말소리와 음성과학 제6권 제4호 (2014.12.31)
ISSN 2005-8063, pp. 91~99
www.speechsciences.or.kr
http://dx.doi.org/10.13064/KSSS.2014.6.4.091

발화방식에 따른 미국인 남성 영어모음의 스펙트럼 특성과 포먼트 대역

Spectral Characteristics and Formant Bandwidths of English Vowels by American Males with Different Speaking Styles

양 병 곤[1)]
Yang, Byunggon

ABSTRACT

Speaking styles tend to have an influence on spectral characteristics of produced speech. There are not many studies on the spectral characteristics of speech because of complicated processing of too much spectral data. The purpose of this study was to examine spectral characteristics and formant bandwidths of English vowels produced by nine American males with different speaking styles: clear or conversational styles; high- or low-pitched voices. Praat was used to collect pitch-corrected long-term averaged spectra and bandwidths of the first two formants of eleven vowels in the speaking styles. Results showed that the spectral characteristics of the vowels varied systematically according to the speaking styles. The clear speech showed higher spectral energy of the vowels than that of the conversational speech while the high-pitched voice did the same over the low-pitched voice. In addition, front and back vowel groups showed different spectral characteristics. Secondly, there was no statistically significant difference between B1 and B2 in the speaking styles. B1 was generally lower than B2 when reflecting the source spectrum and radiation effect. However, there was a statistically significant difference in B2 between the front and back vowel groups. The author concluded that spectral characteristics reflect speaking styles systematically while bandwidths measured at a few formant frequency points do not reveal style differences properly. Further studies would be desirable to examine how people would evaluate different sets of synthetic vowels with spectral characteristics or with bandwidths modified.

Keywords: English vowels, spectral characteristics, formant bandwidths, speaking styles

1. 서론

일반적으로 모음의 음향적 특징은 포먼트값에서 나타난다. 또한 포먼트값을 통해 입 벌림 정도와 혀의 위치 등 성도의 대략적인 모양을 추정할 수 있다. 지금까지의 대부분의 연구는 모음의 포먼트값 측정에 치중하였으며, 포먼트값의 세기를 나타내는 대역(bandwidth)값이나 모든 주파수에 걸쳐서 나타나는 스펙트럼 특성에 대해서는 발음통제나 측정된 수만 개의 데이터를 처리하는 과정의 어려움 때문에 국내외의 연구 자료가 부족한 편이다. 포먼트값은 어느 정도 발화방식에 대한 정보를 제공하기는 하지만, 발화방식에 대한 좀 더 자세한 정보는 제공해 주지 않는다. 일부 연구에서는 스펙트럼 정보에서 첫 번째 두 개의 배음차이를 이용해서 발화방식을 추정하였지만, 전체적인 스펙트럼 정보에 비해 일부분에 해당하기 때문에 음성합성에 응용하기 어렵다. 보통 포먼트 합성기를 이용하여 합성음을 만들 때는 한두 개 배음의 정보보다는, 각 포먼트의 세기 정보를 대역값이나 세기값으로 입력해주어야 하고, 이 정보를 통해 어느 정도 다양한 음질의 음성을 합성할 수 있다. 그러나 다양한 발화방식을 구현하려면, 좀 더 체계적인 정보가 필요하다. 덧붙여, 대역값은 포먼트의 세기를 나타내는 중요한 자료이긴 하지만, 현재의 음성분석소프트웨어에서 대역값의 측정은 여전히 문제가 많은 것으로 여겨진다. 어떤 문제점이 있는지를 살펴보는 것도 대역값을 이용하는 앞으로의 연구를 위해 필요할 것으로 생각된다.

이 논문에서는 9명의 미국인 남성이 또렷한 발음과 대화체

1) 부산대학교, bgyang@pusan.ac.kr

접수일자: 2014년 11월 7일
수정일자: 2014년 12월 12일
게재결정: 2014년 12월 14일

로 두 가지 높이의 음계를 실어 발음했을 때 모음의 평균 스펙트럼 에너지의 특성과 포먼트의 세기를 나타내는 대역값을 분석해보고자 한다. 이러한 연구는 지금까지 포먼트에 대한 연구에 덧붙여, 발화방식에 따른 모음 스펙트럼의 자세한 정보와 포먼트 대역측정에 필요한 기초적인 정보를 제공할 수 있을 것으로 기대된다.

2. 이론적 배경

Fant(1970)의 음원-여과기 이론(source-filter theory)에 따르면 사람의 발화된 음성(P(f))은 성대에서 발생하는 음원인 성문스펙트럼(S(f))과 성도의 모양과 크기의 변화에서 나타나는 성도공명(T(f)), 입술 밖으로 방출되면서 얼굴의 표면에서 작용하는 방사효과(R(f))가 포함되어 나타난다고 한다. <그림 1>은 이러한 관계를 자세히 보여준다(Borden, Harris, & Raphael, 2002, Figure 5.13; Kent & Read, 2002, Figure 2-5 & 2-16; Stevens, 1998, Figure 3.3; Mannell, 2014, Figure 8 & 10을 참고하여 만듦). T(f)부분은 이 연구의 스펙트럼 에너지의 이해를 돕기 위해 저자가 각 포먼트마다 분해하여 점선으로 나타내었다. 음원은 한 옥타브마다 12 dB씩 감소하게 되는 성문스펙트럼을 만든다. 방사효과는 입안의 기류가 입술 밖으로 분출하면서 얼굴 표면으로 만들어진 일종의 반사판이 옥타브당 6 dB씩 스펙트럼 에너지를 증가시키는 효과를 나타내어, 최종음성출력에서 스펙트럼 에너지가 옥타브당 6 dB씩 감소하는 모양으로 나타나게 된다고 한다. 그림의 첫 번째 음원스펙트럼의 급작스런 기울기에 비해 마지막의 최종 출력 스펙트럼은 높은 주파수로 갈수록 서서히 기울여지는 모양을 보이고 있다.

