

한국인 영어학습자의 지각 모음공간과 발화 모음공간의 연계1)

A Link between Perceived and Produced Vowel Spaces of Korean Learners of English

양 병 곤²⁾

Yang, Byunggon

ABSTRACT

Korean English learners tend to have difficulty perceiving and producing English vowels. The purpose of this study is to examine a link between perceived and produced vowel spaces of Korean learners of English. Sixteen Korean male and female participants perceived two sets of English synthetic vowels on a computer monitor and rated their naturalness. The same participants produced English vowels in a carrier sentence with high and low pitch variation in a clear speaking mode. The author compared the perceived and produced vowel spaces in terms of the pitch and gender variables. Results showed that the perceived vowel spaces were not significantly different in either variables. Korean learners perceived the vowels similarly. They did not differentiate the tense-lax vowel pairs nor the low vowels. Secondly, the produced vowel spaces of the male and female groups showed a 25% difference which may have come from their physiological differences in the vocal tract length. Thirdly, the comparison of the perceived and produced vowel spaces revealed that although the vowel space patterns of the Korean male and female learners appeared similar, which may lead to a relative link between perception and production, statistical differences existed in some vowels because of the acoustical properties of the synthetic vowels, which may lead to an independent link. The author concluded that any comparison between the perceived and produced vowel space of nonnative speakers should be made cautiously. Further studies would be desirable to examine how Koreans would perceive different sets of synthetic vowels.

Keywords: English vowels, vowel space, Korean learners, gender, perception, production, link

1. 서론

모음은 입벌림 정도와 혀의 전후위치에 따라 달라지는데, 영어의 모음체계와 한국어의 모음체계가 다르고, 입안에서의 조음기관인 혀의 움직임을 조절하기 어렵기 때문에 한국인 영어학습자의 지각과 발화에 어려움이 많은 편이다. 이 연구는 한국인 영어학습자가 모음의 음질이 서서히 변해가는 합성음으로 만들어진 모음공간에서 이상적인 영어모음으로 선택한 지각 모음공간과, 지각실험에 참여한 동일한 학습자가 또렷하

게 발음한 발화 모음공간이 어떤 관계를 보이는지 탐구함으로써 한국인 영어학습자의 모음지각과 모음발화와의 관계를 규명하고, 궁극적으로는 영어모음학습의 문제점을 찾아보고자 한다. 구체적으로 이 연구에서는 프라트(Praat)의 포먼트 합성기로 실제 음성에 가까운 합성음을 만들어 한 컴퓨터 화면에 제시하고, 이를 실험 참여자가 하나씩 들어보고 화면에 제시된 목표영어모음에 가장 가까운 이상적인 합성음을 택하게 한다. 합성음은 미국인 남성의 피치와 여성의 피치값을 대표하는 두 가지 종류로 제시한다. 발화실험은 지각실험에 참여했던 대상자들이 목표모음이 들어간 단어를 서로 다른 음높이로 또렷하게 발음한 것 가운데, 좀 더 높은 피치의 이상적인 발음의 음향적인 측정값을 구해서 만든 발화 모음공간을 지각 모음공간과 비교해보기로 한다.

이러한 연구는 지금까지 지각과 발화간의 관계를 살펴보기 보다는 지각이면 지각만을, 발화이면 발화만을 다룬 개별 연구들과는 달리, 지각과 발화간의 연계에 대한 통합적인 연구

1) 이 논문은 2013년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2013S1A5A2A01015538)

2) 부산대학교, bgyang@pusan.ac.kr

접수일자: 2014년 7월 31일

수정일자: 2014년 9월 12일

게재결정: 2014년 9월 22일

방법을 통해 기존 연구에서 볼 수 없었던 새로운 결과를 제시할 수 있을 것으로 기대된다.

2. 이론적 배경

지금까지 많은 연구에서 모음의 지각에 대한 다양한 논의가 진행되었는데 일부 연구자들은 음향적으로 정의된 모음공간에 나타나는 포먼트 유형이 충분한 정보를 제공하는 것으로 주장했다(Diehl, Lotto, & Holt, 2004; Stevens, 1998). 다른 연구자들은 사람의 음성지각은 청취자 자신의 머릿속에 들어 있는 조음동작을 참조한다고 했다(Fowler, 2003; Liberman & Mattingly, 1985; Liberman & Whalen, 2000). 이런 지각연구의 방법에서 주로 음향적 특징을 모음의 어느 지점으로 보느냐에 따라 관점이 달라지기도 한다. 예를 들어, 음향적인 포먼트 유형에 초점을 둔 연구들은 모음 구간 전체에서 중심부에 나타나는 포먼트의 변화가 해당 모음을 결정하는 충분한 정보라고 주장한다. 이에 반해 두 번째 집단들은 문맥적인 차이에 따라 중심부의 포먼트값이 변하며 오히려 시작과 끝부분의 조음 동작이 모음을 결정하는 중요한 정보를 가지고 있다고 본다. 이들은 모음의 중앙부분을 묵음으로 대체한 다음 청자에게 어떤 모음인지 판단하게 한 연구를 통해 모음의 가장자리정보도 중요하다라는 결론을 제시했다(Strange, 1989; Strange, Jenkins, & Johnson, 1983). 이러한 연구결과에서 보면 모음의 포먼트가 모음판단에 중요한 요소임을 알 수 있다. 또 다른 모음지각의 중요한 문제는 화자 정규화(Speaker normalization)에 관한 것이다. 일반적으로 사람들은 성도의 길이가 다른 사람들이 발음한 음성의 음향적 포먼트값은 차이가 많이 나지만, 자연스럽게 적응하여 쉽게 해당모음으로 지각한다. 이에 반해 컴퓨터로 사람의 음성을 처리할 때는 같은 모음이라도 남녀노소에 따라 음향적 포먼트값은 아주 다른 범위의 값으로 나타나게 되어 자동음성인식과 같은 장치에서 많은 오류를 보인다. 모음발음에 대한 연구들에서는 남성과 여성이 발음한 모음의 포먼트값에 일정한 비율의 차이가 있음을 보고했다(Peterson, & Barney, 1952; Hillenbrand, Getty, Clark, & Wheeler, 1995; Yang, 1996). 남성에 비해 상대적으로 성도가 짧은 여성의 포먼트값이 높게 나타난다. 물론 발화방식과 방언의 차이에서도 남녀의 차이가 생길 수 있다. 이러한 비언어적인 차이를 최대한 통제하여 수집한 남녀 포먼트값은 아주 정확하게 서로의 값을 예측할 수 있는 관계를 구할 수 있다(Yang, 1990, 1996). 이런 관계를 이용해서 남성과 여성의 포먼트값을 정규화시키려는 여러 가지 연구가 진행되었지만(Flynn, 2011), 여전히 한계점을 보이고 있다. 정규화의 두 가지 접근 방법은 첫째, 음향적 측정값을 화자사이의 성도길이 차이 비율만큼 나누거나 곱하여, 신체적 차이를 조정하여 서로 비교할 수 있는 값으로 만들거나, 둘째 음향적 측정값을 mel이나 bark와 같은 지각적

