

웨이블릿 필터뱅크를 이용한 자동차 소음에 강인한 고립단어 음성인식

Robust Speech Recognition with Car Noise based on the Wavelet Filter Banks

이대종 · 곽근창 · 유정웅 · 전명근

Dae-Jong Lee, Keun-Chang Kwak, Jeong-Woong Ryu and Myung-Geun Chun

충북대학교 전기전자 및 컴퓨터공학부 컴퓨터정보통신연구소

요 약

본 논문에서는 웨이블릿 서브밴드 필터링기법을 이용하여 다중의사 결정기법에 기반을 둔 외부 잡음에 강인한 고립단어 음성인식 알고리즘을 제안하고자 한다. 음성인식에 있어서 외부잡음은 음성인식 알고리즘의 인식률을 저하시키는 주요 원인으로 지적되므로 음성인식기의 성능을 향상시키기 위해서 무엇보다도 잡음에 강인한 음성인식 알고리즘의 개발이 절실히 요구되고 있다. 제안된 알고리즘의 타당성을 검증하기 위하여 다양한 자동차 소음하에서 한국어 단독 숫자음 10단어의 인식률 변동을 알아 보았다. 그 결과 현재 음성인식 기법으로 널리 쓰이고 있는 벡터양자화 알고리즘만을 적용한 경우에 비해 9~25%의 향상된 인식률을 보였다.

Abstract

This paper proposes a robust speech recognition algorithm based on the wavelet filter banks. Since the proposed algorithm adopts a multiple band decision-making scheme, it performs robustness for noise as the presence of noisy severely degrades the performance of speech recognition system. For evaluating the performance of the proposed scheme, we compared it with the conventional speech recognizer based on the VQ for the 10-isolated korean digits with car noise. Here, the proposed method showed more 9~25% improvement of the recognition rate than the conventional VQ algorithm for the various car noisy environments.

Key words : Robust speech recognition, wavelet filter bank, car noise

1. 서 론

컴퓨터와 더불어 입력 매체도 함께 발전하여, 이제는 키보드와 마우스에서 벗어나 터치 스크린, 음성 인식, 시각인식 등으로 발전을 하고있다. 이들 중 음성은 인간의 지적 활동과 상호 커뮤니케이션의 가장 보편적 수단으로 가장 효율적이고 자연스러운 의사전달 수단이다. 이러한 음성인식의 궁극적 목표는 인간의 대화하는 능력을 가장 유사하게 모델링하며 그 성능을 극대화하는 방향, 즉 사람과 사람 사이의 자연스러운 대화를 인간과 기계 사이에 실현하는 것이다[1].

이러한 음성인식의 상용화를 위해 우선 시도된 분야가 안정성과 편리성을 강조하는 자동차에서의 음성인식기 개발이었다. 최근에 차량내에서 핸드폰 사용에 따른 사고를 방지하기 위하여 핸드프리를 사용하고 있지만 이

또한 전화를 걸 때 수동으로 조작해야 한다는 문제점을 지니고 있다. 따라서, 차량의 각종 장치의 조작을 음성으로 대신하게 된다면 주행시의 편의성과 안전성을 높일 수 있다. 그러나, 자동차 내에서의 음성인식의 어려움은 차량 주행시 발생하는 소음이나 주변환경의 잡음에 의해 인식률이 매우 저조해진다는 문제점을 지니고 있어 이를 극복할 수 있는 방법의 개발이 시급한 실정이다[2,3].

현재까지 잡음에 강인한 음성인식 알고리즘에 관한 연구는 여러 접근방법이 사용되고 있는데, 우선 음성인식의 전 단계에서 잡음을 제거하는 방법으로 잡음에 강인한 특징추출과 거리척도인 SMC(Short-time Modified Coherence)[4], RASTA(Relative SpecTrAl)[5] 처리와 동적인 특징파라미터[6], 캡스트럼 사상척도[7] 등의 방법과 음성신호에 포함된 잡음을 제거하는 방법인 스펙트럼 차감법[8,9], 청각기관의 특성을 이용하는 방법[10] 등이 있다. 이러한 방법들은 음성인식기와 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 잡

접수일자 : 2001년 6월 19일

완료일자 : 2001년 12월 12일

본 연구는 정보통신부 대학기초연구지원사업에 의해 일부 지원되었습니다.

음을 제거하므로, 추정된 잡음의 형태가 음성인식기에 입력되는 잡음 음성에 포함된 잡음과 상이한 특성을 나타낼 경우에는 효과적인 잡음제거가 불가능하다는 문제점을 지니고 있다.

