한국지능시스템학회 논문지 2010, Vol. 20, No. 2, pp. 208-213

응급상황에서의 음성인식을 위한 필터기 구현

Implementation of Speech Recognition Filtering at Emergency

조영임 · 장성순
Young Im Cho and Sung Soon Jang

수원대학교 IT대학 컴퓨터학과
E-mail: ycho@suwon.ac.kr · veteranus@paran.com

요 약

일반적으로 음성인식 시스템의 사용에 가장 저해되는 요소에는 배경 잡음을 들 수 있다. 잡음은 음성인식 시스템의 성능을 저하시키고, 이로 인해 사용 장소의 제약을 많이 받게 되는 이유가 된다. 이런 잡음의 영향을 해결하기 위해 본 논문에서는 음질 향상에 목적을 두고 신호단계에서부터 잡음성분을 제거하는 필터 중 FIR필터의 대역통과를 이용하여 일반적으로 사람의 음성 주파수 영역과 잡음 영역을 추출한 정보를 토대로 Wiener 필터를 구현, 그 성능을 향상하여, 전송되어지는 음성신호구간에서 잡음구간과 음성구간에 따라 잡음을 유연하게 처리하도록 구현하였다.

키워드 : 응급상황, Wiener 필터, 잡음제거

Abstract

Generally, the mal factor for speech recognition is the background noise in speech recognition. The noise is the reason to reduce the speech recognition performance. Owing to the fact, the place to recognize is very important. To improve the recognition performance from the sound having noise, we implemented the noise filtered Wiener filter at the signal process step which adopted the FIR filter. In FIR filter, it deal with the filtered speech signal which is appropriate frequency range of human speech frequency range. Therefore, we make the recognition system distinguish between noise and speech sound from the incoming speech signal.

Key Word : Emergency, Wiener filter, Noise filtering

1. 서 론

음성은 인간이 사용하는 가장 보편적인 수단이며, 유비쿼터스 시대의 도래와 함께, 각종 정보기기들과 인간과의 통신을 보다 효율적으로 이루어지게 되는 기술로 연구되어 지고 있다[1,2].

그러나 일반적으로 실내 환경만 아니라 실외 환경에서 발생할 수 있는 외부환경의 경우 주변에 소음이 생기는 잡음환경에 처해 있으며, 응급 상황이 발생 시에 잡음으로 인하여 제한된 환경에서 음성인식시스템의 성능보다 크게 저하되는 문제점이 발생된다.

이러한 문제점은 인식 시스템이 학습된 환경과 실제로 인식 시스템이 구현되는 환경에서의 음성 정보가 가지는 특성의 차이에서 오는 것이다. 마이크의 특성, 주변의 소음, 거리상의 문제 등 다양한 요소들이 인식 성능을 낮추게 된다. 그 중에 주변의 소음은 자동차 소음, 주위 사람들에 의한 잡음, 거리에서 일상적으로 나오는 잡음 등 다양한 형태로 발생하여, 인식 시스템에서 인식해야 하는 음성에 합쳐져 인식 시스템의 정확성을 떨어뜨리며, 잘못된 인식 결과를 가져오게 하는 문제점을 가지고 있다.

이처럼 음성인식 시스템의 성능은 학습환경과 구현환경에서의 차이에서 온다. 그리고 그 차이는 신호단계(signal process), 특징벡터단계(feature space process), 모델단계(model process) 으로 구분되는데, 신호단계에서의 차이가 음성인식의 큰 차이를 발생하고 있다[3,4].

따라서 본 논문에서는 처음 단계인 신호단계에서 들어오는 음성정보를 MATLAB[5]을 이용하여 디지털 필터를 구현하여, 잡음 제거를 하는 것을 연구 목적으로 고려하였으며, 이를 위하여, 일차적으로 FIR 필터의 대역통과 특징을 이용하여 음성구간과 잡음구간을 구한 후 Wiener 필터를 구현 및 적용하여 음성인식 시스템의 성능을 개선하는 것을 목표로 하고자 한다. 2장에서는 구축된 음성인식 시스템과 필터링 시스템에 관한 설명을 하고, 3장에서는 제안하는 잡음제거 필터링에 관한 것을 논할 것이다. 그리고 4장에서는 실험결과에 대한 고찰과 5장에서는 결론을 서술 할 것이다.

접수일자 : 2009년 11월 30일
완료일자 : 2010년 3월 31일
본 논문은 경기도지역협력연구센터 사업의 일환으로 수행되었음(GRRC 수원 2009-B3, u-City 보안감시기술협력센터).

