

필터뱅크를 이용한 한국어 숫자음 인식에 관한 연구

김홍식 한득영
건국대학교

A Study on the Recognition of Korean Digits using Filter-Bank

Hong-Sik Kim Deuk-Young Han
Kon-Kuk University

ABSTRACT

This paper is concentrated on the recognition of Korean Digits. The speech signals of each of digits are fed into computer through the 18 bandpass filters, AD converter. Spectrum input data are analyzed and used. BASIC program language is used for recognition performance and the result of recognition is outputted to computer screen and printer.

In this paper, the strength and weakness of filter-bank analysis method is described and the technique of real-time recognition is argued. In this experiment, Ratio of recognition for speaker dependent recognition was about 97% and recognition time was also satisfied. Therefore, A way of speaker independent recognition will be presented and using for special communication in the future.

1. 서론

음성정보처리를 위한 하드웨어 및 소프트웨어 기술이 진보하여 기계가 직접 사람의 말을 알아들을 수 있게 하는 것을 음성인식이라 한다. 그 오랜 옛날부터 사람이 기계사이 상호간의 필요성을 깨닫는 느낌에 따라 편리한 기기를 만들어 왔으나 상호 대화에 있어서는 화자의 감성이나 솔직한 표현인 음성을 완벽하게 도입시키지 못하였다.[1]

18세기초 인간의 음성 메카니즘을 모델화 하려는 시도가 있었고, 그 후반기에 Kratzenstein, Von Kempelen 등이 풍무를 이용하여 기계적으로 합성하려는 시기도 하였다.[1][13] 그 후 컴퓨터가 점점 발전됨에 따라 인간과 기계 사이의 대화 매체로 음성을 이용하고자 하는 필요성이 점차 늘어나, 이를 충족시키기 위한 노력으로 1950년 이후 음성의 자동인식에 관한 연구가 시작되었다.[14] 그리고 1960년대 중반에 디지털 신호처리 기술이 음성신호 처리에 응용되면서 인식 연구분야의 새로운 토대가 마련되어 70~80년대에 이르기까지 많은 음성인식의 연구가 진행되고 있다.[5][16]

또한 음성인식은 여러가지 상업적 시스템에 이용되고 있는데, 이러한 상업적 시스템은 본질적으로 있어서 많이 발전되어 있고 소음에서도 발생자의 혼동에 따라서 95% 넘는 인식율을 보이고 있다. 음성인식의 고전적 형태는 단어인식이고, 컴퓨터 과학의 발전에 따라 문장단어를 인식하고 단어를 연속적으로 간단히 하는 능력을 주어서 결합할 수 있게 하였다.[5]-[7]

음성인식은 세가지로 나누어져 연구되고 있다. 그 첫째분야는 격리 단어의 인식으로서 명료하게 발음하는 단어를 인식하는데 현재 여러가지 시스템들이 실험화 되어 있으나 격리단어 인식용 위해서는 각 단어 사이를 구분하여 발음해야 하므로 실제 사용에는 어려움이 있다. 두번째분야는 연결단어 인식으로서 이 인식 시스템은 좀더 자동스럽게 발음하는 연결단어를 다루므로 격리단어 인식시스템의 단점을 어느정도 보완할 수 있다. 세번째분야는 연속음성 인식으로서 이 경우는 자연스럽게 발음하는 문장 구절을 인식하는 것이다.

지금까지의 데이터 처리 방법으로는 DFT변환, FFT변환, Cepstrum 분석, 선형예측분석(Linear Prediction Analysis)등의 디지털 처리 방법들을 많이 이용하였으나, 메모리 용량과 실시간 처리에 있어서는 많은 문제점들이 노출되고 있다.

본 논문은 한국어 숫자음의 실시간 분석 및 인식에 관한 연구로서 필터뱅크를 이용하여 한국어 숫자음 /중/ 부대 /구/ 자치를 분석 및 인식 하였고 사람과 기계와의 대화가 용이 해 지도록 하여 전화의 위치를 만들 수 부가시키는 데 의의가 있다.

