

# 영어 단모음의 지각 실험\*

양 병 곤  
(동의대학교)

Yang, Byunggon. 2000. A perceptual experiment of English monophthongs. *Korean Journal of Linguistics*, 25-4, 609-628. This study focused on how human beings discriminate vowel pairs with varied formant frequencies. 27 Korean and American subjects participated in the AX discriminatory tests of 376 English vowels synthesized. Fant's (1972) bandwidth equations were employed to dynamically tune to the varied formant values. Then, the center formant frequency values of each subject for which each subject perceived the same vowel quality were collected. Results showed that the center formant frequency calculated on the data of both Korean and American groups matched exactly the formant frequency values of the synthesis model. This seems to reflect the accurate human auditory mechanism. Future studies on the human auditory scale using synthesized speech stimuli will be desirable to develop an auditory scale applicable to speech recognition. (Donguei University)

## 1. 머리말

조음음성학, 청각음성학, 음성공학 등을 포함하는 현대의 음성과학 연구 분야에서 가장 문제가 되는 부분은 바로 음성 신호의 변화다. 음성신호는 말하는 사람의 신체적인 특징과, 정서상태, 주위 잡음 등이 모두 섞여져서 최종적으로 청자의 귀에 들어가게 된다. 물리적 음성의 특징은 바로 이러한 화자의 비언어적인 요소들이 포함되어 있기 때문에 언어단위로서의 음성과 1 대 1 대응되는 요소를 찾기 힘들다. 특히, 화자는 환경에 적응하여 발음을 조절하게 된다 (Lindblom and Maddieson, 1988; Lindblom and Engstrand, 1989; Lindblom, 1990). 즉, 혼자 말할 때는 조용히 불분명하게 말할 것이고, 상대방이 귀가 먼 노인이라면 화자는 거기에 맞춰 더 또렷하고 크게 발음하여 의미를 전달하려고 할 것이다. 공식적인 모임에서 연설을 할 때는 평소 친구에게 하듯이 말하지 않는다. 또한, 여러 가지 음들이 결합되는 과정에서 발성기관인 혀와 턱, 입술 등의 움직임이 연결될 때 음

---

\*이 연구는 1999학년도 동의대학교 일반연구과제 연구 조성비에 의하여 연구되었음.

성끼리의 상호작용에 의해 변형될 것이다. 주위 잡음의 유무에 따라서도 발생 방식이 달라진다. 예를 들어, 소음이 많은 환경에서는 화자는 점점 더 목소리를 높여 발음하게 되고, 이런 조음 동작은 평소에 조용한 곳에서 대화할 때와는 아주 다른 물리 음향적 특징을 갖게된다. 그래서 실제 음성적으로나 심리적으로는 동일한 음성이라도 결과에서는 엄청나게 다른 발음이 되어 나타나게 된다.

청자들은 이러한 다양한 변화에도 불구하고 쉽게 발음을 듣고 이해한다. 이렇게 음성과는 무관한 신체적인 특징과 비언어적인 요소를 제거하고 화자가 의도하는 음성특징만을 골라내는 과정을 정규화(Normalization)라고 한다 (Fant, 1968). 정규화 과정은 두 가지로 나눌 수 있는데 조음적 측면에서 접근하는 방법과 청각적 측면에서 접근하는 방식으로 나눌 수 있다. 조음적 정규화 방식에는 균일축척 방식(Nordstrom and Lindblom, 1975)과 비균일 축척 방식 (Fant, 1975)이 있는데 근본 원리는 사람마다 발생하는 성도의 모양과 크기가 다르기 때문에 음향적 결과가 달라진다고 보고 이를 같은 모양과 크기로 환원시키는 작업에 해당한다. 그래서 성도의 길이가 큰 남성을 기준으로 정규화시킬 때는 보다 짧은 여성의 성도길이와의 비율을 구해 동일한 길이로 재환산하거나, 각 모음마다 구강과 인강의 비율이 달라지기 때문에 모음마다 다른 환산값을 부여하여 보다 더 정밀한 정규화를 추구하기도 했다. 하지만 연구 대상이 많아질수록, 이러한 정규화 함수는 변화하게 된다. 실제 이들 함수를 구하기 위해 연구된 자료는 인구어 중심으로 진행되었기 때문에 국어 자료나 다루지 않았던 외국어의 자료와 비교해볼 때 다른 값을 가지게 될 것이다 (Yang 1992, 1996). 그것은 바로 서로 다른 성도 모양과 크기를 가진 각국의 화자들의 관계를 하나의 함수로 나타낼 수 없기 때문이다. 이러한 한계를 극복하기 위해서는 청자의 입장에서 문제를 해결해보는 것이 필요하다. 사람의 청각기관인 귀는 쉽게 남녀의 차이를 극복하고 필요한 음성정보를 뽑아낸다. 이러한 정규화방식을 추구하는 과정에서 사람이 듣는 방식으로 음향신호를 변형하는 방안과 포먼트값 사이의 관계를 이용하려는 방안 (Syrdal and Gopal 1986; Miller 1989)이 제시되었다. 예를 들어, 음향분석에서 나온 결과를 청각척도로 변형하면 신체적인 특징이 많이 걸러지는 결과를 가져오기 때문에 일종의 정규화가 진행되었다고 할 수 있다. 소리의 높이에 대한 청각척도에는 mel이 있는데 이것은 음향심리학 실험에서 1,000 Hz (40 dB) 음을 1,000 mel로 하고 비교할 음이 이것의 두 배가되면 2,000 mel, 반이 되면 500 mel로 지정하였다 (Fant, 1960). mel과 비슷한 청각척도로는 bark가 있는데 이것은 어떤 주파수를 가장 효과적으로 차폐할 수 있는 임계주파수역 (critical band)을 측정한 것이다. 이 임계주파수역은 저주파에서는 약간의 변화에도

쉽게 감지되며 고주파에서는 아주 넓은 값의 변화에도 둔감한 특성을 보인다 (Zwicker and Terhardt, 1980). 이 두 가지 척도에 대한 비교와 논의는 양병곤 (1997)을 참고하기 바란다. 이러한 척도는 컴퓨터에서 분석된 음향 물리적 단위가 사람이 듣는 청각척도와 선형적인 관계가 없음을 보여주며 이런 차이 때문에 음성분석 결과에서 일관된 음성의 특성을 찾지 못하는 원인이기도 하다. 음성인식기도 일부화자에게는 잘 적용되지만 화자가 달라지면 성능이 떨어지는 이유도 이런 점 때문이라고 할 수 있다. 하지만 이들 청각 척도 자체가 소수의 피험자에 의한 실험결과이고, 또한 단순한 정현파의 조작을 통해 설정한 척도이기 때문에 복잡한 여러 가지 배음으로 구성된 음성에 적용하기에는 부적절하다고 할 수 있다. 사람의 음성은 기본적으로 서너 개의 포먼트로 특징을 정의할 수도 있지만, 각 포먼트의 음향특성은 인접한 많은 수의 배음에 의해 결정된다. 따라서, 배음 끼리 상호 작용하여 차폐효과와 진폭상승 효과 등을 일으키게 된다. 이러한 복잡한 관계를 반영하지 않는 청각척도는 대체적인 청각유형을 보일 수는 있으나 음성인식과 합성 등에 응용하기에는 불완전하다고 할 수 있다. 덧붙여, 음원의 특성인 기본주파수 값과 각 포먼트 사이의 관계를 이용하여 절대적인 포먼트 값보다는 상대적인 비율값을 중심으로 영어 모음을 구분하려는 시도 (Syrdal and Gopal, 1986)와 인접한 자음의 변화에 따른 상대적인 모음 포먼트의 변화를 이용하여 자음을 예측하려는 Sussman 외 (1991)의 시도를 보면 절대적인 청각척도보다는 상대적인 척도를 연구할 필요성을 확인해준다. Syrdal and Gopal (1986)은 상대적인 비율값에 의해 Peterson and Barney (1952)의 모음을 구분했는데 이 결과는 그 당시 피험자들이 직접 들어서 확인한 것과 차이가 없음을 보인 바 있다.