이러한 기존의 잡음에 강한 음성인식 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 음성인식 성능향상을 위하여 웨이블릿 기법을 이용한 새로운 특징벡터를 산출하는 방법[12,13]과 웨이블릿 서브밴드 필터링을 이용하는 방법[14] 등이 제안되었다. 전자의 경우 신호를 프레임으로 나눈 후 프레임내에서 웨이블릿 변환을 거쳐 각 대역의 에너지를 합한 것을 특징벡터를 산출하는 방법[12]과 정적인 정보는 기존의 켈프스트림 계수를 이용하고 동적인 정보는 웨이블릿 변환을 통하여 얻어진 에너지를 이용하여 특징벡터를 산출하는 방법[13]이 있다. 그러나, 위에서 서술한 두가지 방법은 외부 잡음이 없는 이상적인 신호에 대해서 다루었으며, 외부잡음에 대한 강인함에 관한 언급은 없다. [14]인 경우는 웨이블릿 필터링을 통과한 신호에 대하여 기존의 켈프스트림 계수를 산출하는 방법으로 기존의 켈프스트림을 이용한 경우보다 1~2%의 성능향상을 보였으므로 실제적으로 적용하는데는 어려움이 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. 2장에서는 웨이블릿 기본이론을 설명하고 3장에서는 기존의 벡터양자화 기법 그리고 본 논문에서 제안된 알고리즘을 설명한다. 4장에서는 음성인식 실험을 통해 제안한 방법의 타당성을 검토하고, 마지막으로 5장에서 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 신호 분석은 시간영역의 신호를 주파수영역으로 변환하는 기법을 사용한다. 가장 널리 사용되는 방법인 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 기저함수로 윈도우복수 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 변환이라고 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 대

한 해석은 어렵다는 단점이 있다[11].

웨이블릿 변환은 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 직교변환의 일종으로서 식 (1)과 같이 정의할 수 있으며, 시평면 신호 $x(t)$ 에 대하여 다중 윈도우 기능을 제공함으로써 다중분해능 해석을 가능하게 한다.

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

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

식 (1)의 웨이블릿 변환은 식 (2)와 같은 역변환식으로 나타낼 수 있으며, 웨이블릿 변환된 신호는 마더 웨이블릿을 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.

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

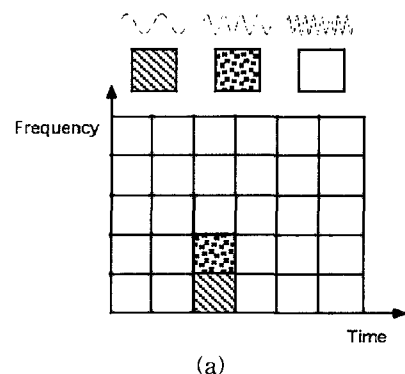
여기서, $h_{m,n}(k) = a_0^{-m/2} h(a_0^{-m}k - nT), \quad a = a_0^m, \quad b = na_0^m T$

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

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 1과 같이 신호의 고주파 성분을 고립시켜 해석할 경우에는 주기가 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 구조와 고역 통과 부분도 필터뱅크로 확장하는 구조를 가지는 웨이블릿 패킷 구조



(a)

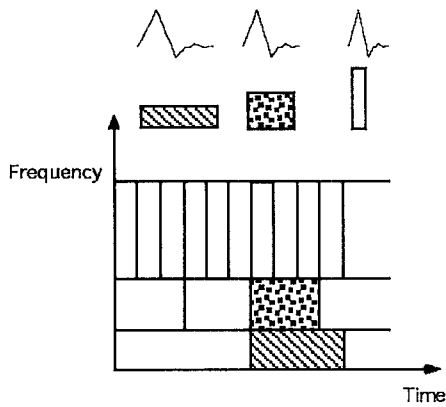
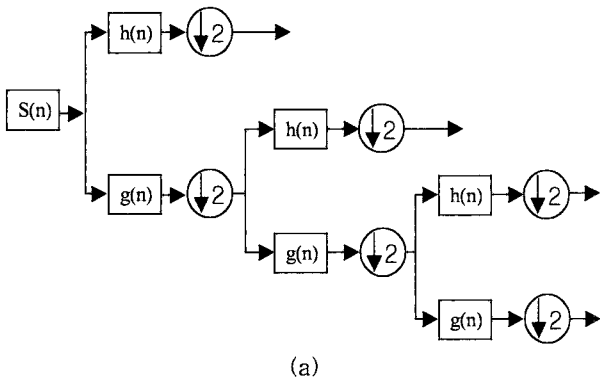


