

청지각적 음성평가의 신뢰도와 정확도 향상을 위한 집중적 훈련 프로그램의 효과에 관한 연구*

최 양 규**

대구대학교 언어치료학과

강 선 정

대구대학교 언어치료학과 석사과정

《 요 약 》

주관적 음성평가는 음성장애 진단에서 기기를 이용한 객관적 음성평가 못지않게 중요하게 다루어지고 있다. 음성평가가 요구되는 임상현장에서 평가의 신뢰도와 정확도는 평가자가 갖추어야 할 필수적인 능력이다. 이 연구는 음성평가 경험이 없는 실험참여자를 대상으로 단기간에 집중적으로 실시되는 컴퓨터로 자동화된 청지각적 음성평가 훈련 프로그램이 주관적 음성평가의 신뢰도 및 정확도를 향상시키는 지를 알아보려고 한 것이다. 청지각적 훈련프로그램에 참여하는 20명의 실험집단과 훈련에 참여하지 않은 20명의 통제집단의, GRBAS의 하위척도인 R, B, S에 대한 사전·사후 음성 평가 결과를 비교하여 청지각적 훈련의 효과를 검증하였다. 실험에 사용한 훈련프로그램은 E-PrimeTM을 통해 제작되었으며, 난이도에 따라 3단계로 구성되었다. 훈련의 각 단계에서 정확도가 80% 이상 도달하면 다음 단계로 진행할 수 있도록 설계되었다. 실험 결과, 청지각적 훈련을 받은 실험집단은 훈련 후의 사후평가에서 훈련 전의 사전평가보다 신뢰도와 정확도, 모두 향상되었으나, 훈련을 받지 않은 통제집단은 그렇지 않았다. 마지막으로 단기간에 음성평가 능력을 향상시킬 수 있는 효과적인 청지각적 훈련 시스템 구현에 지침이 될 수 있는 개선 사항이 논의되었다.

주제어 : 청지각적 훈련, 음성평가, 음성장애, GRBAS, 신뢰도, 정확도

* 이 논문은 2009학년도 대구대학교 학술연구비 지원에 의하여 연구되었음.

** 제1저자, 교신저자(ygchoi@daegu.ac.kr)

I. 서 론

본 연구는 개인용 컴퓨터를 이용하여 자동화한 집중적인 청지각적 훈련 프로그램을 통해서, 주관적인 음성평가 결과의 신뢰도와 정확도를 향상시킬 수 있는지를 검토하고, 향후에 보다 신뢰롭고 정확한 음성평가를 위한 자동화된 체계적인 청지각적 훈련 프로그램 개발을 위한 지침을 마련하는 것에 궁극적인 목표를 두고 있다.

음성평가는 음성장애 진단을 위한 기본적인 과정으로서 MDVP™ 등과 같은 기기를 이용한 객관적인 평가와 아울러 GRBAS(Grade, Roughness, Breathiness, Asthenic, Strained) 등과 같은 주관적인 청지각적 평가가 병행되어야 한다. 기기를 통한 객관적인 음성평가는 중요한 과정이지만 음성 표본 추출(녹음) 환경의 가변성과 분석과정에 적용된 신호처리 알고리즘의 한계로 인해서 오류가 발생하는 것을 피할 수가 없다. 때로는 기기를 이용한 분석 결과보다는 숙련된 평가자에 의한 청지각적 평가 결과가 더 신뢰로운 경우가 많다. 음성의 음질과 같은 복잡한 음향적 특성을 판정 또는 감별하기 위해서는 기기보다는 훈련된 임상가의 청각이 뛰어나고(안희영, 1996), 음향학적 또는 공기역학적 측정에 비해 청지각적 평가가 더 유용하다(Moran & Gilbert, 1984). 따라서 기기를 통한 객관적 평가를 보완하는 차원에서 청지각적 평가는 중요하므로, 음성장애의 정확한 진단을 위해서는 병리적 음성의 특징에 대한 임상가의 숙련된 청지각적 판단 능력은 필수적이다.

청지각적 음성평가의 이러한 중요성에도 불구하고 논란이 되는 것은 바로 평가의 주관성에 내재된 문제들 때문이다. 주관적 평가의 결과는 평가자 간에 그리고 평가자 내에서 평가 시마다 결과가 다르게 나타날 수 있다는 신뢰도의 문제점이 지적되어 왔다(De Bodt & Wuyts, 1997; Holmberg & Oates, 2009; Kreiman, Gerratt, & Berke, 1994; Wuyts & De Bodt, 1999). 따라서 청지각적 평가의 신뢰도를 향상시키기 위한 방안이 모색되고 있다. 임상가의 청지각적 평가의 신뢰도는 음성평가 경험 정도와 비례하므로(김향희 등, 2005), 신뢰로운 음성평가를 위해서는 많은 경험이 필요하다. 임상경험이 없는 초보적인 음성평가자의 신뢰도를 향상시키기 위해서는 표준화된 음성표본을 통한 반복적인 훈련이 필수적이므로(손진호, 2008), 집중적인 청지각적 훈련이 음성평가 능력 향상을 위한 방안으로 제안되어 왔으며, 여러 연구들에서 청지각적 훈련의 효과가 입증된 바 있다. 일정 기간 동안 청지각적 훈련을 시켰을 때 평가자간 신뢰도가 높아졌으며(McWilliams & Philips, 1979), 훈련 프로그램에 의해 경험이 풍부해진 청취 평가자는 훈련받지 않은 평가자보다 더 일관된 신뢰로운 평가 결과를 보였다(Gelfer, 1988). 병리적 음성에 대한 청지각적 평가 경험이 없는 평가자들이 집중적인 듣기 훈련을 받은 후, 음성 표본에 대해 병리적인 정도를 평가하게 했을 때, 거친 정도(R; roughness)와 기식 정도(B; breathiness)에

대한 평가의 신뢰도가 훈련 전보다 현저하게 향상되었다(Eadie & Doyle, 2005).