따라서, 이 논문에서는 이러한 복잡한 음향신호에도 불구하고 쉽게 필요한 음성정보를 뽑아 처리하는 사람의 청각기관에 초점을 두고, 합성한 모음을 청자가 어떻게 받아들이는지를 구별실험을 통해 추적해보고 아울러 기존의 청각척도와는 어떤 관계가 있는지를 살펴보고자 한다. 사람마다 다양한 포먼트 주파수를 가진 모음 발음을 하면서도, 들을 때는 어떤 일정한 평가기준을 가지고 동일한 모음으로 판단한다고 여겨진다. Yang (1996)에서 한국인과 미국인의 포먼트 값에서 추정된 성도길이의 차이를 보았는데, 지각에는 어떤 차이가 있는지도 비교해보기로 한다. 구체적으로는 어떻게 청자들이 이러한 포먼트 값의 변화에 대해 기준을 설정하여 구별해나가는 지 다음과 같은 연구 질문을 중심으로 살펴보기로 한다.

1. 미국인과 한국인 남녀가 동일한 영어 모음으로 판단하는 포먼트 값의 범위는 어떠한가?

2. 남녀화자의 구별 값과 범위에 어떤 차이가 있는가?
3. 원래의 합성과일에 사용한 대역 값은 청자의 지각 범위에 영향을 미치는가?
4. 동일하게 지각하는 범위와 기존의 청각척도는 어떤 상관 관계가 있는가?

이런 질문에 답하기 위해서 저자는 양병곤 (1996)에 제시된 9개의 영어 모음의 포먼트 값을 이용하여 자연스러운 영어모음을 합성하고, 이들 각각의 모델에서 포먼트 값을 임의의 간격으로 변화시킨 것과 쌍을 이뤄 27명의 한국인과 미국인 남녀화자에게 들려주고 동일한 모음으로 듣는 범위를 구했다. 이어서 동일한 모음 범위의 중심값을 구하여 합성용 모델의 포먼트 값과 비교하여 보고 또한 이 범위와 청각척도의 관계를 살펴보고자 한다.

## 2. 연구방법

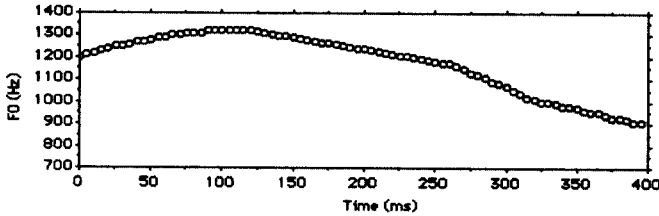
### 2.1. 피험자

피험자는 미국 텍사스 오스틴 주립대학에 재학하고 있는 대학원생으로 건강하고 청각에 지장이 없는 사람들을 골랐다. 한국인 남성 6명, 한국인 여성과 미국인 남녀 각각 7명씩 모두 27명이 두 쌍의 음질이 동일한지 여부를 응답하는 AX 구별실험에 참가했다. 미국인은 미국에 태어나서 생활한 원어민이고 한국인은 유학생으로 일정한 수준의 TOEFL 성적을 받아서 이 대학에 입학허가서를 받았기 때문에 영어구사능력이 매우 우수한 편이다. 미국인 여성의 나이의 평균은 26세이고 키는 163 cm이고, 남성의 나이와 키는 평균 29세이고 182 cm이다. 한국인 남성의 평균은 33세, 175 cm이고 여성은 30세 162 cm이다.

### 2.2. 합성음 생성과 실험과정

음성 합성은 포먼트 합성소프트웨어인 SenSynPPC1.0을 이용하여 파워 피씨 컴퓨터에서 합성했다. 국어 단모음과 이중모음의 합성 방법에 대한 개관은 양병곤 (1995, 1996)을 참고하길 바란다. 합성용 포먼트 파라미터는 Yang (1996)의 미국인 남성화자의 포먼트 값의 평균을 사용했다. 합성된 모음의 지속시간은 400 ms이며, 합성한 뒤 실제 모음으로 발음되는 부분은 300 ms에 해당한다.

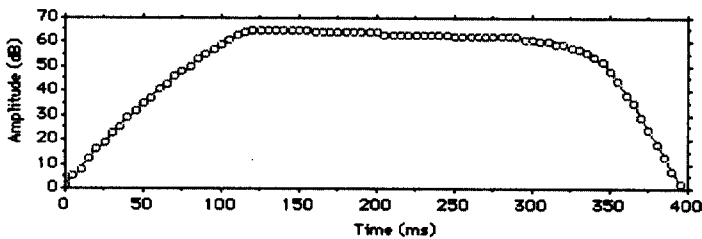
그림 1. 피치값의 변화 모델. x축은 시간을 나타내며 y축은 피치값을 10배하여 나타내었다.



합성기의 파라미터 중 피치는 자연스런 말소리의 특성을 최대한 반영하기 위해 미국인 전체 남성화자의 피치값을 평균하여 얻은 132 Hz가 최대값이 되게 하고 두 명의 미국인 화자의 발음을 분석하여 그림 1과 같이 서서히 이 최대값으로 올라갔다가 떨어지는 모델을 설정하여 자연스러운지를 점검한 뒤 이것을 기본으로 사용했다. 그림 1에서는 120 Hz에서 시작하여 110 ms 지점에서 132 Hz로 도달했다가 서서히 줄어들어 90 Hz에서 끝나도록 했다.

진폭값도 그림 2와 같이 0 dB에서 시작하여 100 ms 지점에서 65 dB로 빠르게 상승한 뒤 서서히 떨어지도록 했다. 이렇게 합성한 음성을 여러 번 들어보고 가장 자연스럽게 되도록 연구하여 최종 모델로 선정했다.

그림 2. 진폭값 모델. x축은 시간을 나타내며 y축은 진폭값을 나타내었다.