- 음성 다이얼링의 장점으로는 다음과 같은 사항을 하에서 다이얼링이 가능하다.
- (1) 손과 눈이 자유롭기 때문에 다른 작업을 하면서 다이얼링이 가능하다.
- (2) 눈이 잘 보이지 않는 사람도 자유로운 다이얼링이 가능하다.
- (3) Key입력보다 2배 이상 빠르게 입력된다.
- (4) 사람이름, 회사이름, 서어비스 이름등에 다이얼링이 가능하도록 하는 경우 전화번호의 기억이나 전화번호부의 조치가 필요 없다.
- (5) 전화기에서 다이얼링 기구를 제거 함으로써 소형화를 도모할 수 있다.
- (6) 위키 기술을 확립 함으로써 전화기로부터 음성에 의한 정보 입력을 가능하게 하여 각종 전화 서어비스의 확대를 도모할 수 있다.

(7) 또한, FA(Factory Automation), OA(Office Automation)는 물론 HA(House Automation)에도 응용이 가능하다. 이러한 자동음성다이얼링은 기본적으로 전화통신망에 설치되기 때문에 여러가지 많은 문제점을 내포하게 되어 그 실연에 어려움이 뒤따르고 있는데 특히 음성인식 기술이 중요한 단점이 된다.

II. 음성의 분석

II-1. 음성의 생성

음성의 물리적인 표현인 음성파형은 인간의 발성기관에 의하여 발생하는데 발성기관의 구조는 성대(vocal cord)와 성도(vocal tract)로 구분되고, 음성발생과정은 음원의 발생, 발음, 소리의 방사 등 3분야로 구성되어 있다. 음성은 폐에서 나오는 공기의 압력에 의해 성대에서 일정한 진폭으로 단속되어 만들어진 기본음이 성문(groties)에서 입술에 이르는 성도의 여러 곳에서 산란되어 만들어진다. 이와같은 구조에서 만들어진 음성은 유성음과 무성음의 두 종류로 크게 구분할 수 있다.

유성음은 림스 발생기의 출력을 성도필터를 통했음 때의 림스 줄음으로서, 무성음은 백성갑음 발생기의 출력을 성도필터에 통했음 때 일어나는 필터 출력으로서 각각 모델화 할 수 있다.[15]

II-2. 음성의 분석방법

음성을 인식하려면 음성파형으로부터 인식에 유리한 물리적 매개변수인 음성 파라미터를 추출해 내는 음성신호의 분석이 최 우선 과제이다. 음성신호에 있어서, 격리단어를 인식하기 위한 시스템은 일반적으로 크게 특징추출, 형태의 유사성, 결정방법의 세가지 방법을 거친다. 특징을 추출하기 이전에 격리된 음성분석을 위하여 전처리(preprocessing)과정을 수반해야 하며, 인식을 하기위해서는 가장 기본적인 것으로 첫번째 매우 복잡한 입력 음성신호 데이터의 특징을 포함하는 기준패턴을 추출하여 컴퓨터에 저장하는 것이고, 두번째는 입력된 음성신호 패턴이 미리 저장된 기준패턴의 어느 것과 유사한지를 비교하고, 세번째는 이들 결과로부터 어떤 단어로 결정할 것인가를 선택하는 것이다.

음성 데이터의 판별법은 형태 일치법과 판별규칙에 의한 방법 등으로 나눌 수 있으며, 본 논문의 실험에서는 형태 일치법을 사용하여 인식을 실행하였다.

음성신호의 특징은 시간에 따라 비교의 천천히 변화한다는 사실이다. 이러한 사실을 음성신호의 짧은 부분이 격리되어 진행되는 다양한 단시간(short time) 처리방법을 유도해 낼 수 있는데, 대부분의 단시간 처리기술은 다음과 같은 수식으로 나타낼 수 있다.[4]

$$Q_n = \sum_{m=0}^{n-1} T(x(m)) w(n-m) \quad (1)$$

여기서, T는 계승을 나타내고

$$W(n) = 1, \quad 0 \leq n \leq N-1$$

$$= 0, \quad \text{otherwise 이다.}$$

TX(n)은 요구되는 주파수 대역을 분리하기 위해 선행필터를 통한 후의 음성신호 변환이다. 또한 Qn은 TX(n)을 갖는 계승의 형태로 임펄스 응답 h(n)=w(n)을 가진 선형 시불변 시스템의 출력으로 해석될 수 있다.

