필터뱅크 분석법을 사용한 한국어 음소의 인식에 관한 연구

남 문 현 ° 주 상 규
건국대학교 전기공학과

A Study on the Recognition System of Korean Phonemes Using Filter-Bank Analysis

Moon-Hyon, Nam and Sang-Gyu, Ju
Dept. of Electrical Engineering, Kon-Kuk University, Seoul

- ABSTRACT -

The purpose of this study is to design a phoneme-class recognition system for Korean language using filter-bank analysis and zero crossing rate method. First, the speech signals are seperated in 16 bandpass filters to obtain short-time spectrum of speech signals, and digitized by 16-ch A/D converter. And then, with the set of features which extracted from patterns of ratios of each channel energy level to overall energy level, the decision rules are made for recognize unknown speech signal. In this experement, the recognition rate was about 93.1 percent for 7 vowels under multitalker environment and 74.4 percent for 10 initial sounds at single speaker.

1. 서 론

컴퓨터를 이용한 자동 음성인식의 기술은 인간의 이해 능력에까지 도달하는 것이 궁극적인 목표가 된다. 이러한 음성인식의 기술과 학문은 과거 20여년전부터 컴퓨터 과학의 발전에 따라 분절 단어를 인식하고 단어를 연속적으로 간단히 하는 능력을 주어서 결합을 할 수 있게 되었다[1]-[4]. 그리고 광범위한 어휘의 인식을 위해서 음소별 인식의 필요성이 절실해지고 있으며[5], 인식된 음소를 결합하여 단어로 변환하는 것이 연구되고 있다[6], [7].

본 논문에서는 한국어를 기본 문법과 같이 초성, 중성 및 종성으로 구분하여 초성자음 열 개와 단모음 일곱 개를 음소별로 규칙에 의해 인식하는 시스템을 구성하여, 광범위한 어휘의 인식 및 실용성에 접근을 시도하였다. 실시간 처리를 위하여 간단한 하드웨어와 짧은 실행시간의 소프트웨어로 구성할 수 있는 필터뱅크 방식을 사용하였으며, 무성음과 유성음의 구분을 위하여 영교차측정법을 보조적으로 사용하였다.

2. 음성의 분석 방법

음성을 인식하려면 음성신호를 해석하여 음성이 가지고 있는 물리적 매개변수를 추출해 내어 음성신호를 분석하는 것이 우선 과제이며, 주파수 영역에서의 분석방법과 시간 영역에서의 분석방법이 있다[8]. 본 연구에서는 시간영역의 분석방법으로 영교차율 측정법을 채택하였고, 주파수 영역의 분석을 위하여 필터뱅크 분석법을 사용하였다.

(1) 영교차 측정

신호의 디지탈 표현에 대한 상황에서 영교차는

$$\mathrm{sgn}\,[X(n)] = \mathrm{sgn}\,[X(n-1)] \quad \text{------ (1)}$$

일 때 샘플 상수 사이에서 발생한다. 신호가 주파수 f_o인 교류파형 (sinusoid)이라면 영교차의 평균수는

$$Nz = 2 f_o \text{ crossings} / s \quad \text{---------- (2)}$$

이다. 마찰음의 스펙트럼은 3 kHz 이상에서 집중되는 대 반하여 유성음의 에너지는 3 kHz 이하에서 집중되는 경향이 있어, 영교차 측정은 음성의 특별한 분절(segment)이 유성음인지 무성음인지를 결정하는 데 자주 사용된다.

(2) 필터뱅크 분석

음성의 해석에는 파형의 주파수 스펙트럼 형태에서 각 음소에 따라 뚜렷한 형성을 보이는 포르만트(formant) 주파수를 분석하는 방법중 가장 보편적인 방법의 하나이다. 근접한 Q개의 대역통과 필터를 분석하고자 하는 대역을 포괄하도록 배치하여 실시간으로 단시간 스펙트럼의 개략적인 형태를 얻을 수 있다는 장점과, 스펙트럼의 형태가 음성의 특성을 잘 나타내어 널리 사용된다[3], [9]-[12].