2. 음성인식 시스템

본 논문에서 기 구축한 음성인식 시스템은 플랫 렉시콘과 렉시컬 트리를 기반으로 하여 빠른 인식속도 요구하는 상황인지가 가능하도록 설계하여, 응급상황 발생시에 효과

적인 대처가 가능하도록 CCTV 환경을 구축하였다[6]

기본적으로 음성인식 시스템은 그림 1에서와 같이 총 6단계에 걸쳐 구성된다. 1단계는 음성신호를 전기신호로 변환하여 디지털화하여 전송하는 음성입력 단계이며, 2단계는 주위 잡음을 제거하고 음성신호를 분리하여 음성이 있는 구간을 찾아내게 되는 전처리 단계이다. 3단계는 음성인지모델을 통하여 음성인식에 유용한 특징을 뽑아내는 특징추출 단계이며, 4단계는 음성 인식 훈련 과정으로 표준 패턴 DB를 생성하는 단계이다. 5단계는 미리 생성된 기준패턴과 입력되는 음성을 비교하여 가장 비슷한 것을 인식결과로 결정하는 음향모델 단계인 탐색과정이다. 마지막으로는 이러한 인식결과를 원하는 응용에 적용하여 사용자 인터페이스 기술을 이용하게 되는 단계이다.

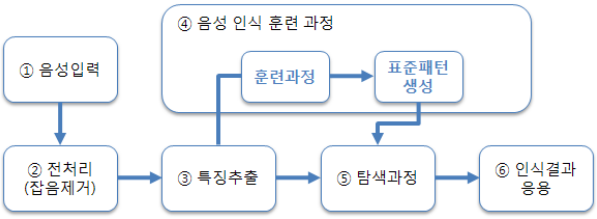

그림 1. 일반적인 음성인식 시스템 구조
Fig. 1. The general structure of speech recognition system

그림 1에서 ②전처리(잡음제거) 과정에서는 CCTV나 기타 장치에 전송되어 오는 아날로그 음성신호를 디지털 음성신호로 변환 후 디지털 필터(Digital filter)를 사용한다. 디지털 필터는 신호에 대한 잡음 특성, 성능, 시스템 구성의 용이성 등으로 인하여 널리 사용되고 있다. 디지털 음성신호에서 필터란 들어오는 신호 입력에 대해 특정 주파수 대역에 대해서는 출력으로 내보내지 않고 차단시키는 회로를 의미한다. 즉, 어떤 음성 신호에 대한 스펙트럼을 원하는 주파수 대역만큼 제한시키는 주파수 선택회로를 뜻한다. 여기서 원하는 주파수 대역은 통과대역(passband)이 되고 원치 않는 대역은 차단대역(stopband)이 된다.

이러한 필터 중 피드백 과정의 유무에 따라 IIR(infinite impulse response) 와 FIR(finite impulse response)으로 구분되며 오차의 영향이 적은 FIR 필터를 선택하였으며 이러한 필터를 거친 후 최종적으로 잡음제거를 위한 필터링으로 Wiener 필터, 칼만필터[7] 등을 많이 사용하고 있다. 이 중 응급상황은 짧은 음성 신호 구간 내에서 상황을 인지해야 하는 상황이 발생하므로, 일반적으로 짧은 구간에서 정상성의 주파수 스펙트럼을 가진다는 가정 하에 음성 정보를 추정하는 Wiener 필를 이용하여 시스템을 많이 사용하고 있다. 모델기반 Wiener 필터의 구성도는 다음 그림 2와 같다[8,9].

일반적인 모델 기반 Wiener 필터의 구현 과정은 구하고자 하는 음성을 $\hat{s}(t)$ 라고 하고, $s(t)$와 $n(t)$를 각각 잡음이 포함된 음성과 잡음이라 하고, Wiener 필터를 $g(t)$

일반적인 모델 기반 Wiener 필터의 구현 과정은 구하고자 하는 음성을 $\hat{s}(t)$ 라고 하고, $s(t)$와 $n(t)$를 각각 잡음이 포함된 음성과 잡음이라 하고, Wiener 필터를 $g(t)$라 하면,

$$\hat{s}(t) = g(t) * (s(t) + n(t)) \quad (1)$$

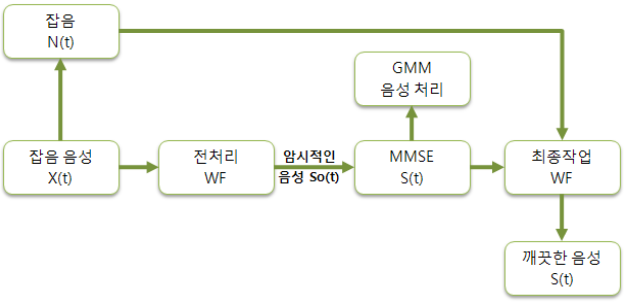

그림 2. 모델 기반 Wiener 필터의 구성도
Fig. 2. The structure of model based Wiener filter

식(1)처럼 $\hat{s}(t)$ 를 구하고자 하는 것이다. 이때, $s(t)$로부터 N(t)의 추정치를 구하고 이것을 이용해 $\hat{s}(t)$ 의 근사치를 얻는다는 것이다. 또한 $\hat{s}(t)$ 에 더 가까운 근사치를 얻기 위해 음성의 보편적인 특성을 나타내는 GMM을 이용하는데, 이것은 식(2)로 표현된다.