음성파형으로부터 인식에 유효한 음성 파라미터를 추출하는 방법, 즉 음성 분석방법에 대하여 간단히 언급하기로 한다.

II-2-1. 시간영역에서의 분석방법

음성신호를 디지털 형태로 가능한 한 정확하게 부호화 하여 재구성하여 음성모형의 매개변수와 특성을 일정 구간 내에서 표현할 수 있으며, 음성파형의 성질이 본질 사이에서 변화하기 때문에 일반적으로 균등하게 시간 구간에 놓여진 음성의 짧은 본질을 해석함으로써 시변상의 음성을 분석할 수 있다.

- (1) 최대치 측정
유성음의 음성신호는 기본주파수에서 주기적으로 발생하는 최대치의 연속에 의해 특징 지어진다. 분석은 최대치기 비교적 작으며, 식별할 수 있는 모델로 나타내지 않는다. 따라서 무성음과 유성음 본질의 구분이 가능한 반면 짧은 구간에서 음성신호는 정확히 주기적이지 않아서 다른 해석이 나올 수 있는 단점이 있다.
- (2) 에너지 측정
단시간 내에서의 에너지로 구해 유성, 무성음 분석 등에 이용한다. 시불변 시간신호 x(m)에 대한 이산시간 신호의 에너지는 다음과 같이 정의되고,[4]

$$E = \sum_{m=0}^{n-1} x^2(m) \quad (2)$$

단시간 에너지는 다음과 같이 간단히 나타낼 수 있다.

$$E_n = \sum_{m=n-N+1}^n x^2(m) \quad (3)$$

음성에 있어서 무성음 부분의 진폭 크기는 일반적으로 유성음에 비해 매우 작다. 단시간 에너지는 이러한 진폭변화를 고려할 수가 없으며, 일반적으로 음성의 단시간 에너지는 다음과 같이 정의할 수가 있다.

$$Q_n = \sum_{m=n}^{\infty} [x(m) w(n-m)]^2 \quad (4)$$

$$= \sum_{m=n}^{\infty} x^2(m) h(n-m)$$

여기서, $h(n)=w^2(n)$ 이다. 단시간 에너지 E_n 의 중요한 성질은 유성, 무성음 부분을 뚜렷하게 구분하는 것이고, 이 에너지 함수를 유성음 무성음으로 되는 시간을 알 수 있고, 매우 높은 고음에 있어서 침묵(silence)을 구별하는데 사용할 수 있다. 식 (4)의 단시간 에너지 함수가 가지는 특징은 제곱형태로 계산하기 때문에 큰 신호 준위에 너무 민감하다는 것이다. 이러한 문제점을 보완하기 위한 간단한 방법으로 평균치 함수를 정의하면 다음과 같다.

$$M_n = \sum_{m=n}^{\infty} |x(m)| w(n-m) \quad (5)$$

여기서, 음성신호의 절대치 합은 제곱의 합 대신에 계산된다. 에너지와 평균치 함수의 대역폭이 지역성과 필터의 대역폭이므로 이러한 함수들을 음성신호처리 주기적으로 샘플링 할 필요가 없다는 것이고, 표준 필터 설계 방법 중 하나를 사용해서 지역성과 필터를 설계하면 된다.

(3) 영고차 측정

음성신호의 부호만 남겨 진폭을 1 bit로 양자화 한 것으로, 수리적으로 해석이나 처리 테스트의 결과, 음성정보를 잘 보호하고 있는 것으로 확인되었다. 스펙트럼 영역에서의 무제한 주파수 성분 혹은 대역이 잘 되므로 적은 스펙트럼 스퀴어법의 대역적인 특징을 표현하는 데 적합하다. 대부분의 음성이 넓은 주파수 스펙트럼을 갖기 때문에 음성에 영고차 측정의 해석은 매우 정확하지는 않다. 그럼에도 불구하고 마찰음 스펙트럼은 3KHz 이상에서 집중되는 데 반하여 유성음 에너지는 3KHz 이하에서 집중되는 경향이 있어, 영고차 측정은 음성의 특별한 본질(segment)이 유성음인지 무성음인지를 결정하는 데 자주 사용된다.

