

개별화자 음성의 특징 파라미터 분석*

An Analysis of Phonetic Parameters for Individual Speakers

교도홍**

Do-Heung Ko

ABSTRACT

This paper investigates how individual speakers' speech can be distinguished using acoustic parameters such as amplitude, pitch, and formant frequencies. Word samples from fifteen male speakers in their 20's in three different regions were recorded in two different modes (i.e., casual and clear speech) in quiet settings, and were analyzed with a Praat macro scrip. In order to determine individual speakers' acoustical values, the total duration of voicing segments was measured in five different timepoints. Results showed that a high correlation coefficient between F_1 and F_2 in formant frequency was found among the speakers although there was little correlation coefficient between amplitude and pitch. Statistical grouping shows that individual speakers' voices were not reflected in regional dialects for both casual and clear speech. In addition, the difference of maximum and minimum in amplitude was about 10 dB which indicates a perceptually audible degree. These acoustic data can give some meaningful guidelines for implementing algorithms of speaker identification and speaker verification.

Keywords : Acoustic parameters, Praat, speaker identification, speaker verification

I. 서론

언어음은 인간의 화자의 입을 통하여 매우 다양하게 나타나므로 연구하기가 매우 까다롭다. 분절적인 속성을 갖고 있으면서도 동시에 비분절적인 속성을 함께 가지고 있는 것이 언어음인 것이다. 다시 말하면, 비교적 까다롭기는 하지만 언어음은 개별음으로 분절화(segmentation) 할 수는 있으나 주어진 환경에 따라 서로 상당한 영향을 주고받는다. 예를 들면, 동시조음(coarticulation) 현상은 하나의 개별음이 이웃하는 음뿐만 아니라 음절의 경계, 단어의 경계뿐만 아니라 구(phrase)의 경계를 넘어 이웃 음에 영향을 줄 수 있기 때문에, 단순한 평균값만으로 분석하여 응용하기 어렵다. 이러한 문제를 효과적으로 해결하기

* 본 연구는 한국과학재단 특정기초연구(1999-2-302-016-5) 지원으로 수행되었음.*

** 한림대학교 국어국문학과

위해서 음성인식이나 합성에서 분절단위(segmentation unit)를 여러 가지 방식으로 적용해 보고 있는 것이다.¹⁾

화자에 따른 변화는 화자 고유의 피치와 성도(vocal tract)의 길이 또는 모양에 기인하기 때문에 화자의 개별음성은 음향음성학적 또는 조음음성학적 방식으로 연구될 수 있다. 예를 들면, 성도의 변화가 음성 신호의 포먼트(formant)에 반영된다고 볼 때 포먼트의 변화를 정규화함으로써 개별화자의 음성 특징을 객관화시킬 수 있을 것이다.

음향음성학적으로는 FFT(Fast Fourier Transform)에 의해서 분석되는 피치(pitch), 진폭(amplitude), 포먼트 주파수(formant frequency) 등과 같은 파라미터들을 분석함으로써 특징 추출을 시도할 수 있고, 조음음성학적으로는 성도의 특성(Vocal Tract Trait), 비음성(Nasality), 성문하압(Subglottal Air Pressure), 구강 vs. 비강압력(Oral-to-Nasal Pressure) 등을 분석함으로써 특징 추출이 가능할 수 있다.

화자들 사이에서 나타나는 개별음성의 차이를 음성학적으로 살펴보려는 노력은 주로 개별화자의 평균값에 의존하고 있다. 따라서 이러한 연구는 개별화자의 일반적인 경향성(tendency)을 알아보는 데 한정될 수밖에 없고 개별음성이 갖는 역동성을 밝히는 것이 미흡할 수밖에 없다. 때문에 여러 사람을 대상으로 많은 데이터를 분석한다하더라도 화자인식이나 화자 확인의 기초가 되는 화자의 특징 추출에 크게 도움이 되지 않는다.

주지하는 바와 같이 인식(recognition)이나 확인(verification)에 가장 널리 쓰이고 있는 통계적 모델링 방법인 HMM(Hidden Markov Model)은 분절음(sound segment)을 10 msec으로 끊어서 분석하고 있다. 물론, 분절된 작은 구간별로 음성특징 파라미터별로 나누어 통계적으로 처리하는 것이다(Ramachandru, et. al. 1997; Marcus, et. al. 1995). 이 연구는 HMM의 분석 기법에서와 같이 분석 대상이 되는 음을 몇 개의 구간으로 나누어 화자들 사이의 개별음성이 과연 몇 가지 파라미터에 의해서 결정될 수 있는가를 알아보기 위한 새로운 시도이다. 즉, 이 연구에서는 기존의 연구방식과 다르게 하나 하나의 분절음 또는 어절에 대하여 음향 파라미터를 구간별로 나누어 분석하고자 하였다.

