

웨이블릿 필터뱅크에 기반을 둔 강인한 화자식별 기법

이 대 종[†] · 광 근 창^{††} · 유 정 웅^{†††} · 전 명 근^{†††}

요 약

본 논문에서는 웨이블릿 서브밴드 필터링기법을 이용하여 다중의사 결정기법에 기반을 둔 잡음에 강인한 화자식별 알고리즘을 제안한다. 제안된 방법은 잡음이 첨가된 음성신호를 웨이블릿 서브밴드 필터뱅크를 이용하여 각 주파수 대역별로 신호를 분리한 후 개별적인 대역별로 인식 알고리즘을 수행하기 때문에 어떤 서브밴드에서의 노이즈 영향이 상대적으로 적으므로 대역제한된 형태로 주어지는 일반적인 주변잡음이 있는 환경하에서 우수한 성능을 보일 수 있도록 시스템을 구성하였다. 제안된 알고리즘은 화자인식 기법으로 널리 쓰이고 있는 벡터양자화 알고리즘만을 적용한 경우에 비해 15~60%의 향상된 인식률을 보였다.

A Robust Speaker Identification Method Based on the Wavelet Filter Banks

Dae-Jong Lee[†] · Keun-Chang Kwak^{††} · Jeong-Woong Ryu^{†††} · Myung-Geun Chun^{†††}

ABSTRACT

This paper proposes a robust speaker identification algorithm based on the wavelet filter banks and multiple decision-making scheme. Since the proposed speaker identification algorithm has a structure performing the identification algorithm independently for each subband, the noise effect of an subband can be localized. Through this process, we can obtain more robust results for the environmental noises which generally have band limited frequency. In the experiments, the proposed method showed more 15~60% improvement than the vector quantization method for the various noisy environments.

키워드 : 강인한 화자인식(Robust speaker recognition), 정보보호(information security), 생체인식(biometrics), 웨이블릿 필터뱅크(wavelet filter bank)

1. 서 론

오늘날 우리는 급속도로 발전하는 정보사회에 살고 있다. 이러한 정보사회는 정보의 생성, 저장, 처리, 운반, 검색기능이 상호 연결된 통신망 환경하에서 다양한 형태의 정보 서비스가 이루어지는 사회이다. 한편 이러한 정보화 사회가 주는 편리함과 유익성에 비례하여 매우 위험하고 파괴적인 역기능이 뒤따르고 있다. 특히, 인가받지 않은 불법 사용자에 의한 정보시스템의 파괴, 개인신상 비밀의 누설 및 유출, 불건전 정보의 유통 등과 같은 정보화의 역기능으로 인해 고통받고 있다. 이는 향후에 정보통신 시스템과 네트워크가 더 개방되고, 용량과 성능 그리고 연결성이 강화되는 추세에 비추어, 그 취약성도 비례하여 증대될 것으로 예견

되고 있다[1].

시스템 보안을 위한 방법 중 패스워드에 의한 개인확인인증은 가장 많은 운영체제에서 지원되고 있는 구현이 용이하고 별도의 특수장치가 필요 없는 시스템 보안의 방법이다. 그러나, 개인이 아무리 자신의 패스워드 관리를 잘 한다 하여도 그것이 타인에게 누설될 수도 있으므로 이러한 패스워드에 의한 시스템 보안을 대체할 수 있는 새로운 개인확인 및 인증방법이 시급히 요구되고 있는 실정이다[2].

이러한 기존의 단순한 형태의 개인확인 및 검증방법의 한계를 극복하여, 절도나 누출에 의하여 도용될 수 없으며 변경되거나 분실할 위험성도 없는 새로운 형태의 신분검증방법에 대한 연구 분야가 바로 바이오메트릭스(Biometrics) 분야이다. 생체계측이라고 번역되는 바이오메트릭스는 “자동화된 방법으로 특정개인의 특성을 검증하거나 신분을 인식하기 위해 측정가능한 물리적 특성 또는 개인의 생물학적 특징을 연구하는 학문”으로 정의될 수 있다. 위와 같은 바이오메트릭스 기술에서 대상으로 하고 있는 인간의 생체

※ 본 연구는 한국과학재단의 목적기초연구(R02-2000-00397) 지원으로 수행되었음.

† 정 회 원 : 한국전기안전공사 전기안전시험연구원

†† 정 회 원 : 충북대학교 전기전자 및컴퓨터공학과 박사후 연구원

††† 정 회 원 : 충북대학교 전기전자 및 컴퓨터공학부 컴퓨터정보통신연구소 교수

논문접수 : 2002년 2월 25일, 심사완료 : 2002년 7월 2일

학적 특성은 크게 지문이나 얼굴, 홍채, 음성인식 등이 사용되고 있다[3, 4].

바이오펜트릭스 기술에서 사용하고 있는 대상들 중 먼저, 지문(Fingerprint)에 대해 살펴보자. 각 개인의 지문은 땀샘이 용기되어 일정한 흐름을 형성한 것으로, 그 모양이 개인마다 서로 다를 뿐만 아니라 태어날 때의 모습 그대로 평생동안 변하지 않음이 증명되어 실생활에 사용되어져 왔다. 인식에 필요한 개인별 특징추출에 사용되는 방법으로는 주파수 공간에서의 Fourier, Wavelet 변환, 또는 신경회로망이나 퍼지논리에 의한 것 등을 들 수 있다. 그러나, 지문에는 기준 좌표축이 존재하지 않으므로 임의로 회전되어 채취된 지문에 대한 처리가 어려우며, 또한 지문은 신체의 일부로서 유연성을 가지고 있어 채취할 때마다 그 모양이 달라 보이며, 지문은 손상될 수 있어 고성능의 개인 인증시스템에는 그 문제점이 지적되고 있다[5].