척도로 변환하여 동일한 모음은 비슷한 값의 분포를 보이고 있음을 제시하려 했다(Yang, 2009). 이렇게 변환한 값들은 남녀성별의 차이가 없는 모음마다 비슷한 포먼트값을 가지게 되어 음성처리시스템을 하나로 만들 수 있는 장점이 있다. 하지만, 이런 실험에 참가한 화자마다 다른 방언적인 요소나 실험상의 오류가 개입될 여지도 있다. 이런 모음지각에 대한 연구들은 모두 영어 모국어화자를 대상으로 연구를 했는데, 영어를 외국어로 학습하는 한국인화자의 지각유형에 대한 연구는 부족한 편이다. 정해진 포먼트값을 이용하여 서서히 모음의 음질이 변하도록 합성한 음의 쌍을 들려주고 같은지 여부를 물어보는 연구는 있었지만(Yang, 2006), F1와 F2의 모음공간에 일어날 수 있는 수백 개의 실제음에 가까운 합성음을 하나씩 들어보고 가장 이상적인 모델을 찾는 지각실험 연구는 국내에서 아직까지 시도해본 사례가 없는 것으로 여겨진다.

마지막으로 모음지각과 발음에 대한 연계성을 연구한 사례를 살펴보면, Johnson, Flemming & Wright(1993)는 10명의 여성과 4명의 남성을 대상으로 영어모음 지각과 발화의 모음공간을 비교해보았다. 이들의 모음지각실험은 Method of Adjustment(MOA)라는 방법을 이용했는데, 이는 단순히 한 자극 쌍에 대한 참여자의 반응을 얻기 보다는 수백 개의 합성음을 참여자가 눌러보고 자신의 머릿속에 들어있는 가장 이상적인 모음을 점진적으로 찾아가는 방법이다. 그들이 지각과 발화에서 구한 포먼트값으로 만든 모음공간을 비교해본결과 큰 차이가 있음을 발견하고, 이를 해석하는 방법으로 hyperspace effect라는 이론을 제시했다. 이 이론은 사람의 머릿속에는 지각공간과 같이 과장된 이상적인 모음공간이 존재하고 있으며, 이 공간은 발화하는 과정에서 줄어든다는 것이다. 이에 대해 Whalen, Magen, Pouplier, Kang & Isakarous(2004)는 그들의 실험이 방법상 오류를 범한 것으로 혹평했다. 합성음에 대한 지각실험결과는 합성음 자체가 최적의 발음이 아니기 때문에 좀더 가장자리의 값들을 택한 것이고, 지각실험 결과가 비슷하게 나타난 것은 한 화자의 성도모델로 합성한 음성이었기 때문이라고 지적했다. 이들의 논쟁은 아직까지 미완의 결론으로 남아있다. 덧붙여, 신동진 & Iverson(2014)은 한국인 영어학습자의 발화 상에 나타나는 모음삽입현상이 지각실험에서 나타나는 모음삽입현상과 연관이 있는지 조사했는데, 유의미한 연관성이 없다고 결론지었고, 이런 결과는 발화와 지각에서 사용되는 전략이 다르다는 연구(Kabak & Idsardi, 2003)를 뒷받침한다고 지적했다. 이러한 지각과 발화간의 관계는 실험대상이나 조사환경에 따라 달라질 수도 있는데, 본 연구에서는 이들의 연계성에 대한 또 다른 정교한 결과를 보여줄 것으로 기대된다.

3. 연구방법

3.1 참여자

이 연구의 참여한 한국인 영어학습자는 대학생 16명(남자 8명, 여자 8명)이다. 이들 가운데 10명은 언어학을 전공하였고 6명은 영어교육을 전공하였으며, 각 전공의 남녀비율은 반반씩 차지한다. 이들의 토익점수의 평균은 858점으로 다소 높은 편이고 점수의 범위는 최고 970점에서 최저 670점으로 분포되어 있으며 표준편차는 104점이다. 이들은 주로 부산 경남 출신의 대학생들로 특별히 지역이나 방언을 통제하지 않고 임의표본으로 선정하였고, 지각실험과 발화실험에 모두 참가했다.

3.2 지각실험과 자료분석

지각실험에 사용한 합성음은 프라트의 KlattGrid(Klatt & Klatt, 1990)를 이용해서 4개의 포먼트값을 입력하여 실제 사람의 발화에 가깝게 만들었다. 포먼트값은 기존의 연구(Johnson, Flemming, & Wright, 1993; Whalen, Magen, Pouplier, Kang, & Isakarus, 2004)에서 제시한 범위의 값을 이용하였다. 피치값은 미국인남성의 기본값인 110 Hz와 미국인여성의 기본값인 230 Hz로 만들었다(Yang, 1996). 피치값의 차이를 둔 것은 피치에 따라 지각유형에도 달리 나타날 것인지 알아보기 위함이었다. 모음의 음질을 결정하는 두 가지 중요한 음향변수인 F1과 F2의 모음공간에 나타날 수 있는 모음을 만들기 위해 F1은 250 Hz에서 F2는 800 Hz부터 가능한 조합을 만들고, F3는 Nearey(1989)에 사용한 회귀함수를 이용하여 정하고, F4는 3500 Hz를 부여하되 F3가 3500 Hz 이상이 되면 300 Hz를 더하여 자연스런 모음이 되도록 했다. 이렇게 두 가지의 피치에 해당하는 합성음세트를 각각 330개씩 (F2 가로축 22개, F1 세로축 15개) 만든 뒤, 플라잉팝콘(Parasys)이라는 멀티미디어제작 도구를 이용해서 각 화자들의 반응을 자동으로 수집하는 시스템을 만들었다. 플라잉팝콘으로 만들어진 컴퓨터 화면에는 맨 위에 영어모음이 포함된 목표단어가 하나 제시되어 있고, 바둑판처럼 배열된 330개의 초록색 바둑알모양의 단추가 나타나있고, 맨 아래에는 얼마나 자연스럽게 이상적인 모음인지를 선택하는 7개의 숫자(1번은 가장 나쁜 모음, 7번은 가장 이상적인 모음)가 적혀있는 자연성평가 단추가 있고, 마지막으로 이 화면에 대한 모든 반응을 저장하는 단추가 있다. 실험 참여자는 화면에 나타난 330개의 단추 가운데 4개의 가상공간으로 나눈 지점의 중심에 있는 단추를 먼저 눌러서 이 단추와 연결된 합성음을 들어보고, 화면 맨 위에 주어진 단어의 영어모음에 가까운 단추를 중심으로 상하좌우의 단추를 건너뛰어 가며 하나하나 클릭하여 들어보면서 330개 합성음세트 가운데 가장 이상적인 모음을 하나 선택하고, 연이어 자연성 정도를 나타낸 숫자를 선택한다. 저장단추를 누르면 마지막으로 선택한 합성음의 F1과 F2 정보와 자연성평가값이

