

한국인의 영어 문장 발음에 대한 한국인/원어민/ILT(Interactive Language Tutor) 평가 점수 사이의 상관관계

이석재, 박진규*
연세대학교 영어영문학과, 동아시테크(주) 기술연구소*

Correlations between pronunciation test scores given by Korean/Native/ILT(Interactive Language Tutor) raters against the Korean-spoken English sentences

Seok-Chae Rhee, Jeon Gue Park*
Dept. of English Language and Literature, Yonsei University,
R&D Center/Dong-A Seetech Co., Ltd*

scrhee@yonsei.ac.kr, jeongue@donga.co.kr

Abstract

This study carried out an experimental English pronunciation assessment to see the differences in the relationship between the different rater categories. The result shows that i) correlation between Korean and Native American raters is high($r=.98$) enough to be considered reliable, ii) previous instructions about assessment rubric and the knowledge about English phonetics and phonology exert little influence on the rating scores, iii) correlation between the automatic ILT(Interactive Language Tutor) rating using speech recognition technology and Natives' rating is stronger than that between ILT and Koreans' rating.

1. 서론

영어교육에 있어 (한국인) 개개인 학습자의 발음에 대한 평가 및 집중 교육은, 제 7차 교육과정 시행 이후 음성 언어(영어)로서의 의사소통이 영어 교육의 중요한 목표로 대두되면서, 그 중요성이 날로 더해가고 있다. 그러나 사람에 의한 발음 평가는 평가 대상 집단의 수가 확대되면 확대될수록 그 시간적 경제적 이유로 인하여 한계에 다다르게 되며, 이러한 사정으로

인하여 음성정보기술(Speech Information Technology)의 언어교육에의 응용을 통한 발음의 자동 평가에 관한 기술 개발([1], [2], [3], [4], [5])이 그 당위성을 갖는다. 이러한 배경 하에서 본 논문은 한국인이 발화한 영어문장의 발음 정도를 평가하려는 노력으로 개발 중인 동아시테크(주) 기술연구소의 ILT(Interactive Language Tutor) 음성인식엔진을 통한 한국인의 영어 발음 평가 결과와 한국인 및 영어 원어민의 동일한 문장에 대한 발음 평가 간의 상관관계를 분석하여 평가자 범주 간의 특징과 그 의미를 찾아보려고 한다. 이러한 작업은 궁극적으로 원어민 평가자와 한국인 평가자, 한국인 평가자 중에서도 음성학 훈련을 받은 평가자와 그렇지 않은 평가자, ILT 평가와 인간에 의한 평가를 비교하여, 기계 평가의 신뢰성을 높이고 예상 문제점들을 미리 짚어 보아 향후 평가 방법의 선택과 기계평가 기술 개발의 기초 자료로 삼기 위함이다.

2. 실험 평가 절차

2.1. 평가 대상 음성 DB

평가의 대상이 되는 원 음성은 한국인의 영어 발화 음성 DB인 K-SEC(Korean-Spoken English Corpus [6])으로부터 가져왔다. K-SEC은 지역간, 연령간, 남녀간 구분을 지어 총 336명(현재 15명의 원어민 성인과 4인의 원어민 어린이 녹음이 부가로 포함되어 있음)의

한국인이 동일한 영어 어휘, 문장, 이야기 등을 발화한 낭독체 정형 영어 음성 DB인데, 여기서 한국인 어린이 10명과 원어민 어린이 2명의 문장 녹음을 무작위로 추출하여 평가 대상으로 삼았다. 무작위 선택된 발화자의 문장 녹음은 DAT로부터 16kHz, 16bit linear PCM, 모노로 샘플링 하여 PC로 옮겼으며, PC 상에서 수작업으로 문장 단위로 파일화 하였다. 아래 표 1은 문장 1에 대한 음성 파일 집합의 예를 보여주는데, 평가자가 파일 이름을 기준으로 화자의 원어민/한국인 구분을 할 수 없도록 파일 이름을 설정하였다. 이런 방법으로 문장 단위로 각 폴더에 저장된 음성 파일들(문장당 12명의 발화자)은 CD로 옮겨져 각 평가자에게 배부되어 평가 되었다. 결국 평가자들은 36개 문장 x 12명 발화자= 432개의 음성 파일에 대한 평가를 내린 결과 된다.

