2003)은 미국인과 한국인의 영어 발화를 수집하여 모델을 학습한 뒤 인식 네트워크를 구성하여 교정할 발음을 검출하는 시스템을 구축하고 성능을 평가하였다. (Park, 2003)은 영어 발화의 유창성을 자동 진단하는 음성인식 엔진을 개발하였다.

상기 연구들에서는 대부분 진단 시스템이 발성 내용

1)서울시립대학교 heisco@hanmail.net
2)서울시립대학교 ms@uos.ac.kr
3)서울시립대학교 hjyu@uos.ac.kr, 교신저자
4)서울시립대학교 jkyoonhan@gmail.com
5)서울시립대학교 jookeng@uos.ac.kr
이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(KRF-2007-321-A0015)

접수일자: 2010년 5월 1일
수정일자: 2010년 6월 10일
게재결정: 2010년 6월 22일

(transcription)을 알고 있다고 가정하였다(Cucchiarini, 2000; Franco, 1997; Kim, 2002; Kim, 2003; Neumeyer, 1998; Neumeyer, 1996; Park, 2003). 이러한 방법을 이용할 경우 사용자가 사전에 지정된 몇몇 발화만 진단 받을 수 있는 제약을 받거나, 자신이 발성한 내용이 무엇인지 입력해야 하는 불편함이 발생할 수 있다. (Moustroufas, 2007)는 이러한 문제를 극복하기 위하여 시스템이 사용자의 발성 내용을 모르는 상태에서 제 2언어 학습자의 발화를 진단할 수 있는 시스템을 구축하였다.

본 연구에서는 (Moustroufas, 2007)가 제안한 방법에 기반하여 한국인 화자의 영어 발화를 유창성 · 발음의 정확성 · 억양의 정확성 면에서 각각 진단하고 통합 점수를 도출하는 시스템을 구축하였다. 하지만 기존의 방법에서는 원어민의 발화 모델과 제 2언어 학습자의 모국어 발화 모델을 구성하였는데 반해, 본 연구에서는 원어민의 발화 모델과 제 2언어 학습자의 제 2언어 발화 모델을 구성하여 실험하였다. 또한 기존의 연구에서 화자 단위(speaker level) 발화 진단 성능에 비해 상대적으로 저조하였던 문장 단위(sentence level) 발화 진단 성능을 개선하기 위하여 정제(purification) 알고리즘(Anguera, 2006)을 적용하였다. 발화 수준이 높은 비원어민의 제 2언어 발화는 원어민의 발화와 유사할 것이므로, 이를 이용하여 생성한 두 모델은 완전한 독립성을 보장하기가 어렵다. 이러한 점은 성능에 악영향을 미칠 수 있으므로, 정제 알고리즘을 통하여 모델 학습 데이터에서 원어민 발화와 비원어민 발화간의 유사한 특징 프레임을 제거함으로써 성능을 개선하고자 하였다.

본 논문의 구성은 다음과 같다. 다음 장에서는 본 연구를 수행하기 위해 수집한 데이터베이스를 소개한다. 3장에서는 기존의 방법과 제안한 방법을 설명한다. 4장에서 시스템 성능을 평가하고 결과를 분석한 뒤, 마지막 장에서 결론을 맺는다.

2. 데이터베이스 구축

2.1 녹음

본 연구를 위해 원어민과 비원어민(한국인)의 영어 발화를 수집하였다. 자연스러운 발화를 유도하기 위하여 2인이 1조가 되어 서로 마주보고 대화하는 형식으로 녹음하였다. 5~10개의 문장으로 구성된 135개의 대화문을 4 set으로 구분하여 각 조의 화자들이 1 set씩 발성하도록 하였다. 원어민 24명(12조)과 비원어민 20명(10조)의 발화를 수집하였다. 녹음은 조용한 사무실 환경에서 이루어졌다. set 별 구성 화자 수는 <표 1>과 같다.

