[4-9]

실시간 음성 특징 추출을 위한 필터 뱅크의 설계

조 동 완 • 홍 희 식 * 한 영 렬
한양대학교 전자통신공학과
* 동양공업전문대학 통신과

Filter Bank Design for Real-Time Speech Parameters Extraction

Cho, Dong Wan • Hong, Hee Sik * Han, Young Yearl
Dept. of Electronic Commun. Engineering, Hanyang Univ.
* Dept. of Commun., Dongyang Technical College

ABSTRACT

In this paper, 16ch. filter bank for real time speech parameters extraction in frequency domain analysis is designed and implemented. Experiments are made in isolated word recognition with changing a center frequency, bandwidth, number of channels and some analog processing modules.

It meeted with the best results that number of channels was 13, Q value was 8.5 and analysis time was 10 msec. Preemphasis resulted badly and the system of small number of channels had a comparatively high recognition rate.

Ⅰ. 서론

음성 통신(Speech Communication)은 사람과 사람간에 공간 채널에서 음성을 이용하여 정보를 전달하는 것이다. 이에 대하여 음성 통신의 송,수신자 중 일부 또는 전부를 전자장치로 바꾼 것을 전자 음성 통신 (Electronic Speech Communication) 이라 한다. 이때 송,수신자의 일부만 전자장치로 바꾼 것은 Man-Machine Communication이라고 부른다.[8] 현재 이러한 목적으로 연구되고 있는 분야는 음성 합성(Speech Synthesis), 음성 인식(Speech Recognition), 화자 인식(Speaker Recognition), 음성 부호화(Speech Coding)의 4가지로 구분된다. 이 4가지를 통틀어 음성 신호 처리(Speech Signal Processing)라 한다. 음성 인식은 기술적인 어려움 때문에 1970년대 부터 본격적인 연구가 시작되었지만 실용화가 시작된 것은 최근의 일이며 아직 해결해야 할 문제가 많다.

음성 인식은 수신자(청취자)를 대신하는 것인데 화자에 상관없이 음성 신호에서 공통된 의미 즉 운질(韻質) 정보를 파악하는 것이다. 음성 인식 시스템에는 음소의 음성학적 계층 구조에 의한 양자택일적인 판정을 반복하여 최종 인식 판정을 하는 형식과 그림 1과 같이 음성의 음향적 특징에 관한 표준 패턴을 미리 기억시킨 뒤에 이를 입력 음성과 비교하여 유사도가 큰 것을 인식 결과로 하는 형식이 있다.

표준 패턴을 사용한 인식 시스템은 일반적으로 인식율이 좋고 불특정의 발성자에 적응하도록 학습기능을 부가할 수 있어서 많은 연구가 되고 있다. 반면에 단어 단위로 인식을 할 경우에는 어휘수가 증가하면 표준 패턴의 기억용량 및 인식 판정에 필요한 계산량이 늘어나는 단점이 있다.

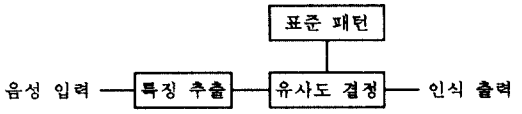

그림 1. 표준 패턴을 사용한 인식 시스템

현재의 음성 인식 연구는 표준 패턴과 음성 신호간의 단시간 스펙트럼 유사도나 거리(Distance)를 이용하는데 집중되고 있다. 단시간 스펙트럼 정보를 뽑아내는 방법에는 10-30 채널의 대역 필터 뱅크를 이용하는 것, FFT에 의해 직접 스펙트럼을 구하는 것, LPC 분석에 의한 것이 있다.

본 논문에서는 제한된 어휘의 격리단어 인식 시스템에 쓸 실시간 음성 파라미터를 추출하기 위해서 16채널 필터 뱅크 방식을 채택하여 실제 제작하고 실험, 분석하였다.