필터뱅크 분석 시스템에서 음성신호는 보통 100에서부터 3000~8000Hz의 차단주파수까지를 포괄하는 Q개의 대역통과 필터를 통과한다. 필터의 수 Q는 5~32까지를 사용하고 필터의 간격은 1000Hz까지는 선형이고 1000Hz 뒤에는 대수적으로 증가시킨다. 각 대역통과 필터의 출력은 비선형요소 (예를 들면, 전파정류기)를 통과하고 저역통과

필터를 거쳐 해당대역의 음성신호의 에너지에 비례하는 신호로 변환된다.

(3) 음성인식 순서

음성인식 시스템은 기본적으로 (a) 특징 추출, (b) 형태 유사성 추출, (c) 결정 방법의 세 단계를 거친다. 1단계 처리는 데이타를 축소시키는 단계로서 음성신호의 특징을 가장 작게 변형시킨다. 다음 단계는 시험형태와 기준형태 사이의 시간 배열 사이의 형태 유사성과 시간길이 등을 정의한다. 마지막 단계는 미지의 시험 형태와 기준 형태를 연결시켜 선택하는 과정이다. 음성의 판별법으로는 형태일치법과 판별규칙에 의한 방법의 두 가지로 나눌 수 있으며, 형태일치법은 각 발음의 형태를 시스템에 미리 저장되어 있는 표본과의 거리를 계산하여 가장 일치하는 것으로 결정하는 방법으로서 표본의 변경이 유리한 반면 판단시간이 길어지는 단점이 있다. 반면, 판별규칙의 사용은 각 발음의 절대적이거나 상대적인 특징을 알고리즘으로 구성하여 필요한 특징만을 비교하여 음소를 결정한다[9]. 이 방법은 규칙의 수정이 용이하지 못한 반면에 고속의 실행시간을 장점으로 갖고 있다. 일반적으로 단어 단위의 인식에는 주로 형태일치법이 사용되고 있으며, 음소인식을 위한 본 실험에서는 판별규칙을 사용하였다.

3. 음성인식 시스템의 설계

(1) 시스템 구성

본 논문에서 구성한 시스템의 구성도는 그림 1과 같다. 다이나믹 마이크 MC-100을 통해 음성신호가 입력되며, 고역의 에너지가 저역의 에너지보다 작은 것을 보상하기 위해 앞단에 프리엠파시스를 둔다. 다음에 전치증폭기에서 신호를 증폭하고 완충증폭기를 통해 각 대역통과 필터에 신호를 공급한다. 대역통과 필터, 전파 정류기, 저역통과 필터로 1 채널을 구성하여 모두 16개의 채널로 구성된다. 각 채널에서 출력되는 에너지와 영교차율 변환기의 출력이 아날로그 멀티플렉서와 A/D 변환기를 통해 컴퓨터에 입력되어 표본 추출 및 판별과정에 이용된다.

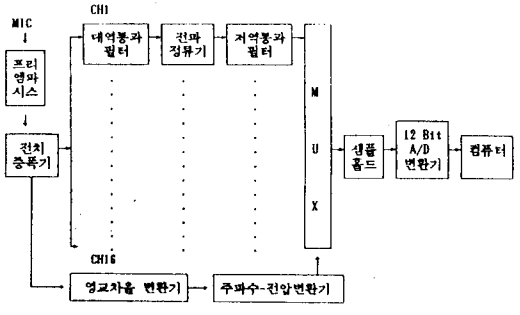


그림 1. 음성인식 시스템의 구성도

Fig. 1. Experimental system for speech recognition

표 1. 각 대역통과 필터의 중심주파수 및 대역폭

Table 1. Center frequency and band-width of each bandpass filters

채널	중심주파수	대역폭	Q
0	200	100	2
1	300	100	3
2	400	100	4
3	500	100	5
4	600	100	6
5	700	100	7
6	800	100	8
7	900	100	9
8	1,000	100	10
9	1,150	150	7.7
10	1,300	150	8.7
11	1,500	200	7.5
12	1,800	300	6
13	2,200	400	5.5
14	2,800	600	4.7
15	3,600	800	4.5