한편, 얼굴영상인식의 문제에서는 많은 응용분야와 다양한 연구에도 불구하고 만족할 만한 신뢰도를 보장받지 못하고 있다. 왜냐하면 정지영상의 경우는 영상으로부터의 얼굴분리가 어렵고, 동영상상을 이용한 경우는 시간정보를 이용하기 때문에 얼굴의 분리는 가능하여 정지영상상을 이용한 경우의 문제가 없으나, 영상의 질이 몹시 저하된다는 문제가 따른다. 또한 노화에 따른 얼굴의 변형이나 머리카락의 길이, 표정의 변화, 주위 조명에 많은 영향 등으로 인한 특징 추출의 문제들로 인해 아직까지 개선할 사항이 많은 상황이다.

위와 같은 단점 때문에 음성을 이용한 개인인식 및 인증에 대한 연구가 진행되어 왔었다. 그러나 이 경우 소음이나 주변환경의 잡음으로 인하여 화자인식 시스템의 심각한 성능저하가 발생하게 된다[6-10]. 이러한 문제점을 극복하기 위하여 현재까지 잡음에 강인한 화자인식 알고리즘에 관한 연구는 여러 접근방법이 사용되고 있다. 이러한 방법들로는 화자의 음성신호에 포함된 잡음을 제거하는 방법인 스펙트럼 평균 차감법(Cepstral Mean Subtraction)[7], 잡음을 선형변환의 형태로 추정하는 캡스트럼 선형 변환법(Affine Transform of Cepstrum)[8] 등이 있다. 이러한 방법들은 화자인식기와 독립적인 처리가 가능하고 비교적 계산량이 적다는 장점이 있으나, 시간에 따라 다양하게 변화하는 잡음을 적절히 처리할 수 없다는 단점이 있다. 특히, 잡음처리를 위해 가장 대표적으로 사용되는 스펙트럼 차감법인 경우 음성이 존재하지 않는 구간에서 추정된 잡음을 잡음환경에서 차감하여 잡음을 제거하므로, 추정된 잡음의 형태가 화자인식기에 입력되는 잡음 음성에 포함된 잡음과 상이한 특성을 나타낼 경우에는 효과적인 잡음제거가 불가능하다는 문

제점을 지니고 있다. 그 외에, 위너 필터(Wiener filter)[9]와 칼만 필터(Kalman filter)[10]를 이용한 음질 향상의 연구도 진행되고 있지만, 이 방법 또한 잡음의 형태가 천천히 변하는 통계적 특성을 갖는 비정상(nonstationary)인 경우 음질 향상면에서 좋은 결과를 얻을 수 없다.

이러한 기존의 잡음에 강한 화자인식 알고리즘들의 문제점을 피하기 위하여 본 논문에서는 뛰어난 주파수 분해능력을 갖고 있기 때문에 현재 다양한 분야에 적용되어 좋은 성능을 보여주고 있는 웨이블릿을 이용하고자 한다[11]. 최근 들어 화자인식 성능향상을 위하여 웨이블릿 기법을 이용한 새로운 특징벡터를 산출하고 이 특징벡터를 신경회로망 기법을 이용하여 학습시키는 방법[12-14]에 관한 연구가 많이 진행되고 있지만, 이러한 방법들은 학습데이터의 크기에 따라 인식을 변동이 크게 나타나기 때문에 대용량 화자인식기를 구축하는데 있어 문제점을 지니고 있다.

따라서, 본 논문에서는 잡음이 첨가된 음성신호를 웨이블릿 서브밴드 필터뱅크를 이용하여 각 주파수 대역별로 신호를 분리한 후 개별적인 대역별로 인식 알고리즘을 수행하기 때문에 어떤 서브밴드에서의 노이즈 영향이 다른 서브밴드에서는 적으므로 일반적으로 대역제약된 형태로 주어지는 주변잡음이 있는 환경하에서 우수한 성능을 보일 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 웨이블릿 기본이론을 설명하고 제3장에서는 기존의 벡터양자화 기법 그리고 본 논문에서 제안된 알고리즘을 설명한다. 제4장에서는 화자인식 실험을 통해 제안한 방법의 타당성을 검토하고, 마지막으로 제5장에서 결론을 맺는다.

2. 웨이블릿 변환을 이용한 신호해석

일반적으로 실생활에서 접하게 되는 대부분의 신호는 시간축과 신호의 크기를 나타내는 진폭축으로 표현된다. 이러한 신호를 시간영역에서만 분석하는 경우 신호가 포함하고 있는 정보를 충분히 해석하기 어렵기 때문에 신호분석은 시간영역의 신호를 주파수영역으로 변환하는 기법을 사용한다. 가장 널리 사용되는 변환방법은 푸리에 변환으로 오늘날 신호처리, 영상압축, 물리학 등 많은 분야에서 원신호 데이터가 지니고 있는 특성을 추출해 내기 위하여 기저함수의 가중합으로 원신호를 표현하는 기법으로 널리 이용되고 있다.