Ⅱ. 필터 뱅크의 설계

Ⅱ-1. 음성 신호의 분석

음성 신호에서 파라미터를 뽑아내는 방법에는 신호의 시간적 변화를 이용하는 시간 영역상의 분석법과 주파수 성분을 이용하는 주파수 영역상의 분석법이 있다. 시간 영역상의 분석에서는 선형 예측(Linear Prediction), 편자기상관(Partial Autocorrelation), Pitch 주기, 영교차율(Zero Crossing Rate) 등이 사용되고 주파수 영역상의 분석에는 단시간 스펙트럼, Formant(모음의 고조파 성분 중에서 스펙트럼의 상대적 크기가 큰 주파수), Cepstrum, 기본주파수 등이 사용된다.[8]

주파수 영역상의 분석 중에서 밴드 패스 필터 뱅크에 의

한 방법은 실시간 음성 파라미터 추출이 가능하며 비교적 값싸고 쉽게 하드웨어로 구현이 가능하므로 음성 인식 시스템의 실용화를 위해서 많이 연구되고 있다. 또 필터 뱅크 방식은 LPC에 비해 잡음에 강하고 산출된 스펙트럼의 모양이 실제 음성의 음운 특성을 잘 표시할 뿐 아니라 마찰음과 음성의 과도 특성(Transient Response)에 관한 정보까지 제공해 주는 장점이 있다.

필터 뱅크 분석 모델은 그림 2처럼 대체적으로 100Hz - 6KHz의 음성 대역을 수용할 수 있는 k개의 밴드 패스 필터와 정류기, 저역 통과 필터로 구성된다. 음성 신호는 대역 필터를 통과한 다음 정류기와 저역 필터를 거쳐 그 대역의 음성 신호의 에너지에 해당되는 출력으로 나타난다. 임의의 표본 시간 t_i에서 각 대역 필터로부터 얻어지는 출력 $x_1(t_i)$, $x_2(t_i)$, $\cdots$, $x_k(t_i)$는 t_i에서의 필터 뱅크의 특징 벡터 $X(t_i)$가 된다.[3,6]

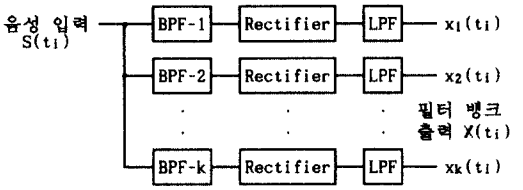

그림 2. 필터 뱅크의 구성도.

$$X(t_i) = \left[x_1(t_i),\ x_2(t_i),\ \cdots,\ x_k(t_i) \right] \quad \cdots\cdots\cdots \quad (1)$$

만일 특정의 음성 입력 신호가 t_1에서 t_M까지 존재한다면 특징 벡터 $X(t_i)$의 시계열(Time Sequence)은 특정의 음성 입력 신호에 대한 패턴 메트릭스 P를 형성한다.

$$P = \begin{bmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ \vdots \\ X(t_M) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots\cdots & x_k(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots\cdots & x_k(t_2) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x_1(t_M) & x_2(t_M) & \cdots\cdots & x_k(t_M) \end{bmatrix} \quad \cdots\cdots \quad (2)$$

음성의 주파수 특성을 이용한 인식 시스템에서는 패턴 매트릭스 P를 주파수 영역 파라미터로 삼아 멀티플렉서와 A/D 컨버터를 이용하여 디지탈 부호로 바꾸고 미리 저장된 표준 패턴 매트릭스와 비교하여 유사도를 산출한다. 이 유사도의 대소에 따라 인식 판정을 수행하는데 유사도를 계산하는 연산량은 많은 편이지만 그 정확도는 아주 우수한 것으로 입증되고 있다. 특히 신호 패턴과 표준 패턴 간의 발성 지속 시간의 차에 기인하는 부정합 문제를 동적 계획법(DP: Dynamic Programming) 방식으로 해결하는 DTW (Dynamic Time Warping) 앨고리듬이 개발된 뒤로는 그 정확도가 실용적인 수준에 이르게 되었다.[1,8]

본 논문에서는 SCF(Switched-Capacitor Filter)를 사용하여 16채널의 밴드 패스 필터 뱅크를 제작하고 그 앞단에 마이크로 받은 음성 신호를 증폭, 조절하는 전치 증폭부를, 뒷단에는 A/D 컨버터를 연결하여 그림 3과 같은 음성 인식 시스템을 구성하였다.

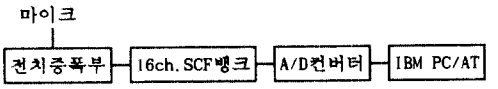

그림 3. 16채널 SCF 뱅크 음성 인식 시스템.

Ⅱ-2. SCF의 원리[5,7,11]

1970년대 후반부터 비상한 관심을 모은 SCF (Switched-Capacitor filter) IC는 대표적인 IC 필터로 1980년대 중반이 되어서야 제품이 생산되었다.

