

합성한 한국어 이중모음의 지각실험 연구*

양 병 곤
(동의대학교)

Yang, Byunggon. 1996. A Perceptual Study of Korean Diphthongs Synthesized. *Korean Journal of Linguistics*, 21-3, 829-843. Acoustic parameters of 10 Korean diphthongs produced by a healthy male subject were analyzed to synthesize diphthongs by a formant synthesis software until each synthesized diphthong was perceived almost the same as the naturally produced diphthong. Then, each of the first three formant values of the diphthong files for synthesis was uniformly or partially modified and presented to 10 male and female listeners. Results showed that the listeners had a certain perceptual range for each formant of the diphthong: The lower the formant, the wider the perceptual range. In the uniformly modified and synthesized diphthongs, similar perceptual ranges to those of the previous study on Korean monophthongs were found. This study will be helpful in developing automatic speech synthesizers and speech recognition devices. (Donggeui University)

1. 머리말

오늘날 첨단 기기의 발달과 더불어 인간과 기계와의 연계, 예를 들어, 자동통역기와 같은 장치의 개발에 상당한 투자를 하고 있다. 그러나 이런 장치의 개발에 필요한 기본적인 연구자료가 충분히 확보되지 않아서 어느 정도의 발전을 이루고 난 다음에는 더 진척이 되지 않는 경우가 많다. 다시 말해서, 한 분야에 꾸준한 발전을 이루기 위해서는 상당한 기본 연구가 밑바탕이 되어야 한다. 예를 들어, 복잡하고 다양한 언어처리를 가능하게 하려면 각 자모음에 대한 기본적인 자료가 확보되어 있어야 한다. 이렇게 부족한 자료 가운데 하나가 이중모음인데, 이중모음에 대한 연구는 적절한 합성기나 분석기와 같은 실험도구의 부족으로, 또한 전이음의 시작부터 중핵모음의 도달부까지 계속하여 변하기 때문에 일정한 결론을 내리기가 어렵다는 이유에서 현재 연구된 자료가 별로 없다.

이중모음은 학자들마다 조음적, 지각적, 음향적 기준 또는 이들 기준의 조합에 따라 다르게 정의하고 있는데(전상범 1988; 조성식 1990; Catford

*이 연구는 1995학년도 한국학술진흥재단 자유공모과제 연구조성비에 의하여 연구되었음.

1977; Gerber 1972; Ladefoged/Maddieson 1990; Trager/Smith 1951), 이들 정의의 공통된 사실은 이중모음이란 두 개의 변별적인 요소가 있고 그 사이에 빨리 변하는 전이부가 있다는 것이다. 따라서, 이중모음이란 두 개의 다른 목표음을 가지고 그 사이의 최단 거리 부분을 전이하는 부분으로 정의할 수 있다(양병곤 1993). 이들 이중모음의 종류에는 조음기관인 혀의 이동 방향과 이동 거리, 음향적 성분소 및 울림도 등과 같은 기준에 따라 여러 가지로 나눌 수 있다. 특히, 혀의 이동 방향에 따라 구분하면 한국어에서는 전이음 /j, w/ 다음에 울림도가 큰 중핵모음이 오는 /와, 야/ 등과 같은 전이부와 한 개의 중핵모음으로 구성된 상승이중모음이 많이 쓰이고, 영어에서는 그 반대의 구조인 /aw, ai/와 같은 하강이중모음이 많이 쓰이고 있다. 이 때 두드러진 모음부분이 음절의 중핵이 되고 약한 도입부의 전이음은 음절의 부음으로 분류하기도 한다(김재민 외 1985: 355).

이중모음에 대한 지각 실험 연구에는 Gay(1970)와 Bond(1978, 1982) 등이 있는데 이들은 주로 이중모음의 전이 부분에 대한 지각실험을 했다. Gay(1970)의 실험결과에서는 주파수 변화보다는 전이부의 지속시간 변화가 모음과 이중모음의 구분 단서이고, 포먼트 주파수의 변이속도는 이중모음의 고정된 특징임을 밝혔다. 덧붙여, 전이음의 구체적인 변화 경로와 지속시간이 이중모음의 구별에 중요한 단서임을 밝혔다. Bond(1978)는 미국영어의 /ar, au, or/를 합성하여 안정부분의 지속시간을 고정하고 전이음 부분의 지속시간을 6-150ms까지 변화시켰을 때의 인식실험을 했는데, 각 목표모음 사이의 간극을 48ms로 했을 때 두 개의 모음으로 판단했고, 10ms일 때 모두 이중모음으로 판단했다. 그의 연구는 청자들이 전이음 지속시간의 절대 값뿐만 아니라 발화속도에 따른 전이음 지속시간을 추정하여 이중모음을 구별함을 밝혔다. Bond(1982)는 전이음의 지속시간을 비교적 길게 하면 앞뒤 안정된 모음부분에 상관없이 이중모음으로 판단했고, 전이음이 10ms이라도 앞뒤의 안정 부분이 길면 이중모음으로 판단함을 밝혔다. 포먼트 변화의 지속시간을 바꾸었을 때의 지각적 판단에 대한 연구로는 O'Connor 외(1957)과 Liberman 외(1956)이 있는데, 이들은 전이부의 지속시간이 서로 다른 소리를 구분하는 중요한 단서로 작용함을 밝혔다. O'Connor 외(1957)에서는 /j, w/의 전이부가 완전 모음이라는 인상을 주지 않게 하려면 40ms를 넘을 수 없음을 밝혔다. 또한 100ms 이상의 포먼트 전이구간은 이중모음의 인상을 준다는 것을 밝혔다. Liberman 외(1956)은 폐쇄음-전이음-이중모음의 연속을 전이부의 지속시간의 변화에 따라 지각되는 부분을 조사했다. /we/는 변이시간이 150-200ms일 때 두 개의 다른 모음의 중첩인 /ue/로 지각되기 시작했다. 이는 형성음 전이부의 지속시간이 이중모음 식별의 중요한 단서임을 보여준다. 지금까지의 이중모음의 지각실험 연구는 주로

