

한국어 단모음의 성별, 연령별 특징변화 및 인식

(Changes in Features of Korean Vowels with Age and Sex of Speakers and Their Recognition)

李 勇 主* 金 敬 泰* 車 均 鉉**

(Yong Ju Lee, Kyung Tae Kim and Kyun Hyon Tchah)

要 約

음소인식에서의 개인차 해결을 위한 기초자료로서 모음을 특징지워주는 포먼트 주파수와 피치주파수를 대상파라미터로 하여 우리말 단모음의 연령, 성별에 따른 이들 특징량의 변화를 다량의 음성데이터에 의해 분석하였고 또한 개인차 흡수에 대하여 고찰하였으며 각종 통계적 거리에 의한 인식실험을 수행하여 다음과 같은 결과를 얻었다.

- 1) 연령·성별에 따른 피치주파수 변화는 어린이의 경우 남녀간의 차는 인정하기 어렵고, 변성기 전후의 차는 여자가 약 0.2oct, 남자가 0.9oct 정도로 분석되었다.
- 2) 연령에 따른 포먼트 주파수 변화는 모음에 따라 정도의 차이는 있으나 연령이 높아짐에 따라 대체로 낮아지는 경향을 보였다.
- 3) 피치와 포먼트 주파수간에는 연령변화를 매개로 한 간접적인 상관은 있으나 직접적인 상관이 있다고 보기는 어렵다.
- 4) 피치와 포먼트주파수에 의한 인식실험에서 성별·연령별로 다양한 발성자를 대상으로 할 때는 피치도 유효한 인식파라미터로 작용한다.

Abstract

As the basic analysis to solve the within-and cross-speaker variability in phoneme based speech recognition, changes in pitch and formant frequencies of 8 Korean vowels with age and sex of speaker has been investigated by analyzing a large number of samples.

Conclusions obtained are as follows:

- 1) Changes in pitch frequency with age and sex of speaker for children are hard to distinguish and the difference of before and after the voice change is analyzed approximately 0.2 oct. for female an 0.9 oct. for male.
- 2) While most of the formants of vowel considerably change with the age of speaker, the change becomes smaller as the age becomes older.
- 3) While there is an indirect correlation between pitch and formant with change in age, it is hard to see a direct correlation.
- 4) When the objects of the recognition experiment by pitch and formants are various speakers in each age and sex, pitch also works as an efficient recognition parameter.

*正會員, 韓國電子通信研究所
(Electronics and Telecommunications Research Institute)

**正會員, 高麗大學校 電子電算機工學科
(Dept. of Elec. & Comp. Eng., Korea Univ.)
接受日字: 1988年 4月 29日

I. 서 론

기계가 음성을 자동으로 인식하는 음성인식 기술은 반도체 및 신호처리 기술의 발전에 힘입어 상당한 성과를 거두고 있어서 수백 단어의 인식이 가능한 시스템들이 각 분야에서 점차 응용이 확대되어

가고 있다.¹¹⁾

이러한 음성인식의 궁극의 목표는 자연스럽게 발성된 연속음성을 인식 또는 이해하는 것으로 이에 대한 각국의 연구 또한 활발하다.¹²⁾

발성자를 한정시키지 않고 연속음성을 인식하기 위해서는 인식의 단위를 단어 이하의 음소와 같은 미소 단위로 해야하는데 이러한 음소단위 인식을 위해서는 개인차의 문제와 음소간의 조음결합이 가장 큰 어려움으로 지적되고 있다. 그런데 인식의 대상을 단모음만으로 한정시켜 본다면 조음결합에 관계없이 개인차만을 고려할 수 있을 것이고 이러한 단모음의 인식은 음소단위의 음성인식에 기초가 된다.

본 연구에서는 발성자에 의존하지 않는 우리말 단모음의 인식을 위하여 각 단모음의 피치 및 포먼트 주파수가 연령 및 성별에 따라 어떻게 변화하며, 이들 모음을 서로 구별하려 할 때 각 특징량이 어느 정도 유효한지, 그리고 이를 이용한 개인차의 정규화 가능성을 실험적으로 검토하여 인식 알고리즘 개발에 필요한 자료를 얻고자 하였다.

이에 관한 외국의 연구예로는 Fant,¹³⁾ Kasuya,¹⁴⁾ Nakagawa,¹⁵⁾ Fujisaki¹⁶⁾ 등이 있으나 우리말을 대상으로 한 관련연구로는 피치 및 포먼트 주파수를 따로 측정한 예¹⁷⁾¹⁸⁾들과 이를 이용하여 식별을 시도한 예¹⁹⁾²⁰⁾²¹⁾²²⁾²³⁾²⁴⁾²⁵⁾²⁶⁾²⁷⁾²⁸⁾²⁹⁾³⁰⁾³¹⁾³²⁾³³⁾³⁴⁾³⁵⁾³⁶⁾³⁷⁾³⁸⁾³⁹⁾⁴⁰⁾⁴¹⁾⁴²⁾⁴³⁾⁴⁴⁾⁴⁵⁾⁴⁶⁾⁴⁷⁾⁴⁸⁾⁴⁹⁾⁵⁰⁾⁵¹⁾⁵²⁾⁵³⁾⁵⁴⁾⁵⁵⁾⁵⁶⁾⁵⁷⁾⁵⁸⁾⁵⁹⁾⁶⁰⁾⁶¹⁾⁶²⁾⁶³⁾⁶⁴⁾⁶⁵⁾⁶⁶⁾⁶⁷⁾⁶⁸⁾⁶⁹⁾⁷⁰⁾⁷¹⁾⁷²⁾⁷³⁾⁷⁴⁾⁷⁵⁾⁷⁶⁾⁷⁷⁾⁷⁸⁾⁷⁹⁾⁸⁰⁾⁸¹⁾⁸²⁾⁸³⁾⁸⁴⁾⁸⁵⁾⁸⁶⁾⁸⁷⁾⁸⁸⁾⁸⁹⁾⁹⁰⁾⁹¹⁾⁹²⁾⁹³⁾⁹⁴⁾⁹⁵⁾⁹⁶⁾⁹⁷⁾⁹⁸⁾⁹⁹⁾¹⁰⁰⁾가 있으나 대부분 성인남자 또는 여자만을 대상으로 하고 있다.