그림 4는 MOS FET와 커패시터로 된 스위치드-커패시터의 회로도와 동작원리를 나타낸 것이다. MOS FET는 V_{GS}가 스레숄드 전압(Threshold Voltage) V_T보다 크고 작음에 따라 R_{DS}가 작거나 매우 크게 된다. 즉 MOS FET는 V_{GS}에 의해 스위치의 역할을 하게 된다. 그러므로 그림 4 (a)의 회로에서 클럭 ϕ_1,ϕ_2를 (b)처럼 서로 역위상으로 가하면 MOS FET가 번갈아 On, Off 되므로 그림 (c)와 같이 된다.

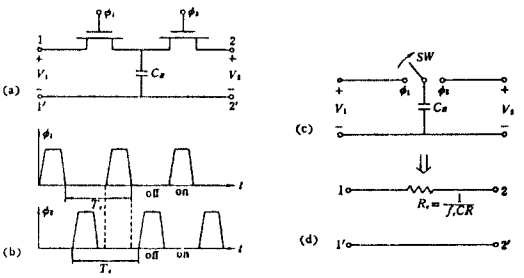

그림 4. 스위치드-커패시터의 원리

(c)에서 스위치가 일정 주기 T_c(클럭의 주기)로 왔다갔다하면, 왼쪽에 있을 때는 커패시터 C_R이 V_1까지 충전되고 오른쪽에 있을 때는 V_2까지 방전된다. 따라서 한주기 동안에 이동되는 전하 Q_c는

$$Q_c = C_R(V_1 - V_2) \quad \cdots\cdots\cdots\cdots\cdots \quad (3)$$

클럭의 주파수가 $f_c = \frac{1}{T_c}$ 이므로 T_c 동안에 흐르는 전류는 평균적으로

$$I(t) \equiv \frac{\Delta q}{\Delta t} \simeq \frac{C_R(V_1 - V_2)}{T_c} = \frac{V_1 - V_2}{\frac{1}{C_R f_c}} \quad \cdots \quad (4)$$

식 (4)는 $\frac{1}{C_R f_c}$ 에 해당하는 저항에 흐르는 전류를 나타내므로 그림 4 (c)의 회로는 등가적으로 (d)와 같이 된다. 즉 (a)의 회로는 $R_c = \frac{1}{f_c C_R}$ 인 등가 저항으로 표현될 수 있다. 이때 결국은 MOS FET 스위치가 시간적으로 샘플링 작용을 하는 것이기 때문에 샘플링 주파수인 f_c가 클수록 식 (4)의 근사식이 잘 맞게 된다. 실용상으로는 f_c가 신호의 최고 주파수보다 10배 이상은 되어야 한다. 본 논문에서 사용한 SCF는 f_c가 신호의 최고 주파수보다 50배, 100배 높게 선택할 수 있게 되어 있다.

능동 RC 필터에서는 R 또는 C를 조정해야 하고 비교적 큰 용량의 C가 쓰인다. 그러나 SCF IC에서는 R의 값을 매우 크게 할 수 있으므로 C의 값은 수 pF정도면 된다. 반도체 기술로 MOS IC에서 수 pF의 커패시터를 정확히 만드는

것이 가능하고 등가저항 Rc는 클럭 주파수 fc에의해서 정확히 가변되기 때문에 정밀한 가변 필터를 만드는 것이 가능하다.

반면에 MOS FET가 스위칭을 함에 따라 필연적으로 스위칭 노이즈가 발생하므로 보통의 모놀리딕 SCF IC에서는 S/N비가 80 - 90dB로 제한된다. 그러므로 고품질의 오디오 신호나 소신호 입력에는 적합하지 않다. 또 샘플링 정리에 따라 이론적으로는 fc/2까지의 신호를 다룰 수 있지만, Aliasing을 피하고 등가저항 Rc의 오차를 줄이기 위해서 실제로는 fc/10까지로 취급 신호의 최대 주파수가 제한되므로 고역 특성이 제한을 받는다. 따라서 SCF 필터는 주로 오디오 대역에서 사용된다.

SCF는 몇가지 결점이 있지만 오디오 대역 특히 음성 대역에서 요구되는 물리적 특성은 충분히 커버할 수 있어서 CODEC(Pulse-Code Modulation Coder-Decoder), Tone Encoding과 Decoding, Echo Cancellation, 음성 인식, 음성 합성 등에 주로 사용된다.

Ⅱ-3. 전치증폭부

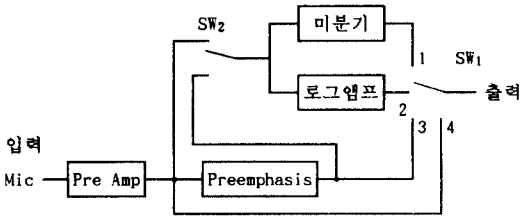

그림 5. 전치 증폭부.

