

## 발화방식에 따른 미국인 남성 영어모음의 피치와 포먼트 궤적

### Pitch and Formant Trajectories of English Vowels by American Males with Different Speaking Styles

양 병 곤<sup>1)</sup>  
Yang, Byunggon

#### ABSTRACT

Many previous studies reported acoustic parameters of English vowels produced by a clear speaking style. In everyday usage, we actually produce speech sounds with various speaking styles. Different styles may yield different acoustic measurements. This study attempts to examine pitch and formant trajectories of eleven English vowels produced by nine American males in order to understand acoustic variations depending on clear and conversational speaking styles. The author used Praat to obtain trajectories systematically at seven equidistant time points over the vowel segment while checking measurement validity. Results showed that pitch trajectories indicated distinct patterns depending on four speaking styles. Generally, higher pitch values were observed in the higher vowels and the pitch was higher in the clear speaking styles than that in the conversational styles. The same trend was observed in the three formant trajectories of front vowels and the first formant trajectories of back vowels. The second and third trajectories of back vowels revealed an opposite or inconsistent trend, which might be attributable to the coarticulation of the following consonant or lip rounding gestures. The author made a tentative conclusion that people tend to produce vowels to enhance pitch and formant differences to transmit their information clearly. Further perceptual studies on synthesized vowels with varying pitch and formant values are desirable to address the conclusion.

**Keywords:** English vowels, pitch, formants, trajectory, American males, speaking styles

#### 1. 서론

지금까지 영어모음 발음을 다룬 많은 연구들이 주로 실험실에서 또렷하게 발음한 음성을 대상으로 발표되었다. 또렷한 발음은 음성분석을 하기도 쉽고 결과도 안정적으로 나타나 연구하기에 아주 이상적인 모델이다. 하지만, 일상생활에서 음성의 목표는 자신의 의사를 상대방인 청자에게 전달하는 것이기 때문에, 상대방이 어떤 사람인가에 따라, 또 환경이 어떤가에 따라 늘 변한다(양병곤, 2008a). 예를 들어, 늘 대화하는 친

구가 아닌 잘 알아듣지 못하는 노인에게 말할 때, 또는 조용한 사무실이 아닌 시끄러운 시장에서 대화해야 할 때, 화자는 평소의 발화방식을 바꿔서 발음하게 된다. 덧붙여, 방언이나 개인발음습관에 따른 언어적 차이를 보이기도 하지만 화자의 감정적인 상태나 상대방의 조건에 따라 비언어적인 차이를 보이기도 한다. 예를 들어, 때로는 감정에 복받쳐 아주 높은 피치로 발음하기도 하고, 때로는 우울한 기분을 반영하듯 낮은 목소리로 피치의 변화를 보이기도 한다. 영어모음은 자음에 비해 화자의 정서와 관련된 음향적인 특징을 많이 간직하고 있어 이런 다양한 발화방식에 따른 음향적 변화를 잘 살펴볼 수 있다.

보다 듣기 좋은 음성을 만들기 위한 알고리즘을 개발하고, 또렷한 음성 훈련이나 음성처리에 필요한 기술을 지원하기 위한 목적으로 Smiljanić & Bradlow (2010)는 또렷한 발화방식에 따라 화자와 청자의 발음의 특징이 어떻게 변하는지 기존의

1) 부산대학교, bgyang@pusan.ac.kr

이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음.

접수일자: 2012년 1월 25일

수정일자: 2012년 3월 10일

게재결정: 2012년 3월 12일

연구들을 개괄적으로 살펴보고, 해결되지 않은 문제점들과 앞으로의 연구 방향을 제시하고 있다. 그들에 따르면 녹음 자료가 실제상황의 대화라기보다는 실험실에서 연구자의 지시에 따라 녹음되면서, 화자마다 다르게 지시사항을 이해하게 되어 차이를 보일 수 있고, 녹음 대상자도 언어적인 훈련을 받은 사람과 받지 않은 사람에 따라 달라질 수 있음을 지적했다. 녹음은 보통 한번은 대화체로 한번은 또렷한 발음으로 두 번씩 목록을 읽게 하는 연구가 많았다(Krause & Braida, 2002; Ferguson & Kewley-Port, 2002; Ferguson, 2004; Smiljanić & Bradlow, 2005). 이런 서로 다른 발화 방식을 영어로는 conversational speaking style과 clear speaking style로 구분하여 표현하는데, conversational은 casual(평상 발음)로 나타내기도 한다. 영어사전의 단어별 정의를 이용하면, casual한 발음은 평소에 늘 하던 방식으로 발음하고 특별히 세부적인 발음에 신경을 쓰지 않는 반면, clear한 발음은 상대방이 잘못 알아듣지 않도록 신경을 써서 선명하고 또렷하게 하는 발음이라고 할 수 있다. 이 논문에서는 각각을 대화체 발음과, 또렷한 발음으로 나누어 표현하기로 한다. Smiljanić & Bradlow (2010)에 따르면 기존 연구에서 또렷한 발음을 실현하기 위해 참가자에게 청각장애가 있는 사람이나 비원어민 화자에게 말하는 것으로 상상하고 또렷하고 정확하게 발음하라고 지시했고, 이러한 지시사항에 따라 녹음한 뒤 음향적인 분석을 해본 결과 큰 차이를 보이지 않았다는 결과도 제시했다 (Wassink, Wright & Franklin, 2006; Smith, 2007; Uther, Knoll & Burnham, 2007). 이에 반해, 서로 다른 발화방식의 발음을 유도해 내기 위해, Harnsberger, Wright & Pisoni (2008)는 생략된 발음 (reduced), 낭독 발음 (citation), 과장된 발음 (hyperarticulated)의 세 가지 발화방식으로 7개 단어 전후의 영어문장을 녹음하게 했다. 이들의 연구에서 생략된 발음은 녹음 도중에 단기 숫자 기억 실험에 집중하게 함으로써 참가자가 자신의 발음의 정확성에 신경을 쓰지 않도록 하고, 낭독 발음은 실험실에서 읽는 음성으로, 과장된 발음은 문장을 한 번 더 세심하게 읽는 것으로 했다. 이들은 녹음된 음성을 한 음성학자가 듣고 그 차이가 얼마나 되는 지 백분율 (%)로 평가한 점수를 근거로 10명 중 6명이 세 가지 다른 방식으로 구분하여 발음해낼 수 있었고, 모든 화자가 일관성 있게 발음하지 못한 숫자 기억 방식은 좀 더 개선할 여지가 있음을 보고 했다. 이들의 연구에서 참가자의 음성에 대한 비교평가가 결론에 영향을 주기 때문에 매우 중요한데, 단 한 명의 음성학자의 평가를 사용했고, 자세한 평가기준이나 개인의 편차 등에 대한 설명이 부족하여 그런 평가의 타당성이나 신뢰도 면에서 적절한 지는 의심스럽다. 예를 들어, 그 음성학자는 생략된 발음과 낭독 발음의 차이를 6%에서 91%까지 질적으로 다르다고 평가한 반면, 생략된 발음과 과장발음, 낭독 발음과 과장된 발음은 모든 화자에 대해 모두 100% 질적으로 차이가 난다고 기록하고 있다