2.2 원어민 청취 평가(human scoring)

3명의 원어민 평가자가 비원어민(한국인)에게서 수집한 각 발화를 문장별로 청취하고 진단하였다. 이 때, 각 발화를 유창성 · 발음의 정확성 · 억양의 정확성을 기준으로 각각 1~5점으로 진단하도록 하였다. 이렇게 청취 진단한 점수 목록에 대하여

표 1. set별 구성 화자 수.
Table 1. The number of speakers for each set.

set	원어민			비원어민 (한국인)			총계
	남성	여성	계	남성	여성	계	
1	3	3	6	4	2	6	12
2	4	2	6	6	0	6	12
3	2	4	6	0	2	2	8
4	4	2	6	2	4	6	12
계	13	11	24	12	8	20	44

평가자간 상관관계(correlation)의 계수와 개방 상관관계(open correlation)의 계수를 계산하였다. 평가자간 상관계수를 구하는 식은 다음과 같다.

$$\frac{\sum_i (A_i - A)(B_i - B)}{(A_i - \bar{A})^2 \sum_i (B_i - \bar{B})^2} \quad (1)$$

여기서 A_i, B_i는 각각 평가자 A, B가 평가한 각 발성의 점수이며, $\bar{A}$, $\bar{B}$는 각각 평가자 A, B가 평가한 전체 점수의 평균이다.

수집한 데이터베이스를 이용하여 구축한 시스템의 최대 성능을 추정하기 위하여 개방 상관관계를 계산하였다. 본 연구의 목적은 원어민 교사를 대신하여 학습자의 영어 발화를 자동 진단하는 시스템을 만드는 것이다. 따라서 이상적인 자동 진단 시스템은 사람(원어민 평가자)과 유사한 결과를 도출할 것이다. 개방 상관관계는 한 평가자가 평가한 점수 목록을 자동 진단 시스템의 결과라 가정하고, 다른 평가자들의 평균 점수 목록간의 상관계수를 계산하여 구한다. 상관계수 계산 결과는 <표 2>와 같다. 여기서 모든 개방 상관계수의 평균이 0.84였으므로, 이를 본 연구에서 구축한 시스템의 목표 성능으로 계획하였다.

표 2. 평가자간 상관계수 및 개방 상관관계
Table 2. Correlations and open correlations between raters.

평가 척도	평가자	1	2	3
유창성	1	1.00	0.81	0.76
	2		1.00	0.83
	3			1.00
개방 상관계수		0.82	0.88	0.84
발음의 정확성	1	1.00	0.77	0.80
	2		1.00	0.82
	3			1.00
개방 상관계수		0.82	0.84	0.87
억양의 정확성	1	1.00	0.79	0.74
	2		1.00	0.77
	3			1.00
개방 상관계수		0.81	0.84	0.80

원어민 청취 평가 결과의 점수 분포는 <표 3>과 같다. 수집한 비원어민의 발화는 3점(보통 수준) 및 2점(낮은 수준)의 비율이 많았고 1점(매우 낮은 수준) 및 5점(매우 높은 수준)의 비율이 적었다.

표 3. 모든 평가자 / 모든 척도의 점수 분포 (%)
Table 3. Distribution of scores across all raters and all types for all scores (%)

점수	1	2	3	4	5
유창성	1.18	28.48	57.93	12.14	0.27
발음	8.64	34.40	47.06	9.73	0.17
억양	2.11	36.54	51.98	9.17	0.21
전체 비율	3.98	33.14	52.32	10.35	0.22

각 평가자가 각 평가 척도에 대하여 진단한 결과의 평균 및 표준편차는 <표 4>와 같다.

표 4. 평가자별 점수의 평균과 표준편차
Table 4. Means and standard deviations of scores from each rater

평가자		1	2	3	평균
유창성	평균	2.92	2.77	2.77	2.82
	표준편차	0.66	0.65	0.65	0.65
발음	평균	2.75	2.41	2.58	2.58
	표준편차	0.76	0.78	0.78	0.78
억양	평균	2.78	2.60	2.69	2.69
	표준편차	0.66	0.66	0.68	0.67

3. 자동 발성 진단 시스템

3.1 가우시안 혼합 모델을 이용한 발화 진단 방법

본 절에서는 음성인식 및 화자인식에서 음소 또는 화자를 모델링하는데 주로 이용하는 가우시안 혼합 모델(Gaussian mixture model)을 통해 자동으로 발화 점수를 진단하는 방법(Moustroufas, 2007)에 대해서 기술한다.

