

모음지각에서 화자 변산성 처리: 정규화 대 본보기

최 양 규 · 신 현 정

춘해대학 유아특수치료교육과 부산대학교 심리학과

모음지각에서 전통적인 정규화모형은 화자 변산성을 제거되어야 할 소음으로 처리하는 반면, 본보기모형은 화자 변산성 그 자체를 중요한 정보로 활용할 수 있다. 이 연구는 모음지각에 대한 정규화모형을 비판적으로 검토하고 본보기모형을 그 대안으로 제시하려는 것이었다. Peterson & Barney(1952)의 자료를 가지고 정규화모형과 본보기모형(GCM과 ALCOVE)을 정량적으로 대비시키는 시뮬레이션 연구를 수행하였다. 정규화모형과 본보기모형을 서로 대비시킨 시뮬레이션 연구에서 본보기모형은 모음식별과 화자성별인식에서 각각 90.13%와 93.3%의 인식률을 보인 반면, 정규화모형은 각각 87.2%와 88.7%의 인식률을 보였다. 이 결과는 본보기모형이 모음지각에서 화자 변산성을 보다 효과적으로 처리하는 모형일 가능성을 시사한다.

말소리(speech)의 기본단위인 음소는 자음과 모음으로 나누어진다. 자음은 모음과 결합해야만 그 음가를 발휘하는 데 비해 모음은 독자적인 말소리로 작용할 수 있다. 따라서 모음은 말소리의 핵이라 할 수 있겠다. 모음은 성대(vocal cords), 성도(vocal tract), 턱, 혀, 입술 등 조음기관의 작용에 따라 달라진다. 모음산출과정을 설명하는 발성원-여과기 이론(source-filter theory)에 따르면 음성의 산출은 폐로부터 밀려나오는 기류가 성대를 진동시켜 만든 성대파가 후두에서 입술 끝까지의 성도를 거쳐 지나오는 동안 성도의 특정 부위의 넓이나 길이에 의해 어떤 주파수 대역을 공명시켜 강화하거나 약화시키는 여과과정을 거쳐 이루어

진다(Fant, 1960; Lieberman & Blumstein, 1988; Müller, 1848).

그런데 실생활에서 접하는 발성된 모음들은 동일한 모음일지라도 굉장히 다양한 F1과 F2¹⁾ 값을 갖는다(그림 1 참조). 게다가 어떤 경우는 동일한 F1과 F2 값일지라도 다른 모음으로 지각되는 경우가 있다. 이것은 바로 화자의 성별, 나이, 방언, 발화속도, 어투 등과 같은 개인마다 서로 다른 특성 때문에 발생한다. 이렇게 동일하게 지각되는 모음도 물리적으로 다양한 값을 가질 수 있게 하는 특성을 화자 변산성(speaker variability)이라 한다.

모음의 음향특성은 성대의 진동과 여과기 역할을 하는 성도의 모양과 길이에 의해서 결

이 연구는 1998년도 부산대학교 학술연구조성비를 지원받아 수행된 연구임.

교신저자 주소: 신현정, 부산시 금정구 장전동 부산대학교 심리학과, 〒 609-735

(e-mail: hjshin1@hyowon.pusan.ac.kr; 최양규, ygchoi@choonhae.ac.kr)

1) F0, F1, F2, F3은 각각 기본주파수와 제 1포먼트, 제 2포먼트, 제 3포먼트의 주파수를 의미한다.

위인 음소(phoneme)간의 불변성(invariance)을 찾기가 어렵다는 것이다. 말소리 지각 이론들은 이 불변성 문제를 필수적으로 다루고 있다. 불변성 문제를 해결하는 방식에 따라 말소리 지각 이론은 크게 두 부류로 나누어진다. 하나는 운동이론으로서 말소리의 지각과 산출간의 연계를 전제하고 있으며, 말소리 지각은 말소리의 산출과정에서 나타나는 조음동작을 참조함으로써 이루어진다고 본다(Liberman & Mattingly, 1985; Miller, 1990). 또 다른 이론은 청각이론(auditory theory)으로서 말소리 지각 과정을 일반적인 청지각(auditory perception)과 동일한 과정으로 설명하려고 한다(Diehl & Kluender, 1989; Stevens & Blumstein, 1981).

더 구체적으로 말해서 운동이론은 말소리의 음향 속성 자체보다 뇌로부터 발성기관(예컨대 턱, 입술, 혀)에 내려지는 조음 운동명령(motor command)에서의 불변성을 찾고 있다. 동일한 음소일지라도 화자나 말의 빠르기, 동시조음(coarticulation)되는 주변의 음소에 따라 말소리 자체의 음향적인 속성은 달라지는 반면, 그 음소를 발음하기 위한 조음과정은 동일하다고 보기 때문이다. 운동이론에 따르면, 말소리 지각과정은 말소리를 조음동작(articulatory gesture)으로 변환하는 특성화된 음성모듈(phonetic module)을 통해서 이루어지며, 이 음성모듈은 말소리의 지각과 산출을 모두 중재하고 있다. 이러한 맥락에서 볼 때, 말 지각과 말 산출은 조음 운동체계(motor system)를 통해 서로 직접적으로 연계되어 있다고 할 수 있다. 즉, 말소리의 지각과 산출은 입력과 출력이라는 방향성에서는 서로 반대이지만, 조음과정 혹은 조음동작이라는 공통적인 과정을 통해서 진행된다는 것이다. 그러나 특성화된 음성모듈의 처리방식, 조음 운동명령 등이 구체화되고 정교화되지 않은 상태이기 때문에 운동이론에 의한 불변성 해결 과정은 아직까지 검증될 수 없는 가설적인 수준에 머물러있다. 반면에 청각이론은 조음동작과 상관없이 말소리 자체에 대한

음향적 처리를 통해서 말소리지각에서의 불변성을 찾으려 하였다.

모음지각과정에서 모음식별에 앞서 동일모음에 대한 음향특성의 절대값(예컨대, F0와 포만트 주파수)의 화자간 변산성을 감소시키는 가설적인 지각과정을 모음 정규화(vowel normalization) 또는 화자 정규화(speaker normalization)라고 한다. 청각이론과 운동이론은 모두 정규화 과정을 전제하고 있다고 볼 수 있는데 운동이론은 대뇌의 운동영역에서의 조음 운동명령에 의한 정규화를, 청각이론은 모음의 여러 가지 음향특성을 분석하고 처리하는 방식에 의한 정규화를 각각 제안하고 있다. 따라서 모음정규화는 모음지각에서의 불변성 혹은 항등성 문제와 관련하여 지금까지 모음지각에 관한 많은 연구들에서 중요하게 다루어져 왔다(Johnson, 1997).