$$P(s) = \sum_{k}^{K} p(k) N(s; \mu_k; \sum k) \quad (2)$$

식(2) 의 가정으로부터 모델 기반 Wiener 필터는 아래 순서로 설계된다.

① 입력된 현재의 프레임에서 통계기반 VAD를 이용해 잡음구간을 판별하고 잡음구간이면 잡음모델을 이전 값에서 갱신한다.

② Decision-directed Wiener 필터를 이용해 전처리-WF 블록에서 임시적인 깨끗한 음성을 추정한다.

③ 앞의 과정에서 얻어진 추정치를 이용해 가지고 있는 GMM의 각 Gaussian에 대한 사후확률을 계산하고, 이것을 이용해 MMSE 방법에 따라 최종 작업 WF 후 깨끗한 음성을 추정한다.

④ 추정된 깨끗한 음성과 ①에서 얻은 잡음 모델을 이용해 최종적인 Wiener 필터를 설계한다.

⑤ 얻어진 Wiener 필터로 현재 프레임을 처리하여 깨끗한 음성을 만들고, 다음 프레임은 단계 ①부터 위의 과정을 반복해서 처리한다.

이러한 과정을 거쳐 나온 음성만을 사용하여 그림 1에서 이후 과정인 ③특징추출, ④음성 인식 훈련과정, ⑤탐색과정, ⑥인식결과 응용을 거쳐서 음성인식 과정을 거치도록 된다.

3. 향상된 잡음제거 필터

이절에서는 본 논문에서 제안하는 향상된 잡음제거필터 방법에 대해 소개하고자 한다. 본 논문에서 제안하는 방식은, CCTV에서 전송되는 모든 소리를 음성인식에 사용되지 않으며, 기본적으로 음성에 필요한 에너지를 가지고 있는 소리 정보에 대해서 감지를 하여, 이를 인식에 사용되는 디지털 신호로 저장하는 과정을 거치도록 설계하고자 한다.

이렇게 제안하게 된 동기는, 아날로그 신호인 음성을 음성인식 시스템에 적용하기 전에 불필요한 잡음 등 인식에 필요하지 않는 신호를 제거하기 위해 디지털 필계를 고려하였으며, 최종적으로 성능이 우수한 FIR Wiener 필터를 선택하여 적용하고자 한다.

사람의 음성이 300-3400khz에 집중되어 있다는 점에 착안하여 전송되는 음성데이터를 FIR 필터[11,12]의 특징을 이용하여 통과대역(본 논문에서는 음성 발화 구간), 저지대역, 천이대역을 결정하여, 그 후 과정에서의 소요시간을 줄이며 전반적인 성능향상을 얻고자 한 점이 특징이다.

다음 식(4)는 기본적인 FIR 필터의 유도식이다.

$$y[x]=\sum_{k=0}^{N-1}h[k]x[n-k] \tag{4}$$

여기에서 $x[n]$과 $y[n]$은 입력되는 음성 정보 및 필터링 후 출력 음성정보를 나타내며, $h[n]$은 필터의 유한 충격 응답(Finite impulse Response) 특성이며, N은 필터의 차수를 의미한다. 그러나 위 식으로 FIR 필터가 구현되면 입력되는 정보와 계수들의 곱해진 후 한꺼번에 더해지는 과정을 거쳐야 하므로 잡음제거에 소요되는 처리속도를 만족하기 힘들기 때문에 곱셈을 과정을 제거하기 위하여 식(4)에 bit-serial 알고리즘[13]을 적용하여 식(5)와 같이 표현된다.

$$y[x]=\sum_{k=0}^{N-1}(\sum_{j=0}^{M-1}h_j[k]\cdot 2j)x[n-k] \tag{5}$$

여기서 h_j,N,M,은 각각 계수 h의 j번째 비트, 탭수, 계수 비트수를 나타낸다. bit-serial 알고리즘은 승수의 LSB로부터 MSB로 쉬프트 시키면서 피승수를 곱한 결과에 그전에 계산된 부분곱을 누적시키는 방법이다. 곱셈 연산을 위한 총 계산되는 사이클 수를 줄이기 위해서 식(5)의 짝수 부분과 홀수 부분에 대해서 나누어 bit-serial 알고리즘을 적용하면 식(6)과 같이 표현된다.