따라서 본 연구에서는 성장기의 남녀 각 연령층을 주대상으로 하여 다량의 음성표본을 이용, 통계적 분석방법에 의해 성별·연령에 의한 각 특징량의 변화 경향, 발성자의 정규화 및 인식실험을 수행하였다.

II. 음성자료 및 특징추출

1. 음성자료

본 연구에서 사용하는 음성자료의 발성자는 녹음 당시 대전에 거주하는 국민학교·중학교, 고등학교에 재학중인 7~18세 사이의 남녀학생 각 5명(계 120명)과 20대 이후의 성인남자 18명, 여자 6명이며 특히 12세 이전의 발성자는 변성기 전, 13세 이후는 변성기 후기 또는 완전히 끝난 자를 대상으로 하였다.

대상어휘는 /아/, /어/, /오/, /우/, /으/, /이/, /에/, /애/의 8모음으로 하였다.

일반적으로 발음이 익숙하지 않은 단모음을 평소의 음높이로 발성시켰을 경우, 발성자 고유의 피치 및 포먼트 주파수가 안정되게 연결될 수 있을지 의문의 여지가 있으나 마이크로폰을 향한 발성에 습관

이 붙은 3회째 부터는 비교적 안정된 결과를 얻었다는 문헌(4)의 연구예에 따라 여기서도 랜덤한 순서로 배열한 8모음을 미리 수회 발성연습을 시킨후 최종적으로 5회 발성한 것을 녹음하여 이중 3회째 또는 4회째의 발성을 사용하였다.

이 음성을 4.5KHz의 저역통과 필터를 거쳐 10KHz로 샘플링한 12bit의 디지털 데이터를 컴퓨터에 수록하였고 수록된 각 모음의 정상부분 1frame만을 실험 대상으로 하였다.

2. 특징추출

1) 피치 주파수

피치의 추출법에는 여러가지가 있으나 여기서는 LPC분석의 잔차신호의 자기상관 함수에 의한 방법¹⁵⁾과 의외의 추출 에러를 배제하기 위하여 스펙트로그램에서의 단위시간당의 상대펄스 수를 세는 방법¹⁶⁾을 함께 사용하였다.

먼저, 각 성성자의 8모음중 1모음만 스펙트로그램에 의해 피치를 측정해 두고 8모음 전체의 중심프레임의 피치를 잔차상관법으로 측정한다. 이때 프레임에 따라 상대펄스 수에 비해 전혀 의외의 측정결과가 나오는 경우는 인접의 다른 프레임을 측정한다. 이렇게 해서 측정된 8모음의 평균을 그 발성자의 피치로 삼았다.

2) 포먼트 주파수의 추출

음성파의 주파수 스펙트럼 상에는 성도의 전달함수의 극과 영에 의해 여러개의 공진의 산과 반공진의 골짜기가 생긴다. 이들 공진과 반공진을 포먼트 및 엔티포먼트라 하며 각종 음운은 주로 포먼트, 엔티포먼트의 주파수로 규정되고 특히 모음은 낮은쪽에서 2~3개의 포먼트주파수로 특징 지워진다(이하, F1, F2, F3로 약칭한다).

포먼트주파수의 추출법에는 스펙트로그램에 의한 관찰법등 여러가지가 있는데 여기서는 컴퓨터를 이용하는 방법중 가장 정밀하다고 알려진 AbS(analysis by synthesis)법¹⁷⁾¹⁸⁾을 사용하였다.

이 AbS법은 음성신호로부터 분석한 스펙트럼과 음성생성 모델에 의해 합성한 스펙트럼을 비교하여 그 오차정보에 의해 합성을 위한 포먼트주파수 등의 파라미터를 갱신하는 과정을 반복하여 점근적으로 포먼트주파수를 구하는 방법이다. 따라서 이 AbS법은 계산시간은 소요되지만 음성생성 모델을 이용하므로 가장 정도 높은 포먼트 주파수의 추정이 가능하다.¹⁹⁾

본 실험에서의 분석조건은 다음과 같으며, 분석절차 및 추출 예를 그림 1과 그림 2에 보였다.

-저역필터의 차단주파수 : 4.5KHz

-분석계산기 : 16bit micro computer

- 표본화 : 10KHz, 12bit A/D
- 분석시간길이 : 30ms
- window : Hamming window
- 고역강조 : 1 차 디지털 필터
- 기본주파수 추출 : 선형예측 잔차상관법
- AbS 분석 파라미터 (성대 음원 파라미터)
- Fi (i=1,5) (제1~5포먼트주파수)
- Fc (고차 보정주파수)

- AbS 합성모델

$$\log S(f) = HI(f) + \{\alpha \log f + \beta + \log f / (f^2 + BO^2)\}$$

$$+ \sum_{i=1}^5 \log (F_i^2 + B_i^2 / 4) /$$

$$\sqrt{\{(f + f_i)^2 + B_i^2 / 4\} \{(f - F_i)^2 + B_i^2 / 4\}}$$

$$+ \{R5(f/F_c)^2 + L5(f/F_c)^4\}$$

- 제 1 항 : 저역필터 주파수 특성 HI(f)
- 제 2 항 : 음원특성항
- 제 3 항 : 성도특성항
- 제 4 항 : 고차보정항

B_i : 포먼트 대역폭 B_i = α (0.015F_i + 25)

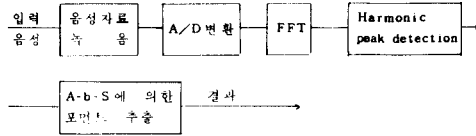


그림 1. 분석절차의 블록선도
Fig. 1. Block diagram of analysis process.