신뢰로운 청지각적 평가를 위해서는 평가자의 반복 훈련에 사용되는 표준화된 음성 표본이 필요하다(Sören, 1992). 일본에서는 1975년, 음성언어의학회 발성기능 검사법위원회에서 병리적 음성 표본에 대해 각 척도를 평가한 것을 집계하여 제작한 훈련용 음성표본을 이용하여 예비 언어임상가를 대상으로 청지각적 반복 훈련을 실시하고 있다. 그런 반면, 국내에서는 청지각적 평가의 중요성에도 불구하고 표준화된 음성표본에 대한 연구가 미흡한 실정이고, 청지각적 평가의 신뢰도를 향상시키기 위한 체계적인 음성평가 훈련 프로그램이 없으며, 음성평가의 신뢰도 향상에 관한 연구들 또한 미흡한 실정이다.

음성의 병리적인 정도를 평가하는 청지각적 평가 도구로는 GRBAS, CAPE-V (Consensus Auditory Perceptual Evaluation of Voice), 버팔로 음성 프로파일 (Buffalo Voice Profile), 음성프로파일 분석(Vocal Profile Analysis) 등, 여러 가지 평가 도구들이 있으나, 이 중에서 GRBAS 척도가 현재 가장 널리 사용되는 청지각적인 평가 도구로서(Kreiman, Gerratt, & Berke, 1992), 간편하여 쉽고 빠르게 평가할 수 있기 때문에 청지각적 음성평가 도구 중 가장 우수하다고 알려져 있다(양윤수, 2003). GRBAS는 종합애성도(Grade; 이하 G), 조조성(Roughness; 이하 R), 기식성(Breathiness; 이하 B), 무력성(Asthenic; 이하 A), 긴장성(Strainedness; 이하 S), 이 다섯 가지 항목에서 음성을 평가하는 척도인데, G항목은 전반적인 인상에 대한 것이고, A항목은 음성평가 항목으로서 부적절하다는 보고가 있다(손진호, 2008). 따라서 본 연구에서는 GRBAS 척도 중, G와 A, 이 두 항목을 제외하고, 구체적인 음질특성에 대한 항목인 R, B, S 항목을 청지각적 음성평가 항목으로 채택하여 사용하였다.

본 연구는 숙련된 전문가 집단의 일치된 평가 결과에 근거하여 제작된 훈련용 음성 표본을 이용하여 단기간의 집중적인 청지각적 음성평가 훈련 프로그램을 제작한 후, GRBAS 척도의 항목들 중에서 조조성, 기식성, 긴장성, 즉 R, B, S에 대한 평가 항목에서의 신뢰도 및 정확도 향상 효과를 알아보려고 하였다.

II. 연구 방법

1. 실험 참여자

D대학교에서 언어치료를 전공하고 있는 1, 2학년 학생들 중에서 청력이 정상

(양이 청력역치 20dB이하)이고 이전에 음성평가와 관련된 과목을 이수한 적이 없었으며, 본 실험과 유사한 실험에 참여한 적이 없는 40명이 실험에 참여하였다(남자 10명, 여자 30명). 이들 40명 중에서 20명은 청지각적 훈련을 받지 않는 통제집단에, 나머지 20명은 청지각적 음성평가 훈련을 받는 실험집단에 각각 무작위로 배정되었다.

2. 음성 표본

2009~2010년에 서울 ○○병원에 내원하면서 음성장애로 진단된 환자들이 음성 검사실에서 편안한 상태로 /아/ 모음을 약 3초간 연장 발성한 1,000여개의 병리적 음성 D/B에서 무작위로 200개의 표본을 1차로 선정하였다. 이렇게 선정된 200개의 음성 표본을, 임상현장에서 5년 이상의 음성평가 경력이 있는 음성장애 전문 언어치료사 2명에게 GRBAS의 R, B, S 항목에 대해 심한 정도에 따라 정상(0점), 경함(1점), 중간정도(2점), 심함(3점)을 평정하도록 의뢰하였다. 평가된 200개의 음성 표본 중에서 평가자간에 평가값이 완전히 일치하는 80개의 음성 표본을 2차로 선정 후, 이 중에서 20개는 사전 또는 사후 평가를 위한 음성표본으로, 60개는 청지각 훈련용 음성표본으로 최종 선정하였다.