(2) 회로의설계

가) 대역통과 필터의 설계

본 연구에서는 음성신호를 실시간으로 처리하기 위한 1단계 처리 방법으로 음성신호를 16개의 대역통과 필터로 분리한다. 각 대역통과 필터에 연산증폭기를 사용하였으며 특성은 표 1에 나타내었다. 그리고 각 필터회로에는 중심주파수와 이득을 조정하는 가변저항을 사용하였으며, FCR-6B 주파수 응답 추적기를 이용하여 각 필터를 교정하였다.

나) 영교차율 변환기

음성신호가 무성음인지 유성음인지를 구분하는 보조적인 방법으로 신호가 영 준위를 교차하는 빈도를 측정한다. 연산증폭기의 포화특성을 이용하여 영준위를 교차하는 신호를 양전원과 음전원의 크기로 증폭하여 준다. 다음에 주파수의 측정은 주파수-전압 변환기를 사용하여 주파수를 아나로그 준위로 변환시켜서 A/D 변환기를 이용하여 컴퓨터에 입력시킨다.

(3) 컴퓨터와의 접속

신호의 해석 및 인식을 위한 컴퓨터는 8 bit APPLE II 개인용 컴퓨터를 사용하였으며, 신호의 입력은 12 bit 아날로그-디지탈 변환기를 통해 입력되고 해석 결과의 출력은 HP-7475A 플로터와 도트 프린터를 사용하였다.

4. 음성인식 실험

(1) 음성의 분석 그림 2, 3은 발음된 한 음절의 신호를 A/D 변환시켜 플로터로 파형을 출력한 것이며, 그림 4는 16개의 채널을 통한 음성 에너지의 시간에 따른 변화이다. 구간 t1이 초성 자음에 해당되는 시간으로 VOT(Voice Onset Time)라 하며 음성이 모음으로 안정되기까지의 시간이다.

유성음은 기류가 성대를 통과함과 동시에 긴장된 막 사이에서 공기가 진동을 일으키며 지나가는 반면에, 무성정지음에서는 성대가 폐쇄된 막을 열어 공기를 통과시킨 후에 진동을 울린다. 곧, 폐쇄된 막이 열리고 진동으로 소리내는 시간 VOT가 유성음 [g], [d], [b]에서는 영에 가깝고, 무성음 [k], [t], [p]에서는 상대적으로 오랜 시간이 경과한다. 우리는 일반적으로 VOT가 30msec보다 작으면 유성음으로, 그보다 길면 무성음으로 판별한다. 이러한 VOT의 차이를 각 대역통과 필터를 통과한 음성 에너지의 시간에 따른 변화를 나타낸 그림 4에서 관찰할 수 있다. 유성음은 VOT가 짧아 처음부터 뚜렷한 포르만트의 형성을 보여주는 반면, 무성음은 포르만트의 형성이 흐려진다[17].

또한, 처음의 초성에 해당되는 부분과 중성의 에너지가 모음에 비해 매우 작음을 알 수 있다. 따라서 초성과 중성의 구분은 신호의 크기로 구분할 수 있게 된다. 본 연구에서는 이것을 이용하여 트리거 레벨 V1을 설정하여 발음의 시작을 포착하여 일정한 시간동안 자음부분을 표본화한 다음 다시 모음의 구분 레벨 V2를 초과하면 모음으로 인식하였다. 이렇게 함으로써 발음이 자음에서 모음으로 변하는 구간을 취하지 않게 되어 보다 정확한 인식이 이루어지도록 하였다.