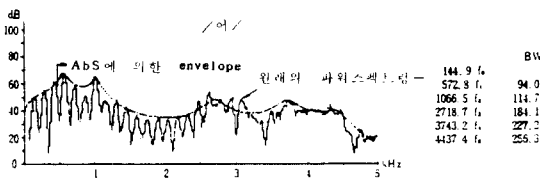


그림 2. AbS에 의한 포먼트주파수 추출예
Fig. 2. Formant frequency extraction using AbS method.

II. 성별, 연령별 피치 및 포먼트주파수의 변화

1. 연령과 피치주파수의 관계

피치주파수와 다른 음운 파라미터와의 관계를 보기 위하여 먼저 연령에 따른 피치주파수의 변화를 측

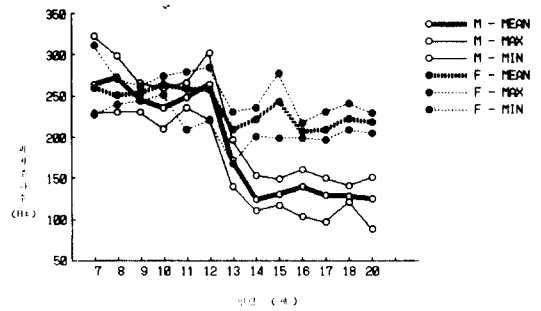


그림 3. 성별, 연령별 피치주파수의 변화
Fig. 3. Relation between the pitch frequency and the age of the speakers.

정한 결과를 그림 3에 보였다.

변성기 전의 남녀간의 피치주파수 차는 발견할 수 없었고 여성의 경우 변성기 전후의 피치주파수의 차는 약 0.2oct였다.

한편, 남성의 변성기에 의한 피치주파수의 차는 0.9oct 정도로 거의 1oct의 차를 보이고 있다.

이에 따르면, 어린이와 성인남자의 구별에 피치주파수가 유효한 파라미터로 생각되나 어린이와 성인 여자는 명확하게 구별하기 어렵다. 이러한 피치주파수의 변화는 성장에 따른 상대크기의 변화와 상관이 있을 것으로 추측되어 신체발육을 나타내는 신장 및 체중과 피치주파수의 상관계수를 구해 보았다.

신장 : 피치주파수 (-0.74)

체중 : 피치주파수 (-0.71)

이 결과에 의하면 강한 상관이 있는 것으로 나타나고 있다.

연령층을 7~8세와 17~18세로 고정시켰을 경우의 관계를 그림 4에 보였다. 이 결과로 보아 신장·체중이 피치와 직접 상관이 있다고 보기는 어렵고, 신장 및 체중과 피치간에는 연령변화를 매개로한 간접 상관만이 인정된다고 생각된다.

2. 연령과 포먼트주파수

포먼트주파수의 연령별 평균치 변화의 F1, F2 및 F2 F3 도상에서의 궤적을 그림 5에 보였다.

모든 모음의 F1, F2, F3는 원점을 향한 궤적의 경향을 보이고 있는데 이는 연령에 따라 발성기관의 구조적 변화가 일정한 것은 많으나 전체적인 성도의 길이는 연령에 따라 늘어나므로 공진주파수는 낮아지는 결과로 해석된다.

여기서, /애/와 /에/의 궤적이 전 연령층에 걸쳐서 겹쳐지는 것은 녹음시에 이를 구별하여 발성하지

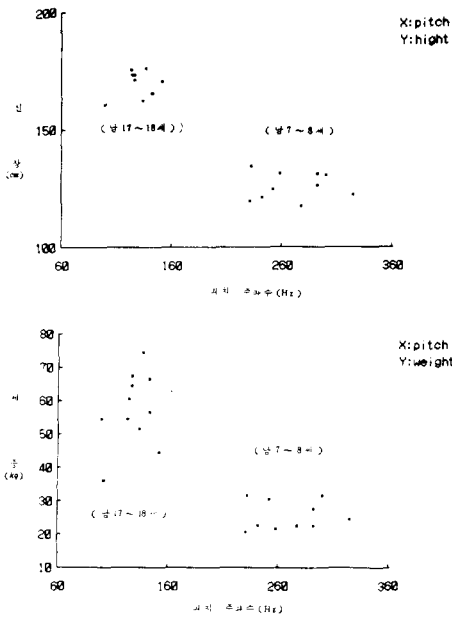


그림 4. 연령층을 고정했을 때의 신장, 체중 및 피치와의 관계
 Fig. 4. Relation between hight, weight and pitch when the age of speaker is fixed.

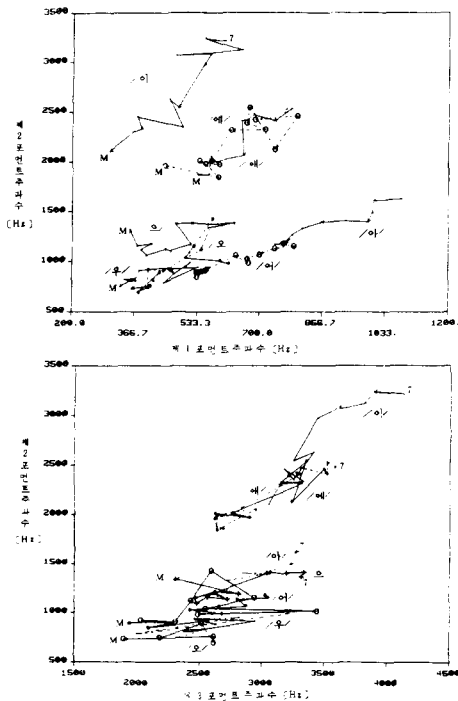


그림 5. 연령별 포먼트주파수의 평균치 변화
 Fig. 5. Changes of formant frequencies with age.

