구관조 음성모방의 음향학적 분석을 통한 음성인식에 대한 고찰

The Study of Voice Perception with Formant Analysis of Two Myna Bird's Voice Imitation

이 옥 분* · 정 옥 란**
Ok-Bun Lee · Ok-Ran Jeong

ABSTRACT

This study was an attempt to determine acoustic characteristics in myna bird's notes. Two myna birds' sounds imitating a normal male voice in his late 20's were sampled and analyzed. The analyses included the mean values of F1, F2, F3 and pitch contours.

The results were as follows;

First, there was a significan difference in the mean values of F1, F2, and F3 in isolatd vowel /a/ and /i/ between the myna birds' sounds and the human voice. However, there was no apparent difference in pitch contour of their formants. Second, there was a difference in pitch contour of their formants in their sentence ("hn-nyung-ha-se-yo?" meaning "How are you?") production. Namely, the myna birds' pitch contour was located higher than that of the human's.

Keyword: Myna bird, Formant Analysis, Pitch Contour

1. 서 론

구관조는 사람의 목소리를 흉내낼 수 있는 새로 널리 알려져 있다. 사람의 목소리 색깔이나 발음이 거의 흡사하여 어떤 단서가 주어지지 않는 한 음성모방의 샘플만을 듣고서는 새의 목소리라고 정확하게 판단하기가 쉽지 않다. 구관조는 일정기간 동안 전문가의 훈련을 통해 반복적으로 발성 및 발음하는 흉내를 연습한다. 계속되는 반복학습을 통해 사람의 말을 만들어 나가는 것이다. 이 구관조는 다른 종류의 새들의 소리도 흉내내는 것이 가능하다. 물론 이것은 어디까지나 모방에 국한 된 것이며, 이러한 음성모방을 통해서 사람들과 동일한 의사소통 메시지를 주고 받지는 못한다. 중요한 것은 구관조들이 반복학습 통해서 사람의 목소리의 색깔과 발음을 어떻게 인식하고 그것을 모방하여 산출하게 되는 과정에 대한 것이다. 물론, 구관조가 모방이 가능한 발화의 어절 수는 한정적이다. 그러나, 사람의 사고하는 뇌 구조와는 엄연히 다른 구관조가 사람의 목소리를 흡사하게 모

* 한림대학교 언어청각학부 교수
** 대구대학교 언어치료학과 교수

방할 수 있을 정도로 음성인식을 하고 소리를 낸다는 점에 착안해 볼 때, 음성인식 기술에 흥미로운 정보를 제공하게 될 것이다.

현대에는 음성합성, 음성인식에 대한 전문영역이 급속한 속도로 발전해 가고 있다. 이러한 기술의 혜택을 일상생활에서도 손쉽게 접하고 있다. 의사소통 과학의 분야에서도 사람들의 말소리에 대한 보다 전문화되어 있는 음향분석의 발전을 토대로 언어장애 영역뿐만 아니라 지능 로봇의 개발의 영역에도 많은 도움을 주고 있다. 말소리 음향분석의 여러 유형 가운데 포만트 값의 분석을 조음기관의 운동특성 즉, 구강 내 혀의 위치 변화, 조음점 접촉 변화 등의 특성을 추정할 수 있고, 알려진 대로 공명강의 면적을 알게 된다. F1 값은 구강 내 위치하는 혀 높낮이와 인두강과 상관관계가 있고, F2 값은 구강 내 혀의 전·후 위치와 관련이 있다(Ryalls & Behren, 2000; 김병욱, 1996). 이러한 포만트의 값은 모음이 분석 자료가 된다. 즉 말소리의 명료도와 화자의 음색을 지각하는 데 있어서 중요한 자질이 된다(김기호 외, 2004).

이러한 점에서 기존의 기본적인 포만트 분석 기준을 토대로 구관조의 음성모방에 대한 포만트 분석은 음성인식에 대한 흥미로운 정보를 제공할 것이다. 구관조가 사람과 흡사한 음성을 모방할 수 있다는 것은 화자가 산출하는 모음의 포만트의 값을 추정하여 이를 토대로 구강 내에서 혀 위치를 변화를 적절하게 시키고 있다는 것을 추정해 볼 수 있다. 한계점은 구관조가 화자의 음성을 인식하는 뇌 구조 시스템을 시각적으로 밝힐 수는 없다는 것이다. 그러나, 가능한 것은 구관조가 화자의 음성 모방된 샘플을 분석하여 조음산출 방식을 좀더 객관적으로 파악할 수 있다는 점이다. 즉, 일반인들의 말소리 산출결과를 토대로 말소리의 인식과정을 추정해 내듯이 구관조들이 음성을 인식하는 방식을 추측해 볼 수 있다.

