합성한 한국어 단모음의 지각실험 연구[1]

양 병 곤
(동의대학교)

Yang, Byunggon. 1995. **A Perceptual Study of Synthesized Korean Monophthongs.** *Korean Journal of Linguistics,* 20-3, 127-146. Acoustic parameters of each vowel produced by a healthy male subject were analyzed to synthesize the six vowels by a formant synthesis method until each corresponding synthesized vowel was perceived as almost the same as the naturally produced vowel. Then, each of the first three formant values of the final synthesized vowel was modified and presented to the subject that recorded the tokens and 20 male and female subjects. Results showed that the subject responded to the synthesized vowels with the same standard as he produced. The 20 male and female subjects showed a strong correlation with a consistent response pattern on the synthesized vowels. These results will be helpful in developing automatic speech synthesizers and speech recognition devices. **(Dongeui University)**

1. 머리말

오늘날 정보사회의 발달과 더불어 인간과 기계와의 연계에 관한 연구가 상당히 진척되었다. 특히, 인간과 기계와의 통신의 필요성이 부각되면서, 구미 각국에서는 벌써 인간의 발성을 받아들여, 통역하는 음성 인식과 합성기술의 결정체인 장비를 이미 개발했으며, 지금은 이들 기기를 보다 정확하고 광범위하게 활용하려는 연구에 박차를 가하고 있다. 하지만 한국어에 관한 음성분석과 합성에 대한 연구는 기초적인 자료의 축적도 부족하고 또한,정밀한 합성기를 통

[1]이 연구는 1994학년도 동의대학교 자체 학술 연구 조성비에 의하여 연구 되었음.

하여 합성한 음성에 대한 지각적인 실험연구는 별로 없는 실정이다.

현재 국내에 몇 가지 상용화된 합성기가 있는데 이들이 사용하는 모음 합성은 주로 반음절이나 온음절을 그저 연결시켜서 발성하게 함으로써 기계적인 합성음의 수준에 불과하다. 따라서, 합성된 음은 기계적인 음으로서 이용자에게 거부감을 주고, 이를 자연스러운 음으로 개선하는 데에도 한계가 있다. 국외의 합성 및 지각실험 동향은 Gay(1970), Bond(1982) 등의 선행연구가 있으나 이들의 합성 방법은 Pattern Playback이나, Rockland Digital Speech Synthesizer를 사용했는데, 주로 제 1,2,3 포먼트만을 사용하고, 자연스런 변화폭이나 기타 변수의 조절에 대한 언급이 없다. 비록 미세한 지각적 역할을 하는 제 4 이상의 포먼트 값도 합성음의 자연성을 일부 떨어지게 만든다. 따라서, 이들의 결과가 연구하고자 하는 부분에 대한 청취자의 판단이라기 보다는 오히려 낮은 수준의 합성음에서 나오는 기타 요소의 개입이 판단에 영향을 미쳤으리라 추정된다.

또한 현재 한국어 합성과 인식기 개발에 상당한 투자를 하고 있음에도 괄목할 만한 성과가 나오지 않는 이유 중 하나는 한국어에 대한 음성학적 지식이 부족한 상태에서 행한 실험에서 나온 정확치 못한 파라메터를 사용하는데서 비롯된다고 여겨진다. 특히 지각적인 실험결과를 반영하지 않고 음향 물리적인 단위에만 매달려 해결책을 찾는 데서 문제가 있을 수도 있다. 다시 말해서, 주관적인 사람의 청각기관은 기계와 같이 정확하게 물리적인 음성의 차이를 감지하지 않으며, 고주파보다는 저주파 영역에서 민감한 지각 체계를 보이고 있다 (양 병곤, 1993). 따라서, 사람의 지각 경계선에 대한 연구 결과가 있으면 이러한 시행착오 과정을 상당히 줄일 수 있을 것이다.

본 연구에서는 건강하고 정상적인 청력을 가진 피험자가 발성한 여섯개의 한국어 단모음의 포먼트 및 기본주파수, 진폭, 지속시간 등 네가지 매개변수를 음향학적으로 정밀 분석한 뒤, 이들 파라메터를 이용하여 원음과 거의 구분이 되지 않도록 포먼트합성기를

사용하여 음성을 합성한다. 이렇게 합성된 음의 파라메터중 제 1,2,3 포먼트를 임의의 간격으로 변화시켜서 녹음한 피험자와 대학에서 음성학을 공부하는 정상적인 청력을 가진 20명의 남여 학생들에게 들려주어 지각실험을 행한다. 이들이 자연스럽게 합성된 단모음으로 판단하는 포먼트의 변화 범위가 어떤지를 밝힘으로써, 단모음의 합성 및 자모음 결합에 의한 후속적인 합성 연구에 대한 지각실험의 중요한 기초자료와 연구방법론을 제시함으로써 앞으로 보다 많은 지각실험 연구를 유도한다. 또한, 음성인식에 대한 인식 경계선에 대한 자료를 제시함으로써 이를 활용한 음성합성 및 음성자동 인식장치 개발에 기여하고자 한다.