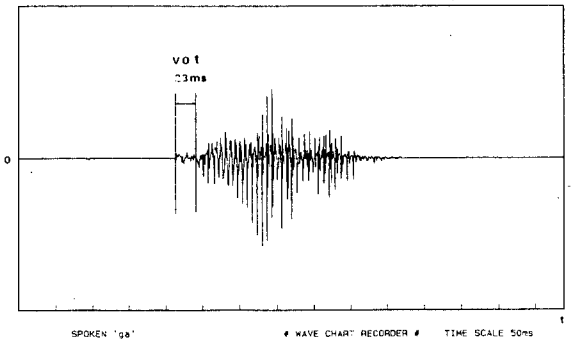


그림 2. 유성음의 파형
Fig. 2. Waveform of the voiced

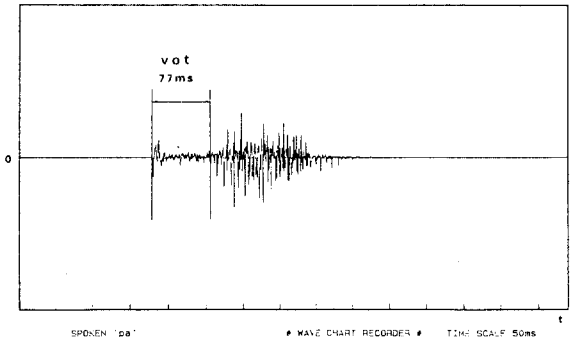


그림 3. 무성음의 파형
Fig. 3. Waveform of the unvoiced

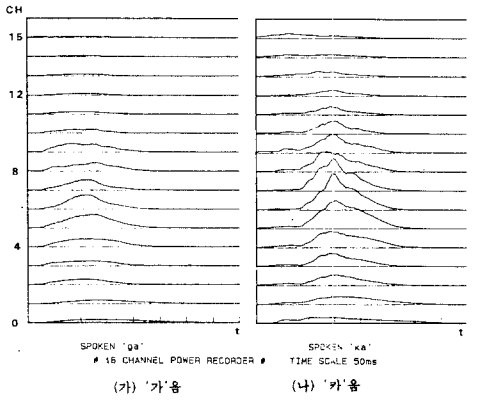


그림 4. 유성음과 무성음의 채널별 비교
Fig. 4. Records of the voiced and the unvoiced

한국어의 대표적인 단모음은 학자에 따라 차이가 있지만 [아], [어], [오], [우], [으], [이], [에], [애] 등을 들 수 있으나 [에]와 [애] 사이에는 뚜렷한 특징이 없다. 이는 발음의 유사성이 많기 때문이며 실제로 두 발음을 정확히 발음하는 사람도 많지 않다[13]. 따라서 본 연구에서는 [에]로서 [애] 발음을 대표하여 일곱 개의 단모음을 분석 및 인식하였다.

본 연구에서는 각 발음을 필터뱅크 분석방법으로 여러 번 플로터에 중첩을 하여 도시한 후 각 자음 및 모음의 특징을 분석하였다. 그림 5는 모음의 분석 결과로써 기존의 분석결과와 매우 유사한 결과를 얻을 수 있었으며 초성 자음의 분석 결과는 그림 6과 같다.

초성자음은 발성신호를 트리거하여 처음 10 msec 동안의 주파수 스펙트럼과 영교차율을 측정하여 표본을 추출하였으며, 모음은 [아]를 갈이 발음하였다.

(2) 음성 인식의 실험

인식의 방법은 대역통과 출력을 분석한 결과를 종합하여 각 채널의 에너지 구성비와 영교차율의 특징으로 각 자음 및 모음간의 판별규칙을 작성하였다. 결과의 출력은 모니터상에 나타내거나 프린터로 출력하였으며, 특히 인식 프로그램과 분석 프로그램을 일체화하여 잘못 인식되는 경우 재분석 및 판별규칙의 수정을 용이하게 하였다.