컴퓨터에 저장되고, 다음 화면으로 넘어간다. 이런 자료 수집 방법의 특징은 제한된 수의 합성음이나 자극 쌍에서 구할 수 없는 연구대상자 자신의 머릿속에 들어있는 모음체계에 가까운 결과를 구할 수 있다는 점이다. 목표단어는 모두 39개로 처음 3개는 odd, heed, who'd를 넣어서 지각실험에 적용하도록 유도하였고, 실제 지각분석에 사용한 나머지 36개는 두 개의 세트로 나뉘는데, f0를 110 Hz로 설정하고 남성의 이미지를 보이면서 남성의 목소리로 반응하라고 지시한 18개 (9개 모음에 대해 2번씩 반응)와 f0를 230 Hz로 설정하고 여성의 이미지를 보이면서 여성의 목소리로 반응하라고 지시한 18개이다. 참여자의 개별 지각반응시간은 약 30분에서 50분이 걸렸다. 지각 측정 자료는 총 1728개(36개 모음x2개 포먼트와 자연성평가값x16명)이다.

지각실험의 자료분석은 먼저 모든 참여자의 결과를 엑셀로 열어서 f0기준 110 Hz와 230 Hz로 나눈 뒤, 두 번씩 반응한 모음별로 정렬했다. 이어서 참여자마다 가장 이상적인 모음으로 두 번 선택한 합성음의 F1과 F2의 값을 자연성평가값을 반영한 최종값으로 구했다. 예를 들어, 한 참여자의 f0가 110 Hz인 heed의 첫 번째 지각반응으로 F1이 293 Hz이고 F2가 2454 Hz인 합성음을 7점의 자연성평가값을 선택했고, 두 번째 합성음은 F1이 336 Hz이고 F2가 2914 Hz인 합성음에 6점의 자연성평가값을 선택했다고 가정하면, 이 반응에 대한 F1의 최종값은 $\{(293 \times 7) + (336 \times 6)\} \div (7+6) = 313$ Hz이고, F2의 최종값은 $\{(2454 \times 7) + (2914 \times 6)\} \div (7+6) = 2666$ Hz가 된다. 이렇게 계산함으로써 자연성평가값의 비중을 반영한 자료를 만들어 통계 비교에 사용했다. 마지막으로 개별 지각반응 결과를 남녀집단별로 평균과 표준편차값을 구했다. 통계처리는 집단별로 정규분포를 가정할 수 없다고 판단하여 SPSS(v.20)의 비모수통계인 Mann-Whitney의 U검정을 사용하여 비교했고, 유의수준은 0.05로 하였다.

3.3 발화실험과 자료분석

이 연구에 사용한 발화 녹음자료는 hVd의 환경에 모음을 넣은 단어로 heed, hid, head, had, who'd, hood, awed, odd, hudd이다. 분석에 사용하지 않는 맨 처음 3개를 제외하고 이 단어들을 두 번씩 되풀이해서 나타나도록 인쇄한 뒤 참여자들에게 녹음하게 했다. 연구자는 참여자에게 “시끄러운 길거리에서 어린아이나 잘 듣지 못하는 사람에게 말하듯이 천천히 입을 크게 벌리고 또렷하게 발음”하도록 요구했다. 발음환경은 “Say heed, heed, heed, heed, again”과 같이 4번을 음높이 변화를 주어 첫 번째와 세 번째 단어는 좀 더 높은 피치로 두 번째와 네 번째는 낮은 피치로 발음하게 하였다. 녹음된 자료 가운데, 지각실험에서 구한 가장 이상적인 발음에 버금가는 음성을 구현하기 위해 높은 피치로 발음한 첫 번째와 세 번째의 발음 가운데 하나의 음향적 특징을 수집했다. 음성녹음은

조용한 연구실에서 GoldWave(v.5.69)를 실행하고 컴퓨터에 연결된 쎄나이저 헤드셋마이크(Sennheiser PC151)를 착용한 참여자가 데스크탑 컴퓨터에 Mono PCM(Pulse-Code Modulation) 방식으로 16비트의 진폭해상도에 44.1 kHz의 표본속도로 녹음했다. 분석과정은 녹음된 파일을 프라트에 하나씩 불러온 후, 다음의 프라트 분석 스크립트를 이용하여 피치, F1, F2값을 측정하고 그 결과를 텍스트 파일로 저장했다.

```
soundName$=selected$("Sound")
Edit
editor Sound 'soundName$'
  Spectrogram settings... 0 5000 0.005 30
  Formant settings... 5500 5 0.025 30 1
  Pitch settings... 75 500 Hertz autocorrelation automatic
  for i from 1 to 30
  clearinfo
  pause Stop this analysis?
  pause Select a vowel segment to analyze...
  Move start of selection to nearest zero crossing
  start=Get start of selection
  Move end of selection to nearest zero crossing
  end=Get end of selection
  Play... start end
  dur='end'-'start'
  onethirdpoint=start+dur/3
  Select... onethirdpoint-0.0125 onethirdpoint+0.0125
  f0=Get pitch
  f1 = Get first formant
  f2 = Get second formant
  print
'soundName$'tab$'f0:0'tab$'f1:0'tab$'f2:0'tab$'tab$'tab$'
  pause Write a vowel name into the info window
endfor
Close
endeditor
```