모음 발음에 대한 고전적인 연구는 주로 성도의 공명특성인 포먼트에 치중하여 왔다(Peterson & Barney, 1952; Hillenbrand, Getty, Clark, & Wheeler, 1995; Yang, 1990, 1996; Assman & Katz, 2000). 이들 연구의 영어모음에 대한 구체적인 포먼트값에 대해서는 Kent & Read(2002)나 강석한(2007)을 참고하기 바란다. 실제 음성합성에서 해당 모음의 포먼트값을 입력하고 대역값은 B1=60 Hz, B2=90 Hz, B3=120 Hz, B4=150 Hz 등의 임의의 값을 넣어도 자연스러운 모음으로 합성되기도 한다(Klatt & Klatt, 1990). 하지만, 두 개의 포먼트가 서로 접근하는 모음 포먼트에 대해서는 대역값을 좀 더 큰 값으로 조정해야 과잉 스펙트럼 에너지로 생기는 최종 합성음의 음성파형이 부자연스럽게 찌그러지거나 높아지는 현상을 방지할 수 있다. 이러한 현상의 원인은 <그림 1>의 두 번째 성도공명에서 각 포먼트값의 공명이 그림과 같이 겹쳐져 있어서 그 결과로 나타나는 스펙트럼 에너지는 주변 포먼트가 서로 모이는가 멀리 떨어져있는가에 따라 달라지기 때문이다(Kent & Read, 2002:28 Figure 2-17). 스펙트럼 에너지의 분포에 대해 Kent & Read(2002)가 설명한 기본적인 원리는 F1이 가장 낮은 주파수에서 가장 높은 주파수까지 영향을 주게 되므로 이 값이 내려가거나 올라가게 되면 F1 자체의 에너지도 내려가거나 올라가게 되며 동시에 F2 이상의 모든 전체 스펙트럼 에너지가 낮아지거나 높아지게 된다고 한다. 더욱이, 각각의 포먼트공명값이 합쳐져서 최종 출력발음이 결정되기 때문에, 두 개의 포먼트값이 서로 모이면 각 포먼트의 세기도 강하게 된다고 한다. 이런 관계를 생각해보면, 모음마다 포먼트가 낮은 주파수에서 F1과 F2가 서로 접근해 있거나(예, [o, u]), 보다 높은 주파수에서 F2와 F3이 서로 접근해 있는(예, [i]) 경우에 따라 스펙트럼 에너지의 분포도 달라짐을 알 수 있다.

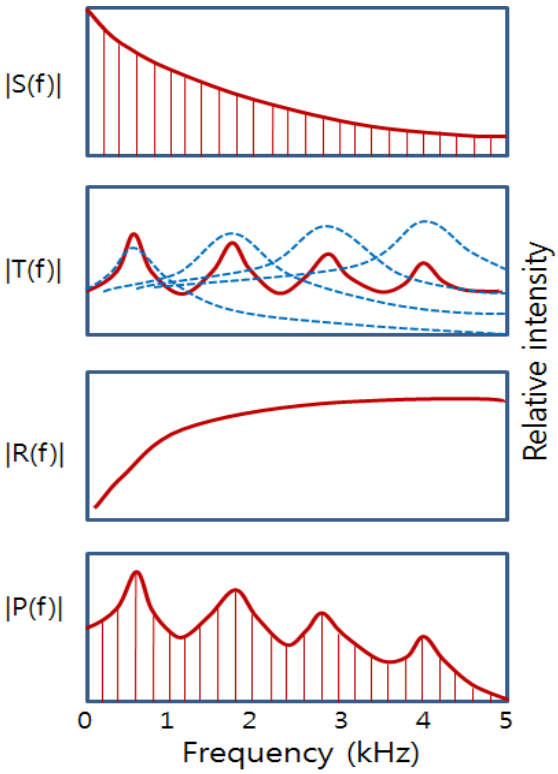

그림 1. 발음(P(f))의 음원(S(f)), 성도공명(T(f))과 방사효과(R(f))의 분해도. 성도공명특성은 개별포먼트로 분해하여 점선으로 나타내었음.

Figure 1. Schematic diagram of speech output(P(f)), source(S(f)), filter(T(f)) and radiation effect(R(f)). Dotted lines in the filter indicate individual formant transfer function. Adapted from Mannell(2014).

이런 각 포먼트의 정점의 세기를 직접 dB값으로 구하기도 하지만, 대역값을 구하여 나타내기도 한다. 대역값은 <그림 2>와 같이 가장 강한 에너지가 보이는 포먼트 중심주파수값에서 3 dB아래의 주파수 영역의 범위를 말한다(Mannell, 2014, Figure 9를 참고하여 만듦). 그림에서 B1은 B2에 비해 상대적으로 좁고 에너지의 세기가 높다. 대역값은 발화된 음성의 특정 시간지점에서 구하게 되는데, 순수하게 성도공명에 의한 개별 포먼트값의 크기를 반영하기 보다는, <그림 1>의 두 번째에

서 보이듯이 모든 포먼트전이함수가 반영된 실선으로 나타난 전이함수곡선(Transfer function curve)에서 구한다.

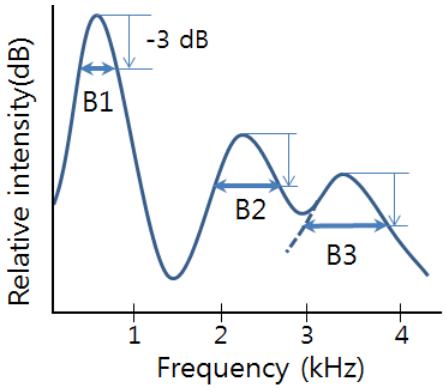

그림 2. 포먼트 대역값(B1, B2, B3)의 측정
Figure 2. Measurements of formant bandwidths(B1, B2, B3). Adapted from Mannell(2014).