3. 도 구

청지각 훈련 프로그램의 자동화를 위해서 E-Prime™(Psychology Software Tools, v2.0)이 설치된 펜티엄IV급 컴퓨터가 음성자극제시 및 응답점검과 정오 피드백을 위해 사용되었다. 음성 표본은 무작위 순으로 제시하고, 실험참여자의 반응을 기록하고, 정오 피드백을 제시하는 등, 전반적인 실험 통제에는 E-Prime™을 이용하여 개발한 실험용 훈련 프로그램을 펜티엄IV급 컴퓨터에 설치하여 사용하였다. 기존의 GRBAS 척도에서 G와 A, 이 두 항목을 제외하고, 구체적인 음질특성에 대한 항목인 조조성, 기식성, 긴장성, 즉 R, B, S 항목에 대해 0~3점까지 사전·사후 평가와 청지각적 훈련 과정에서 평정하도록 하였다. 각 평가 항목에서 음질이 정상이면 0, 가장 나쁜 상태이면 3으로 평정하게 된다.

4. 절 차

실험은 D대학교의 언어치료학과 음성분석실에서 이루어졌으며, 한 번에 1명씩

실험에 참여하였다. 실험은 ① 음성평가(GRBAS)에 대한 일반적인 교육, ② 사전평가, ③ 청지각적 훈련, ④ 사후평가, 이렇게 모두 4단계의 절차로 구성되었다. 청지각 훈련은 실험집단에게만 실시되고 통제집단에게는 실시되지 않았다.

모든 실험참여자들은 사전평가를 실시하기 전에 음성평가에 대한 개념을 숙지하도록 음성평가에 대한 교육을 받았다. 교육은 청지각 평가 항목인 R(roughness; 조조성), B(breathness; 기식성), S(strainedness; 긴장성) 각각에 대한 개념을 익히도록 해당하는 음성표본을 청취하는 내용으로 구성되었다(부록 I 참조). 일본음성언어의학회에서 제작된 훈련용 음성표본이 이 교육단계에 사용되었다. 약 30분 정도 소요되는 평가교육을 받은 후 사전 평가를 실시하였다.

사전평가에서 실험참여자는 20개의 평가용 음성표본을 각각 무작위순으로 청취한 후 그 음성표본의 R, B, S를 평가하여 0점(정상)에서 3점(심한)까지 평가기록지에 기록하였다. 각 음성표본은 3초간 제시되었는데 실험참가자가 충분히 평가할 수 있도록 연속적으로 3회를 제시하였다. 신뢰도 측정을 위해서 30분 정도의 휴식 후, 동일한 음성 표본에 대해 한 번 더 평가하도록 하였다.

청지각적 훈련프로그램을 개발하는 한 과정인 예비조사에서 5명의 참여자를 대상으로 음성평가를 실시한 결과, 음성표본에서 R, B, S 중에서 항목의 요소들이 중복될수록 평가의 정확도가 낮아지는 경향성을 발견하였는데, 이것에 근거하여 실험집단에게만 실시되는 청지각적 훈련은 난이도에 따라 3단계로 구성하였다. 단계 1에서는 R, B, S 중에서 한 항목의 성분만 포함된 음성표본, 단계 2에서는 두 개의 항목 성분이 중복되는 음성표본, 단계 3에서는 세 개의 항목성분 모두가 들어있는 음성표본이 각각 20개씩 사용되었다. 평가 결과의 채점을 위해서 음성평가 전문 언어치료사의 평가 점수를 정답으로 사용하였다. 컴퓨터 화면 가운데 <스페이스바를 누르시오>라는 지시문이 나타나고 실험참여자가 스페이스바를 누르면 훈련 프로그램이 시작되었는데, 빈 화면 가운데 시행횟수를 나타내는 숫자가 나타나면서 음성자극이 제시되었다. 이 때 이 음성 표본을 청취한 후 키보드로 0에서 3까지의 숫자를 입력하면 정오의 피드백으로서, 정답과 일치하면 “CORRECT”, 정답과 다른 값이면 “INCORRECT”가 화면 중앙에 나타났다. 각 단계에서 20회의 평가시행 중에서 전문가의 평가결과와 일치하는 시행이 16회(80%) 이상일 때 다음 단계로 진행되고 그렇지 못하면 그 단계를 반복 수행하도록 하였다.

청지각적 훈련의 모든 단계에서 정답률 80% 이상의 기준에 달한 경우에 청지각적 훈련을 모두 마친 것으로 보고, 마지막 사후평가를 사전평가와 동일한 절차로 실시하였다. 전체적인 청지각적 훈련 기간은 사람에 따라 달랐으나 평균적으로 총 6시간정도 소요되었으며, 약 2시간씩 3일에 걸쳐 나누어 실시되었다.

Ⅲ. 결 과

결과 분석을 위한 자료로서 전체적인 평가자 내 신뢰도는 사전 또는 사후평가에서 동일한 음성표본에 대한 2회의 평가에서 R, B, S 모두가 일치하는 시행의 회수로 측정하고, 전체적인 정확도는 R, B, S 모두에서 음성평가 전문 언어치료사가 부여한 값과 일치하는 시행 회수의 2회 평균값으로 측정하였으며, 동일한 절차로 R, B, S, 각각에 대한 평가 정확도를 측정하였다. 이렇게 측정한 신뢰도와 정확도를 토대로 다음과 같이 결과를 분석하였다.