$$y[x]=\sum_{k=0}^{N-1}\sum_{j=0}^{\frac{M}{2}-1}(h_{2j}[k]\cdot 2^{2j}+h_{2j}[k]\cdot 2^{2j+1})x[n-k] \tag{6}$$

식(5)에서 적용된 필터식은 NM 사이클을 요구하나 제안되는 식(6)의 알고리즘은 $\frac{NM}{2}$ 사이클이 걸리므로 산술적으로 2배의 속도 향상 효과를 얻을 수 있다.

이렇게 이러한 FIR 필터의 특성을 고려하여 음성신호와 잡음을 효과적으로 구분하여 예측되는 희망 출력과의 오차를 최소로 하는 Wiener 필터링을 거치게 하였다.

그 이후에 원본 데이터와 추출한 데이터의 차이를 구하여 노이즈 신호를 추출하는 과정을 거친 후, 추출된 노이즈 신호와 원본 데이터를 이용하여 식(1)의 Wiener 필터를 적용하여 노이즈 제거를 위한 필터를 설계하였다.

그림 3에서 보여주는 바와 같이 입력되는 음형 파형을 음성의 주파수 대역을 구분하여 FIR 필터를 적용하여 음성 발화구간만을 구분하여 전반적으로 Wiener 필터를 거치는 시간을 줄이는 효과를 얻을 수 있다.

일반적인 Wiener 필터도 수학적 표현을 보면 현재 및 과거(즉 시간 지연된)의 데이터와 필터 계수들과의 곱셈과 덧셈으로 이루어져 있으며, 이들 소자들의 전달 함수와 수학적 표현식들로 설계가 가능하다, 현 연구 목표 안 에서는 물리적인 상황들(동작 안정성, 감도, 전송되는 데이터의 안정성)을 일차적으로 차이가 없다고 가정한 상태에서 고려한다면 동작시간이 빠르거나 적은 개수의 소자를 사용하여 필터의 동작 시간을 적게 하는 것이 바람직하다고 할 수 있다.

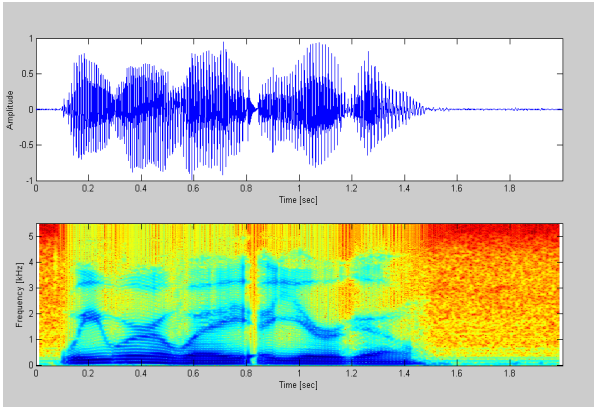

그림 3. FIR 필터링 적용 음성의 진폭/진동
Fig. 3. The frequency/oscillation of FIR filtering

최종적으로 잡음을 제거하는 Wiener 필터는 아래와 같이 수행된다.

$$So(w)=H(w)S(w) \tag{6}$$

식(6)처럼 기본 Wiener 필터의 경우 잡음이 포함된 음성신호 $S(w)$와 잡음을 제거한 음성신호 $So(w)$, Wiener 필터의 추정함수 $H(w)$를 이용하여 얼마나 효과적으로 $H(w)$를 구하는 것인가가 주요 연구 목표이다. 본 연구에서는 이를 위하여 다음 식(7)과 같이 제안하는바이다.

$$H(w)=\frac{P_s(w)}{P_s(w)+P_d(w)} \tag{7}$$

식(7)에서 $Ps(w)$는 원음성 신호의 음성 스펙트럼을 나타내며, $Pd(w)$는 잡음 신호의 음성 스펙트럼을 뜻한다. 이처럼 필터링 과정을 거치면 원음성 신호의 음성 스펙트럼을 추정하는데 오차가 생기게 되며, 이 오차를 줄이기 위해, 계수를 먼저 곱한 뒤에 시간 지연에 따른 계산을 하는 것을 고려한 아래와 같은 식을 제안하는 것이다.

$$H(w)=\left(\frac{P_s(w)}{P_s(w)+\alpha\cdot P_d(w)}\right)^{\beta} \tag{8}$$

파라메터 값 α,β을 이용하여 각 신호들의 평균의 제곱형태로 계산하여 그 오차값을 줄이는 방안을 도입한 것이다.