2. 음성 합성기

음성합성기는 원음을 생성하는 성대와 이를 걸러서 원하는 소리로 만드는 공명기인 성도로 이뤄진 인간의 발화 생성 장치를 모델 (Fant, 1970) 로 음성을 생성한다. 초기에는 조음기관을 직접 모델로 하여 음성을 합성하는 조음합성기 (articulatory synthesizer) 가 쓰였다. 조음합성기에 사용하는 파라메터는 x-선 촬영과 직접 눈으로 관찰한 것에서 나온 조음기관의 모양과 위치이다. 다시 말해서, 주로 혀나 입술 등의 움직임을 측정하여 성도의 모양을 추정하고 이 값을 이용하여 합성기에 응용하였다. 그러나 조음적인 합성은 사람의 발성에 대한 이해를 상당히 도왔지만, 여전히 불충분한 지식때문에 발전에 어려움이 있다. 예를 들어, x-선 촬영은 성도의 측면만을 보여주고 입체적인 변화는 포착하기 어렵다 (Denes and Pinson, 1993). 비록 음향적인 특징은 횡단면과 성도의 길이만으로 어느 정도 조사해 낼 수 있지만 여전히 불완전한 정보임에는 틀림없다. 또한 폐에서 나오는 공기의 흐름, 성대의 길이와 긴장도, 성도의 모양이나 조음기관의 움직임 등에 대한 세부적인 내용을 모두 이해해야만 완전한 합성기 모델을 만들 수 있다.

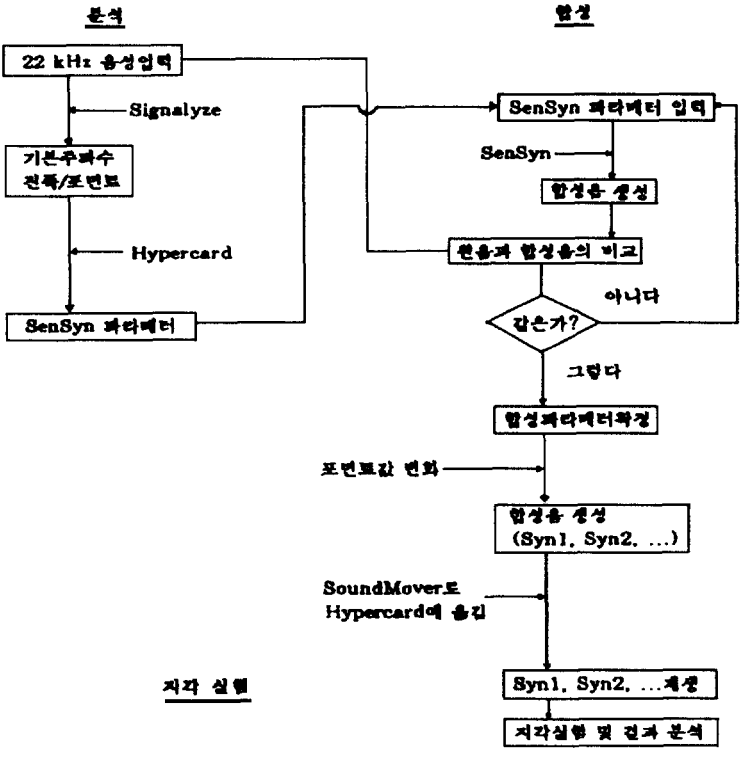
이어서 1950년대의 문양재생기 (pattern playback)가 나오면서 음성의 변화를 전기적인 조절로 모방하게 되었다. 이것은 투명한 벨트 위에 붓으로 인광물질을 사용하여 포먼트의 궤적을 그려나간 뒤 이를 스펙트로그램 생성의 반대과정으로 처리하여 음을 합성하는 방법이다. 이를 통해 음성에 대해 상당히 이해할 수 있게 되었다. 특히 음향적인 변화와 지각적인 변화 사이에는 상당한 차이가 있음을 밝혀낸 점에서 중요한 연구결과를 도출했다 (Denes and Pinson, 1993).

1960년대에 들어와서는 디지탈 컴퓨터의 등장으로 이러한 단순한 포먼트의 궤적을 그리는 대신에 보다 섬세하고 정확한 음성합성기로서 포먼트 합성기가 나왔는데 포먼트 합성기는 유성음용 전자음 생성기 (electrical buzz generator) 와 무성음용 소음 생성기 (hiss generator)로 되어 있고 성도의 포먼트를 모방하기 위해 서너 개의 전자 공명회로를 갖추고 있다. 이런 공명회로는 직렬로 연결하거나 병렬로 연결할 수 있도록 되어있다. 직렬합성기 (cascade synthesizer) 는 대역을 조정하여 포먼트의 진폭을 정밀하게 조정할 수 있으며 모음의 합성에 적합하다. 그러나 자음과 비음화된 모음은 병렬합성기 (parallel synthesizer) 로 합성한다 (Kent and Read, 1992). 현재의 포먼트 방식에 의한 합성기는 60여 가지에 해당하는 다양하고 복잡한 파라메터를 조절하여 원음에 거의 가까울 정도로 합성할 수 있도록 되어 있다. 특히, 이들 파라메터를 독립적으로 조절할 수 있으므로 여러가지 변화를 통한 정밀한 음성 합성 및 지각실험에 적합하다. 실제, 포먼트방식에 의한 합성기를 사용했을 때, 거의 자연음과 구분할 수 없을 정도로 정밀한 합성이 가능하다는 연구결과가 나와있다 (Holmes, 1973). 따라서, 본 논문에서는 포먼트방식에 의한 정밀한 음성합성기를 이용하여 음성을 합성했다.