이 스크립트는 프라트 개체창에 불러온 음성파일을 편집창에 열어준다. 한 음성파일에 한 참여자가 발화한 단어가 21개로 이어져 있기 때문에 잠깐 화면을 정지하여 되풀이해서 측정할 것인지 연구자에게 묻도록 하여, 확인을 누르면 분석이 진행되고, 더 이상 분석할 음성파일이 없으면 취소를 눌러 분석을 끝낸다. 확인 단추를 누르기 전에 연구자는 편집창에 나타난 포먼트의 궤적이 스펙트로그램의 에너지가 집중된 검은 띠의 중앙에 적절히 위치해 있는지 확인하고, 만약 포먼트의 위치가 부정확할 경우에는 연구자가 확인 단추를 누르기 전에

포먼트 갯수 설정을 남성일 경우에는 기본을 5개로, 여성일 경우에는 4개로 하고 1개 단위로 증감하여 검은 띠의 중심부로 포먼트 궤적이 지나가도록 한다. 구체적으로 영어모음 가운데 F1과 F2가 서로 가까이 접근하는 후설원순모음이 들어간 발음에서는 1개를 증가시킨 경우가 많았다. 이어서 연구자가 마우스로 분석할 모음구간을 선택하면, 전체모음지속시간의 1/3에 해당하는 지점의 앞뒤 각각 12.5 ms구간을 컴퓨터가 선택하여 파일이름과 피치값, F1과 F2를 구하여 정보창에 입력해준다. 1/3지점을 선택한 이유는 모음발화 후 목표모음에 도달하는데 약간의 시간이 소요될 것으로 보고 시작점보다 약간 뒤의 위치를 잡았고, 무엇보다도 발화자의 발화속도에 관계없이 일관된 비율을 적용함으로써 서로 비교할 수 있는 측정값을 구하기 위해서이며, 25 ms의 측정구간을 선택한 이유는 한 시간지점의 측정값보다는 보다 안정적인 측정값을 구할 수 있기 때문이다. 이어서 정보창에 선택한 모음정보를 입력하기 위해 잠깐 멈춘다. 처음 발화파일을 열어 실행할 때 편집창 내에서만 작동하는 New editor script를 클릭해서 다음과 같이 입력한 뒤 실행하여 분석한 모음이 들어간 영어단어가 추가 입력된 정보창의 내용을 result 파일에 덧붙여 쓰게 했다.

```
vowelName$="heed"
print 'vowelName$'newline$'
```

이 짧은 스크립트는 다른 이름으로 저장하여 두고, 분석한 모음이 들어간 단어를 매번 수정하면서 나머지 음성분석에 활용했다.

이렇게 측정한 발화자료는 총 864개(9개 모음x2개 포먼트와 피치값x16명x2번 발화)이고, 두 번 발화의 평균을 구해 통계비교에 사용했다. 자료 처리과정은 먼저 남녀집단별 지각과 발화포먼트값은 엑셀에 불러와 전체평균값을 구한 뒤 R 그림으로 나타냈다. 그래프의 단위는 모두 프라트에서 제시하고 있는 hertzToBark에 제시한 공식을 엑셀에 아래의 공식 (1)을 이용하여 지각적인 차이를 보여주는 Bark로 변환하였다.

$$(1) 7 * \ln\left(\frac{\text{Hertz}}{650} + \sqrt{1 + \left(\frac{\text{Hertz}}{650}\right)^2}\right)$$

통계처리는 지각실험에서와 같이 SPSS 20로 Mann-Whitney의 U검정을 사용하였고, 회귀분석은 R(v.3.1.1)로 처리했다.

4. 지각실험 결과

한국인 영어학습자의 지각실험 결과는 <표 1>에 나타나 있다. 이 표는 남성집단과 여성집단을 분리하여 각 피치세팅에 따른 지각모음의 평균값과 표준편차를 보여준다.

표 1. 한국인 영어학습자의 피치를 달리한 합성음에 대한 지각실험에서 구한 포먼트값의 평균과 표준편차(괄호)의 분포. 마지막 글자 m은 남성의 자료를 나타내고, f는 여성의 자료를 나타냄.

Table 1. Distribution of means and standard deviations (in parentheses) of formant values of English vowels perceived by Korean English learners on two f0 sets of synthetic stimuli. The final letter *m* indicates male data while that of *f* does female data.

Vowels	f0	F1m	F2m	F1f	F2f
i		315(40)	2604(137)	325(57)	2473(216)
ɪ		346(73)	2443(266)	309(55)	2512(242)
ɛ		681(74)	2258(323)	634(121)	2102(245)
æ		677(96)	2183(291)	598(109)	2062(279)
u	110	334(40)	972(111)	339(62)	1121(406)
ʊ		341(39)	1079(151)	340(71)	1099(442)
ʌ		596(128)	981(60)	553(76)	1203(210)
ɔ		595(162)	1057(83)	633(125)	1195(495)
ɑ		632(122)	1024(125)	651(104)	1100(175)
i		334(34)	2541(272)	329(51)	2509(180)
ɪ		318(57)	2649(219)	324(75)	2570(204)
ɛ		638(85)	2184(223)	663(130)	2132(164)
æ		762(65)	2099(283)	658(132)	2118(249)
u	230	362(53)	1025(134)	336(77)	1006(170)
ʊ		349(48)	1049(210)	367(75)	1000(130)
ʌ		650(120)	1142(172)	637(91)	1126(158)
ɔ		616(137)	1081(66)	703(133)	1286(537)
ɑ		682(161)	1073(59)	723(114)	1094(139)

<표 1>에서 보면 대체로 남성과 여성집단에 약간의 차이를 보이지만 그다지 크지 않음을 알 수 있다. 개별 모음별 포먼트값을 살펴보면 일부는 110 Hz에서보다 230 Hz에서 높게 나타났는데, 110 Hz를 기준으로 해서 상승한 항목은 36개 가운데 25개로 69%나 된다. 포먼트별로 보면 F1에서 상승한 경우는 남성과 여성집단에서 15개이어서 83%를 차지하고 F2에서는 10개로 55.6%가 된다. 양적인 값과 음적인 값이 혼합되어 있어 산술적인 평균을 구하기가 어렵기 때문에, 각 포먼트별로 평균값을 구해 비교해보면, 110 Hz의 경우에는 F1m이 502 Hz, F2m이 1622 Hz, F1f가 487 Hz, F2f가 1652 Hz로 나타났고, 230 Hz의 경우에는 F1m이 523 Hz, F2m이 1649 Hz, F1f가 527 Hz, F2f가 1649 Hz로 나타났다. 이렇게 전체 포먼트 평균값으로 보면 F1에서는 남녀 모두 21 Hz, 40 Hz 높은 값을 택했으나, F2에서는 남성집단에서만 27 Hz 높아졌고, 여성집단에서는 3 Hz 낮아진 결과를 보였다. 이러한 결과는 이상적인

모음을 선택할 때 F1에서 아주 미미하지만 피치값의 영향을 받은 것으로 볼 수 있다.