잡음이 포함된 음성 정보의 처리를 위하여 Wiener 필터 과정을 거치지만, 이러한 과정으로 시간 지연이 발생하여 목적에 부합되지 못하기 때문에, 이러한 문제점을 해결하기 위하여, 본 논문에서는 식(8)에서 제시한 식을 그림 2에서 제시된 모델 기반 Wiener 필터의 단계 ①에서 통계기반의 VAD[10]를 이용시에 최적화를 통하여, 시간 지연과 잡음 제거의 성능간의 관계를 고려한 식 (9)로 수정 제안한 후 설계하고자 한다. 비대칭 윈도우의 적용으로 잡음 제거 시에 소요되는 시간을 최소화하기 위한 것이다. 일반적으로 잡음의 통계량은 그 자체로 정상적이라고 판단이 가능하지만, 음성 정보의 통계량과 비교한다면 차이를 알 수 있을 정도로 비정상을 확인 할 수 있으므로, 그 차이를 구분하여 Wiener 필터의 최적화를 고려하였다.

$$H(n)=\begin{cases}0.54-0.46\cos\left(\dfrac{2\pi n}{P_1}\right), & 0 \le n < n_0\\ \cos\left(\dfrac{2\pi(n-n_0)}{P_2}\right), & n_0 \le n < N\end{cases} \quad (9)$$

식 (9)에서 P_1,P_2,는 비대칭 창함수의 왼쪽 및 오른쪽 부분을 나타내기 위한 주기값이며, n_0 및 N은 최대치가 존재하는 위치 및 창함수 전체의 길이를 나타낸다.

이렇게 잡음이 제거된 음성신호를 바탕으로 구축된 음성인식 DB를 바탕으로 응급상황 인식 및 탐지에 사용되게 된다. 이 때 기본적으로 음소를 기반으로 단어를 인식하고자 DB를 구축하는 것을 기본으로 구축하였다.

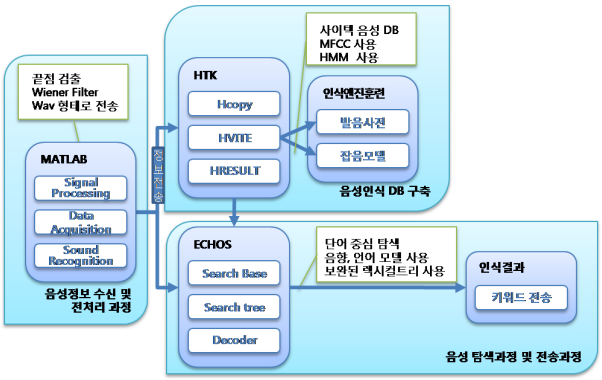

그림 4. 본 논문에서 구축된 음성인식 시스템 개요
Fig. 4. The overview of proposed speech recognition system

그림 4에서는 그림 1의 음성인식 시스템의 기본구조에서 제안하는 방법을 적용한 본 논문에서 구축된 음성인식 시스템의 전반적인 구조를 나타낸다. 제안된 FIR Wiener 필터의 경우는 MATLAB으로 구현하였으며, 그 이후 과정의 경우는 HTK와 ECHOS를 사용되었다. 구축된 음성인식 시스템의 특징은 음향모델을 중심으로 하여 단어(keyword)을 탐지하는 것을 우선으로 선정하였으며, 이러한 단어 중심의 인식 시스템에서는 플랫 렉시콘, 렉시컬 트리를 이용한다. 렉시컬 트리는 메모리 사용은 효율적이지만, 언어모델 확률값의 적용 지연과 단어간 모델링 구현의 복잡성이 존재하기 때문에, 트리 복사 알고리즘을 구현하였다. 이 음성인식 시스템에서는 렉시컬 트리가 가지고 있는 단점을 보완하기 위해 단일 음소로 이루어진 단어에 대해서는 렉시컬 트리를 구성할 때, 별도의 병렬적인 구조를 갖도록 설계하여 문제점을 해결하였다.

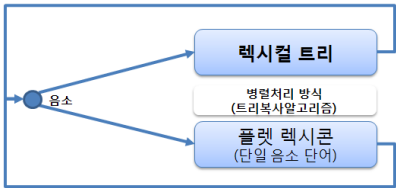

그림 5. 렉시컬 트리에 의한 문제점 해결 전략
Fig. 5. The solution strategy of lexical tree problem

이러한 일련의 과정을 통하여 인식된 결과는 사용자 인터페이스로 전송하게 구축되어 있다.