(Harnsberger 등, 2008, p. 327, Table 1 참고).

미국인 남성들의 영어모음 포먼트에 대한 궤적을 연구한 양병곤 (2009)에서는 역동적으로 변하는 포먼트의 음향적 특징의 중요성을 인식하고, Hillenbrand, Getty, Clark & Wheeler (1995)가 인터넷에 게시한 28명의 미국인 남성 화자들이 발음한 모음의 시작과 끝부분을 제외하고, 상대적으로 비슷한 구간인 6개 지점에서  $f_0$ , 처음 세 개의 포먼트값, 강도값 등을 측정했다. 이런 연구 방법은 모음의 포먼트 측정값을 한 지점에서만 선택하여 나타낼 때 놓치기 쉬운 중요한 동적인 정보를 살펴볼 수 있는 장점이 있기 때문에 이 연구에서도 그런 분석 방법을 적용했다. 양병곤 (2009)의 연구 결과에서는 포먼트 궤적을 통해, F1은 전설모음과 후설모음을 구분하는데 1차적인 변수로 나타났고, F2는 따라오는 자음인 /d/를 지향하는 공동조음의 영향을 관찰했고, F3은 모음별로 주목할 만한 차이를 보이지 않았다고 보고 했다. F1과 F2로 나타낸 모음공간에 각 포먼트의 궤적을 그려서 살펴본 결과 시작부분에서 차이를 보였고, 끝부분은 따라오는 자음을 향한 동작이 전설모음보다는 후설모음의 이동거리가 길게 나타났다. 덧붙여, 모음마다 지속시간과 피치값을 구해 비교해본 결과, 대체로 개방모음이 상대적으로 지속시간이 길었고, 고모음의 피치값이 높았으며, 강도에서는 큰 차이를 보이지 않았다고 보고했다. 이 논문에서는 9명의 미국인 남성이 또렷한 발음과 대화체로 두 가지 다른 높이의 음계를 실어 발음했을 때 모음 피치와 포먼트의 궤적에 어떤 음향적인 변화를 보이는지 전체적인 경향을 살펴보고자 한다. 이러한 발화 방식에 따른 음향적인 궤적의 변화는 기존의 또렷한 발음으로 연구된 결과에 덧붙여, 다양한 환경에서 자동으로 음성을 인식하는 장치의 개발과 또렷한 발음을 내기위한 전략을 찾는 데 필요한 기초자료를 제공할 것으로 기대된다.

## 2. 연구 방법

### 2.1 피험자와 녹음자료

피험자는 미국인 남성 9명으로 뉴욕과 뉴저지 출신이 3명이고, 코네티컷 주 출신이 2명, 미주리와 오하이오, 뉴올리언즈, 텍사스 주 출신이 각각 1명이다. 이들은 Yale대학(원)생과 Haskins Lab의 연구원으로 평균 연령은 26.2세이며 18세에서 38세까지 다양하다. 이들의 키의 범위는 168 cm에서 185 cm이었으며, 평균은 181.1 cm로 다소 큰 편에 속한다. 이들 가운데 4명은 언어학을 전공으로 졸업했거나 재학 중이며, 음성학에 대한 기본적인 지식을 갖추고 있었다. 이들이 구사하는 방언의 특징을 알아보기 위해 Hillenbrand 등 (1995)에서도 논의된 바와 같이 awed - odd의 두 모음을 구분하여 듣고 발음할 수 있는지에 대해서 질문했는데, 2명의 대학생을 제외하고는 모두 구분한다고 응답했다. 녹음 자료는 Whalen, Magen,