일반적으로 앵무새나 구관조의 음성모방에서 특정한 포만트를 찾을 수 없는 것으로 알려져 있다. 아마, 이러한 관심이 구체적인 실행으로 이어지지 않았기 때문일 수도 있다. 그 이유는 앵무새의 구강구조가 크지 않아 사람들과 유사한 공명강 자체를 가지지 못하기 때문이라는 것이다. 그러나, 사람들의 성대모사에 있어서도 공명강의 크기 자체가 결정적인 요소라고 판단하지 않는다. 이것은 사람의 성도는 성별, 연령을 포함한 각 화자의 특성에 따라 성도의 특징이 어느 정도 다르기 때문에 절대적인 주파수가 제시될 수는 없다(Borden & Harris, 1980). Ryalls & Begrens(2000)은 이러한 개인 간, 또는 개인 내 차이에도 불구하고 음향적으로 다른 신호들을 언어학적으로 동일한 음가로 지각할 수 있는 것은 청자가 음을 인식하는데 있어 끊임없이 언어학적 판단을 적용시키려 하기 때문이라고 하였다. 따라서 음성인식은 복잡하고 변화하는 음향적 정보에서 특정한 음향 유형을 단서로 음성을 지각하고 언어적으로 범주화하는 과정이라고 할 수 있다.

구관조나 앵무새의 소리는 의미가 있는 발화라고 여겨질 만큼 지각적으로 뚜렷한 특징을 지니고 있다. 따라서 그들이 산출하는 울음소리에는 청자인 사람이 듣고 인식을 할 수 있도록 하는 특정한 음향적 단서가 포함되어 있을 것이며, 그것은 사람의 음성에서 찾을 수 있는 음향적 단서와 유사할 것이라 유추할 수 있다.

따라서, 이 연구에서는 정상 성인 화자와 동일하게 인식되는 구관조들의 모방음성에 대한 포만트 분석을 실시하여 사람 말소리와는 어떠한 차이점이 있는가를 알아보고자 한다. 이 연구결과는 구관조가 사람의 말소리를 인식하여 처리하는 과정에 있어서 그 음성인식에 미치는 영향을 알아보는 데 중요한 정보를 제공하리라 사료된다. 이에 따른 구체적인 연구문제는 다음과 같다.

첫째, 단일 모음 수준(아, 이)에서의 구관조 모방음성 간, 정상 화자간의 포만트 값포만트 구성에서 차이가 있는가?

둘째, 연속 발화(안녕하세요)에서의 구관조 모방음성 간, 정상화자간의 포만트 구성에서 차이가 있는가?

2. 연구방법

2.1 연구대상

대구광역시 소재 화원유원지에서 사육되는 구관조로 생후 10 년(A)과 9 년(B)된 2 마리를 연구대상으로 하였다. 두 마리 모두 생후 6 개월 이전부터 매일 오전, 오후 일정한 시간에 규칙적으로 음성모방 훈련을 받았으며, 사육사는 모두 성인 남성이었으나, 음성 모방훈련은 주로 성인여성에 의해 받아왔다.

2.2 연구 도구 및 절차

구관조의 음성모방 소리는 SONY TCM-20DV 녹음기를 사용하여 SK Slim120 테이프에 녹음하였다. 소리의 표집은 유리문으로 외부와 통제되어 있는 사육실 내부에서 표집하였고, 구관조와 녹음기와의 거리는 정방향 40-50 cm였다. 울음소리의 음향학적인 분석은 Praat(version 4.1.4)을 이용하였다.

3. 연구결과

3.1 구관조 음성모방의 모음 포만트 분석 결과

구관조 A, B를 대상으로 표집한 성인화자 음성모방 샘플에 대한 포만트 분석 결과는 표 1과 표 2와 같다. 음성모방 샘플의 선정은 연구자들이 정확히 모음 '아'와 '이'만을 각각 5 차례 걸쳐 선택한 값의 평균값을 구하였다. 이 포만트 분석 자료는 그림 1과 같다.