못한 발성자가 대부분이었던 것에 기인한다고 본다.
 F1, F2도상에서 볼때 고연령층(성도가 긴 발성자)의 /아/는 저연령층(성도가 짧은 발성자)의 /어/와, 고연령층의 /어/는 저연령층의 /우/와 /오/에 겹치고 있다.

그러나 우리는 실제로 이를 잘 구별해서 지각할 수 있으므로 포먼트의 절대위치와는 상관없이 음운이 지각됨을 알 수 있다. 편의상 7세~12세까지를 어린이, 13~15세까지의 남자를 소년, 이후는 성인으로 분류했을 때 각 그룹의 포먼트주파수의 평균치, 그리고 성인남자의 포먼트주파수에 대한 각 그룹의 포먼트주파수의 변화량의 비를 계산한 양 R_i 를 표 1에 나타

표 1. 포먼트주파수의 그룹별 평균
 Table 1. Averages of formant frequencies of vowels spoken by children boys and female and male adults.

VOWEL	P	F1	R1	F2	R2	F3	R3
/우/	CHILD	494	39.5	944	55	2654	8.5
	BOY	400	13	889	8.3	2640	8
	FEMALE	449	26.8	1044	27.2	2770	13.3
	MALE	354	0	821	0	2445	0
/어/	CHILD	596	60.6	3208	36	3915	21.6
	BOY	476	28.3	2501	6	3313	2.9
	FEMALE	598	61.2	2845	20.6	3725	15.7
	MALE	371	0	2359	0	3219	0
/아/	CHILD	1047	39	1607	35.6	3255	26.1
	BOY	783	4	1245	5.1	2715	5.2
	FEMALE	957	27	1493	26	3008	16.5
	MALE	753	0	1185	0	2581	0
/애/	CHILD	748	31.2	2532	28.4	3575	29.7
	BOY	596	4.6	2024	2.6	2748	-0.3
	FEMALE	674	18.2	2370	20.2	3199	16
	MALE	570	0	1972	0	2757	0
/어/	CHILD	764	42	1129	30.8	2910	35.5
	BOY	573	6.5	934	8.2	2564	19.4
	FEMALE	676	25.7	1012	17.3	2662	23.9
	MALE	538	0	863	0	2148	0
/에/	CHILD	750	33.7	2460	28.3	3490	32.8
	BOY	603	7.5	2058	7.3	2961	12.7
	FEMALE	662	18	2313	20.6	3118	18.6
	MALE	561	0	1918	0	2628	0
/오/	CHILD	540	34.3	1069	42.9	2576	-0.5
	BOY	413	2.7	832	11.2	2158	-16.6
	FEMALE	484	20.4	1052	40.6	2662	2.8
	MALE	402	0	748	0	2589	0
/으/	CHILD	559	45.2	1391	20.4	3239	22.3
	BOY	431	11.9	1102	-4.6	2575	-2.8
	FEMALE	495	28.6	1372	18.8	2989	12.8
	MALE	385	0	1155	0	2649	0

내었다. 이 결과에 의하면 성인남자를 기준으로 할 때, 어린이의 /어/, /에/, /애/는 F1, F2, F3의 변화 비율이 거의 일정하여, 이들 모음에 대한 조음 형태가 성인남자와 비슷하며 성도 길이에 대한 의존도도 비슷하다는 것을 알 수 있다.

각 모음의 그룹별 평균치 변화를 모음사각도의 형태로 보인 것이 그림 6이다. 여기서도 모음사각도가 개략적인 형태를 유지하면서 원점을 향하여 이동하고 있음을 알 수 있다.

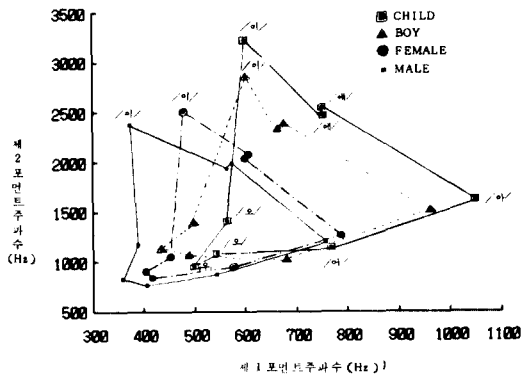


그림 6. 연령 그룹별 F1, F2의 변화

Fig. 6. Changes of the first and second formant frequency with age group.

3. 피치와 포먼트주파수의 관계

피치와 포먼트주파수간에 상관 정도를 보기 위하여 표 2에 각 모음의 피치와 F1, F2, F3의 상관계수를 보였다.

이를 보면 모음에 따라 정도의 차이는 있으나 피치가 낮은 발성자는 피치가 높은 발성자에 비해 포먼트

표 2. 피치와 포먼트주파수의 상관계수
Table 2. Correlation coefficient between pitch and formant frequency.

	F ₁	F ₂	F ₃
/우/	0.57	0.13	0.1
/이/	0.52	0.78	0.58
/아/	0.71	0.61	0.62
/애/	0.69	0.66	0.68
/어/	0.73	0.58	0.43
/에/	0.66	0.6	0.69
/오/	0.69	0.3	0.1
/으/	0.71	0.47	0.55

트주파수도 낮게 나타나고 있는데 이러한 결과가 연령변화에 의한 것인지, 아니면 자연적인 피치와의 상관에 의한 것인지를 보기 위하여 /아/ 모음을 대상으로, 연령층을 7~8세 어린이와 성인남자로 고정했을 때의 피치 및 포먼트 주파수의 관계를 보인 것이 그림 7이다. 발성의 개인차에 비해 피치와 포먼트 주파수간의 상관은 찾아보기 어렵다. 따라서 본 자료에 의하면 피치와 포먼트주파수 간에는 직접 상관보다는 연령차를 매개로 간접적인 상관만이 인정된다.