3. 음성 분석 및 합성

본 연구에서는 건강하고 정상적인 청력을 가진 경상도 출신의 피험자가 느린 속도로 세번씩 발음한 여섯 개의 단모음 /아, 이, 우, 에, 오, 어/를 사용했다. 이들 모음은 /에/를 제외하고는 방언적인 차이가 크지 않으므로 선택을 했다. 세번 가운데 변화가 많은 녹음 시작부와 끝부분은 제외하고 중간부분의 음성을 주로 분석에 활용했다. 음성 분석, 합성, 지각 실험의 모든 과정은 그림 1과 같다.

그림 1. 음성 분석, 합성, 지각실험 과정



음성 입력은 Macintosh의 Signalyze 2.45에 MacRecorder용 Digitizer를 사용하여 22 kHz 표본속도로 양자화하여 입력시켰다. 입력된 신호에서 기본주파수의 분석은 자기상관법 (autocorrelation)에 의하여 수집했다. 수집 간격은 마우스를 이용하여 음절의 총지속시간 (Total Time:TT)을 구한 뒤 이를 다음 공식 1)에 의하여 적어도 80 개 이상의 기본주파수값을 구할 수 있도록 했다.

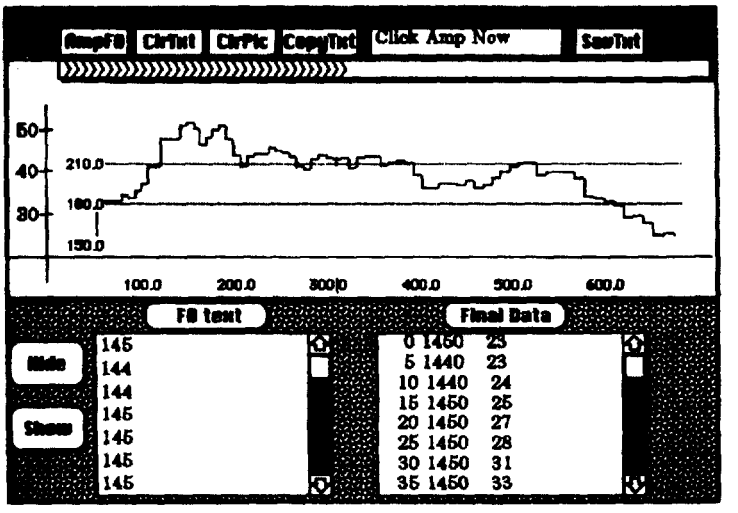
공식 1) 기본주파수 수집 간격 (msec) = (5 X TT) /400

공식 1)에서 400은 합성기에 들어갈 총 지속시간이며 5는 합성기의 파라메터 생성 간격이다. 이렇게 구한 수집 간격은 약 5~10 msec 마다 하나의 값을 취한 것이 된다. 또한 기본주파수의 범위를 100~200 Hz 범위로 제한하여 수집하였으며 이 때 범위를 벗어난 극단치는 파형을 확대하여 음파의 각 성대 떨림사이의 시간을 측정하여 수정했다. 이렇게 수집된 자료를 아스키 화일로 저장했다. 이 화일을 Hypercard에 불러와, 앞뒤에 생성된 무한값을 제거한 뒤 80여 개의 값을 그림 2와 같이 Hypertalk로 만든 기본주파수 및 진폭입력 프로그램에 입력하고, 거기에 덧붙여, Signalyze에서 분석하여 생성한 각 모음의 진폭윤곽 (envelope) 을 옮겨와서 커서로 궤적을 따라가면서 자동으로 진폭을 입력했다.

이렇게 완성된 자료의 윗부분에 Klatt 합성기의 기본값을 더하여 이를 하드 디스크에 아스키 화일로 저장했다. 다음으로는 Klatt & Klatt (1990) 의 포먼트합성방식을 채택한 Macintosh용 합성소프트웨어인 SenSyn 1.0으로 열어서 기본주파수와 진폭에 대한 화일을 자동으로 입력하고, 동시에 Signalyze에서 분석한 네 개의 포먼트 값을 일일이 입력했다. 각 모음의 포먼트 값은 Signalyze를 이용해 해당 모음의 스펙트로그램을 생성하여 모음의 포먼트 위치를 확인한 뒤 안정된 부분에 해당하는 총 지속시간의 1/3 지점에서 120 Hz 대역의 스펙트럼을 생성하여 구했다. 이 때, 이 시점의 좌우의 스펙트럼도 분석하여 동시에 한 화면에 나타내어 비교하면서

정밀한 값을 구했다.

그림 2. 기본주파수 및 진폭입력 프로그램의 예



또한 포먼트는 제1에서 제6포먼트까지 수치를 구하여 합성에 이용했고, 낮은 강도의 포먼트일 경우에는 기본값을 입력하고 주파수역이 넓도록 높은 대역값을 입력했다. 표 1은 이렇게 생성한 모음 [이]의 실제 합성하기 직전의 60가지 파라메터 설정부분과 시간의 변화에 따른 기본주파수와 진폭의 변화된 부분의 입력 파일의 예를 보이고 있다. (각 파라메터의 정의에 대한 설명은 Klatt and Klatt (1990)을 참고하기 바람.)