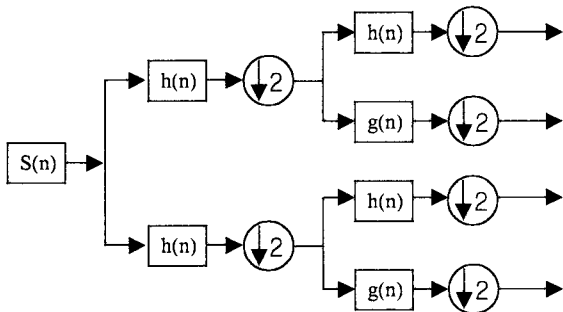
그림 1. 시간-주파수 평면에서 윈도우 형태 : (a) 푸리에 변환, (b) 웨이블릿 변환(Daubechies)

Fig. 1. Window form for time-frequency plane

로 구현될 수 있다[11]. 그림 2에서는 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.



(a)



(b)

그림 2. 웨이블릿의 구조 (a) 옥타브 밴드 (b) 웨이블릿 패킷
Fig. 2. Wavelet structure (a) Octave band (b) Wavelet packet

3. 웨이블릿 필터뱅크를 이용한 자동차 소음에 강인한 고립단어 음성인식

벡터양자화를 이용한 일반적인 음성인식 시스템은 그림 3에서 보는 바와 같이 음성 검출부, 음성 분석부, 훈련부, 인식부로 구성되어 있다. 음성인식 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 쉐프트럼 등의 계수를 구하는 부분이다. 훈련부에서는 음성인식 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 입력된 음성데이터를 훈련과정에서 만들어둔 각 단어에 대한 코드북과 비교하여 벡터 공간상의 거리가 가장 가까운 것을 선택하여 그 코드북에 해당하는 단어를 인식으로 결정하는 부분이다[1].

위와 다르게 그림 4에서는 본 논문에서 제안한 웨이블릿 필터뱅크를 이용한 음성인식기의 단어 인식과정을 보여 주고 있다. 여기서, 웨이블릿 필터뱅크의 구조는 주파수 대역을 균등하게 분할하는 방식인 Daubechies의 기저함수를 이용한 웨이블릿 패킷구조방식을 사용하였으며, 필터뱅크의 출력수는 4개로 한정하였다.

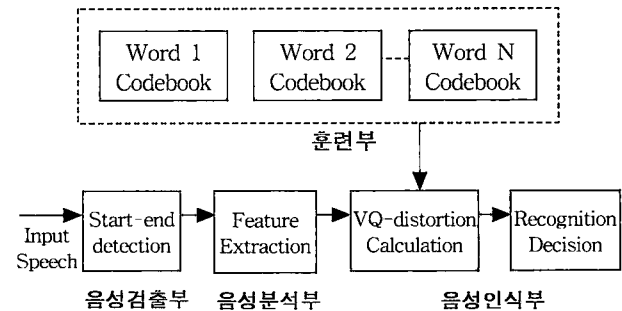


그림 3. 벡터양자화에 의한 음성인식

Fig. 3. Speech recognition based on vector quantization

각각의 필터에서 출력되는 음성신호는 음성 분석부에서 특징벡터를 계산한 후 기존의 벡터양자화 기법과 동일하게 K-means 알고리즘을 이용하여 독립적인 코드북을 미리 만들어 놓는다. 인식과정에서는 인식하고자 하는 음성신호가 입력되면 웨이블릿 변환을 하여 주파수별로 음성신호를 분할한 후 음성 분석부에서 각각의 주파수 대역에 대한 특징벡터를 계산한다. 이렇듯 음성 분석부에서 계산된 특징벡터는 미리 벡터별로 만들어 놓은 코드북과의 거리를 계산한 후 독립적인 인식률을 산출한다. 여기서 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 단어들에 대한 소속정도를 모두 가지고 있다. 이러한 각 단어에 대한 소속도를 정규화 하기 위하여 각 단어의 선택된 프레임수를 총 프레임으로 나누었다. 마지막으로 최종 인식단계에서는 각각의 소속도를 합산한

후, 그 중에서 소속도가 가장 높은 단어를 인식단어로서 선정한다.