표 1. 모음 /아/로 지각되는 소리의 포만트 주파수(단위 Hz)

	F1			F2			F3		
	A	B	정상성인 화자*	A	B	정상성인 화자*	A	B	정상성인 화자*
M	893.6	1095.6	775	1618.2	1519.0	1357	2953.2	2315.0	2540
SD	69.19	127.12	738	46.04	67.61	1372	59.14	217.82	2574

* 양병곤(1996)의 연구결과 인용

표 1에서 보는 바와 같이, 모음 /이/에 대한 음성모방에서의 포만트 분석결과 구관조 A와 B의 F1 값(Hz)의 평균치는 893.6과 1095.6이고, F2값은 1618.2와 1519.0이며, F3값은 2953.2와 2315.0이다.

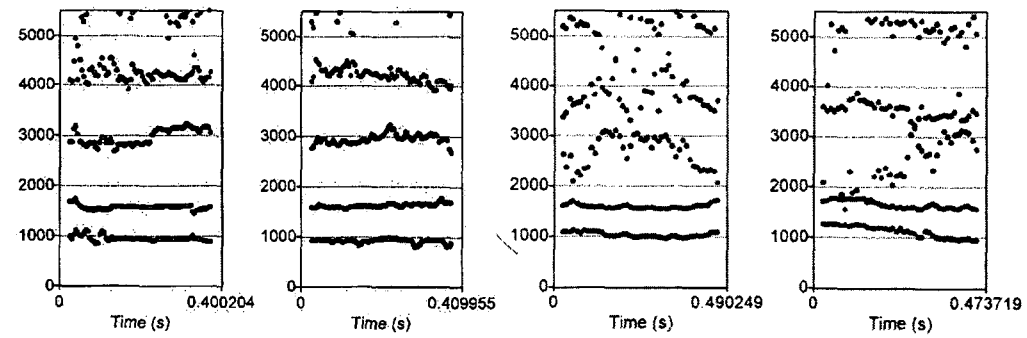
표 2. 모음 /이/로 지각되는 소리의 포만트 주파수(단위 Hz)

	F1			F2			F3		
	A	B	정상성인 화자*	A	B	정상성인 화자	A	B	정상성인 화자
M	876.8	834.2	386	2474.4	1793.6	1730	2955.4	2170.0	2612
SD	147.18	140.86	341	131.52	56.75	2219	115.07	70.65	3074

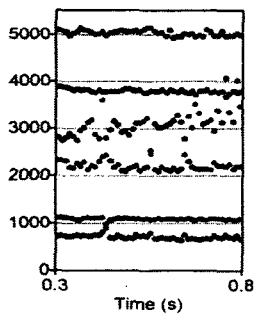
* 양병곤(1996)의 연구결과 인용

모음 /이/에 대한 구관조 A와 B의 F1값의 평균치(Hz)는 876.8과 834.2이며, F2값은 2474.4와 1793.6이고, F3값은 2955.4와 2170.0이다. 분석 결과에서 구관조 A와 B간에도 동일화자의 모방음성임에도 산출된 포만트 값의 차이가 현저하게 존재함을 제시하고 있다. 이 결과는 정상성인 화자들을 대상으로 한 양병곤의 연구(1996)에서 정상인 화자의 모음 포만트 값에도 절대적인 수치가 존재하지 않음을 제시하였다.

그림 1은 구관조 A와 B의 각각 모방음성에 대해 각각 2 개의 음성샘플(/아)을 분석한 것을 그림으로 제시하였다.