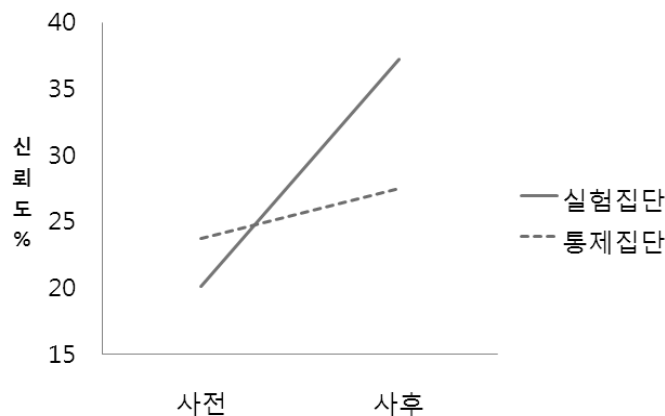
1. RBS 전체 평정에서 평가자 내 신뢰도

사전·사후평가에서 R, B, S값, 모두에서 평정값이 일치한 횟수를 토대로 평가자 내 신뢰도를 제시한 것이 <표 1>과 <그림 1>이다.

<표 1> RBS 평정에서 사전·사후평가의 평가자 내 신뢰도 평균

집단	사전평가	사후평가
실험집단 (n=20)	4.1(20.1%) 1.97 ^a	7.45(37.25%) 2.50
통제집단 (n=20)	4.75(23.75%) 2.46	5.50(27.5%) 2.54

^a표준편차



<그림 1> RBS 평정에서 평가자 내 신뢰도

RBS 평정의 평가자 내 신뢰도에서 청지각 훈련을 받지 않은 통제집단의 경우에 사전·사후 간에 큰 변화가 없었지만, 청지각 훈련을 받은 실험집단에서는 사후평가(37.25%)에서 사전평가(20.1%)보다 17.2% 더 향상되었다. 혼합설계 변량분석을 실시하여 평균 차이의 유의성을 알아보았다(표 2).

<표 2> RBS에서 평가자 내 신뢰도에 대한 변량분석

변산원	df	F
집단(2)	1, 38	1.17
사전·사후(2)	1, 38	20.32*
집단(2) × 사전·사후(2)	1, 38	8.27*

* $p < .01$

RBS에 대한 평가자 내 신뢰도에서 집단의 주효과는 유의하지 않았으나 [$F(1,38) = 1.17$, $p > .05$], 사전·사후 간의 주효과와 사전·사후와 집단간의 상호작용효과는 유의하였다[각각, $F(1,38) = 20.32$, $p < .01$; $F(1,38) = 8.27$, $p < .01$]. 상호작용효과가 유의하여, 집단별로 사전·사후 간의 차이를 t 검증을 통해 유의성 검증을 하였으나 실험집단에서만 유의하였다 [$t(19) = 4.915$, $p < .001$]. 이 결과는 청지각적 훈련 프로그램이 전반적인 신뢰도 향상에 효과가 있음을 시사한다.

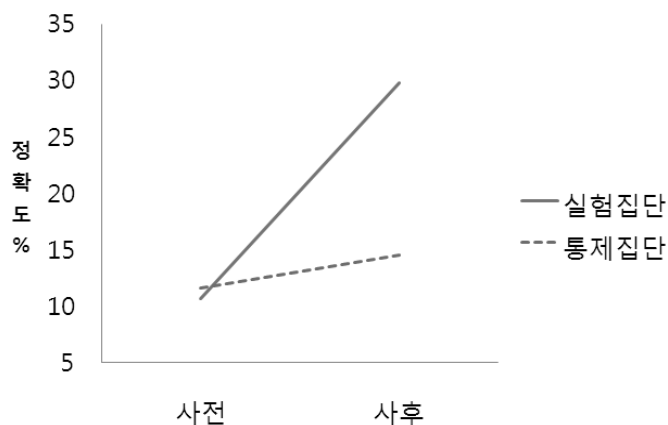
2. RBS 전체 평정의 정확도

RBS 평정의 사전·사후평가에서 R, B, S값 모두에서 전문가의 정답과 일치한 시행의 횟수를 토대로 정확도를 제시한 것이 <표 3>과 <그림 2>이다.

<표 3> RBS 평정에서 사전·사후평가의 정확도 평균

집단	사전평가	사후평가
실험집단 (n=20)	2.15(10.7%) 1.89 ^a	5.95(29.75%) 1.90
통제집단 (n=20)	2.35(11.6%) 1.75	2.90(14.5%) 1.33

^a표준편차



<그림 2> RBS 평가의 정확도

RBS 평정의 정확도에서 청지각 훈련을 받지 않은 통제집단의 경우에 사전·사후 간에 큰 변화가 없었지만 청지각 훈련을 받은 실험집단에서는 사후평가(29.75%)에서 사전평가(10.7%)보다 약 19.1% 더 향상되었다. 혼합설계 변량분석을 실시하여 평균 차이의 유의성을 알아보았다(표 4).

<표 4> RBS 정확도에 대한 변량분석

변산원	<i>df</i>	<i>F</i>
집단(2)	1, 38	11.12*
사전·사후(2)	1, 38	33.48*
집단(2) × 사전·사후(2)	1, 38	22.04*

**p*<.01

RBS의 정확도에서 집단의 주효과, 사전·사후 간의 주효과, 사전·사후와 집단 간의 상호작용효과 모두가 유의하였다[각각, $F(1,38)=11.12$, $p<.01$; $F(1,38)=33.48$, $p<.01$; $F(1,38)=22.04$, $p<.01$]. 상호작용효과가 유의하여 집단별로 다시 사전·사후 간의 차이를 *t* 검증을 통해 유의성 검증을 하였으나 실험집단에서만 유의하였다 [$t(19)=7.09$, $p<.001$]. 이러한 결과 역시, 청각적 훈련 프로그램이 신뢰도와 마찬가지로 평가의 정확도 향상에도 효과가 있음을 시사한다.