이 때 해당 포먼트의 대역은 주변에 있는 여러 개의 포먼트 값이 겹쳐져 나타나기 때문에, 하나의 포먼트로 이뤄진 가상의 음성에 대해서는 쉽게 측정할 수 있지만, 실제 사람의 음성과 같이 여러 개의 포먼트로 나타난 <그림 2>와 같은 가상의 스펙트럼에서 B3을 측정할 때 왼쪽의 낮은 주파수의 경계지점은 임의의 선을 추정해서 대역값을 구해야 하는데 일관성 있는 값을 구하기 어렵다. Praat에서도 두 개의 포먼트가 가까이 접근하여 잘못된 포먼트 지점에서 구한 대역값이 타당하지 않는 값으로 나타나기도 한다. 이 연구에서는 다소 불안정한 대역값이지만, 스펙트럼 정보를 가장 간략하게 요약하여 보여주므로 일단 측정하여 서로 다른 발화방식에서 어떤 차이를 보이는지 또 어떤 측정상의 문제점이 있는지 살펴보고자 한다.

발화방식은 대다수의 연구가 또렷한 발음으로 준비된 문장이나 단어를 읽은 연구가 많았고, 일부는 대화체로 발음하여 발성의 차이를 비교했다(Krause & Braida, 2002;Ferguson & Kewley-Port, 2002;Ferguson, 2004;Smiljanić & Bradlow, 2005). 발화방식에 따라 음향적인 변수가 어떻게 달라지는지를 알아보기 위해 양병곤(2012)은 미국인 남성 9명이 또렷하게 발음하거나 대화체로 발음한 11개의 음성에서 피치값과 포먼트값의 궤적을 살펴본 결과, 피치궤적에서는 각 모음별로 뚜렷한 차이를 보였는데, 높은 음계로 발음한 경우는 높게, 낮은 음계로 발음한 경우는 낮게 나타났고, 또렷한 발음방식에서 대화체보다는 높은 피치값을 보였음을 보고했다. 한편 포먼트 궤적에서는 또렷한 발음방식의 전설모음에서는 모든 포먼트값에서 대화체보다는 높은 값을 보였지만, 후설모음에서는 F1에서 다소 높지만, F2에서는 오히려 낮은 경우도 있어서 일관성이 없다고 보고했다. 실제 포먼트측정값은 모음의 특징을 주파수값으로 요약하여 나타내긴 하지만, 서로 다른 발화방식에 대해 스펙트럼상 얼마의 세기로 발화되었는지에 대해서는 보여주지 못할 것으로 예상된다. 따라서 이 연구에서는 이들 미국인이 발음한 음성자료에서 조사하지 못했던 모음의 포먼트 대역값과 스펙트럼 에너지에 대해 좀 더 상세하게 살펴보고자 한다. 이러한 연구는 포먼트 측정에서 찾을 수 없는 또 다른 음향적 특징을 살펴볼 수 있을 것으로 기대된다.

화자확인을 위한 음향적 변수를 찾는 과정에서 양병곤과 강선미(2002)는 스펙트럼 에너지를 이용하여 포먼트 정보에서 볼 수 없는 더 많은 정보를 추출하여 모델을 만들어 화자구별을 할 수 있는 방안을 모색했다. 이들의 연구에서는 개인별 발화에 대해 좁은 대역의 스펙트로그램을 만들어 3300 Hz까지 150 Hz마다 장기평균스펙트럼값(Long-term average spectra: Ltas)을 구했다. 좁은 대역 스펙트로그램은 넓은 대역 스펙트럼에 비해 시간마다 측정값의 변화가 다소 안정적이어서 실제 조음기관의 유연하고 느린 연속적인 변화를 반영하는 것으로 보았다. 일반적으로 음성분석은 단절된 구간의 정보를 분석용 창을 씌어서 처리하여 나타내며 주변값의 변화가 반영되지 않아 급작스럽게 변하는 값이 나타나는 경우가 많다. 이들의 연구에서는 비록 비슷한 발성기관을 가진 화자의 구별을 하기 위해, 이들이 발음한 숫자음에 포함된 모음 구간의 스펙트럼 정보의 상관관계와 절대값의 차이를 구했다. 이러한 관계와 차이로 화자를 확인하는데 어느 정도 도움이 되었지만 완전한 모델 설정에는 숫자음외에 또 다른 단어를 활용한 연구를 제안한 바 있다.

한편 Boersma & Kovacic(2006)는 크로아티아의 민속노래를 부르는 세 가지 방식의 스펙트럼 특징을 분석하여 이들 간에 어떤 차이를 보이는지 연구했다. 이들은 12명의 남성으로 이뤄진 직업가수들의 노래에서 스펙트럼 에너지를 측정할 때, 피치값의 직접적인 영향을 제거한 Ltas를 구한 다음 주성분분석과 가수들의 포먼트와 스펙트럼의 기울기를 통해 세 가지 발성방식의 차이를 구별해 낼 수 있었다고 보고했다. 그 결과 klapa 방식은 말하듯이 노래하며, ojkanje 방식은 산악에서 멀리까지 전해질 수 있는 아주 큰소리로 피치값의 두 배지점에 하나와, 3.5 kHz 주변에 넓게 나타나는 두 개의 스펙트럼 정점이 나타났고, tarankanje 방식은 성대를 짜듯이 누르고 비음이 섞인 평평한 스펙트럼을 보였다고 한다. 피치값을 수정한 이런 분석방식의 결과는 첫 배음이 아주 낮은 값으로 나타나는 부분이 보정되고, 포먼트 정점이나 포먼트간의 스펙트럼이 낮은 계곡부분의 갑작스런 변화부분이 제거됨으로써 부드러운 스펙트럼의 변화모양을 그려준다(Boersma & Kovacic, 2006의 FIG 1과 FIG 4 참고). 이러한 결과는 양병곤과 강선미(2002)에서도 지적했던, 넓은 대역스펙트럼의 급작스런 변화를 상당히 완화시키는 효과를 주게 된다. 이들의 연구에서 서로 다른 노래방식의 차이가 스펙트럼에서 잘 나타난 것과 같이 또렷한 발음과 대화체와 같은 발화방식의 차이도 스펙트럼에서 드러날 것으

로 예상된다.