비교적 간단히 조작할 수 있는 지속시간의 변화에 대한 지각 연구이고 주로 분석에 치중하였기 때문에 포먼트 합성 방식에 의한 정밀한 파라미터 조정을 통한 실험이 거의 없다고 할 수 있다.

따라서, 본 연구에서는 정상적인 피험자가 발성한 이중모음의 기본주파수, 진폭, 포먼트 등의 음향적 파라미터를 정밀 분석하여 원음과 같이 합성한 뒤, 합성용 파일에서 이중모음의 중요한 음향적 파라미터인 각 포먼트를 아래위로 일정하게 증감시키거나, 이중모음의 중요한 단서인 전이부만을 일부 변형하여 합성한 뒤, 원음과 가까운 합성음과 비교하여 들려줌으로써, 동일하게 지각하는 범위를 찾아내고, 아울러, 이중모음의 그룹별로 분류하여 이들의 특성을 분석한 뒤, 이 특성을 통합하여 합성한 음의 지각에 관한 분석을 한다. 구체적으로, 본 연구에서는 다음과 같은 연구 문항에 대한 결과를 지각실험을 통해 밝히고자 한다.

- (1) 각 포먼트를 전 구간에 걸쳐 일정한 간격으로 변화시켰을 때, 동일한 이중모음으로 지각하는 범위는?
- (2) 세 개의 포먼트 중 한 개의 전이음 구간을 변화시키고 나머지 두개를 그대로 두었을 때, 동일한 이중모음으로 지각하는 범위는?
- (3) 전이음 /j/로 시작하는 이중모음인 /야, 여, 요, 유/ 그룹과 전이음 /w/로 시작하는 /와, 위, 웨, 위/ 그룹의 특징과 각 그룹의 평균값을 취하여 합성했을 때, 어떤 음으로 들리는가?

2. 음성분석 및 합성

본 연구에서는 건강하고 정상적인 청력을 가진 경상도 출신의 피험자가 느린 속도로 또박또박 발음한 열 개의 이중모음 /야, 여, 요, 유, 예, 와, 위, 웨, 위, 의/를 사용했다. 이들의 음성분석 합성 지각실험의 모든 과정은 단모음의 합성과정(양병곤 1995)과 유사하다. 다만, 이중모음은 시간에 따른 포먼트의 변화가 심하므로 이러한 동적인 변화를 보다 정밀하게 포착하는데 치중했다. 음성입력은 Macintosh의 Signalize 2.45에 22kHz의 표본속도로 마이크로 입력했다. 입력시 고주파 부분의 스펙트로그램의 선명도를 확보하기 위해 고성능 오디오의 필터기능을 이용하여 고주파 부분을 약간 증폭시키고 저주파 부분은 감폭시켰다. 음성분석은 총 지속시간을 구해 적어도 80개 이상의 기본주파수 값을 자기상관법에 의해 수집했으며 단모음에서 사용한 Hypertalk로 작성된 프로그램에 덧붙여 이중모음의 특성인 포먼트 전이음의 특징을 포착하기 위한 새로운 프로그램을 제작했다. 그림 1은 포먼트 전이부에 관한 자동 추적 프로그램의 화면 일부를 보이고 있다.