표 1 문장1(Miss Henry drank a cup of coffee)의 파일

AJH_s01.wav	LBK_s01.wav
HDK_s01.wav	LTH_s01.wav
HHJ_s01.wav	PAR_s01.wav
KMH_s01.wav	PHM_s01.wav
KTF_s01.wav	SBK_s01.wav
KYH_s01.wav	TTL_s01.wav

다양한 문장의 유형(평서문, 의문문, 명령문, 감탄문, 단문, 복문, 등)으로 구성된 평가 대상 문장(즉, 녹음된 문장)들은 아래에 제시되어 있다.

1. Miss Henry drank a cup of coffee.
2. What are you looking for?
3. Put your toys away right now.
4. The dancing queen likes only the apple pies.
5. I closed the door and waited for the bus.
6. Hit the ball with this bat.
7. Thirteen years later, Mary met him at the same place.
8. Jenny walked home from school in the rain.
9. Raise your right hand, if the teacher calls your name.
10. Nothing can stop me now.
11. How long does it take to the next bus stop?
12. My brother is coming on Friday.
13. I believe the class begins at ten.
14. The cats should have eaten the hotdog.
15. People couldn't sleep well last night because of the noise.
16. Elizabeth wanted to be a judge.
17. Children like candy very much.
18. It's my sister who talked to the kid.

19. The police took the cab to Seoul.
20. What a surprise!
21. Is the turtle alive?
22. I have friends who are just like me.
23. I see the moon shining brightly all over the lake.
24. If a tree could talk, what would it say?
25. Did he fail the test again?
26. The students are going to the zoo.
27. Call me tonight around seven.
28. What's the weather like today, sunny or cloudy?
29. You like orange juice, don't you?
30. The players sent the cap to Susan.
31. Heat the water to thirty degrees.
32. I can't remember the scene vividly.
33. When did he go skiing, Wednesday or Thursday?
34. We went to London, Paris, Cairo, and Boston.
35. Where do you think you left it?
36. They suspect that the suspect killed Ted.

2.2. 평가자(Rater) 범주

평가자는 4 범주로 구분되었다. 제1범주는 ILT 음성 인식엔진이고, 제2범주는 영어 원어민 화자, 제3범주는 음성학 훈련을 받은 연세대학교 영어영문학과 소속 한국인 전문가이며, 제4범주는 음성학 훈련을 받지 않은 만 20세 이상의 한국인 비전문가들이다.

표 2 평가자 구분

구분	수	항목 설명	이하 약칭
1	1기	ILT 음성인식엔진	ILT
2	3인	영어 원어민	Nat
3	5인	한국인 전문가	Kor(s)
4	19인	한국인 비전문가	Kor(g)

ILT 음성인식엔진에 의한 평가(ILT 기계평가)는 동아시테크(주) 기술연구소에서 행하여졌으며, 점수 산출은 음향모델과의 유사성(triphone 및 uniphone)을 나타내는 정규화된 음소 사후 확률(normalized phone posterior) 값과 정규화된 음소 지속 시간(phone-based duration time) 값을 기반으로 원어민에게는 일괄적으로 95점을 부여한 상태에서 이와 비교되는 한국인 발화에 대한 점수 부여 작업을 100점 만점으로 실시하였다(자세한 과정은 [1]을 참조 할 것). 그 외 인간에 의한 모든 평가는 평가자 각자의 사무실이나 가정에서 컴퓨터를 통한 음성 재생으로 행하여졌다(평가자 상호간의 교류는 없었다). 평가지는 사전에 형식을 통일하여 제작, 배부하였다.

영어 원어민 평가자와 한국인 전문가 평가자들에게는 평가 시 영어의 무슨 음성 음운적 특징에 주의를 기울여야 하는지 각 문장에 따른 평가 항목(rubric)을 설명하였다.¹⁾ 하지만 평가 자체는 총체적(holistic) 평가로서 평가 항목을 의식하면서 재생되는 문장을 듣고 가장 영어답게 발음하였으면 A, 가장 영어답지 못한 발음은 E를 부여하고 그 사이에 B, C, D를 구분지어 총 5단계로 평가하도록 하였다.