정규화를 통한 모음지각

청각이론의 입장에서 기본주파수와 포만트 주파수에 근거하여 화자 변산성을 제거하거나 감소시키기 위한 여러 가지 정규화모형이 제안되었다. Gerstman(1968)은 화자가 발생할 수 있는 포만트 범위상의 상대적 위치로 포만트 값을 정규화하는 범위 정규화(range normalization)를 제안했다. 예컨대, 화자의 F1 범위가 300-700Hz일 때 특정 모음의 F1 값이 500Hz라면 범위의 중간 위치에 해당하므로 그 모음은 0.5의 값을 가진다. Joos(1948) 그리고 Potter & Steinberg(1950)는 포만트들간의 차이값을 이용하는 방법을 제안했다. 예컨대, 어떤 모음의 F2가 1500Hz이고 F1이 500Hz라면 그 차이값인 1000Hz라는 값으로 변환하는 것이다.

최근에는 기본주파수와 포만트 주파수들을 특정한 척도를 이용하여 변환시킨 후, 변환값들을 적절히 조합함으로써 화자 변산성을 제거하려는 모형들이 제기되어왔다. Syrdal & Gopal(1986) 그리고 Traumlüller(1981)는 포만트 값을 Bark 척도(임계대역 척도)로 변환한

후 그 값들 간의 차이를 이용하는 방법을 사용한다. Miller(1989)와 Nearey(1989)는 포만트의 log값간의 차이를 이용하여 정규화를 구현하였다. 이 모형들의 공통점은 첫째, 모음인식에 앞서 음성신호의 청각 표상이 어떤 식으로든 수정되는 과정을 가정하고 있다는 점 그리고 둘째, 변환된 포만트들간의 차이값을 이용하고 있다는 점이다. 그러나 이 모형들은 사용하고 있는 변환척도에서 큰 차이를 보이고 있다.

Syrdal & Gopal(1986)이 사용하고 있는 Bark 척도는 심리음향학적 연구들로부터 유도된 척도로서 주관적으로 동일한 크기로 들리는 대역폭을 말한다. 입계대역은 청각 시스템의 필터의 대역폭을 나타내는데, 1 입계대역(1 Bark)은 달팽이관 기저막의 어떤 지점에서든 1.3mm를 차지한다. 대역폭은 500Hz까지는 주파수와 비례하여 증가하지만 그 이상의 영역에서는 주파수의 대수값에 비례한다(Zwicker & Terhardt, 1980). 주파수를 Bark 척도로 변환하는 등식은 다음과 같다.

$$Z=13\arctan(0.76 \times F)+3.5\arctan(F/7.5)^2 \quad (1)$$

등식 (1)에서 Z는 Bark로 표시한 입계대역이고 F는 주파수이다. Bark는 저주파수 영역에서는 대수척도(logarithmic scale)와 다르다.

Syrdal & Gopal(1986)의 모형에 따르면 모음식별은 두 단계를 거쳐 진행된다. 첫 단계에서 포만트 값이 Bark 척도로 변환되고 다음 단계에서 세 차원 즉 B3-B2, B2-B1, B1-B0²⁾ 값에 따라 모음이 분류된다. B1-B0 차원은 모음의 고저성을 나타내며 3 Bark 이내이면 고모음으로, 3 Bark 이상이면 저모음으로 분류된다. B3-B2 차원은 모음의 전후성을 나타내며 3 Bark 이내이면 전모음으로, 3 Bark 이상 그들의 모형에 따라 판별분석하여 얻은 모음유목화(vowel classification) 혼동행렬이 Peterson

(1986)은 Peterson & Barney(1952)의 자료들과 Barney(1952)가 수집한 모음들에 대한 사람들의 청취결과와 큰 차이가 없음을 보였다.

반면에 log 척도를 사용하는 Miller(1989)의 모형에 따르면 모음은 다음과 같은 3차원의 청지각 공간에 배열된다.

$$x=\log(F3/F2) \text{ 또는 } \log(F3)-\log(F2) \quad (2)$$

$$y=\log(F2/F1) \text{ 또는 } \log(F2)-\log(F1) \quad (3)$$

$$z=\log(F1/SR) \text{ 또는 } \log(F1)-\log(SR) \quad (4)$$

여기서 SR은 감각 참조점(sensory reference)으로써, 화자의 F0를 다음과 같이 등식 (5)에 의해 변환한 것이다.

$$SR=168 \times (F0/168)^{1/3} \quad (5)$$

동일모음의 경우 이 3차원상의 값은 화자간에 일정한 값을 갖는다고 가정하고 있기 때문에 다른 화자에 의해 발음된 동일 모음은 이 3차원 공간에서 서로 가깝게 모여 있게 된다.

정규화의 문제점과 대안

이런 식으로 정규화를 통해서 화자 변산성 문제를 해결하고자 한 시도들은 말소리의 언어적 내용을 음소와 같은 추상화된 언어적 단위로 부호화하는 이상적인 범주를 찾는 데 강조점을 두어온 전통적인 분석적 접근(analytic approach)의 영향 때문이라 하겠다. 이러한 접근은 청자를 둘러싼 환경자극들의 변산성을 무시한 채 물리적 속성 자체에서 어떤 불변성(invariance)을 찾으려 한다. 이런 시각에서 화자 변산성은 음향속성에서 제거되어야 할 소음으로 취급되어 왔다. 정규화는 다양한 입력에서 동일한 형태를 생성하는 것, 즉 물리적으로 다른 자극을 동일한 표상형태로 변환하는 것을 의미하며, 이에 따른 자극의 변산성 감소

2) B0, B1, B2, B3은 각각 F0, F1, F2, F3을 Bark 척도로 변환한 값을 말한다.