각 모음에 대한 대역값은 실제 음성에서 측정하기 어렵다 (Fujimura and Lindqvist, 1971; Klatt, 1980). 실제 음성은 음원인 성대의 떨림과 성도의 벽면을 이루는 인강과 구강의 습기 찬 표면 등의 영향을 받고 또한 얼굴 표면에 구멍이 뚫린 듯이 입에서 음성이 방사되는 효과가 가미되어 최종 음성 출력의 포먼트 대역은 이런 모든 변수들이 반영되어 있고 서로 이

웃하는 포먼트의 전과 효과도 첨가되어있기 때문이다. 따라서, Fujimura와 Lindqvist (1971)의 연구에서는 남녀 각각 3명의 피험자가 각각의 모음을 발음하는 모양을 하게한 채, 성대 바로 윗부분에 해당하는 목에 외부적으로 정현파음을 낮은 주파수에서 높은 주파수로 서서히 변화하도록 하면서 공명주파수 값을 추적했다. 그 결과 남녀의 대역 값에 차이가 있으며 제 1포먼트 (F1)에서는 낮은 주파수에서 오히려 대역값이 높다는 결과를 얻었다. 그러나 대체적으로 주파수가 높아질수록 더 높아지는 모양을 보였다. 이 논문에서는 넓은 범위로 포먼트 값을 변화시켜야 하기 때문에 Fant의 대역 값 환산 공식을 이용했다 (Fant, 1972:47).

$$(1) B1=15*(500/F1)^2+20*(F1/500)^{0.5}+5*(F1/500)^2$$

$$(2) B2=22+16*(F1/500)^2+12000/(F3-F2)$$

$$(3) B3=25*(F1/500)^2+4*(F2/500)^2+10*F3/(F4-F3)$$

이 공식들은 F1의 대역은 F1에 영향을 반영하고 제 2포먼트 (F2)의 대역은 F1 값, 제 3포먼트(F3)와 F2의 차이를 반영하여 나타낸 것으로 상대적으로 거리가 넓을 수록 좁아지며 거리가 좁을수록 대역값이 넓어지도록 함으로써 두개의 포먼트가 겹쳐 서로 상승시키는 효과가 발생하는데 (Klatt, 1980), 이러한 과도한 상승효과를 줄여주는 역할을 한다. 특히, 지각 실험을 하기 위해 포먼트 값을 임의의 간격으로 변화시킬 때 동일한 포먼트 대역을 사용하게 되면 합성된 결과음이 과도한 신호값으로 찌그러진 결과가 나타나거나 너무 작은 음성이 되기도 하므로 이렇게 역동적으로 변하는 대역값 추출 방식은 자연성을 살리는데 도움이 된다. 다만, 이 공식은 실제 음성에서 뽑은 결과를 통계적으로 처리하여 얻은 값이기 때문에 발생 기관의 크기가 다른 외국인 화자에게 적용할 때는 또 다른 공식이 필요할 것으로 여겨진다. 이렇게 구한 포먼트 대역값과 각 모음의 포먼트 값은 표 1에 나타나 있다.

표 1에서 보듯이 미국영어의 모음 가운데 9개의 모음만을 합성했다. 모음사각도에서 후설 저모음에 해당하는 [o]음은 이중모음인 [ou]의 1/3지점의 포먼트 값을 이용하여 음성을 합성했고 나머지는 모두 단모음 발음 구간에서 1/3지점에 해당하는 포먼트 값을 나타낸다. F4는 3,700 Hz로 고정하고, 대역값을 350 Hz로, F5는 4,300 Hz (대역값 400 Hz), F6는 4,990 Hz (500 Hz)를 부여하여 약간의 고주파 음을 첨가함으로써 자연스러운 합성음이 되도록 했다. 사전 실험에서 이런 대역값을 부여했을 때 보다 낮은 포먼트가 모음의 음질을 주로 결정하기 때문에 F4이상의 포먼트의 변화는 모음의 음질에는 거의 영향을 주지 않는 것으로 나타났다.

표 1. 음성합성에 사용한 모음의 포먼트 주파수 (F1, F2, F3)와 Fant의 공식에서 계산된 대역 값 (B1, B2, B3).

모음	F1	B1	F2	B2	F3	B3
æ	687	70	1743	150	2497	320
ɔ	663	90	1026	100	2527	80
e	531	60	1900	90	2561	200
i	286	45	2317	200	3033	400
ɪ	409	50	2012	100	2671	140
o	498	80	1127	70	2375	70
u	333	65	1393	110	2282	140
ʊ	446	80	1331	100	2380	80
ʌ	592	80	1331	50	2494	140
평균	494	69	1576	108	2536	174

Fant의 공식으로 구한 대역값은 앞서 공식에서도 나타나 있지만, F1의 대역은 F1값에 영향을 받고, F2, F3의 대역은 인접한 포먼트 값을 반영하도록 되어있다. 실제 결과에서도 F1과 B1의 상관계수는 0.665이며, F2와 B2의 계수는 0.669로 높게 나타난다. 무엇보다도 F2-F1의 차이를 구한 뒤 B1과의 상관계수를 구해보면 -0.977로 매우 높은 역상관관계를 나타내준다. 이것은 바로 Fant의 공식이 인접한 포먼트 일수록 대역값을 높게 하여 전이에 의한 영향력을 줄이는 결과를 극명하게 보여준다고 할 수 있다. 실제 음성합성에서 서로 가까운 포먼트에 대해 낮은 대역 값을 부여하면 음이 찌그러지는 결과를 가져옴을 관찰할 수 있었다. 앞으로 보다 정밀한 합성에 의한 지각 실험을 통해 자연스러운 대역값을 구하는 방식을 연구할 필요가 있다. 표 2는 모음 [ɛ]의 합성용 파라미터 값을 보여준다.

표 2. 모음 [ɛ]의 합성용 파라미터 값. 기본주파수와 진폭의 변화는 그림 1과 그림 2를 참고하기 바람.

```

Synthesis specification for file: 'model ɛ'
SenSyn Version 1.0
Total number of waveform samples = 4400
CURRENT CONFIGURATION: 60 parameters
SYM V/C MIN VAL MAX DESCRIPTION
-----
DU C 30 400 5000 Duration of the utterance, in msec
UI C 1 5 20 Update interval for parameter reset, in msec
SR C 5000 11128 20000 Output sampling rate, in samples/sec
NF C 1 5 6 Number of formants in cascade branch
    
```