이와는 대조적으로 한국인 비전문가 평가자들에게는 평가 항목에 관하여 일체 언급 없이 자신의 기준대로 A, B, C, D, E로 구분지어 들리는 영어 문장에 대한 발음 정도를 평가하도록 지시하였다.

평가 사전 지시에서는 반드시 각 문장 당 12인 전화자(원어민 2인 포함)의 음성을 모두 들은 후에 다시 재생 시키며 문장 당 평가를 하도록 하였다.

사람에 의한 평가지는 평가 개시 후 약 1주일 이후에 회수하여 5단계 등급을 점수화 하였다. 점수는 일

표 3 문장 1, 2의 평가자 범주별 평균 점수

문장 번호	일련 번호	발화자	Kor(s) 5인 m.	Kor(g) 19인 m.	Nat 3인 m.	ILT
1	1	A J H	59	62	72	68
	2	H D K	75	72	78	93
	3	H H J	71	76	72	76
	4	K M H	71	73	62	59
	5	K T F	95	95	95	95
	6	K Y H	71	78	78	77
	7	L B K	73	74	72	76
	8	L T H	81	83	72	69
	9	P A R	69	70	72	66
	10	P H M	71	76	75	77
	11	S B K	69	71	75	76
	12	T T L	93	94	95	95
2	1	A J H	67	69	68	57
	2	H D K	79	80	75	86
	3	H H J	81	82	75	79
	4	K M H	71	72	68	69
	5	K T F	95	93	95	95
	6	K Y H	71	77	75	69
	7	L B K	73	74	75	63
	8	L T H	77	82	72	69
	9	P A R	75	78	72	54
	10	P H M	71	74	72	53
	11	S B K	81	82	78	74
	12	T T L	95	95	92	95

률적으로 A=95, B=85, C=75, D=65, E=55로 한 후, 이를 바탕으로 각 평가자 별 분류 및 통계 작업에 들어갔다. 아래는 상관계수 비교 작업에 들어가기 전에 문장 1, 2에서 각 발화자에 대한 평가자 범주 내 각 평가자들의 평가 평균값을 보여 준다 (이와 같은 작업이 문장 1-36까지 행하여 졌음).

3. 결과 및 토의

3.1. 상관계수 비교

본 장에서는 평가자 범주 사이의 평가 결과 상호 관련성 비교를 위하여 단순 상관계수(correlation coefficient) 추출 결과를 제시한다. 평가자 범주 간 관련성 정도 비교 대상은 아래 표 4에 명시된 (a) 기계평가 대 한국인 전문가 평가, (b) 기계평가 대 한국인 비전문가 평가, (c) 기계평가 대 한국인 평가 전체, (d) 기계평가 대 원어민 평가, (e) 기계평가 대 인간에 의한 평가, (f) 한국인 전문가 평가 대 한국인 비전문가 평가, (g) 원어민 평가 대 한국인 전문가 평가, (h) 원어민 평가 대 한국인 비전문가 평가, (i) 원어민 평가 대 한국인 전체 평가이다.

표 4 상관관계 측정 대상 평가자 범주

1) 평가 항목은 [7]을 참조하여 다음 요인들을 문장별로 평가자에게 설명한 것이다: 순음성 폐쇄음과 마찰음의 구분; 저해음 유·무성의 구분; 저해음 유·무성의 차이와 앞모음 길이의 상관성; 탄설음화의 양상; [s] 다음에 폐쇄음의 기음(aspiration) 정도; 단어 내 위치에 따른 무성파열음의 기음 정도; 파열음의 단어 내 위치에 따른 파열 유·무와 정도; 치경과 경구개치경 마찰음과 파찰음의 구분 정도; 저해음의 부분적 무성음화의 정도; 연음의 양상; 원순음화; 자음-비음 연결체; 모음 사이의 무성폐쇄음 발음; 공명음 [n], [r], [l]의 구분; 축약모음(schwa)의 양상; 리듬; 억양; 휴지(pause).