3. 연구방법

3.1 참여자와 녹음자료

이 연구에 참여한 미국인 9명은 Yale대학(원)생과 Haskins Lab의 연구원으로 신체 특징과 방언에 대해서는 양병곤(2012)에 자세히 제시되어 있다. 서론에서도 언급했듯이 이전의 연구에서는 발화한 모음의 피치값과 포먼트값의 궤적을 발화방식에 따라 비교하였지만, 보다 중요한 스펙트럼 정보에 대한 자세한 연구가 필요하여 동일한 음성자료를 이용해 이 연구를 수행하게 되었다. 발화한 모음은 /(h)Vd/의 문맥에서 사용한 11개의 모음이 들어간 단어다(heed [i], hid [ɪ], aid [eɪ], head [ɛ], had [æ], hud [ʌ], odd [ɑ], awed [ɔ], owed [oʊ], hood [ʊ], who'd [u]). 이 가운데 이중모음이 포함된 aid [eɪ]와 owed [oʊ]는 측정지점에 따라 모음값의 변화가 생기지만, 전체지속시간에서 1/3 지점을 택하여 일관성 있게 측정하면 서로 비교가 될 것으로 여겨져 모두 포함했다. 이 단어들은 각 화자마다 "I say **(h)Vd**, (h)Vd, **(h)Vd,** (h)Vd, now"의 문장에 넣어 대화체로 두 번씩 전체 단어목록을 뒤섞어 발음했고, 이어서 또렷한 발음양식으로 두 번씩 뒤섞어 발음했다. 대화체는 친구에게 말하듯이 자연스레 말하게 했고, 또렷한 발음은 시끄러운 장소에서 귀가 어두운 사람에게 말하듯이 또렷하게 발음하되, 진하게 표시된 단어는 높은 음(솔)의 높이로 발음했고, 보통체의 단어는 낮은 음(도)로 발음하게 했다. 결국, 동일한 모음으로 만들어진 문장이 연이어 나타나지 않게 4개 단어들을 두 번씩 되풀이 발음하여 1인당 모두 16번씩 녹음이 되었고 녹음된 단어는 1584개다(9명 x11개모음x2가지발음방식x2번발음x4개단어).

3.2 자료분석

음향분석 도구는 Praat(version 5.3.83)을 이용했고, 포먼트 대역은 양병곤(2012)에 사용된 스크립트를 부록과 같이 변형하여 측정한 다음, 모든 화자별로 엑셀을 이용해서 평균값을 구했고, 스펙트럼 특성은 Praat의 Ltas(Pitch corrected)를 이용하여 구한 뒤, 모든 자료를 하나의 파일로 만들어 발화방식과 음높이, 혀의 전후위치에 따른 차이를 조사했다. 구체적으로 자료분석 절차를 살펴보면, 1584개의 발음에서 792개를 택해서 각 모음지속시간을 연구자가 선택하면 스크립트에 의해서 모음구간의 1/3에 해당하는 지점의 앞뒤로 25 ms구간이 선택되어 제1, 제2포먼트의 대역값(B1, B2)을 추출하고 이어서, 100 Hz간격으로 50개의 Ltas(Pitch corrected)를 추출하여 파일로 저장한다. 이들이 발화한 11개의 모음을 두 가지 발화방식으로 음높이를 바꾸어 발음한 4개의 발음에서 주로 첫 번째와 세 번째 발음을 분석대상으로 하고, 대역값이 문제가 있을 경우에는 두 번째와 네 번째의 발화를 분석했다. 이렇게 측정된 대역값에서 동일한 발화 두 번의 평균값을 구해서 최종대역값자료로 정했다.

대역값의 측정에서는 포먼트 설정을 어떻게 하는가에 따라 값이 달라지는 경향이 있었다. 구체적으로 <그림 3>은 1번 남성화자가 발음한 who'd의 음성파형과 스펙트로그램을 보여준다. 포먼트 설정에서 포먼트 개수를 5로 지정했을 때 B1은 196 Hz, B2는 498 Hz를 정보창에 나타냈는데, 포먼트 개수를 6으로 지정했을 때는 B1은 87 Hz, B2는 213 Hz로 나타났다. 이 모음 [u]는 F1과 F2가 가까이 접근하는 후설원순모음이라 6개로 지정했을 때 바른 값이 구해지므로, 두 번째의 값을 최종포먼트 대역값으로 선정했다. 이 연구에서는 후설원순모음인 경우에는 포먼트개수를 조정하여 적정값을 구하였다. 앞으로 포먼트값이 바르게 측정된 이후에 대역값이 적정하게 구해질 수 있다고 여겨지므로 대역값의 측정에 주의할 필요가 있다.

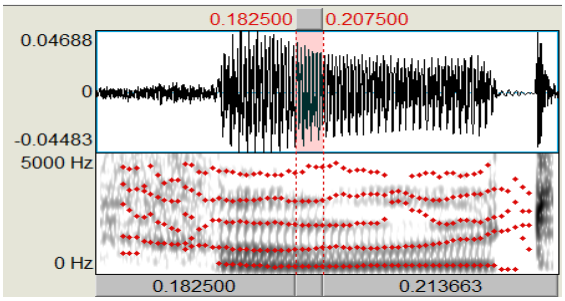

그림 3. 미국인 남성(m1)이 발음한 who'd의 음성파형과 스펙트로그램. 점선 사이의 선택부분에서 대역값(B1, B2)을 구했음.

Figure 3. Waveform and spectrogram of *who'd* produced by an American male(m1). Bandwidths(B1, B2) were measured at the highlighted section between dotted lines.

스펙트럼 에너지는 총 39600개의 수치값으로 나타나 있는 792개의 Ltas파일을 R의 열병합기능(cbind)을 이용해서 하나의 파일로 만든 다음 엑셀에서 각각 대화체와 또렷한 발음, 높은 음과 낮은 음, 전설모음과 후설모음, 단모음과 이중모음 집단으로 나누어 평균값을 구하고 그래프로 나타내어 비교했다.