(4) 단시간 자기상관 분석

불연속 시간신호 $x(n)$ 의 자기상관함수는 다음과 같이 정의한다.

$$\Phi(m) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n)x(n+m) \quad (6)$$

자기상관함수는 어떠한 과정에서 구성을 나타내 데 유용하며 음성에서도 예외는 없다. 또한 자기상관함수는 $m=0$ 인 부분에서 날카롭게 피크를 나타내며, m 이 증가함에 따라 급격히 떨어진다. 음성은 한 정적 신호가 아니나 음성신호의 특성은 비교적 오랜 시간동안 적당히 남아 있다. 음성신호의 짧은 본질에서 동작하는 단시간 자기상관함수는 아래의 식들에 대해서

$$\Phi(m) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n+m) \quad 0 \leq m \leq N-1 \quad (7)$$

으로 정의된다. 또 원래신호의 기본주기 만큼 지연된 경우 현저한 피크를 보여 기본 주파수의 검출에 널리 사용된다.

II-2-2. 주파수 영역에서의 분석방법

인어목서의 음성 특징은 주로 성도(vocal tract)의 공진계에 의존하는 바 크며, 공진의 표현은 음성 스펙트럼이나 성도의 주파수 전달함수가 보다 명확하다. 또한 인간의 청각에서의 음성처리는 주파수 분석에 바탕을 둔다는 사실을 감안하면 스펙트럼 분석의 중요성을 더 잘 알 수 있다. [13]

(1) bandpass filter bank에 의한 분석방법

음성인식을 위해 사용하는 가장 보편적인 방법 중의 하나로, 분석하는 주파수의 범위를 인접하는 여러개의 대역필터로 얻을 수 있는 것이 편리한 것으로, 실시간에 음성의 대역적인 단시간 스펙트럼 파형을 얻을 수 있는 장점과 스펙트럼의 형태가 음성의 특성을 잘 나타내어 현재에도 음성인식에서의 용도가 많다. [4][6][10]-[12]

주파수 분석에는 광대역 분석과 협대역 분석으로 나눌 수 있는데, 어떤 주파수 분석에서든 대역폭 B와 분석시간 T는 다음과 같은 조건을 만족시켜야 한다. [16]

$$BT > 1 \quad (8)$$

따라서 대역폭이 좁을수록 분석시간은 증가하며 이러한 이유로 협대역 분석은 음향기기의 성능시험이나 기계진동 분석과 같이 신호 내용이 시간에 따라 거의 변하지 않는 정상신호를 대상으로 할 때 주로 이용된다.

(2) Fourier 변환

이산적 Fourier 변환(Discrete Fourier Transform : DFT)과 고속 Fourier 변환(Fast Fourier Transform : FFT)은 시간영역 음성신호를 주파수 영역의 스펙트럼으로 변환하는 가장 일반적인 수법이라 할 수 있다.

II-2-3. Cepstrum 분석

본래는 지진파의 해석에 쓰이던 수법으로 음원의 기본주기 검출에 사용된 이래, 주요한 음성분석법이 되어 음성연구 전반에 널리 쓰이고 있다.

II-2-4. 선형예측분석(LPC분석)

음성생성모델의 파라미터를 음성파형으로부터 직접 구하는 방법으로 제시된 이래, 음성의 분석 합성 및 인식용 거의 전반에 걸쳐 현대 음성분석 수법 중 가장 유력하고 광범위하게 이용되고 있다. 선형예측계수, PARCOR분석에 의한 반사계수(부반사계수), 최근의 선스펙트럼쌍(Line Spectrum Pair : LSP) 등의 파라미터가 모두 LPC분석에 기반을 두고 있다.