우선 원어민의 발화를 이용하여 원어민 발화 모델 ^{TARGET} 을 생성한다. 비원어민 학습자가 T시간 동안 발성한 영어 발화를 $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T\}$ 라 할 때, 다음과 같이 로그 유사도(log-likelihood)를 계산한다.

$$\log f(X \mid \phi_{TARGET}) = \sum_{k=1}^{T} \log f(\mathbf{x}_k \mid \phi_{TARGET}) \quad (2)$$

이와 마찬가지로 비원어민의 모국어 발화를 이용하여 모국어 발화 모델 ^{SOURCE} 을 생성하고 로그 유사도($^{\kappa} \mid \phi_{SOURCE}$)를 계산한다. 최종적인 자동 진단 점수(machine score)는 다음과 같이 구할 수 있다.

$$\frac{(X \mid \phi_{TARGET})}{T} - \frac{L(X \mid \phi_{SOURCE})}{T} \quad (3)$$

단, 본 연구에서는 ^{SOURCE} 를 생성할 때 비원어민 화자의 영어 발화(set 2, 4)를 이용하여 모델을 학습하였다.

3.2 정제 알고리즘(purification algorithm)

(Anguera, 2006)는 화자 군집화(speaker clustering)를 수행할 때, 각 군집 내에서 해당 화자의 발성이 아닌 특징 프레임(short pause, occlusive silences 등)을 걸러내기 위하여 정제 알고리즘을 사용하였다.

본 연구에서는 프레임 단위 정제 알고리즘을 통하여 앞에서 소개한 특징뿐만 아니라 원어민 발화와 비원어민 발화 간의 유사한 특징 역시 함께 제거함으로써 ^{TARGET} 과 ^{SOURCE} 를 보다 효과적으로 모델링 할 수 있을 것으로 기대하였다. 이는 다음과 같은 과정을 통해 수행한다.

1. ^{TARGET} 을 생성할 발화의 모든 특징 벡터를 기준으로 앞 뒤 m개의 인접한 특징 벡터를 포함하는 길이 Q (Q = 2m + 1) 크기의 모든 프레임에 대하여 로그 유사도($^{\kappa} \mid \phi_{SOURCE}$) 를 계산한다.
2. 정제 후 유지할 프레임의 비율(P%)만큼을 제외하고, 유사도가 높은 순으로 나머지 프레임을 제거한다(정제 과정). 이 때, 앞뒤 Q개의 특징 벡터를 모두 제거하는 것이 아니고 기준이 되는 중앙의 특징 벡터 하나만 제거한다(<그림 1> 참조).
3. 정제된 발화를 이용하여 ^{TARGET} 을 다시 생성한다.
4. ^{SOURCE} 를 생성할 발화에 대해서도 로그 유사도($^{\kappa} \mid \phi_{TARGET}$) 를 계산하여 정제 과정을 거쳐 ^{SOURCE} 를 다시 생성한다(이 때, ^{TARGET} 은 1~3 단계에서 정제하기 이전의 모델임).

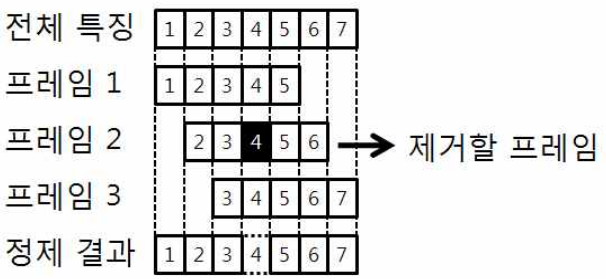

그림 1. 정제 과정
Figure 1. The purification process

4. 시스템 성능 평가

4.1 실험 설계

수집한 데이터베이스에서 원어민의 영어 발화를 ^{TARGET} 의 학습 데이터로 이용하였다. 비원어민의 영어 발화 중 일부(set 2, 4)는 ^{SOURCE} 의 학습 데이터로 이용하였고, 나머지(set 1, 3)는 테스트 데이터로 이용하였다. 충분히 길게 발성한 발화에 대해서도 시스템 성능을 평가하기 위해, 테스트 데이터에서 6초 이상의 발화만 선별하여 추가 실험을 수행하였다. 이 때 테스트 데이터의 구성은 <표 5>와 같다.