Pouplier, Kang & Isakarus (2004)에서 /hVd/의 문맥에서 사용한 11개의 모음이 들어간 단어들이다 (heed [i], hid [ɪ], aid [eɪ], head [ɛ], had [æ], hud [ʌ], who'd [u], hood [ʊ], owed [oʊ], awed [ɔ], odd [ɑ]). 이 단어의 일부는 Peterson & Barney (1952)나 Hillenbrand 등 (1995)의 고전적인 연구에서도 쓰인 /hVd/ 환경에 들어가 있는데, 그 중에서 일부는 초성인 h 자음이 없어서 발음하기 편하고 일상생활에서 많이 나타나는 단어들을 사용했다. 이 단어들은 “I say hVd (So), hVd (Do), hVd (So), hVd (Do), again.”이라는 문장 속에 넣어 음의 높낮이에 따라 어떤 차이를 보이는지 관찰하기 위해 음악에 사용되는 음계 가운데 높낮이의 차이를 어느 정도 나타낼 수 있는 솔음과 도음을 들어 설명하며, 높은 음조, 낮은 음조의 순서로 두 번씩 발음하게 했다. 발화방식으로는 대화체와 또렷한 발음의 두 가지 형태로 나누어 두 번씩 목표음을 발음하게 했다. 발음 목록은 A4용지 두 장에 대화체용과 또렷한 발음용으로 각각 나눠서 지시문과 25개의 단어를 인쇄하여 녹음에 활용했다. 25개 단어에서 처음 3개 단어 heed, who'd, odd는 발음에 적용되지 않은 상태로 보고 목록에는 넣었지만 분석에서는 제외된 자료이고, 나머지 22개는 11개의 모음이 들어간 단어를 각각 두 번씩 배치했다. 녹음 순서는 대화체 음성을 먼저하고, 또렷한 발음을 뒤로 미뤘다. 사전 연구에서 또렷한 발음을 먼저 했을 경우에는 여러 번 발음을 하면서 일종의 습관이 형성되어, 대화체 음성을 발음하다가 갑자기 또렷한 발음으로 바꾸어 발음하는 피험자가 있었기 때문이다. Smiljanić & Bradlow (2010)에서도 지적된 바와 같이 모호한 지시사항은 화자들의 개인적인 가정이나 기준에 따라 발음의 변화가 많아지기 때문에 가능하면 변화를 줄이기 위해 구체적인 발음 요청을 했다. 대화체 음성에 대한 지시사항에서는 아주 친한 친구에게 전화로 빨리 말하듯이 발성해 달라고 했고, 일부 화자가 천천히 또렷하게 말하는 경우에는 실험자가 들어보고 다시 좀 더 빠르게 발음해 달라고 부탁하기도 했다. 또렷한 발음은 어린아이나 잘 듣지 못하는 사람에게 시끄러운 거리에서 말한다고 상상하며 발음해 달라고 했다. 특히, 고모음 보다는 입벌림이 많은 저모음이 들어간 had, odd가 나오는 경우에는 Harnsberger, Wright & Pisoni (2008)에서 과장된 발음을 얻는데 사용한 방법과 같이 한 번 더 입벌림을 크게 하고 또렷하게 발음해 달라고 요구했는데, 대체로 피험자들은 대화체보다는 입술이나 입벌림 동작이 눈에 띄게 달라졌음을 관찰할 수 있었다.

녹음은 방음이 된 녹음실에서 피험자가 Sennheiser PC330 헤드셋을 착용하고, PCM-M10 (Sony) 녹음기에 Stereo LPCM (Linear Pulse-Code Modulation) 방식으로 16비트의 진폭해상도에 22050 Hz의 표본속도로 녹음한 다음, SENS RF711 (Samsung)로 옮겨 분석에 사용했다.

## 2.2 음향분석

음향분석과 처리도구로 Praat(version 5.2.21)와 GoldWave (v5.58)를 이용했고, 포먼트 분석은 이전의 연구를 참고하여 (양병곤, 2008a), 양병곤 (2009)에 제시된 분석용 스크립트를 약간 수정하여 구했다. 분석과정을 간단히 서술하면, 먼저 녹음기에서 받은 음성자료를 GoldWave를 이용해 mono로 바꾸고, 각 화자별로 4번의 발음을 한 단위로 저장했다. 이어서, Praat를 열어 폴더자료 읽기 스크립트로 음성파일을 모두 개체창에 불러온 다음, 개체창의 음성분석 스크립트를 이용하여 선택된 음성을 편집창에 열어서 스펙트로그램과 피치곡선 음성파형 등을 근거로 연구자가 지정한 모음 구간을 6등분하여 시작점을 포함하여 7개 지점의 피치와 F1, F2, F3로 나타나는 세 개의 포먼트를 구했다. 음계를 높은 경우와 낮게 한 경우가 두 번씩 나타나기 때문에 측정값은 두 개 중 하나씩만 구했고, 나머지 하나의 음향적 궤적을 참고로 했다. 대체로 또렷한 발음은 첫 번째 두 개의 발음에서 구했고, 대화체는 두 번째 두 개의 발음에서 구하는 원칙을 따랐지만, 일부 음성에서는 직접 들어보고 두 번의 발화 중 더 또렷하거나 대화체로서 자연스러운 음성을 택했다. 포먼트값을 구하는 기본 설정은 5000 Hz까지 5개를 정해두고, 두 개의 포먼트가 가까이 접근하는 모음인 [oʊ, ʊ, u] 등에서는 5.5나 6을 지정하여 분리된 값을 가질 수 있도록 수정하여 값을 구했다. 대체로 남성의 포먼트는 Praat에서 검은 띠로 나타나는 에너지 중심점을 지나가는 추정값에 에러가 적지만, 가끔씩 F2나 F3값에서 잘못된 추정값을 보인 경우에는 바로 수정하였다. 피치값에서도 아주 적은 경우이지만, 피치반 값 에러 (pitch halving error)로 인해 실제 f0값의 반으로 처리된 값들이 포함되어 있어서 이를 수정하였다 (양병곤, 2009). 측정지점은 연구자가 선택한 모음구간의 시작부분과 끝부분에서 음성파형의 중앙부를 향해 각각 22.5 ms를 들어간 지점의 지속시간을 구한 다음 이를 6등분하여 구한 7개의 시간점에서 좌우 12.5 ms를 포함하는 25 ms의 창에서 평균 포먼트값을 구했다. 참고로 측정된 모음의 지속시간은 또렷한 발음이면서 높은 음계로 발음한 경우는 평균 274 ms (표준편차, 77 ms), 낮은 음계로 발음한 경우는 평균 263 ms (표준편차, 68 ms)를 나타냈고, 대화체 발음이면서 높은 음계로 발음한 경우는 평균 195 ms (표준편차, 50 ms), 낮은 음계로 발음한 경우는 188 ms (표준편차, 46 ms)를 보였다. 측정 위치 가운데 3번째는 모음구간의 1/3지점에 해당하며 기존의 연구에서 사용한 측정값과 비교해 볼 수 있다 (Peterson & Barney, 1952; Hillenbrand 등, 1995; Yang, 1996). 덧붙여, 25 ms라는 측정구간은 여러 번 측정하면서 임의로 정한 값이다. 이 측정구간을 좁히면 포먼트값에 에러가 많아지고, 넓히면, 인접한 값들이 평균값에 영향을 주어 원래의 값을 대표하지 않을 수도 있다. 음성은 우리가 측정하듯이 단절된 구간의 연속이라기보다는 발화자가 조음동작 목표를 가지고 여러 개의 조음 기관을 움직이면서 발음하기 때문에 단절된 시간점보다는 어느 정도의 지속시간으로

된 구간을 선택하여 구간 평균값을 구해야 에러를 줄일 수 있다. 앞으로 가장 타당한 포먼트 평균값을 측정할 수 있는 지속시간에 대해 보다 정밀한 연구가 필요하다.