로 인한 정보의 손실이 수반된다. 따라서 자극의 변산성은 음성 신호에서 지각에 방해를 주는 소음에 불과한 것으로 여겨졌다. 그러나 반

드시 그런 것은 아니다. 여러 연구들은 화자의 목소리에 관한 매우 세밀한 부분까지 정보가 부호화되고 저장되어 음성인식에 영향을 주고

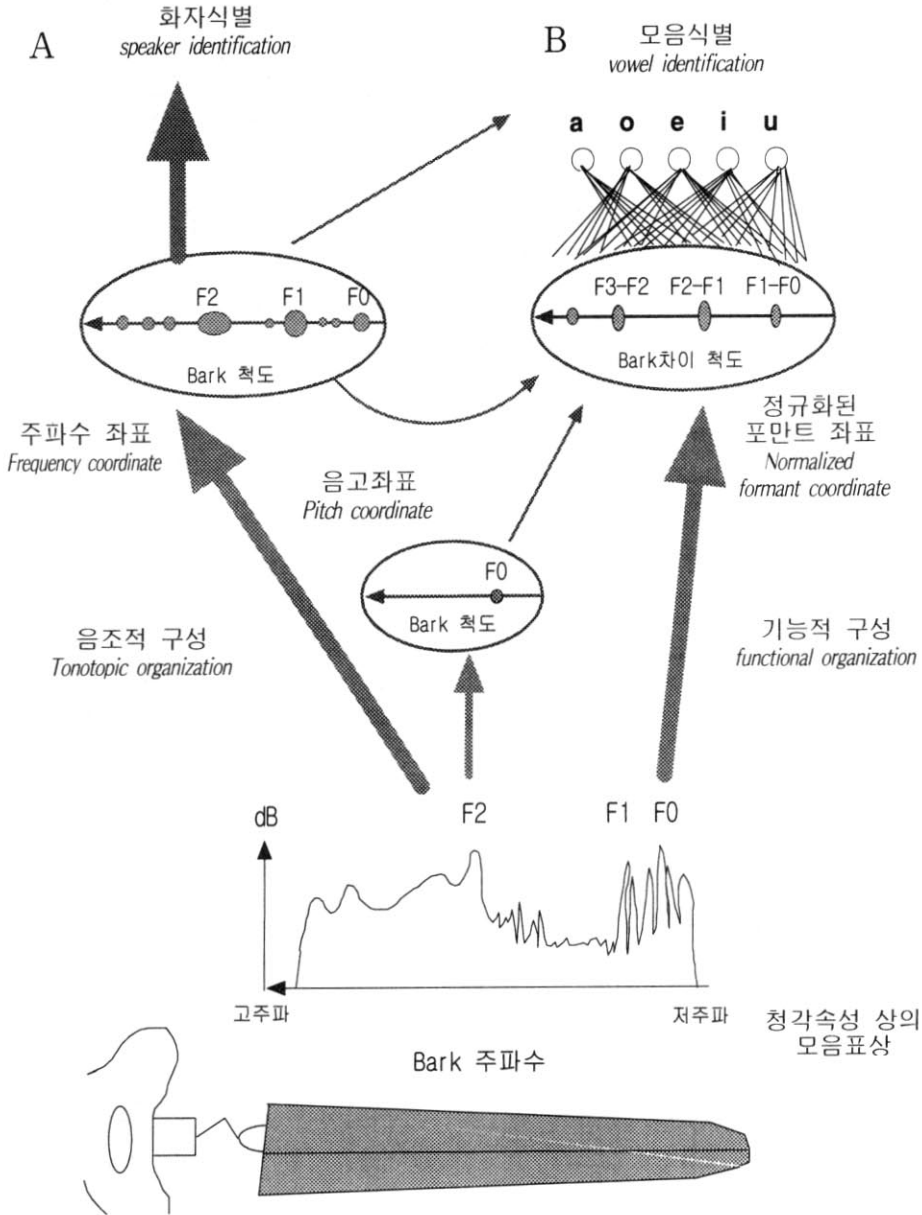


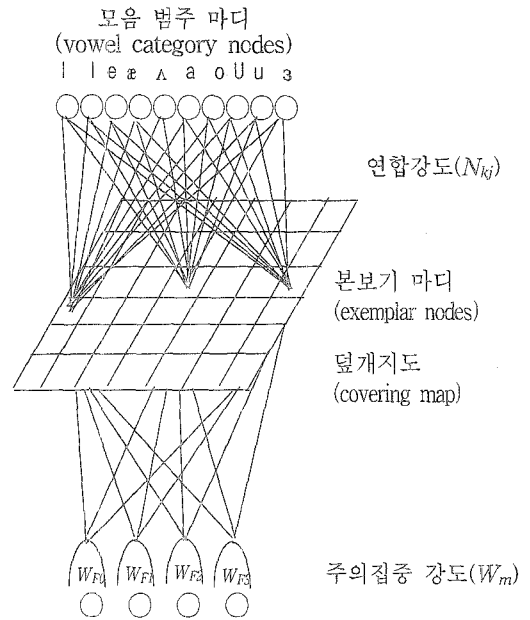
그림 2. 화자식별(A)과 모음식별(B)에 관한 정보가 동시에 병렬적으로 처리되는 모형의 한 예. Bark척도로 변환된 주파수의 음조적 구성(tonotopical organization)은 화자 식별을 위한 단서가 되고, Bark 척도로 변환된 포먼트 주파수들 간의 관계정보, 즉 차이값은 모음식별을 위한 단서가 된다(Hirahara & Kato, 1992).

있음을 보여 준다(Goldinger, 1992; Goldinger, Pisoni, & Logan, 1991; Nygaard, Sommers, & Pisoni, 1994, 1995; Palmeri, Goldinger, & Pisoni, 1993; Sommers, Nygaard, & Pisoni, 1994). 다시 말해서 성별, 방언, 말속도 등 화자의 음성에 관해 학습한 세부적인 내용들은 지각의 결과로 없어지지 않고 여러 가지 지각 차원에서 다양한 방식으로 부호화되고, 장기기억에 저장되어 오히려 음성인식과 회상을 촉진시킨다(Pisoni, 1997). 즉, 화자 변산성이 지각 과정에서 정보를 풍부하게 지니고 있다는 말이다.

사람은 음성의 세부적인 부분을 잘 파지하고 또한 화자의 성별, 방언, 말속도, 정서, 말투 등 화자의 목소리에 관한 많은 정보를 부호화하는 것으로 보인다. 따라서 모음항등성은 변산성을 감소시켜 자극에 관한 세부적인 정보 손실을 가져오는 처리과정, 즉 정규화를 통해 이루어진다고 보는 전통적인 정규화모형은 사람의 모음지각을 설명하는 데는 생태학적 타당도(ecological validity)가 떨어진다고 볼 수 있다.

그림 2는 말소리에 나타나는 화자-특정적 정보(speaker-specific information)의 처리와 언어적 처리가 병렬적으로 진행됨을 보여주는데, 포먼트 주파수값은 화자 식별(A) 정보로 이용되고 포먼트들간의 상대적인 차이값은 모음 식별(B)을 위한 정보로 사용된다. 두 정보는 동일한 음성신호로 동시에 전달되고 처리된다(Hirahara & Kato, 1992). 따라서 화자식별을 위한 화자-특정적 정보가 모음식별을 위해 사용될 수 있다. 다시 말해서 화자-특정적 정보 혹은 화자 변산성이 모음식별을 위한 단서가 될 수 있다는 말이다.

모음식별도 일종의 범주화과정이라고 볼 때, 화자-특정적 정보 또는 화자 변산성 처리의 문제는 범주화과정을 설명하기 위해 제안된 모형의 하나인 본보기모형(exemplar model)에서 해결의 실마리를 찾아볼 수 있다. 본보기모



자극의 청각속성 차원 마디 (stimulus dimension nodes)

그림 3. 연결망구조의 본보기모형(ALCOVE)에 의한 모음지각. 각 청각속성 마디는 모든 본보기 마디들에 연결되어 있고, 각 본보기 마디는 범주 마디들에 연결되어 있다(성별식별의 경우에 범주 마디들은 남자, 여자, 아동 등의 각 성별범주에 해당한다).