그림 1. 모음 [아]의 스펙트로그램의 F_1 과 F_2 의 동적인 포먼트 전이부분을 포착하기 위한 프로그램

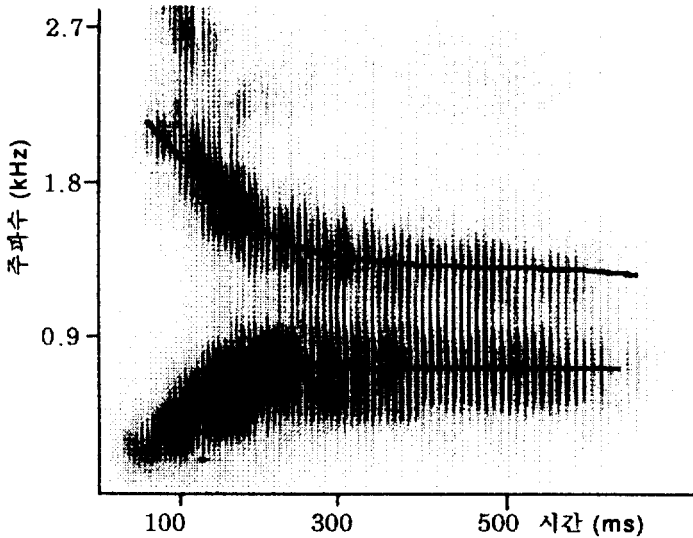


그림 1에서처럼 포먼트 궤적 추적을 통해 보다 정밀한 값을 구하기 위해 간에 스펙트로그램을 120Hz대역으로 설정하여 분석한 F_1 , F_2 를 한 화면에 확대하여 포먼트 띠의 중심부를 미끄러지듯이 일정한 속도에 의해 마우스를 움직이면 자동적으로 포먼트 값이 입력되도록 제작했다. 덧붙여, 각 포먼트 궤적의 변화율을 모두 같은 속도로 수집할 수 있도록 스펙트로그램 위에 궤적을 따라가는 기준점을 일정한 시간 간격으로 화면에 표시하게 하여, 그 점을 따라 가면서 현재 입력되고 있는 시간점을 확인하므로써, 너무 빠르거나 느리지 않게 하여 거의 같은 시간점에서 F_1 - F_3 가 입력되도록 했다.

이어서 Klatt합성기의 기본값을 파일의 앞부분에 붙인 뒤 이를 하드 디스크에 아스키 화일로 저장하고, Klatt/Klatt(1990)의 포먼트 합성 방식을 채택한 Macintosh용 합성 소프트웨어인 SenSyn 1.0로 합성했다. 합성용 기본값 설정은 400ms의 지속시간을 정해서 총 8000개의 표본값을 갖게 하고, 6개의 포먼트값을 줄 수 있도록 지정했다. F_3 와 F_4 가 아주 가까울 때 방사 함수에 의한 과도한 신호의 찌그러짐이 있을 것을 예상해 500Hz 전후의 적절한 거리를 유지하는 범위 내에서 조절했다. 표 1은 이렇게 작성한 이중모음 [아]의 합성하기 직전의 60가지 기본 파라미터 설정부분과 각 시간점이 5ms씩 증가함에 따라, 기본주파수(f_0), 진폭(AV), 포먼트(F_1 , F_2 ,

F₃)의 변화되는 값을 보인다(각 약어 설명은 Klatt/Klatt(1990)을 참조하기 바람).

표 1. 이중모음 [야]의 합성용 화일의 각 파라미터 설정값의 예

Synthesis specification for file: 'ya.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	620	1300
B1	v	30	70	1000	DF1	v	0	0	100
DB1	v	0	0	400	F2	v	550	1200	3000
B2	v	40	90	1000	F3	v	1200	2950	4800
B3	v	60	90	1000	F4	v	2400	3700	4990
B4	v	100	130	1000	F5	v	3000	4300	4990
B5	v	100	150	1500	F6	v	3000	4990	4990
B6	v	100	200	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	f_0	AV	F_1	F_2	F_3
0	1300	26	263	2398	3136
5	1300	27	287	2380	3105
10	1300	30	305	2359	3073
15	1310	32	329	2326	3042
20	1320	35	359	2304	3010
25	1330	37	386	2283	2978
30	1340	38	413	2250	2957
(중간 부분 생략)					
370	987	28	739	1283	3003
375	977	25	728	1272	3007
380	966	23	728	1272	3010
385	956	21	728	1272	3014
390	945	19	739	1272	3017
395	934	16	739	1272	3021

이렇게 생성된 화일을 LC630으로 합성했으며 각 화일 하나 당 약 5초 정도 소요되었다. 주파수 대역은 F_1 - F_3 각각 100Hz 전후로 했으며 이들 합성음과 원음의 포먼트, 기본주파수, 진폭 등의 음향학적 특징을 비교함과 동시에 녹음한 피험자에게 합성된 음을 여러 번 들려주어 원음과 차이가 없을 때까지 되풀이해서 파라미터를 조정하여 열 개의 최종 합성용 화일을 만들었다.