SS C 1 2 3 Source switch (1=impulse, 2=natural, 3=LF model)  
 RS C 1 8 8191 Random seed (initial value of random # generator)  
 SB C 0 1 1 Same noise burst, reset RS if AF=AH=0, 0=no,1=yes  
 CP C 0 0 1 0=Cascade, 1=Parallel tract excitation by AV  
 OS C 0 0 20 Output selector (0=normal,1=voicing source,...)  
 GV C 0 60 80 Overall gain scale factor for AV, in dB  
 GH C 0 60 80 Overall gain scale factor for AH, in dB  
 GF C 0 60 80 Overall gain scale factor for AF, in dB  
 F0 V 0 1000 5000 Fundamental frequency, in tenths of a Hz  
 AV V 0 60 80 Amplitude of voicing, in dB  
 OQ v 10 50 99 Open quotient (voicing open-time/period), in %  
 SQ v 100 200 500 Speed quotient (rise/fall time, LF model), in %  
 TL v 0 0 41 Extra tilt of voicing spectrum, dB down @ 3 kHz  
 FL v 0 0 100 Flutter (random fluct in f0), in % of maximum  
 DI v 0 0 100 Diplophonia (alt periods closer), in % of max  
 AH v 0 0 80 Amplitude of aspiration, in dB  
 AF v 0 0 80 Amplitude of frication, in dB  
 F1 v 180 531 1300 Frequency of 1st formant, in Hz  
 B1 v 30 60 1000 Bandwidth of 1st formant, in Hz  
 DF1 v 0 0 100 Change in F1 during open portion of period, in Hz  
 DB1 v 0 0 100 Change in B1 during open portion of period, in Hz  
 F2 v 550 1900 3000 Frequency of 2nd formant, in Hz  
 B2 v 40 90 1000 Bandwidth of 2nd formant, in Hz  
 F3 v 1200 2561 4800 Frequency of 3rd formant, in Hz  
 B3 v 60 200 1000 Bandwidth of 3rd formant, in Hz  
 F4 v 2400 3700 4990 Frequency of 4th formant, in Hz  
 B4 v 100 350 1000 Bandwidth of 4th formant, in Hz  
 F5 v 3000 4300 4990 Frequency of 5th formant, in Hz  
 B5 v 100 400 1500 Bandwidth of 5th formant, in Hz  
 F6 v 3000 4990 4990 Frequency of 6th formant, in Hz (applies if NF=6)  
 B6 v 100 500 4000 Bandwidth of 6th formant, in Hz (applies if NF=6)  
 FNP v 180 280 500 Frequency of nasal pole, in Hz  
 BNP v 40 90 1000 Bandwidth of nasal pole, in Hz  
 FNZ v 180 280 800 Frequency of nasal zero, in Hz  
 BNZ v 40 90 1000 Bandwidth of nasal zero, in Hz  
 FTP v 300 2150 3000 Frequency of tracheal pole, in Hz  
 BTP v 40 180 1000 Bandwidth of tracheal pole, in Hz  
 FTZ v 300 2150 3000 Frequency of tracheal zero, in Hz  
 BTZ v 40 180 2000 Bandwidth of tracheal zero, in Hz  
 A2F v 0 0 80 Amp of fric-excited parallel 2nd formant, in dB  
 A3F v 0 0 80 Amp of fric-excited parallel 3rd formant, in dB  
 A4F v 0 0 80 Amp of fric-excited parallel 4th formant, in dB  
 A5F v 0 0 80 Amp of fric-excited parallel 5th formant, in dB  
 A6F v 0 0 80 Amp of fric-excited parallel 6th formant, in dB  
 AB v 0 0 80 Amp of fric-excited parallel bypass path, in dB  
 B2F v 40 250 1000 Bw of fric-excited parallel 2nd formant, in Hz  
 B3F v 60 300 1000 Bw of fric-excited parallel 3rd formant, in Hz  
 B4F v 100 320 1000 Bw of fric-excited parallel 4th formant, in Hz



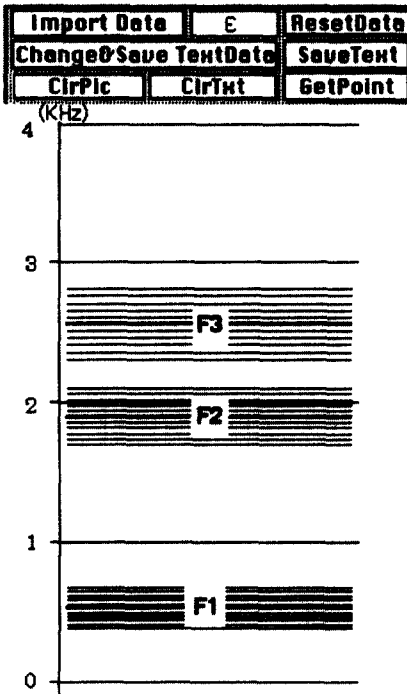
B5F v 100 360 1500 Bw of fric-excited parallel 5th formant, in Hz  
 B6F v 100 1500 4000 Bw of fric-excited parallel 6th formant, in Hz  
 ANV v 0 0 80 Amp of voice-excited parallel nasal formant, in dB  
 A1V v 0 60 80 Amp of voice-excited parallel 1st formant, in dB  
 A2V v 0 60 80 Amp of voice-excited parallel 2nd formant, in dB  
 A3V v 0 60 80 Amp of voice-excited parallel 3rd formant, in dB  
 A4V v 0 60 80 Amp of voice-excited parallel 4th formant, in dB  
 ATV v 0 0 80 Amp of voice-excited par tracheal formant, in dB

합성파일을 생성하여 여러 번 파라미터를 변화시켜 원래의 음성과 거의 가까운 발음이 되었을 때 이를 모델로 설정하였고, 실제 합성된 모음 [ε]와 발성음의 스펙트로그램을 살펴보면 성문 파동이 아주 자연스럽게 움직이는 실제음성과 거의 비슷한 모양이고 음질도 거의 같았다. 이어서, 이 모델을 중심으로 각 포먼트값을 Hypertalk를 이용하여 변화시켜 컴퓨터에 저장시켰다. 그림 3은 모델 [ε]의 포먼트 값과 중심선에서 아래위로 다섯 개씩 차례로 변화한 합성 파일을 저장한 흔적을 보여주고 있다. 이 프로그램은 세 포먼트 값의 차이를 자동으로 계산해주어서 이를 바탕으로 실험자가 포먼트 값 아래위로 몇 개의 파일을 생성할 것인지 결정하게 해준다. 사람의 귀가 저주파는 매우 상세하게 구분하고 고주파 부분은 상당한 변화에도 둔감한 청각 특성을 보이기 때문에, F1에서는 30 Hz 간격으로 하고 F2에서는 40 Hz, F3에서는 50 Hz 간격으로 기준 포먼트 값 아래위로 여러 개의 합성음을 만들었다. 단, 인접한 포먼트 값은 넘지 않도록 하여 모두 1,109개의 파일을 생성하였고, 이것을 합성하여 두 명의 미국인에게 들려주어 비슷한 모음으로 지각하는 범위를 구한 뒤 이 범위 아래위로 변화시켜서 완전히 다른 모음으로 들릴 때까지 충분한 수의 합성음을 배치하여 총 362개의 실험용 합성음을 정하고 열 개의 동일한 모음 쌍을 추가하여 청자의 지각판단의 유효성을 확인했다.

F3의 변화폭은 매우 넓고 중복되기 때문에 실험시간을 줄이기 위해 세 개의 모음(ɔ, i, u)에 대해서만 합성했다. 이어서 합성한 9개의 모음을 각각 엑셀에서 임의의 난수를 발생하게 하여 이 순서대로 배열함으로써 모든 포먼트 값안에서 서로 섞이도록 하여 Praat3.8 프로그램을 이용하여 기준음과 변화음의 순서로 재생시켰다. 두 합성음의 발음 간격은 약 200 ms이고 이어서 1.2 초 동안 피험자가 동일한지 여부를 표시할 반응시간을 넣어 Sony 미니디스크 (MZ-R30)에 녹음했다. 이 미니디스크에 녹음된 음성과 컴퓨터에서 직접 재생된 음성을 필자가 몇 번 들어보아 차이가 없음을 확인했다.

각 실험에서는 376개의 임의의 순서쌍의 합성음을 헤드폰으로 피험자가 볼륨을 직접 조정하여 들었다. 청각 실험은 발음 장치된 UT의 언어학과 음성실험실과 심리학과 실험실을 이용했다.

그림 3. Hypertalk에 의한 합성파일 생성 흔적. 굵은 선은 모델[c]의 세 개의 포먼트 지점을 나타내며 아래위의 가는 선은 변화시킨 흔적. x축은 시간을 나타내고 y축은 주파수를 나타낸다.