구분	X	Y
a	ILT 기계평가	Kor(s) 한국인 전문가
b	ILT 기계평가	Kor(g) 한국인 비전문가
c	ILT 기계평가	Kor(total) 한국인 전체
d	ILT 기계평가	Nat 원어민
e	ILT 기계평가	Hum 인간 전체
f	Kor(s) 한국인 전문가	Kor(g) 한국인 비전문가
g	Nat 원어민	Kor(s) 한국인 전문가
h	Nat 원어민	Kor(g) 한국인 비전문가
i	Nat 원어민	Kor(t) 한국인 전체

상기의 상호 관련성 비교는 각 문장 별로 총 36개 문장 전체에 대하여 행하여 졌으며 그 결과는 아래 종합적 표 5에 제시되었다.

아래 표로부터 우리는 모든 문장에 대한 종합적 결과를 놓고 볼 때 각 평가자 범주 간 상관 정도에 있어, (f) 한국인 전문가 대 한국인 비전문가의 상관관계가 가장 높고($r=0.98$), 한국인과 원어민 평가의 상관 정도를 보여주는 (g), (h), (i)에서는 같은 계수 값을 가지면서 상관 정도가 높다($r=0.92$). 이것이 뜻하는 바는 평가 항목에 대한 사전 설명을 들은 한국인 전문가 평가 결과와 그렇지 않은 한국인 비전문가 평가의 결과에 있어 거의 같은 평가를 내린다는 것을 말하고, 원어민 외국인 평가 결과와 한국인 평가 결과 사이에서도 상당히 관련 정도가 높아 유사한 평가 결과가 나왔다는 것을 뜻한다. 또한 이 결과는 한국인의 발음 평가를 상대적으로 심한 경제적 시간적 손실을 감수하며 굳이 외국인으로 하여금 평가하게 할 이유가 없다는 것으로 해석할 수도 있다. 유사한 해석으로 [8]은 본 연구 결과를 보충한다.

표 5 각 문장에 대한 평가자 범주 간의 상관계수

다음으로 기계평가와 사람에 의한 평가를 비교해 보면, (c) 기계평가 대 한국인 전체 평가의 평균 r 값은 0.67이고 (d) 기계평가 대 원어민평가 평균 r 값은 0.72로서 ILT 음성인식엔진에 의한 평가가 한국인에 의한 평가보다는 원어민에 의한 평가와 보다 관련성이 있음을 보여주었다(편차 값 $r=0.05$). 이는 영어 발음에 대한 기계 평가의 궁극적 목표를 한국인의 관점보다는 원어민의 관점을 대신하여 기계가 신뢰성 있는 평가를 하는 것이라고 볼 때 긍정적 결과라고 볼 수 있다. 이는 또한 [2]에서 나타난 SRI의 결과보다 그 정도는 크지 않지만 우월한 상관관계를 보여준다.

3.2. 낮은 상관계수를 보이는 문장의 관찰

아래는 상기의 표 5를 각 상관계수가 높은 순서로부터 낮은 순서로 크게 6등급으로 나누면서(동그라미 선으로 표시) 재정리한 것인데, 이를 통하여 어떠한 문장