II. 실험 방법

2.1 실험 단어

실험 단어는 연세말뭉치 연구(서상규 1998)에서 조사된 4,200만 어절 가운데 순위 200 이내의 단어를 선정하였다. 일차적으로 선정된 100 단어를 각 음운 환경별로 나누어 음운 환경이 동일한 단어는 제외하여 50단어를 선정하였으며, 이 중에서 각 음운 환경을 대표할 수 있는 18개 단어를 1차 분석 대상으로 삼았다(표 1 참조). 음운 환경은 2음절어에서 두 번째 음절이 자음으로 끝나는 경우(CV+CV형, CV+CVC형, CVC+CVC형), 2음절어

1) 국어에서 인식 및 합성에서 주로 이용되었거나 현재 이용되고 있는 분절단위로는 50개 내외의 음소유사음소 단위(Phonemoid = Phoneme-like Unit), 음절의 안정된 구간(즉, 모음구간)을 반으로 나누어 이용하는 반음절 단위(Demisyllable Unit), 한 음절의 안정된 구간과 다음 음절의 안정된 구간을 이용하는 2음절단위(Dissyllable Unit), 목표음소의 앞 뒤음소를 함께 분석하는 3음소 단위(Triphone Unit), 4음소 단위(VCCV) 등이 있다.

에서 첫 번째 음절이 자음으로 끝나는 경우(CVC+CV형), 그리고 3음절어의 경우(CV+CV+CV형, CVC+CVC+CVC형)로 나누어 가능한 여러 가지 환경을 고려하여 단어를 선정하였다.

표 1. 단어 목록

	유성음	/ㄱ/받침	/ㅇ/받침	/ㅇ/받침	/ㄹ/받침	/ㄴ/받침	/ㅂ/받침
2음절어	가지	교육	바람	모양	사실	당신	모습
	여자	한국	지금	세상	서울	일본	방법
2음절어	/ㄴ/받침		/ㄴ/받침		3음절어	유성음	받침
	문제		친구			이야기	선생님

종합적인 연구에서는 18개 단어를 15명의 화자에 의해서 3회 반복한 내용을 다루고 있으나, 이 논문에서는 표 1의 단어목록 가운데 구간별 분석 단위에 일관성을 보여주는 4개의 단어(교육, 모양, 서울, 당신) 가운데 특정한 분절음으로 제한하여 분석의 대상으로 삼았다. (표 2 참조) 이는 화자에 따라 분절의 양상이 다르게 나타나서 통계처리가 곤란한 경우가 생기기 때문이다.

표 2. 단어와 분석대상

단어	분석대상	분석대상	
		CASUAL SPEECH	CLEAR SPEECH
교육		ㅍ[jo]+ㅍ[ju]	ㅍ[ju]
모양		ㄴ[o]+ㅍ[ja]+ㅇ[ŋ]	ㅍ[ja]+ㅇ[ŋ]
서울		ㄱ[ə]+ㄷ[u]+ㄹ[l]	ㄱ[ə]
당신		ㅣ[i]+ㄴ[n]	ㅣ[i]+ㄴ[n]

2.2 피험자 및 녹음 방법

실험에는 우선 서울·춘천 출신의 20대 남자 15명(지역별 각 5명)이 피험자로 참여하였다. 녹음은 강선미(1997)에 의해서 개발한 CSDB(ver. 1.5)를 이용하였으며, 피험자들로 하여금 고립된 환경에서 각각의 단어를 두 가지 방식으로 읽도록 하였다. 이 연구가 화자 확인을 목표로 한 만큼 화자들은 고립된 환경에서 자연스럽게 발음한 경우(casual speech)와 분명하게 발음한 경우(clear speech)로 나누어 발음하였다. 일정한 발화속도를 유지하기 위해서 피험자들은 자신들의 음성을 녹음하기 전에 Wave 파형으로 입력된 음성으로 어떤 것이 CASUAL SPEECH이고 어떤 것이 CLEAR SPEECH인가를 인지하도록 2-3차례 지각훈련을 하였으나 CASUAL SPEECH인 경우는 본인이 가지고 있는 본래의 속도를 자연스럽게 유지하도록 주문하였다. CSDB 프로그램은 피험자가 발음한 파형이 입력되면 자동으로 시작점과 끝점을 끊어 파일로 저장해주며, 잘못된 발음은 언제든지 재녹음할 수 있도록 설계되었다. 이 연구에서는 4개 단어만을 목표로 하였지만, 표 2에서 보듯

이 총 분석대상 음절수는 585개(13어절×3회 반복×15명)이었지만 이것을 다시 최종분석 대상이 되는 분절음으로 잡고 하나의 분절음을 다시 5구간으로 나누면 데이터의 양이 상당히 많아지게 된다.

2.3 분석 방법

2.3.1 시작점과 끝점 추출

실험에서 관심을 둔 음향적인 파라미터는 각 음절의 지속시간(duration), 유성구간의 포먼트 주파수(F_1 , F_2)와 진폭(Amplitude), 피치(Pitch), 폐쇄자음의 VOT 값이었는데 이 논문에서는 유성 구간의 포먼트 주파수, 진폭 및 피치에 대하여 논의하기로 한다. 양병곤(2000)은 녹음된 데이터를 Praat 프로그램을 이용하여 음향 파라미터의 값을 다섯 구간으로 나누어 자동으로 추출하도록 설계하였는데, 이 연구 결과는 이를 바탕으로 한 것이다. Praat를 이용한 분석에서는 진폭 값과 피치 값은 음절경계에 관계없이 시작점과 끝점에서 각 10 msec씩 안쪽으로 들어가 각 파라미터의 값을 자동으로 구하였다.