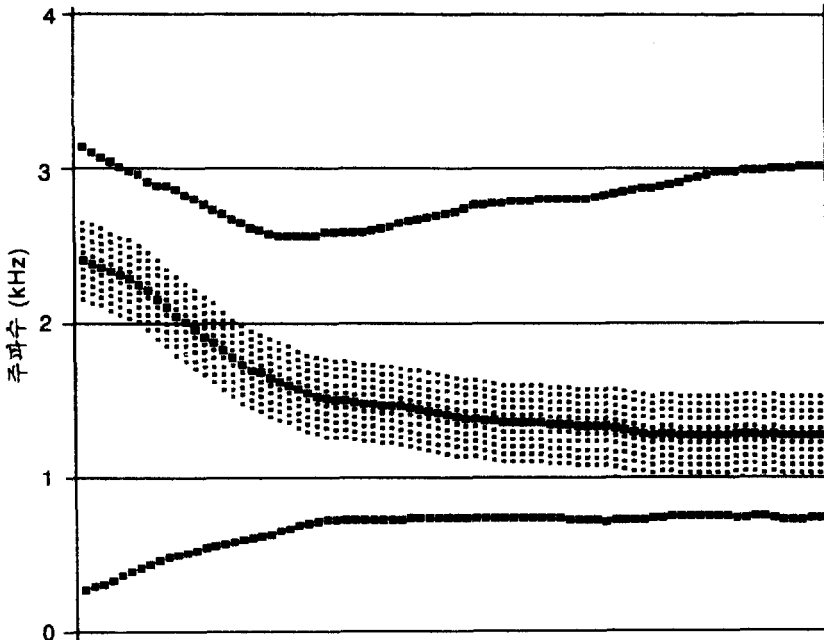
3. 지각실험 및 분석

이중모음의 원음과 거의 동일한 최종 합성용 화일인 아스키 화일을 각 연구 문항에 맞게 Hypertalk로 프로그램을 만들어 변형한 뒤 이를 SenSyn 1.0로 합성하여 청자에게 들려주었다. 지각실험 방법은 단모음 논문(양병곤 1995)과 같은 요령으로 합성한 음을 지각실험용 Hypertalk 프로그램에 하나씩 옮긴 뒤 원음과 동일하다고 판단된 합성음과 포먼트를 약간씩 변형한 합성음을 0.7초 간격으로 들려주고 피험자는 이들 두 음이 같게 들리는지 다르게 들리는지 응답지에 표시하는데 2.7초의 여유를 주고 다음 소리를 들려주는 방식으로 실험했다. 이들 응답지의 결과를 통합하여 평균을 구해서 실험문항에 대한 분석을 시도했다. 지각실험 1,2에서는 동의대학교 영문과 남아 학생 5명씩 모두 10명이 조용한 연구실에서 컴퓨터로 생성된 합성음을 들으면서 응답지에 표시했다. 청취실험은 약 40분정도 소요되었다.

지각실험 1

연구 문항 (1)에 대한 지각실험을 하기 위해 F_1 , F_2 는 각각 50Hz 간격으로, F_3 는 100Hz의 간격으로 최종 합성용 화일의 값을 중심으로 인접한 포먼트와 겹쳐지지 않을 때까지 일정한 간격으로 변형하여 음성을 합성했다. 사전 연구에서 그다지 다른 지각적 인상을 주지 않는 넓은 지각범위를 갖는 F_3 는 /여, 위, 의/만 변형하여 들려줬다. 그림 2의 큰 사각형은 실제 모음 [야]의 원음을 분석한 F_1 , F_2 , F_3 의 위치를 보여주고 있고, 작은 사각형은 F_2 의 아래위로 50Hz 간격으로 각각 5개씩 포먼트를 변형했을 때의 위치를 표시한 것이다. 결국, F_1 과 F_2 에서는 500Hz범위의 변화를 준 것이 되고, F_3 에서는 1000Hz의 변화폭을 갖게 된다. 이들 값은 단모음에 관한 사전연구에서 동일하게 판단한 지각범위(F_1 150Hz; F_2 300Hz; F_3 800Hz; 양병곤 1995)를 훨씬 넘는 값이기 때문에, 보다 더 넓은 범위의 변형은 하지 않았다.

그림 2. 이중모음 [야]의 합성용 세 포먼트와 F_2 를 아래위로 5개씩 변형한 모습



각 이중모음에 대한 남녀 10명의 지각실험의 결과를 평균한 값은 표 2와 같다. 각 포먼트 앞의 - 부호는 해당 포먼트를 50Hz씩 감소시켜서 합성하여 지각 실험한 자료에 대한 평균이고, + 부호는 해당 포먼트를 50Hz씩 증가시켜서 합성한 자료에 대한 평균이다. 평균 (1), (3)은 /j/ 그룹의 평균이며, 평균 (3), (4)는 /w/ 그룹의 평균이다. 이 표에서는 각 유형에 따른 차이가 있는지 알아보기 위해 /j/로 시작하는 그룹과 /w/로 시작하는 모음을 따로 구분하여 평균을 구해 보았다.