그러나, 푸리에 변환은 변환에 사용되는 정현파함수가 무한한 범위를 갖는 신호이기 때문에 시공간 영역에서 발생하는 불규칙 신호의 발생 시점을 정확히 찾아낼 수 없으며 신호의 크기도 파악하기 어렵다는 단점이 있다. 그래서 기

저함수로 윈도우 복수 정현파를 사용함으로써 이러한 문제를 해결하고자 하였다. 즉, 공간과 주파수 두 영역 모두에서 신호의 변화에 대한 정보를 표현할 수 있는 기저함수를 이용하였다. 이 변환을 STFT(Short Time Fourier Transform), 윈도우 푸리에 변환(Windowed Fourier Transform)이라고 부른다. 그러나, STFT는 신호의 모든 구간에서 동일한 윈도우가 적용되기 때문에 해상도가 시-주파수 영역에서 같아지게 되어 동일한 종류의 신호 해석에는 문제가 없지만 서로 다른 종류의 신호에 대한 해석은 어렵다는 단점이 있다[11].

웨이블릿 변환(Wavelet Transform)은 비주기적인 신호분리가 가능한 Daubechies, Coiflet, Haar, Symmlet 등과 같은 웨이블릿 계열의 기저함수를 사용하여 신호를 해석한다. 또, 자료를 해석하는 해상도가 시간축과 진폭축에 따라 다양한 형태의 윈도우를 이용하여 분석하기 때문에 원신호로부터 다양한 주기와 진폭을 갖는 패턴을 동시에 해석할 수 있는 장점을 갖고 있다. 웨이블릿 변환은 직교 변환(orthogonal transform)의 일종으로서 식 (1)과 같이 정의할 수 있으며, 시평면 신호 $x(t)$ 에 대하여 다중 윈도우(multi window) 기능을 제공함으로써 다중분해 해석이 가능하게 된다.

$$CWT_x(\tau, a) = \frac{1}{\sqrt{a}} \int x(t) h^*\left(\frac{t-\tau}{a}\right) dt \quad (1)$$

$$x(t) = c \int_{a>0} \int CWT(\tau, a) h_{a,\tau}(t) \frac{da d\tau}{a^2} \quad (2)$$

식 (1)의 웨이블릿 변환은 식 (2)와 같은 역변환식으로 나타낼 수 있으며, 웨이블릿 변환된 신호는 마더 웨이블릿(mother wavelet)을 크기변환 하거나 이동시킨 함수 $h((t-\tau)/a)$ 에 대하여 신호 $x(t)$ 를 내적한 것과 같은 기능을 갖는다. 식 (1)과 (2)로부터 크기변환 및 이동변수 a, τ 를 이산화하여 식 (3) 및 식 (4)와 같이 컴퓨터에 의해 쉽게 구현될 수 있는 형태인 이산 웨이블릿 변환 및 역변환을 구할 수 있다.

$$c_{m,n} = \sum_{k=-\infty}^{\infty} x(k) h^*_{m,n}(k) \quad (3)$$

여기서, $h_{m,n}(k) = a_0^{-m/2} h(a_0^{-m} k - nT)$, $a = a_0^m$, $b = na_0^m T$

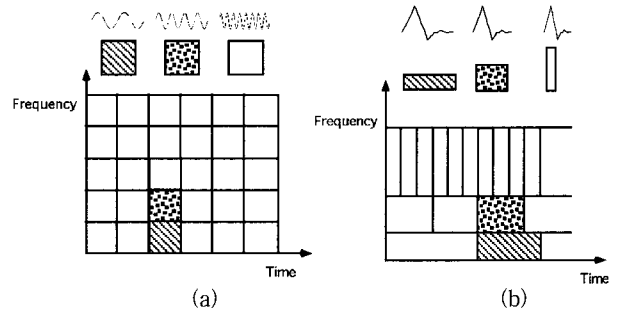
$$x(n) = c \sum_j \sum_k c_{j,k} h_{j,k}(n) \quad (4)$$

여기서, c 는 정규화 상수이다.

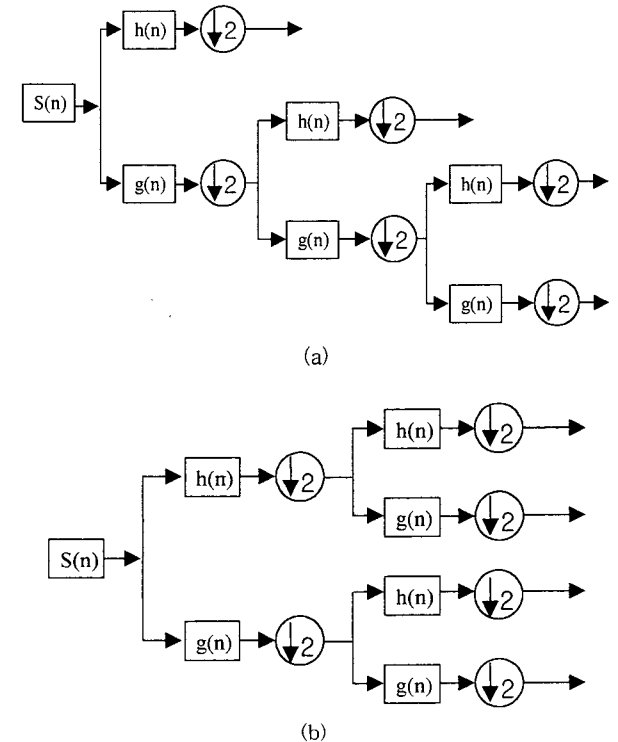
웨이블릿 변환은 STFT가 갖는 단점을 해소하기 위하여 신호의 특성에 따라 사용하는 윈도우의 크기를 유동적으로 변화시킬 수 있는 기저함수를 사용한다. 즉, (그림 1-(b))와 같이 신호의 고주파 성분은 고립시켜 해석할 경우에는 주

기가 짧은 기저함수를, 저주파 성분을 세밀하게 해석하고자 할 경우에는 주기가 긴 기저함수를 사용한다.