2.3.2 자동측정의 문제

자연스럽게 발음(casual)한 음성 샘플에서 두 번째 음절이 모음으로 시작되거나 유성음화가 되어 2음절어가 한 덩어리로 묶여지는 경우에도 음절수에 관계없이 다섯 구간으로 나누어 분석된다. 그림 1에서 보는 바와 같이, 단어 [서울]에서 [ɪ+ɪ+ɾ]이 유성음으로 연결되어 있기 때문에 하나의 단위로 끊어지는 경우가 있다. 이와 같이 자동측정을 하는 경우에는 Praat macro Scrip는 음소의 경계를 무시하고 무조건 5개 구간으로 분할된 값을 준다. 이때에 음소의 경계뿐만 아니라 음절의 경계도 없어지게 되어 음성학적 해석에는 문제가 생길 수 있으나 단어를 대상으로 한 화자의 확인(verification)에는 아무런 문제가 없게 된다(고도홍, 2000).

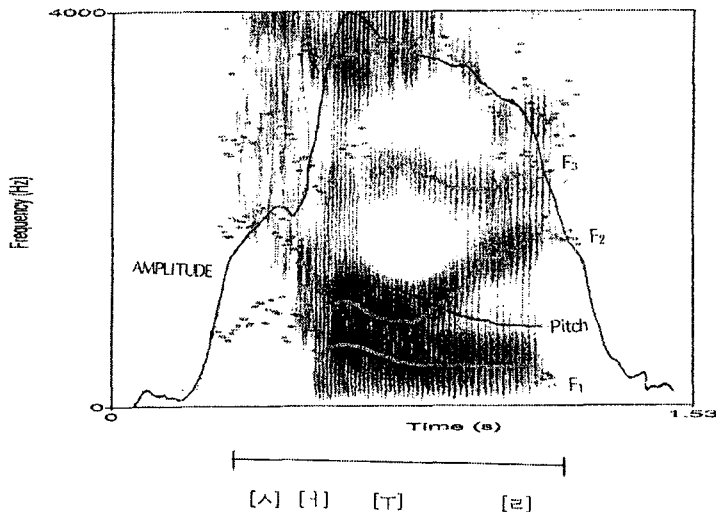


그림 1. 음절의 경계가 모호한 경우: 서울 화자가 CASUAL SPEECH로 발음한 단어 [서울]

2.3.3 발화속도(Speaking Rate)의 조절

음향 파라미터를 비교하기 위해서는 우선 화자들의 발음속도가 매우 중요하기 때문에 녹음하기 전에 화자들로 하여금 두 가지 경우에 대한 Reference를 제시하는 의미에서 sample recording을 들려주었으나, 가능한 한 화자 개인이 갖고 있는 지속시간의 특성이 그대로 유지될 수 있도록 하였다. 3개 지역 화자들 사이에 지속시간의 편차는 크지 않았으며, 4개 단어를 대상으로 실험에서는 CASUAL SPEECH의 경우 평균 580 msec(SD: 33)이었으며, CLEAR SPEECH에서는 평균 1,010 msec(SD: 52)이었다.

III. 결과 및 토의

3개 지역의 화자들²⁾을 대상으로 이들이 진폭(amplitude), 피치(pitch), 포먼트 주파수(formant frequency)를 측정하고 Duncan Grouping을 이용하여 각 화자들의 음성 특징을 비교하고자 하였다. 같은 단어를 3번 반복하여 측정대상 구간(유성음)을 다섯 구획으로 나누어 평균값(mean)을 측정하였다. 각 파라미터별로 결과를 알아보고 그 결과가 의미하는 것이 무엇인가 알아보고자 한다.

3.1 CASUAL SPEECH

표 3에서 보여주는 바와 같이, 단어 ‘교육’의 [교+育]에 대하여 진폭 값과 피치값은 화자 간 또는 화자의 출신 지역에 관계없이 편차가 비교적 안정적으로 나타나고 있다. 화자 간의 평균 진폭값의 차(max-min)는 약 10 dB 정도로 청각적으로 느낄 수 있는 차이를 보여주고 있다. 이에 비하여 화자 간의 평균 피치값은 1명의 화자(12번 화자)를 제외하면 40 Hz 이내의 차이를 보여주고 있다.

개별화자의 특성을 알아보기 위한 통계적 Grouping에 의하면, 진폭에 있어서 15명의 남성 화자들 가운데 10명의 화자(10,8,4,5,15,13,3,11,9,1)가 Group A에 속하고, 7명의 화자(15,13,3,11,9,1,14)가 Group B에 속하며, 6명의 화자(3,11,9,1,14,7)가 Group C에 속하고, 또 다른 6명의 화자(1,14,7,6,2,12)가 Group D에 속한다. 여기서 Group에 속한다는 것은 통계적으로 동질성이 강한 집단으로 간주할 수 있음을 의미한다.