3.1절에서 소개한 진단 방법에 따라 자동 진단 점수 LDIFF(machine score)를 계산하고, 원어민 청취 점수(human

표 5. 테스트 데이터 구성
Table 5. Test data set

	전체 발화	긴 발화(6초 이상)
발화 수	1616개	255개
평균 발화 길이	4.17초	7.5초

score)와의 상관계수를 측정하였다. 또한 각 평가자별 / 척도별 점수를 정규화한 뒤 평균한 통합 점수와의 상관계수를 측정하였다. 어떤 점수 s 의 평균이 $\bar{s}$ 이고 표준편차가 σ_s 일 때 정규화된 점수 z 는 다음과 같이 구하였다.

$$= \frac{s - \bar{s}}{\sigma_s} \quad (4)$$

4.2 특징 추출

본 연구에서는 구축한 진단 시스템의 성능을 최대화할 수 있도록 실험적으로 최적의 특징 추출 방법을 선별하였다. <표 6>의 결과는 혼합 수 8, 16, 32, 64, 128, 256개로 학습한 모델을 각각 이용하여 발화 단위(sentence level) 진단 실험을 수행한 뒤 가장 높은 상관계수를 취한 것이다. 단, 특징 추출시 window size는 25ms로 고정하고 10ms씩 shift하였다. 표에서 '12_MFCC'는 12차 MFCCs(mel-frequency cepstral coefficients), '12_LPC'는 12차 LPCs(linear prediction coefficients), 'E'는 에너지, 'ZE'는 발화 단위로 정규화한 에너지, 'P'는 피치, 'ZP'는 발화 단위로 정규화한 피치, 'D'는 delta, 'A'는 delta-delta를 의미한다. '(13d)', '(39d)', '(42d)', '(78d)'는 전체 특징이 각각 13차원, 39차원, 42차원, 78차원임을 뜻한다.

표 6. 특징별 원어민 청취 점수와 자동 진단 점수의 상관관계
Table 6. Correlations between human and machine scores with various feature vectors.

특징별 상관계수	유창성	발음	억양	통합 점수
12_MFCC+E (13d)	0.235	0.190	0.246	0.316
12_MFCC+ZE (13d)	0.305	**0.237**	0.299	0.393
12_MFCC+ZE+D+A (39d)	0.350	0.212	0.298	0.399
12_MFCC+ZE+ZP +D+A (42d)	**0.370**	0.219	**0.304**	0.407
12_LPC+ZE+ZP +D+A (42d)	0.162	0.145	0.185	0.223
6_MFCC+6_LPC+ZE+ZP +D+A (42d)	0.263	0.189	0.225	0.305
12_MFCC+12_LPC+ZE+ZP +D+A (78d)	0.241	0.204	0.260	0.331

'12_MFCC+E (13d)'와 '12_MFCC+ZE (13d)'를 비교한 결과 에너지를 정규화하는 것이 성능 향상에 도움이 되는 것을 확인하였다. delta 및 delta-delta를 추가('12_MFCC+ZE+D+A (39d)') 하자 발음의 정확성에 대한 진단 성능은 소폭 감소했지만 유창성에 대한 진단 성능은 향상되었다. 여기에 피치를 더하여 실험한 결과('12_MFCC+ZE+ZP+D+A (42d)') 상관계수가 더 높아졌다. 반면에 LPC를 이용한 경우는 MFCC만 사용한 경우에 비하여 별다른 성능 향상을 확인할 수 없었다.

표 7. 혼합 수에 따른 상관관계(12_MFCC+ZE+ZP+D+A (42d))
Table 7. Correlations according to number of mixtures.

혼합 수	유창성	발음	억양	통합 점수
8	0.310	0.194	0.257	0.355
16	0.343	0.189	0.256	0.353
32	0.362	0.187	0.272	0.383
64	**0.370**	**0.219**	**0.304**	0.407
128	0.333	0.204	0.295	0.392
256	0.312	0.206	0.284	0.377

12차 MFCCs에 정규화한 에너지와 피치를 더하고, delta 및 delta-delta를 추가한 42차원 특징을 사용하였을 때 발음의 정확성을 제외하고는 상관계수가 가장 높았으므로 이후의 실험에서는 모두 이 특징을 사용하였다. 이 때, 혼합 수에 따른 상관계수는 <표 7>과 같다.