이산 웨이블릿 변환은 고역 통과 부분을 한 단계의 필터뱅크로 구성하고, 저역 통과 부분을 계속적인 필터뱅크로 확장하는 옥타브 밴드(octave-band) 구조와 고역 통과 부분도 필터뱅크로 확장하는 구조를 가지는 웨이블릿 패킷(wavelet packet)구조로 구현될 수 있다[11]. (그림 2)에서는 옥타브 밴드 구조와 웨이블릿 패킷구조를 보이고 있는데, 여기서 $g[n]$ 은 저역 통과 필터를, $h[n]$ 은 고역 통과 필터를 각각 나타내며, 마더 웨이블릿으로부터 구성됨을 알 수 있다. 또한 $\downarrow 2$ 는 샘플의 개수를 1/2로 줄이는 데시메이션(decimation)을 나타낸다.



(그림 1) 시-주파수 평면에서 윈도우 형태 : (a) 푸리에 변환, (b) 웨이블릿 변환(Daubechies)



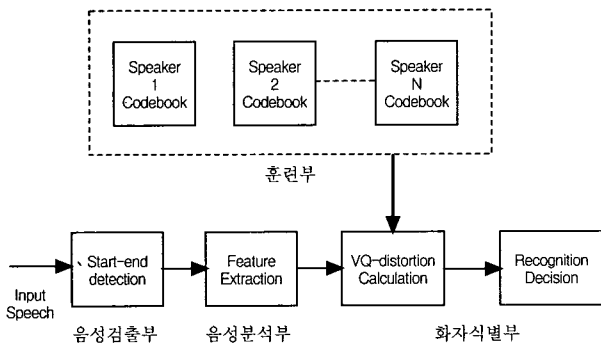
(그림 2) 웨이블릿의 구조 (a) 옥타브 밴드 (b) 웨이블릿 패킷

3. 웨이블릿 필터뱅크를 이용한 강인한 화자식별

벡터양자화를 이용한 일반적인 화자인식 시스템은 (그림 3)에서 보는 바와 같이 음성 검출부, 음성 분석부, 훈련부, 인식부로 구성되어 있다. 화자인식 시스템을 구성하는 각 부분을 간단히 설명하면, 우선 음성 검출부에서 음성의 시작점과 끝점을 검출하며, 음성 분석부에서는 각 단어들의 음성에 대한 분석을 거쳐서 최종적으로 각각 선형예측계수와 멜주파수 캡스트럼 등의 계수를 구하는 부분이다. 훈련부에서는 화자인식 알고리즘에 따라 여러 가지 기법이 있으며, 벡터양자화 알고리즘인 경우 음성 분석부에서 얻어진 특징벡터들을 K-means 클러스터링 알고리즘을 이용하여 각각의 음성을 대표하는 코드북을 만드는 과정이다. 인식부에서는 식 (5)에서 보인 바와 같이 유클리디안 방법을 이용하여 입력된 화자데이터를 훈련과정에서 만들어둔 각 단어에 대한 코드북과 비교하여 벡터 공간상의 거리가 가장 가까운 것을 선택하고 선택된 코드북에 해당하는 화자를 인식으로 결정하는 부분이다[15].