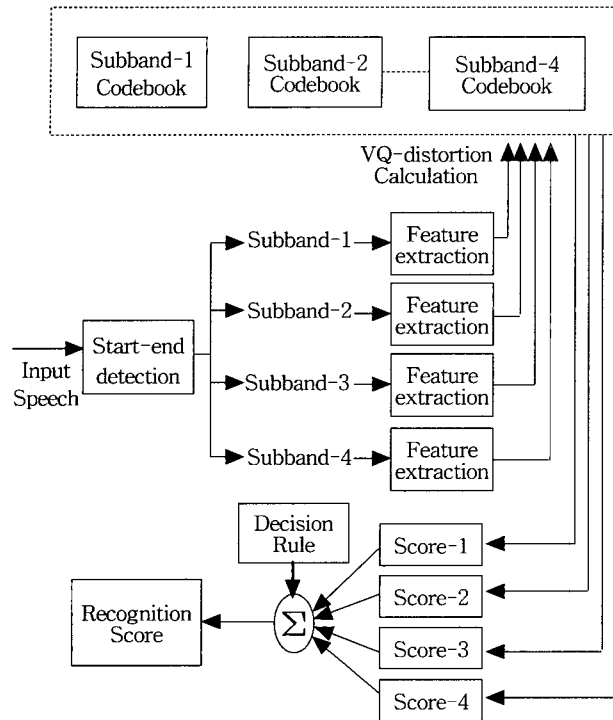


그림 4. 웨이블릿 필터뱅크를 이용한 자동차 소음에 강인한 음성인식기

Fig. 4. Robust speech recognition with car noise based on the wavelet filter banks

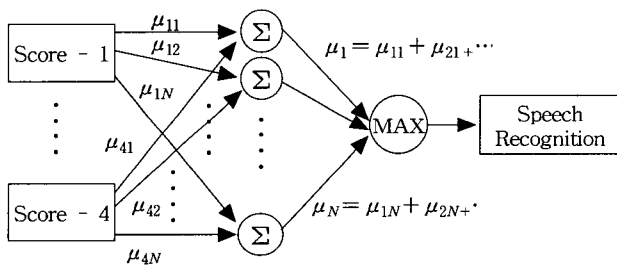


그림 5. 다중밴드에서의 의사결정 방법

Fig. 5. Multiple band decision-making method

위 과정을 그림 5에 나타내었는데, 그림에서 μ_{ij} 는 서브밴드 i 에서 단어 j 의 소속도를 의미하고 μ_k 는 k 의 단어가 인식될 소속도를 나타낸다.

4. 자동차소음 하에서 고립단어 인식 실험 및 결과

4.1 자동차 소음 및 주파수 특성

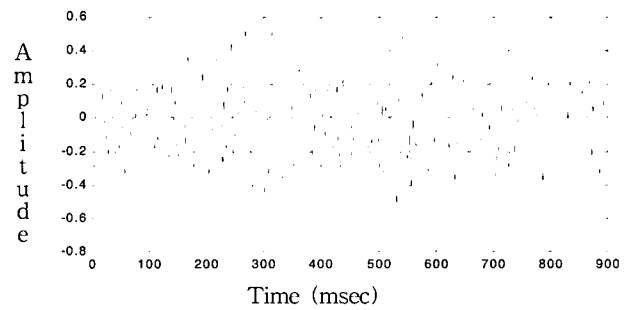
본 논문에서 제안한 알고리즘의 유용성을 평가하기 위하여 자동차 내에서의 차량소음을 다각적으로 녹음·분석을 하였다. 자동차내에서의 녹음은 발생할 수 있는 모든 경우를 고려하여 표 1에 보인 바와 같이 6가지 경우에 대하여 녹음을 하였다. 모의실험은 표 1에서 보

인 I ~ VI의 차량소음의 크기를 변동시키는 방법으로 각각의 SNR을 변화시킨 노이즈를 잡음에 없는 상태에서 녹음한 고립단어(영~구/공)에 첨가시킨 후 인식률을 조사하였다.