이렇게 수집한 피치, 포먼트값의 자료는 22,176개 (9명의 화자 x 11개 모음 x 4개 측정값 x 7개 측정지점 x 4가지 발화방식 x 2회 반복)이다. 이 값들은 마이크로소프트 엑셀을 이용해서 각 모음별로 정돈한 뒤 모든 화자에 대해 측정지점별로 평균값을 구해 R (v2.14.0) 그래프로 나타냈다. 통계적인 비교 분석은 측정지점별로 할 수 있으나 아직까지 피치나 포먼트값의 절대적인 차이가 지각적으로 얼마만큼 유의미한 결과를 보일 것인지 참고할 만한 결과가 없기 때문에 다음으로 미루기로 한다. 다시 말해서, 통계적으로 유의미한 차이가 난다고 하더라도 양병곤 (2006)의 합성음 연구에서와 같이 AX구별방법을 이용하여 연속된 신호의 차이를 지각적으로는 느끼지 못하는 범위를 구해본 결과 9개모음에 대해 어떤 모음인가에 따라 다소 차이는 있지만, 평균적으로 미국인 남성들은 F1에서 102 Hz, F2에서 220 Hz, F3에서 457 Hz의 차이에도 두 개의 합성음을 동일하게 지각하는 결과를 보였고, 미국인 여성들은 각각 115 Hz, 244 Hz, 565 Hz의 범위를 보였다. 이러한 기준은 각 주파수 영역에 대해 사람의 지각이 선형적으로 변하기보다는 지각척도의 경향과 같이 주파수가 높아질수록 비선형적인 분포를 보일 것으로 예상된다. 우선은 비선형적인 분포치로 변환하여 통계 처리하는 방법이 가능하지만, mel (Fant, 1973)이나 bark (Zwicker & Terhardt, 1980)도 적용할 만한 완전한 지각척도라고 하기에는 아직 이라고 생각한다.

### 3. 결과 분석 및 논의

결과 분석은 먼저 각 모음별로 구한 평균 피치 궤적을 살펴보고 이어서 포먼트 궤적을 다루고자 한다. 숫자로 된 데이터를 모두 제시하기 보다는 그림으로 나타내어 발화방식별로 음향적인 특성에서 어떤 변화를 보이는 지 비교해 보기로 한다. 모음은 11개를 한꺼번에 하나의 그림으로 그려서 논의하기 보다는 세부적인 부분을 어느 정도 자세히 나타내기 위해 전설모음과 후설모음 집단으로 각각 나눠서 다룬다.

#### 3.1 또렷한 발음과 대화체로 발음한 전설모음의 피치 궤적

<그림 1>은 9명의 남성 화자들이 또렷한 발음과 대화체로 발음한 5개 전설모음의 피치 궤적을 보여준다.

<그림 1>에서 보면 또렷한 발음과 대화체에서 각 모음별로 뚜렷한 피치 궤적의 차이를 관찰할 수 있다. 먼저 음계를 높게 발음한 모음과 낮게 발음한 모음의 주파수값에서 차이를 보이고 있다. 높은 음계로 발음한 전설모음의 피치값의 평균은 148 Hz이고, 낮은 음계로 발음한 경우는 136 Hz이다. 전설

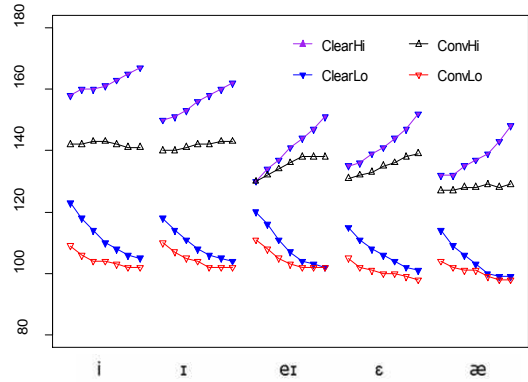


그림 1. 미국인 남성 화자들의 또렷한 발음과 대화체로 발음한 전설모음의 피치궤적  
Figure 1. Pitch trajectories of front vowels produced by American male speakers in clear and conversational styles

모음을 대화체로 높은 음계로 발음한 경우는 108 Hz인데 비해, 낮은 음계로 발음한 경우는 103 Hz를 보였다. 측정지점별로 차이가 달라지기 때문에 개별 측정모음별로 차이를 구하여 평균을 내본 결과 또렷한 발음에서는 평균 39 Hz 높게 발음한 반면, 대화체 발음에서는 33 Hz를 보였다. 또렷한 발음에서는 보다 많은 노력을 기울이게 되어 대화체보다 6 Hz 더 많은 변화를 보였을 것으로 추정된다. 또렷한 발음을 높은 음계로 발음한 경우와 대화체 방식을 높은 음계로 발음한 차이는 평균 11 Hz이고, 또렷한 발음을 낮은 음계로 발음한 경우와 대화체를 낮은 음계로 발음한 음향적인 차이는 평균 6 Hz이다. 이는 또렷한 발화방식에서 화자마다 특별한 조음동작을 활용하여 결과적으로는 피치 차이를 대화체 보다는 거의 두 배에 이를 정도로 많이 나타내게 했다고 말할 수 있다. 각 모음별로 살펴보면 전설 고모음에서 저모음으로 갈수록 또렷한 발음에서 피치값이 점점 낮아짐을 알 수 있고, 이런 변화는 대화체로 높은 음계를 발음한 모음에서도 유지되는 편이나 낮은 음계에서는 그다지 뚜렷한 차이를 보이지 않으며 대화체 낮은 음계에서는 대체로 비슷한 높이를 보이고 있음을 알 수 있다. 단모음과는 달리 이중모음 [er]에서는 또렷한 발음과 대화체 높은 음계의 궤적의 출발점이 일치하고 낮은 음계에서는 끝점이 일치하고 있다. 대체로 높은 음계로 발음한 경우에는 시작점보다는 끝점이 올라가는 경향을 보이고 있고, 낮은 음계에서는 끝점이 내려가는 경향을 보이고 있다. 이러한 경향은 개인적인 발화방식에 따라 달라질 수 있는 부분이지만, 피치와 관련된 조음 동작을 이용하여 서로 다른 방향으로 향하게 함으로써 결과적으로는 피치값의 차이를 발화방식의 구별에 활용한 것으로 분석해 볼 수 있다. Liberman & Pierrehumbert (1984)는 영어 문장의 중간에서는 추상적인 기준점을 두고 그곳까지 서서히 내려가다가, 문장 끝에서는 예상보다 상당히 내려가는 피치를 보이는데, 이것을 final