$$d(x, y) = (x - y)W^{-1}(x - y)^T, \quad W = I \quad (5)$$

여기서, x : 기준 코드북내의 특징벡터
 y : 입력음성의 특징벡터

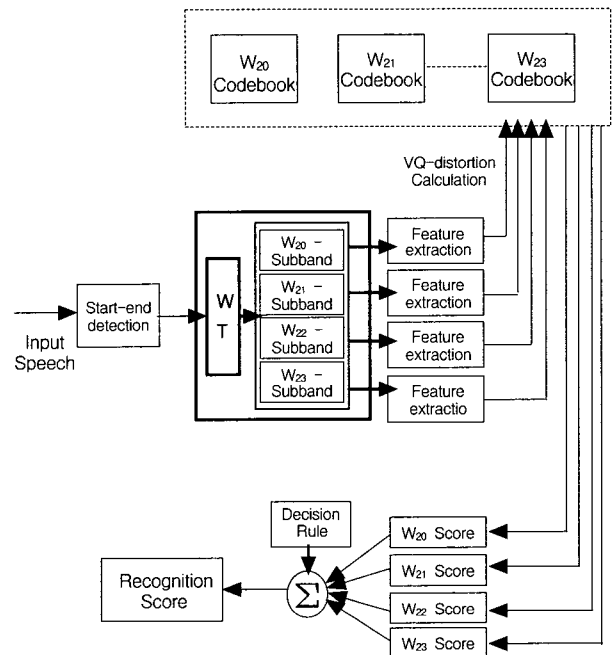


(그림 3) 벡터양자화에 의한 화자식별

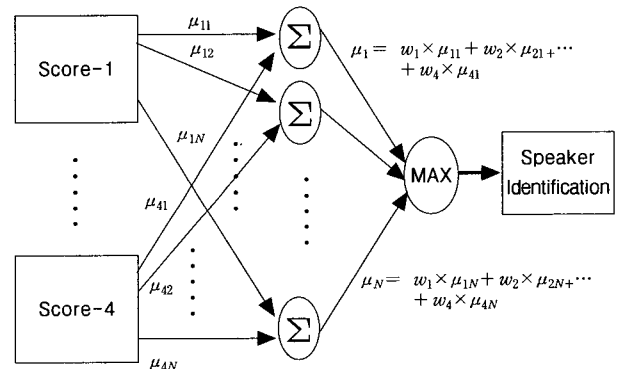
위와 다르게 (그림 4)에서는 본 논문에서 제안한 웨이블릿 필터뱅크를 이용한 화자인식기의 인식과정을 보여 주고 있다. 여기서, 웨이블릿 필터뱅크의 구조는 주파수 대역을 균등하게 분할하는 방식인 웨이블릿 패킷구조방식을 사용하였으며, 화자인식을 위해 사용된 필터는 (그림 2-(b))와 같은 4개의 필터뱅크로 구성되었고 이중 다양한 실험에서 가장 낮은 인식률을 보인 대역을 제외하고자 한다. 그리고 가장 널리 사용되는 Daubechies 함수를 이용하여 신호를 해석하였다.

각각의 필터에서 출력되는 화자의 음성신호는 음성 분석부에서 특징벡터를 계산한 후 기존의 벡터양자화 기법과 동일하게 K-means 알고리즘을 이용하여 독립적인 코드북

을 미리 만들어 놓는다. 인식과정에서는 인식하고자 하는 화자의 음성신호가 입력되면 웨이블릿 변환을 하여 주파수 별로 음성신호를 분할한 후 음성 분석부에서 각각의 주파수 대역에 대한 특징벡터를 계산한다. 이렇듯 음성 분석부에서 계산된 특징벡터는 미리 बैं크별로 만들어 놓은 코드북과의 거리를 계산한 후 독립적인 화자 인식률을 산출한다. 최종적으로는 각각의 대역별에서 출력된 화자인식률을 이용하여 최종 인식률을 결정하기 위한 다중밴드에서의 의사결정방법을 결정해 주어야 한다. 본 논문에서는 (그림 5)와 같은 의사결정법칙을 제안한다.



(그림 4) 웨이블릿 필터뱅크에 기반을 둔 강인한 화자식별



(그림 5) 다중밴드에서의 의사결정 방법

(그림 5)에서 보는 바와 같이 각 대역별에서 산출된 인식률은 음성신호를 프레임으로 나누고 각각의 프레임에서 얻어진 특징벡터와 코드북과의 거리계산에 의하여 산출되기

때문에 어느 특정 단어에 대한 정보만을 가진 것이 아니라 인식하고자 하는 대상에 대한 소속도를 가지고 있다. 이러한 소속도를 μ_{ij} 로 표현했으며, 여기서 μ_{ij} 는 서브밴드 i 에서 단어 j 의 소속도를 나타낸다. 마지막으로 최종인식 단계에서는 각각의 소속도를 합산하여 나타낸 μ_k (k 의 단어가 인식될 소속도)를 구할 수 있으며, 이 중에서 소속도가 가장 높은 화자를 인식대상으로 선정한다. 본 논문에서는 각 밴드별의 인식률에 따른 가중치 w 를 [1 1 0]으로 설정하여 최고주파 성분에 대한 영향을 무시하였다.