4. 시뮬레이션 및 결과 분석

음성인식에서의 성능을 평가하기 위해 SITEC에서 제작한 스피치 단어 음성 데이터베이스를 사용하였으며, !6kHz/16 bit로 녹음되어있으며, 음향 모델 학습에 500명 인원의 음성을 사용하였다. 추후 비교 음성은 !6kHz/16 bit의 형태로 마이크 혹은 CCTV에서 전송되는 데이터를 기준으로 사용하였다.

또한, 구축된 시스템에서 수집되는 모든 음성정보를 음성인식에 이용하는 것은 이미 서론에서 언급된 사실과 같이 부정확한 결과를 가져오며, 소요 시간 또한 급격히 증가 되는 것을 알 수 있다. 3절에서 제안하는 FIR 필터를 이용하여 음성인식에 필요한 음성이라 추정되는 주파수 영역의 음성 정보만을 일차적으로 추출하는 고정을 거치는 과정을 통하여 전반적인 처리 시간을 줄이는 효과를 얻을 수 있다.

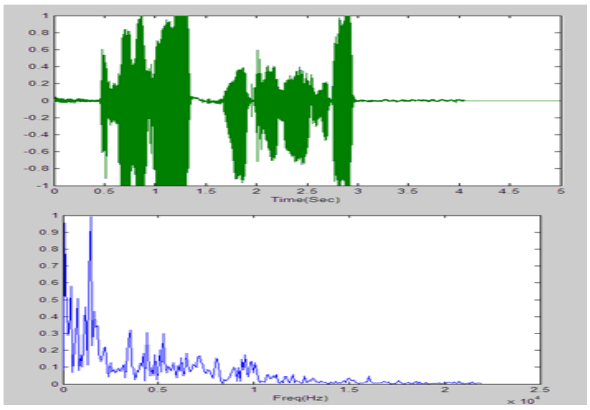

그림 6. FIR Wiener 필터링 이전 음성 정보
Fig. 6. The speech information before using FIR filtering

일차적인 필터링 과정 이후에는 제거하지 못하는 잡음의 경우는 Wiener 필터를 통하여 최종적으로 잡음을 제거하는 과정을 거치며, 그림 6는 FIR 필터의 통과대역을 거쳐서 음성이라고 추정되는 음성정보의 스펙트럼과 주파수 영역을 보여주고 있다. 이 음성 정보를 FIR 필터링의 결과를 이용하여 잡음 정보를 추출하고 이것을 Wiener 필터를 그림 7과 같이 구현하였다.

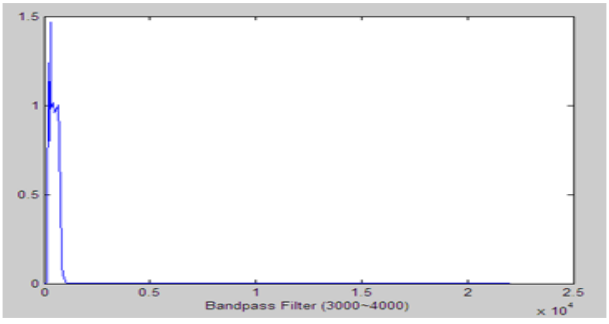

그림 7. FIR Wiener 필터 구현 파형
Fig. 7. The waveform of FIR Wiener filter

이렇게 얻어진 Wiener 필터를 적용하여 그림 8과 같은 음성 정보를 최종적으로 얻을 수 있다.

FIR Wiener 필터를 적용하여 배경 잡음을 어느 정도 제거됨을 알 수 있다. 그림 7와 그림 8의 파형을 스펙트럼상에서 육안으로 비교 관찰 결과를 알아 볼 수 있으며, 필터 적용 전/후의 음원을 청취한 결과 잡음 제거 효과를 알 수 있었다.

이렇게 저장된 음성정보들만을 이용하여 Wiener 필터를 MATLAB에서 구현하여 그림 1에서와 같이 음성 정보에 포함된 잡음을 제거하는 과정을 통하여, 정확한 인식효과를 가져 올 수 있다.