II-3. 한국어 숫자음의 표준음성 및 분석

II-3-1. 숫자음의 표준음성

본 논문은 한국어 숫자음의 음성인식인데 인식을 위한 설정표준음성은 다음과 같다.

분류	음성	데이터
모음 (Vowels)	[]	일, 이, 칠
	[]	삼, 사, 팔
	[...]	오, 궁
	[...]	육
자음 (Consonent)	[...]	구
	[...]	일, 칠, 팔
	[...]	삼
	[...]	삼, 사
	[...]	칠
[...]	팔	
[...]	구, 궁	

II-3-2. 숫자음의 분석

/궁/ 부터 /구/ 까지의 발성상의 특징을 보면 다음과 같다. 예비본류로서의 유성음과 무성음 판별에 도움을 주고 10개의 음성입력 데이터들 가운데 총성 부분에서 /S/를 제외한 나머지 유성자음(/일/, /삼/, /칠/, /팔/, /궁/)과 모음(/이/, /사/, /오/, /구/)이라는 사실도 음성분석 처리에 있어 끝점검출(End-Point Detection) 알고리즘 설계에 유용하다. 그 외에 다음과 같은 사실이 있다.

- (1) 한국어 단독 숫자음은 단음절이다.
- (2) 단독 숫자음의 초성은 자음이 있어서 무성자음 /ʃ/, /s/, /ʃ/, /s/ / /이다.
- (3) 단독 숫자음의 종성은 모음 / / 와 비음 / m, /, / / 과 같은 유성자음이다.
- (4) 단독 숫자음 / 육 / 의 종성 / ʃ / 은 거의 묵음에 가깝다.

II-4. 음성인식 순서

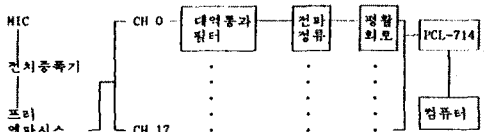
음성인식시스템은 기본적으로 3 단계를 거친다. 1단계에서는 데이터를 축소시키는 단계로서 음성신호의 특징을 가장적절 변경시키는 특징추출단계이다. 그 다음 단계는 시험형태와 기준형태 사이 시간변별 사이의 형태 유사성과 시간절차를 정의한다. 마지막 단계는 시험형태와 기준형태를 연결시켜 선택하는 방법이다.

음성신호 → 특징추출 → 형태유사성 → 결정방법 → 결과출력



III. 음성인식 시스템의 설계

본 논문에서 실험한 음성인식 시스템의 구성도는 그림 3-1과 같다.



<그림 3-1 음성인식 시스템 구성도>

III-1. 음성입력부

음성입력부에서는 아날로그 데이터가 처리되며 잡음의 영향을 방지하기 위해 비선형 가폭적 필터가 있다. 또 마이크에서 이 프리엠파시판으로 들어오는 신호선 및 이 프리엠파시판에서 나가는 신호선은 절도선을 사용하고 있다. 이러한 종류의 회로는 비교적 잡음을 약하므로 각 오퍼엠파시의 전진핀에 본연의 잡음에 잡음을 어느정도 방지할 수 있다. 일반적으로 마이크의 출력은 매우 작으므로 음성입력을 크게 증폭할 필요가 있다. 엠파시에서는 1단의 증폭회로로 몇백배 되는 증폭률도 취할 수 있으나 잡음에 대한 것 중도 고려하여 본 실험에서는 3단으로 나누어 증폭한다. 또 4단계의 오퍼엠파시는 고역을 강조하는 필터이다. 사람의 음성은 젊은 입 안에서 갑자기 넓은 공간으로 나올 때 높은 주파수 성분이 많이 야해져 버려 고역을 어느정도 강조해 두는 편이 인식을 실험할 때 편리하므로 이러한 방법을 채택했다.

III-2. 필터뱅크부

본 연구에서는 음성신호를 실시간으로 처리하기 위한 1단계 처리방법으로 대역통과 필터로 이루어진 필터뱅크를 사용했다. 필터뱅크는 18개의 필터 회로로 되어있다.