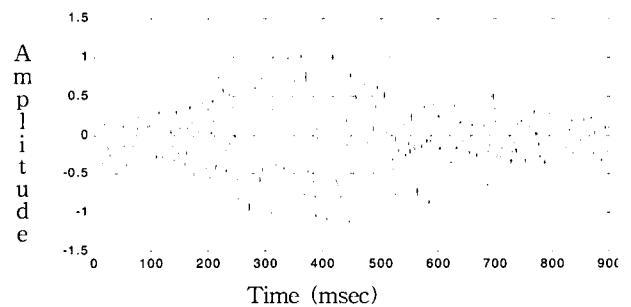
표 1. 차량소음의 종류

Table 1. Environment conditions in a car noise

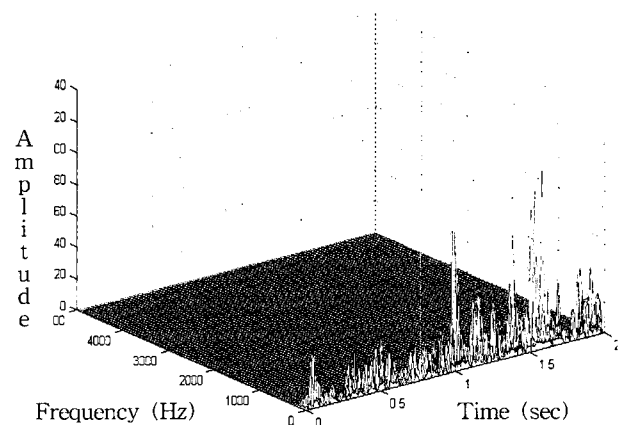
실험	차량소음의 종류
I	공회전에 발생한 차량 소음
II	공회전에 발생한 차량소음 + 에어컨 소음
III	속도 30km/h에서의 차량 소음
IV	속도 60km/h에서의 차량 소음
V	속도 60km/h + 에어컨 소음
VI	속도 60km/h + 에어컨 + 라디오 소음



(a)



(b)



(c)

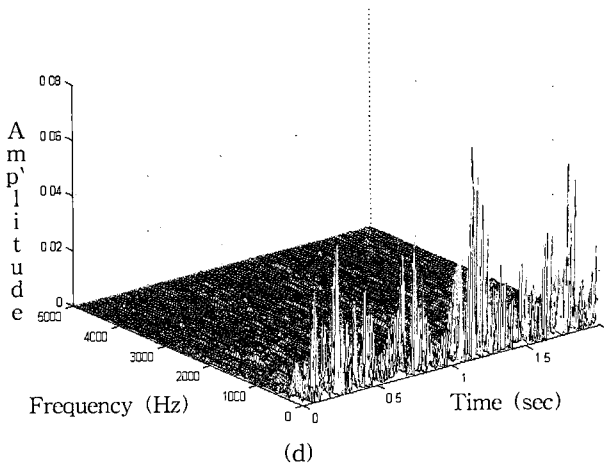


그림 6. 차량 소음의 크기 및 주파수 특성 (a) 실험조건 (VI)에서의 노이즈 크기, (b) 원 신호에 노이즈 (a)가 첨가되었을 때의 파형 (c) 노이즈 (a)에 대한 주파수 분석, (d) 실험조건 (I)에서 녹음된 노이즈의 주파수 분석

Fig. 6. The Amplitude and frequency characteristics of a car noise (a) Noise Amplitude (VI) (b) Noisy signal (c) Frequency characteristics of (a) (d) Noisy signal frequency (I))

그림 6의 (a)~(c)는 실험 (VI)인 경우의 노이즈, 노이즈가 첨가된 신호 그리고 노이즈에 대한 주파수를 나타내었고, 그림 6의 (d)는 공회전 시의 차량소음을 녹음한 실험 (I)의 주파수 분석을 나타내었다. 그림 6의 (c)와 (d)에서 알 수 있는 바와 같이 공회전 또는 주행시 발생한 차량 소음인 경우 대부분 저주파수 대역에 집중되어 있다는 것을 알 수 있었다. 위와 같은 자동차소음하에서 제안된 시스템의 성능을 평가하기 위하여 사용된 음성 데이터는 잡음이 최대로 억제된 상황에서 녹음된 우리말 숫자음 10개 /영/에서 /구/을 대상으로 실험하였다. 남성 화자 10명이 각각 3회씩 발음한 음성신호 중 2개는 학습, 또는 기준패턴을 만들기 위해서 사용하였고, 나머지 1개는 인식실험을 위해서 사용하였다. 녹음된 음성데이터의 샘플링 주파수는 10kHz이며, 기준패턴인 코드북의 사이즈는 32로 정하였다. 음성신호의 특징파라미터는 약 20ms 구간에서 음성신호가 정상(stationary)이라는 가정 아래 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 data에서 10차의 멜켵스트럼 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[1].

음성의 시작점과 끝점을 검출한 후 음성의 고주파성분을 나타내기 위하여 일반적으로 식(5)와 같은 고역통과 필터를 이용한 전처리(pre emphasis) 과정을 거치는데, 웨이블렛 기법을 이용하는 경우 이러한 전처리과정

이 음성인식률에 어느 정도의 영향을 미치는가에 대하여 알아보기 위하여 전처리가 있는 경우와 없는 경우의 인식률을 구하였으며, 웨이블렛 필터뱅크의 출력수를 4개로 선택한 후 음성인식에 최적의 필터뱅크 구조를 알아보기 위하여 옥타브 밴드 구조와 웨이블렛 패킷 구조 방식에 대해서 각각의 인식률을 알아보았다.