표 2. 일정한 간격으로 포먼트를 변화시켰을 때의 합성음에 대한 지각실험 결과 평균표

모음	-F ₁	+F ₁	-F ₂	+F ₂	-F ₃	+F ₃
야	40	120	180	220		
여	40	70	90	220	400	480
요	30	50	150	200		
유	80	60	220	200		
예	40	80	150	200		
평균(1)	46	76	158	208	400	480
와	10	90	130	130		
워	50	70	94	230	500	480
웨	60	100	210	220		
위	50	40	220	250		
평균(2)	43	75	164	208	500	480
의	70	110	210	160	360	360
남자평균	44	76	160	208	450	480

모음	-F ₁	+F ₁	-F ₂	+F ₂	-F ₃	+F ₃
야	50	70	140	170		
여	50	80	60	200	200	380
요	20	50	100	170		
유	80	50	130	200		
예	30	50	170	200		
평균(3)	46	60	120	188	200	380
와	30	90	130	120		
워	60	60	120	130	440	500
웨	30	70	180	150		
위	70	40	170	210		

평균(4)	48	65	150	153	440	500
의	70	90	180	210	400	340

여자평균	49	64	138	177	347	407
=====						
전체남여평균	48	72	152	190	383	423

위의 표에서 살펴볼 때, F₁에 대하여 동일하게 지각하는 변화범위는 전체남여자료를 모두 평균한 값에서 아래위로 증감시킨 값을 더했을 때 F₁에서는 120Hz이고, F₂의 변화폭은 342Hz, F₃의 변화폭은 806Hz이다. 이들 값은 단모음의 실험결과에서 나타난 동일하게 지각한 변화폭과 거의 일치하는 값이다. 이는 청자들이 단모음과 이중모음 거의 동일한 청각범위를 갖고 있는 것을 뒷받침하는 자료이다. 남여 각 그룹 내에서의 이중모음의 그룹별 차이는 별로 없는 것으로 나타났다. 특히, 이중모음 [의]의 경우에는 남여 각각 F₁과 F₂에서 보다 넓은 청각범위를 보이고 있음을 알 수 있다. 이는 이중모음 [의]의 발음상의 불안정함에 의해 발성 범위도 넓어진 것과 연관이 있을 것이다. 남여 자료에서 이중모음 [요]와 [와]의 경우는 아주 낮은 F₁때문에 매우 예민한 청각범위를 보이고 있다. 덧붙여, 저주파에서 고주파로 올라갈수록 동일하게 지각하는 범위는 높아져 가고 있다. 예를 들어, 전체 남여 평균값에서 볼 때 F₁을 50Hz씩 감소시켰을 때의 지각범위는 평균 48Hz인데 반해, F₁을 50Hz씩 증가시켰을 때의 지각범위는 평균 72Hz이고, F₂와 F₃에서도 점점 더 증가하고 있는데 이는 바로 사람의 청각 영역이 비선형적으로 증가함을 밝힌 기존 연구자료를 뒷받침하는 것이다(Zwicker 1962).

지각실험 2

연구문항 2에 대한 해답을 얻기 위해 각 이중모음의 F₁과 F₂의 전이음 구간을 50Hz씩 아래위로 5 개씩 변형하여 모두 20 개(500Hz 범위)의 합성 화일을 만들었다. 이 때 총지속시간 400ms 중 전이부가 위치한 약 125 ms에 해당하는 25개의 포먼트값을 2Hz씩 선형적으로 증감시켜서 변화하도록 Hypertalk로 프로그램화했다. 그림 3의 큰 사각형은 이중모음 /야/의 F₁을 나타내고 작은 사각형은 아래위로 변형된 값을 보여준다.

표 4는 이렇게 변형하여 합성한 화일을 10 명의 남여 피험자에게 들려주고 원음에 가까운 합성음과 비교시켜 동일하다고 판단한 결과의 평균값을 보여준다. 표기방식은 지각실험 1과 동일하다.

그림 3. 모음 /야/의 F₁ 변환 그림

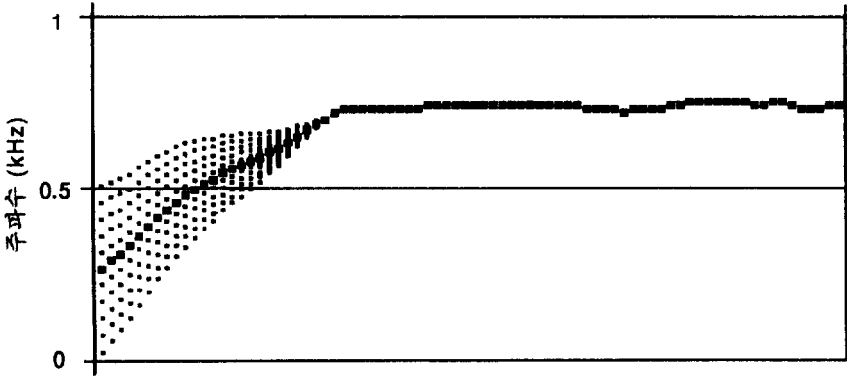


표 4. 포먼트의 일부를 변형하여 들려주었을 때의 지각실험 결과

모음	-F ₁	+F ₁	-F ₂	+F ₂	-F ₃	+F ₃
야	250	290	350	210		
여	150	240	550	240	480	460
요	140	120	460	200		
유	90	200	420	230		
예	140	80	250	150		
평균(1)	174	186	406	206	480	460
와	190	350	230	300		
워	250	210	250	370	500	440
웨	200	220	230	570		
위	120	140	290	180		
평균(2)	190	230	250	355	500	440
의	170	140	260	400	480	500
남자평균	181	206	337	272	490	450