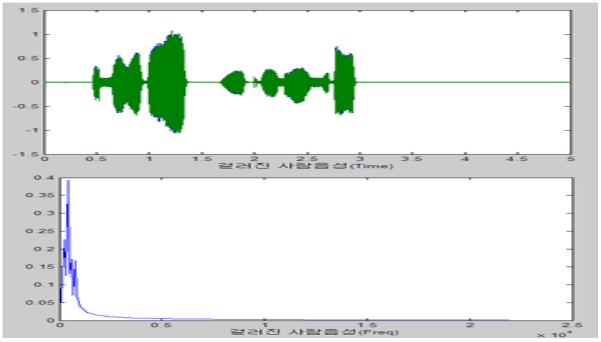

그림 8. FIR Wiener 필터링 이후 음성 정보
Fig. 8. The speech information after using FIR filtering

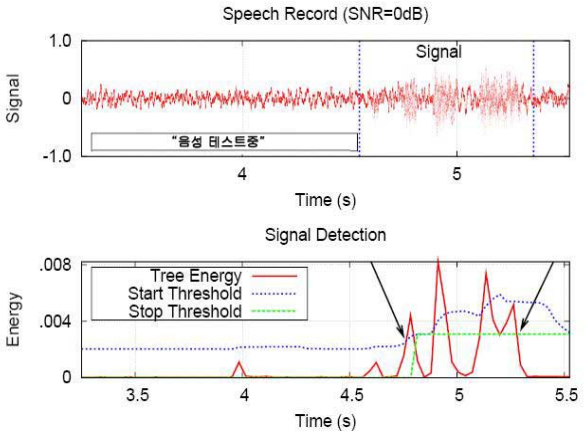

그림 9. 음성 감지 테스트
Fig. 9. Signal detection test

이와 같이 필터링 전 음성 정보들의 불필요한 부분, 즉 인식 결과에 큰 영향을 끼치지 못하는 음성의 부분이나 소리가 발생시에 일어나는 각종 잡음, 혹은 CCTV에서 정보를 전송하면서 생기는 각종 잡음들을 필터링을 하여, 인식률의 상승효과를 가져 올 수 있는 결과를 확인 할 수 있다. 또한, 단어 모델의 사용 여부가 인식률과 인식 시간 모두에 큰 영향을 미치고 있음을 알 수 있었다. 즉, 탐색하는 단어들의 연관관계를 고려하는 단어 모델을 사용하는 경우 그렇지 않는 경우보다 보다 나은 인식률을 가져오는 것을 알 수 있었다.

잡음이 제거된 음원을 사용하여 구축된 음성인식 DB를 이용하여 ECHOS를 이용하여 two-pass 탐색 시에 정방향 탐색(bigram)시와 정방향과 역방향(trigram) 시와의 비교결과, 다음 표1와 같이 단어 모델의 사용 여부가 인식률과 인식 시간 모두에 큰 영향을 미치고 있음을 알 수 있었다. 즉, 탐색하는 단어들의 연관관계를 고려하는 단어 모델을 사용하는 경우 그렇지 않는 경우보다 보다 나은 인식률을 가져오는 것을 알 수 있었다.

또한, 이와 같은 결과를 바탕으로 HMM의 기본적인 탐색방법(좌에서 우로)을 따르는 정방향 탐색 시보다는 음소들 간의 관계를 고려하는 역방향 탐색이 단어 인식률이 8% 이상 상승으로 단어의 인식 시에 음소들 간의 관계를 고려하여 시스템을 구축하여 탐색하는 것이 정확한 인식률을 얻을 수 있음을 알 수 있었다. 그러나 음성인식 DB에서 가장 일치하는 단어열을 찾는 확률이 높아지는 것이 결과적으로는 탐색 시간의 증가도 가져오고 있음을 알 수 있었다.

CCTV에서 전송되는 모든 소리를 음성인식 탐색 구간에서 이용하지 않고, FIR Wiener 필터를 이용하여 음성발화 구간과 잡음이 제거된 구간만을 선택하여 상황인지를 위한 프로세싱 타임의 감소와 함께, 보다 빠른 상황 대처가 가능하게 된다.

표 1. 음성인식 시스템의 인식 결과
Table 1. The recognition result of proposed system

탐색방법	단어간 모델 사용	단어 인식률 (%)	인식 시간 (sec/문장)
정방향	X	77.2	5.4
정방향 + 역방향	X	80.1	6.3
정방향	O	88.9	21.0
정방향 + 역방향	O	90.0	22.1

5. 결 론

제한된 환경 내 에서 처리되는 음성 인식 시스템에서는 노이즈 제거 필터의 처리가 용이하지만, 본 논문에서 고려하는 CCTV 환경처럼 많은 정보가 수집되는 환경에서는 수집되는 음성 정보에 잡음의 포함이 클 수밖에 없으며, 이 잡음을 효과적으로 제거하지 못한다면, 음성인식 시스템의 성능에 큰 영향을 가져올 수 있다는 사실을 알 수 있었다.