3. R, B, S 개별 항목에 대한 평정의 정확도 분석

R, B, S의 각 개별 항목에서의 정확도를 분석한 결과, 청지각적 훈련 프로그램을 받은 실험집단은, 훈련 후에 훈련 전보다 향상되었으나 훈련을 받지 않은 통제집단의 경우에는 그렇지 않았다. 이 결과는 세 가지 모든 항목의 정확도에서 청각적 훈련 프로그램이 향상 효과가 있음을 시사한다. 자세한 것은 다음과 같다.

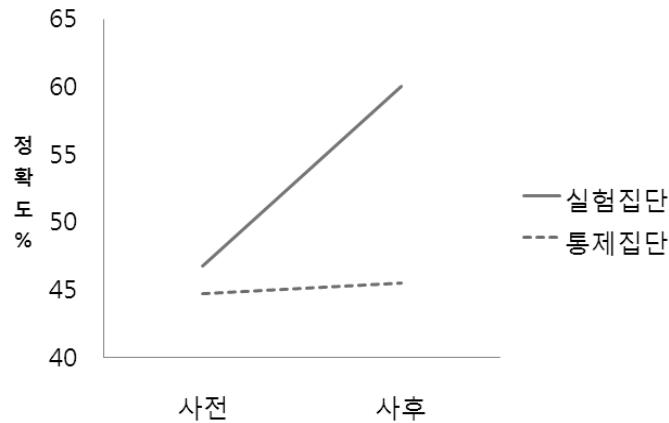
1) R(Roughness) 평정의 정확도

사전·사후평가에서 R값에서 전문가의 정답과 일치한 시행의 횟수를 토대로 정확도를 제시한 것이 <표 5>와 <그림 3>이다.

<표 5> R 평정에서 사전·사후평가의 정확도 평균

집단	사전평가	사후평가
실험집단 (n=20)	9.35(46.75%) 2.13 ^a	12.00(60%) 2.29
통제집단 (n=20)	8.95(44.75%) 2.50	9.10(45.5%) 3.09

^a표준편차



<그림 3> R 평가의 정확도

R값 평정의 정확도에서 청지각 훈련을 받지 않은 통제집단의 경우에 사전·사후 간에 큰 변화가 없었지만 청지각적 훈련을 받은 실험집단에서는 사후평가(60%)에서 사전평가(46.75%)보다 약 13.25% 더 향상되었다. 혼합설계 변량분석을 실시하여 평균 차이의 유의성을 검증한 결과, R에 대한 정확도에서 집단의 주효과, 사전·사후 간의 주효과, 사전·사후와 집단 간의 상호작용효과 모두가 유의하였다[각각, $F(1,38)=7.64, p<.05$; $F(1,38)=6.88, p<.05$; $F(1,38)=5.49, p<.01$]. 상호작용효과가 유의하여 집단별로 사전·사후간의 차이를 t 검증을 통해 유의성 검증을 하였으나 실험집단에서만 유의하였다 [$t(19)=3.40, p<.01$].

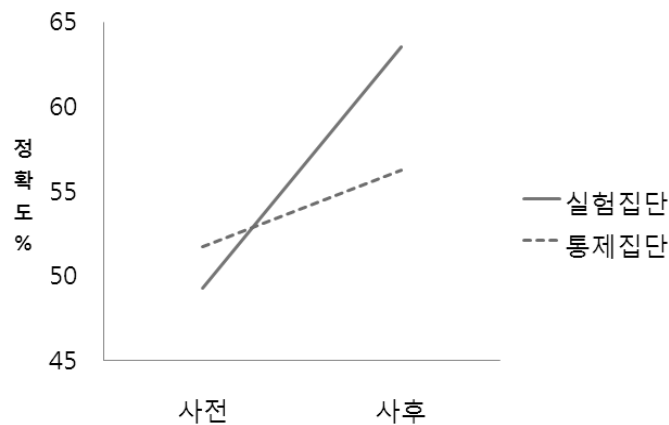
2) B(Breathiness) 평정의 정확도

사전·사후평가에서 B에 대한 평정값이 전문가의 정답과 일치한 시행의 회수를 토대로 정확도를 제시한 것이 <표 6>과 <그림 4>이다.

<표 6> B 평정에서 사전·사후평가의 정확도 평균

집단	사전평가	사후평가
실험집단 (n=20)	9.85(49.3%) 2.79 ^a	12.70(63.5%) 2.10
통제집단 (n=20)	10.35(51.75%) 3.39	11.25(56.3%) 3.02

^a표준편차



<그림 4> 평정의 정확도

B값 평정의 정확도에서 청지각 훈련을 받지 않은 통제집단의 경우에 사전·사후 간에 큰 변화가 없었지만 청지각 훈련을 받은 실험집단에서는 사후평가(63.5%)에서 사전평가(49.3%)보다 약 14.2% 더 향상되었다. 혼합설계 변량분석을 실시하여 평균 차이의 유의성을 알아본 결과, B에 대한 정확도에서 사전·사후 간의 주효과와 사전·사후와 집단간의 상호작용효과가 유의하였다[각각, $F(1,38)=15.34, p<.01$; $F(1,38)=4.15, p<.05$]. 상호작용효과가 유의하여 집단별로 사전·사후간의 차이를 t 검증을 통해 유의성 검증을 하였으나 실험집단에서만 유의하였다 [$t(19)=4.23, p<.001$].