문장 번호	a	b	c	d	e	f	g	h	i
	ILT / Kor(s)	ILT / Kor(g)	ILT / Kor(t)	ILT / Nat	ILT / HUM	Kor(s)/ Kor(g)	Nat / Kor(s)	Nat / Kor(g)	Nat / Kor(t)
1	0.71	0.65	0.69	0.89	0.79	0.97	0.8	0.79	0.8
2	0.87	0.84	0.86	0.81	0.86	0.97	0.94	0.91	0.93
3	0.45	0.44	0.44	0.54	0.49	0.98	0.91	0.88	0.9
4	0.6	0.63	0.62	0.63	0.63	0.98	0.95	0.95	0.95
5	0.58	0.61	0.6	0.66	0.62	0.99	0.98	0.97	0.98
6	0.64	0.64	0.64	0.76	0.69	0.99	0.93	0.91	0.92
7	0.74	0.76	0.75	0.85	0.81	0.98	0.87	0.85	0.87
8	0.79	0.77	0.79	0.82	0.81	0.97	0.93	0.95	0.94
9	0.78	0.84	0.81	0.9	0.86	0.98	0.91	0.91	0.91
10	0.71	0.8	0.76	0.82	0.79	0.98	0.92	0.93	0.93
11	0.75	0.77	0.76	0.75	0.77	0.99	0.93	0.94	0.94
12	0.5	0.41	0.46	0.58	0.5	0.97	0.9	0.9	0.9
13	0.76	0.72	0.74	0.75	0.75	0.99	0.94	0.96	0.95
14	0.71	0.68	0.69	0.78	0.73	0.98	0.97	0.96	0.97
15	0.73	0.71	0.72	0.77	0.74	0.98	0.93	0.96	0.95
16	0.59	0.66	0.62	0.69	0.65	0.98	0.95	0.95	0.95
17	0.59	0.68	0.64	0.72	0.68	0.95	0.87	0.93	0.91
18	0.73	0.76	0.74	0.75	0.75	0.99	0.97	0.97	0.97
19	0.76	0.78	0.77	0.82	0.81	0.99	0.87	0.83	0.85
20	0.53	0.57	0.55	0.59	0.57	0.98	0.94	0.94	0.95
21	0.73	0.74	0.73	0.8	0.77	0.99	0.89	0.89	0.89
22	0.68	0.69	0.69	0.71	0.7	0.95	0.96	0.94	0.96
23	0.72	0.72	0.72	0.7	0.72	0.99	0.97	0.96	0.97
24	0.6	0.59	0.6	0.68	0.63	0.99	0.96	0.97	0.97
25	0.53	0.62	0.57	0.77	0.64	0.99	0.93	0.96	0.95
26	0.64	0.67	0.66	0.71	0.69	0.99	0.92	0.94	0.93
27	0.73	0.84	0.79	0.87	0.83	0.94	0.9	0.94	0.93
28	0.57	0.6	0.59	0.69	0.63	0.97	0.89	0.94	0.92
29	0.72	0.69	0.71	0.75	0.73	0.99	0.97	0.95	0.96
30	0.53	0.44	0.49	0.63	0.54	0.99	0.9	0.84	0.88
31	0.64	0.62	0.64	0.56	0.62	0.97	0.92	0.9	0.92
32	0.82	0.78	0.81	0.71	0.8	0.97	0.89	0.81	0.85
33	0.38	0.4	0.39	0.51	0.43	0.97	0.96	0.94	0.96
34	0.66	0.55	0.61	0.63	0.63	0.98	0.89	0.88	0.89
35	0.5	0.43	0.47	0.44	0.47	0.97	0.89	0.87	0.88
36	0.81	0.83	0.82	0.82	0.83	0.99	0.96	0.95	0.96
m.	0.66	0.66	0.67	0.72	0.69	0.98	0.92	0.92	0.92

들에서 평가자 범주 상호 간에 상관도가 낮은 결과가 나온 것인지를 쉽게 확인할 수 있다(괄호 안에 있는 수가 문장 번호를 뜻하며, 지면의 제한 상 (c), (d), (e), (i) 상호 간의 결과만을 제시한다).