피치에 있어서 1명의 화자(12번)가 여성의 피치에 근접하는 173 Hz를 보여주고 있기 때문에 혼자서 Group A 매우 독특한 특성(173 Hz)을 보여주고 있으며, 5명의 화자(6,10,15,13,7)가 같은 Group B에 속하고, 5명의 화자(10,15,13,7,8)가 같은 Group C에 속하며, 3명의 화자(8,5,11)가 같은 Group D에 속하고, 7명의 화자(5,11,1,14,3,9,2,4)가 같은 Group E에 속하며, 끝으로 6명의 화자(1,14,3,9,2,4)가 같은 Group F에 속한다.

이러한 통계적 Grouping을 통하여 화자별 특성을 생각해 볼 때, 3개 지역의 공통적인 특성이 특별히 반영되지 않음을 의미한다.

2) 이 실험에 참여한 15명의 피험자들은 대전(5명), 서울(5명), 그리고 춘천(5명) 지역 출신이다. 이들은 각각 1-5번은 대전화자, 6-10번은 서울화자, 11-15번은 춘천화자를 가리킨다.

표 3. [교육]에서 [ㄹ+ㅍ]에 대한 진폭과 피치의 관계 (CASUAL)

순번	진폭 (dB)			피치 (Hz)		
	Mean	화자	Duncan Grouping	Mean	화자	Duncan Grouping
1	78.4	10	A	173.4	12	A
2	78.4	8	A	130.7	6	B
3	77.2	4	A	126.1	10	C B
4	77.0	5	A	123.1	15	C B
5	74.9	15	B A	121.3	13	C B
6	74.9	13	B A	118.7	7	C B
7	74.5	3	B A C	115.7	8	C D
8	73.1	11	B A C	106.7	5	E D
9	72.0	9	B A C	106.6	11	E D
10	70.8	1	B D A C	101.7	1	E F
11	70.3	14	B D C	100.9	14	E F
12	69.9	7	D C	97.7	3	E F
13	68.7	6	D	97.2	9	E F
14	68.3	2	D	95.0	2	E F
15	67.6	12	D	90.3	4	F

그림 2는 진폭과 피치의 상관관계를 잘 보여주고 있다. 단어 '교육'의 [ㄹ+ㅍ]에서 진폭과 피치는 큰 관계가 없는 것으로 나타나고 있다. 즉, 진폭이 크면 피치가 낮거나, 진폭이 작으면 피치가 높다고 할 수 없다. 물론, 진폭이 크면 피치가 높거나, 진폭이 작으면 피치가 낮게 나타나지도 않는다.

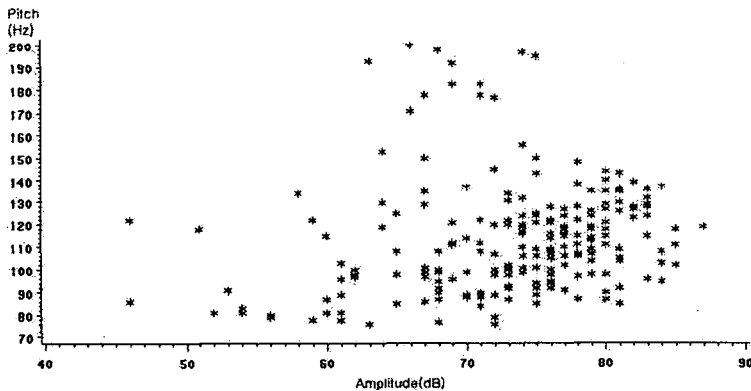


그림 2. 교육[ㄹ+ㅍ]의 진폭과 피치의 상관관계 (CASUAL, p<.007)

표 4는 단어 '교육'의 [ɟ+ɳ]에 대하여 F₁ 평균값과 F₂ 평균값이 역시 화자의 출신 지역과 관계가 없이 나타나고 있음을 보여주고 있다.³⁾

표 4. [교육]에서 [ɟ+ɳ]에 대한 F₁과 F₂의 관계 (CASUAL)

순번	F ₁ (Hz)			F ₂ (Hz)		
	Mean	화자	Duncan Grouping	Mean	화자	Duncan Grouping
1	453.1	14	A	1988.5	3	A
2	437.8	1	B A	1813.9	2	B A
3	417.8	10	B D A	1659.4	10	B
4	387.6	3	B D A	1638.2	8	B C
5	382.5	12	E B D A	1583.2	4	B C D
6	370.7	13	E B D A	1532.4	6	C D
7	369.0	7	E B D A	1518.6	14	F C D
8	350.8	6	E B D C	1441.4	9	E F C D
9	344.9	11	E B D C	1422.5	7	E F C D
10	342.1	5	E B D C	1414.7	1	E F G D
11	341.8	9	E B D C	1337.4	13	E F G D
12	329.9	8	E D C	1282.3	5	E F G
13	313.7	15	E D	1256.1	11	F G
14	302.7	2	E D	1251.9	12	F G
15	285.1	4	E	1138.0	15	G

화자 간의 평균 F₁의 차(max-min)는 161 Hz로서 비교적 높게 나타나고 있으며, 평균 F₂의 차는 대략 850 Hz로 큰 폭의 차이를 나타내고 있다.⁴⁾ 그리고 통계적 Grouping에 의하면, 포먼트 주파수(F₁, F₂)는 지역적인 공통성을 나타내고 있지 않다.