형을 모음지각에 적용해본다면 기억에 저장된 본보기가 모음정보뿐만 아니라 화자-특정적 정보를 함께 지니고 있으며, 모음의 범주화는 기억 속에 표상된 본보기와의 유사성에 근거한다고 볼 수 있다. 따라서 입력되는 모음자극에 대한 지각은 화자간의 차이를 제거하기 보다 기억 속에 표상된 기존의 본보기들과의 유사성 평가에 화자 변산성 정보를 이용하는 모음지각과정을 제안할 수 있다.

정규화를 가정하지 않는 모음지각: 본보기 모형

본보기모형은 말소리 지각 현상보다는 일반

적인 범주화과정을 설명하기 위해 제안된 모형이지만 모음지각을 범주화의 일종으로 본다면 모음지각현상을 설명하는 데에도 적용할 수 있다. 본보기모형을 모음지각에 적용할 때, 앞서 기술한 정규화과정을 상정하지 않고 즉, 화자변산성 정보를 제거하는 것이 아니라 오히려 효과적으로 이용하여 화자 변산성을 처리하는 모음지각 모형이 가능해진다.

본보기모형(Estes, 1993; Hintzman, 1986; Kruschke, 1992; Nosofsky, 1986, 1988, 1991; Nosofsky, Kruschke, & Mckinley, 1992)에서는 범주가 경험한 본보기들의 집합으로 표상된다고 가정한다. 즉, 어떤 추상적인 범주 원형을 가정하지 않는다. 범주화될 사례와 각 범주의 기억된 본보기들과의 비교를 통해 범주화 과정이 진행되며, 범주화 결정은 각 범주 본보기들과의 유사성의 총화에 근거한다는 것이다.

지금까지 범주화과정을 설명하기 위한 여러 가지 본보기모형들이 제안되어 왔다. Nosofsky (1984, 1986, 1987)는 대표적인 본보기모형의 하나인 Medin & Schaffer(1978)의 맥락모형(context model)을 확장시킨 일반화맥락모형(Generalized Context Model: 이후 GCM)을 제안하고 있다. GCM은 다차원척도법(MDS)에 근거한 전형적인 본보기모형의 하나이다. 최근에는 Kruschke(1992)가 GCM의 본보기 표상 방식과 Gluck & Bower(1988a; 1988b)가 제안한 연결망모형의 오류-주도적 학습방식(error-driven learning)을 결합하여, 범주학습과정을 보다 잘 설명할 수 있는 연결망구조의 본보기모형인 ALCOVE(Attention Learning COVERing map)를 제안하였다. 그림 3은 ALCOVE를 모음 범주화과정에 적용한 예를 보여주고 있다. 그림 3에서 보듯이 본보기는 청각속성(auditory property)과 범주를 연결짓는 부분이다. 청각속성은 말초 청각시스템(peripheral auditory system)

의 출력이고, 범주는 본보기가 저장될 때 붙여지는 언어적 명칭, 화자의 성별, 화자의 이름과 같은 것들이다. 범주화될 자극이 주어지면 그 자극의 청각속성이 각 본보기들의 청각속성과 비교되는데 자극과 본보기들간의 유사성에 따라 본보기의 활성화 수준이 좌우된다. 유사성이 크면 본보기의 활성화 수준이 높아져서 그 본보기가 속한 범주, 즉 그 본보기와 연합된 범주마디가 활성화된다. 특정 범주에 속하는 모든 본보기들의 활성화 총량은 자극이 그 범주의 사례로 범주화될 증거로 받아들여진다. 이것은 모든 종류의 범주에 다 적용될 수 있으므로 이 모형은 모음인식과 화자인식에 모두 적용될 수 있다.

GCM과 ALCOVE³⁾에서는 다차원 공간상에 위치한 사례들간의 심리적 거리를 유사성의 지표로 채택한다. 각 본보기들은 다차원 공간에서 하나의 점으로 표상되며, 유사성은 거리의 단조감소함수라고 가정한다. 심리적 다차원 공간상에서 사례간의 거리가 가까울수록 사례간의 유사도는 커지게 된다. 두 점간의 거리(d_{ij})는 등식 (6)과 같이 가중치를 준 Minkowski의 멱함수(power function)를 사용하여 계산된다 (Carroll & Wish, 1974).

$$d_{ij} = \left[\sum_{m=1}^n W_m |X_{im} - X_{jm}|^r \right]^{1/r} \quad (6)$$

여기서 X_{im} 은 청각차원 m 에서 사례 i 가 갖는 심리적 값이며, $W_m(0 \leq W_m \leq 1, \sum W_m = 1)$ 은 차원 m 에 주어지는 선택적 주의집중의 가중모수치(주의집중강도)이다. 모든 가중치들의 합이 일정하다고 가정하는 것($\sum W_m = 1$)은 제한된 처리용량체계의 가정을 반영한다. r 의 값은 사례를 구성하는 차원의 유형에 달려있다. 사례간의 유사성(S_{ij})은 등식 (7)의 방식으로 계산된다.

3) ALCOVE는 오류-주도적 학습과정을 제외하고는 GCM과 동일하다고 할 수 있다.

$$S_{ij} = \exp(-cd_{ij}^p) \quad (7)$$

여기서 $p = 1$ 인 지수감소함수는 변별이 용이한 사례들을 유목화하는 상황에 적합한 반면, $p = 2$ 인 가우스함수는 매우 유사한 사례들을 집중적으로 변별학습하는 지각적 변별 상황에서 적합한 것으로 보인다(Nosofsky, 1985, 1986, 1987; Shepard, 1986, 1987). c 는 심리적 공간에서의 전반적인 변별력을 반영하는 민감도 모수치이다. 자극 i 가 기억 속의 본보기 j 를 활성화시키는 정도(a_{ij})는 등식 (8)로 계산될 수 있다.

$$a_{ij} = S_{ij} + e_j \quad (8)$$

여기서 e_j 는 오차(error)로서 평균 0을 중심으로 정규분포를 이루고 있다. 자극 i 가 모음범주 C_k 에 속할 증거(evidence)는 범주 C_k 의 본보기 j 들의 활성화 총화로 결정되며 이는 곧 범주마다 k 가 활성화되는 정도($E_{k,i}$)이다. 그것을 등식으로 나타내면 다음과 같다.

$$E_{k,i} = \sum_j N_{jk} a_{ij}, j \in C_k \quad (9)$$

여기서 N_{jk} 는 각 본보기 j 와 그 본보기가 속한 범주 k 간의 연합강도(association strength)를 말한다.⁴⁾ 연합강도가 범주마다 다르면 범주화에 반응편파가 발생할 수 있다. 주의집중 가중치 W_m 에 따라 어떤 자극차원 상의 변동을 무시할 수도 있다. 이 주의집중 가중치를 다르게 함으로써 화자 정규화효과를 낼 수도 있다. 민감도 모수치 c 는 다차원 공간상에서 멀리 떨어져 있는 본보기의 영향을 제한하는 기능을 한다. 다차원 공간상의 거리를 유사성으로 변환하는 함수가 지수함수이므로 멀리 떨어진

본보기의 영향력은 거의 없어진다. 본보기모형에서 화자를 구분해주는 청각속성들은 본보기에 저장되어 있기 때문에 범주화되어야 할 자극과 가장 유사한 본보기는 같은 화자나 소리가 비슷한 화자에 의해 발화된 본보기들이다.