4.2 실험결과

일반적으로 음성신호의 주파수 스펙트럼이 6dB/oct의 기울기로 감소되는 특성 때문에 고주파성분을 나타내기 위하여 6dB/oct로 입력신호를 보상하여 주는 식 (5)와 같은 전처리과정을 하는데, 웨이블렛인 경우 선택된 이러한 전처리 과정이 인식률에 어떠한 영향을 미치는가에 대하여 알아보았다.

$$H(z) = 1 - 0.95z^{-1} \quad (5)$$

노이즈가 없는 상태에서 각 단어의 인식률 실험결과를 표 2에 나타내었는데, 표에서 알 수 있는 바와 같이 옥타브 밴드와 웨이블렛 패킷구조 모두 각각의 주파수 대역에서 전처리과정을 생략한 경우가 인식률이 높음을 알 수 있었다. 또한, 옥타브 밴드 구조보다 웨이블렛 패킷구조가 우수한 성능을 보였다.

표 2. 전처리 과정의 유·무에 따른 인식률 비교
Table 2. Recognition rate as pre-emphasis processing

[단위 : %]

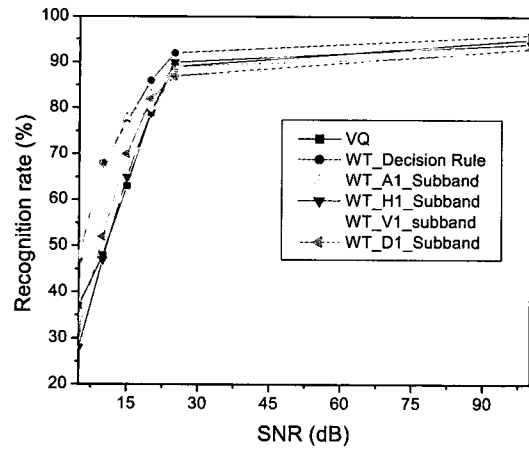
구 분	전처리 (training data)			전처리 (test data)			
	유	무		유	무		
		대역별	최종인식률		대역별	최종인식률	
VQ	99.0	99.5	99.5	95.0	94.0	94.0	
octave band	A	98.0	99.0	99.5	74.0	85.0	94.0
	D	95.5	99.5		66.0	94.0	
	D	27.0	85.0		20.0	38.0	
	D	30.5	76.0		27.0	34.0	
wavelet packet	A	99.0	99.5	99.5	93.0	92.0	96.0
	H	97.0	99.5		66.0	94.0	
	V	99.5	99.5		94.0	91.0	
	D	97.0	99.5		70.0	93.0	

위와 같이 웨이블렛을 적용한 경우 전처리 과정이 인식률을 저하시키는데, 그 이유는 웨이블렛 변환을 하기 위해 원신호에 일종의 필터인 마더웨이블렛의 크기변환과 천이과정을 필요로 하는데, 식 (5)와 같은 전처리과정을 하면 원신호가 가지고 있던 각각의 대역별 신호가 유실되기 때문이다.

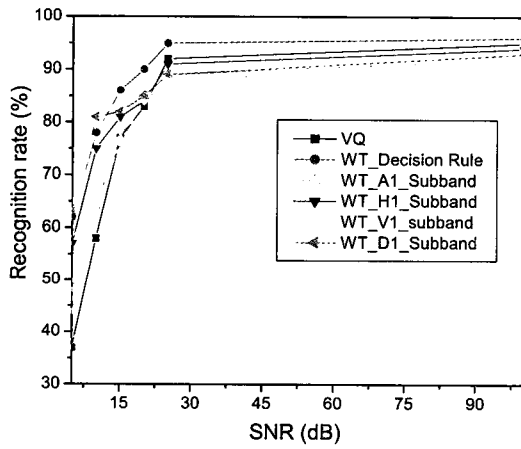
표 2에서 주목할 만한 사항은 각각의 대역별에서 출력된 인식률보다 최종적으로 각 대역별 인식단어의 소속도를 종합적으로 합한 다중의사 결정방법을 적용한 경우 인식률이 향상됨을 알 수 있었다.