모음	-F ₁	+F ₁	-F ₂	+F ₂	-F ₃	+F ₃
야	230	250	390	240		
여	40	190	520	220	440	360
요	110	180	550	240		
유	40	210	360	230		
예	140	120	240	140		

평균(3)	152	190	412	214	440	360
와	200	290	240	330		
위	230	150	250	380	480	460
웨	170	230	230	520		
위	90	160	400	170		
평균(4)	173	208	280	350	480	460
의	110	150	220	430	440	450

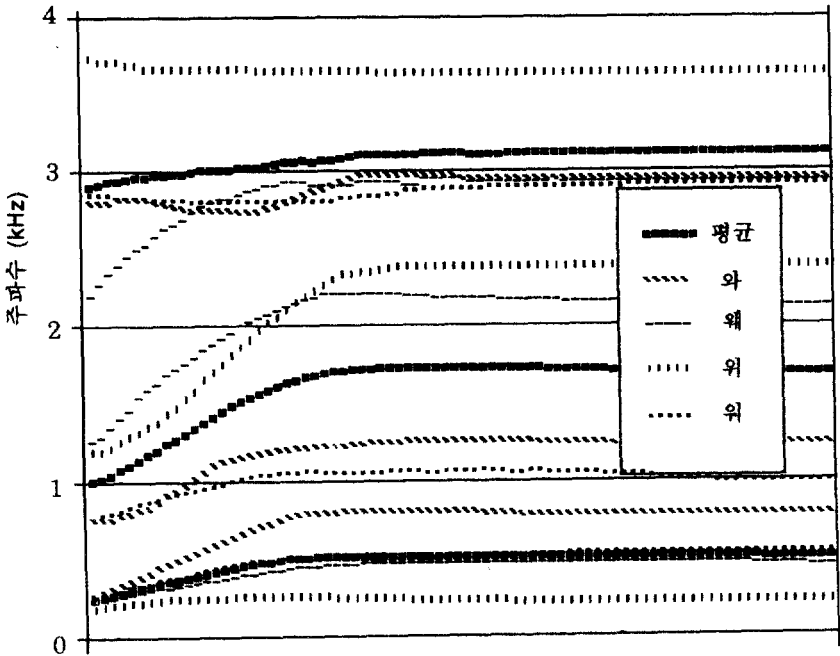
여자평균	161	198	353	274	460	410
=====						
전체남여평균	168	196	335	288	470	445

위 표에서 F₁에 대해 동일하게 지각하는 전체남여평균은 364Hz이고 F₂에서는 623Hz, F₃은 915Hz의 범위를 갖는다. F₁, F₂는 일정간격으로 변화시켰을 때 동일하게 지각하는 범위의 약 두 배에 해당하는 값이다. 이런 결과는 그림 3에서 보듯이 포먼트의 전이부만 선형적으로 변화시켜 포먼트 전이의 모양이 일부 남아있고 또 중핵모음 부분의 포먼트는 고정되어 있으므로 지각실험상 동일하게 인지하는 범위가 넓어졌다고 할 수 있다. 여기서도 남여 그룹의 차이는 거의 없다. 이중모음 그룹별로 분류한 값에서도 큰 차이는 없으나 한가지 주목할 것은 F₂에서 /j/ 그룹과 /w/ 그룹의 포먼트를 증가시켜서 합성한 값에 대한 지각범위와 포먼트를 감소시켜서 합성한 값에 대한 지각범위의 크기가 서로 반대로 되어있음을 볼 수 있다. 이는 각 그룹별 모음의 자체적인 높이에서 기인한다. 다시 말해서, /j/ 그룹은 F₂가 위에서 아래로 내려오는 모양이므로 포먼트 시작부 아래 저주파 부분에 대해 상당한 변화에도 동일하게 지각하는 범위가 넓다고 할 수 있으며, 그 반대로, /w/ 그룹은 아래에서 위로 올라가는 모양을 하고 있으므로 저주파보다는 고주파부분에 대한 보다 넓은 지각 공간이 있기 때문일 것이다. 그러나, 아래위 값을 합치면 거의 비슷한 지각범위를 보여주고 있으므로 여전히 F₂의 지각범위는 일치한다고 말할 수 있다. 아울러, 앞서 지각실험 1에서는 저주파에서 고주파로 갈수록 동일하게 지각하는 포먼트 주파수의 범위가 비선형적으로 증가하고 있었는데, 일부를 변화시킨 본 실험의 결과에서 F₂의 남여 총평균에서 이를 벗어난 듯이 보인다. 하지만, F₂의 평균값은 /j/ 그룹과 /w/ 그룹의 평균을 내었기 때문에 이러한 결과가 나왔다. 다시 말해서, /j/ 그룹에서는 F₂의 위치가 [이]에 가까운 고주파이므로 여기서는 변화폭이 넓고 /w/ 그룹은 모음 [우]에 가까운 저주파 포먼트를 가지므로 변화폭이 좁을 수밖에 없다. 따라서, 지각실험 1, 2에서는 여전히 비선형적인 포먼트 주파수의 범위를 보이고 있다고 말할 수 있다.