시뮬레이션 연구: 정규화모형 대 본보기모형

시뮬레이션 연구에서는 첫째, Miller(1989) 그리고 Syrdal & Gopal(1986)이 제안한 정규화모형과 본보기모형간의 모음식별에서의 성과를 비교하여 보았다. 둘째, 사람들의 수행결과에 대한 두 모형의 적합도를 비교하였다. 부가적으로 성별식별에서의 성과도 서로 비교해 보았다.

모형검증에 사용된 D/B. Peterson & Barney (1952)의 자료를 이용하여 모음지각에서 정규화에 관한 많은 연구들이 행해졌다(Disner, 1980; Fourakis & Miller, 1987; Miller, 1989; Nearey, 1989; Syrdal & Gopal, 1986; Watrous, 1990, 1991). 이 시뮬레이션 연구에서는 Watrous(1991)가 이용했던 것과 동일한 Peterson & Barney (1952)의 자료를 인터넷 사이트에서 가져와서 이용하였다.⁵⁾ 이 자료에는 남자 33명, 여자 28명, 아동 15명 등, 모두 76명이 모음 10개⁶⁾를 2회씩 발음한 총 1,520개 모음발음의 F0, F1, F2, F3 자료와 화자의 성별자료가 포함되어 있다. 1,520개의 모음 자료를 첫 회에 발음한 것과 두 번째 발음한 것으로 각각 760개씩 반으로 나누고 받은 학습단계에, 나머지 받은 성능평가단계에 사용하였다.

4) ALCOVE에서는 연합강도와 주의집중강도가 오류-주도적 학습과정을 통해서 학습된다.

5) www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/speech/database/pb/pb.tgz

6) [i], [I], [e], [æ], [ʌ], [a], [o], [U], [u], [ɜ].

정규화에 의한 모음 판별분석

F0와 포만트 값 그리고 포만트들간의 차이 값을 주파수로 변환하거나 혹은 Syrdal & Gopal(1986) 또는 Miller(1989)가 했던 것처럼 Bark척도나 log척도로 변환한 값을 사용하여 Peterson & Barney(1952)의 자료를 10개의 모음으로 분류하는 판별분석을 실시하였다. 자료의 첫 번째 받은 판별식을 얻는 데 사용되었고, 나머지 받은 모형의 분류성능을 평가하는 데 사용되었다. 사례수가 같으므로 각 모음집단의 사전확률은 동일하게 하였다. 판별분석에는 SPSS 7.5가 사용되었다. 그 결과가 표 1에 나와 있다. 표 1의 결과를 보면 Bark척도나 log척도로 변환하여 정규화했을 때의 성과가 정규화하지 않은 주파수를 이용한 경우보다 우수한 것으로 나타났다. 이것은 척도변환이 정규화의 효과를 가져오기 때문으로 볼 수 있겠다. 그리고 Bark나 log값으로 변환한 경우 포만트들간의 차이값을 이용했을 때가 포만트 값들을 그대로 이용했을 때보다 좀더 우수한 성과를 보였다. 정규화하지 않은 포만트 주파수값을 이용한 경우에는 오히려 포만트들간의 차이값을 이용했을 때 수행이 더 저조했다. 이것은 포만트들간의 차이값은 척도변환한 것이어야 정규화 효과를 발휘한다는 것을 말해준다. 그리고 log척도를 사용하는 것이 Bark 척도를 사용하는 것 보다 약간 우수한 것으로 나타나고 있는데(1~1.3% 차이) 이것은 log척도가 Bark척도보다 우수한 정규화 효과를 지

니는 것으로 해석할 수 있겠다.

정규화에 의한 화자 성별 판별분석

검증에 사용된 D/B는 남자 33명, 여자 28명, 아동 15명에게서 얻은 자료이며, 각 자극에 대한 성별자료가 포함되어있는데 모음 대신 성별이라는 것 외에 유목화(classification)하는 면에서는 동일한 상황이므로 모음식별과 같은 방법으로 화자의 성별을 분류하기 위한 판별분석을 부가적으로 실시하였다. 과제는 각 자극을 남성, 여성, 아동 중 하나로 분류하거나 범주화하는 것이다. 그 결과가 표 1에 제시되었다. 성별인식에서는 포만트들간의 차이값을 사용한 경우를 제외하면 모두 비슷한 수행을 보였다. 포만트들간의 차이값을 사용한 경우, 모음식별의 경우와는 달리 수행이 저조했다(Bark척도와 log척도의 경우 42.1%와 38.6%로서 우연수준인 33.3%보다 약간 상회하는 정도이다). 이 결과는 그림 2에서 볼 수 있는 것처럼 모음식별에는 포만트들간의 상대적인 차이값이 사용되고, 화자식별에는 포만트 주파수값이 사용된다는 점을 고려해 본다면 당연한 결과일 것이다. 포만트값 그대로 사용하는 것보다 포만트들간의 상대적인 차이값을 사용하는 것이 모음식별에는 효과적이거나 화자 성별식별에는 덜 효과적인 것은 그만큼 화자 변산성을 감소시켰기 때문일 것이다. 정규화 정도를 평가하기 위해서 Disner(1980)가 사용한 기준은 화자간의 차이를 최소화시킨 정도와 모음범주

표 1. 주파수, Bark, log척도로 정규화한 모형들의 모음 식별율과 성별 식별율.

척도유형	사용된 모수치	모음 식별율	성별 식별율
주파수값	F0, F1, F2, F3	82.1%	89.1%
	F1-F0, F2-F1, F3-F2	79.9%	64.9%
Bark 척도 (Syrdal & Gopal, 1986)	B0, B1, B2, B3	85.8%	87.5%
	B1-B0, B2-B1, B3-B2	86.2%	42.1%
log 척도 (Miller, 1989)	L0, L1, L2, L3	87.1%	88.7%
	L1-L0, L2-L1, L3-L2	87.2%	38.6%

간의 차이를 극대화시킨 정도의 비(比)이다. 이런 기준에서 볼 때 포맷트들간의 상대적인 차이값을 사용하는 것이 포맷트값을 그대로 사용하는 것보다 정규화 효과가 크다고 하겠다.