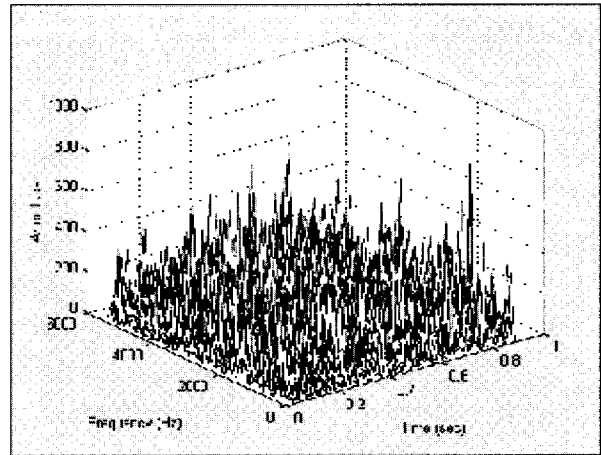
4. 노이즈하에서의 화자인식 실험 및 결과

4.1 실험환경 및 화자 음성데이터 구성

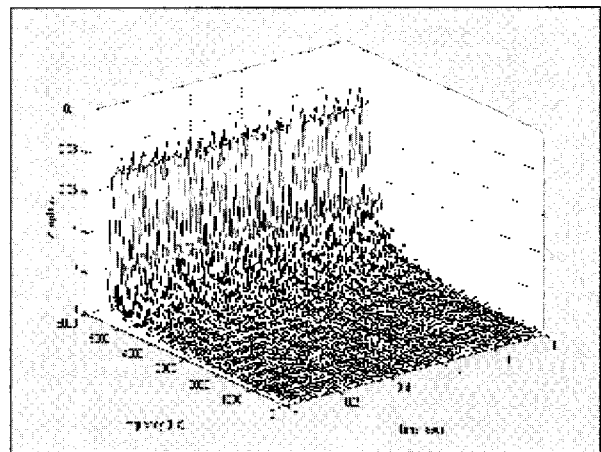
본 논문에서는 주위환경 또는 잡음에 따라 급격하게 저하되는 인식성능을 개선하기 위하여 웨이블릿 필터뱅크를 기초로 한 잡음에 강인한 벡터양자화 음성인식기를 제안하였다. 제안된 시스템의 성능을 평가하기 위하여 (그림 6)에서 보는 바와 같이 랜덤 노이즈와 5kHz의 특정주파수 대역에 한정된 노이즈를 첨가하였다. 일반적으로 노이즈에 대한 화자인식률을 알아보기 위하여 (그림 6-(a))와 같은 랜덤 노이즈를 사용하지만[7-10], 본 논문에서는 화자인식률에 악영향을 끼치는 특정 고주파성분의 노이즈에 대해서 제안된 알고리즘이 어느 정도 효과가 있는가를 알아보기 위하여 (그림 6-(b))와 같은 특정 노이즈에 대해서도 인식률 조사를 알아보았다. 인식률 변동에 따른 인식률 비교를 위하여 잡음이 없는 상태에서 녹음한 화자의 음성단어(숫자음: 영~구)에 각각의 SNR에 맞는 노이즈를 첨가하였다.

화자의 음성 데이터는 잡음이 최대로 억제된 상황에서 녹음된 우리말 숫자음 10개 /영/에서 /구/을 대상으로 실험하였다. 남성화자 7명과 여성화자 3명이 각각 10회씩 발음한 음성신호 중 2개는 학습, 또는 기준패턴을 만들기 위해서 사용하였고, 나머지 8개는 화자인식실험을 위해서 사용하였다. 녹음된 음성데이터의 샘플링 주파수는 10kHz이며, 기준패턴인 코드북의 사이즈는 32로 정하였다. 일반적으로 음성신호의 특징파라미터는 약 20ms 구간에서 음성신호가 정상(stationary)이라는 가정아래 20ms의 프레임 단위로 구하여 지는데, 본 논문에서도 20ms의 Hamming window를 사용하였고, 프레임 양 끝단의 신호정보를 보충하기 위하여 10ms씩 중첩을 시켜서 윈도우를 이동시켰다. 이렇게 Hamming window를 사용하여 원 신호를 프레임 단위로 분할한 후 각각의 프레임에 포함된 데이터에서 13차의 멜켵스트림 계수를 구하였다. 벡터 양자화 과정에서 음성의 시작점과 끝점을 검출하는 정확하게 검출하는 것은 매우 중요한데,

본 논문에서는 Rabiner와 Sambur에 의해 제안된 단시간 평균에너지(Short-time average energy)와 단시간 영교차율(Short-time zero crossing rate)을 이용한 알고리즘을 사용하였다[15].



(a)



(b)

(그림 6) 노이즈의 크기 및 주파수 특성 (a) 랜덤 노이즈에 대한 주파수 분석, (b) 특정 주파수 노이즈에 대한 주파수 분석

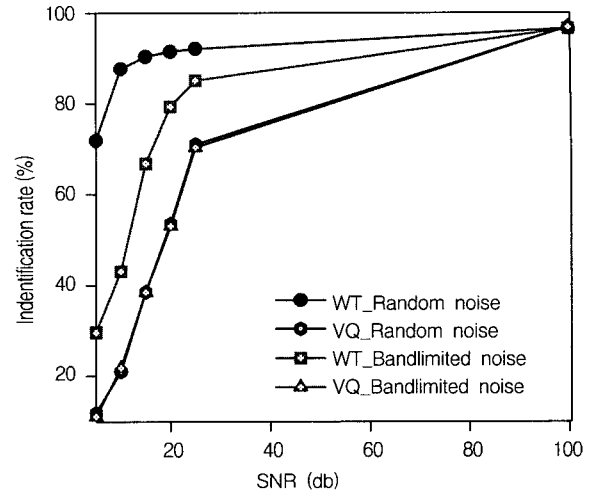
음성의 시작점과 끝점을 검출한 후 음성의 고주파성분을 나타내기 위하여 일반적으로 $H(z) = 1 - 0.95z^{-1}$ 와 같은 고역 통과 필터를 이용한 전처리(pre-emphasis) 과정을 거치는데, 웨이블릿 기법을 이용하는 경우 전처리과정을 하면 원신호가 가지고 있던 각각의 대역별 신호가 유실되기 때문에 사용하지 않았다.