또한, 노이즈 제거 과정에서 소요 시간이 오래 걸린다면 응급 상황에서 이를 효과적으로 이용할 수 없으므로, 소요 시간과 효과적인 노이즈 제거가 필요한 이유이다. 이를 위하여, FIR 필터와 Wiener 필터를 적용하여, 수집된 음성 정보 중에 효과적으로 잡음 제거하여, 깨끗한 음성 정보를 가지고 음성 인식에 이용될 수 있는 과정이 필요하다는 것이다. 이처럼 잡음제거 시에 걸리는 시간 단축 뿐 만 아니라, 인식에 필요한 음성 정보의 오류를 적게 하는 것이 중요하기에, FIR 필터와 Wiener 필터의 적용이 효과적으로 작동함을 알 수 있었다.

그러나 본 연구에서 확인된 문제점으로는 주변 잡음 중에서도 3절에서 고려한 사람의 일반적인 음성대역인 300-3400khz에서도 발생할 수 있다는 사실이다. 이러한 배경 잡음이 포함되는 경우라면 FIR 필터를 사용 시에도 음

성영역의 정보만을 추출하는 과정이 확실치 않으며, 또한 잡음제거가 확실치 않다는 단점을 알 수 있다. 추후 연구에서는 이러한 부분에 대한 수정과 함께, 예상치 못하는 잡음의 제거를 위한 연구를 할 것이다.

참 고 문 헌

[1] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, "Toward Conversational Human-Computer Interaction", *AI Magazine*, vol. 22, no. 4, pp.27-37, 2001.

[2] H. Kruegle, "CCTV Surveillance", *Analog and Digital Video Practices and Technology*, Elsevier, pp.227-239, 2007.

[3] Y.Gong, "Speech Recognition in Noisy Environments: A Survey", *Speech Communication,* vol.16, no.3, pp.261-291 1995.

[4] C.-H.Lee, "On Stochastic Feature and Model Compensation Approaches to Robust", Speech Recognition, *Speech Communication,* pp.29-47. 1998.

[5] 김경수, *MATLAB 신호처리 및 이미지처리*, 아진, pp.213-250, 2007

[6] 조영임, 장성순, "CCTV 응급상황에 따른 지능형 음성인식 시스템 구현 ", *한국지능시스템학회 논문지*, 제19권, 제3호, pp.415-420, 2009.

[7] 김지곤, "최소 최대 추정량과 베이즈 추정량으로서의 Kalman 필터에 관하여", *자연과학연구,* 제5호, pp.21-30, 1995.

[8] 강점자, 강병옥, 정호영, 정훈, 이윤근, 신성장동력 산엉용 대어휘 음성인식 기술 및 응용. *전자통신동향분석*, vol. 23, no. 1, pp 70-76, 2008.

[9] Doclo, S., Rong Dong, Klasen, T.J., Wouters, J., Haykin, S., Moonen, M., "Extension of the multi-channel Wiener filter with ITD cues for noise reduction in binaural hearing aids", *Applications of Signal Processing to Audio and Acoustics,* vol. 16, no. 16, pp 70-73, 2005.

[10] 장준혁, 김동국, 김남수, "음성검출기의 설계에 있어 새로운 통계모델과 접근방법", *Telecommu nications Review*, 제15권, 1호, pp. 201-209, 2005.

[11] 류탁기, 박구현, 홍대식, 강찬언, "주파수 영역에서의 간단한 zero-forcing 기법을 이용한 속도 적응형 채널 추정기법", *한국통신학회논문지*, 제31권, 제1호, pp.38-47, 2006

[12] 박윤식, 장준혁, " 주파수영역에서 Soft Decision 기반의 음향학적 반향 제거", *Telecommunications Review*, 19권, 5호, pp.837-844, 2009.

[13] Robert E.Morley, Jr. Gray E. Christensen, Thomas J. Sullivan, Orly Kamin, " The Design of a Bit-Serial Coprocessor to Perform Mulitiplication and Divison on a Massively Parallel Architecture", *in Proc IEEE, The 2nd Syposium on the Frontiers of Massively Parallel Computation, Farifax, U.S.A*, pp.419-422, 1998

저 자 소 개

조영임(Young Im Cho)
1988년 : 고려대학교 컴퓨터학과 학사
1990년 : 고려대학교 컴퓨터학과 석사
1994년 : 고려대학교 컴퓨터학과 박사
현재 : 수원대학교 컴퓨터학과 교수

관심분야 : 뉴로-퍼지시스템, 에이전트, 상황인지, u-City
E-mail : ycho@suwon.ac.kr

장성순(Jang Sung Soon)
2008 : 수원대학교 컴퓨터학과 학사
현재 : 수원대학교 컴퓨터학과 석사과정

관심분야 : 인공지능, 정보검색, 음성인식, 유비쿼터스 시스템