lowering으로 불렀다 (Truckenbrodt, 2004). 그림에서 피치값이 서로 모이는 부분인 일종의 추상적인 기준점으로 작용했을 수도 있다. 이 연구에서는 발화문장의 시작부에 I say를 넣고 가운데 목표단어를 4개 넣은 뒤 문장 끝에는 now를 배치하였기 때문에 네 번째 단어라도 어말에 따른 피치의 변화는 큰 영향을 주지 않았던 것으로 판단되지만, H-L의 되풀이되는 음높이의 두 개의 어구로 나누어 발음하는 과정에서 상대적으로 H로 발음한 단어의 모음 끝의 피치는 올리고, L로 되는 부분은 끝을 내려서 발음했을 여지도 남아 있다. 앞으로 다른 발화환경이나 다른 성별의 화자에서도 이런 유형을 보이는지 연구해 볼 필요가 있다.

### 3.2 또렷한 발음과 대화체로 발음한 후설모음의 피치 궤적

<그림 2>는 9명의 남성 화자들이 또렷한 발음과 대화체로 발음한 6개 후설모음의 피치 궤적을 보여준다.

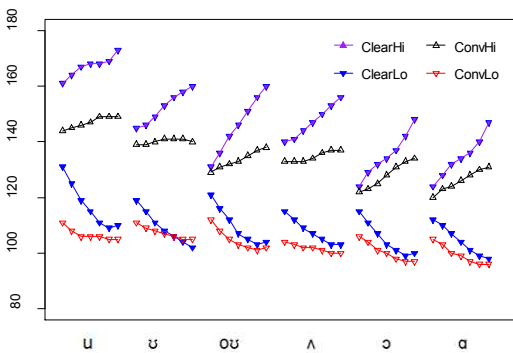


그림 2. 미국인 남성 화자들의 또렷한 발음과 대화체로 발음한 후설모음의 피치 궤적  
Figure 2. Pitch trajectories of back vowels produced by American male speakers in clear and conversational styles

<그림 2>에서도 전설모음의 피치 궤적과 비슷한 경향을 보이고 있다. 먼저 모든 측정시점값의 평균을 살펴보면 또렷한 발음을 높은 음계로 발음한 모음의 피치값의 평균은 143 Hz이고, 낮은 음계로 발음한 경우는 132 Hz이다. 대화체를 높은 음계로 발음한 경우는 전설모음보다 1 Hz 낮은 107 Hz인데 비해, 낮은 음계로 발음한 경우는 전설모음과 같은 103 Hz를 보였다. 전체적으로 또렷한 발음에서는 평균 38 Hz 높게 발음한 반면 대화체 발음에서는 32 Hz로 6 Hz의 차이를 보였다. 또렷한 발음을 높은 음계로 발음한 경우와 대화체를 높은 음계로 발음한 차이는 평균 12 Hz이고, 또렷한 발음을 낮은 음계로 발음한 경우와 대화체를 낮은 음계로 발음한 차이는 평균 6 Hz이다. 따라서 후설모음의 피치값에서도 화자가 조음동작을 통해 대화체 보다는 또렷한 발화방식에서 거의 두 배에

이를 정도로 변화를 보였다고 할 수 있다. 각 모음별로 살펴보면 후설 고모음에서 저모음으로 갈수록 또렷한 발음에서 피치값이 점점 낮아짐을 알 수 있고, 이런 변화는 전설모음에서와는 달리 후설고모음 긴장이완모음쌍인 [u]-[ʊ]와 이중모음 [ou]와 [ʌ]를 제외하고는 거의 같은 변화를 보이고 있다. 전설모음에서와 같이 이중모음 [ou]와 [ʊ]의 높은 음계로 발음한 모음에서 출발점이 거의 동일한 편이고 측정시점점에서 뒤로 갈수록 피치값이 올라가는 경향을 보이고 있다. 낮은 음계로 또렷하게 발음하거나 대화체로 발음한 경우에는 측정시점점에서 뒤로 갈수록 피치값이 내려가는 경향을 보이고 있다. 피험자들이 자신들의 발음을 발화방식에 따라 구별하기 위해 이런 피치차이를 낼 수 있는 조음동작을 사용했을 것으로 추정된다. 후설모음에서도 고모음에서 피치가 높고, 저모음으로 갈수록 피치값이 낮아지는 경향을 보이고 있다. 이러한 전설모음과 후설모음의 고유한 피치값의 분포는 세계 31개 언어에서 고모음 [i, u]와 저모음 [a, ʌ]의 피치값을 조사한 Whalen & Levitt (1995)의 연구에서도 밝혀진 바 있다. 이들은 각각의 언어에서 구한 f0값들은 모음을 발음하면서 화자가 의도적으로 성대의 진동을 조절하기 보다는 자연스럽게 차이가 나게 되었다고 제안한다. 이와는 달리 Diehl & Klunder (1989ab)나 Diehl (1991)은 고모음의 F1이 낮으므로 f0를 올림으로써 F1와 f0사이의 차이가 더 나게 하고 저모음에서는 그 반대로 f0를 내려서 높은 F1과 더 잘 구별할 수 있도록 의도적으로 발음한다고 주장한 바 있다. Whalen & Levitt (1995)은 이들의 주장이 윤상갑상근(Cricothyroid)을 조절하여 피치를 바꿀 수 있다는 사실에 근거했다면, 단순히 한 두 개의 근육의 조절로 피치를 바꾸기는 힘들고, 의도적인 발음을 하기 힘든 생후 6개월 된 유아들의 발음에서도 고유한 피치값이 나타나는 것 (Whalen, Levitt, Hsiao & Smorodinsky, 1995)을 고려해서 자연스런 현상으로 보는 것이 옳다고 주장했다. 이 연구에서는 두 가지 발화 방식을 가정하고 발음을 시켰기 때문에 자연스런 현상으로 나타나는 결과와 화자의 의도적인 노력이 함께 섞였을 것으로 판단된다. 고모음과 저모음 모두에서 또렷한 발음의 피치값이 높게 나타난 현상만을 볼 때, 화자는 의도적으로 조음동작을 통해 또렷한 발음의 피치를 높이려 했다고 결론을 내릴 수 있다. 의도적인 노력의 정도가 달라질 경우에는 어떤 피치의 변화가 나타날 지는 앞으로 연구해볼 필요가 있을 것이다.