최종대역값자료는 B1과 B2를 각각 396개씩 구했다. 이 과정에서 Praat에서 대역값을 측정할 때 여러 화자의 측정값 분포로 살펴보았을 때 명백한 에러로 판정되는 부분이 있어서, 일단은 250 Hz이상은 모두 새로 점검해보고 필요하면 수정했다. 이어서 대역값들은 각각 대화체와 또렷한 발음, 높은 음과 낮은 음, 전설모음과 후설모음, 단모음과 이중모음 집단으로 나누어 그래프로 결과를 나타내고, 통계적인 차이는 집단별로 정규분포를 가정할 수 없다고 판단하여 R(v.3.1.1)의 비모수통계인 Wilcoxon의 순위합 검정을 사용하여 비교했고, 유의수준은 0.05로 하였다.

4. 스펙트럼 에너지 분석 결과 및 논의

<그림 4>는 대화체와 또렷한 발음 방식에서 구한 스펙트럼 에너지의 평균곡선을 나타낸다.

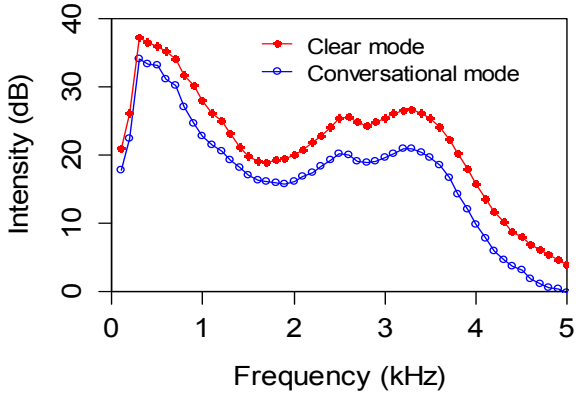

그림 4. 또렷한 발음과 대화체로 발음한 모음의 스펙트럼 특성 비교

Figure 4. Spectral characteristics of vowels produced in clear or conversational speaking styles

이 스펙트럼 특성 곡선을 보면 채워진 원으로 나타낸 또렷한 발음의 경우가 전반적으로 빈 원으로 나타낸 대화체 음성보다는 에너지가 높게 나타나 있다. 전체 평균값으로는 대화체가 16.9 dB이고, 또렷한 발음은 21.5 dB로 약 4.6 dB의 차이를 보인다. 값의 범위는 대화체가 -0.01 dB에서 34.0 dB까지 분포되어 있고, 또렷한 발음은 3.9 dB에서 37.1 dB로 분포되어 있다. 이러한 차이는 예상된 것으로 발화자가 더 또렷하게 발음하려는 시도에서 모든 주파수 영역에 걸쳐서 스펙트럼 에너지가 증폭된 것을 알 수 있다. 또렷한 발음과 대화체 음성의 차이가 많은 주파수는 900 Hz의 5.7 dB정점이 있고 살짝 내려갔다가 2500 Hz부터 5.1 dB이상이 4700 Hz까지 이어져 있다.

높은 음과 낮은 음으로 발음한 스펙트럼 에너지의 특징은

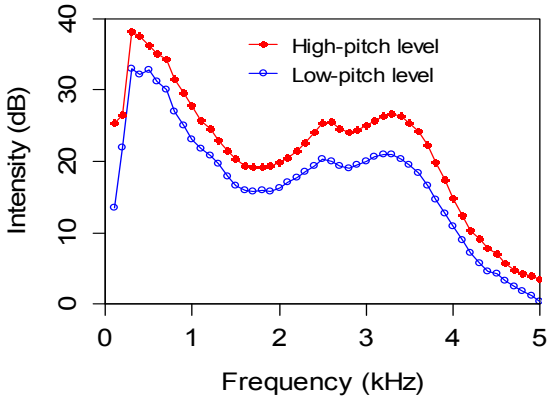

그림 5. 높은 음과 낮은 음으로 발음한 모음군의 스펙트럼 특성 비교

Figure 5. Spectral characteristics of vowels produced in high- or low-pitch levels

<그림 5>에 나타나 있다.

그림에서 보면 발음한 음높이가 높을수록 스펙트럼 에너지가 높게 나타난다. 전체 평균값으로는 높은 음에서 17.1 dB이고, 낮은 음에서는 21.4 dB로 약 4.3 dB의 차이를 보인다. 스펙트럼값의 변화 범위는 낮은 음에서는 0.3 dB에서 33.0 dB까지 분포되어 있고, 높은 음에서는 3.5 dB에서 38.1 dB까지 분포되어 있다. 이러한 차이는 근본적으로 음높이를 올리면서 전체적인 스펙트럼 에너지가 증폭되어서 생긴 것으로 보인다. 음높이에 의한 차이가 가장 많은 지점은 100 Hz에서 11.8 dB의 차이가 나고 이어서 300 Hz와 400 Hz에서 5 dB이상을 유지하다가 3 dB까지 내려갔다가, 2500 Hz에서 3800 Hz까지 다시 5 dB이상 차이를 유지했다. 첫 배음에서 큰 차이가 난 것은 Praat의 Ltas(Pitch corrected)분석 방식에서 보정이 완전하게 되지는 않았을 것으로 추정되는데 앞으로 다른 음높이를 보이는 대상자의 연구를 통해 검토할 필요가 있다.

마지막으로 전설모음 군과 후설모음 군으로 나누어 스펙트럼 에너지의 특성을 살펴보면 <그림 6>과 같다.