<그림 1>은 한국인 영어학습자의 지각실험 결과를 좀 더 직관적으로 관찰할 수 있게 한다.

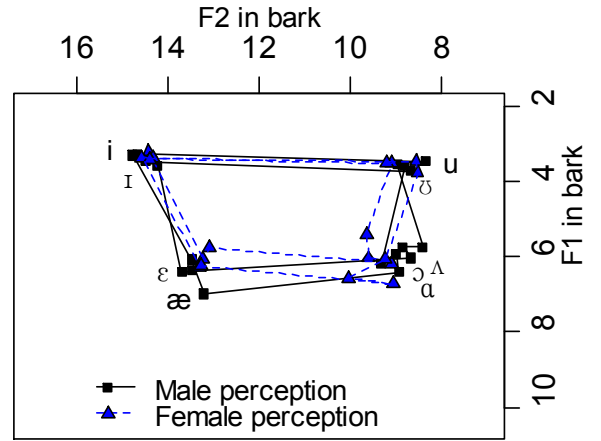


그림 1. 남자(사각형)와 여자(삼각형) 집단별 영어합성모음지각공간. 피치값이 110 Hz와 230 Hz인 두 벌에 대한 반응을 집단별로 연결하여 나타냄. x축은 F2를, y축은 F1을 나타내며 단위는 모두 bark임.

Figure 1. Perceived vowel spaces of male(rectangles) and female(triangles) Korean learners. Two responses to f0 value settings (110 Hz, 230 Hz) were connected peripherally. x-axis indicates F2 values while y-axis does F1 values in bark.

<그림 1>은 남녀집단의 포먼트값을 bark단위로 변환하여 한꺼번에 제시했는데, 전체적인 경향을 살펴보면 다음과 같다. 먼저 전설고모음의 긴장모음과 이완모음쌍들([i-ɪ, u-ʊ])이 영어 원어민들의 발음과 달리 구분되지 않고 비슷한 위치에 모여 있음을 관찰할 수 있다. 덧붙여 전설저모음인 [e-æ]도 남성집단에서는 다소 차이를 보이지만, 여성집단에서는 거의 근접한 모양을 보이고 있다. 후설저모음에서도 모음들이 서로 뒤섞여있다. 이 그림에서는 평균값으로만 나타냈기 때문에 집단별 차이를 알기 어렵다.

남녀집단간의 지각실험에서 나타난 모음 포먼트값들 간에 어떤 통계적 차이를 보이는지 알아보기 위해, Mann-Whitney의 U검정으로 비교해보았다. 먼저, 남녀집단 모두의 각 모음별 F1과 F2로 된 18개의 변수를 검정필드로 하고 피치값(110 Hz와 230 Hz)을 집단으로 처리해본 결과 모든 항목에서 유의미한 차이를 보이지 않았다. 이어서, 남녀집단 모두의 각 모음별 F1과 F2를 검정필드로 하고 남녀집단별로 비교해본 결과 had 발음인 [æ]의 F1값에서만 유의미한 차이를 보였다 ($U=72.5, p<0.05$). <그림 1>에서 보면 여성집단보다는 남성집단의 모음 [æ]에 대한 F1 평균값이 다소 높은 것을 알 수 있다. F1값이 높으면 입벌림 동작을 크게 하여 더 구분되게 발

음한 것으로 유추할 수 있다(Pickett, 1987). 이번에는 남녀집단을 따로 나누어, 각 모음별 F1과 F2값들을 검정필드로 하고 피치별로 비교해본 결과, 남성집단과 여성집단에서는 모든 항목에서 유의미한 차이를 보이지 않았다. 이러한 결과는 남녀 집단별로 피치별로 동일하게 합성음을 지각하였다고 말할 수 있다.

5. 발화실험 결과

<표 2>는 한국인 영어학습자들이 발음한 영어모음의 피치와 포먼트값 및 표준편차를 나타낸다.

표 2. 한국인 영어학습자들이 발음한 영어모음의 피치와 포먼트 평균값 및 표준편차(괄호 안)의 분포. 마지막 글자 m은 남성의 자료를 나타내고, f는 여성의 자료를 나타냄.

Table 2. Distribution of means and standard deviations(in parentheses) of pitch and formant values of English vowels perceived by Korean English learners. The final letter m indicates male data while that of f does female data.

Vowels	f0m	F1m	F2m	f0f	F1f	F2f
i	183	358(23)	2307(173)	296	399(55)	2982(69)
ɪ	183	358(33)	2281(155)	299	408(57)	2897(158)
ε	171	654(61)	1847(147)	276	848(39)	2186(174)
æ	170	670(71)	1835(157)	285	845(41)	2153(228)
u	184	395(33)	914(64)	309	487(56)	1056(193)
ʊ	186	408(8)	1008(127)	299	481(78)	1085(160)
ʌ	168	666(66)	1026(97)	270	861(56)	1280(82)
ɔ	160	703(98)	1055(122)	262	808(69)	1190(103)
ɑ	162	702(96)	1047(102)	262	838(103)	1190(72)

<표 2>에서 보면 남성집단의 피치값의 평균은 175 Hz이고 표준편차는 9.9 Hz이고 범위는 160 Hz에서 186 Hz에 걸쳐져 있으며, 여성집단의 평균은 285 Hz이고 표준편차는 16.9 Hz이고, 범위는 262 Hz에서 309 Hz로 나타났다. 이러한 결과는 Yang(1996:252)에서 구한 한국인 남녀집단의 평균과 비슷한 분포를 보인다(남자 10명의 평균값 169 Hz, 표준편차 25 Hz; 여자 10명의 평균값 269 Hz, 표준편차 29 Hz). 이어서 모음별 포먼트값의 관계를 알아보기 위해 R의 회귀분석을 실시해본 결과, 모든 남성의 포먼트와 모든 여성의 포먼트 사이의 관계는 기울기가 1.258이고 절편은 -50.8을 나타냈다($r^2=0.991$, $p<0.05$). 이러한 관계는 결국 여성집단의 F1과 F2값이 남성의 값보다 25.8% 더 높다는 것으로 해석할 수 있다. 이러한 한국인 남녀집단의 신체적인 성도길이의 차이는 Yang(1996, Figure 3)에서 회귀분석을 통해 처리한 결과 기울기가 1.171로 나타나서 여성의 포먼트값이 17.1% 높다는 결과와 비슷하다. 이 연

구에서는 젊은 대학생 남녀를 대상으로 했고, Yang(1996)의 연구에서는 좀 더 나이든 대학원생 남녀를 대상으로 했기 때문에 피치와 기울기에서 차이가 났을 것으로 추정된다.