합성한 음은 AX구별 실험 쌍으로 배열하여 피험자의 섬세한 구별능력을 측정했다. 응답지에는 다섯 개의 연습 문제와 함께 각 항목을 들으면서 같은 모음으로 들릴 때는 동그라미를 다른 모음으로 들릴 때는 빗금을 치게 했다. 실험 시간은 약 30분 동안 소요되었다. 실험당시 피험자들은 처음 모든 모음을 한 쌍씩 듣고 나서 5개의 연습 문제를 표시하면서 서서히 빠르게 적용해 나갔다. 각각의 문항 사이에는 숫자음을 넣어서 어떤 항목을 실험하고 있는지 알려주었고 서로 다른 모음 사이에는 종소리와 같은 정현파음을 두 개씩 네 개를 올려서 다음 모음으로 이동하도록 유도했다. 피험자들은 실험하고 있는 부분을 충분히 이해할 수 있었고, 짧은 시간에 실험

방식과 속도에 적응하여 너무 빠르거나 느리다고 생각하지는 않았다.

임의의 순서로 된 응답지를 주파수 값의 순서로 재정렬한 뒤 다음과 같이 처리했다. 첫째 동일하다고 표시한 모음 쌍이 하나 건너 두 개 이상 위치해있을 때는 그 사이에 다르다고 표시된 것은 잘못으로 보고 두 개 이상 위치한 쌍 가운데 중심주파수에서 먼 쪽의 값을 취했다. 다르다고 표시한 쌍의 위아래에 동일하다고 판단한 경우에는 그 사이의 주파수 값을 취했다. 이것은 청각 기관의 정밀도를 최대한 반영하기 위한 것이며 피험자마다 다르지만 약 2-3개 지점에서 이런 모양이 발견되었다.

### 3. 결과와 토론

표 3과 표 4는 14명의 미국인 남녀가 동일한 모음 값으로 판단하는 주파수의 범위와 중심주파수 값의 평균을 보여준다. 표 5와 표 6은 13명의 한국인 남녀의 동일한 모음 값으로 판단하는 주파수의 범위와 중심주파수 값의 평균을 보여준다. 표 3에서 F1의 표준편차를 보면 33 Hz로 매우 좁으며 F2에서 F3값으로 갈수록 약간 높아지기는 하지만 대체로 안정된 구별 결과를 보여주고 있음을 알 수 있다. 동일하게 지각하는 범위도 평균을 중심으로 큰 변화가 없으며 고주파로 갈수록 약간 증가하는 경향을 보이고 있다. 이러한 경향은 표 4에서 표 5, 표 6에 이르기까지 비슷하게 나타나고 있는데, 앞으로 상세히 분석해보겠지만, 일반적으로, 모국어가 다른 집단의 지각임에도 불구하고 매우 유사한 청각구별 결과를 볼 수 있다. 미국인 남성화자에서는 모음 /æ/에서 매우 높은 편차를 보인 것을 제외하고는 거의 비슷한 값을 보이고 있으며, 여성에게서는 거의 변화가 없음을 알 수 있다. 이 값을 제외하고는 대체로 여성의 주파수 중심값에 대한 편차가 더 많다고 할 수 있다.

표 3. 7명의 미국인 남성이 동일한 모음으로 판단한 주파수 범위의 중심값 (F1m, F2m, F3m)과 주파수 범위 (R1m, R2m, R3m)의 평균값. 괄호 안에 표준편차값을 표시했다.

모음	F1m	R1m	F2m	R2m	F3m	R3m
æ	715 (112)	159 (69)	1752 (30)	217 (105)		
ɔ	661 (20)	90 (35)	1050 (19)	134 (51)	2527 (50)	657 (348)
ε	550 (50)	111 (51)	1897 (35)	177 (124)		
ɪ	405 (14)	86 (36)	2043 (30)	189 (30)		
i	273 (16)	86 (32)	2337 (94)	303 (197)	3022 (105)	293 (234)
o	522 (15)	90 (35)	1158 (34)	223 (100)		
u	329 (28)	103 (49)	1402 (51)	371 (124)	2286 (30)	393 (93)
ʊ	431 (17)	90 (55)	1325 (15)	160 (73)		
ʌ	590 (22)	107 (49)	1315 (61)	209 (69)		
평균	497 (33)	102 (46)	1587 (41)	220 (97)	2612 (62)	448 (225)

표 4. 7명의 미국인 여성이 동일한 모음으로 판단한 주파수 범위의 중심값 (F1m, F2m, F3m)과 주파수 범위 (R1m, R2m, R3m)의 평균값. 괄호 안에 표준편차값을 표시했다.

모음	F1f		R1f		F2f		R2f		F3f		R3f	
æ	688	(51)	199	(130)	1743	(37)	240	(142)				
ɔ	660	(15)	114	(57)	1025	(18)	157	(73)	2538	(78)	893	(358)
ɛ	555	(18)	128	(61)	1914	(43)	246	(162)				
ɪ	428	(17)	99	(69)	2012	(60)	217	(142)				
i	269	(10)	94	(21)	2348	(82)	326	(269)	3067	(94)	368	(233)
o	509	(14)	107	(42)	1150	(29)	229	(89)				
u	329	(19)	94	(70)	1407	(65)	349	(184)	2279	(42)	436	(246)
ʊ	450	(11)	94	(47)	1331	(37)	194	(96)				
ʌ	594	(10)	107	(52)	1318	(32)	237	(120)				
평균	498	(18)	115	(61)	1583	(45)	244	(142)	2628	(71)	566	(279)

표 5. 6명의 한국인 남성이 동일한 모음으로 판단한 주파수 범위의 중심값 (F1m, F2m, F3m)과 주파수 범위 (R1m, R2m, R3m)의 평균값. 괄호안에 표준편차값을 표시했다.

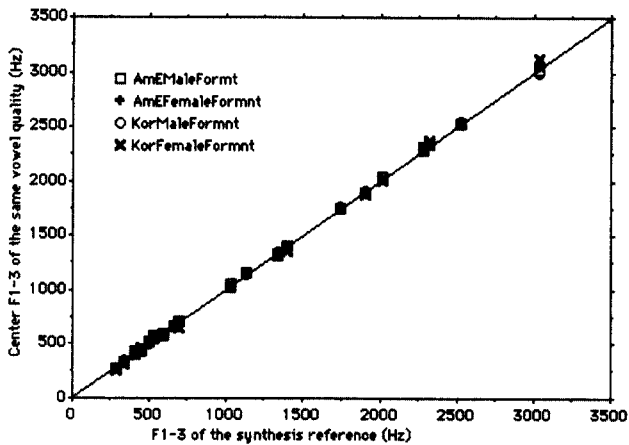
모음	F1m		R1m		F2m		R2m		F3m		R3m	
æ	695	(52)	150	(114)	1760	(23)	207	(69)				
ɔ	661	(18)	105	(41)	1034	(52)	123	(37)	2552	(164)	700	(315)
ɛ	576	(23)	55	(52)	1907	(24)	240	(51)				
ɪ	414	(26)	75	(37)	2042	(17)	300	(49)				
i	271	(9)	100	(31)	2350	(86)	387	(249)	2996	(52)	525	(186)
o	508	(15)	90	(19)	1144	(20)	193	(39)				
u	341	(25)	105	(73)	1386	(52)	373	(155)	2307	(32)	467	(133)
ʊ	451	(20)	130	(56)	1331	(28)	240	(110)				
ʌ	585	(31)	95	(44)	1339	(19)	297	(56)				
평균	500	(24)	101	(52)	1588	(36)	262	(91)	2618	(83)	564	(211)

표 6. 7명의 한국인 여성이 동일한 모음으로 판단한 주파수 범위의 중심값 (F1m, F2m, F3m)과 주파수 범위 (R1m, R2m, R3m)의 평균값. 괄호안에 표준편차값을 표시했다.