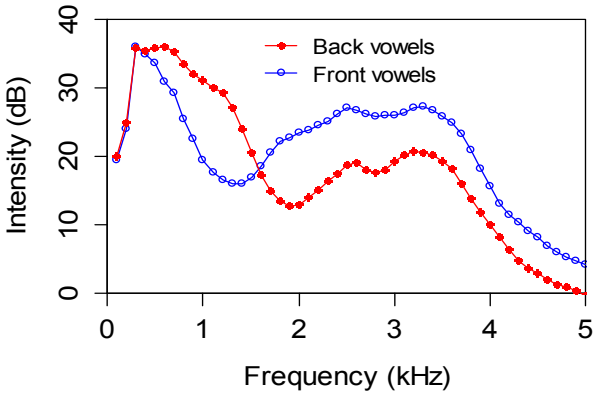

그림 6. 전설모음과 후설모음 군에 따른 스펙트럼 특성 비교

Figure 6. Spectral characteristics of front or back vowel groups

전설모음 군의 F1과 F2는 후설모음 군에 비해 떨어져 있어서 그림과 같이 낮은 주파수 영역에 스펙트럼 정점이 있고 후설모음 군은 상대적으로 높은 스펙트럼 에너지가 모여 있어서 (Kent & Read, 2002) 높게 나타난다. 하지만, 중간의 주파수 영역부터 높은 주파수 영역으로 갈수록 전설모음 군의 스펙트럼 에너지가 절대적으로 높게 나타난다. 전체평균값을 구해서 비교해보면 전설모음 군은 23.0 dB가 되고 후설모음 군은 18.1 dB를 보였다. 전설모음 군의 스펙트럼 에너지 범위는 4.2 dB에서 35.9 dB에 걸쳐 분포되어 있고, 후설모음 군은 -0.3 dB에서 35.9 dB 값을 보였다. 후설모음 군에서 전설모음 군을 뺀 차이는 1200 Hz에서 가장 큰 차이인 12.7 dB를 보였고, 600 Hz에서 1400 Hz까지는 5 dB이상의 차이를 보였으며, 1700 Hz에서 4700 Hz까지는 모두 -5 dB이상의 차이를 보였다.

Boersma & Kovacic(2006)는 크로아티아의 노래 방식에 따라 스펙트럼 정점의 위치나 모양이 달라지고, 피치값의 두 배 지점에 하나와 특정한 주파수 영역 주변에 스펙트럼 정점을 보고했는데, 노랫말에서 어떤 모음이나 자음이 더 많은가에 따라 결과가 달라졌을 것으로 예상된다.

지금까지 살펴본 스펙트럼 에너지의 분포 특성을 요약해 보면, 모음의 스펙트럼은 발화양식에 따라 차이를 보이며, 전설모음과 후설모음과 같이 발화모음의 특성에 따라 달라짐을 알 수 있다. 따라서 음성인식과 음성합성에서 이런 정보를 활용하면 발화특성을 어느 정도 추정할 수 있을 것으로 여겨지며, 동시에 합성음을 만들 때 화자의 발화양식의 특징을 반영하는데 활용할 수 있을 것으로 생각된다. 덧붙여, 전설모음과 후설모음 등으로 구분된 자극모음을 어떤 비율로 선택하는가에 따라 발화방식에 대한 연구결과가 달라질 수도 있음을 알 수 있다.

5. 포먼트 대역값 분석 결과 및 논의

포먼트 대역값 측정에 문제점이 있긴 하지만, 발화방식별로 구한 포먼트 대역값의 전체 총평균은 B1이 89.5 Hz이고, B2가 101.3 Hz로써 그 차이가 11.8 Hz로 나타났고, B1의 최소값과 최대값으로 나타난 범위는 17.0 Hz에서 269.0 Hz까지, B2는 26.0 Hz에서 233.0 Hz로 분포되었다. 이러한 결과는 F1이 F2에 비해 스펙트럼 에너지의 정점이 높음을 보여준다. 이러한 경향은 <그림 1>에서 제시된 음원의 스펙트럼이 비록 방사특성에 의해 증폭되었지만, 여전히 옥타브당 6 dB씩 떨어지는 경향이 그대로 반영되어 주파수가 상대적으로 높은 F2에서 포먼트 대역값이 높아지게 되었다고 할 수 있다. 덧붙여, 앞의 스펙트럼 특성 그림들에서 보듯이 스펙트럼 곡선도 모두 F1주변의 정점을 시작으로 해서 주파수가 증가함에 따라 서서히 스펙트럼의 세기가 낮아지는 모양에서도 확인할 수 있다.

이번에는 발화방식에 따라 포먼트 대역값의 분포를 살펴보자. 먼저 대화체와 또렷한 발음에서 구한 두 개의 포먼트 대역값의 분포는 <그림 7>에 상자도표로 나타나 있다.

이 그림에서 보면 포먼트 대역값에서 큰 차이를 보이지 않는다. 대화체 음성의 B1과 B2 평균은 각각 88.2 Hz와 103.6 Hz이고 또렷한 음성의 B1과 B2 평균은 각각 90.8 Hz와 99.0 Hz이다. 그림에서 상자 안의 굵은 띠는 중앙값(median)을 나타내고, 대화체 음성의 B1의 중앙값이 더 높게 나타나는데, 평균값(mean)에서는 오히려 또렷한 음성의 B1이 더 높게 나타낸다. 이러한 결과는 또렷한 음성의 극단치들이 평균값을 중앙값보다 높아지게 한 것으로 보인다. Wilcoxon의 순위합으로 집단간 통계적 차이를 검정해본 결과 B1에 대해서는 W=19570, p=0.978이었고, B2에 대해서는 W=18560, p=0.360으로 모두 유의미한 차이를 보이지 않았다. 여기서 포먼트 대역값이 작을수록 포먼트의 에너지가 높다(Fant, 1970)는 점을 고려해본다면,

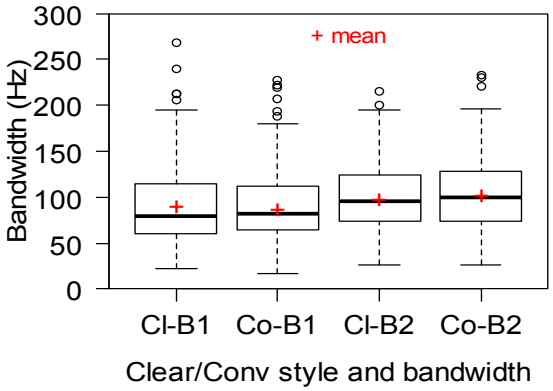

그림 7. 또렷한 발음과 대화체의 영어모음 발음에서 보인 대역값의 분포.
Figure 7. Bandwidths of English vowels produced in clear or conversational styles.