각 필터는 Ch0, Ch1, ..., Ch17 로 되어 있으며 가장 낮은 주파수를 통과 하는 필터 중심주파수를 250 Hz 까지 하고있다. 그후 1/4 옥타브씩 중심주파수를 증가 시키고 4790 Hz 까지 중심 주파수를 변화 시키고 있다. 각 대역통과 필터의 특징표는 다음과 같다.

채널	중심 주파수(Hz)	대역폭(BW)	선택도(Q)
0	250	50	5
1	297	59.4	5
2	351	70.2	5
3	425	85	5
4	503	100.6	5
5	596	119.2	5
6	704	140.8	5
7	832	166.4	5
8	1010	202	5

9	1200	240	5
10	1410	282	5
11	1670	334	5
12	2020	404	5
13	2390	478	5
14	2830	566	5
15	3350	670	5
16	3960	792	5
17	4790	958	5

각 대역폭과 필터 회로의 각 저항 R 값과 커패시터 C 값들을 구하기 위하여 우선 대역폭 B 와 공진 주파수 ω_c 를 선정하고 $Q = \omega_c / B$ 로서 선택된 Q 를 구한다. 설계를 간단히 하기 위하여 $C_1 = C_2 = C$ 를 선정 한 후 다음과 같이 R_1, R_2, R_3 값을 구한다.

$$R_2 = \frac{2}{BC}$$

$$R_1 = \frac{R_2}{2A\omega_c}$$

$$R_3 = \frac{R_2}{4Q^2 - 2A\omega_c}$$

여기서, R_3 는 양수, A는 공진 주파수에서 최대전압 이득이며, $4Q^2 > 2A\omega_c$ 이어야 한다.

III-3. 소프트웨어 구성
프로그램은 크게 두부분으로 구성되어 있는데 음성신호의 각 스펙트럼을 컴퓨터로 읽어들이는 베이직프로그램과 입력된 음성신호 데이터를 필터 일치부에 의하여 처리하는 인식프로그램으로 구성되어 있다. 결정방법은 마이크를 통하여 입력된 단어와 기준 단어 사이의 차이로 입력된 단어를 인식하는 형태인식법을 사용 하였다.

IV. 음성인식 실험

본 논문의 음성인식 실험에서는 시작점음성과 묵음의 식별을 위해 프래기 프로그램을 만들었으며 실험시에는 잡음과 소음이 가장 적은 때를 골라 인식실험을 하였다. 인식방법으로는 입력된 음성신호를 각 채널별 백준주파수 크기로 단독 숫자들의 데이터로 입력시킨후 최자가 반응하는 음성의 데이터와 비교하여 가장 형태가 일치하는 것을 판별하여 인식 하였다.

여기에서 사용한 음성데이터는 각 채널별의 스펙트럼 크기로 크게 발음 형태와 작게 발음 형태의 변화를 고려하여 각 채널별 상대백준을 주파수 크기로 구했다. 구해진 음성데이터를 가지고 먼저 입력된 기준 패턴과 비교하여 그 차이가 가장 작은것을 골라 인식을 하였다.

V. 결과 및 결론

본 실험은 각 채널의 출력값을 절대치로 이용하지 않고 전체 출력값중의 구성비로된 상대적인 비율로 해석을 일치함으로써 소리의 크기와 잡음에 큰 영향을 받지않는 시스템을 구성 할수 있었다. 음성신호를 주파수 스펙트럼으로 변환하는 대역폭과 필터를 채널화하여 복잡한 분석 과정을 피할수 있어서 메모리 용량문제나 실시간 처리문제에 적절하게 대처할수 있었다.

인식결과 ->

	1	2	3	4	5	6	7	8	9	0	인식율(%)
1	47	1					2			1	92
2		50									100
3			45	5							90
4			3	47							94
5					49				1		98
6						50					100
7	3						47				94
8								50			100
9									50		100
0										50	100
										평균	97

인식실험 결과를 살펴보면 기준데이터 작성에 관한 발생자에 대한 단음절 숫자음 인식율은 98.0%였으며 이는 변화자 독립적인 방법이기 때문에 여러사람을 대상으로 각각의 음성데이터를 아주쉽게 입력시켜 인식 실험을 하였다.