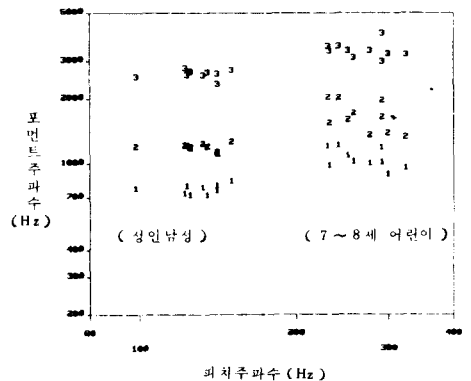


그림 7. 연령층을 고정했을 때의 피치 및 포먼트의 상관(/아/의 경우)

Fig. 7. Relation between formant frequencies and pitch frequencies when the age of speaker is fixed.

IV. 인식실험 및 검토

1. 피치 및 포먼트주파수에 의한 식별의 가능성
일반적으로 F1, F2에 의해 대부분의 모음이 특징 지워진다고 한다. 그러면 우리말의 경우, 성별, 연령별로 다양한 발성자의 모음을 F1과 F2만으로 어느 정도 식별이 가능한지를 보기 위하여 본 실험에서 사용한 모든 측정치의 F1, F2의 분포를 그림 8에 보였다. 여기서 타원은 대상 데이터의 90%를 포함하는 확률타원을 나타낸다. (이하같음)

이 결과에 의하면 /에/와 /애/는 발성의 불명확 때문이라 하더라도 /어/, /오/, /우/, /으/의 분포가 상당부분 겹쳐지는 것으로 보아 F1, F2만으로는 충분히 구별할 수 없을 것으로 보인다.

아울러 F2, F3의 분포를 그림 9에 보였다. 여기서도 /이/, /에/, /애/와 기타 모음간의 대분류는 어느

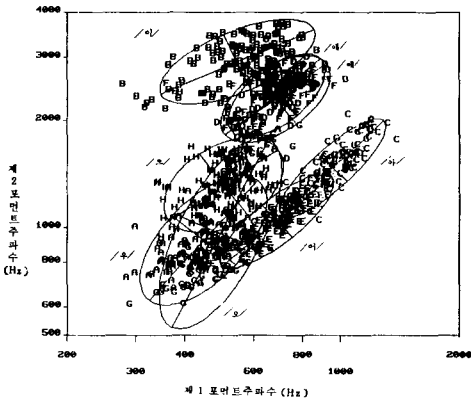


그림 8. 각 연령층의 F1, F2 분포 (대상데이터수 = 902)
 Fig. 8. F1-F2 diagram of 8 Korean vowels spoken by children, youth and adults.

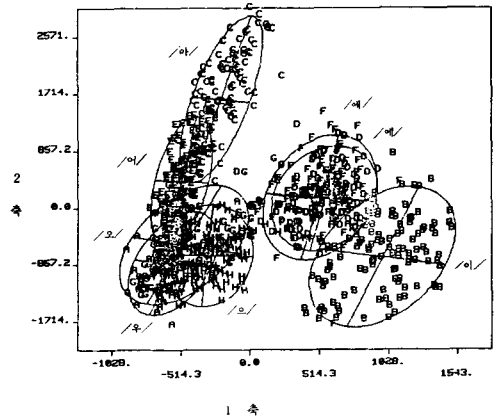


그림 10. 제 1, 2, 3 포먼트주파수에 의한 주성분 분석
 Fig. 10. Principal component analysis using first 3 formant frequencies.

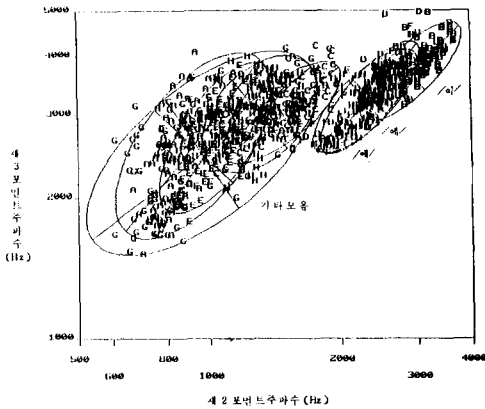


그림 9. 각 연령층의 F2-F3 분포 (대상데이터수 = 902)
 Fig. 9. F2-F3 diagram of 8 Korean vowels spoken by children, youth and adults.

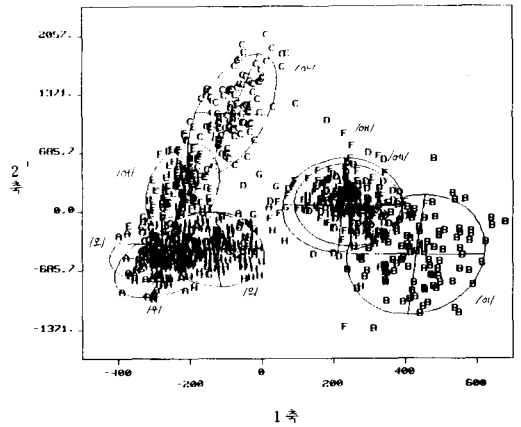


그림 11. 피치 및 제 1, 2, 3 포먼트주파수에 의한 주성분 분석
 Fig. 11. Principal component analysis using pitch and first 3 formant frequencies.

정도 가능해 보이나 세부적 식별은 곤란하다.

F1, F2, F3를 대상으로 선형관별함수에 의한 식별 가능성을 검토하기 위하여 주성분 분석(29) 한 후 기여율이 큰 제 1 축과 제 2 축을 도식한 것이 그림 10이다.

/아/와 /어/를 비롯하여 분포가 보다 뚜렷해지고 있으나 /오/, /우/, /으/는 역시 구별이 어렵다.