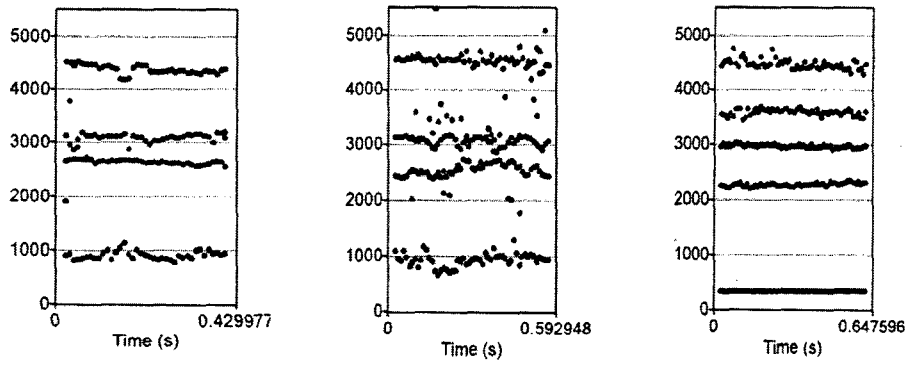
구관조 A 구관조 B



정상성인(연구자) 모음 /아/

그림 1. 구관조 A, B의 모음 /아/의 포만트 분석

모음 /아/에 대한 모방음성의 포만트 형성에서 구관조 A는 거의 일정한 패턴을 가지고 있으며, 구관조 B는 다소 간의 포만트 형성 모양에 차이가 있다. 그러나 비교적 두 구관조 모두 일관성 있는 포만트를 형성하고 있다. 모음에 따른 F1, F2의 상대적인 포먼트 형태는 구관조 간, 구관조 내에서 일관성 있게 나타났다. 반면에 F3의 주파수는 구관조 내에서는 일관되게 나타났으나, 구관조간에 차이가 있었다. 또한 동일 환경에서 수집한 연구자(남성화자)의 실제 모음 /아/ 산출시 표집된 포만트와 유사 형태를 갖추고 있다. 이러한 결과는 그림 2에서 보는 바와 같이, 모음 /이/에서도 동일한 패턴으로 나타났다.

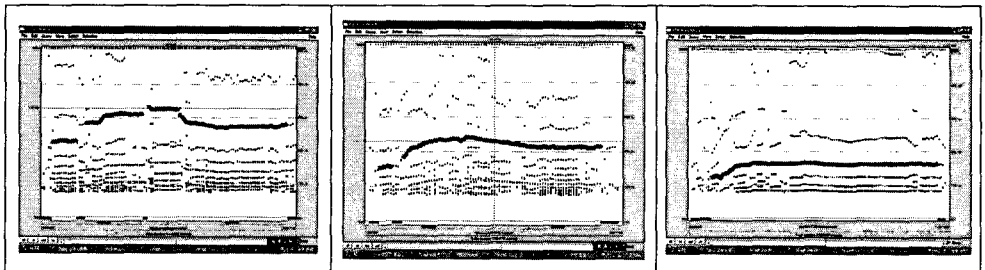


구관조 A 구관조 B 정상 성인 화자

그림 2. 구관조 A, B의 모음 /아/의 포만트 분석

3.2 연속 발화(안녕하세요)에서의 구관조 간, 정상화자간의 포만트

그림 3은 '안녕하세요'의 구관조 모방음성과 정상성인 남성 화자(구어 모델링제시)의 말 샘플에 대한 음도 변화 패턴을 tracking 한 것을 제시하였고, 이 발화의 구관조 모방음성의 유사성을 확인할 수 있도록 그림 3에 말샘플의 종결어미에 해당하는 '-에요'에 해당하는 포만트 구조의 변화를 제시하였다. 구관조 A, B와 구어 모델링 화자의 말 샘플에서 음도 변화 패턴을 비교하였다.

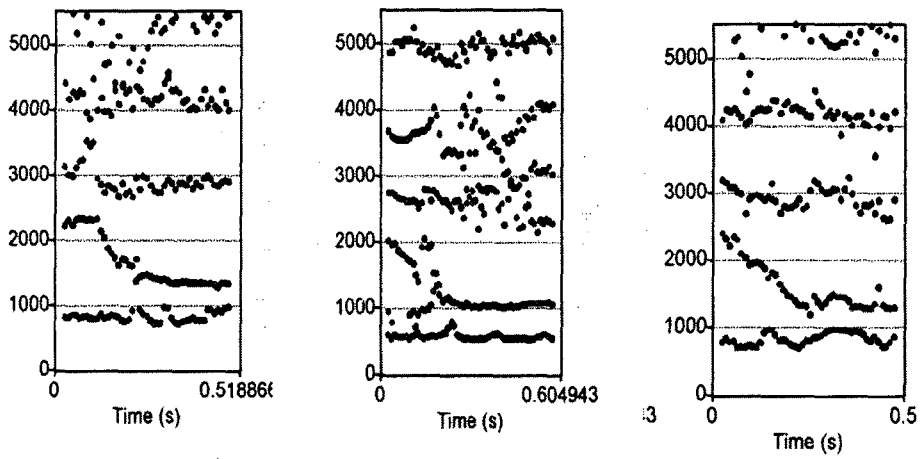


구관조 A	구관조 B	성인 화자(모델링 제시)

그림 3. '안녕하세요'의 구관조 모방음성과 정상 화자 발화 포만트 분석결과

음도 변화 분석 결과, 연구자들의 청지각적 판단결과(녹음된 구관조의 모방음성과 성인화자의 목소리 구별하기 힘듦)와는 매우 상이하게 구관조 간 구관조 내, 성인 남성과의 비교에서 산출된 자음의 음향 분석결과가 모두 다르게 나타났다. 구관조 소리 "안녕하세요"에서 유성자음 /ㄴ/의 생략이 두드러졌고, 무성자음 /ㅅ/의 유성음화가 나타났다. 그러나 전체 표집(구관조 A:14, 구관조 B: 8)에서 이러한 음운변화가 동일하게 일어나지는 않았다.