본보기모형에 의한 모음식별과 화자 성별 식별

정규화모형은 자동음성인식 시스템 구현을 위해 인식에 방해가 되는 화자간 변산성을 제거하고 있지만, 사람의 경우 화자-특정적 정보를 말소리 지각에 사용한다는 많은 증거들이 여러 연구들에서 제시되었다(Pisoni, 1997 참조). 정규화를 가정하지 않고 사람의 경우처럼 화자 변산성 그 자체를 모음인식을 위한 유용한 정보로 간주하는 모형을 찾는다면 GCM과 ALCOVE와 같은 본보기모형을 들 수 있다.

이 연구에서는 등식 (6), (7), (8), (9)에 근거한 본보기모형으로 모음식별과 화자성별 식별 성능을 평가하였다. 등식 (6)에서 심리적 거리를 측정하는 데에 분리가능한 차원(separable dimension)인 경우에는 도시구획척도($r=1$)가 사용되거나 통합된 차원(integral dimension)인 경우에는 가중치가 있는 유클리드 거리가 흔히 사용된다(Nosofsky, 1987; Shepard, 1964). 따라서 청각 자극의 심리적 거리를 측정하는 데는 유클리드 척도($r=2$)를 사용하여 심리적 거리를 측정하였다. 각 모음자극은 F0, F1, F2, F3의 주파수값에 따라 4개의 차원 값을 가지는데, F0, F1, F2, F3 차원에 대한 주의집중 가중치는 각각 W_{F0} , W_{F1} , W_{F2} , W_{F3} 이다. 등식 (7)에서 변별이 용이한 사례들을 유목화하는 상황에서는 $p = 1$ 인 지수감소함수가 적합하기 때문에 이 실험에서는 지수감소함수($p=1$)로 유사성을 도출하였다. 등식 (8)에서 오차

(e_j)를 0으로 고정시켰다. 등식 (9)에서 각 본보기와 본보기가 속한 범주간의 연합강도(N_{jk})는 1로, 속하지 않은 다른 범주와는 0으로 고정시켰다.⁷⁾ 등식 (9)에 의한 증거가 가장 큰 범주로 자극을 유목화하였다.

평가를 위한 자료는 앞서 정규화모형들을 검증하는 데 사용된 자료와 동일하였다. 총 1,520개의 모음으로 구성된 Peterson & Barney (1952)의 자료를 760개씩 두 세트로 나누어 첫 세트는 학습에, 두 번째 세트는 성능평가에 사용하였다. 학습시행과 평가시행은 각각 다음과 같다.

모음식별을 위한 학습시행에서는 첫 세트의 모음자료 760개중에서 한 개씩 추출하여 유목화할 자극으로 사용하고 나머지 759개 자료를 본보기로 사용하였다. 이런 식으로 모음자료 760개를 10개의 모음 중 하나로 유목화하는데 최대의 수행을 보이는 자유모수치(free parameter)를 구하였다. 이렇게 구한 자유모수치의 값들이 표 2에 나와있다. 평가시행에서는 첫 세트 760개 모음자료를 본보기로 사용하고 두 번째 세트의 모음자료들을 유목화해야 할 자극으로 사용하였다. 학습시행에서 구한 자유모수치를 적용하여 수행정도를 평가하였다.

성별식별을 위한 학습시행과 평가시행은 세 개의 성별로 유목화했다는 것을 제외한 나머지는 동일한 절차로 수행되었다. 그 결과, 모음식별과 화자성별식별에서 정확율은 각각 90.13%, 93.3%로서 전체적인 수행정도는 앞서 논의한 정규화모형보다 우수함을 보여주었다.⁸⁾

표 2에서 본보기모형의 F0, F1, F2, F3에 대한 가중치를 살펴보면 모음식별의 경우 F0, F1, F2는 각각 .377, .322, .215로서 서로 비슷한 값이나 F3에 대한 가중치는 .085로서 다른 것보다 비중이 낮다. 이것은 그림 1에서 볼 수

7) 이런 제약 하에서 ALCOVE와 GCM은 동일하다.

8) ALCOVE의 경우, 오류-주도적 학습시행을 통해 학습시켰을 때, 학습자극에 대해서는 시행에 따라 수행이 증가하여 최고치에 달했지만, 학습시행에 따른 새로운 자극에 대한 범주화의 수행정도는 증가하다가 다시 점차 감소하는 포물선 모양을 그렸다. 그러나 새로운 자극에 대한 최대수행은 GCM과 동일하였다.

표 2. 본보기모형을 적용한 모음식별과 성별 식별 시뮬레이션 실험의 모수치와 정확식별율

모수치	모음식별	성별식별
c	.087	.097
W_{F0}	.377	.788
W_{F1}	.322	.034
W_{F2}	.215	.047
W_{F3}	.085	.132
식별율	90.13%	93.3%

c는 민감도 모수치이고 W_{F0} , W_{F1} , W_{F2} , W_{F3} 는 각각 F0, F1, F2, F3 차원에 주어지는 주의집중 가중치이다.

표 3. 사람들의 수행(Peterson & Barney, 1952)에 대한 정규화모형과 본보기모형의 적합도

모형	SSE	설명변량 (%)
정규화모형		
F0, F1, F2, F3	.349	94.19
F1-F0, F2-F1, F3-F2	.222	96.60
자유 모수치 B0, B1, B2, B3	.151	97.74
B1-B0, B2-B1, B3-B2	.211	96.75
L0, L1, L2, L3	.219	96.68
L1-L0, L2-L1, L3-L2	.149	97.78
본보기모형		
자유 모수치 c, W_{F0} , W_{F1} , W_{F2} , W_{F3}	.100	98.61

SSE는 사람의 수행결과에 나타난 관찰확률과 모형의 예언확률간의 편차 자승의 합이다.

있듯이 모음이 주로 F1, F2 차원에서 구분될 수 있고, 그 반면 F3은 F1과 F2에 비해 모음간의 차이가 비교적 적다. 따라서 F1과 F2가

모음지각에 중요하게 사용됨을 시사하고 있는데, 이 결과는 모음의 경우 두개의 지각적 차원이 F1과 F2에 상응함을 보인 최양규와 신현정(1998)의 연구결과와도 일치한다. 그리고 F0는 화자의 성별에 의해 크게 좌우되며 F1과 F2에도 상당한 영향을 준다. 다시 말해서 다양한 화자로부터 발생된 모음에 대한 지각은 F0, F1 그리고 F2에 의해 주로 결정됨을 시사받을 수 있다. 성별식별의 경우 4개의 모수치 중 F0에 대한 가중치가 .788로 가장 큰 값을 가지는 것은 F0가 성별에 따라 가장 크게 좌우되므로 성별식별에 가장 큰 단서가 된다는 선행 연구결과와도 일치한다(Johnson, 1990).