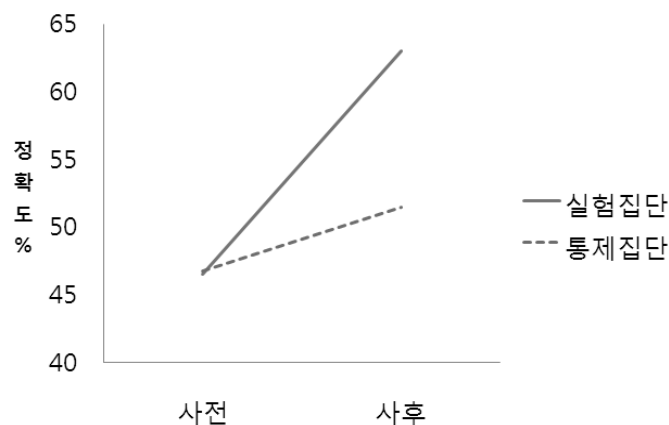
3) S(Strained) 평정의 정확도

사전·사후평가에서 S에 대한 평정값이 전문가의 정답과 일치한 시행의 횟수를 토대로 정확도를 제시한 것이 <표 7>과 <그림 5>이다.

<표 7> S 평정에서 사전·사후평가의 정확도 평균

집단	사전평가	사후평가
실험집단 (n=20)	9.30(46.5%) 2.38 ^a	12.60(63%) 2.64
통제집단 (n=20)	9.35(46.75%) 2.32	10.30(51.5%) 2.69

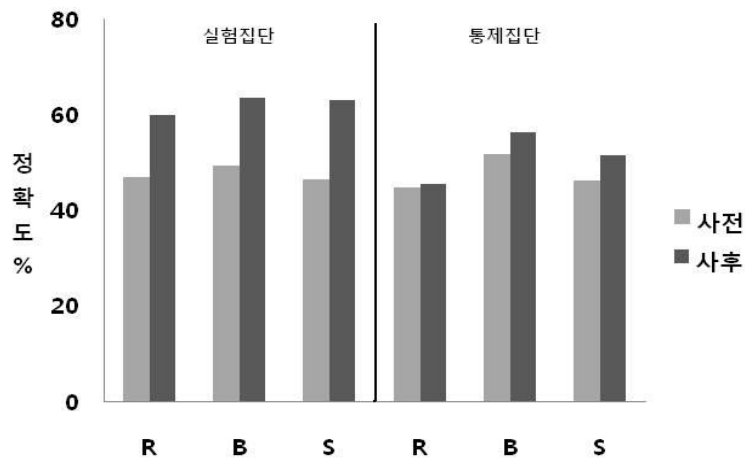
^a표준편차



<그림 5> S 평정의 정확도

S값 평정의 정확도에서 청지각 훈련을 받지 않은 통제집단의 경우에 사전·사후 간에 큰 변화가 없었지만 청지각 훈련을 받은 실험집단에서는 사후평가(63%)에서 사전평가(46.5%)보다 약 16.5% 더 향상되었다. 혼합설계 변량분석을 실시하여 평균 차이의 유의성을 알아본 결과, S에 대한 평가 정확도의 집단의 주효과, 사전·사후 간의 주효과, 사전·사후와 집단간의 상호작용효과 모두가 유의하였다[각각, $F(1,38)=3.92, p<.05$; $F(1,38)=14.51, p<.01$; $F(1,38)=4.44, p<.05$]. 상호작용효과가 유의하였기에 집단별로 사전·사후 간의 차이를 t 검증을 통해 유의성 검증을 하였으나 실험집단에서만 유의하였다 [$t(19) = 4.02, p<.01$].

4) R, B, S 간의 평정 정확도의 비교



<그림 6> 집단별 R, B, S 평정 정확도의 사전·사후 비교

<그림 6>은 집단별로 R, B, S 각각에 대한 평정 정확도를 사전·사후로 대비해 놓은 것이다. 청지각 훈련을 받은 실험집단에서 사전·사후 평균 점수를 비교해보면 단기간에 각 평가 항목은 모두 향상되었는데, S(16.5%), B(14.25%), R(13.25%) 순으로 향상되었다. 이 결과는 청지각적 훈련이 상대적으로 S항목에서 가장 효과적이고 R과 B는 S에 비해 좀 더 집중적인 훈련이 요구됨을 시사해준다.

IV. 논 의

E-Prime™을 이용하여 자동화된 청지각적 음성평가 훈련프로그램의 음성평가 능력의 향상 효과를 검증하기 위해 실시된 본 연구의 결과를 요약하면 다음과 같다.

첫째, 청지각적 훈련프로그램은 평가자 내 신뢰도 향상에 효과가 있었다. 청지각적 훈련을 받지 않은 통제집단은 사전·사후 간의 유의한 차이가 없었으나 청지각적 훈련을 받은 실험집단에서는 전반적인 신뢰도에서 훈련 후(37.25%)에는 훈련 전(20.1%)보다 17.2% 더 향상되었다.