모음	F1f		R1f		F2f		R2f		F3f		R3f	
æ	657	(27)	180	(104)	1749	(28)	217	(76)				
ɔ	661	(42)	124	(47)	1027	(30)	140	(49)	2527	(69)	857	(354)
ɛ	576	(59)	99	(41)	1877	(44)	251	(105)				
ɪ	428	(21)	90	(24)	2015	(21)	269	(72)				
i	260	(11)	111	(23)	2368	(84)	423	(115)	3119	(73)	429	(191)
o	526	(27)	99	(48)	1153	(38)	211	(38)				
u	324	(34)	120	(30)	1359	(76)	297	(105)	2311	(62)	500	(132)
ʊ	459	(16)	137	(38)	1337	(41)	229	(50)				
ʌ	577	(31)	99	(29)	1328	(37)	257	(55)				
평균	496	(30)	118	(43)	1579	(44)	255	(74)	2652	(68)	595	(226)

표 5와 표 6에서 한국인들도 미국인과 비슷한 정도의 편차를 보이고 있으며, 중심주파수 값의 평균도 거의 일치하는 값을 보인다.

그림 4. 합성모델로 사용한 음성의 포먼트 중심주파수를 x축으로 하고 미국인 남녀, 한국인 남녀의 값을 y축에 둔 분포 그림. 중심선은 서로 일치하는 지점을 그래프의 꼭지점을 임의로 이은 것임.



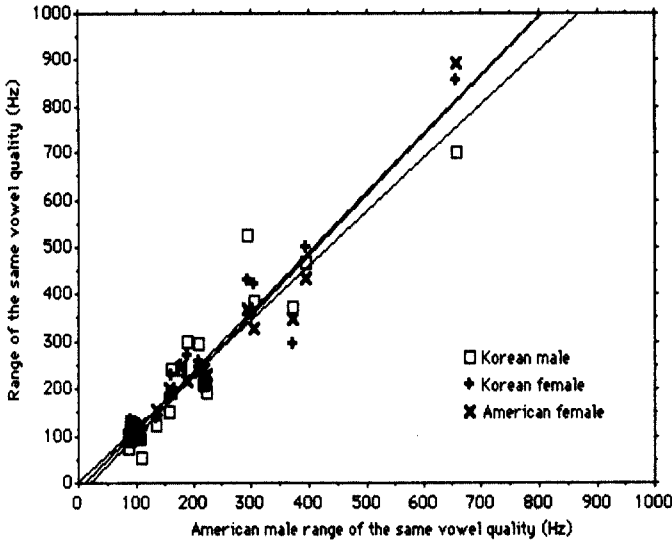
한국인 남녀에서는 모음 /æ/를 동일하게 지각하는 범위에 많은 차이를 보여 여전히 국어 발음에서 모음 /에/와 /애/의 구분이 적은 특성을 반영하

고 있음을 알 수 있다. F3의 범위에서는 후설모음 /ɔ/에서 모두 높은 편차를 보이고 있으며 범위도 매우 높다.

이러한 관계를 그림과 통계분석을 통해 보다 명확히 살펴보자. 먼저, 합성 모델로 사용한 음성의 포먼트 주파수와 피험자들이 동일한 모음으로 판단한 각 그룹의 포먼트 중심주파수에 대해 실시한 결과는 그림 4와 같다.

그림 4에서 살펴보면 미국인 남녀, 한국인 남녀의 중심값 주파수는 거의 일치하는 직선에 모두 분포되어 있음을 알 수 있다. 비록 표 3~6에서 보듯이 약간의 차이는 있지만 놀라울 정도로 정확한 청각구별 특성을 알 수 있다. 특히, 한국 유학생들의 중심값도 미국인과 동일한 분포를 보이는데 이것은 이들이 우수한 영어 모음 구별능력을 갖추었기 때문으로 여겨지는데 앞으로 영어유창도가 낮은 그룹에서도 이렇게 나타나는지 연구해볼 필요가 있다.

그림 5. 미국인 남성화자가 동일한 모음으로 지각하는 범위의 평균값을 x축으로 하고 미국인 여성, 한국인 남녀의 값을 y축에 둔 분포 및 회귀곡선 그림.



이어서, 미국인 남성화자가 동일한 모음으로 지각하는 범위의 평균값에 대해, 미국인 여성, 한국인 남녀화자집단을 StatView SE+로 회귀분석을 하여 나타내면 그림 5와 같다. 그림 5를 보면 F1에서는 허용하는 주파수 값

의 범위가 매우 좁고 F2에서는 더 넓어졌으며 F3에서는 매우 넓은 영역을 차지하고 있다. 대체로 남성화자는 항상 여성화자보다 낮은 범위를 보이고 있으므로 이것은 남성화자의 성도 길이가 길기 때문에 전반적으로 낮은 포먼트 주파수를 가지고 있으므로 생긴 결과로 볼 수 있다.

보다 세밀한 분석을 하기 위해 네 집단의 중심주파수값과 동일한 모음으로 판단한 범위 사이의 관계를 회귀분석을 실시한 뒤 회귀함수의 기울기, 절편과 예측값( $r^2$ )으로 나타내보면 표 6과 같다.

표 6. 네 집단이 각각 동일한 모음으로 지각하는 주파수 범위의 중심값을 회귀분석한 결과 나타난 함수의 기울기, 절편과 예측값. 두 그룹 가운데 앞의 그룹을 x축에 뒤의 그룹을 y축에 두고 회귀분석했음.

비교대상	기울기	절편	예측값( $r^2$ )
한국인 여성 대 미국인 여성	0.967	-2.076	0.957
한국인 남성 대 미국인 여성	0.980	3.244	0.858
한국인 남성 대 한국인 여성	1.032	1.123	0.930
미국인 남성 대 한국인 남성	1.147	4.058	0.876
미국인 남성 대 한국인 여성	1.262	-10.670	0.927
미국인 남성 대 미국인 여성	1.272	-22.617	0.962