4.2 실험 결과

(그림 7) 및 <표 1>에서는 본 논문에서 제안한 웨이블릿 패킷구조의 웨이블릿 기법을 적용한 경우 노이즈의 SNR

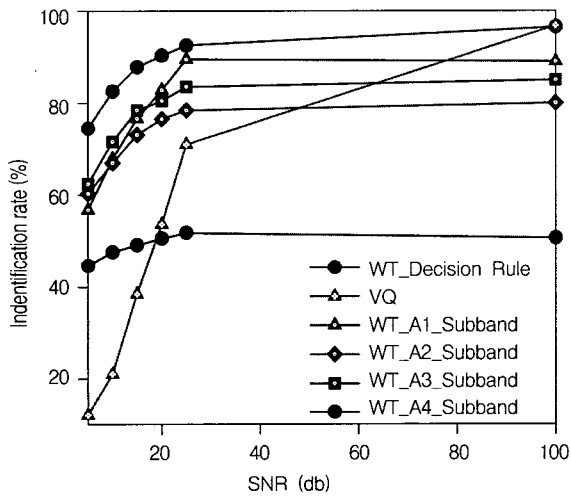
(Signal-to-Noise Rate) 변동에 따른 인식률을 나타내었다. 기존의 VQ기법을 적용한 경우보다 본 논문에서 제안한 기법을 적용한 경우 랜덤노이즈를 첨가한 경우 인식률이 최소 15%에서 최대 20%까지 향상되었고, 고주파성분의 특정 노이즈를 첨가한 경우는 최소 20%에서 최대 60%까지 인식률이 향상된 것으로 나타났다.

기존의 VQ기법은 (그림 7)에서 보는 바와 같이 랜덤노이즈 또는 특정주파수 성분을 갖고 있는 고주파 노이즈에 상관없이 어떠한 노이즈에도 인식률에 민감한 반응을 보이고 있는 것으로 나타났다. 또한, <표 1>에서 보는 바와 같이 웨이블릿을 이용한 경우 A4대역에서 인식률이 매우 저조해 최종 인식률 하락을 가져오는 요인이라 할 수 있다. 따라서, 본 논문에서 제안된 알고리즘은 매우 낮은 A4대역을 제외한 A1, A2, A3 대역을 이용하였고 최종 인식결정은 (그림 4) 및 (그림 5)와 같은 구조로 이루어졌다.

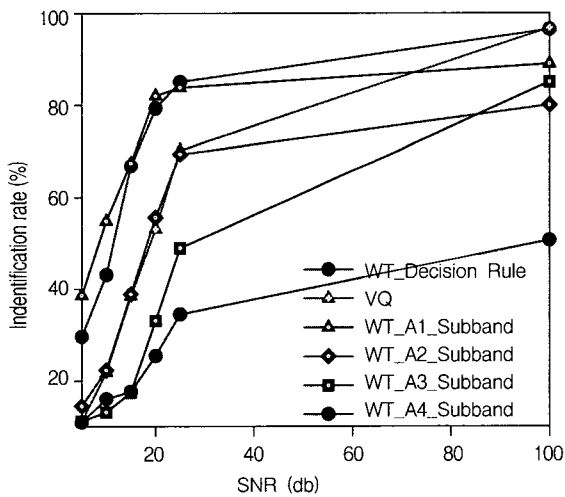


(c)

(그림 7) 노이즈 변동에 따른 인식률 : (a) 랜덤노이즈하에서의 인식률, (b) 특정 노이즈 하에서의 인식률, (c) 최종 인식률 비교



(a)



(b)

<표 1> 노이즈 변동에 따른 인식률 비교

(단위 : %)

SNR	VQ	Wavelet Filtering					
		A1 Band	A2 Band	A3 Band	A4 Band	Decision Rule	
Random Noise	∞	96.8	89.0	80.0	85.0	50.63	96.5
	25	70.3	83.8	69.5	48.9	34.5	85.0
	20	53.1	82.0	55.6	33.1	25.4	79.3
	15	38.5	67.5	38.9	17.4	17.6	66.8
	10	21.8	54.8	22.3	13.1	16.0	43.1
	5	11.1	38.6	14.4	11.1	10.9	29.6
Band Limited Noise	∞	96.8	89.0	80.0	85.0	50.63	96.5
	25	71.0	85.5	78.4	83.5	51.75	90.3
	20	53.6	82.75	76.5	80.5	50.5	88.0
	15	38.5	76.6	73.1	78.4	49.1	84.1
	10	21.0	67.9	67.0	71.6	47.6	76.8
	5	11.9	56.8	60.3	62.4	44.6	67.3

본 실험결과 본 논문에서 제안한 웨이블릿의 서브밴드 필터링 기법을 이용하여 화자인식을 하는 경우 랜덤노이즈 성분보다는 특정주파수 성분을 갖고 있는 노이즈 성분에 대해서 매우 우수한 결과를 보이고 있다. 이와 같이 특정 주파수성분의 노이즈 성분에 대해서 본 알고리즘의 인식률이 향상된 것은 (그림 7)에서도 알 수 있는 바와 같이 잡음이 집중된 고주파대역의 A4대역에서는 노이즈에 민감한 반응을 보인 반면 이 보다 저대역인 A1, A2, A3 영역에서는 상대적으로 잡음의 영향에 강한 특성을 보이기 때문이다. 따라서, 본 논문에서 제안된 알고리즘은 보통 대역제한된 형태로 표현되어지는 주변잡음이 있는 환경 하에서 좋은

성능을 발휘할 수 있음을 알 수 있다.

4. 결 론

본 논문에서는 잡음환경하에서 인식능력을 향상시켜주는 새로운 화자인식 알고리즘을 구현하였다. 제안된 알고리즘은 인식하고자 하는 화자의 음성을 웨이블릿 패킷구조를 사용하여 각각의 주파수 대역별로 분할한 후 벡터양자화 기법을 이용하여 개별적인 인식률을 조사한 후 최종적으로 인식결정 방법에 의해 고립 단어를 인식하는 구조로 되어 있다.