또렷한 음성이 B2에서 낮게 나온 점은 앞의 스펙트럼 정보와 일치하지만, B1에서는 오히려 2.6 Hz 높게 나타난 이유는 극단치 또는 Praat 자체의 측정값의 오류에 따른 결과이거나 스펙트럼 정보에 비해 포먼트 주파수에 해당하는 한 지점에서 측정하는 과정에서 전설모음과 후설모음 등의 정보가 뒤섞이면서 나타난 문제점으로도 여겨진다.

<그림 8>은 녹음한 음높이에 따른 포먼트 대역값의 차이를 보여준다.

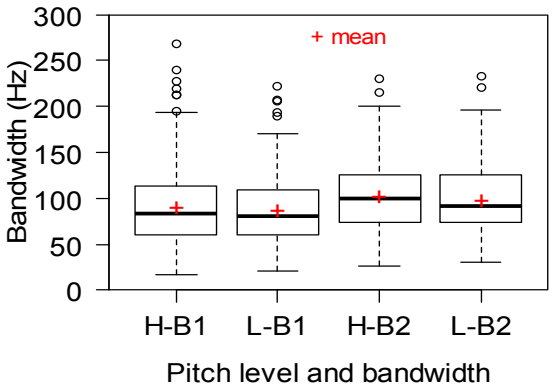

그림 8. 높은 음과 낮은 음으로 발음한 영어모음 대역값의 분포.
Figure 8. Bandwidths of English vowels produced in high- or low-pitch levels.

먼저 높은 음의 발음에서 B1과 B2의 평균은 각각 91.3 Hz와 103.0 Hz이고 낮은 음의 발음에서 B1과 B2의 평균은 각각 87.7 Hz와 99.7 Hz로 나타났다. 높은 음의 발음보다 낮은 음의 발음에서 포먼트 에너지가 약간 높게 나타난 것은, 높은 음으로 발음했을 때 피치와 반비례하는 F1의 관계를 생각해보면 (Pickett, 1987), F1이 상대적으로 낮은 주파수에 나타나게 되고,

<그림 1>의 음원스펙트럼과 앞 절의 <그림 4, 5, 6>에 나타난 에너지 분포 곡선으로 추정해볼 때, 보다 낮은 주파수에서는 에너지가 높아지고, 대역값도 낮아져 이런 경향을 보이는 것으로 해석할 수도 있다. 앞으로 더 정교한 분석을 위해서는, 근본적으로 Praat에서 구한 측정값이 옳아야 하고, 다음으로는 그림에서 나타나는 극단치를 제거한 다음 전체적인 경향을 살펴봐야 할 것으로 생각된다. 통계적으로는 음높이에 따른 집단별 차이를 Wilcoxon의 순위합으로 검정해본 결과 B1에 대해서는 W=19983, p=0.738이었고, B2에 대해서는 W=20449, p=0.457로 모두 유의미한 차이를 보이지 않았다.

마지막으로 전설모음과 후설모음으로 나누었을 때 대역값을 분석한 결과는 <그림 9>에 나타나 있다.

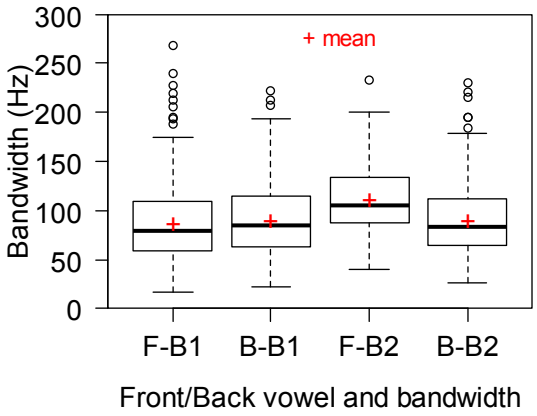

그림 9. 전설모음 군과 후설모음 군으로 나눈 모음의 대역값의 분포.

Figure 9. Bandwidths of English vowels divided into front and back vowel groups.

그림에서 전설모음의 B1과 B2의 평균은 각각 87.9 Hz와 113.2 Hz이고 후설모음의 B1과 B2의 평균은 각각 90.8 Hz와 91.4 Hz로 나타났다. 모음의 위치에 따른 집단별 차이를 Wilcoxon의 순위합으로 검정해본 결과 B1에 대해서는 W=17984, p=0.199로 유의미한 차이를 보이지 않았는데, B2에 대해서는 W=26292, p<0.05로 유의미한 차이를 보였다. 전설모음에서 B1값은 후설모음에 비해 낮게 나타났는데, 이는 전설모음의 F1부분이 후설모음의 F1에 비해서 낮은 주파수로 음원에서 들어오는 에너지가 상대적으로 높기 때문으로 여겨진다. 마찬가지로 후설모음의 F2는 전설모음의 F2에 비해 낮은 주파수에 해당하므로 포먼트 대역값이 높게 나타났다.

지금까지 대역값을 살펴본 결과 전반적으로 B1이 B2에 비해 낮은 값을 보였고, 이는 F1의 스펙트럼 세기가 상대적으로 크다는 것을 나타내고, B2에서는 전설모음과 후설모음의 차이가 유의미하게 나타났지만, 나머지 비교에서는 모두 차이가 없음을 알 수 있었다. 덧붙여, 전설모음과 후설모음의 구성비에 따라 대역값의 분포도 달라질 수 있음을 알 수 있었다.

6. 요약 및 결론

이 연구에서는 발화방식이 달라질 때 영어모음의 스펙트럼의 특성과 포먼트 대역이 어떻게 변하는지를 알아보기 위해 미국인 남성이 대화체와 또렷한 발음, 높은 음과 낮은 피치로 발음한 영어모음의 처음 두 개의 포먼트에 대한 대역값과 장기평균스펙트럼값을 구해서 비교해 보았다. 그 결과는 다음과 같다.

첫째, 모음의 스펙트럼 특성은 발화방식에 따라 차이를 보였으며, 모음의 조음위치에 따라 달라짐을 알 수 있었다. 대화체에 비해 또렷한 발음이 더 높은 스펙트럼 에너지 분포를 보였으며, 피치가 높을수록 스펙트럼 에너지도 전반적으로 높게 나타났다. 덧붙여, 전설모음 군과 후설모음 군의 스펙트럼 곡선은 포먼트의 분포에 따라 서로 다른 모양을 나타냈다.