4.3 모델 학습 데이터 선정의 적절성 평가

(Moustroufas, 2007)는 TARGET모델($_{TARGET}$)을 학습하기 위하여 원어민 발화 음성 코퍼스를 사용하고, SOURCE 모델($_{SOURCE}$)을 학습하기 위하여 제 2언어 학습자의 모국어 발화 음성 코퍼스를 사용하였다. 그러나 본 연구에서는 SOURCE 모델을 학습할 때 제 2언어 학습자의 제 2언어 발화 음성을 사용하였다. 이렇게 모델 학습 데이터의 선정에 차이점을 둔 것이 적절했는지 판단하기 위하여 추가 실험을 진행하였다. <표 8>에서 'baseline'은 4.2절에서 가장 좋은 성능을 보인 결과이고, '비교실험'은 (Moustroufas, 2007)와 유사하게 모델 학습 데이터를 선정하여 'baseline'과 동일한 방법으로 실험한 것이다. '비교실험'에서는 TARGET 모델 학습에 TIMIT(The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus) 코퍼스의 TRAIN데이터 전체(4620개 발화)를 사용하였고, SOURCE 모델 학습에 서울말 낭독체 발화 말뭉치(국립국어 연구원 개발)의 20대 남녀 화자의 발화를 화자별로 20개씩 임의 선택(800개 발화)하여 사용하였다.

표 8. 모델 생성 방법에 따른 비교실험
Table 8. Correlations between human and machine scores according to methods of building models.

학습 데이터별 상관계수	유창성	발음	억양	통합 점수
baseline	**0.370**	**0.219**	**0.304**	0.407
비교실험	0.245	0.156	0.236	0.300

'비교실험'의 경우에는 학습 데이터의 수가 다르고 녹음 채널이 일치하지 않으므로 두 방법의 효용성을 객관적으로 단순 비교할 수는 없다. 하지만 실험 결과 본 연구에서 사용한 방법이 더 높은 진단 성능을 보였으며 절대적인 상관계수가 기존 연구와 유사한 수준이었으므로, 본 연구의 접근 방법 또한 유효하다고 판단하였다.

4.4 정제 알고리즘을 적용한 진단 시스템 성능 평가

정제 알고리즘을 적용하여 시스템의 성능을 개선하기 위해서는 파라미터를 최적화해야 할 필요가 있다. 정제 알고리즘의 파라미터는 기준 특징 벡터로부터 앞뒤 m개의 특징 벡터로 구성되는 프레임의 크기 Q(Q = 2m + 1)와 정제 후 유지할 프레임의 비율 P이다. 파라미터별로 문장 단위 진단 실험을 수행한 결과는 <표 9>와 같다. <그림 2>는 통합점수만 비교한 그래프이다.

통합점수의 상관계수만 계산한 경우, 정제 후 유지 비율을 75%~85%로 하였을 때 시스템 성능이 향상되었다. 반면에 90%를 유지('p:0.9 m:2')하거나 70% 이하만 유지('p:0.7 m:2', 'p:0.65 m:2', 'p:0.6 m:2')한 경우는 점차 성능이 하락하였다. 프레임 길이는 5개의 특징 벡터를 포함('p:0.8 m:2')하도록 구성하였을 때 가장 좋았다.

정제 알고리즘의 파라미터를 최적화할 경우 유창성에 대해서는 상관계수가 상대적으로 5.3%('p:0.8 m:2), 통합점수에 대해서는 3.8%(p:0.85 m:2) 소폭 향상되었다. 하지만 발음의 정확성 및 억양의 정확성에 대해서는 성능 향상 효과가 미미하였다(최적의 경우 각각 1.6%, 0.9% 향상).

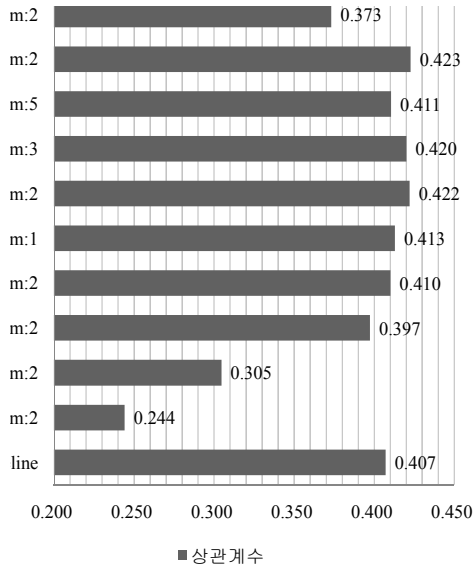

그림 2. 파라미터별 상관관계(통합점수).
Figure 2. Correlations between total scores on different parameters.