### 3.3 또렷한 발음과 대화체로 발음한 전설모음의 포먼트 궤적

<그림 3>은 9명의 남성 화자들이 또렷한 발음과 대화체로 발음한 5개 전설모음의 포먼트 궤적을 보여준다.

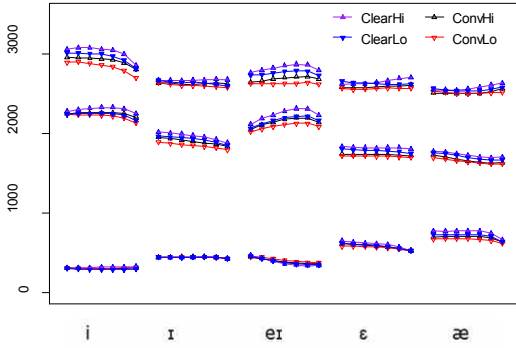


그림 3. 미국인 남성 화자들의 또렷한 발음과 대화체로 발음한 전설모음의 포먼트 궤적

Figure 3. Formant trajectories of front vowels produced by American male speakers in clear and conversational styles

이 그림은 네 가지 발화방식을 한꺼번에 그려서 다소 복잡하지만, 발음의 전체적인 음향적 특징을 한 눈에 비교하고 파악할 수 있는 이점이 있다. 먼저 전설모음에서는 기존의 모음 연구에서와 같이 입벌림 정도가 점점 많아지는 저모음으로 갈수록 F1이 높아지고 있고, F2와 F3은 고모음에 비해 저모음으로 갈수록 낮아지는 경향을 관찰할 수 있다 (Peterson & Barney, 1952; Hillenbrand 등, 1995; Yang, 1990, 1996).

이중모음 [eɪ]에서는 뒤에 있는 이완모음 [ɪ]의 위치를 향하여 세 개의 포먼트 동작을 살펴볼 수 있는데, <그림 3>에서는 첫 번째의 긴장모음 [i]와 두 번째에 있는 이완모음 시작부분의 중간쯤에 해당하는 위치로 끝나고 있음을 볼 수 있다. [ɛ]와 [æ]는 F1에서 가장 차이가 많이 나고 이어서 F2, F3으로 갈수록 차이가 적어진다. 그림에서 큰 차이를 보이지 않지만, 각 포먼트별로 모든 7개의 측정지점에서 구한 값들의 평균을 보면 또렷한 발음을 높은 음계로 발음한 경우의 F1은 평균 501 Hz이고, 낮은 음계로 발음한 경우는 485 Hz, 대화체를 높은 음계로 발음한 경우는 484 Hz, 낮은 음계로 발음한 경우에는 479 Hz로 발화방식에 따라 달라짐을 알 수 있다. F2에서도 각각 2015 Hz, 1964 Hz, 1942 Hz, 1904 Hz로 비슷하게 나타났고, F3에서도 각각 2753 Hz, 2711 Hz, 2671 Hz, 2633 Hz로 나타났다. 이런 경향은 또렷한 발화방식에 더 많은 조음노력을 기울여, 피치값에 이어 포먼트값에서도 대화체보다는 높은 값을 보여서 결과적으로는 발화방식 사이에 음향적인 차이를 두려는 경향이 반영됐다고 말할 수 있다. <그림 3>에서 이러한 경향은 특히 긴장모음 [i]와 이중모음 [eɪ]가 주도하고 있음을 알 수 있다. 덧붙여, 이완모음 [ɪ]의 F1에서는 거의 차이가 없으며 F3에서도 전반부에서는 차이가 적음을 볼 수 있다 (Yang, 2010). 일반적으로 F1값이 높아진 것은 입벌림 동작을 크게 할 때 F1값이 높아지고, F2는 혀가 앞으로 갈 때 높아지기 때문에, 또렷한 발음을 하기위해 화자들이 조음동작을

평소보다는 입을 더 많이 벌리고 대화체의 모음을 발음할 때 보다는 혀를 상대적으로 앞쪽 또는 뒤쪽으로 더 많이 움직여 발음한 결과로 해석할 수 있다 (Pickett, 1987). 덧붙여, 이러한 차이는 주파수가 낮은 포먼트값에서 보다는 높은 포먼트값에서 차이를 보였는데, 이는 미국인 남성이 발음한 모음의 포먼트값을 15% 정도 올리면, 미국인 여성의 포먼트값에 해당할 정도로 주파수대에 따라 균일하게 증가한 관계를 보였던 Yang (1996)의 연구 결과에서와 같이, 이 연구에 참여한 남성 집단 내에서도 조음 방법을 다르게 하여, 포먼트값이 어느 정도 비율적으로 증가했다고 볼 수 있다.

### 3.4 또렷한 발음과 대화체로 발음한 후설모음의 포먼트 궤적

<그림 4>는 미국인 남성 화자들이 또렷한 발음과 대화체로 발음한 6개 후설모음의 포먼트 궤적을 보여준다.