둘째, R, B, S 모두에서 정확한 평가를 한 경우와 R, B, S 각 항목별로 정확하게 평가한 자료를 토대로 정확도를 분석한 결과, 청지각적 훈련을 받은 실험집단은 모든 경우에 정확도에서 유의한 향상을 보였으나 통제집단은 유의한 향상을 보이지 않았다. 실험집단은 RBS 전체 평정의 정확도에서 훈련 후(29.75%), 훈련 전(10.7%)보다 약 19.1% 더 향상되었고, R항목의 평정 정확도에서도 훈련 후(60%), 훈련 전(46.75%)보다 약 13.25% 더 향상되었으며, B항목의 정확도에서 훈련 후(63.5%), 훈련 전(49.3%)보다 약 14.2% 더 향상되었고, S항목의 평정 정확도에서 훈련 후(63%), 훈련 전(46.5%)보다 약 16.5% 더 향상되었다. 이러한 결과들은 국외에서 수행된 선행연구들과 일치한다(Eadie & Doyle, 2005; Gelfer, 1998; Kreiman et. al., 1992; McWilliams & Philips, 1979).

셋째, 이 연구에서 평가항목으로 다루었던 R, B, S 항목별 향상 여부를 살펴보면 실험집단에서는 모두 유의미한 향상을 보였는데, S에 대한 평가 정확도가 가장 많이 향상되었다. 이것은 동일한 향상을 위해서는 B와 R에 대한 훈련이 S에 비해 보다 더 집중적이고 많은 시행이 요구됨을 시사해 준다.

청지각적 음성평가는 앞서 언급하였듯이 주관적이기 때문에 신뢰도의 문제를 항상 안고 있다. 이러한 제한점을 극복하기 위해 평가의 신뢰도와 정확도를 높일 수 있는 체계적인 방안이 시급하다. 국외에서는 청지각적 훈련에 대한 연구가 많이 수행되어왔으며, 일본음성의학회에서는 훈련용 음성표본을 제작하여 표준화한 바 있다. 하지만 국내에서는 관련 연구나 청지각적 훈련프로그램에 대한 연구가 미흡한 실정이다.

따라서 본 연구에서는 청지각적 평가의 단점인 신뢰도 문제를 보완하고자 훈련 프로그램을 개발하였고, 이를 통한 훈련으로 평가자 내 신뢰도와 음성평가의 정확도에서 모두 향상된 결과를 얻었다. 이러한 결과는 보기에 따라서는 그 효과가 미미하다고 여길 수도 있으나, 6시간 정도의 단기간에 수행된 훈련의 효과라는 점을 고려해 볼 때, 청지각적 음성평가를 위한 훈련프로그램의 효과성이 만족할만한 수준은 아니더라도 어느 정도 입증된 것이라 볼 수 있고, 훈련기간을 늘이고 좀 더 체계적

이고 효과적인 청지각적 훈련프로그램 개발과 함께 이와 관련된 추후 연구가 수행된다면 더 좋은 결과가 나올 것으로 예상된다.

후속연구에서 보완되어야하는 사항에 대해 제언을 하자면, 첫째, 집중적으로 진행되기는 하나 6시간 정도의 훈련은 숙련된 전문가의 경험에 비추어 본다면 상당히 짧은 기간이라고 볼 수 있다. 본 연구 결과에 비추어볼 때, 차후 연구에서 좀 더 장기간의 훈련이 진행된다면 초보자를 전문가 수준의 능력에 까지 향상시킬 수 있을 것으로 여겨진다.

둘째, 본 연구는 참여집단과 샘플 수집을 위한 전문적인 평가집단의 수가 제한적인 상황에서 탐색적으로 수행된 연구이기에 연구 결과의 일반화에 한계가 있을 수 있다. 향후, 훈련자로 하여금 보다 다양한 병리 음성표본을 경험할 수 있는 음성표본이 수집되어야할 것이다.

셋째, 본 연구에서 프로그램개발에 사용한 도구인 E-Prime은 일반적인 실험프로그램 제작용이어서 개발에 한계가 있으며 차후에는 JAVA 라든지, Director와 같은 전문적인 소프트웨어 개발 도구를 사용한다면 더욱 전문적인 훈련용 소프트웨어 제작될 수 있을 것이다. 또한 개인용 컴퓨터가 소형화되어 휴대가 용이한 스마트폰 형태로 된 갤럭시나 아이폰에 청지각적 훈련 프로그램을 어플리케이션으로 개발한다면 여러 가지 측면에서 유용할 것으로 여겨진다.

넷째, 이 연구에서는 연구의 편의상, R, B, S 항목에 국한하여 평가항목으로 훈련에 사용하였다. 추후 연구에서는 평가능력 향상이 요구되는 다양한 주관적인 평가항목이 추가된 훈련프로그램을 구성한다면 더욱 훌륭한 청지각적 프로그램이 가능하게 될 것이다.