초성자음의 인식에서 [가], [카], [하] 및 [사], [자], [차]의 초성은 각각 유사한 형태로 나타났으며 영교차율로서 구분되었다. [나], [다], [라], [바]의 초성은 각기 주파수 스펙트럼의 특징으로 구분되었으며 [마]는 [나]와 유사하여서 [나]로 대표하였다. 초성의 자음은 단독으로 발음이 힘들어 [아]를 갈이 발음하여 기준을 작성하였는데, 이렇게 만들어진 판별 기준으로 [어], [이], [에] 등 다른 모음의 앞에서 발음될 경우에도 인식률은 약간 떨어지지만 판별이 가능하였다.

모음의 인식에서는 [아], [이]는 특징이 뚜렷하여 거의 완벽한 인식을 얻을 수 있었다. [어], [에]는 발성자에 따라 약간의 혼동이 있었으며 [오], [우], [으]는 특징의

경계에서 상호간에 오인식이 나타났으나 판별기준의 수정으로 인식률을 높일 수 있었다.

5. 결과 및 고찰

본 논문의 특징은 음성신호의 특징 추출방법과 음소 판별방법에 있다. 각 채널의 에너지를 절대치로 이용하지 않고 전체 에너지 중의 구성비로 된 상대적인 비율로 해석 및 인식을 함으로써 소리의 크기와 잡음에 큰 영향을 받지 않는 시스템을 구성할 수 있었다. 시간 영역의 음성신호를 주파수 스펙트럼으로 변환하는 방법으로서 대역통과 필터를 채택하여 복잡한 계산과정을 피하였으며, 판별규칙에 의한 인식으로 실시간의 해석 및 인식을 수행할 수 있었으며 인식실험의 결과는 다음과 같다.

1) 표 2의 인식결과에서 보인 대로 기준 형태의 작성에 관여한 단일 발성자에 대한 단모음 인식률은 98.0%였으며, 불특정 발성자에 대한 인식률(5명의 남성이 각 발음을 10회씩 발성)은 88.9%였으며, 몇번의 발음 연습 후에는 93.1%로 향상되었다.
2) 초성 자음의 인식에 있어서는 자음에 관한 기존 분석 자료의 부족과 실제로 자음의 인식이 난이도가 높음으로 인해 단일 발성자조건에서 표 3의 결과와 같이 74.4%의 인식률을 기록하였으며, 5명의 임의 발성자에게서도 50% 정도의 인식률을 얻을 수 있었다.
3) 초성자음중에서 [ㄱ], [ㅋ], [ㅎ]은 주파수 스펙트럼의 유사성이 많아서 영교차율로만 세 발음을 구분하게 되어 세 발음간의 인식에 오류가 많았으며, [ㅅ], [ㅈ], [ㅊ]에서도 같은 결과로 나타났다. 따라서 각 세 발음을 하나의 발음으로 대표하여 여섯 개의 자음만으로 인식할 경우는 인식률이 88.6%로 높아지게 된다.

6. 결론

본 연구는 한국어의 음소별 인식에 관한 것으로서 필터뱅크 분석법과 영교차율 측정법을 사용하여 음성의 분석 및 인식 장치를 구성하고 일곱 개의 단모음과 열 개의 초성자음에 관하여 인식률을 조사하였으며 연구 결과를 요약하면 다음과 같다.