표 6에서 살펴보면 한국인 남성은 미국인 남성에 비해 약 1.14배 이상 넓은 허용범위를 보이고 있으며 한국인 여성과 미국인 여성은 두 그룹 모두 1.26배 이상 넓은 허용범위를 보여서 이것은 앞서 각 피험자의 평균 키의 비율을 생각해볼 때 상당히 신체적인 비율을 반영하고 있음을 알 수 있다. 앞서 미국인 여성과 한국인 여성의 키의 평균은 거의 비슷하게, 162 cm와 163 cm였고, 한국인 남성의 평균은 175 cm였고, 미국인 남성의 키는 평균 182 cm였다. 키가 클수록 성대와 성도의 크기도 비례하여 크고 이에 따른 포먼트와 피치값은 낮아지게 된다 (Fant, 1970). 이런 포먼트의 비율 값은 양병곤 (1996:255, Figure 3)의 논문에 제시된 성도길이의 차이와 연관됨을 알 수 있다. 앞으로 이들 각 집단들의 발성 자료를 분석하여 성도의 길이를 나타내는 포먼트 값과 지각 특성을 비교해 볼 계획이다. 기울기 값이 1보다 크기 때문에 대체로 남성화자의 범위가 낮고 여성화자의 범위가 높게 나타나 있으며, 저주파에서 고주파로 갈수록 동일하게 지각하는 모음의 범위가 넓어짐을 알 수 있다. 회귀함수의 절편 값은 만일 동일한 성도의 길이와 모양을 가진 사람이라면 0에 가깝게 나타나겠지만, 다른 값을 보이기 때문에 남녀의 차이를 나타내는 변수로 여겨지는데 여기서는 그렇게 많은 차이를 보이지 않는다. 또 한가지 주목할 만한 점은 미국인 남

성과 여성집단의 허용범위에 대한 예측값 (0.962)이 매우 높고 한국인 남성과 한국인 여성집단사이에 예측값 (0.930)이 높다는 점이다. 이것은 각 언어내에 속하는 집단의 지각방식이 거의 비슷함을 보여주고 있으며, 반면에 한국인 남성과 미국인 남성을 비교한 경우에는 약간 낮은 예측값을 보여주고 있다. 그러나 한국인 여성과 미국인 여성의 비교에서는 보다 높은 예측값 (0.957)이 나타났다. 이 실험에 참가한 한국 유학생들의 영어능력이 매우 높은 점을 고려해볼 때 이들이 영어의 모음 지각 범위를 이미 습득하여 원어인에 가깝게 도달했기 때문으로 추정되는데, 앞으로 영어발음 구별 능력이 낮은 그룹에 대한 실험을 통해 이를 밝혀볼 필요가 있을 것이다. Diehl(1989)은 여성의 기본주파수 값이 매우 높기 때문에 각 포먼트 주위에 적은 수의 배음이 배치되어 발음되고 이에 따라 청각적인 구별능력도 달라졌음을 지적한 바 있다. 그래서, 여성은 남성보다 넓은 범위의 포먼트 변화에 대해 동일하게 지각하는 결과를 나타내게 된다고 해석할 수도 있다.

덧붙여, 원래의 합성과일에 사용한 대역값과 미국인의 지각범위의 상관관계는  $r^2=0.093$ 으로 거의 관계가 없는 것으로 나타났다. 반면 F3을 제외하고 F1과 F2만을 비교했을 때는 0.368로 다소 상관성이 있음을 보였지만, 이것은 아마 낮은 포먼트 일수록 대역값이 좁게 밀집되어서 통계적으로 상관성을 보이는 결과를 보였을 것이다. 이런 분석 결과는 합성을 자체의 대역값이 개별음의 구별에 큰 영향을 미치지 않았음을 증명해준다고 할 수 있다.

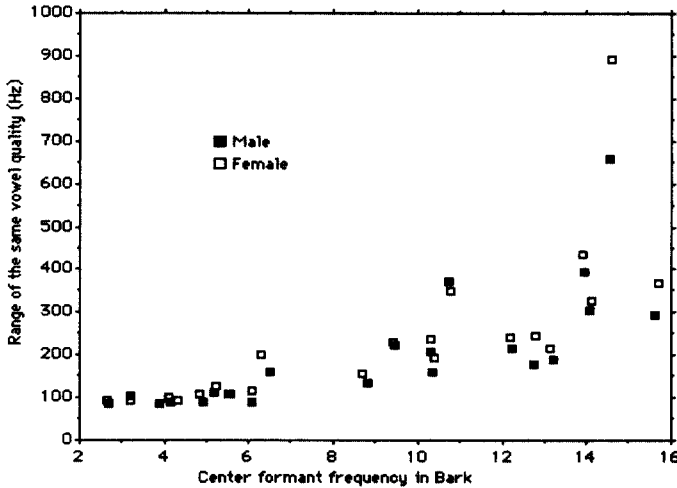
그러면 이렇게 동일하게 지각하는 범위와 청각척도와는 어떤 관련이 있는지 알아보자. 앞서 소개했듯이 사람의 지각특성은 여러 가지 척도로 나타나는데 비록 정현파를 이용한 실험에서 나온 결과이긴 하지만 이들 네 그룹이 지각한 범위와 어떤 관계가 있는지를 살펴봄으로써 청각척도의 문제점을 확인할 수 있을 것이다. 그림 6과 7은 각 포먼트 주파수를 Bark 척도로 환산한 값과 동일하게 지각하는 범위와의 회귀곡선을 보여주고 있다. 참고로 공식 4를 이용하여 음향주파수를 Bark로 환산했다. 여기서 주파수 F는 kHz 단위이다.

$$(4) Zc = 13 \arctan(0.76 * F) + 3.5 \arctan (F/7.5)^2$$

이 분포를 살펴보면 Bark값이 높을 수록 동일하게 지각하는 범위도 넓어짐을 알 수 있으나 선형적으로 완전히 예측할 수 있는 관계를 보이지는 않는다. 즉, 사람의 청각 척도와 비슷한 비선형적인 관계를 대체적으로 나타낼 뿐이지 정확히 일치하는 척도는 아니라고 할 수 있다.



그림 6. 미국인이 동일하게 지각한 모음의 포먼트 중심값을 바크로 환산한 값과 동일하게 지각하는 모음의 주파수 범위.

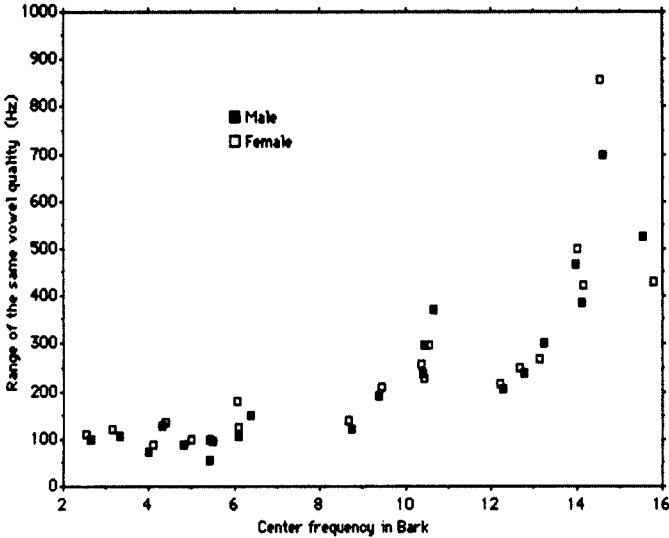


이것은 바로 이 논문의 실험과 Bark 척도를 구한 실험에서 사용한 자극 쌍의 차이에서 나타난 것으로 여겨진다. Bark는 정현파를 소음 신호와 섞어서 들려주고 소음 대 신호의 비율을 조절함으로써 막 신호를 들을 수 있는 경계값을 구한 반면 이 논문에서는 여러 개의 정현파로 구성된 복합음에 의한 자극쌍을 사용했다. 앞서서도 지적했듯이, 단순한 정현파의 비교 실험에서 구한 척도는 실제 합성음에서와 같이 다양한 배음과 포먼트가 상호 작용하여 차폐와 진폭 상승 효과를 주는 음성과 완전히 일치하기는 어렵다. 앞으로 이런 합성 음성신호를 이용하여 사람의 청각 척도를 구한 뒤 실제 음성 인식에 적용해볼 필요가 있다.