<그림 2>는 이들 남성과 여성집단별 포먼트 평균값으로 만들어진 발화 모음공간을 bark척도로 변환하여 보여준다.

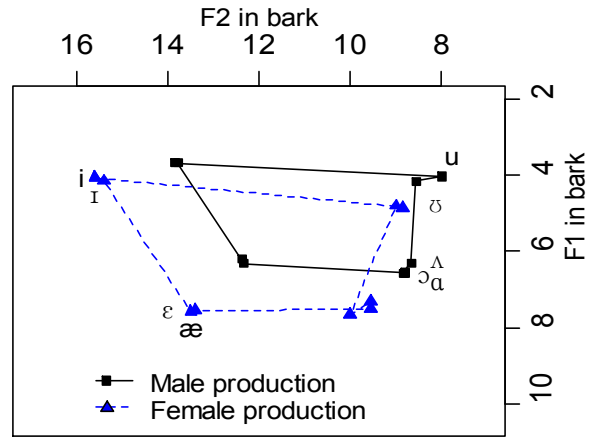


그림 2. 한국인 영어학습자인 남성(사각형)과 여성(삼각형) 집단의 발화 모음공간. x축은 F2를, y축은 F1을 나타내며 단위는 모두 bark임.

Figure 2. Produced vowel spaces of male(rectangle) and female(triangle) Korean English learners. x-axis indicates F2 values while y-axis does F1 values in bark.

<그림 2>에서 보면 남성과 여성집단 모두 전설과 후설 고모음에서 구분되지 않음과 후설 고모음에서 구분되지 않음을 관찰할 수 있다. 또한 전설저모음쌍도 구분이 되지 않으며 후설저모음에서도 모음들이 서로 모여져 있음을 알 수 있다. 이러한 경향은 한국인 영어학습자의 공통된 문제점으로 이미 여러 연구에서 지적되었다(Kent & Read, 2002; Ladefoged, 2001; Yang, 1996; 강석한, 2007). 양병곤(2008)은 원어민과 한국인들이 발음한 영어모음의 한 측정지점에서 구한 포먼트값보다는 여러 개의 포먼트값으로 된 궤적을 비교해보고, 한국인남성이 영어 긴장모음과 이완모음을 다르게 발음하기 위해서는 입벌림과 혀 움직임 좀 더 뚜렷하게 차이를 두어 발음해야 한다고 지적했다. 덧붙여, 양병곤(2013)에서는 전설모음에서 상대적인 거리 비율이 원어민과의 차이를 보인다고 지적했는데, 이런 문제점들을 교정하기 위해서는 입벌림을 강제로 실현시키는 보조장치를 이용하는 방안을 제안했다(양병곤, 2010).

6. 지각 모음공간과 발화 모음공간의 비교

이 연구에서는 한국인 영어학습자의 지각실험에서 나타난 지각 모음공간과 발화실험에서 구한 발화 모음공간을 살펴보

고 평균값의 비교에서 볼 수 없는 개별값들을 통계적으로 비교할 경우에 어떤 문제점이 있는지 살펴보고자 한다.

앞 절의 지각실험결과에서 보았듯이 남녀의 지각패턴이 피치값에 따라 통계적으로는 차이가 없지만 약간의 절대적인 차이를 보였기 때문에, 남성집단에서는 f0가 110 Hz에 반응한 지각모음에서 구한 F1과 F2값을 이용하고, 여성집단에서는 f0가 230 Hz에 반응한 값을 이들의 실제 발화음성과 함께 비교했다.

<그림 3>에서 보면 남성집단의 지각공간이 발화공간에 비해 약간 적은 면적을 차지하고 있음을 알 수 있다. 또한 영어 고모음의 긴장이완모음쌍에서도 지각과 발화가 비슷한 유형을 보이고 있다. 전설저모음에서는 비록 F1에서는 차이가 없지만, F2에서는 많은 시각적인 차이를 보인다. 후설저모음에서도 발화와 지각이 거의 비슷한 위치에 몰려있다.

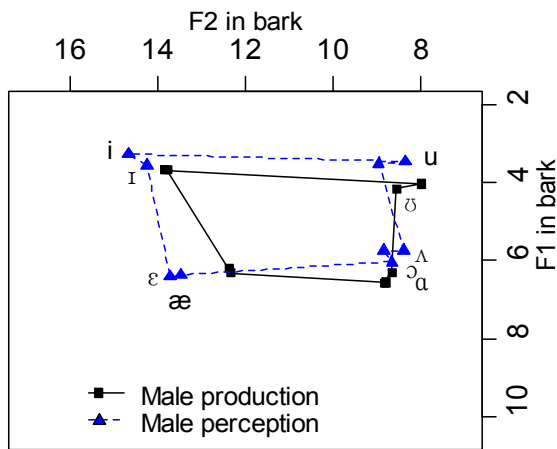


그림 3. 한국인 남성 영어학습자집단의 지각(사각형)과 발화(삼각형)모음공간 비교

Figure 3. Perceived(rectangle) and produced(triangle) vowel spaces of male Korean English learners. x-axis indicates F2 values while y-axis does F1 values in bark.

<그림 4>에서는 남성집단에서 보았던 것과는 달리 여성집단의 지각공간이 발화공간에 비해 아주 넓은 면적을 차지하고 있음을 알 수 있다. 또한 영어고모음의 긴장이완모음쌍에서도 발화와 지각이 비슷한 유형을 보이고 있지만 남성집단에 비해 모음지점간 거리가 멀다. 나머지 모음들에서도 많은 시각적인 차이를 보인다. 이러한 차이는 집단별 편차를 반영하지 않은 단순한 평균값의 비교라 할 수 있다.