1) 선형예측법, 고속 퓨리에 변환법, 자기상관법 등에 비해 필터뱅크 분석법은 알고리즘이 간단하고 소프트웨어 실행 시간이 빨라 실시간 인식에 적합함을 알 수 있었다.
2) [아], [어], [이]는 평균 98%의 인식률을 보여 인식이 거의 완벽하였으며, [오], [우], [으]는 혼동이 있어 인식률이 90% 정도였다. 이것은 [오], [우], [으]의 주파수 스펙트럼의 형태가 유사하기 때문이라고 생각한다.
3) 단일 발성자(판별기준 작성자)와 임의 발성자 간의 인식률은 약 10%의 차이를 보였으며, 임의 발성자도 교정후에는 인식률이 증가하였다.
4) 초성자음의 인식결과는 단모음의 경우에 비해 낮았으며 이것은 자음의 인식 난이도가 높기 때문이라 생각한다.
5) 본 연구 결과를 통하여 한국어의 기계인식에 대한 가능성을 확인하였으며, 본 연구를 토대로 인식 장치를 세분화하면 음성의 분석에 대한 보다 상세한 정보를 얻을 수 있으리라 생각한다.

본 실험에서는 초성자음이 특정의 모음 앞에 올 경우를 기준하여 판별기준을 작성하여서 다른 모음이 이어질 경우에는 인식률이 다소 저하되는 것을 관찰할 수 있었다. 따라서 각 모음의 앞에 오는 자음의 변화를 세밀히 분석하여 보다 정확한 인식을 얻어야 하겠으며, 또한 유성음과 무성음간의 오인식이 많았는데 보다 안정된 구별방법이 연구되어야 함을 알았다. 대역통과 특성이 우수한 필터를 대규모 집적회로로 구성하여 보다 많은 채널로 세분하여 분석하면 보다 좋은 인식 결과를 얻을 수 있을 것이며, 다른 분석법과의 병렬처리도 가능하리라 생각된다. 또한, 발성자에 의존하지 않는 대용량 어휘의 인식을 위하여는 음절 또는 음소별 인식에 관한 연구가 많이 이루어져야 하겠으며, 아울러 이들 음소를 결합하여 구절로 변환하는 알고리즘의 개발이 필요하다고 생각한다. 본 연구결과는 한국어 음성으로 작동하는 각종 정보처리기계, 산업용기계 및 음성인식 로보트 등의 연구에 크게 기여하리라 생각한다.

참 고 문 헌

[1] Shuzo Saito and Kazuo Nakata, "Fundamentals of Speech Signal Processing," Academic Press, 1985.

[2] A. R. Smith and M. R. Sambur, "Hypothesizing and Words for Speech Recognition," Trends in Speech Recognition, Prentice Hall Inc., pp.139-165, 1980.

[3] Lawrence R. Rabiner and Stephen E. Levinson, "Isolated and connected word recognition - theory and selected applications," IEEE Trans. Communication, vol. COM-29, No. 5, pp.621-659, May 1981.

[4] George M. White and Richard B. Neely, "Speech Recognition Experiments with Linear Prediction, Bandpass Filtering, and Dynamic Programing," IEEE Trans. Accust., Speech, and Signal Process., vol. ASSP-24, pp.183-188, April 1976.

[5] George M. White, "Speech Recognition: A Tutorial Overview," Computer, vol. 9, pp.40-53, May 1976.

[6] R. L. Kashyap and Mahesh C. Mittal, "Recognition of Spoken Words and Phrases in Multitalker Environment Using Syntactic Methods," IEEE Trans. Computers, vol. C-27, No. 5, pp.442-451, May 1978.

[7] A. M. Derouault and B. Merialdo, "Natural Language Modeling for Phoneme-to-Text Transcription," IEEE Trans. Pattern Analysis and Mashine Intelligence, vol. PAMI-8, No. 6, pp.742-749, Nov. 1986.

[8] Ronald W. Schafer and Lawrence R. Rabiner, "Digital Representa- tions of Speech Signals," Proc. IEEE, vol. 63, pp. 662-677, Apr. 1975.

[9] S. R. Hyde, "Automatic Speech Recognition: A Critical Survey and Discussion of the Literature," Human Communication: A Unified View, McGraw Hill, New York, pp. 399-438, 1972.

[10] Bishnu S. Atal, "Automatic Recognition of Speakers from Their Voices," Proc. IEEE, vol. 64, pp. 460-475, Apr. 1976.