정규화모형과 본보기모형의 적합도 분석

사람들의 식별반응(Peterson & Barney, 1952)에 대한 정규화모형과 본보기모형의 예언력을 비교하기 위해 적합도 분석을 하였다. 그러기 위하여 사람들의 수행과 각 모형의 수행결과로 나타난 혼동행렬(부록참조)에서 관찰확률과 예언확률간의 편차 자승의 합(SSE)을 지표로 모형의 적합도를 평가하였다. 그 결과가 표 3이다. 표 3을 보면 본보기모형이 가장 적은 SSE를 보이고 있으며 사람들의 반응에 대한 설명력도 98.61%로 정규화모형에 비해 높음을 알 수 있다⁹⁾. 또한 부록에 제시된 혼동행렬표를 보면 사람들의 결과(Peterson & Barney, 1952)와 본보기모형의 반응패턴이 유사함을 살펴볼 수 있다.

논 의

이 연구는 화자 변산성을 제거하는 전통적

9) 표 3의 결과를 해석할 때, 유의해야 할 사항이 있다. 본보기모형의 설명력이 여러 정규화모형들보다 우수한 것으로 보이지만, 직접적인 비교를 할 수 있는 것은 아니다. 정규화모형들은 3개 또는 4개의 자유모수치를 사용하는 반면, 본보기모형은 5개를 사용하고 있기 때문이다. 따라서 본보기모형의 설명력이 우수한 이유가 모형 자체에 있는 것이 아니라, 보다 많이 상정한 자유모수치에 의한 것일 가능성도 있다.

인 정규화모형보다는 화자 변산성 정보를 활용할 수 있는 범주화모형의 하나인 본보기모형을 모음지각에 적용함으로써 불변성 또는 항등성 문제해결의 최대 관건인 화자 변산성을 효과적으로 처리하는 방안을 제시하고자 하였다. 그러기 위해 첫째로, Peterson & Barney(1952)의 자료를 D/B로 이용하여 두 모형의 모음식별 성과를 비교하였다. 그 결과 본보기모형이 90.13%의 수행을 보여서 정규화모형(85.8~87.2%)보다 우수하였다. Peterson & Barney(1952)의 연구에서 사람들의 모음식별 정확반응율이 94.4%이었음을 고려한다면 본보기모형의 수행결과가 사람의 수행과 큰 차이가 나지 않음을 알 수 있다. 둘째, 사람들의 수행결과에 대한 정규화모형과 본보기모형의 적합도를 비교한 결과에서도 본보기모형이 정규화모형보다 높은 적합도를 보였다.

이러한 결과는 GCM이나 ALCOVE같이 MDS에 기반을 둔 본보기모형이 화자 변산성을 효과적으로 처리하는 모형일 가능성을 시사해 준다. 다시 말해서 정규화를 통해서 화자 변산성을 제거하지 않고서도 모음의 지각과정이 설명될 수 있음을 시사받을 수 있다. 따라서 본보기모형의 계산과정이 기존의 자동인식시스템 개발을 위해 제안된 정규화모형보다 더 효과적인 모음인식 알고리즘이라고 평가할 수 있다. 아울러 본보기모형에서 화자 변산성을 모음지각에서 중요한 정보로 사용한다는 점은 여러 연구들에서의 사람의 음성인식 수행을 보다 잘 설명할 가능성이 있다. 다시 말해서 이 모형은 정규화모형보다 생태학적 타당도가 높은 모형이라 할 수 있다. 마지막으로 이 연구의 음성인식에 관한 공학적인 함의를 제언해보면 다음과 같다.

본보기모형: 효과적인 모음인식을 위한 알고리즘

이 시뮬레이션 연구는 본보기모형이 화자

변산성을 제거하는 정규화 과정을 거치지 않고서도 화자 변산성을 효과적으로 처리할 수 있음을 보여주었다. 다시 말해서 본보기모형은 모음지각에서의 항등성 또는 불변성 문제에 대한 적절한 해답을 제공해 준다.

따라서 앞으로 본보기모형은 말지각을 효과적으로 설명하는 경쟁력있는 모형으로 등장할 가능성이 있는 것으로 여겨진다. 그러기에 앞서 이 모형은 해결해야 하는 한가지 커다란 문제점을 가지고 있다. 본보기모형이 지니고 있는 가장 큰 문제는 경형에 의해 끊임없이 접하는 수많은 본보기들을 어떻게 저장하느냐 하는 문제이다. 이 문제는 개개의 본보기들 중 일부만 저장하는 방법을 통해 해결될 수 있다. 그런 해결 가능성을 연결주의적 본보기모형인 ALCOVE에서 찾아 볼 수 있다. 이 모형에서는 본보기들 대신에 덮개지도(covering map)를 상징하고 있다. 덮개지도란 본보기를 표상하는 심리적인 다차원공간상에서 임의의 제한된 수의 본보기 마디(node)들이 특정 밀도를 유지하면서 무선적으로 퍼져있는 공간을 말하는데, 그림 3은 본보기 공간을 표상하고 있는 덮개지도를 예시하고 있다. 이 덮개지도에서 각 마디들은 청각 속성들의 조합에 상응한다. 각 청각 속성마다 주의집중 가중치(attention weight)가 하나씩 있어서 각 청각 속성을 덮개지도에 대응시키는 것을 관장한다. 덮개지도의 각 마디들과 각 범주간에도 연합 가중치(association weight)가 하나씩 있어서 덮개지도상의 각 본보기 마디들을 범주마디에 대응시키는 것을 관장한다. 청각 속성과 범주가 많을수록 덮개지도는 보다 복잡해지고 연합 가중치의 수는 커지더라도 일반적인 본보기모형과는 달리 덮개지도의 크기는 제한되어 있다. ALCOVE 모형에서 주의집중 가중치와 연합가중치는 정확 범주화시행에서 주어지는 피드백을 통해 학습된다. 따라서 이 모형에서 입력되는 본보기들은 그 자체의 저장에 아니라 가중치 수정의 형태로 입력된다. 이렇게 ALCOVE

같은 연결주의적 형태를 통해서 본보기모형은 모음지각뿐 아니라 새로운 모음의 획득과정도 설명 가능한 모형이 될 수 있을 것이다. 경험에 의한 수많은 본보기들을 처리할 수 있도록 그리고 경험과 학습의 효과가 반영되도록, 다시 말해서 새로운 언어범주를 학습하여 획득하거나 기억 속에 표상된 기존의 본보기의 재범주화가 가능하도록 ALCOVE 같은 연결주의적 본보기모형을 수정 보완하여 모음지각에 적용한다면 모음학습도 가능한 훌륭한 모음인식의 알고리즘으로 될 가능성이 있다고 하겠다.