동일한 방법으로 긴 발화에 대한 진단 성능이 향상되는지 확인하기 위하여 6초 이상 발성한 테스트 데이터에 대해서 추가 실험을 수행하였다(<표 10>).

6초 이상의 긴 발화에 대하여 진단 성능을 비교한 결과, 정제 알고리즘을 사용한 경우(p:0.8 m:2) 유창성, 발음의 정확성, 억양의 정확성, 통합 점수에 대하여 상관계수가 상대적으로 각각 7.3%, 2.5%, 6.3%, 6.9%씩 증가하였다. 이를 통하여 짧은 발화보다 긴 발화에 대해 정제 알고리즘의 성능 향상 효과가 커지는 것을 확인할 수 있었다.

표 9. 여러 진단점수에 대한 파라미터별 상관관계
Table 9. Correlations between scores on different parameters.

	유창성	발음	억양	통합점수
baseline	0.370	0.219	0.304	0.407
p:0.6 m:2	0.264	0.112	0.142	0.244
p:0.65 m:2	0.293	0.137	0.216	0.305
p:0.7 m:2	0.319	**0.222**	**0.307**	0.397
p:0.75 m:2	0.366	0.214	0.302	0.410
p:0.8 m:1	0.378	0.221	0.302	0.413
p:0.8 m:2	**0.390**	0.209	0.300	0.422
p:0.8 m:3	0.381	0.218	0.303	0.420
p:0.8 m:5	0.365	0.216	0.302	0.411
p:0.85 m:2	0.385	0.208	0.305	0.423
p:0.9 m:2	0.338	0.197	0.270	0.373

표 10. 긴 발화(6초 이상)의 상관관계.
Table 10. Correlations on long utterances (equal or more than 6 secs).

	유창성	발음	억양	통합 점수
baseline	0.428	0.275	0.341	0.477
p:0.8 m:2	**0.460**	**0.282**	**0.363**	**0.510**

5. 결론 및 토의

본 논문에서는 영어 학습자의 영어 발화 수준을 자동 진단하는 시스템을 구축하고 개선 방안을 연구하였다. 이를 위해 먼저 원어민 24명 및 비원어민(한국인) 20명 화자의 영어 발화를 수집한 뒤, 3명의 원어민 평가자로 하여금 문장별로 유창성, 발음의 정확성, 억양의 정확성을 청취 평가하도록 하였다. 발성 내용을 모르는 상태에서 사용자의 발화를 자동 진단하는 시스템을 구축하기 위해 수집한 데이터베이스를 이용하여 TARGET 모델과 SOURCE 모델을 각각 가우시안 혼합 모델로 학습하였다. 구축한 진단 시스템은 유창성, 발음의 정확성, 억양의 정확성에 대한 진단 점수와 통합점수를 함께 도출하도록 하였다.

시스템의 성능을 최대화 할 수 있도록 다양한 특징 추출 방법을 이용하여 비교 실험을 수행한 결과, 12차 MFCCs에 정규화한 에너지와 피치를 더하고, delta 및 delta-delta를 추가한 42차원 특징을 사용하였을 때 가장 높은 성능을 보였다. 그러나 delta 및 delta-delta를 추가할 경우 발음의 정확성에 대해서는 상관계수가 하락하는 결과가 나타났으므로 향후 연구에서는 진단 척도별로 최적의 성능을 보이는 시스템을 각기 구분하여 실험할 필요성이 있을 것으로 판단된다.

본 연구에서는 기존의 방법과 달리 SOURCE 모델을 제 2언어 학습자의 모국어 발화로 구성하지 않고, 제 2언어 학습자의 제 2언어 발화로 구성하였는데, 비교실험을 통하여 이 방법의

유효성을 확인하였다. 그러나 학습 데이터 수가 다르고 채널이 일치하지 않으므로 객관적으로 어느 방법이 더 우수한지를 입증하기 위해서는 추가적인 검증 실험이 필요하다.

성능을 개선하기 위하여 정제 알고리즘을 적용한 결과, 각 척도별 진단 성능이 소폭 향상되었다. 긴 발화를 진단할 경우 성능 향상 폭이 더욱 커지는 것을 확인하였다. 6초 이상의 발화로 테스트하였을 때, 목표(상관계수 0.84)대비 성능 61%(통합점수 기준)를 달성하였다.