[11] L. C. Pols, "Real-Time Recognition of Spoken Words," IEEE Trans. Comput., vol. C-20, pp. 972-978, Sep. 1971.

[12] D. Raj Reddy, "Speech Recognition by Mashine: A review," proc. IEEE, vol. 64, pp. 501-531, Apr. 1976.

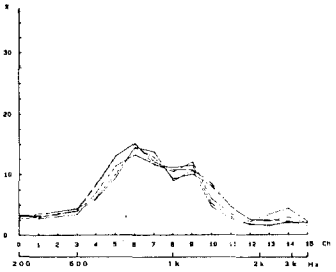
[13] 이용주, "한국어 단모음의 분석및 인식에 관한 고찰," 한국 전자통신 연구소, 전자통신, vol. 8, No. 1, Apr. 1986.

[14] 김중규, 안수길, "Vocal Tract의 Dynamic parameter에 의한 한글 단모음간의 Affinity에 관한 연구," 한국음향학회, 한일합동 음향학 학술발표회 논문집, pp. 108-113, 1981.

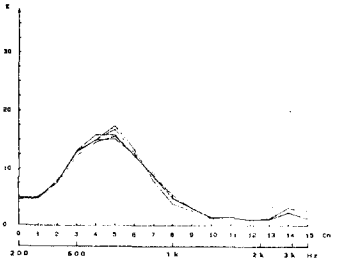
[15] 이병수, "한국어 음성의 분석및 규칙합성에 관한 연구," 건국대학교 대학원, 박사학위논문, 1987.

[16] 남문현, 전기회로와 신호, 자유아카데미, pp. 107-134, 331-362, 447-491, 1987.

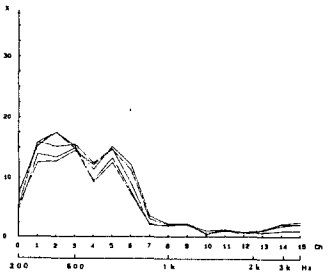
[17] 조 명한, 언어 심리학, 대우학술총서. 인문사회과학 17, 민음사, pp. 49-50, 1985.