표 1. 모음 [이]의 합성용 파라메터가 설정된 예

```
Synthesis specification for file: 'i.par'
SenSyn      Version 1.0
Total number of waveform samples = 8000
CURRENT CONFIGURATION:
   60 parameters
```

SYM	V/C	MIN	VAL	MAX	SYM	V/C	MIN	VAL	MAX
DU	C	30	400	5000	UI	C	1	5	20
SR	C	5000	20000	20000	NF	C	1	6	6
SS	C	1	2	3	RS	C	1	8	8191
SB	C	0	1	1	CP	C	0	0	1
OS	C	0	0	20	GV	C	0	60	80
GH	C	0	60	80	GF	C	0	60	80
F0	V	0	1000	5000	AV	V	0	60	80
OQ	v	10	50	99	SQ	v	100	200	500
TL	v	0	0	41	FL	v	0	0	100
DI	v	0	0	100	AH	v	0	0	80
AF	v	0	0	80	F1	v	180	330	1300
B1	v	30	150	1000	DF1	v	0	0	100
DB1	v	0	0	400	F2	v	550	2520	3000
B2	v	40	70	1000	F3	v	1200	3230	4800
B3	v	60	80	1000	F4	v	2400	3600	4990
B4	v	100	100	1000	F5	v	3000	4500	4990
B5	v	100	200	1500	F6	v	3000	4990	4990
B6	v	100	500	4000	FNP	v	180	280	500
BNP	v	40	900	1000	FNZ	v	180	280	800
BNZ	v	40	900	1000	FTP	v	300	2150	3000
BTP	v	40	900	1000	FTZ	v	300	2150	3000
BTZ	v	40	900	2000	A2F	v	0	0	80
A3F	v	0	0	80	A4F	v	0	0	80
A5F	v	0	0	80	A6F	v	0	0	80
AB	v	0	0	80	B2F	v	40	250	1000
B3F	v	60	300	1000	B4F	v	100	320	1000
B5F	v	100	360	1500	B6F	v	100	1500	4000
ANV	v	0	0	80	A1V	v	0	60	80
A2V	v	0	60	80	A3V	v	0	60	80
A4V	v	0	60	80	ATV	v	0	0	80

Varied Parameters:

Time	F0	AV
0	1660	30
5	1720	32
10	1700	34
15	1700	35
(중간부분 생략)		
385	1180	22
390	1180	19
395	1180	13

이렇게 생성된 각 화일은 PowerPC로 합성하였으며 각 화일 하나당 약 1분 정도 소요되었다. 주파수 대역은 기본값을 그대로 사용하여 합성한 뒤 이를 Signalyze로 분석하여 원래의 음성파형에 대한 스펙트로그램과 시각적으로 비교함과 동시에 녹음한 피험자에게 음을 여러 번 들려주어 원음과 차이가 거의 없다는 판단이 날 때까지 되풀이해서 파라메터를 조절했다. 주파수대역은 포먼트 궤적의 가늘고 굵은 정도를 조정할 수 있게 되어있다. 이런 수정 과정을 거쳐 최종적으로 확정한 포먼트 주파수 및 주파수대역값은 다음 표 2에 나타나 있다.

표 2. 합성기에 입력한 각 모음의 포먼트 주파수 및 주파수대역. F_1은 제1포먼트를 나타내며, B_1은 제1포먼트의 주파수대역을 나타낸다.

모음	F_1	B_1	F_2	B_2	F_3	B_3	F_4	B_4
이	330	150	2520	70	3230	80	3600	100
아	800	110	1345	120	2710	120	3800	150
오	540	90	900	110	2600	150	3500	100
우	370	80	730	90	2600	60	3550	100
에	550	60	2100	90	2900	150	3600	300
어	620	70	1200	90	2950	90	3600	140

다음으로는 원음에 가까운 합성음의 포먼트 주파수값을 다음과 같이 각각 변화시켜서 지각 실험용 합성음을 생성했다. F_1은 표 2의 각 원음의 포먼트 주파수값에서 아래 위로 50 Hz간격으로 200~950 Hz 범위내에서 (F_2)보다는 낮게 합성했다. F_2는 50 Hz간격으로 각 원음의 포먼트 주파수값의 아래 위 500 Hz범위내에서 합성했다. F_3의 경우에는 50 Hz간격의 변화에 대해 지각적으로 별 차이를 보이지 않았던 사전 실험에 바탕을 두고 100 Hz간격으로 각 원음의 주파수값의 아래 위 500 Hz범위에서 합성음을 만들었다. 이렇게 생성한 화일은 표 3의 예와 같이 총 270개였다.

표 3. 합성한 음성의 포먼트 및 주파수 대역의 예. 원음에 가까운 합성음의 파라메터 중 제 1,2,3 포먼트를 임의의 간격으로 변화시켰다.