그림 3은 F₁과 F₂의 상관관계를 보여 주고 있다. F₁과 F₂는 p<.0002에서 상관력이 있음을 나타내고 있다. 화자 2명(10번 화자, 13번 화자)의 경우에는 화자 15명 가운데 F₁에 있어서나 F₂에 있어서 동일한 순위로 나타났지만, 일반적으로 F₁이 높아질수록 F₂가 낮아지고 F₁이 낮아질수록 F₂가 높아지는 음의 상관력을 갖고 있다고 할 수 있다.

3) 이 연구에서는 화자들을 대전 서울, 춘천 3개 지역으로 한정해서 지역적인 특성이 통계에 잘 반영이 되지 않고 있지만 다른 지역출신의 화자들을 비교해야만 지역별 특성에 대한 구체적인 결론을 낼 수 있다고 본다.

4) 통계적으로 F₁에서 약 20-40 Hz, F₂에서 약 80-100 Hz, 그리고 F₃에서는 약 120-150 Hz는 오차의 범위로 간주할 수 있는 수치이다.

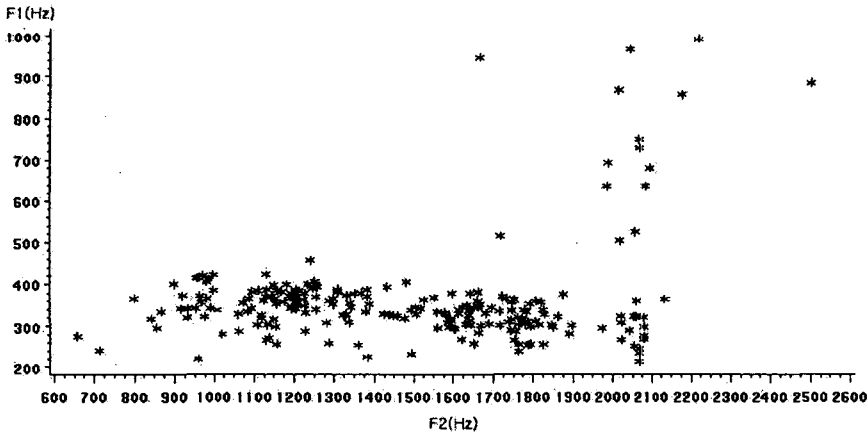


그림 3. '교육(교+ㅠ)': F₁과 F₂의 상관관계
(CASUAL, p<.0002)

3.2 CLEAR SPEECH

표 5에서도, 단어 '당신'의 [l+l]에 대하여 진폭 값과 피치값은 화자간 또는 화자의 출신지역에 관계없이 편차가 비교적 안정적으로 나타나고 있다. 화자간의 평균 진폭값의 차(max-min)는 약 9.8 dB 정도로 CASUAL SPEECH의 경우와 마찬가지로 청각적으로 느낄 수 있는 음의 차이이다. 이에 비하여 화자 간의 평균 피치값은 1명의 화자(12번 화자)를 제외하면 44 Hz 이내의 차이를 보여주고 있다.

개별화자의 특성을 알아보기 위한 통계적 Grouping에 의하면, 진폭에 있어서 15명의 남성 화자들 가운데 7명의 화자(5,15,1,13,4,14,8)가 Group A에 속하고, 11명의 화자(15,1,13,4,14,8,11,9,6,10,7)가 Group B에 속하며, 11명의 화자(1,13,4,14,8,11,9,6,10,7,3)는 Group C에 속하고, 10명의 화자(4,14,8,11,9,6,10,7,3,12)는 Group D에 속하며, 그리고 8명의 화자(8,11,9,6,10,7,3,12,2)는 Group E에 속하여 통계적으로 동질성이 강한 집단으로 분류할 수 있다.

CLEAR SPEECH의 피치에 있어서도 15명의 남성 화자들 가운데 1명(12번 화자)이 Group A로 피치가 예외적으로 높게(159.5 Hz) 나타나고, 3명의 화자(6,15,8)가 Group B로 나타나고, 5명의 화자(15,8,10,7,13)가 Group C로 나타나며, 5명의 화자(8,10,7,13,15)가 Group D로 나타나고, 7명의 화자(7,13,5,11,1,2,4)가 Group E로 나타난다.

CASUAL SPEECH의 경우에서와 같이, 이와 같은 통계적 Grouping은 지역적인 특성을 제대로 반영하고 있지 않다는 것을 의미한다.