그림 5는 마이크로 받은 음성 신호를 증폭하고 몇가지 아날로그 신호 처리를 하기 위해서 본 논문에 사용한 전치 증폭부이다. 실험에는 가청주파수(20~20,000Hz) 범위에서 평탄한 주파수 특성을 갖는 콘덴서 Mic를 사용하였다. 음성 신호의 dynamic range를 줄이고 단음절 명료도를 높이기위해 loudness 곡선을 고려한 Preemphasis 회로를 부가하여 특정 고음역을 boost 시켰다.또한 로그 앰프와 미분기를 병렬로 구성하여 적합한 인식 파라미터를 찾는데 도움이 되게하였다.

전치 증폭부에서 로그 앰프를 사용한 이유는 사람의 감각의 변화량은 자극의 변화율에 비례한다는 웨버-페흐너(Weber-Fechner)의 법칙에 기인한다. 즉 ΔR(감각의 변화량) $\propto \frac{\Delta S}{S}$ (자극의 변화율)가 된다. 따라서 미소감각의 변화량은 다음과 같게 된다.

$$dR = K\frac{dS}{S} \quad (K는 감각의 종류에 따른 비례 상수) \quad (5)$$

식 (5)를 적분하면

$$R = K\cdot\log S + C \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (6)$$

이때 적분상수 C는 감각 R이 0일 때의 자극 So(소리의 세기인 경우에는 최소가청한이 됨)가 되므로 식 (6)은

$$R = K\cdot\log\frac{S}{S_o} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (7)$$

즉 감각의 크기는 자극의 변화에 로그를 취한 값에 비례한다.

프리엠파시스 회로는 외이(Outer Ear)의 주파수 특성과 등청감곡선을 보정하기 위한 필터로 되어있다. 외이에서는 귓구멍 부분이 얇은 접시처럼 움푹 패어 있어서 5KHz 부근에서 완만한 공진특성을 가지며 외이도에서는 2.5-3KHz에 공진점을 갖는다. 외이도의 공진특성은 음의 입사각에는 무관하지만 귓구멍 부분은 상당한 영향을 받는다. 또 귓바퀴는 3KHz 이상의 소리에서 입사각에 따라 복잡하게 주파수 특성이 변동한다. 그 결과 고막에서는 2-6KHz의 대역에서 10dB 정도의 음압상승이 나타난다.[2] 외이 전체의 주파수 특성은 그림 6에 나타나있다.

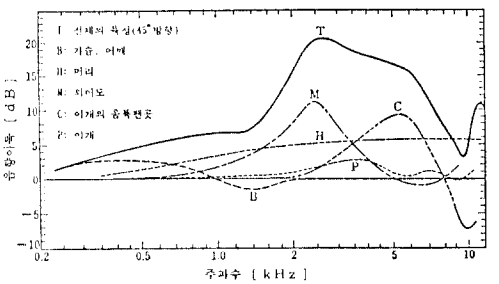

그림 6. 외이의 음향 특성 (Show, 1974)

사람이 특정 세기, 단일주파수의 순음만을 들었을 때 느끼는 소리의 크기를 실제 실험을 하여 그려놓은 것이 그림 7의 등청감곡선이다. 이것은 벨 연구소의 플레쳐와 먼슨이 최초로 작성하였으므로 플레쳐-먼슨(Fletche-Munson)곡선이라고도 부른다. 그림 7은 로빈슨과 닷손이 재작성한 등청감곡선이다.[9] 등청감곡선을 보면 그림 6의 외이의 음향 특성과 비슷하지만 10KHz 이상에서는 차이가 있음을 알 수 있다. 이로써 실제 청각의 주파수 특성은 외이에 큰 영향을 받지만 그외의 기관의 영향도 큼을 알 수 있다.

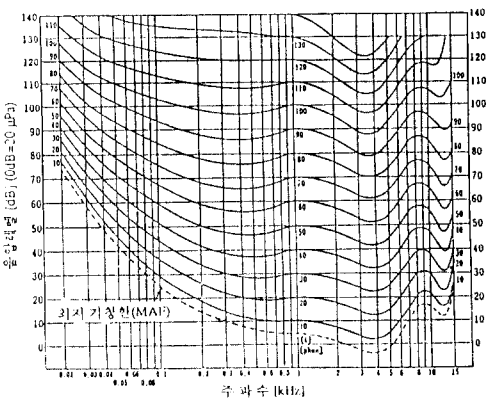

그림 7. 등청감곡선(Robison - Dadson)

음성 신호에서는 고조파 성분이 6KHz를 넘어가는 것은 별로 없고 6KHz 이하의 부분에서는 그림 6과 그림 7의 특성이 비슷한 것에 착안하여 3.5KHz에서 10dB, 400Hz에서 4dB의 공진 특성을 가지며 6KHz 이상에서 -18dB/oct, 200Hz 이하에서 -12dB/oct의 감쇄 특성을 갖는 필터를 구성하였다.