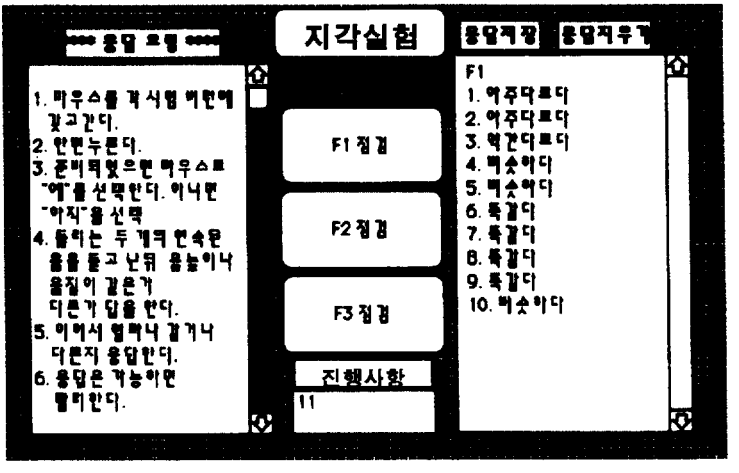
모음	F_1	B_1	F_2	B_2	F_3	B_3	F_4	B_4
이	200	150	2520	70	3230	80	3600	100
이	250	150	2520	70	3230	80	3600	100
이	300	150	2520	70	3230	80	3600	100
이	350	150	2520	70	3230	80	3600	100
이	400	150	2520	70	3230	80	3600	100
(중간부분 생략)								
아	800	110	1500	120	2800	120	3800	150
아	800	110	1500	120	2900	120	3800	150
아	800	110	1500	120	3000	120	3800	150
아	800	110	1500	120	3100	120	3800	150
아	800	110	1500	120	3200	120	3800	150

4. 지각실험 및 결과 분석

4.1. 녹음한 피험자의 지각실험

앞에서 합성한 음성은 원음과 비교하여 응답하는 방식으로 지각실험을 행했다. 실험환경은 Macintosh LC475 에 Hypertalk로 실험환경에 대한 프로그램을 그림 3과 같이 만들었다. 각 모음에 대한 합성음은 AIFF (Audio Interchange File Format)으로 되어 있으므로 이를 먼저 SoundWave파일로 변환시켜 SoundMover를 이용하여 Hypercard에 옮겼다. 피험자에게 각 모음마다 포먼트가 변형된 합성음을 조용한 방에서 주파수 반응 대역이 20~20,000 Hz인 EMMA-300 헤드폰을 이용하여 들려주었다.

그림 3. 지각 실험을 위한 Hypercard 실험 환경 화면



피험자는 0.5초 간격으로 두 개의 신호를 듣고 이를 판단했다. 사전 실험에 의하면 두 신호의 간격이 너무 가깝거나 멀 때에는 중첩이나 단기기억의 문제점 때문에 판단에 어려움이 있음을 발견했다. 판단은 먼저 "비슷하다"와 "다르다"중 하나를 마우스로 선택한 뒤 "비슷하다"를 택했을 때는 "꼭같다"와 "비슷하다" 중 하나를, "다르다"로 선택했을 때는 "다르다"와 "아주 다르다" 중 하나를 반드시 선택하도록 되어있다. 피험자가 선택하면 1초후에 다음 신호가 재생되도록 하고, 마지막 응답이 끝나면, 피험자의 답안을 모두 모아서 하나의 텍스트 화일이 자동으로 생성되도록 Hypertalk로 프로그램을 만들었다.

지각 실험 결과를 살펴보면 다음과 같다. 먼저, 원음과 비교해 봤을 때, "비슷하다"나 "꼭같다"를 선택한 합성음의 주파수 범위는 다음 표 4와 같이 나타났다.

표 4. 원음과 합성음의 비교 지각 실험 결과. 두 개의 주파수값 중 앞의 수치는 비슷하다고 느끼기 시작하는 주파수이며, 뒤의 수치는 다르다로 느끼기 시작하는 주파수 직전의 값이다.

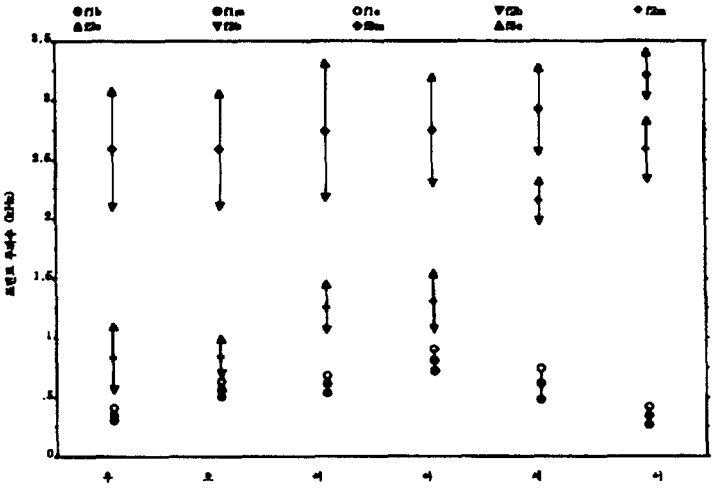
모음	F_1	범위	F_2	범위	F_3	범위
이	200~400	200	2350~2650	300	3000~3400	400
우	200~350	150	600~ 800	200	2100~3100	1000
에	400~550	150	1800~2200	400	2300~3300	1000
오	450~550	100	650~ 900	250	2100~2900	800
어	500~700	200	1150~1350	200	2500~3300	800
아	800~900	100	1100~1550	450	2400~3200	800
	평균	150		300		800