지각실험 3

연구문항 3에 대한 해답을 구하기 위해, 전이음 /j/로 시작하는 이중모음인 /야, 여, 요, 유/ 그룹과 전이음 /w/로 시작하는 /와, 위, 웨, 위/ 그룹을 한 그림에 표시하고, 또 이들 그룹의 평균값을 내어 보았다. 그림 4, 5는 각 그룹의 평균값과 이중모음의 각 포먼트를 나타낸 것이다.

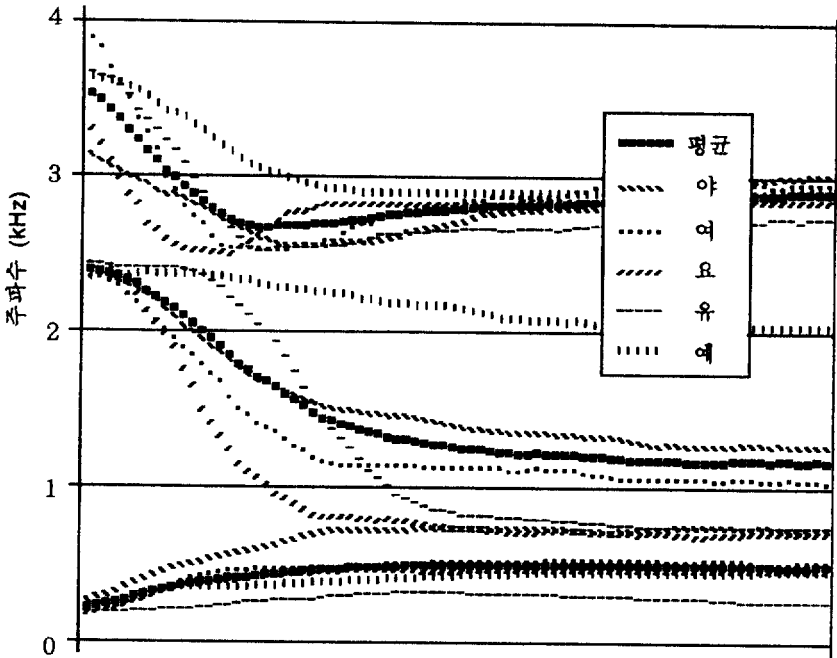
그림 4. /w/로 시작하는 이중모음의 그룹



/w/로 시작하는 그룹에서는 모음 /우/의 포먼트에 가까운 전이음 /w/로 시작되므로, F_1 은 아주 낮은 주파수에서 시작되었고, /위/를 제외하고는 이어지는 중핵부분의 모음의 포먼트로 상승하는 경향을 보인다. F_2 에서도 전이음 시작점은 동일하지만 이어지는 중핵부분의 모음으로 상승하고 있다. F_3 에서는 /웨/를 제외하고 평탄한 포먼트를 보이고 있다. 따라서, 이 그룹의 특징은 F_1 , F_2 모두 상승하는 것이다. 이들 값들을 모두 평균하여 합성했을 때 전이음 구간에서는 [우]에 가까운 음으로 중핵부분은 [어]로 되어 [위]소리로 들렸다. 그러나, F_1 에서는 평균값과 일치했지만, F_2 에서는 중핵

부가 고주파인 [이]와 [에], 저주파인 [아]와 [어]의 평균값을 취했기 때문에 어느 쪽에도 가깝지 않는 F_2 가 되어 선명치 못한 [워] 소리가 되었다. 일정한 간격으로 변화시켰을 때 [워]를 동일하게 지각하는 F_2 의 범위가 앞서의 실험에서 약 300Hz 전후인 것을 감안해 보면, 이런 범위를 훨씬 벗어난 F_2 값은 음질을 바꾸기에 충분하다.