표 5. [당신]에서 [ㅣ+ㄴ]에 대한 진폭과 피치의 관계 (CLEAR)

순번	진폭 (dB)			피치 (Hz)		
	Mean	화자	Duncan Grouping	Mean	화자	Duncan Grouping
1	72.0	5	A	159.5	12	A
2	69.0	15	B A	134.4	6	B
3	67.9	1	B A C	121.8	15	C B
4	67.8	13	B A C	118.5	8	C B D
5	67.3	4	B D A C	115.5	10	C E D
6	67.2	14	B D A C	113.7	7	C F E D
7	66.6	8	E B D A C	106.6	13	G C F E D
8	66.2	11	E B D C	103.7	5	G F E D
9	66.0	9	E B D C	99.5	11	G F E
10	64.8	6	E B D C	98.9	1	G F E
11	63.6	10	E B D C	98.8	2	G F E
12	63.5	7	E B D C	96.5	4	G F
13	62.3	3	E D C	95.5	9	G
14	61.6	12	E D	94.2	14	G
15	61.2	2	E	90.4	3	G

그림 4는 CASUAL SPEECH에서와 같이 단어 '당신'의 [ㅣ+ㄴ]에서 진폭과 피치는 큰 관계가 없는 것으로 나타나고 있다. 즉, 진폭이 크면 피치가 낮거나, 진폭이 작으면 피치가 높다고 할 수 없다. 물론, 진폭이 크면 피치가 높거나, 진폭이 작으면 피치가 낮게 나타나지도 않는다.

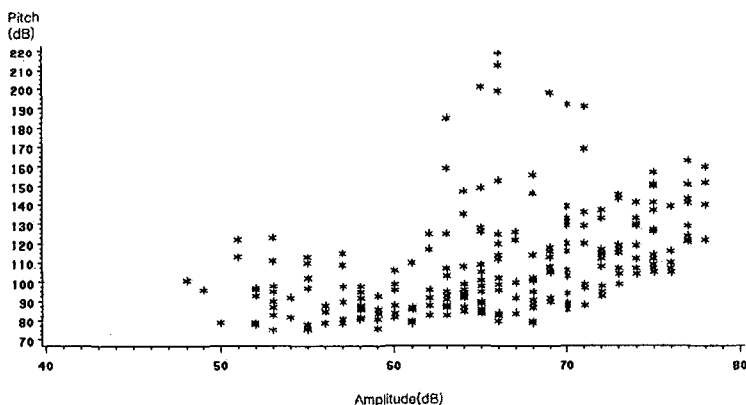


그림 4. 당신 [ㅣ+ㄴ]의 진폭과 피치의 상관관계 (CLEAR SPEECH, p<.0001)

표 6은 단어 '당신'의 [ㅣ+ㄴ]에 대하여 F₁ 평균값과 F₂ 평균값이 역시 화자의 출신 지역과 관계가 없이 나타나고 있음을 보여주고 있다. 통계적 Grouping에 의하면, CASUAL

SPEECH에서와 같이 F_1 과 F_2 에 있어서도 포먼트 주파수에 지역적인 공통성 나타나고 있지 않다.

표 6. [당신]에서 [l, ɫ]에 대한 F_1 과 F_2 의 관계 (CLEAR)

순번	F_1 (Hz)			F_2 (Hz)		
	Mean	화자	Duncan Grouping	Mean	화자	Duncan Grouping
1	388.5	13	A	1983.9	12	A
2	384.1	12	A	1885.0	2	B A
3	374.8	10	A	1760.5	6	B C
4	325.4	2	B	1754.9	9	B C
5	321.1	3	B	1748.3	5	B C
6	319.7	11	B	1747.7	8	B C
7	307.4	6	C B	1746.7	1	B C
8	302.0	8	C B D	1706.6	7	C
9	298.3	1	C E B D	1702.0	10	C
10	294.5	9	C E B D	1698.0	13	D C
11	286.0	5	C E B D	1668.4	11	D C
12	278.0	14	C E B D	1653.3	14	D C E
13	262.1	4	C E D	1564.3	4	D E
14	257.3	7	E D	1529.5	15	E
15	252.7	15	E	1529.1	3	E

그림 5는 F_1 과 F_2 의 상관관계를 보여 주고 있다. F_1 과 F_2 는 $p < .0001$ 에서 상관력이 있음을 나타내고 있다. 화자 2명(12번 화자, 14번 화자)의 경우에는 화자 15명 가운데 F_1 에 있어서나 F_2 에 있어서 동일한 순위로 나타났지만, 일반적으로 F_1 이 높아질수록 F_2 가 높아지고 F_1 이 낮아질수록 F_2 도 낮아지는 양의 상관력을 갖고 있다고 할 수 있다.

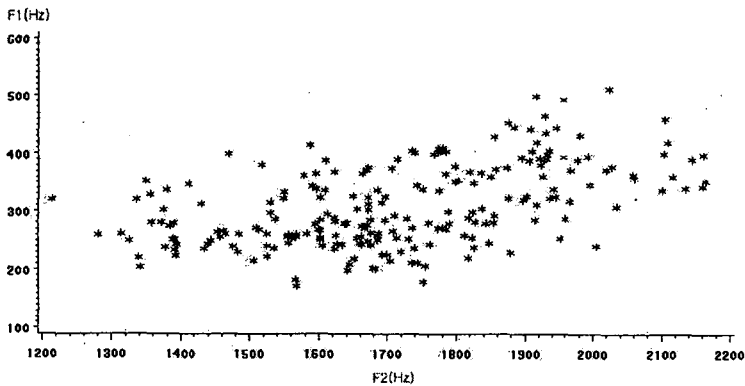


그림 5. '당신(l, ɫ)'의 F_1 와 F_2 의 상관관계 (CASUAL SPEECH, $p < .0001$)

3.3 4개 단어의 파라미터별 상관관계

CASUAL의 경우에, 교육, 모양, 서울, 당신 4개 단어는 진폭 vs. 피치는 $p < .0002$ 에서 매우 약한 상관관계를 나타내고 있다. (그림 6 참조) 그러나 CLEAR SPEECH의 경우에는 진폭 vs. 피치 거의 직선상으로 나타나 서로 상관관계가 없음을 보여주고 있다.