이러한 포만트 구조 변화의 패턴이 다소 상이한 듯 산출되었으나, 그림 4에 제시된 '-에요'의 말소리 분석에서는 성인 화자의 포만트 구조 변화와 거의 흡사한 모양의 결과가 산출되었다.



구관조 A 구관조 B 성인화자(모델링 제시)

그림 4. '-에요' 에 대한 포만트

4. 결론 및 토의

이 연구는 구관조의 모방음성 포만트 구조를 분석하여 정상 성인 화자간의 어떠한 차이점이 있는 가를 알아보므로써 구관조의 모방음성 산출의 특성을 연구하는데 의의가 있었다. 또한 이 결과를 토대로 음성인식 분야에 대한 고려점을 제시하고자 하였다. 이에 따른 연구결과는 다음과 같다.

첫째, 단일 모음 수준(아, 이)에서의 구관조 모방음성 간, 정상 화자간의 포만트 값에서 차이가 나타났다. 모음 /이/에 대한 음성모방에서의 포만트 분석결과 구관조 A와 B의 F1값(Hz)의 평균치는 893.6과 1095.6이고, F2값은 1618.2와 1519.0이며, F3값은 2953.2와 2315.0이다. 모음 /이/에 대한 구관조 A와 B의 F1값의 평균치(Hz)는 876.8과 834.2이며, F2값은 2474.4와 1793.6이고, F3값은 2955.4와 2170.0이다. 분석 결과에서 구관조 A와 B간에도 동일화자의 모방음성임에도 산출된 포만트 값의 차이가 현저하게 존재함을 제시하고 있다. 이러한 결과는 정상 성인을 대상으로 한 양병곤의 연구(1996)에서도 동일하게 제시된 사항이다. 즉, 정상인 화자 개인 간에도 많은 차이가 존재하기 때문에, 모음 포만트 값에도 절대적인 수치가 존재하지 않다는 그의 견해와도 같이 구관조 간에

도 일치하지 않은 포만트 값은 타당한 경향인 것 같다. 그러나, 모음 /아/와 /이/에 대한 모방음성의 포만트 형성에서 구관조 A는 거의 일정한 패턴을 가지고 있으며, 구관조 B는 다소 간의 포만트 형성 모양에 차이가 있었다. 그러나 결과적으로 비교적 두 구관조 모두 일관성 있는 포만트를 형성하고 있다는 것이 나타났다. 즉, 모음에 따른 F1, F2의 상대적인 포먼트 형태는 구관조 간, 구관조 내에서 일관성 있게 나타났다. 반면에 F3의 주파수는 구관조 내에서는 일관되게 나타났으나, 구관조 간에 차이가 있었다. 무엇보다 동일 환경에서 수집한 연구자(남성화자)의 실제 모음 /아/ 산출시 표집된 포만트와 유사 형태를 갖추고 있다는 사실이 나타났다. 이 결과는 청지각적으로 두 연구자가 구분할 수 없을 정도의 유사한 모방음성이 절대적인 포만트 값에 근거하기 보다는 포만트 구조의 형태에 의존하고 있다는 사실을 의미한다. 이는 비록 사람과 다른 구조의 공명강을 가지고 있다하더라도 구관조의 혀 조절 능력이 뛰어나다는 점을 감안 해 볼 때, 사람과 유사한 모방음성 산출의 가능성이 구강내 혀 위치 변화가 매우 중요한 요소임을 시사하고 있다.

이상의 결과를 보다 뒷받침하기 위해 연속발화에서의 구관조 모방음성을 분석한 두 번째 결과에서는 구관조 간 모방음성 그리고 구관조와 정상화자간의 포만트 구성에서도 첫 번째 결과와 유사한 경향이 나타났다. '안녕하세요'의 구관조 모방음성과 정상성인 남성 화자(구어 모델링제시)의 말 샘플에 대한 음도 변화 패턴을 tracking의 결과, 구관조들 간 그리고 성인 화자간의 모방음성의 pitch contour의 모양은 거의 유사하게 산출이 되었다. 단, pitch contour가 형성된 주파수대가 구관조들이 성인 화자에 비해 다소 높게 위치하고 있었다. 또한 모방음성 샘플에서 말 샘플의 종결어미에 해당하는 '-에요'에 해당하는 포만트 구조의 변화에서 구관조 간, 그리고 구관조와 성인 화자간의 포만트 구조 모양이 거의 일치한 형태를 띠고 있었다. 이러한 결과는 첫 번째 연구문제에 따른 결과와 상응하는 것으로, 구관조의 비교적 발달된 혀의 조음 능력이 최대한 화자의 음성과 유사한 공명구조를 형성하는 게 가능하다는 것을 시사하고 있다.