그림 7에서는 본 논문에서 제안한 웨이블렛 패킷구조의 웨이블렛 기법을 적용한 경우 차량 소음의 SNR 변동

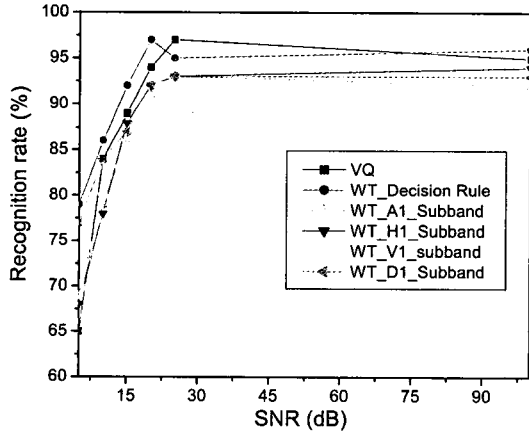
에 따른 인식률을 나타내었다. 기존의 VQ기법을 적용한 경우보다 본 논문에서 제안한 기법을 적용한 경우 인식률이 최소 9%에서 최대 25% 향상된 것을 알 수 있었다. 이와 같이 인식률이 향상된 것은 그림 7에서도 알 수 있는 바와 같이 자동차 소음이 집중된 저주파대역의 A1대역에서는 노이즈에 민감한 반응을 보인 반면 고대역의 D1대역에서는 잡음의 영향에 강인한 특성을 보이기 때문이다. 따라서, 본 논문에서 제안된 알고리즘은 보통 대역제한된 형태로 표현되어지는 주변잡음이 있는 환경 하에서 우수한 성능을 발휘함을 알 수 있었다.



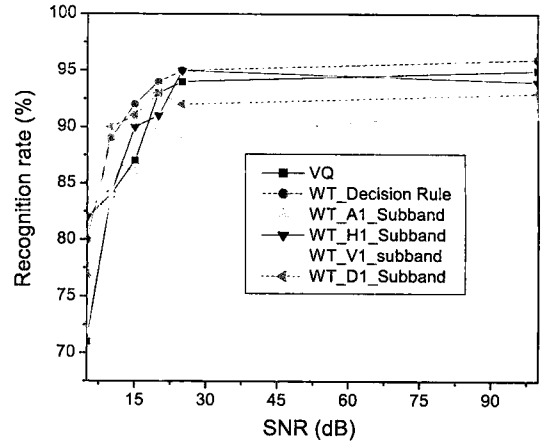
(a)



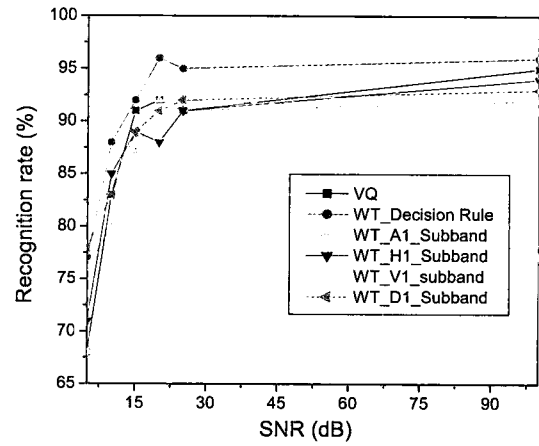
(b)



(c)



(d)



(e)

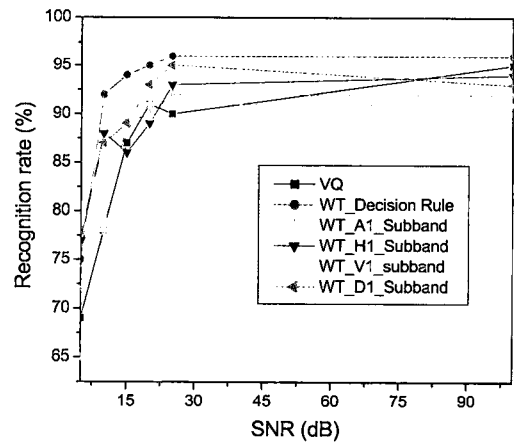


그림 7. 노이즈 변동에 따른 인식률 : (a) 실험조건(I), (b) 실험조건(II), (c) 실험조건(III), (d) 실험조건(IV), (e) 실험조건(V), (f) 실험조건(VI)
Fig. 7. Recognition rate according to the variation of a car noise