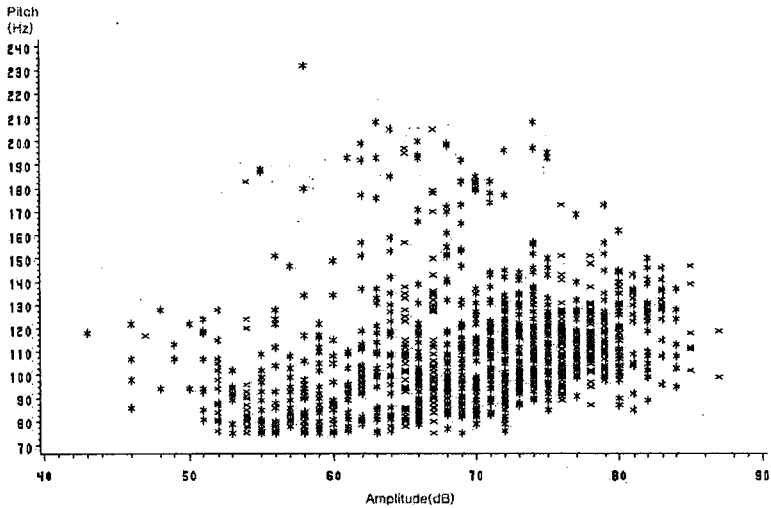


그림 6. 진폭과 피치의 상관관계(4개 단어: 교육, 모양, 서울, 당신)
(CASUAL SPEECH, $p < .0002$)

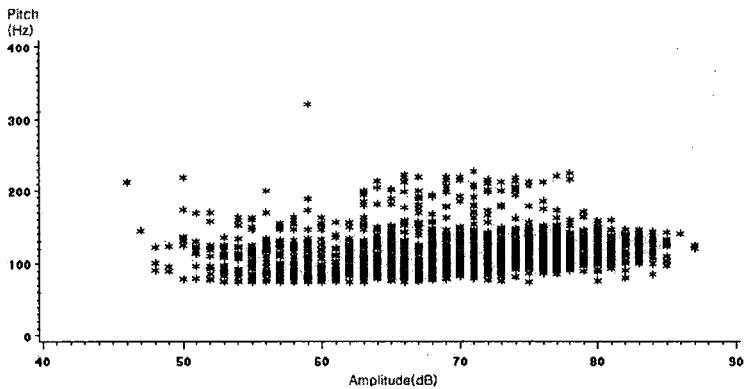


그림 7. 진폭과 피치의 상관관계(4개 단어: 교육, 모양, 서울, 당신)
(CLEAR SPEECH, $p < .0001$)

진폭과 피치는 CASUAL이나 CLEAR SPEECH의 경우에 모두 상관관계가 거의 또는 전혀 없는 것으로 나타나지만, CASUAL SPEECH에서 포먼트 주파수 F_1 vs. F_2 에서는 $p < .6095$ 에서 상당히 상관관계가 있음을 나타내 주고 있으며(그림 8 참조), CLEAR SPEECH에서도 F_1 과 F_2 는 $p < .0001$ 에서 상관관계가 있다.(그림 9 참조)

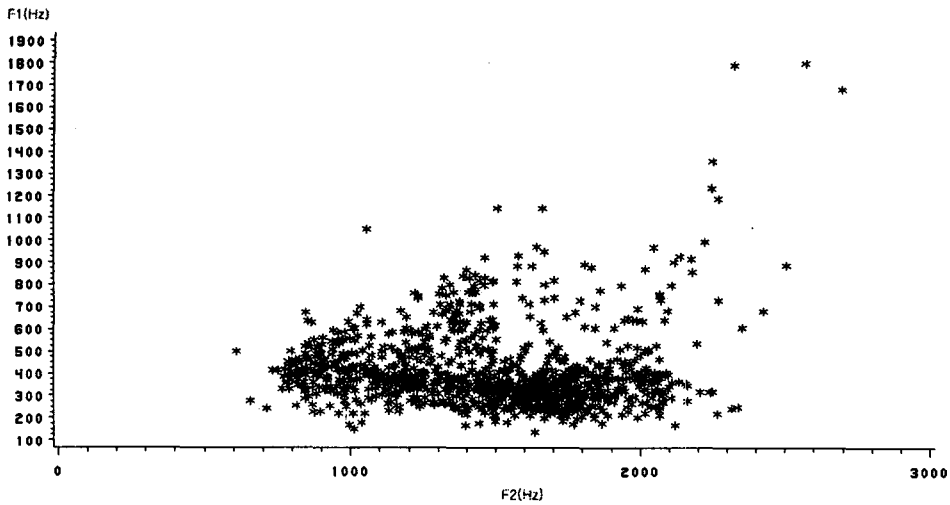


그림 8. F₁와 F₂의 상관관계(4개 단어: 교육, 모양, 서울, 당신)
(CASUAL SPEECH, $p < .6095$)