포먼트 주파수와 함께 피치까지 포함하여 주성분 분석한 결과를 그림 11에 보였는데 분포가 더욱 집중되어 식별 가능성을 높여주고 있다. 이상의 결과에

의하면 전 연령층의 발성자를 대상으로 하여 단번에 식별하는 데는 한계가 있음을 알 수 있다.

2. 개인차 정규화

포먼트주파수를 이용한 모음의 자동인식과 관련하여 개인차를 정규화 하고자 하는 여러가지 시도가 있는데²⁾ 이 중 몇가지의 예를 본 측정데이터에 적용하여 정규화의 효과를 살펴본다. 단 여기서 인용한 좌표는 참고문헌등에 자주 인용되는 것을 이용하였을 뿐, 객관적으로 가장 우수한 정규화 좌표라고는 단

언 할 수 없다.

입의의 F1, F2, F3에 대하여

Iri 좌표는 다음과 같이 정규화 좌표를 계산한다.

$$\alpha = -\frac{10}{3} \log_{10} \left(\frac{F_2^2}{F_1 \cdot F_2 \cdot F_3} \right)$$

$$\beta = \frac{10}{3} \log_{10} \left(\frac{F_3^2}{F_1 \cdot F_2 \cdot F_3} \right)$$

Fujisaki 좌표는 동일모음을 발생시킬때, 인간의 성도의 형태가 연령, 성별에 의한 성도길이의 차에도 불구하고 본질적으로 비슷하다면, 입술의 방사를 무시했을 때 이론적으로 동일모음의 포먼트주파수 F1, F2, F3의 상대비는 모두 같다고 보고 다음과 같이 극좌표로 나타내었다.

$$r = (F_1^2 + F_2^2 + F_3^2)^{1/2},$$

$$\phi = \tan^{-1} (F_1 / F_2),$$

$$\theta = \tan^{-1} \{ (F_1^2 + F_2^2)^{1/2} / F_3 \}$$

여기서 r방향의 데이터의 분포는 단지 성도길이, 연령, 성별등 개인차에 기인한 것으로, 모음 상호간의 구별에는 거의 공헌하지 않고 (ϕ, θ)가 정규화 좌표가 된다.

PAS(심리적 청공간 좌표)는 합성모음의 청취실험 결과로 구성된 다차원 척도공간에서 도출된 것으로 다음 2가지 파라미터 a1, a2가 각 모음의 정규좌표가 된다.

$$a_1 = \log_{10} [(F_1^{0.4} \cdot F_3^{1.3}) / F_2^{2.1}] + 0.3$$

$$a_2 = \log_{10} [F_3^{1.8} / (F_1^{1.1} \cdot F_2^{0.7})] + 0.8$$

이를 이용하여 좌표변형한 결과를 그림12, 13, 14에 보였다. 다소 정규화 효과가 있음을 시각적으로 비교 확인할 수 있다.

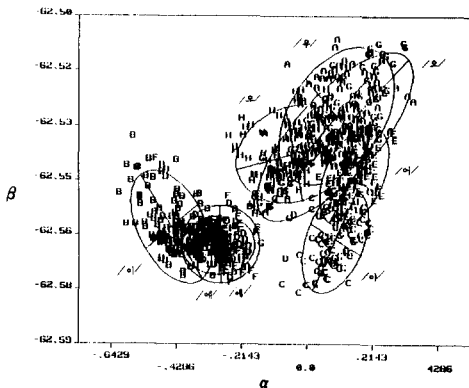


그림12. Iri 좌표에 의한 전체 데이터의 표현
Fig. 12. All data representation by Iri coordinate.

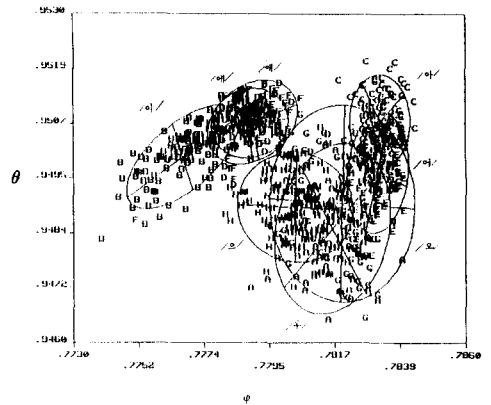


그림13. Fujisaki 좌표에 의한 전체 데이터의 표현
Fig. 13. All data representation by Fujisaki coordinate.

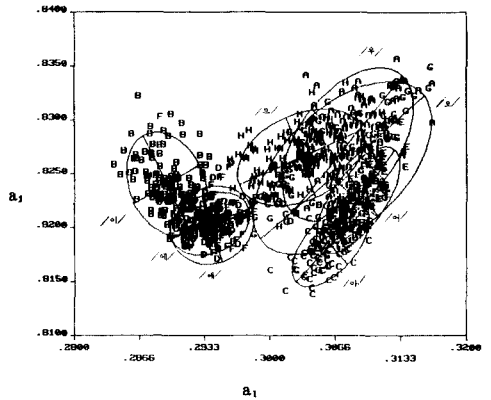


그림14. PAS에 의한 전체데이터의 표현
Fig. 14. All data representation by PAS coordinate.

3. 인식실험

이상의 각종 예비분석을 기초로 통계적 방식에 의해 인식실험을 수행하였다.

본 실험에서는 인식을 위하여 다음과 같은 통계적 거리²²⁾들을 사용한다. 미지입력파라미터를 X라고 카테고리 K에 대한 표준패턴을 μ_k (평균치), W_k (공분산행렬)로 하면

○ 유클리드거리

$$D_E = \sqrt{(X - \mu_k)' (X - \mu_k)}$$

○ 마하라노비스거리

$$D_M = (X - \mu_k)' \cdot W_k^{-1} (X - \mu_k)$$

○ 베이스 거리

$$D_B = (X - \mu_k)^t \cdot W_k^{-1} (X - \mu_k) + \log |W_k|$$

단 t는 전치행렬, W^{-1} 는 공분산의 역행렬, $|W|$ 는 행렬식을 나타낸다. 인식방법으로는, 인식할 입력 파라미터를 포함한 대상그룹 전체로 표준패턴을 만든후 각각의 입력을 인식하는 closed test를 사용하였다.