이상의 결과를 종합해 볼 때, 구관조의 모방음성의 포만트 분석 결과 포만트 값을 절대적인 기준으로 정상화자와 비교하기보다, 포만트 구조의 모양을 근거로 비교하는 것이 구관조의 사람의 말소리 모사능력을 이해하고, 더불어 구관조가 화자의 말소리를 인식하는 방식을 이해하는 데 도움이 될 것이다. 이는 F1, F2, F3의 값은 알려진 대로 구강 내 혀 위치 변화와 밀접한 상관이 있다는 점과 구관조가 혀의 조절능력이 다른 종류의 새들에 비해서 뛰어나다는 점을 함께 감안해 볼 때, 사람과는 매우 제한된 구강을 가진 구관조 모방음성에 대한 음향학적 연구가 음성합성 및 음성인식 분야에 지속적인 도전과제가 될 것이다. 음성인식 분야에서 모음의 F1, F2, F3값의 절대치를 각각 구하는 것과 더불어 각 포만트 값의 넓이와 전체적인 형태가 음성인식의 정확률을 높이는 데 어떤 역할을 하는지 연구되어져야 할 것이다. 또한 이 연구결과에서 제시된 바와 같이 동일한 pitch contour로 모방음성이 가능하고, 자음산출 능력과 관련한 구관조의 말소리 정보처리 능력에 대한 연구가 체계적으로 계획되어져야 할 것이다. 구관조는 인간이 지니는 언어적인 지식이 없거나 있다고 하더라도 매우 낮은 수준일 것으로 가정되는 바, 언어적인 지식과는 별개로 음향적인 정보에 의존하여서도 소리를 학습하고 변별하고 산출할 수 있다는 것을 뜻한다. 그러므로 최근 음성인식 기술에서 만족할 만한 성과를 얻기 위해 언어학적 지식을 접목하려는 시도는 의미 있는 일이고, 지속되어야 할 부분이지만, 순수하게 음향학적인 연구만을 통해서도 해결해야 할 많은 과제들이 많다고 볼 수 있다. 그러나, 구관조의 말소리 기억능력(말소리 길이, 발음의 난이도)에 관한 연구가 함께 이

루어져야 할 것이다. 이 연구에서는 실험대상의 수가 적었고, 모방음성 샘플도 극히 한정적이었기 때문에 성인 화자의 모델링이 되는 말소리 길이, 언어학적 정보처리, 발음의 난이도 등을 고려하지는 못하였다. 추후 연구에서는 이와 같은 미비점을 보완하여 실험을 한다면 보다 명확한 결론과 더불어 음성인식 및 합성 분야에서도 의미 있는 연구가 될 수 있을 것으로 생각된다.

참 고 문 헌

김기호, 양병곤, 고도흥, 구희산. 2004. 음성과학. 서울: 한국문화사.
김병욱. 1997. 응용언어과학. 한국언어치료학회.
양병곤. 2003. 프라트를 이용한 음성분석의 이론과 실제. 만수출판사.
이호영. 1996. 국어음성학. 태학사.
Orden G. J., & Harris, K. S. 1980. *Speech Science Primer Physiology, Acoustics, and Perception of Speech.* WILLIMAS & WILKINS.
Ryalls, J., & Behrens, S. 2000. *Introduction to Speech Science; From Basic Theories to Clinical Applications.* Allyn and Bacon.

접수일자: 2005. 04. 30
게재결정: 2005. 05. 31

▶ 이옥분
강원도 춘천시 옥천동 1번지 (우: 200-702)
한림대학교 자연과학대학 언어청각학부
Tel: +82-33-248-2215
E-mail: oblee72@hallym.ac.kr

▶ 정옥란
대구광역시 남구 대명 3동 2288번지 (우: 712-830)
대구대학교 재활과학대학 언어치료학과
Tel: +82-53-650-8274, Fax: +82-53-650-8259
E-mail: oj@daegu.ac.kr