둘째, 포먼트 대역값을 구해본 결과 전반적으로 B1이 B2에 비해 낮은 값을 보였는데, 이는 음원인 성대의 스펙트럼 특성과 얼굴 방사특성에 의한 스펙트럼 특성이 반영되었을 것으로 추정된다. B2에서는 전설모음과 후설모음 군에서 유의미한 차이를 보였고, 발화방식에 따른 B1과 B2의 통계적인 차이는 없었다.

이러한 결과를 종합해보면 발화방식에 따라 스펙트럼 특성이 조직적으로 달라지며, 서너 개의 포먼트 주파수에서 측정하는 대역값으로는 이러한 차이를 제대로 보여주지 않는다고 결론지을 수 있다.

이 연구의 결과는 음성인식과 합성에서 좀 더 자세한 화자의 발화방식에 대한 정보를 추출하거나 보다 또렷한 음성을 합성하는데 필요한 정보를 제공할 수 있을 것으로 기대되며, 앞으로 합성모음의 스펙트럼 특성이나 대역값을 조정하여 원어민에게 들려주었을 때 발화방식에 대해 어떤 평가를 내릴지 연구해볼 계획이다.

참고문헌

Assman, P.F. & Katz, W.F. (2000). Time-varying spectral change in the vowels of children and adults. *Journal of the Acoustical Society of America,* 71, 975-989.

Borden, G.J., Harris, K.S., & Raphael, L.J. (2003). (3rd ed.). *Speech science primer: Physiology, acoustics, and perception of speech.* Baltimore: Williams and Wilkins.

Boersma, P. & Kovacic, G. Spectral characteristics of three styles of Croatian folk singing. *Journal of the Acoustical Society of America,* 119(3), 1805-1816.

Fant, G. (1970). *Acoustic theory of speech production.* The Hague:

Mouton.

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,* 112, 259-271.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America,* 116, 2365-2373.

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.

Kang, S. (2007). *Acoustic analysis of speech.* (Korean translation) Seoul: Thomson Learning Korea, Pakhaksa.

(강석한 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.

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,* 112, 2165-2172.

Ladefoged, P. (2001). (4th ed.) *A course in phonetics.* Boston: Heinle & Heinle.

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. M. (1987). *The sounds of speech communication: A primer of acoustic phonetics and speech perception.* Austin, Texas: pro-ed.

Mannell, R. (2014). Speech acoustics. Retrieved on October 30 from http://clas.mq.edu.au/speech/acoustics/frequency/ vocal_tract_resonance.html.

Smiljanić, R. & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English, *Journal of the Acoustical Society of America,* 118 (3 Pt 1), 1677-88.

Stevens, K. N. (1998). *Acoustic phonetics.* Cambridge, MA: MIT Press.

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. (2012). Pitch and formant trajectories of English vowels by American males with different speaking styles. *Phonetics and Speech Sciences,* 4(1), 21-28.

(양병곤 (2012). 발화방식에 따른 미국인 남성 영어모음의 피치와 포먼트 궤적. 말소리와 음성과학, 4(1), 21-28.)

Yang, Byunggon & Kang, SunMee. (2002). A study on speaker identification by spectral difference sum and correlation coefficients. *Speech Sciences,* 9(3), 3-16.

(양병곤과 강선미. (2002). 좁은대역 스펙트럼의 차이값과 상관계수에 의한 화자확인 연구. 음성과학, 9(3), 3-16.)

• **양병곤 (Yang, Byunggon)**
부산대학교 영어교육과
부산시 금정구 장전동 30
Tel: 033-649-7816
Email: bgyang@pusan.ac.kr
Homepage: http://fonetiks.info/bgyang

부록: 포먼트 대역과 스펙트럼 에너지 측정 스크립트

```
!Praat script for measuring bandwidths(B1&B2) and longterm
!average spectra after opening all the sound files in the object
!window. Created by Byunggon Yang.
   for i from 1 to 30
    soundID=selected("Sound")
    soundIDplus='soundID'+1
    soundName$=selected$("Sound")
    ltashi$="'soundName$'"+"hi.txt"
    ltaslow$="'soundName$'"+"low.txt"
    print 'ltasName$'
    clearinfo
    select Sound 'soundName$'
     Edit
    editor Sound 'soundName$'
    Spectrogram settings... 0 5000 0.005 30
    Formant settings... 5000 5.3 0.025 35 1
    Intensity settings... 50 100 "mean energy" yes
   !First round for high f0 production
    pause Select vowel(high f0) segment to analyze...
   @bandwidth
    print
'soundName$''tab$'hif0'tab$''b1:0''tab$''b2:0''tab$''timepoint:3''newline$'
    fappendinfo D:\bw\result.txt
   @ltas
    Save as short text file: "D:\bw\'ltashi$'"
   @remover
    select Sound 'soundName$'
     Edit
    editor Sound 'soundName$'
   !second round for low f0 production
    pause Select vowel(low f0) segment to analyze...
   @bandwidth
    print
'soundName$''tab$'lowf0'tab$''b1:0''tab$''b2:0''tab$''timepoint:3''newline$'
    fappendinfo D:\bw\result.txt
   @ltas
    Save as short text file: "D:\bw\'ltaslow$'"
   @remover
     endeditor
    select 'soundIDplus'
   endfor

   procedure bandwidth
    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
    onset='start'+0.0225
    offset='end'-0.0225
    vowsegment='offset'-'onset'
    ratio='vowsegment'/3
    window=0.0125
    timepoint='onset'+'ratio'
    Select... timepoint-window timepoint+window
    b1=Get first bandwidth
    b2=Get second bandwidth
endproc

procedure ltas
 clearinfo
 Select... onset offset
 Extract selected sound (time from 0)
 Close
  endeditor
 select Sound untitled
To Ltas (pitch-corrected): 75, 600, 5000, 100, 0.0001, 0.02, 1.3
 selectObject: "Ltas untitled"
endproc

procedure remover
 Remove
 select Sound untitled
 Remove
endproc
```