표 6 상관계수 순위에 의한 정렬

순위	c	d	e	i
	ILT / Kor(t)	ILT / Nat	ILT / HUM	Nat / Kor(t)
1	0.86(2)	0.9(9)	0.86(9)	0.98(5)
2	0.82(36)	0.89(1)	0.86(2)	0.97(24)
3	0.81(9)	0.87(27)	0.83(36)	0.97(23)
4	0.81(32)	0.85(7)	0.83(27)	0.97(18)
5	0.79(8)	0.82(8)	0.81(8)	0.97(14)
6	0.79(27)	0.82(36)	0.81(7)	0.96(36)
7	0.77(19)	0.82(19)	0.81(19)	0.96(33)
8	0.76(11)	0.82(10)	0.8(32)	0.96(29)
9	0.76(10)	0.81(2)	0.79(10)	0.96(22)
10	0.75(7)	0.8(21)	0.79(1)	0.95(4)
11	0.74(18)	0.78(14)	0.77(21)	0.95(25)
12	0.74(13)	0.77(25)	0.77(11)	0.95(20)
13	0.73(21)	0.77(15)	0.75(18)	0.95(16)
14	0.72(23)	0.76(6)	0.75(13)	0.95(15)
15	0.72(15)	0.75(29)	0.74(15)	0.95(13)
16	0.71(29)	0.75(18)	0.73(29)	0.94(8)
17	0.69(22)	0.75(13)	0.73(14)	0.94(11)
18	0.69(14)	0.75(11)	0.72(23)	0.93(27)
19	0.69(1)	0.72(17)	0.7(22)	0.93(26)
20	0.66(26)	0.71(32)	0.69(6)	0.93(2)
21	0.64(6)	0.71(26)	0.69(26)	0.93(10)
22	0.64(31)	0.71(22)	0.68(17)	0.92(6)
23	0.64(17)	0.7(23)	0.65(16)	0.92(31)
24	0.62(4)	0.69(28)	0.64(25)	0.92(28)
25	0.62(16)	0.69(16)	0.63(4)	0.91(9)
26	0.61(34)	0.68(24)	0.63(34)	0.91(17)
27	0.6(5)	0.66(5)	0.63(28)	0.9(3)
28	0.6(24)	0.63(4)	0.63(24)	0.9(12)
29	0.59(28)	0.63(34)	0.62(5)	0.89(34)
30	0.57(25)	0.63(30)	0.62(31)	0.89(21)
31	0.55(20)	0.59(20)	0.57(20)	0.88(35)
32	0.49(30)	0.58(12)	0.54(30)	0.88(29)
33	0.47(35)	0.56(31)	0.5(12)	0.87(7)
34	0.46(12)	0.54(3)	0.49(3)	0.85(31)
35	0.44(3)	0.51(33)	0.47(35)	0.85(19)
36	0.39(33)	0.44(35)	0.43(33)	0.8(1)

아래 표 7은 최종 관심의 대상인 기계평가 대 평가인 전체 평가 사이의 상관관계가 최하위 등급에 속하는 문장들을 열거한 것이다. 이는 결국 아래 문장들에서 기계평가와 사람에 의한 평가의 차이가 심했던 것으로 이해할 수 있다(r값이 낮은 것부터 열거).

표 7 기계평가 대 인간평가의 상관도가 낮은 문장

ILT / HUM
33. When did he go skiing, Wednesday or Thursday?(0.43)
35. Where do you think you left it?(0.47)
3. Put your toys away right now.(0.49)
12. My brother is coming on Friday.(0.5)
30. The players sent the cap to Susan.(0.54)
20. What a surprise!(0.57)

위 문장들을 대상으로 기계평가와 인간에 의한 평가가 평균 총점에서 많은 차이가 나는 화자를 골라 그 이유를 연구자가 평가 대상 문장을 다시 들어가며 청

취해본 결과, 발화자가 발화 실수로 한 음소나 음절 정도를 발화하였다가 오류를 알아내고는 이를 즉시 취소하고 다시 올바르게 읽은 경우나 실수로 음소 하나 정도가 들어간 경우에 기계 평가는 너무 낮은 점수를 부여하였지만 사람에 의한 평가는 이를 괜찮다고 인정해 주어 상대적으로 더 나은 점수를 부여하였고, 발화자가 문장 내 각 어휘에 대한 확신이 없는 듯 단어 간에 긴 머뭇거린 시간이 경과하고 리듬이나 억양과 같은 운율적 특징이 무시된 경우에도 기계평가는 높은 점수를 부여하였지만, 상대적으로 사람에 의한 평가는 이 경우에 낮은 점수를 주었으며, 기계평가는 단어 사이에서 음운규칙의 잘못된 적용에 관하여 관대하면서도, 생각 단위(thought group) 사이의 극히 정상적인 휴지(pause)에 대하여 감점이 너무 심하였다.