#### 4. 결론

지금까지 이 논문에서는 9개의 미국영어 모음의 포먼트 값을 이용하여 자연음과 비슷하게 합성한 뒤 각각의 포먼트 값을 임의의 간격으로 변화시켜서 모델이 되는 음과 동일한지 여부를 27명의 화자에게 들려주는 지각 실험을 했다.

그림 7. 한국인이 동일하게 지각한 모음의 포먼트 중심값을 바르코 환산한 값과 동일하게 지각한 모음의 주파수 범위.



합성할 때는 주파수 대역값에 따라 출력 음성이 과도하지 않도록 Fant의 대역 값을 구하는 공식을 이용하여 보다 자연스러운 모음을 합성할 수 있었다. Fant의 대역값은 인접한 포먼트 사이가 좁을수록 높은 값을 부여하였고 넓을 수록 낮은 값을 부여하여 포먼트가 모여서 진폭이 상승하는 효과를 조절하는 특징을 볼 수 있었다. 지각실험 결과로는 이들이 동일한 모음으로 지각하는 주파수의 중심 값을 구한 뒤 이것과 합성음 모델에 사용한 포먼트 사이의 회귀분석을 실시했는데, 모국어가 서로 다른 미국인과 한국인 모두에게서 거의 동일한 중심주파수를 나타냈다. 이것은 비록 사람마다 동일하게 지각하는 모음의 범위는 다를지라도 청각적 구별은 동일하며, 복잡한 음성신호에서 필요한 정보를 정확히 골라내는 일종의 정규화와 같은 역할을 하는 것으로 여겨진다. 특히, 실험에 참가한 한국인들은 일정한 수준 이상의 영어구사 능력을 갖추었기 때문에 미국인 집단과 거의 비슷한 기준을 가지고 판단했을 것으로 추정된다. 피험자들이 동일하게 지각하는 범위와 합성파일에 사용한 대역값 사이에는 상관관계가 거의 없었다. 그러나 이들의 지각 범위를 상호 비교해본 결과 매우 비슷한 경향을 보였으며 각 그룹의 평균 키와도 연관됨을 알 수 있었다. 따라서, 앞으로 성도의 신체적 특성을 반영하는 화자의 모음 포먼트 값과 구별에 의한 지각실

험의 중심주파수와와의 관계를 조사해볼 필요가 있을 것이다. 마지막으로 동일하게 지각하는 범위와 청각척도를 살펴본 결과 어떤 선형적으로 예측할 수 있는 관계는 나타나지 않았으나, 대체적으로 저주파에서는 높은 해상도를 보이고 고주파로 갈수록 둔감한 사람의 청각 척도를 반영하고 있음을 확인할 수 있었다. 아마도 정현파에 의한 음성 지각실험에서 나온 청각척도는 복합파로된 합성음을 이용한 지각실험에서 나온 이 논문의 결과는 자극 쌍이 달라서 서로 일치하지 않는 것으로 여겨진다. 앞으로 사람의 음성을 이용한 보다 정밀한 지각실험을 통해 음성인식에 필요한 청각척도를 구하는 실험이 실시되어야 할 것이다.

### 참고문헌

- 양병근. 1995. "합성한 한국어 단모음의 지각실험 연구," 언어 20 (3), 127-146.
- 양병근. 1996. "합성한 한국어 이중모음의 지각실험 연구," 언어 21 (3), 829-843.
- 양병근. 1997. "인간의 청각척도에 관한 고찰," 음성과학 2, 125-137.
- Diehl, R.L. 1989. "Remarks on Stevens' quantal theory of speech," *Journal of Phonetics*, 17, 71-78.
- Fujimura, O. and J. Lindqvist. 1971. "Sweep-tone measurements of vocal-tract characteristics," *JASA* 49, 541-558.
- Fant, G. 1968. "Analysis and synthesis of speech processes," in B. Malmberg (ed.) *Manual of Phonetics*, 243-253. Amsterdam: North Holland.
- Fant, G. 1970. *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Fant, G. 1972. "Vocal tract wall effects, losses, and resonance bandwidths," *STL-QPSR* 2-3, 28-52.
- Fant, G. 1975. "Speech production," *STL-QPSR* 2-3, 1-19.
- Fowler, C.A. 1986. "An event approach to the study of speech perception from a direct-realist perspective," *Journal of Phonetics* 14, 3-28.
- Fujimura, O. and J. Lindqvist. 1971. "Sweep-Tone Measurements of Vocal-Tract Characteristics," *JASA* Vol. 49. No 2 (Pt 2) pp. 541-558.
- Klatt, D.H. 1980. "Software for a cascade/parallel formant synthesizer," *JASA* Vol. 67, 971-995.
- Klatt, D.H. and L.C. Klatt. 1990. "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *JASA* 87, 820-857.
- Lindblom, B. 1990. "Explaining phonetic variation: a sketch of the H-H theory," in W.J. Hardcastle and A. Marchal, eds., *Speech Production and Speech Modeling*. Dordrecht: Kluwer Publishers.
- Lindblom, B. and O. Engstrand. 1989. "In what sense is speech quantal?" *Journal of Phonetics* 17, 107-121.
- Lindblom, B. and I. Maddieson. 1988. "Phonetic universal in consonant systems,"

- in L.M. Hyman, L.M. and Li, C.N. eds.. *Language, Speech and Mind*.  
New York: Routledge.
- Miller, J.D. 1989. "Auditory-perceptual interpretation of the vowel.," *JASA* 85,  
2114-2134.
- Norstroem, P.E. and B. Lindblom. 1975. "A normalization procedure for vowel  
formant data," *Paper 212 at the international congress of phonetic  
sciences in Leeds*, August.
- Sharf, B. 1970. "Critical bands," in J.V. Tobias, eds. *Foundations of Modern  
Auditory Theory*. Vol. I. New York: Academic Press.
- Strange, W. 1989. "Dynamic specification of coarticulated vowels spoken in  
sentence context," *JASA* 85, 2135-2153.
- Syrul, A. K. and H.S. Gopal. 1986. "A perceptual model of vowel recognition  
based on the auditory representation of American English vowels," *JASA*  
79 (4), 1086-1100.
- Yang, B. 1992. "An acoustical study of Korean monophthongs," *JASA* 91, 4,  
2280-83.
- Yang, B. 1996. "A comparative study of American English and Korean vowels  
produced by male and female speakers," *Journal of Phonetics* 24, 245-261.
- Zwicker, E. and E. Terhardt. 1980. "Analytical expressions for critical-band rate  
and critical bandwidth as a function of frequency," *JASA* 68, 1523-25.

부산광역시 부산진구 가야동 산 24  
동의대학교 서양어문학부  
614-714  
E-mail: bgyang@hyomin.donggeui.ac.kr  
Tel. 051-890-1227

접수일자: 2000. 2. 21.  
게재결정: 2000. 5. 20.