향후 연구 주제는 목표 대비 성능을 더욱 높이기 위하여 앞서 언급한 문제점들을 보완하는 것과 3 종류의 진단 척도에 대하여 각각 최적의 성능을 보이는 방법을 따로 구성하여 통합점수를 도출하는 시스템을 구축하는 것이다.

감사의 글

연구를 진행하는 동안 학술대회 등을 통하여 많은 조언을 주신 한국음성학회 회원 여러분들께 감사드립니다.

참고문헌

Kim, M., Kim, H., and Kim, B. (2003). "Performance evaluation of english work pronunciation correction system", The KSPS Spring Conference 2003, pp. 71-74.
(김무중, 김효숙, 김병기 (2003). “한국인을 위한 영어 발음 교정 시스템에 대한 성능 평가”, 2003 대한음성학회 봄 학술대회 pp. 71-74.)

Kim, H. (2002). "An introduntion to 'dr.Speaking' - English pronunciation tutoring system for Korean -", The KSPS 25th Anniversary Conference, pp. 47-50.
(김효숙 (2002). “한국인을 위한 영어발음교정 시스템 'Dr. Speaking' 소개”, 대한음성학회 창립 25주년 기념 학술대회, pp. 47-50.)

Park, J.G., Lee, J.-J., Kim, Y.-C., Hur, Y., Rhee, S.-C., and Lee, J.-H. (2003). "Development of english speech recognizer for pronunciation evaluation", The KSPS Autumn Conference 2003, pp. 37-40.
(박전규, 이준조, 김영창, 허용수, 이석재, 이종현 (2003). “발성평가를 위한 영어 음성인식기의 개발”, 2003 대한음성학회 가을 학술대회 pp. 37-40.)

Anguera, X., Woofers, C., and Hernando, J. (2006). "Purity algorithms for speaker diarization of meetings data", ICASSP, pp. 1025-1028.

Cucchiarini, C., Strik, H., and Boves, L. (2000). "Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms", *Speech Communication, Vol.* 30, No. 2-3, pp. 109-119.

Franco, H., Neumeyer, L., Kim, Y., and Ronen, O. (1997). "Automatic pronunciation scoring for language instruction", ICASSP, pp. 1471-1474.

Moustroufas, N. and Digalakis, V. (2007). "Automatic pronunciation evaluation of foreign speakers using unknown text", *Computer Speech & Language, Vol.* 21, No. 1, pp. 219-230.

Neumeyer, L., Franco, H., Abrash, V., Julia, L., Ronen, O., Bratt, H., Bing, J., Digalakis, V., and Rypa, M. (1998). "Webgrader (tm): A multilingual pronunciation practice tool", Workshop on STiLL, pp. 61-64.

Neumeyer, L., Franco, H., Weintraub, M., and Price, P. (1996). "Automatic text-independent pronunciation scoring of foreign language student speech", ICSLP, pp. 1457-1460.

• **양일호 (Yang, IL-Ho)**
서울시립대학교 컴퓨터과학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5322 Fax: 02-2210-5275
Email: heisco@hanmail.net
관심분야: 음성인식, 화자인식
현재 컴퓨터과학부 대학원 박사과정 재학중

• **김민석 (Kim, Min-Seok)**
서울시립대학교 컴퓨터과학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5322 Fax: 02-2210-5275
Email: ms@uos.ac.kr
관심분야: 음성인식, 화자인식
현재 컴퓨터과학부 대학원 박사과정 재학중

• **유하진 (Yu, Ha-Jin)** 교신저자
서울시립대학교 컴퓨터과학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5322 Fax: 02-2210-5275
Email: hjyu@uos.ac.kr
관심분야: 음성인식, 화자인식
2002~현재 컴퓨터과학부 부교수

• **한혜승 (Han, Hyeseung)**
서울시립대학교 영어영문학부
서울시 동대문구 전농동 90번지
Tel: 010-6411-4508 Fax: 02-2243-4855
Email: jkyoonhan@gmail.com
관심분야: 음성학, 음운론

• **이주경 (Lee, Joo-Kyeong)**
서울시립대학교 영어영문학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5635 Fax: 02-2243-4855

Email: jookyeong@uos.ac.kr
관심분야: 음성학, 음운론
2002~현재 영어영문학부 교수