표 4에서 보듯이 각 모음의 주파수 범위는 F_1에서 F_3으로 갈수록 동일하다고 지각하는 범위가 넓어져 간다. 예를 들어, F_1에서의 비슷하거나 똑같다고 판단하는 주파수 범위의 평균은 150 Hz 인데 반해, F_2에서는 이의 두 배에 해당하는 범위를 갖고, F_3에서는 F_1의 다섯 배가 넘는 넓은 주파수 대역의 변화에도 같게 지각하고 있음을 알 수 있다. 이러한 결과는 지각의 범위가 저주파에서 고주파로 갈수록 넓어진다는 미국인들의 지각범위에 대한 실험 결과와도 일치한다 (Denes and Pinson, 1993). 따라서, 각 포먼트의 주파수를 변형하여 지각실험을 할 때도 이러한 결과를 염두에 두고 필요 이상으로 세부적인 간격으로 나눌 필요가 없다. 표 4에서 모음 [이]의 F_3 범위는 400 Hz로 F_2의 주파수가 이미 2500 Hz근처에 위치해 있으므로 지각적으로 그 범위가 줄어졌기 때문이라고 할 수 있다. 이는 Theory of Adaptive Dispersion (Lindblom, 1990)에서 제시된 바 있는 한 언어내에서 각 모음의 위치는 모음 사이에 충분히 지각적으로 대조가 되도록 배치한다는 지적에서의 충분한 지각적인 거리의 유지와 연관된다. Lindblom(1990)은 세계 여러 언어에 나타나 있는 모음을 조사하면서, 언어 마다 일정한 수의 모음이 있을 때 이들 사이에는 지각적으로 충분한 거리를 유지할 수 있도록 배치된다는 이론을 제안했다. 이런 관찰에 대해 40명의 한국인과 미국인 남여가 발성한 한국어 및 영어의 모음도에서의 포먼트간의 지각적인 거리를 분석해 본 결과 각언어내에서는 지각적인 충분한 거리를 확보하고 있음을 발견했다 (Yang, 1996). 본 자료에서도 모든 포먼트의 자료가 서로 중첩되어 있지 않는 것을 관찰할 수 있다.

표 4에서의 자료와 표 2에서의 자료를 통계적으로 비교해보면 흥미로운 결과를 발견할 수 있다. 표 2의 값과 표 4의 간격 사이의 중심점을 계산한 뒤 이들 사이의 상관관계를 StatView 512+를 이용하여 통계적으로 처리했을 때, F_1은 상관계수가 0.987로 거의 일치하며, F_2는 0.996, F_3는 0.957의 아주 높은 상관계수를 보였다. 이는 결국 우리가 말을 하는 기준과 듣는 기준은 동일하며, 두 가지 다른 행위가 아님을 증명해 주는 중요한 자료라 하겠다. 이들 자료

를 보다 시각적으로 나타내 보기 위해서 위의 표에 나타나 있는 각 모음의 시작 부분(f_nb)과 끝 부분(f_ne)의 값의 평균값(f_nm)을 F_2를 중심으로 정렬한 뒤 이를 그림 4와 같이 나타내 보았다.

그림 4. 녹음한 피험자의 원음과 합성음의 비교 지각 실험 결과를 나타낸 그래프. y축은 제 1,2,3 형성음 주파수를 나타낸다.



F_2는 혀의 이동을 나타내는 중요한 단서인데, 주파수가 낮을수록 후설모음임을 알 수 있고, 높을수록 전설모음으로 나타난다. 실제 그림 4에서는 후설모음인 [우, 오]와 중설모음인 [어,아]와 전설모음인 [에, 이]의 순서로 배열되었다. 모음 [아]는 우리말에서는 모음 삼각도의 꼭지 부분에 해당하기 때문에 전설모음과 후설모음 사이에 위치해 있음을 알 수 있다 (Yang, 1992). 이 그림에서 [에]의 범위는 F_3과 거의 중첩될 정도로 가까이 합성하여도 같은 음으로 들었는데, 이는 [에,애]를 경상도 방언에서는 거의 구분하지 않는 경향 때문으로 여겨진다. 실제 /에,애/의 구분은 경상도 방언뿐만아니라 현재의 한국어에서는 거의 사라진 상태이다. 이 두 모음이 중화 현상을 일으킨 이유는 /에/는 전설모음중 중모음이고 /애/는 전설모

음중 저모음이어서 발음 편이상 전설 중모음으로 통합이 되었기 때문이다.