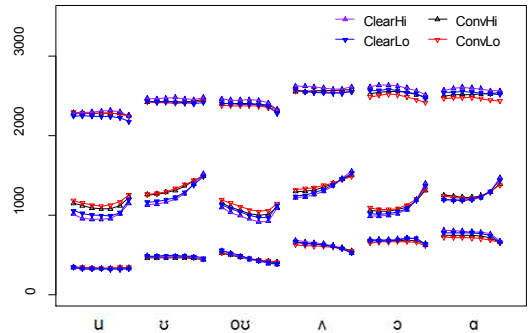


그림 4. 미국인 남성 화자들의 또렷한 발음과 대화체로 발음한 후설모음의 포먼트 궤적

Figure 4. Formant trajectories of back vowels produced by American male speakers in clear and conversational styles

후설모음에서도 기존의 모음 포먼트와 조음기관과의 관계를 제시한 연구 (Pickett, 1987)에서와 같이 입벌림 정도가 점점 많아지는 저모음으로 갈수록 F1이 높아지고 있고, F2는 혀의 이동위치를 보여주는 잣대로 대체로 내려가는 경향을 보이고 있다. F3에서는 원순모음의 영향으로 비원순 저모음에 비해 낮은 값들을 보이고 있는데, 긴장 고모음 [u]의 값이 가장 낮고, 이완모음인 [ʊ]와 이중모음 [ou]는 거의 비슷한 값을 보이고 있다. 여기서도 따라오는 자음 [d]의 지향점 때문에 측정지점마다 많이 달라지긴 하지만, 각 포먼트별로 모든 7개의 측정지점에서 구한 값들의 평균을 구해보면 높은 음계로 또렷한 발음을 한 경우의 F1값은 평균 605 Hz이고, 낮은 음계로 발음한 경우는 601 Hz, 대화체로 높은 음계로 발음한 경우는 589 Hz, 낮은 음계로 발음한 경우에는 579 Hz로 또렷한 발음에서 대화체로 갈수록 주파수가 낮아지는 경향을 보이고 있다. F2

에서는 이와 반대로 각각 1192 Hz, 1206 Hz, 1236 Hz, 1251 Hz로 또렷한 발음에서 대화체로 갈수록 주파수가 점점 올라가는 경향을 나타냈다. F3에서는 각각 2540 Hz, 2489 Hz, 2492 Hz, 2460 Hz로 높은 음계로 발음한 경우가 낮은 음계로 발음한 경우에 비해서 각각 51 Hz에서 32 Hz 더 높다. 이런 경향은 전설모음에서 보았던 또렷한 발화방식에 더 많은 노력을 기울여서, 대화체보다는 높은 포먼트값을 보이는 경향이 F1에는 적용되지만, F2에서는 따라오는 자음 [d]를 발음하기 위한 공동조음동작(Delattre, Liberman & Cooper, 1955)이나, 원순모음에서 입술동작이 첨가되면서 또렷한 발음과 대화체 사이의 F2값의 분포가 반대로 일어났을 것으로 추정된다. 발화방식에 따른 이런 F1과 F2의 값의 분포는 <그림 4>의 후설 고모음 [u, ʊ]와 이중모음 [oʊ]의 궤적에서 뚜렷이 볼 수 있다. 덧붙여, 전설모음에서와 마찬가지로 F1에서는 주파수가 낮은 포먼트값에서 보다는 높은 포먼트에서 차이가 많이 나지만, F2에서는 공동조음동작이나 원순모음에 따른 영향이 더 커서, 이런 경향을 관찰하기 어려운 것으로 여겨진다. 따라오는 자음을 달리 했을 때도 이런 경향을 보이는지, 또 원순성의 정도를 변화시켰을 때는 얼마만큼의 음향적 변화를 보이는지 좀 더 연구해볼 필요가 있다. 덧붙여, 화자마다 발화방식에 따라 자신의 조음동작을 조절하여 결과적으로 포먼트의 음향적인 변화를 보이려는 시도에서 나온 것으로 해석하는 성급한 결론을 내리기 보다는, 비록 대화체와 또렷한 발음의 포먼트 궤적의 방향에서 큰 차이를 보이지는 않았지만, 피험자의 녹음과정에서 발생하는 실험상의 오류가 영향을 미친 것은 아닌지도 검토해볼 필요가 있다. 이 논문에서 발화에 사용한 녹음용 자료의 순서는 높은 음계가 먼저 나오고 낮은 음계로 이동하는 과정에서 앞 절의 논의에서도 다루었듯이, 일부 화자들은 두 개의 H-L, H-L의 형태로 된 어구로 나눠 발음하고, 각각의 어구를 상승-하강의 울조로 실현하는 경우도 볼 수 있었다.

4. 요약 및 결론

이 연구의 목적은 미국인 남성 9명이 발음한 /h)Vd/환경에서 네 가지 다른 발화방식으로 발음한 11개의 영어모음의 피치와 포먼트의 궤적을 살펴보는 것이었다. 구체적으로 피험자들은 또렷한 발음과 대화체 발음으로 음계를 높인 경우와 낮게 한 경우로 나누어 발음했으며 피치와 포먼트의 측정값은 연구자가 선택한 모음의 시작점과 끝점에서 모음 안쪽으로 약간 들어간 지점의 지속시간을 구한 다음 이를 6등분해서 첫 측정지점을 포함한 7개의 지점에서 구한 피치와 포먼트의 궤적을 살펴보았다. 연구결과는 다음과 같다.

첫째, 피치 궤적에서는 발화 방식에 따라 각 모음별로 뚜렷한 피치값의 차이를 보였다. 일반적으로 높은 음계로 발음한 모음의 피치가 높게 나타났고, 낮은 음계로 발음한 모음이 낮

게 나타났다. 또한 고모음에서 저모음으로 갈수록 피치값이 점점 낮아졌다. 또렷한 발음은 대화체 발음보다는 이런 피치 차이가 더 많이 나타났는데, 피험자들이 조음동작을 통해 피치 차이를 더 많이 뱉으로써 더 선명한 발음을 실현하려고 시도했던 것으로 볼 수 있다.

둘째, 포먼트 궤적에서는 기존의 연구에서와 같이 F1은 입벌림 정도를 반영하고 있고, F2는 혀의 위치를 나타냈으며, F3은 입술동작을 반영하고 있음을 알 수 있었다. 발화방식에 따라 전설모음에서는 모든 포먼트에서 또렷한 발화방식에 더 많은 노력을 기울여서 포먼트값에서도 대화체보다는 높은 값을 보였는데 비해, 후설모음의 F1에서는 또렷한 발음이 상대적으로 높게 나타났지만, F2와 F3은 공동조음과 원순모음의 영향으로 또렷한 발음이 낮거나 일관성 있는 경향을 보이지 않았다.