실험결과를 표 3과 표 4에 정리하였다.

4. 결과 및 검토

전체를 일괄인식할 때 포먼트주파수와 피치주파수를 같이 이용하는 것이 포먼트 주파수만을 이용하는 것보다 약 6% 정도 인식률이 상승(bayes거리의 경우) 된다. 그러나 어린이 성인남자 성인여자로 분류하여 인식할 경우에는 피치의 인식률에 대한 기여

표 3. 피치 및 포먼트주파수에 의한 인식률
Table 3. Recognition rates by using pitch and formant frequencies.

()는 e + ε

인식 파라미터	대상		전체	어린이	성인남자	성인여자
	거리	데이타수				
F ₁ , F ₂	유클리트거리		902	474	197	231
	유클리트거리		56.8 (66.5)	64.6 (75.5)	74.6 (85.8)	66.7 (76.2)
	Mahalanobis거리		59.4 (70.7)	67.9 (78.5)	81.2 (91.9)	67.1 (77.5)
	Bayes거리		63.2 (74.2)	69 (80.4)	80.2 (89.8)	71.4 (81.8)
F ₁ , F ₂ , F ₃	유클리트거리		48.1 (56)	56.8 (68)	60.9 (72)	64.5 (74)
	Mahalanobis거리		64.2 (76)	72.6 (82)	79.2 (89)	69.7 (78)
	Bayes거리		66.6 (78)	72.6 (82)	79.2 (89)	74.5 (84)
피치 및 F ₁ , F ₂	유클리트거리		56.8 (66.5)	64.6 (75.5)	74.6 (85.8)	66.7 (76.2)
	Mahalanobis거리		69.1 (80.2)	67.9 (78.7)	82.2 (91.9)	68 (78.4)
	Bayes거리		70.5 (82.6)	69.4 (80)	81.2 (90.4)	69.3 (76.3)
피치 및 F ₁ , F ₂ , F ₃	유클리트거리		48.1 (56)	56.8 (68)	60.9 (71.6)	64.5 (74.5)
	Mahalanobis거리		70.5 (81)	72.2 (81.9)	81.2 (89.8)	71 (78.4)
	Bayes거리		72.1 (84)	73.4 (83.3)	81.2 (90.9)	73.2 (82.7)

표 4. 정규화 좌표변환에 의한 인식률
Table 4. 3 Recognition rates by coordinate conversion.

()는 /예/+/애/

전체 데이타	인식파라미터 거리	F ₁ , F ₂ , F ₃ 의 주성분분석 1, 2축	Iri 좌표	Fujisaki 좌표	PAS
		유클리트거리	60 (71)	63.4 (74)	59.7 (70)
Mahalanobis거리	62.8 (74)	61.2 (72)	56.2 (69)	56.2 (69)	
Bayes거리	64.4 (76)	64.2 (75)	61.8 (73)	62 (73)	

가 별로 크지 않다. 따라서 피치주파수가 주로 대부분 기여하고 있다고 볼 수 있다.

피치와 제1,2 포먼트주파수를 이용하는 것과 제1, 2,3포먼트주파수를 이용한 인식률은 Mahalanobis 거리나 Bayes 거리의 경우 대동소이하나 유클리트거리에 의하면 피치를 이용하는 편이 나은 결과를 보이고 있다.

유클리트거리에 의한 경우 제3포먼트를 이용하지 않는쪽이 이용하는 경우에 비해 전체적으로 인식률이 높은것이 특이하다. Confusion matrix에 의해 검토한 결과 /오/와 /우/간의 오인식이 제3포먼트주파수를 이용하면 더 커지고 있었다. 포먼트 추출시 /오/와 /우/의 포먼트주파수 추출 정도가 다른모음에 비해 낮았던 것이 한 원인으로 생각된다.

각종 정규화 방법은 전체 데이타를 대상으로 할 경우 효과가 비슷하다.

F₁, F₂, F₃를 주성분 분석하여 제1,2축에 의한 인식률을 구하여 보니 정규화 좌표에 의한 인식률과 거의 비슷한 결과를 보이고 있어서 포먼트 주파수만을 인식파라미터로 하여 간단한 선형판별함수에 의해 인식할 경우는 좌표변환이 유리하다고 판단된다.

V. 결론

지금까지 음소인식에 있어서의 개인차 해결을 위한 기초적 분석으로서 피치주파수와 함께 모음을 특징지워주는 포먼트주파수를 Abs법에 의해 추출하여 성별·연령별 이들 특징량의 변화를 고찰하였다. 그 결과를 정리하면 다음과 같다.

- (1) 연령, 성별에 따른 피치주파수의 변화를 보면 변성기전의 어린이는 남녀간의 차이를 인정할 수 없었고 변성전후의 차는 여자가 약 0.2oct, 남자는 0.9oct 정도를 나타내고 있음을 알았다.