Ⅱ-4. SCF 뱅크부

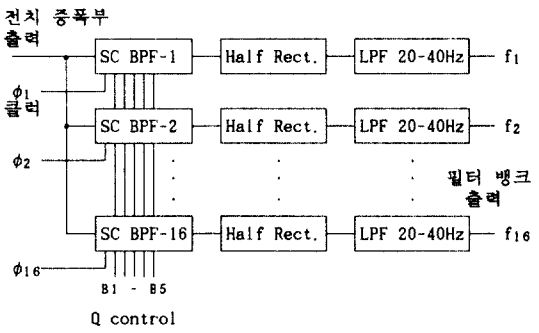

그림 8. 16채널 SCF 뱅크 구성도

그림 8은 각각 다른 센터 주파수를 갖는 16개의 4차 체비세프/버터워스 밴드 패스 필터이다. MF8(National Semiconductor)이라는 SCF IC를 16개 사용하고 190 - 6300Hz의 범위에서 센터 주파수를 1/3옥타브 간격으로 정하였다. 그 이유는 청감 특성상 1/3옥타브 이하의 주파수 변동은 잘 파악되지 않기 때문이다. 또한 Q값도 가변되게 하였다.

SC BPF는 클럭 주파수에 의해 센터 주파수가 임으로 변경될 수 있고 5비트로 된 2진 명령(B_1-B_5)에 의해 Q값을 0.45에서 90까지 조절할 수 있어 매우 편리하다.

SCF Bank의 출력을 각각 이상적인 반파 정류기와 LPF를 거치게 하면 16개 특정 주파수 대역에서 음성 신호의 포락선(Envelope)이 얻어진다.

표 1은 MF8로 만든 BPF의 설계 제원이다. 표 2는 1/3 옥타브 고정 간격 방식 체비세프 모우드에서 Q = 8.5일 때 각 채널의 중심 주파수와 대역폭을 정리한 것이다. 그림 9는 MF8의 결선도이다.

CH.	센터 주파수(Hz)	1/2 대역폭(Hz)
1	195	23
2	250	29
3	312	37
4	390	46
5	500	59
6	624	73
7	780	92
8	1K	118
9	1.25K	147
10	1.56K	184
11	2K	235
12	2.5K	294
13	3.12K	367
14	4K	471
15	5K	588
16	6.24K	734

표 2. 1/3 옥타브 고정간격의 중심 주파수와 대역폭 (체비세프, Q=8.5)

대상 주파수대	190 - 6300Hz
채널 수	16ch.
채널 간격	고정 간격 방식(1/3 Octave)과 임의 간격 방식 혼용
필터 방식	4차 Chebyshev/Butterworth Band pass Filter
필터 IC	Switched-Capacitor Filter
Q(Quality Factor)	임의 가변 (0.45 - 90)
채널간의 상대증폭도	임의 가변 (Frequency Weighting)

표 1. Filter Bank의 설계 제원

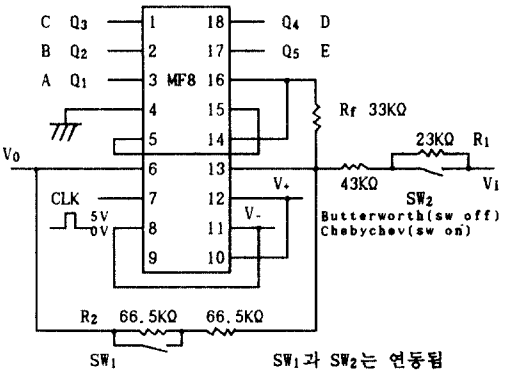

그림 9. MF8의 결선도

포락선을 검출하기 위한 LPF도 SCF(MF4 또는 MF6)를 사용하여 4차 이상의 급준한 필터로 만들었고 차단 주파수는 20-40Hz로 가변되게 하였다. LPF의 차단 주파수가 변경되면 A/D컨버터의 샘플링 주파수도 따라서 변경시켜야 한다. LPF의 차단 주파수를 아주 낮게 할 수 있는 근거는 10ms-30ms의 짧은 시간 동안에는 음성 신호의 변화가 매우 작기 때문이다.