마지막으로 세부적인 집단별 차이를 살펴보기 위해, 한국인과 미국인이 발음한 모음별로 남녀집단별 모음지각과 모음발화공간에 통계적인 차이를 Mann-Whitney의 U검정으로 비교해 보았다. <표 3>은 한국인 영어학습자 남녀집단이 발음한 영어모음의 포트별 집단간 통계적 유의도를 보여준다.

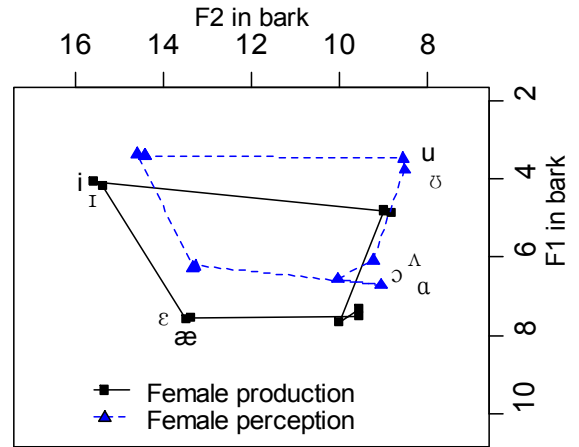


그림 4. 한국인 여성 영어학습자집단의 지각(사각형)과 발화(삼각형) 모음공간 비교

Figure 4. Perceived(rectangle) and produced(triangle) vowel spaces of female Korean English learners. x-axis indicates F2 values while y-axis does F1 values in bark.

표 3. 한국인 영어학습자 남녀집단이 발음한 영어모음포트별 통계 비교 결과

Table 3. Statistical results of formant values on each vowel produced by the Korean male and female English learners(n=16)

Vowels	Males		Females	
	F1	F2	F1	F2
i	0.03*	0.00*	0.03*	0.00*
I	0.72	0.10	0.02*	0.01*
ε	0.72	0.01*	0.00*	0.88
æ	0.88	0.01*	0.00*	0.88
u	0.00*	0.32	0.00*	0.57
ʊ	0.00*	0.28	0.02*	0.33
ʌ	0.28	0.23	0.00*	0.05*
ɔ	0.13	0.88	0.11	0.72
ɑ	0.28	0.72	0.03*	0.16

* $p < 0.05$

<표 3>을 보면 남성화자일 경우에는 6개의 측정값(전체 비교대상의 33%)에 대해 유의미한 차이를 보였지만, 여성화자일 경우에는 11개의 측정값(61%)에 대해 유의미한 차이를 나타냈다. 이러한 차이는 결국 단순히 실험참여자의 성별에 관계없이 지각실험을 하고 이를 발화실험결과와 비교할 경우에 잘못된 결론을 도출할 수 있음을 나타낸다. 다시 말해서, 여성화자들의 대다수가 성도의 길이가 짧은 참여자를 실험대상자로 삼았다면 포트값이 성도의 길이에 반비례하기 때문에 (Pickett, 1987), 유의미한 차이를 훨씬 더 많이 내었을 것으로 예상된다. 이러한 지각과 발화의 차이를 설명하기 위해 Johnson, Flemming, & Wright(1993)는 hyperspace effect라는 화

자의 발화 모음공간에 비해 두뇌 속에 들어있는 지각 모음공간은 과장된 크기를 가지고 있다고 주장하는 이론을 제기했는데, 그 이론의 재검토가 필요함을 알 수 있으며, Whalen 등 (2004)의 비판을 지지한다. 덧붙여, 지각실험결과에서 피치와 성별에 관계없이 동일하게 지각한 것은 이 논문에서 사용한 합성음이 Whalen, Magen, Pouplier, Kang & Isakarous(2004)가 지정한 바와 같이 한 명의 남성화자가 발음한 포먼트값의 조합에 해당하므로 단순히 피치값을 변화시켜 응답을 요구했기 때문에 남성이나 여성에 관계없이 유의미한 차이를 보이지 않는 결과를 가져온 것으로 추정된다. 한편, 비원어민화자에 해당하는 한국인화자의 영어지각 모음공간과 발화 모음공간의 비교에서 긴장이완모음쌍이나 저모음군을 구분하여 지각할 수 없기 때문에 발화 모음공간에서도 구분되어 나타나지 않았다는 점에만 주목한다면 지각과 발화는 연계가 있다고 주장할 수 있다. 지각과 발화의 연계를 주장한 연구로는 일본인 영어학습자에게 /r, l/ 구분 청취훈련을 일정기간 시켰을 때 이들을 더 정확하게 구분하여 발화한 결과를 보고 지각과 발화는 연관성이 높다고 지적했다(Bradlow, Pisoni, Yamada, & Tohkura, 1997). 또한 Rauber, Escudero, Ricardo, Bion, & Naptista(2005)의 브라질 영어학습자들에게 영어모음지각과 발화의 실험을 해본 결과, 정확하게 구분하여 들을 수 있는 모음은 정확하게 구분하여 발음하였고, 그렇지 않는 모음은 발음도 정확하지 못해서 지각과 발화는 연관성이 있다고 보고했다.

그런데, 이 연구에서 사용한 합성음이 일정한 비율간격으로 만들어진 것이기 때문에 남성과 여성집단이 동일하게 지각하였고, 발화한 집단이 남성일 때보다는 여성일 때 통계적으로 유의미한 차이가 더 많다는 것은 참여자의 구성이 누군가에 따라 연계성을 주장할 수도 있고, 반대로 연계성이 없다고 주장할 수도 있게 될 것이다. 다시 말해서, 우연히, 합성음으로 만들어진 모음공간에 버금가는 성도를 가진 참여자가 많을수록 연계성이 높다는 결론을 내리게 될 것이고, 그런 참여자가 적을수록 연계성이 없다는 결론에 이를 것이다. 즉, 모음지각과 발화의 관계를 볼 때 어떤 부분을 보는가, 참여자의 구성은 어떤가에 따라 연관성이 있다고 하거나 그 반대로 연관성이 전혀 없다라고 서로 모순되는 결론을 내릴 수도 있는 양면성이 있을 것으로 여겨진다.