- (2) 연령에 따른 포먼트주파수 변화는 모음에 따라 정도의 차이는 있으나 연령이 높아질수록 대체로 낮아지고 있다.
- (3) 피치와 포먼트주파수간에는 연령별 변화에 의한 측정치로 보면 피치가 낮은 발성자는 포먼트주파수도 낮게 나타나고 있으나 연령층을 고정했을 경우에는 개인차에 비해 양자간의 상관관계를 찾아보기는 힘들었다. 따라서 피치와 포먼트 주파수간에는 연령변화를 매개로 간접적인 상관관계는 있으나 직접적인 상관관계가 있다고 보기는 어렵다.
- (4) 피치와 포먼트주파수를 이용한 인식실험을 각종 통계적 거리를 이용하여 수행한 결과, 포먼트주파수만을 대상으로 할 경우에는 전체를 일괄적으로 인식하는 것보다 어린이, 성인여자, 성인남자로 대별해서 인식하는 것이 유리하다.
- (5) 포먼트주파수와 아울러 피치도 특징파라미터로 하여 인식할 경우 위와같은 대분류의 효과는 거의 없고 인식률은 포먼트주파수와 대분류를 병용한 경우와 대동소이하다. 따라서 연령층이 다양한 경우는 피치주파수가 인식에 기여하는 바가 크다는 것을 알 수 있다.
- (6) 각종 정규화 좌표를 사용한 인식실험결과에 의하면 인식률이 거의 비슷하였고 F1, F2, F3의 주성분 분석에 의한 제1, 2축에 의한 인식률과 비교하여도 1~2% 정도의 차에 불과하였다. 따라서 포먼트주파수만 인식파라미터로 하여 간단한 선형 판별함수에 의해 인식할 경우는 좌표변환이 효과가 있음을 확인하였다.

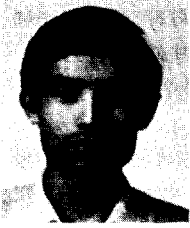
참 고 문 헌

- [1] 이용주의, “음성정보처리 장치의 동향” 전자통신 동향분석 1988. 4.
- [2] 鹿野, 博松(이용주 김경태 역), “음성이해 연구의 동향” 전자통신 9-1, 1987. 3.
- [3] Fant, “Acoustical analysis of speech with application of Swedish,” *Ericson Technics*, 1, 1953.
- [4] 粕谷, 鈴木, 城戸, “年齢, 性別による日本語5母音のヒッチ 周波數とホルマント周波數の變化” *日本音響學會誌*, 24-6, 1968.
- [5] 中川, 白方, 山尾, 坂井, “不特定話者の音聲自動認識のための性別年齢差による話者分類の考察” *日本電子通信學會論文誌* vol. J63-D

no.12, 1980, 12.

- [6] 藤崎, 片桐, 佐藤, “不特定多數話者の定常母音の特徵抽出と認識” *日本音響學會音聲研究會資料* S 77-08, 1977, 6.
- [7] 한국통신기술연구소 “전화통화의 음성분석에 관한 연구” 과학기술처 연구보고서, 1981.
- [8] 최진태, “한글 연속음의 주파수분석적 연구” *대한전자공학회지*, vol. 9, no. 1.
- [9] 박종철, “Sound spectrograph에 의한 우리말 단모음의 분석에 관한 연구” *연세대 석사학위논문*, 1985. 2.
- [10] 李勇柱外 “韓國語の單母音認識に關する一考察” *日本音響學會講演論文集*, 1986. 3.
- [11] 이용주, 차균현, “통계적 방법에 의한 단모음의 인식” *대한전자공학회 하계학술발표회 논문집*, 1986. 7.
- [12] 鄭鉉烈外, “ホルマント情報を利用した韓國語單母音の分析と認識” *日本音響學會講演論文集* 1986. 10.
- [13] 鄭鉉烈外, “韓國語單母音の個人性と方言性に關する檢討” *日本音響學會講演論文集*, 1987.3.
- [14] Bu-11 Kim et al, “Analysis and recognition of Korean vowels,” *Ann. Report of the Eng. Res. Ins. Univ. of Tokyo* vol. 32, 1973.
- [15] Itakura, Saito, “Analysis Telephony based on the Maximum likely hood method,” *Report of 5th ICA*, Tokyo, 1968.
- [16] Potter, G. Kopp, H. Kopp, *Visible speech* Dover Publication, Inc. New York 1966.
- [17] Bell et al, “Reduction of Speech Spectra by Analysis by synthesis techniques,” *J.A.S.A* vol. 33, no. 12, 1961.
- [18] 三輪外, “ペンコンを用いた日本語方言音聲分析システム” *日本岩手大 工學部 研究報告* vol. 37, 1984.
- [19] 藤崎外, “各種のフォルマント周波數抽出誤差の比較” *日本音響學會音聲研究會資料* S73-20, 1903.
- [20] 竹内, 柳井, *多變量解析の基礎* 東洋經濟新報社, 東京, 1972.
- [21] 中田知男, *音聲*, コロナ社 東京, 1977.
- [22] Duda, Hart, *Pattern classification and scene analysis* John Willey & Sons New York, 1973. *

 著 者 紹 介



李 勇 主(正會員)

1954年 1月 17日生. 1976年 2月 고려대학교 전자공학과 졸업 공학사학위 취득. 1987年 8月 고려대학교 전자공학과 졸업 공학석사학위 취득. 현재 고려대학교 전자공학과 박사과정. 1976年 3月~1980年 7月 공군통신장교. 1985年 7月~1986年 1月 일본 도호쿠대학 응용 정보학 연구센터 연구생. 1980年 8月~현재 한국전자통신연구소 음향연구실 선임연구원. 주관심분야는 음성의 분석, 합성, 지각 및 인식, 통화품질, 전기음향 등임.



金 敬 泰(正會員)

1949年 5月 9日生. 1972年 2月 경북대학교 전자공학과 졸업 공학사학위 취득. 1980年 8月 연세대학교 대학원 전자공학과 졸업. 공학석사학위 취득. 1985年 3月 일본 Tohoku대학 전기통신공학과 공학박사학위 취득. 1978年 1月~1986年 1月 한국기계연구소 근무. 1986年 1月~현재 한국전자 통신연구소 근무. 주관심분야는 음성정보처리, 정보이론및 학습기계 등임.



車 均 鉉 (正會員) 第25卷 第9號 參照
현재 고려대학교 전자전산기
공학과 교수