제안된 알고리즘의 성능을 검증하기 위하여 실험한 결과, 기존의 VQ방법에 비해 랜덤노이즈를 첨가한 경우 인식률이 최대 20%에서 최소 15%까지 향상되었고, 고주파성분의 특정 노이즈를 첨가한 경우는 최대 60%에서 최소 20%까지 인식률이 향상되었다. 이로부터 본 논문에서 제안한 웨이블릿의 서브밴드 필터링 기법을 이용하여 화자인식을 하는 경우 보통 대역제한된 형태로 표현되어지는 주변잡음이 있는 환경 하에서 우수한 성능을 발휘하므로 본 연구의 결과는 정보보호를 위한 생체인식 기법의 하나로 적용될 수 있리라 생각된다.

참 고 문 헌

[1] A. Jain, R. Bolle, and S. Pankanti, "Biometrics : Personal Identification in Networked Society," Kluwer Academic Publish, 1999.

[2] 특허청, "2001 신기술동향 조사 보고서-생체인식시스템-", 2001.

[3] 한국전자통신연구원, "생체측정시스템", 2000.

[4] 고현주, 이상원, 전명근, "개인확인 및 인증 알고리즘을 위한 홍채패턴인식", 정보처리학회논문지 C, 제8-C권 제5호, pp. 499-506, 2001.

[5] J. G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," IEEE Tr. on Acoustics, Speech and Signal Processing, Vol.7, pp.1169-1179, 1988.

[6] J. P. Gampbell, Jr., "Speaker Recognition : a Tutorial," Proceedings of the IEEE, Vol.85, pp.1436-1462, 1997.

[7] S. F. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE, Trans., ASSP, Vol.37, No.2, pp.113-120, 1979.

[8] R. J. Mammone, X. Zhang, and R.P. Ramachandran, "Robust Speaker Recognition : a Feature-based Approach," IEEE Signal Processing Magazine, pp.58-71, 1996.

[9] J. H. L. Hansen and M.A. Clements, "Constrained iterative speech enhancement with application to speech recognition." IEEE Trans. Singal Processing, Vol.39, No.4, pp. 795-805, Apr., 1991.

[10] B. G. Lee, K. Y. Lee, and S. Ann, "An EM-based Approach for Parameter Enhancement with an Application to Speech Signal," Signal Processing, Vol.46, pp.1-14, 1995.

[11] Stephane Mallat, A wavelet tour of signal processing, Academic press, 1999.

[12] Francis Phan, M. T. Evangelia, and Samuel Sideman, "Speaker Identification Using Neural Networks and Wavelets," IEEE Engineering in Medicine and Biology, Vol., pp.92-101, 2000.

[13] C. W. Woo, C. P. Lim, and R. Osman, "Development of a Speaker Recognition System using Wavelets and Artificial Neural Networks," IEEE Proceeding on Intelligent Multimedia, Video and Speech Processing, Vol., pp.413-416, 2001.

[14] S. George, A. Dibazar, J, S. Liaw, and T. W. Berger, "Speaker Recognition using Dynamic Synapse Based Neural Networks with Wavelet Preprocessing," IEEE Proceeding on Intelligent Multimedia, Video and Speech Processing, Vol., pp.413-416, 2001.

[15] L. R. Rabiner and B. H. Juang, Fundamentals of speech recognition, Prentice-Hall, 1993.

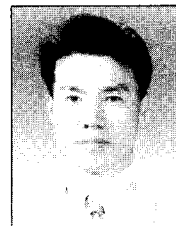
이 대 종



e-mail : leebigbell@hanmir.com
 1995년 충북대학교 전기공학과(학사)
 1997년 충북대학교 전기공학과(공학석사)
 2002년 충북대학교 전기공학과(공학박사)
 2000년~현재 한국전기안전공사 전기안전
 시험연구원

관심분야 : 음성신호처리, 얼굴인식, Biometrics

곽 근 창



e-mail : kckwak@power.chungbuk.ac.kr
 1996년 충북대학교 전기공학과(학사)
 1998년 충북대학교 전기공학과(공학석사)
 2001년 충북대학교 전기공학과(공학박사)
 현재 : 충북대학교 전기전자 및 컴퓨터공
 학부 박사후 연구원

관심분야 : 지능제어, 패턴분류, 얼굴인식



유 정 웅

e-mail : wry@engine.chungbuk.ac.kr
1965년 한양대학교 전기공학과(학사)
1976년 단국대학교 전기공학과(공학석사)
1987년 단국대학교 전기공학과(공학박사)
1969년~1979년 대전개방대학 교수
1979년~현재 충북대학교 전기전자 및
컴퓨터공학부 컴퓨터 정보통신연
구소 교수

관심분야 : 지능제어, 가변구조제어, QFT제어



전 명 군

e-mail : mgchun@cbucc.chungbuk.ac.kr
1987년 부산대학교 전자공학과(학사)
1989년 한국과학기술원 전기 및 전자공학과
(공학석사)
1993년 한국과학기술원 전기 및 전자공학과
(공학박사)

1993년~1996년 삼성전자 자동화연구소 선임연구원
2000년~2001년 University of Alberta 방문교수
1996년~현재 충북대학교 전기전자 및 컴퓨터공학부 컴퓨터
정보통신연구소 교수

관심분야 : Biometrics, 감정인식, 음성신호처리, 얼굴인식