4.2. 남여 학생들의 지각실험

다음으로는 지각 실험자료를 영어음성학 과목을 이수하고 있는 동의대학교 3학년 남여 학생 20명에게 들려주었다. 이들은 주로 부산 경남 지역 출신으로 경상도 방언을 사용하고 있다. 이들의 지각실험은 원음과 가장 가까이 합성한 음과 이 음의 포먼트를 일정 간격으로 변화시킨 음을 0.5초 간격으로 들려주고 비슷한지 다른지를 판단하도록 했다. 두 비교음 사이에는 1초 간격을 두고 컴퓨터에서 재생되는 음성을 컴퓨터에 연결하여 바로 녹음하였으며, 이를 조용한 강의실에서 녹음기로 재생하여 들려 주었다. 각 피험자는 연속으로 제시되는 두 개의 합성된 음을 듣고 "같다" 와 "다르다"의 표시 가운데 하나를 선택하여 주어진 답안지의 해당부분에 동그라미 표시를 하도록 하였다. 지각실험 중 각 포먼트가 바뀔 때마다 피험자에게 재생되고 있는 음이 몇 번인지 실험자가 확인시켜 주었다. 총 29 명의 응답자 중 남여 20명 분을 최종 자료로 설정했다. 이들은 대체로 비슷한 지각범위를 보인 피험자와 남여비율을 고려하여 선정했다. 이들의 실험 결과 중 남자 10명의 자료는 표 5에 여자 10명의 자료는 표 6에 나타나 있다.

표 5. 남자 10명의 지각실험 결과. 원음과 최대한 가까이 합성한 음과 포먼트를 변화시킨 합성음이 비슷하다고 느끼는 최저 주파수 값과 최고 주파수 범위를 나타냈다.

모음	F_1	범위	F_2	범위	F_3	범위
우	300~410	110	570~1100	530	2100~3080	980
오	510~630	120	700~990	290	2110~3060	950
어	540~680	140	1070~1450	380	2170~3310	1140
아	720~910	190	1080~1540	460	2300~3200	900
에	480~740	260	1980~2310	330	2560~3280	720
이	260~420	160	2330~2830	500	3030~3400	370
평균		163		415		843

표 6. 여자 10명의 지각실험 결과

모음	F_1	범위	F_2	범위	F_3	범위
우	300~390	90	560~990	430	2100~3000	900
오	500~730	230	640~930	290	2100~3060	960
어	500~660	160	1110~1470	360	2300~340	1100
아	720~950	230	1040~1490	450	2310~3190	880
에	470~820	350	1980~2490	510	2440~3390	950
이	220~410	190	2290~2970	680	3000~3500	500
평균		208		453		882

이들의 시작점과 끝점의 평균값을 내어 세 꼭지점을 이으면 그림 5, 그림 6과 같다. 그림 5는 남자 10명의 지각 실험 결과를 보이고 있고, 그림 6은 여자 10명의 지각 실험 결과를 보이고 있다.

그림 5. 남자 10명의 지각 실험결과

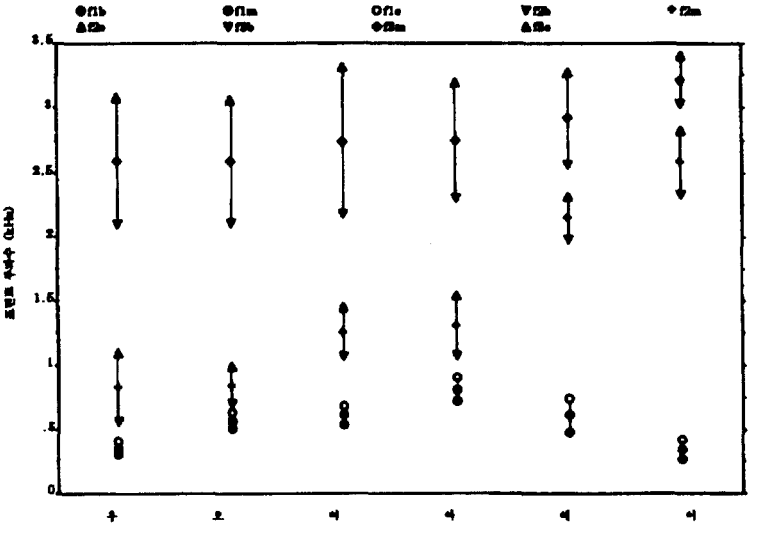
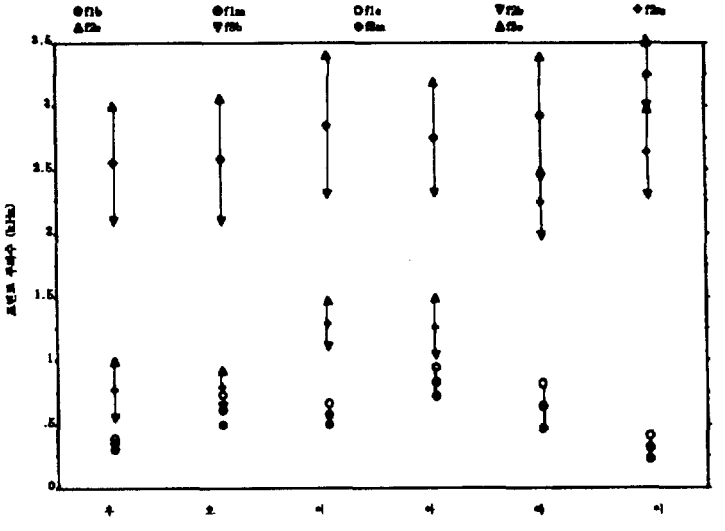