7. 요약 및 결론

이 연구에서는 대학생들로 구성된 한국인 영어학습자의 영어모음 지각과 발화의 연계에서 연구대상이나 방법에 따라 어떤 해석상의 문제점이 있는지 살펴보았다. 지각실험은 16명의 남녀 참여자마다 합성음 대해 반응한 지각 모음공간을 구했으며, 발화 모음공간은 또렷하고 크게 발음한 단어의 모음을 일

관성 있는 측정지점을 적용하여 구했다. 지각 모음공간에서 합성음에 사용한 피치값이나 성별에 차이가 있는지, 발화 모음공간과 비교했을 때는 어떤 차이가 있는지 비교해본 결과는 다음과 같다.

첫째, 지각실험 결과 전반적으로 피치별, 남녀집단별로 거의 동일한 반응을 보였다. 피치에서는 F1에 대해서만 약간의 영향을 주었지만, 통계적으로 유의미한 차이를 보이지 않았다. 이러한 결과는 한국인 영어학습자들이 합성음으로 만들어진 동일한 자극에 대해 비슷하게 지각하였음을 보여준다.

둘째, 발화 모음공간에서는 남녀집단간에 약 25.2%의 차이를 보였다. 이러한 차이는 남녀의 성도길이의 신체적인 차이에서 나온 것이다. 이 비율은 신체적으로 차이를 보이는 참여자의 구성에 따라 달라질 것이다.

마지막으로, 지각 모음공간과 발화 모음공간을 비교해본 결과 여성화자에서는 모음포먼트에 대해 유의미한 통계적 차이가 남성화자에서보다 두 배 정도로 많았다. 이러한 차이는 지각 모음공간이 동일하게 나타났기 때문에, 발화모음공간에서 차이가 주된 요인으로 이미 예상된 결과다.

이러한 결과를 종합해보면 서로 다른 신체적인 조건을 가진 한국인 영어학습자가 지각하고 발화한 모음공간을 비교할 때 매우 주의해야 한다고 결론지을 수 있다. 이런 관계를 인정하지 않고 단순히 지각공간과 발화공간을 비교하는 것은 서로 모순되는 결론을 내리거나 hyperspace effect와 같은 오류를 범할 것이다. 앞으로 한 명의 남성보다는 여성이나 아동의 포먼트비율로 만든 합성음에 대해서는 어떤 지각의 결과를 보이는지 연구해 볼 계획이다.

참고문헌

- Bradlow, A.R., Pisoni, D.B., Yamada, R.A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299-2310.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.
- Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York Papers in Linguistics Series*, 2(11), 1-28.
- Fowler, C. A. (2003). Speech production and perception. In A. Healy & R. Proctor (Eds.), *Comprehensive handbook of psychology, Volume 4: Experimental psychology* (pp. 237-266). New York: John Wiley and Sons.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099-3111.

- Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 69(3), 505-528.
- Kabak, B., & Idsardi, W. (2003). Syllabically conditioned perceptual epenthesis. *Proceedings of the Berkeley Linguistics Society*, 29, 233-245.
- Kang, S. (2007). *Acoustic analysis of speech*. (Korean translation) Seoul: Thomson Learning Korea, Pakhaksa.
(강석환 역 (2007). 음향음성분석론. 서울: Thomson Learning Korea, 박학사.)
- Kent, R., & Read, C. 2002. (2nd ed.) *Acoustic analysis of speech*. San Diego, CA: Singular Publishing Group.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.
- Ladefoged, P. 2001. (4th ed.) *A course in phonetics*. Boston: Heinle & Heinle.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Lieberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4, 187-196.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85(5), 2088-2113.
- Peterson, G. & Barney, H. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America* 24, 175-184.
- Pickett, J. (1987). *The sounds of speech communication: A primer of acoustic phonetics and speech perception*. Austin, Texas: pro-ed.
- Rauber, A.S., Escudero, P., Ricardo, A.H., Bion, R.A.H., & Naptista, O.B. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. *Proceedings of Interspeech*, 2913-2916.
- Shin, Dong-Jin, & Iverson, Paul. (2014). An experimental study of vowel epenthesis among Korean learners of English. *Phonetics and Speech Sciences*, 6(2), 163-174.
(신동진 & Iverson (2014). 한국인 영어학습자의 모음삽입현상에 대한 연구. *말소리와 음성과학*, 6(2), 163-174.)
- Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America*, 85, 2135-2153.
- Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, 74, 695-705.
- Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., & Isakarov, K. (2004). Vowel production and perception: Hyperarticulation without a hyperspace effect. *Language and Speech*, 47(2), 155-174.
- Yang, Byunggon. (1990). *Development of vowel normalization procedures: English and Korean*. Ph. D. Dissertation, The University of Texas at Austin.
- Yang, Byunggon. (1996). A comparative study of English and Korean monophthongs produced by male and female speakers. *Journal of Phonetics*, 24, 245-261.
- Yang, Byunggon. (2006). Discrimination of Synthesized English Vowels by American and Korean Listeners. *Journal of the Korean Society of Speech Sciences*, 13 (1), 7-27.
- Yang, Byunggon. (2008). An acoustical comparison of English tense and lax vowels produced by Korean and American males. *Speech Sciences*, 15(4), 19-27.
(양병곤 (2008). 한국인남성과 미국인남성이 발음한 영어 긴 장-이완모음의 음향적 비교. *음성과학*, 15(4), 19-27.)
- Yang, Byunggon. (2009). English vowel spaces produced and perceived by Americans and Koreans. In Chungmin Lee & G.B. Simpson, Youngjin Kim, (Eds.) *The Handbook of East Asian Psycholinguistics*, Cambridge: Cambridge University Press, pp. 390-397.
- Yang, Byunggon. (2010). A method for correcting English vowel pronunciation by wooden chopsticks. *Phonetics and Speech Sciences*, 2(4), 3-10.
(양병곤 (2010). 나무젓가락에 의한 영어모음 발음교정 방안. *말소리와 음성과학*, 2(4), 3-10.)
- Yang, Byunggon. (2013). A comparative study of relative distances among English front vowels produced by Korean and American speakers. *Phonetics and Speech Sciences*, 5(4), 99-107.
(양병곤 (2013). 한국인과 미국인이 발화한 영어전설모음의 상대적 거리 비교. *말소리와 음성과학*, 15(4), 99-107.)

• 양병곤 (Yang, Byunggon)

부산대학교 영어교육과
부산시 금정구 장전동 30
Tel: 033-649-7816
Email: bgyang@pusan.ac.kr
Homepage: <http://fonetiks.info/bgyang>