VI. 결론

본 연구는 주파수 영역의 필터 뱅크 분석법을 이용하여 한국어 숫자음 분석 및 인식장치를 설계,구성하고 화자독립 인식의 실현을 도모하였다.본 연구의 결과를 보면, 한국어 숫자음의 화자독립인식 및 자동음성인식인식의 가능성을 기대할수있으며 이러한 방법의 연구결과 확장한다면 인식율을 더욱 높일수 있으리라 생각된다.또한 다만 대역폭과 필터이론계, 화자독립이론계 이 발생되는 잡음제거 프로그램 기술의 보완을 한다면 필터이론에서 연결 단어보의 아주 이상적인 음성인식시스템과 자동음성인식시스템의 개발이 가능할수 있을것이다.특히 그 활용성을 살펴보면 자동정보처리 기기, 산업용, 완구용기기와 음성인식로봇등이 있으며 특히 가장 많이 사용되는 분야는 Home Automation라 생각된다.

필터의 설계시 대역폭과 필터로 연산중폭기를 사용하고 그 특성은 비교적 만족할만하였으나 대역폭과특성이 우수한 필터를 대용모집회로도 구성할경우 보다 많은 채널로 세분하여 분석하면 더욱좋은 인식결과를 얻을수 있을것이다.

음성인식기술의 발전을 기대하려면 우선과제가 음성신호처리 기기의 개발과 신호처리 알고리즘을 개발하는 일일것이다.

우리들이 지향하는 사람과 기계간의 음성에의한 자유로운 정보교환이 이루어지기 위해서는 음성인식기술을 보다 발전시켜, 연결단어 발음시 발음 현상과 인식하고자하는 어휘수가 증가할경우 급격히 늘어나는 계산량을 효율적으로 처리할수있는 연결단어 인식시스템에 대해서도 연구하여 음성통신시스템 개발에 박차를 가해야 할것이다.

참고문헌

- [1] Shuzo Saito and Kazuo Nakata, "Fundamentals of Speech Signal Processing", Academic Press, 1985.
- [2] Geoff Bristow, "Electronic speech Recognition", Collins, pp49-129, 1986.
- [3] S. R. hyde, "Automatic Speech Recognition : A Critical Survey and Discussion of the Literature", Human Communication : A Unified View, McGraw Hill, New York, pp.339-438, 1972.
- [4] Ronald W. Schafer and Lawrence R. Rabiner, "Digital Representations of Speech Signals", Proc. IEEE, Vol.63, pp.662-677, Apr. 1975.
- [5] A. R. Smith and M. R. Sambur, "Hypothesizing and Words for Speech Recognition", Trends in Speech Recognition, Prentice Hall inc., pp.139-165, 1980.
- [6] Lawrence R. Rabiner and Stephen E. Levinson, "Isolated and connected word recognition - theory and selected application", IEEE Tran. Communication, Vol. COM-29, no.5, pp.621-659, May, 1981.
- [7] 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, Apr. 1976.
- [8] R. B. Randall, Frequency Analysis, B & k, 1977.
- [9] C. K. Un, "Trends in digital speech communication technologies", Review of KIEE, Vol.11, pp.30-40, Dec. 1984.
- [10] Bishnu S. Atal, "Automatic Recognition of Speakers from Their Voice", 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 Machine : A review", proc. IEEE, Vol.64, pp.501-531, Apr.1976.
- [13] 오 영환, "음성인식", 전자공학회지, Vol.12, No.5, Oct. 1985.
- [14] 정 유현, 이 의택, "음성 디지털 시스템 개발을 위한 예비실험", 한국 전자통신 연구소, 전자통신, Vol. 8, No.1, Apr.1986.
- [15] 이 병수, "한국어 음성의 분석 및 규칙화정에 관한연구", 건국대학교 대학원, 박사학위논문, 1987.
- [16] 김 태형, "한국어 숫자음의 실시간 인식에 관한 연구", 건국대학교 대학원, 석사학위논문, 1989