그림 6. 여자 10명의 지각 실험 결과



위 그림에서 특기할 사항은 앞절의 실험 결과와 같이 경상도 방언에서는 「에」와 「애」의 구분을 거의 하지 않으므로 [에]의 포먼트가

상당히 변해도 청각적으로 동일한 음으로 판단하는 지각범위가 상당히 넓음을 알수 있다. 위 실험 결과는 대체적으로 앞 그림 4와 같은 유형을 보이고 있으며, 이들 시작부분과 끝부분의 남여 각각 평균값을 내어 이를 앞의 녹음한 피험자의 실험결과와 상관비교를 해 본 결과, F_1의 경우에는 남자그룹과는 0.969 여자 그룹과는 0.934의 높은 상관 계수가 나왔으며, F_2에서는 남여그룹 각각 0.996, 0.994였고, F_3에서는 0.913과 0.955가 나왔다. 제 3포먼트에서의 약간 낮은 결과는 모음 [어]에서 남자 그룹의 결과가 약간 낮게 나타났기 때문인데 이는 제 3포먼트의 지각 범위를 생각해 볼 때 그다지 큰 차이가 아님을 알 수 있다. 이들 결과는 상당히 상관도가 높은 수치로써 앞 절의 녹음한 피험자의 실험결과의 신뢰도를 확인해 주고 있다. 또한 남여 각 그룹간의 모든 포먼트값을 비교해 본 결과 0.996의 높은 상관 계수를 보였는데 주관적인 지각실험도 상당히 일관성있는 결과를 보일 수 있음을 알 수 있다.

5. 맺음말

지금까지 이 논문에서는 경상도 방언을 쓰는 한 피험자의 발음을 녹음하여 이와 가장 가까운 합성음을 만든 뒤 이 합성모음의 중요한 음향학적 단서인 포먼트를 일정한 간격으로 변화시켜 음성학자와 남여 각 10명씩의 피험자에게 들려 주는 지각실험을 통해 같은 음으로 판단하는 범위를 찾아내었다. 녹음한 피험자의 실험에서는 우리가 말을 하는 기준과 듣는 기준은 동일하며, 두 가지 다른 행위가 아님을 증명해 주는 중요한 결과가 밝혀 졌다. 또한 남여 피험자의 실험도 이와 거의 일치하는 결과를 보임으로써 주관적인 지각 실험의 신뢰도를 확인할 수 있었다. 특히, 사람의 지각 범위는 저주파부분에서는 대역을 좁게 세분하여 들으며, 고주파에서는 상당한 주파수의 변화에도 동일한 음으로 듣고 있다는 결론을 내릴 수 있으며, 이는 곧 각 포먼트 측정에 있어서 얼마나 정밀한 수치

값이 지각적으로 유효한 값인가를 보여주는 중요한 결과라 하겠다.

본 연구에서는 주로 세 포먼트 중 두개를 고정하고 하나의 포먼트 값만 변형하여 지각실험을 행했는데 앞으로 이들 세 포먼트의 값을 동시에 일정한 간격으로 변화시켜서 지각 실험을 하는 것이 필요할 것이다. 또한 각 단모음의 주파수가 이중모음에 활용될 때 어떠한 변화를 가져오는지 이들을 합성한 뒤 지각실험을 했을 때 어떠한 결과를 가져올것인가에 대한 후속연구는 현재 진행중이다. 지금까지의 연구 결과는 합성한 한국어 단모음에 대한 지각적인 실험의 중요한 기초자료를 제공하고, 또한 국내에서는 거의 없는 합성에 의한 지각실험 연구분야에 상세한 연구방법론을 제시함으로써 보다 많은 지각실험 연구를 유도하고, 음성인식에 대한 인간의 경계선에 대한 자료를 제시함으로써 이를 활용한 음성자동 인식장치 개발에 널리 활용될 수 있을 것이다.

참고문헌

양병곤. 1993. *음성학 입문*. 부산:진영문화사.

Denes, P.B. and E.N. Pinson. 1993. *The Speech Chain*. New York: W.H. Freeman and Company.

Fant, G. 1970. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Gay,T. 1970. "A perceptual study of American English diphthongs." *Language and Speech* Vol. 13 (2), 65-88.

Klatt, D.H. and 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.

Holmes, J. N. 1973. "Influence of glottal waveform on the

naturalness of speech from a parallel formant synthesizer." *IEEE AU-21*, 298-305.

Kent, R.D. and C. Read. 1992. *The Acoustic Aanalysis of speech.* San Diego: Singular Publishing Group, Inc.

Lindblom, B. 1990. "Explaining phonetic variation: a sketch of the H&H theory." *Speech Production and Speech* (W.J. Hardcastle and A. Marchal, eds.), 403-439. Dordrecht: Kluwer Academic Publishers.

Yang, B. 1992. "An acoustical study of Korean monophthongs." *Journal of Acoustical Society of America* 91(4), 2280-2283.

Yang, B. (1996년 4월 인쇄예정). "A comparative study of American and Korean monophthongs produced by male and female speakers." *Journal of Phonetics.*

부산시 진구 가야동 산 24
동의대학교 인문대학 영어영문학과
614-714
E-mail: bgyang@turtle.dongeui.ac.kr
Fax: (051)-890-1582

접수일자: 1995. 5. 16
게재결정: 1995. 9. 15