그림 5. /j/로 시작하는 이중모음의 그룹



/j/로 시작하는 그룹에서는 세 개의 포먼트 중 두 개가 동시에 내려가는 특징을 갖고 있다. 또한, F_1 - F_2 에서는 전이음 시작부가 동일한 위치에서 시작되었고, F_3 에서는 약간의 차이를 보였는데, 앞서의 지각실험에서 F_3 의 지각 범위가 넓은 점을 고려해 볼 때 이는 거의 무시할 정도이다. 이는 모음 /이/의 특성이 전이음의 부분에 공통적으로 나타나기 때문이다. 실제 /j/ 그룹의 평균을 합성하여 들어본 결과 전이음 구간에서는 모음 [이]의 포먼트값에 가깝고 중핵부분에서는 모음 [어]에 가깝게 나타나서 결국 선명한 [여]로 들렸다. 스펙트로그램에서, F_1 과 F_3 는 그림 5에서 보듯이 일치했으며, F_2 에서도 전이음 시작부에서는 완전히 일치하고, 그 이후에는 [여]의 윤곽선을 평행하게 따라가고 있다.

4. 맺음말

본 연구에서는 건강한 화자가 발성한 10 개의 이중모음의 음향적 특징을 추출하여 원음과 최대한 같도록 합성하고, 이 합성용 파일의 파라미터 중 포먼트값을 일정한 간격으로 증감하거나 전이부분을 부분적으로 변환하여 원음과 유사한 합성음과 함께 남녀 대학생 10 명에게 들려주어 같은지 다른지 판단하게 했다. 연구 결과는 일정한 간격으로 변환한 자료에서는 단모음의 결과와 일치하는 값을 보였고, 전이부의 일부만 변환한 자료에서는 F_1 , F_2 에서 약 두 배의 값을 보였다. 지각범위의 예민도에 있어서는 주파수가 상승할수록 무디게 나타나서 청각 영역이 비선형적으로 증가하고 있음을 보였다. 각 이중모음을 전이음 별로 분류하여 평균값을 취하여 합성한 음에서는 전이부는 각 전이음에 가까운 음이 되고 중핵에서는 중모음 /어/에 가까운 구별할 수 있는 소리가 들렸다. 그러나 /w/로 시작하는 그룹에서는 F_2 에서 상당한 차이를 보여서 전체적인 음질은 [워]에서 떨어졌다. 앞으로 두 포먼트를 동시에 일정하게 변형했을 때의 지각적인 범위와 각 포먼트의 중간점을 취하여 합성했을 때의 지각 특성에 관한 연구를 진행할 계획이다.

본 연구의 결과는 합성한 한국어 이중모음에 대한 지각적인 실험의 중요한 기초자료를 제공하고, 또한 합성에 의한 지각실험 연구분야에 상세한 연구방법론을 제시함으로써 보다 많은 지각실험 연구를 유도하고, 음성인식에 대한 인간의 경계선에 대한 자료를 제시함으로써 포먼트 합성 방식에 의한 음성 합성기의 정밀도의 범위 설정과 음성 인식장치에서 동일하게 인식하는 허용범위를 선정할 때 활용될 수 있다.

참고문헌

- 김재민·이계순·한영희·전상범·신상순. 1985. 영어음성학. 서울: 신아사.
 양병곤. 1993. 한국어 이중모음의 음향학적 연구. 부산: 진영문화사.
 양병곤. 1995. "합성한 한국어 단모음의 지각실험 연구," 언어 20-3, 127-146.
 고도홍·구희산·김기호·양병곤 편역. 1995. 음성언어의 이해. 서울: 한신문화사.
 조성식. 1990. 영어학 사전. 서울: 신아사.
 전상범. 1988. 영어음성학. 서울: 을유문화사.
 Bond, Z. S. 1978. "The Effects of Varying Glide Durations on Diphthong Identification," *Language and Speech* 21-13, 253-263.
 Bond, Z. S. 1982. "Experiments with Synthetic Diphthongs," *Journal of Phonetics* 10, 259-264.

- Catford, J. C. 1977. *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.
- Gay, T. 1970. "A Perceptual Study of American English Diphthongs," *Language and Speech* 13-2, 65-88.
- Gerber, S. E. 1972. "Perception of Segmented Diphthongs," in A. Rigault and R. Charbonneau, eds., *Proceedings of the Seventh International Congress of Phonetic Sciences*, 480-492. The Hague: Mouton.
- Ladefoged, P. and I. Maddieson. 1990. "Vowels of the World's Languages," *Journal of Phonetics* 18, 93-122.
- Liberman, A. M., P. C. Delattre, L. J. Gerstman and F. S. Cooper. 1956. "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," *Journal of Experimental Psychology* 52, 127-137.
- O'Connor, J., L. J. Gerstman, A. M. Liberman, P. Delattre and F. S. Cooper. 1957. "Acoustic Cues for the Perception of Initial /w, j, r, l/ in English," *Word* 13, 24-43.
- Trager, G. L. and H. L. Smith. 1951. "An outline of English Structure," *Studies in Linguistics, Occasional Papers 13*. Reprinted by ACLS.
- Zwicker, E. 1962. "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *JASA* 33, 248.

부산시 부산진구 가야동 산 24
동의대학교 인문대학 영어영문학과
614-714
E-mail: bgyang@turtle.dongeeui.ac.kr
Fax: +82-51-890-1209

접수일자: 1996. 1. 1.
게재결정: 1996. 2. 2.