Ⅱ-5. A/D 컨버터부

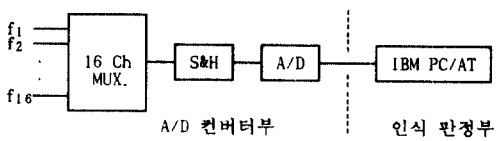

그림 10. A/D 컨버터부

필터 뱅크의 출력은 멀티플렉서와 S&H, A/D 컨버터를 거쳐 디지탈 형태의 패턴 메트릭스 데이터로 변환된다. 본 논문에서는 A/D 컨버터부는 DT2821-F-16SE라는 기존 제품을 사용하였다. 그 제원은 표 3과 같다.

Ⅲ. SCF 뱅크를 이용한 음성 인식 실험 및 분석

Ⅲ-1. 인식 알고리듬과 실험 방법

본 논문에서 사용된 인식 알고리듬은 그림 11과 같다. 2초 동안 받은 데이터 중에서 시작점은 신호의 진폭이 임계값을 연속하여 3번이상 넘을 때 그 첫번째 점으로 하며 끝점은 신호의 진폭이 임계값을 연속으로 10번 이상 넘지 못할 때 그 최초의 점으로 결정하도록 하였다. 그 다음에는

Analog 입력수	16개
입력 level	0 ~ 10 V (unipolor) -10 ~ 10 V (bipolor)
출력 data code	offset binary
Resolution	12 bits
A/D conversion time	2.5 μS
A/D Throughput	250 KHz

표 3. A/D converter DT2821-F-16SE의 제원

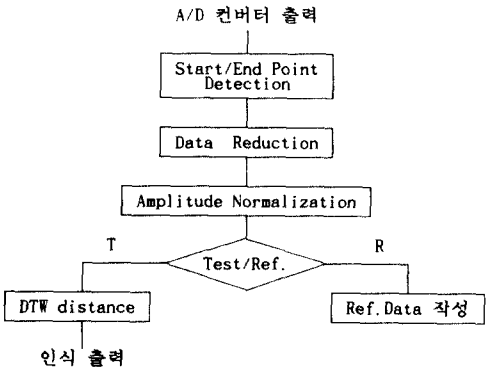

그림 11. 실험에 사용한 인식 알고리듬.

시간축에서 데이터의 양을 ½로 축소하였다. t_1, t_2, t_3 때의 데이터를 산술평균을 취하여 T_1의 데이터를 만들고 t_2, t_3, t_4 때의 데이터로 T_2를, t_3, t_4, t_5 때의 데이터로 T_3를 만들면 데이터의 갯수는 ½로 줄어든다.

마지막으로 발음 때의 진폭 불균형을 보정하기 위해서 진폭 정규화를 한다. 각 채널의 값을 모두 더한 값으로 채널의 값을 나누어 주었는데 모두 더한 값의 크기에 따라 Weighting을 줌으로써 진폭 불균형을 없앴다.

실험 방법은 우선 발성자의 음성을 받아서 표준 패턴을 만든 다음에 이 표준 패턴과 실제 실험용 음성 샘플을 비교한다. 이때 실험 장치의 특성을 변화시킬 때마다 표준 패턴을 일일이 따로 작성해야한다. 비교방법으로는 DTW 알고리듬을 이용하여 각 표준 패턴과 샘플간의 거리(Distance)를 계산하여 거리가 최소인 표준 패턴의 음성으로 판정하도록 하였다.[8]

실험 대상 단어는 대도시와 도청소재지(서울, 부산, 대구, 인천, 대전, 광주, 수원, 춘천, 청주, 전주, 창원, 제주)로 정하였으며 각각 5번씩 발음을 하여 나온 인식 출력의 평균치로 인식 성능을 평가하였다.

Ⅲ-2. SCF 뱅크의 대역폭을 변화시킨 실험

본 실험에서는 채널수 16, 포락선 검출용 저역 통과 필터의 차단 주파수 39Hz(이때 A/D 컨버터의 샘플링 주파수는 100Hz로 둠) 일 때 SCF의 Q값을 2, 4, 8.5, 12.5로 변화시켜 인식율을 비교하였다. Q값이 작으면 밴드 패스 필터의 대역이 서로 겹치고 Q값이 너무 크면 거의 단일 주파수만 통과되므로 음성 스펙트럼의 불균일성에 대처할 수가 없다. 실험의 결과는 Q = 8.5일 때가 가장 좋았다.