이러한 결과에서 영어모음의 피치와 포먼트 궤적은 발화방식에 따라 다양한 변화를 보이며, 화자들은 자신의 정보를 전달하기 위해 조음동작을 통해 음향적인 변수의 차이가 크게 나도록 발음하려는 경향이 있다고 잠정적인 결론을 내릴 수 있다.

앞으로 이 논문에서 따라오는 자음에 의한 공동조음 요인이 또렷한 발화의 후설모음의 포먼트값을 대화체의 값보다는 내려가는 방향으로 영향을 준 것이라면, 다른 자음으로 된 문맥에서 살펴본 피치와 포먼트 궤적이나, 음계의 높낮이 발화 순서를 달리한 경우에 나타난 음향적 측정값의 궤적이나, 실제 상황에서 구한 자연스런 발화에서 구한 값들의 궤적에서는 어떻게 실현되는 지 연구해볼 필요가 있다. 또한 다양한 피치와 포먼트를 조합하여 합성한 음성을 들려주고, 어떤 방식의 발음인지 판단하게 하는 지각실험을 통해 발화방식에 따른 음향적인 특징이 지각 결과와 어떤 관계를 보이는 지 조사해볼 필요가 있다.

참고문헌

Yang, B. (2008a). Formant measurements of complex waves and vowels produced by students, *Speech Sciences*, Vol. 15, No. 3, 39-52.  
 (양병곤 (2008a). 복합음과 대학생이 발음한 모음 포먼트 측정, *음성과학*, 15권 3호, 39-52.)  
 Yang, B. (2008b). *English pronunciation: A new approach using a computer*. Busan: PNU Press.  
 (양병곤 (2008b). *영어발음: 컴퓨터를 활용한 새로운 접근*. 부산: 부산대학교출판부.)  
 Yang, B. (2009). Formant trajectories of English vowels produced by American males, *Phonetics and Speech Sciences*, Vol. 1, No. 3, 65-72.

- (양병곤 (2009). 미국인 남성이 발음한 영어모음의 포먼트 궤적, *말소리와 음성과학*, 1권 3호, 65-72.)
- Delattre, P. C., Liberman, A. M. & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants, *Journal of the Acoustical Society of America*, Vol. 27, 769-773.
- Fant, G. 1973. *Speech sounds and features*. Cambridge, MA: MIT Press.
- Ferguson, S. H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners, *Journal of the Acoustical Society of America*, Vol. 112, 259-71.
- Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners, *Journal of the Acoustical Society of America*, Vol. 116, 2365-2373.
- Harnsberger, J., Wright, R. & Pisoni, D. (2008). A new method for eliciting three speaking styles in the laboratory, *Speech Communication*, Vol. 50, 323-336.
- 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*, Vol. 97, No. 5, 3099-3111.
- Krause, J. C. & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility, *Journal of the Acoustical Society of America*, Vol. 112, 2165-2172.
- Liberman, M. & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length." In M. Aronoff, & R. T. Oehrle (Eds.), *Language, sound, structure: Studies in phonology presented to Morris Halle by his teacher and students* (157-223). Cambridge, MA: MIT Press.
- Peterson, G. & Barney, H. (1952). Control methods used in a study of vowels, *Journal of the Acoustical Society of America*, Vol. 24, 175-184.
- Pickett, J. (1987). *The sounds of speech communication: A primer of acoustic phonetics and speech perception*. Austin, Texas: pro-ed.
- Smiljanić, R. & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English, *Journal of the Acoustical Society of America*, Vol. 118, 1677-1688.
- Smiljanić, R. & Bradlow, A. R. (2010). Speaking and hearing clearly: Talker and listener factors in speaking style changes, *Language Linguist Compass*, Vol. 3, No. 1, 236-264.
- Smith, C. (2007). Prosodic accommodation by French speakers to a non-native interlocutor, *Proceedings of the XVIIth International Congress of Phonetic Sciences*, Saarbrücken, Germany.
- Truckenbrodt, H. (2004). Final lowering in non-final position, *Journal of Phonetics*, 32, 313-348.
- Uther, M., Knoll, M. & Burnham, D. (2007). Do you speak E-N-G-L-I-S-H? A comparison of foreigner- and infant-directed speech, *Speech Communication*, Vol. 49, 2-7.
- Wassink, A., Wright, R. & Franklin, A. (2006). Intraspeaker variability in vowel production: an investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers, *Journal of Phonetics*, Vol. 35, 363-379.
- Whalen, D. H. & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels, *Journal of Phonetics*, Vol. 23, 349-366.
- Whalen, D. H., Levitt, A. G. Hsiao, P. -L. & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9- and 12-month-old French- and English-learning infants, *Journal of the Acoustical Society of America*, Vol. 97, 2533-2539.
- Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., & Isakarous, K. (2004). Vowel production and perception: Hyperarticulation without a hyperspace effect, *Language and Speech*, Vol. 47, No. 2, 155-174.
- Yang, B. (1990). *Development of vowel normalization procedures: English and Korean*. Ph. D. Dissertation, The University of Texas at Austin.
- Yang, B. (1996). A comparative study of English and Korean monophthongs produced by male and female speakers, *Journal of Phonetics*, Vol. 24, 245-261.
- Yang, B. (2006). Discrimination of synthesized English vowels by American and Korean listeners, *Speech Sciences*, Vol. 12, No. 1, 7-27.
- Yang, B. (2010). Formant trajectories of English high tense and lax vowels produced by Korean and American speakers, *Korean Journal of Linguistics*, Vol. 35, No. 2, 407-423.
- Zwicker, E. & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency, *Journal of the Acoustical Society of America*, Vol. 68, 1523-1525.

• 양병곤 (Yang, Byunggon)

부산대학교 사범대학 영어교육과  
 부산시 금정구 장전동 30  
 Tel: 051-510-2619 Fax: 051-582-3869  
 Email: bgyang@pusan.ac.kr  
 Homepage: <http://fonetiks.info/bgyang>  
 관심분야: 음성학, 영어발음교육