5. 결 론

본 논문에서는 잡음환경하에서 인식능력을 향상시켜 주는 새로운 음성인식 알고리즘을 구현하였다. 제안된 알고리즘은 인식하고자 하는 음성을 웨이블렛 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 타당성을 검증하기 위하여 다양한 자동차 소음을 대상으로 모의실험 한 결과, 우선 웨이블렛 필터뱅크의 출력수를 4개로 고정된 후 음성인식에 최적인 필터뱅크 구조를 알아보기 위하여 옥타브 밴드 구조와 웨이블렛 패킷 구조방식에 대해서 각각의 인식률을 알아본 결과 옥타브 밴드 구조보다 웨이블렛 패킷 구조가 모든 주파수 대역에서 우수한 성능을 발휘함을 알 수 있었다. 그리고 웨이블렛 패킷구조의 웨이블렛 기법을 적용한 결과 기존의 VQ 알고리즘만을 적용한 경우에 비하여 인식률이 최소 9%에서 최대 25% 향상된 것을 알 수 있었다. 따라서, 본 논문에서 제안된 알고리즘은 보통 대역제한된 형태로 표현되어지는 주변잡음이 있는 환경 하에서 우수한 성능을 발휘함을 알 수 있었다.

참고문헌

- [1] L. R. Rabiner and B. H. Juang, Fundamentals of speech recognition, Prentice-Hall, 1993.
- [2] B. H. Juang, "Speech recognition in adverse environments", Computer Speech and Language, vol. 5, pp. 275-294, 1991.
- [3] H. W. Ruehl, S. Dobler, J. Weith and P. Meyer, "Speech Recognition in the Noisy Car Environment", Speech Commun., vol. 10, no. 1, pp. 11-22, Feb, 1991.
- [4] D. Mansour and B. J. Juang, "The Short-time modified coherence representation and its application for noisy speech recognition", IEEE Trans. ASSP., vol. 37, no. 6, pp. 795-804, 1989.
- [5] H. Hermansky, N. Morgan and H. G. Hirsh, "Recognition of speech in additive and convolutional noise based on RASTA spectral processing", Proc. ICASSP, pp. 83-86, 1993.
- [6] T. H. Applebaun and B. A. Hanson, "Regression feature for recognition of speech on quiet and in noise", Proc., ICASSP, pp. 985-988, 1991.
- [7] D. Mansour and B. J. Juang, "A family of distortion measures based upon projection operation for robust speech recognition", IEEE, Trans., ASSP, vol. 37, no. 11, pp. 1659-1671, 1989.
- [8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE, Trans., ASSP, vol. 37, no. 2, pp. 113-120, 1979.
- [9] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (nss), hidden markov

models and the projection, for robust speech recognition in cars", Speech Commn., vol. 11, pp. 215-228, 1992.

- [10] O. Ghizta, "Auditory nerve representation as a front end for speech recognition in a noisy environment", Computer, Speech and Language, vol. 1, pp. 109-130, 1986.
- [11] Stephane Mallat, A wavelet tour of signal processing, Academic press, 1999.
- [12] Gowdy, J. N. and Tufekcim, Z. "Mel-scaled discrete wavelet coefficients for speech recognition", Proc., ICASSP, vol. 3, pp. 1351-1354, 2000.
- [13] Yu Hao and Xiaoyan Zhu, "A new feature in speech recognition based on wavelet transform", Proc., ICASSP, vol. 3, pp. 1526-1529, 2000.
- [14] Kidae Kim, Dae Hee Youn, Chul hee Lee, "Evaluation of wavelet filters for speech recognition", Systems, Mans, and Cybernetics, vol. 4, pp. 2891-2894, 2000.

저 자 소 개



곽 근 창(Keun Chang Kwak)

1996년 : 충북대학교 전기공학과(학사)
1998년 : 충북대학교 전기공학과(공학석사)
1998년~현재 : 충북대학교 전기공학과 박사 과정

관심분야 : 지능제어, 패턴분류, 얼굴인식



이 대 종(Dae Jong Lee)

1995년 : 충북대학교 전기공학과(학사)
1997년 : 충북대학교 전기공학과(공학석사)
1998년~현재 : 충북대학교 전기공학과박사 과정 전기안전시험연구원

관심분야 : 음성신호처리, 얼굴인식



유 정 응(Jeong Woong Ryu)

1965년 : 한양대학교 전기공학과(학사)
1976년 : 단국대학교 전기공학과(공학석사)
1987년 : 단국대학교 전기공학과(공학박사)
1969년~1979년 : 대전개방대학 교수
1979년~현재 : 충북대학교 전기전자 및 컴퓨터공학부 컴퓨터 정보통신연구소 교수

관심분야 : 지능제어, 가변구조제어, QFT제어



전 명 근(Myung Geun Chun)

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

E-mail : mgchun@Chungbuk.ac.kr