4. 결론

연구의 결과로 i) 비영어권 한국인 전문가 평가자와 비전문가 평가자 사이의 상관관계가 상당히 높으며 ($r=0.98$), ii) 평가전 평가 항목별 지식의 유무가 평가 결과 상관관계에 거의 영향을 끼치지 않으며, iii) 한국인 평가자와 원어민 평가자 사이의 상관관계도 높으며 ($r=0.92$), iii) ILT 음성인식엔진에 의한 평가가 인간에 의한 평가 중 원어민 평가자 그룹의 결과와 가장 유사했음을($r=0.72$) 알 수 있었다.

평가의 신뢰성을 나타내는 평가 상관도에 있어 인간 평가자에 의한 말하기 능력 평가는 타 능력 평가(독해나(0.9-0.99), 듣기(0.8-89)) 결과에서 보이는 평가 간 신뢰 상관도의 계수가 떨어져 0.71-0.79 범위에 있다는 점을 고려할 때[9], 본 연구에 의한 기계평가와 인간평가의 말하기 평가 상관도가 0.72에 이른다는 점은 ILT 음성인식엔진에 의한 평가에서 영어의 운율적 평가가 고려되지 않았음에도 불구하고 나온 결과로 개발 단계에서 긍정적이라고 여겨진다. 아울러 [10]에 의하면, 영어 발화에 있어 운율적 요인의 올바른 구현 여부는 총체적 영어 발음 능력의 판단에 중요한 역할을 하는 것이므로, 이를 평가할 수 있는 어학 학습 전용 음성인식 엔진의 개발 및 그 성능 검토키가 필요한 실정이다.

참고문헌

- [1] 박전규, 이석재, 이준조, 김영창, 허용수, 이종현, “발성 평가를 위한 영어 음성인식기 개발,” 2003 대 한음성학회 가을학술대회 발표집, 2003.
- [2] Frano, H., Neumeyer, L., Digalakis, V. and O. Ronen, “Combination of Machine Scores for

- Automatic Grading of Pronunciation Quality,” *Speech Communication* 30, pp. 121-130, 2000.
- [3] 백승권, 최정규, 한민수, “영어단어발성시의 오류교정 기술에 관한 연구,” *음성과학회 논문집* No. 8, pp. 83-90, 2000.
- [4] 김무중, 김효숙, 김병기, “한국인을 위한 영어 발음 교정 시스템에 대한 성능 평가,” *2003 대한음성학회 봄 학술대회 발표 논문집*, pp. 71-74, 2003.
- [5] Eekenazi, M., “Pinpointing Pronunciation Errors in Children’s Speech: Examining the Role of the Speech Recognizer,” *2002 PMLA Workshop*, pp. 48-52, 2002.
- [6] 이석재, 이숙향, 강석근, 이용주, “한국인의 영어 음성 코퍼스의 설계 및 구축,” *말소리* 46, pp. 159-174, 2003.
- [7] Celce-Murcia, M., Brinton, D.M. and J.M. Goodwin, *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. New York, NY: Cambridge University Press, 1996.
- [8] Kim, D. D., Lim, J. H., Kim, H. A. and S. Kim, “Native Korean Speaking English Teachers Assessing English Pronunciation,” *Proceedings of 2003 Kate Summer International Conference*, pp. 205-210, 2003.
- [9] Lado, R., *Language Testing*. London: Longman, 1961.
- [10] 이석재, 노지영, “한국 초등 영어 교육에서 문장 강세 및 억양 지도 효과에 관한 연구,” *준비 중. 연세대학교*, 2003.

*본 논문은 정보통신부의 산업기술개발사업 (AA-2002-A3-0275-0001) 지원으로 수행되었습니다.

**본 논문의 준비에 있어 실험에 참여해 주신 다음 분들께 감사드립니다: 조경희, 박신영, 강혜영, 장민희, 박우영, 김현아, 나명희, 전선희, 이경훈, 김계완, 구희정, 김윤정, 김유진, 김민주, 최현덕, 김인선, 최우미, 이민진, 노진희, 김정아, 박성혜, 남기현, 안의정, 최학순, Sarah Jung, Dennis Brittingham, Woo Mee Lee.