Q	2	4	8.5	12.5
인식률	47/60	49/60	50/60	44/60

표 4. Q값에 따른 인식률의 변동.

Q	서울	부산	대구	대전	광주	인천	수원	춘천	청주	전주	창원	제주
2	2.86	8.08	4.84	6.01	4.25	5.84	4.92	5.87	4.33	4.63	5.28	4.48
4	3.47	5.87	6.22	6.92	6.40	7.95	7.67	6.77	4.49	4.52	5.62	5.36
8.5	2.80	7.84	9.99	9.99	9.99	9.99	7.72	9.99	9.30	9.99	9.07	4.80
12.5	3.04	3.62	6.08	5.38	5.75	9.99	5.66	4.00	3.26	3.47	5.28	3.34

표 5. '서울' 데이타의 Distance 5회 평균치.
(Distance의 최대값은 10임)

Ⅲ-3. 채널수를 변경시킨 실험

본 실험에서는, 포락선 차단 주파수를 39Hz로 하고 16채널(Q = 8.5), 13채널(14, 15, 16채널 제거, Q = 8.5), 4채널(2, 5, 8, 11채널 사용, 옥타브 간격, Q = 4)을 사용했을 때의 인식률을 비교하였다. 옥타브 간격으로 할 때에는 음성 주파수 대역을 커버하기 위해서 Q를 4로 바꾸었다. 13채널의 경우는 인식률이 매우 높고 Distance의 값이 일정하게 나왔기 때문에 신뢰성이 높다고 볼 수 있다. 4채널의 경우는 Q = 4와 Q = 8.5가 예상과 달리 인식률이 비슷하였고 채널수 압축의 가능성을 보여주었다. 4채널에서는 목소리를 약간 낮추는 것이 인식률에 유리하였다.

채널수	16	13	4
인식률	50/60	55/60	48/60

표 6. 채널수에 따른 인식률의 변동.

	서울	부산	대구	대전	광주	인천	수원	춘천	청주	전주	창원	제주
1	4.59	4.75	2.18	4.33	2.80	3.41	4.45	3.35	3.11	3.29	3.22	2.47
2	4.96	5.57	3.14	5.24	3.52	4.68	5.36	4.50	3.72	3.55	4.59	3.82
3	4.27	4.65	2.29	4.23	2.51	3.52	4.89	3.85	2.78	2.76	3.56	3.14
4	4.90	5.38	2.77	4.54	3.43	3.97	4.74	4.13	3.49	3.72	3.86	3.42
5	4.91	5.37	2.78	4.83	3.05	4.32	4.46	4.18	3.22	3.58	3.61	2.91

표 7. 13ch 시스템에서 '대구' 데이터의 Distance.

	서울	부산	대구	대전	광주	인천	수원	춘천	청주	전주	창원	제주
서울	4								1			
부산		5										
대구			5									
대전				5								
광주					3				2			
인천						5						
수원							5					
춘천								5				
청주									5			
전주										5		
창원											5	
제주										2		3

표 8. 13ch 시스템의 판정 횟수.

(총 인식 성공 횟수는 55회임)

Ⅲ-4. 음성 분석 구간을 변경시킨 실험

본 실험은 음성 신호가 들어올 때 얼마만한 시간단위로 샘플링을 하는 것이 좋은가, 즉 최적 분석 구간을 찾기위한 것이다. 실험 대상으로 삼은 분석 구간은 10, 12.5, 15, 20mSec이다. 이때 어느 한 순간의 16ch의 출력을 샘플링하는 주파수는 분석 구간에 영향을 받으므로 각각 100, 80, 67, 50Hz로 정하였고 이에 따라 포락선 검출용 LPF의 차단주파수도 39, 31, 20, 20Hz로 변경시켰다. 이 실험은 4ch.(Q = 4)시스템에 대해서만 실시하였다.

분석구간(mSec)	10	12.5	15	20
샘플링주파수(Hz)	100	80	67	50
fc(LPF)(Hz)	39	30	20	20
인 식 률	48/60	43/60	42/60	41/60

표 9. 음성분석구간에 따른 인식률의 변동
(단 4ch, Q = 4임)

Ⅲ-5. 프리엠퍼시스(Ear filter)를 부가한 실험

Ⅱ-3에서 설명한 외이의 주파수 특성과 등청감 특성을 보정하기 위한 프리엠퍼시스를 부가했을 때의 인식률을 16채널(Q = 8.5)과 13채널(Q = 8.5)에 대해서 살펴보았다. 예상하던 결과와는 전혀 반대로 오히려 인식률이 엄청나게 떨어졌지만 역시 13채널을 쓸 때가 16채널 때보다 인식률이 좋은 것은 Ⅲ-3의 실험과 일치하였다. 즉 3kHz 이상의 음성 스펙트럼은 사람이 귀로 들을 때는 중요하지만 인식장치에서는 오히려 방해가 되고 있다고 볼 수 있다.