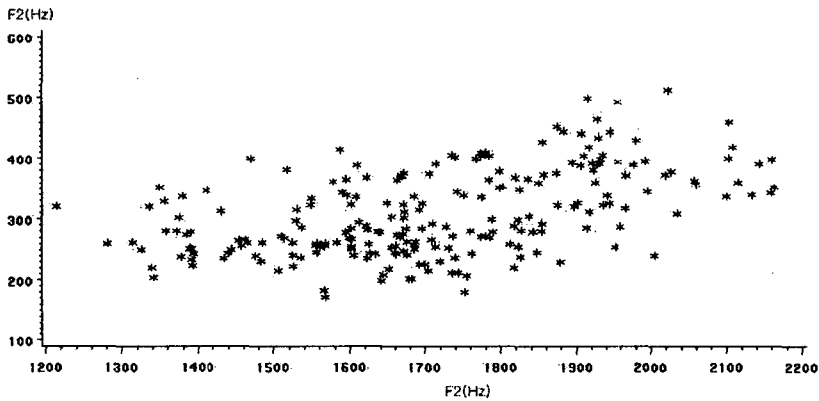


그림 9. F₁과 F₂의 상관관계(4개 단어: 교육, 모양, 서울, 당신)
(CLEAR SPEECH, $p < .0001$)

VI. 결 어

이 연구에서는 진폭, 피치, 그리고 포먼트 주파수가 화자간에 어떠한 양상을 보이는가를 알아보고, 음향 파라미터 상호간 어떠한 상관관계를 갖고 있는지를 알아보고자 하였다. 기존의 연구에서는 화자 개인의 특성을 알아보기 위하여 유성구간 전체를 5개로 나누어 일일이 그 패턴을 분석한 연구는 고도홍(2000)과 양병곤(2000)에 의해서 처음으로 시도된 것이고 이러한 시도는 공학적인 필요에 의해서 구상된 것이다. 이 연구에서 얻어진 결론을 요약하면 다음과 같다.

15명의 남성화자에 대한 연구에서 개별화자가 진폭이나 피치의 범위를 알아본 결과 진

폭의 차는 대략 10 dB 정도이고, 피치의 차는 대략 40 Hz이었다. 통계적 Grouping에 의하면, 3개 지역(대전, 서울, 춘천)의 화자들이 공통적인 특징이 없는 것으로 나타났다. 진폭과 피치와의 상관관계는 거의 없으나 화자별로 F1 vs. F2는 상관력을 보여준다는 사실을 알 수 있었다. 우리는 여기서 몇 가지의 음향 파라미터를 간단히 측정함으로써 개인의 음성 특성을 쉽게 밝혀낼 수 있다고 볼 수는 없겠지만, 개별음성은 특징 추출은 여러 가지 방법으로 모색되어야 할 것으로 본다. 필요가 있다는 것이다. 음성의 특징추출을 위해서는 개별 파라미터의 특정한 값이 개인의 음질에 얼마만큼의 영향을 미치게 하는가에 대한 연구가 뒤따라야 하고, 이를 증명하기 위해서는 지각실험(perceptual experiments)과 합성(synthesis) 등의 연구가 필요하다고 본다.

참 고 문 헌

- 고도홍. 2000. 개별화자의 음성 파라미터 분석. 한국음성과학회 제8회 학술발표회(연세대) 논문집. 75-82.
- 강선미. 1997. 음성신호처리의 이해. 서경대학교 출판부.
- 서상규. 1998. "말뭉치 분석에 기반을 둔 낱말 빈도의 조사와 그 응용." *한글* 242, 225-270.
- 양병곤. 2000. "Praat를 이용한 숫자음의 음향적 분석법. 한국음성과학회 제8회 학술발표회 (연세대) 논문집. 143-148.
- Ramachandrupa, N. V., Sitaram & Thippur Sreenivas. 1997. "Incorporating Phonetic Properties in Hidden Markov Models for Speech Recognition." *Journal of Acoustic Society of America* 102(2), 1149-1158.
- Marcus, S. M. & Ann K. Syrdal. 1995. "Speech: Articulatory, Linguistic, Acoustic, and Perceptual Descriptions." in *Applied Speech Technology* (eds. by Syrdal, A., Bennett, R. & Greenspan, S.), 1-45.

접수일자: 2000. 4. 21.

게재결정: 2000. 5. 27.

▲ 강원도 춘천시 옥천동 1 (우 200-702)

한림대학교 국어국문학과

Tel: (0361) 240-1205, Fax: (0361) 242-6763

H/P: 016-748-1205

e-mail: dhko@sun.hallym.ac.kr