참 고 문 헌

- 최양규, 신현정 (1998). 우리말 단모음의 지각적 차원. 한국실험및인지심리학회 1998년도 여름연구회 발표논문집, 59-68.
- Carrol, J. D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*(vol. 2). San Francisco: W. H. Freeman.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecology of Psychology*, 1, 1-45.
- Disner, S. F. (1980). Evaluation of vowel normalization procedures. *Journal of Acoustical Society of America*, 67, 253-261.
- Estes, W. K. (1993). Concepts, categories, and psychological science. *Psychological Science*, 4, 143-153.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Fourakis, M., & Miller, J. D. (1987). Measurements of vowels in isolation and sentence context. *Journal of the Acoustical Society of America, Supplement 1*, 81, S17(A).
- Gerstman, L. J. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics*, 16, 78-80.
- Gluck, M. A., & Bower, G. H. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, 27, 166-195.
- Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247.
- Goldinger, S. D. (1992). *Words and Voices: Implicit and explicit memory for spoken words*. Research on Speech Perception(Tech. Rep. No.7). Bloomington: Indiana University.
- Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the locus of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 152-162.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Hirahara, T., & Kato, H. (1992). The effect of F0 on vowel identification. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure*, 89-112. Tokyo: Ohmsha Publishing.
- Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America*, 88, 642-654.
- Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson, & J. Mullennix (Eds.), *Talker variability in speech processing*. San Diego: Academic Press.
- Joos, M. (1948). Acoustic phonetics. In *Language and Monograph No. 23*. Linguistic Society of America. Baltimore, MD: Waverly.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech revised. *Cognition*, 21,

- 1-36.
- Lieberman, P., & Blumstein, S. E. (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics*. New York: Cambridge University Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.
- Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, *25*, 114-121.
- Miller, J. L. (1990). Speech perception. In D. N. Osherson, & H. Lasnik (Eds.), *Language: An invitation to cognitive science (vol. 1)*. London: MIT Press.
- Müller, J. (1848). *The physiology of the sense, voice, and muscular motion with the mental faculties* (Trans. by W. Baly). London: Walton and Maberly.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, *85*, 2088-2113.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104-114.
- Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A Choice model analysis. *Perception & Psychophysics*, *33*, 415-432.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 87-108.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700-708.
- Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, *23*, 94-140.
- Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 211-233.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42-46.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, *57*, 989-1001.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1-20.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of Acoustical Society of America*, *24*, 175-184.
- Pisoni, D. B. (1997). Some thoughts on "Normalization" in speech perception. In K. Johnson, & J. Mullennix (Eds.), *Talker variability in speech processing*. San Diego: Academic Press.
- Potter, R., & Steinberg, J. (1950). Toward the specification of speech. *Journal of the Acoustical Society of America*, *22*, 807-820.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54-87.
- Shepard, R. N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, *115*, 58-61.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*,

- 237, 1317-1323.
- Sommers, K. N., Nygaard, L. C. & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of Acoustical Society of America*, 96, 1314-1324.
- Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas, & J. L. Miller (Eds.), *Perspectives on the study of speech*. NJ: Erlbaum, Hillsdale.
- Syrdal A., & Gopal H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79, 1086-1100.
- Traumüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69, 1465-1475.
- Watrous, R. L. (1990). Speaker normalization using second-order connectionist networks. *Journal of the Acoustical Society of America, Supplement 1*, 87, S107.
- Watrous, R. L. (1991). Current status of Peterson-Barney vowel formant data. *Journal of the Acoustical Society of America*, 89, 2459-2460.
- Zwicker, U. T., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of Acoustical Society of America*, 68, 1523-1525.

부록. 모음식별반응에서 사람들의 혼동행렬(Peterson & Barney, 1952)과
본보기모형에 의한 혼동행렬

		모음식별반응(%)									
		[i]	[I]	[e]	[æ]	[ʌ]	[a]	[o]	[U]	[u]	[ɜ]
사람들의 수행결과 (Peterson & Barney, 1952)	원 자 극	[i]	99.9	.0	.0	.0	.0	.0	.0	.0	.0
		[I]	.0	92.9	6.8	.0	.0	.0	.0	.0	.3
		[e]	.0	2.5	87.7	9.2	.0	.0	.0	.0	.5
		[æ]	.0	.0	2.9	96.5	.0	.0	.0	.1	.4
		[ʌ]	.0	.0	.0	.2	87.0	9.9	.7	2.2	.0
		[a]	.0	.0	.0	.0	5.7	92.8	.7	.6	.0
		[o]	.0	.0	.0	.0	.2	.5	96.5	1.7	.9
		[U]	.0	.0	.0	.0	5.3	1.2	1.0	92.2	.0
		[u]	.0	.0	.0	.0	.0	.0	.8	.0	99.2
		[ɜ]	.0	.0	.2	.0	.0	.0	.0	.0	.0
정확식별율: 94.4%											
본보기모형의 수행결과	원 자 극	[i]	96.1	3.9	.0	.0	.0	.0	.0	.0	.0
		[I]	5.3	90.8	3.9	.0	.0	.0	.0	.0	.0
		[e]	.0	10.5	86.8	2.6	.0	.0	.0	.0	.0
		[æ]	.0	.0	13.2	86.8	.0	.0	.0	.0	.0
		[ʌ]	.0	.0	.0	.0	92.1	7.9	.0	.0	.0
		[a]	.0	.0	.0	.0	3.9	89.5	5.3	1.3	.0
		[o]	.0	.0	.0	.0	.0	2.6	94.7	1.3	1.3
		[U]	.0	.0	.0	.0	1.3	.0	2.6	84.2	11.8
		[u]	.0	.0	.0	.0	.0	.0	2.6	10.5	86.8
		[ɜ]	.0	2.6	2.6	1.3	.0	.0	.0	.0	.0
정확식별율: 90.13%											

Processes of Speaker Variability in Vowel Perception: Normalization vs. Exemplars

Yang-Gyu Choi* & Hyun Jung Shin**

*Department of Early Childhood Special Therapeutic Education, Choonhae College

**Department of Psychology, Pusan National University

The purpose of this study was to critically evaluate the normalization models and to suggest the exemplar models as alternatives in explaining the processes of speaker variability in vowel perception. While the normalization models treat the speaker variability as a noise to be removed, the exemplar models use it as one of the important informations in vowel perception. A simulation study was conducted to quantitatively contrast the normalization models and the exemplar models(GCM and ALCOVE) by fitting them to the Peterson & Barney(1952)'s data. The quantitative fits of the exemplar models to the Peterson and Barney(1952)'s data of vowel and gender identification were compared with those of the normalization models. The results showed that predictions of the exemplar models are much better than those of the normalization models. It was suggested that the exemplar models can be more effective modules in processing of the speaker variability in vowel perception. Limitations of the present study and further research problems were discussed in the final section.