채널수	16채널	13채널
인식률	3/60	13/60

표 10. 프리엠퍼시스를 부가한 경우의 인식률.

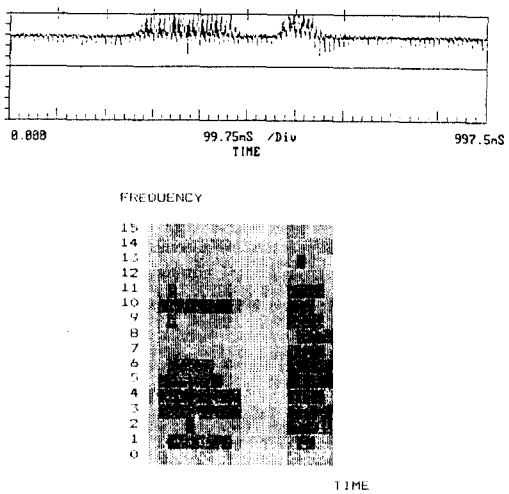

그림 12. '대전'의 시간 파형과 스펙트로그램.

Ⅳ. 결론

채널수를 16으로 고정시키고 SC BPF의 대역폭을 변화시킨 실험에서는 Q = 8.5일 때의 인식률이 가장 높았다. 뿐만 아니라 Distance값의 차이도 Q = 8.5에서 가장 크게 나왔다. 채널수 변경 실험에서는 13채널의 인식률이 90%를 넘어 16채널보다 월등하였고 채널수를 아주 줄였을때에도 상당히 좋은 결과가 나왔다. 음성 분석 구간은 짧을수록 인식률이 높았지만 그만큼 데이터 양이 증가되어 처리속도가 떨어졌다. 프리 엠퍼시스를 부가한 실험에서는 인식률이 급격하게 저하되었다. 프리 엠퍼시스와 마찬가지로 고역으로 갈수록 증폭도가 커지는 미분기, 로그 앰프 역시 같은 결과를 초래하였다. 그러므로 본 논문에서는 13채널, Q = 8.5, 분석구간 10mSec인 밴드패스 필터뱅크를 설계사양으로 제시하고자 한다.

참고 문헌

1. " 한국어 음성인식,합성 시스템 개발에 관한 연구," 한국 전자통신 연구소, 1986.
2. " ISDN을 위한 통화품질 기초 세미나," 한국 전자통신 연구소, 1989.
3. Ngoc C. Bui, Jean J. Monbaron, and Jean G. Michel, "An Integrated Voice Recognition System," IEEE J. Solid-state Circuits, Vol. SC-18, NO. 1, pp.75-80, Feb.1983.
4. Bruce A. Dautrich, Lawrence R. Rabiner, and Thomas B. Martin, "On the Effects of Varying Filter Bank Parameters on Isolated Word Recognition," IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-31, No.4, pp. 793-807, Aug.1983.
5. Alan B. Grebene, Bipolar and MOS Analog Integrated Circuit Design, Wiley-Interscience, 1974, ch.13.
6. Lyon T. Lin, Hsin-Fu Tseng, Douglas B. Cox, Sam S. Viglione, Denton P. Conrad, and Ronald G. Runge, "A Monolithic Audio Spectrum Analyzer," IEEE J. Solid-state Circuits, Vol. SC-18, No. 1, pp. 40-45, Feb. 1983.
7. The Switched-Capacitor Filter Handbook, National Semiconductor, 1985.
8. Douglas O'Shaughnessy, Speech Communication : Human and Machine, Addison-Wesley, 1987.
9. D. Robinson & R. Dadson, "A Redetermination of the Equal-Loudness Relations for Pure Tones," Br. J. Appl. Phys. 7, pp. 166-181, 1956.
10. D. Raj Reddy, "Speech Recognition by Machine : A Review." Proc. of the IEEE, Vol. 64, No. 4, pp. 501-531, Apr. 1976.
11. Rolf Unbehauen, Andrzej Cichocki, Mos Switched-Capaitor and Continuous-Time Integrated Circuits and Systems : Analysis and Design, Springer-Verlag, 1989.