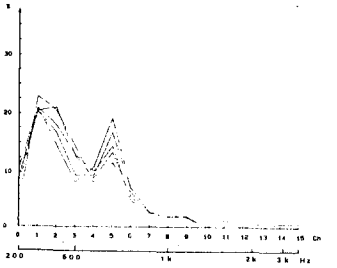
(가) 아 [a]의 분석결과



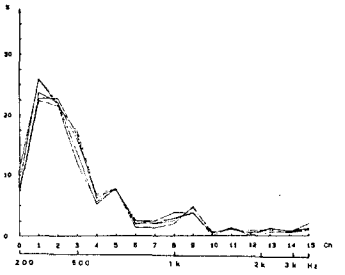
(나) 어 [ʌ]의 분석결과



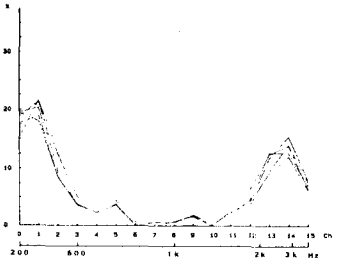
(다) 오 [o]의 분석결과



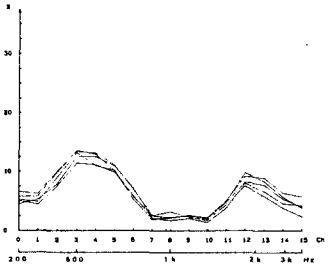
(다) 우 [u]의 분석결과



(바) 으 [ɨ]의 분석결과



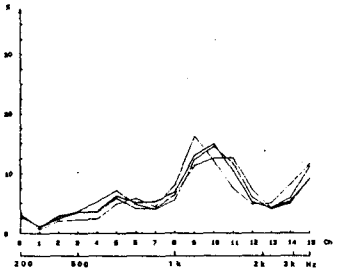
(바) 이 [i]의 분석결과



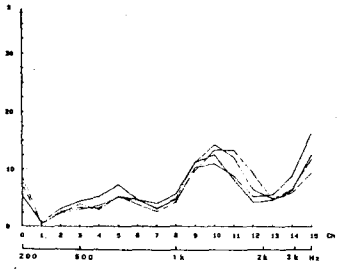
(사) 에 [e]의 분석결과

그림 5. 한국어 모음의 분석 결과

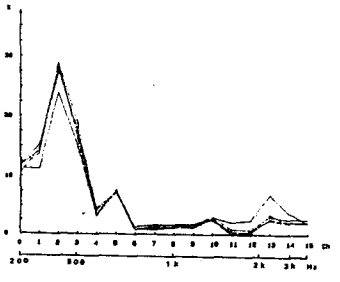
Fig. 5. Korean vowels analysis



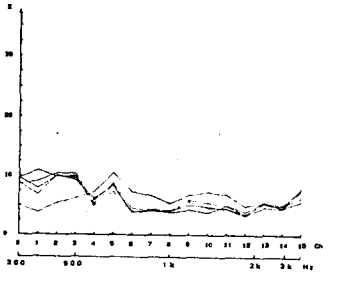
(가) '가'음의 초성 [g]



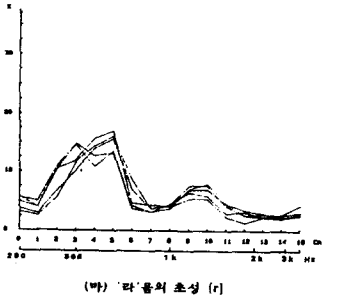
(나) '카'음의 초성 [k]



(다) '나'음의 초성 [n]



(라) '다'음의 초성 [d]



(바) '라'음의 초성 [r]

그림 6. 한국어 자음의 분석 결과

Fig. 6. Korean consonant analysis

표 2. 단모음의 인식결과

Table 2. Vowels recognition rate

(가) 단일 발성자 (50회 발음)

인식 결과 →

	아	어	오	우	으	이	에	인식률 (%)
아 'a'	50							100
어 'ə'		50						100
오 'o'			46	1	3			92
우 'u'				47	3			94
으 'ɨ'					50			100
이 'i'						50		100
에 'e'							50	100
↓ 발음							평균	98.0

(나) 임의 발성자 (5명이 10회)

인식 결과 →

	아	어	오	우	으	이	에	인식률 (%)
아 'a'	50							100
어 'ə'	4	45		1				90
오 'o'		1	38	5	4	2		76
우 'u'			5	39	6			78
으 'ɨ'			1	4	45			90
이 'i'				1		49		98
에 'e'		5					45	90
↓ 발음							평균	88.9

(다) 교정 후의 발성자 (5명이 10회)

인식 결과 →

	아	어	오	우	으	이	에	인식률 (%)
아 'a'	50							100
어 'ə'		48		2				96
오 'o'			45	2	3			90
우 'u'			5	43	2			86
으 'ɨ'			1	3	46			92
이 'i'				1		49		98
에 'e'		5					45	90
↓ 발음							평균	93.1

표 3. 초성자음의 인식결과(단일 발성자 50회)

Table 3. Consonant recognotion rate

인식 결과 →

	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅅ	ㅈ	ㅊ	ㅋ	ㅎ	인식률 (%)
ㄱ 'g'	35		1			1			3	10	70
ㄴ 'n'		49	1								98
ㄷ 'd'	4		28	4	14						56
ㄹ 'r'			4	46							92
ㅁ 'b'	1		4	5	39		1				78
ㅅ 's'			4			41	5				82
ㅈ 'c'		1				8	38	2		1	76
ㅊ 'c''							12	35	3		70
ㅋ 'k'	16								28	6	56
ㅎ 'h'	9		7			1				33	66
↓ 발음										평균	74.4