참고문헌

- 김향희, 이미숙, 김선우, 최성희, 이원용 (2005). 파킨슨증으로 인한 마비말장애의 청지각적 평가에 대한 신뢰도 연구. *음성과학*, 11(4), 129-141.
- 손진호 (2008). GRBAS 음성평가와 음성장애지수. *대한음성언어의학회지*, 19(2), 89-95.
- 안희영 역 (1996). *음성검사법*. 서울: 군자출판사.
- 양윤수 (2003). *음향학적 변수에 기초한 음성장애 중등도 지표*. 박사학위 논문, 전북대학교 대학원.
- De Bodt, M. S., & Wuyts, F. L. (1997). Test-retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. *Journal of Voice*, 11(1), 74-80.
- Eadie, T. L., & Doyle, P. C. (2005). Scaling of Voice Pleasantness and Acceptability in Tracheoesophageal Speakers. *Journal of voice*, 19(3), 373-383.
- Gelfer, M. P. (1988). Perceptual attributes of voice: Development and use of rating scales. *Journal of voice*, 2(4), 320-326.
- Holmberg, E. B., & Oates, J. (2009). Phonetograms, Aerodynamic Measurements, self-Evaluations, and Auditory Perceptual Ratings of Male-to-Female Transsexual Voice. *Journal of voice*, 24(5), 511-522.
- Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic voice quality. *Journal of the Acoustical Society of America*, 96(3), 1291-1302.
- Kreiman, J., Geratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing search*, 35(1), 512-520.
- Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic voice quality. *Journal of the Acoustical Society of America*, 96(3), 1291-1302.
- McWilliams, B. J., & Philips, B. J. (1979). Velopharyngeal incompetence, *Audio Seminars in Speech pathology*. Philadelphia: WB Saunders.
- Moran, M. J., & Gilbert, H. R. (1984). Relation between voice profile rating and aerodynamic and acoustic parameters. *Journal of Communication Disorders*, 17(4), 245-260.
- Sören, F. (1992). Perceptual evaluation. *Journal of voice*, 6(2), 155-158.
- Wuyts, F. L., & De Bodt, M. S. (1999). Is the reliability of a visual analog scale higher than an orinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of voice*, 13(4), 508-517.

A Study on the Improvement of Reliability and Accuracy
in Voice Quality Rating by the Intensive Perceptual
Training Program

Choi, Yang Gyu

Daegu University

Kang, Sun Jung

Graduate School, Daegu University

<Abstract>

The purpose of this study was to examine the effects of the computerized perceptual training program for improvement of the ability to evaluate voice quality. For such purpose, the experiment was performed for evaluating the reliability and accuracy for R, B, S rating of pathological voice samples. 40 college students participated in the experiment. A half of them were allocated in the experimental group with the perceptual training program including three steps according to the difficulty, other half of them allocated in the control group without the perceptual training program. The pre-test and post-test to rate R, B, S scale were taken before and after the perceptual training program. The findings of this study showed that the reliability and accuracy of R, B, S scales rating of experimental group after the perceptual training program were significantly improved than before that program. However the reliability and accuracy of the control group were not improved. The results suggested that the perceptual training program is effective for improvement of the ability to evaluate voice quality. In final section, the guidelines for the effective perceptual training system design and the follow-up researches were discussed.

Key Words

: perceptual evaluation, voice evaluation, reliability improvement, voice quality, voice disorders

논문 접수: 2010. 11. 04 심사 시작: 2010. 11. 10 게재 확정: 2010. 12. 15

<부록 I> 음성평가(RBS)에 대한 일반적인 교육 내용

1. 음성 표본의 각 평가항목(R, B, S)의 개념에 대한 교육 내용

1) 조조성(Roughness; R)의 개념 학습

- ① 조조성(R)은 거친음성이라고 하며, 성대 진동이 불규칙하여 나는 소리로 거친 소리 혹은 불쾌한 소리로 표현할 수 있으며 불규칙한 성대 진동으로 인하여 기본주파수나 진폭이 불규칙하고 낮은 주파수대역에서 잡음이 발생하는 소리임을 학습함.
- ② R을 가지고 있는 전형적인 음성 표본 청취.
- ③ R의 심한정도가 0에서 3까지 다양한 음성 표본의 예들을 청취.

2) 기식성(Breathiness; B)의 개념 학습

- ① 기식성은 성문폐쇄부전으로 인한 공기가 새는 소리로 중음역 이하의 대역의 잡음임을 학습함.
- ② B를 가지고 있는 전형적인 음성 표본 청취.
- ③ B의 심한정도가 0에서 3까지 다양한 음성 표본의 예들을 청취.

3) 긴장성(Strainedness; S)의 개념 학습

- ① 긴장성은 노력성이라고도 하며, 성대가 비정상적으로 과긴장 상태이거나 성대가 과도하게 단단하여 무리하게 힘을 들여서 나는 소리이며 기본주파수가 높고 고음역에서 잡음이 나타남을 학습함.
- ② S를 가지고 있는 전형적인 음성 표본 청취.
- ③ S의 심한정도가 0에서 3까지 다양한 음성 표본의 예들을 청취.

2. 두 개의 평가항목 특성이 혼합된 음성 표본 청취

- 1) 조조성(R)과 기식성(B)이 동반되는 음성 표본 청취
- 2) 조조성(R)과 긴장성(S)이 동반되는 음성 표본 청취
- 3) 기식성(B)과 긴장성(S)이 동반되는 음성 표본 청취

3. 세 가지 평가항목 특성 모두가 혼합된 음성 표본 청취

