잡음에 강한 특징 벡터 및 스펙트럼 차감법을 이용한 음성 인식

Speech Recognition Using Noise Robust Features and Spectral Subtraction

신 원 호*, 양 태 영*, 김 원 구**, 윤 대 희*, 서 영 주***
(Won-Ho Shin*, Tae-Young Yang*, Weon-Goo Kim**, Dae-Hee Youn*, Young-Joo Seo***)

※본 논문은 전자통신연구소의 1995년도 수탁과제연구 지원에 의해 수행되었습니다.

요 약

본 논문에서는 잡음 및 주변 환경에 강인한 것으로 알려져 있는 특징 벡터들을 이용한 인식 성능을 비교하였다. 아울러 스펙트럼 차감법을 적용하여 높은 인식 성능을 얻도록 하였다. 본 논문에서는 환경 변화에 강인한 인식 성능을 얻기 위하여 SMC(Short time Modified Coherence) 분석, 루트(root) 켑스트럼 분석, LDA(Linear Discriminant Analysis), PLP(Perceptual Linear Prediction), RASTA(RelAtive SpecTrAl) 처리 등을 이용하여 인식 실험을 수행하였다. 실험을 위하여 반연속 HMM을 이용한 단독음 인식 시스템을 구현하였고 전시장 및 컴퓨터실의 잡음을 첨가하여 0, 10 및 20dB의 SNR에 대한 인식 실험을 수행하였다. 실험 결과, LPCC(Linear Prediction Cepstral Coefficient)를 이용한 경우에 비하여 SMC나 루트처리를 이용한 멜 켑스트럼(루트_멜 켑스트럼)을 이용한 경우 10dB의 SNR에서 각각 9.86%, 12.68% 향상된 가장 좋은 인식률을 얻었다. 또한 멜 켑스트럼과 루트_멜 켑스트럼을 스펙트럼 차감법과 결합하여 잡음을 제거한 경우 10dB에서 각각 16.7%, 8.4% 향상된 94.91%, 94.28%의 인식률을 얻을 수 있었다.

ABSTRACT

This paper compares the recognition performances of feature vectors known to be robust to the environmental noise. And, the speech subtraction technique is combined with the noise robust feature to get more performance enhancement. The experiments using SMC(Short time Modified Coherence) analysis, root cepstral analysis, LDA(Linear discriminant Analysis), PLP(Perceptual Linear Prediction), RASTA(RelAtiveSpecTrAl) processing are carried out. An isolated word recognition system is composed using semi-continuous HMM. Noisy environment experiments using two types of noises: exhibition hall, computer room are carried out at 0, 10, 20dB SNRs. The experimental result shows that SMC and root based mel cepstrum(root_mel cepstrum) show 9.86% and 12.68% recognition enhancement at 10dB in compare to the LPCC (Linear Predictive Cepstral Coefficient). And when combined with spectral subtraction, mel cepstrum and root_mel cepstrum show 16.7% and 8.4% enhanced recognition rate of 94.91% and 94.28% at 10dB.

I. 서 론

최근 음성 인식 시스템의 실용화가 늘어감에 따라 잡음 환경에서 음성 인식을 하는 문제가 많은 관심의 대상이 되고 있다. 따라서 주위 환경 변화 즉, 잡음에 적응할 수 있는 음성 인식 시스템 구현을 위한 잡음 처리 기술은 실용적인 음성 인식 시스템을 개발하기 위하여 필수적이다.

오늘날 대부분의 시스템은 연속음 인식이나 대용량 어휘, 화자 독립, 잡음 환경, 화자의 피로도나 감정에 따른 변화 등에 의하여 어려움을 겪고 있다. 그러나 고립 단어 인식 시스템의 경우 어휘의 제한에 따라 이들의 문제를 어느 정도 해결할 수 있고 최근, 특징 추출이나 화자 독립성, 잡음에 따른 강인성, 단어 추출(word spotting), 스트레스 보상(stress conpensation) 등의 분야에서 많은 진보가 이루어졌다. 단일 마이크로폰을 사용하는 경우, 잡음에 대처하는 방법으로는 크게 잡음에 강한 특징 벡터와 유사도 측정, 음질 향상, 음성 모델의 보상 방법 등으로 분류할 수 있는데, 본 연구에서는 다음과 같은 두가지 접근 방식을 사용하였다. 첫번째 방법은 잡음에 강인한 특

*연세대학교 전자공학과
Dept. of Electronics Eng., Yonsei Univ.
**군산대학교 전기공학과
Dept. of Electrical Eng., Kunsan National Univ.
***한국전자통신연구소
Electronics and Telecommunications Research Institute.

접수일자 : 1996년 5월 22일

징 벡터를 사용한 인식 시스템의 성능 향상 방법이다. 이러한 방법으로는 잡음에 강인한 루트 켑스트럼 계수(root cepstral coefficient)[1][2] 및 SMC(Short time Modified Coherence)[3] 계수, 청각 특성을 고려한 선형예측계수(perceptually linear prediction coefficient)[4][5], LDA(Linear Discriminant Analysis)[6]를 이용한 특징 벡터 변환, RASTA(RelAtive SpecTrAl) 처리[7][8][9]를 사용한 잡음에 강인한 특징 벡터 등을 사용한 방법이 연구되었다. 또한 청각의 마스킹 효과(masking effect) 효과[10]를 이용한 잡음 제거 방식도 연구되었다.

두 번째 방법은 음성 인식 시스템의 전단에서 잡음을 제거하는 방법으로서 스펙트럼 차감법(spectral subtraction method)[10][11]을 이용하였다. 이러한 방법은 기존 음성 인식 시스템의 성능을 변화시키지 않아도 되는 장점이 있다.

또한 본 연구에서는 다양한 형태의 잡음 처리 기술들의 장단점을 비교하여 인식 성능 향상을 위한 잡음 처리 방법을 제시하였다.

본 연구에서는 음성 인식을 위한 잡음 처리 기술 개발을 위하여, 숫자음 및 몇 개의 명령어를 대상으로 음성 인식 시스템을 구성하여 음성 인식 실험을 수행하였다. 음성 인식 시스템은 반연속(semi-continuous) Hidden Markov Model(HMM)[12]을 사용하여 화자 독립 단독음 인식 시스템을 구성하였다. 잡음은 ETRI에서 제공한 컴퓨터실 잡음과 전시실 부스 잡음을 첨가하여 사용하였다.

제 2장에서는 잡음에 강한 특징 벡터 및 잡음 제거 기술에 대하여 다루었고 제 3장에서는 실험 및 결과에 대하여 다루었고 제 4장에서 결론을 맺었다.

Ⅱ. 잡음에 강한 특징 벡터 및 잡음 제거 기술

2.1 루트 켑스트럼 계수(root cepstral coefficient)

일반적으로 음성 인식에 많이 이용되는 로가리듬 켑스트럼 분석(logaritmic cepstral analysis)은 잡음에 매우 민감하다. 이러한 문제를 극복하기 위하여 루트 켑스트럼 도메인(root-cepstral domain)에서의 분석으로 확장하면 잡음에 민감하지 않은 파라메터를 얻을 수 있다.

Lim[13]은 로그 연산을 root와 power(역 root) 연산으로 근사화하였다. 또한Kobayashi와 Imai[14]는 $(\cdot)^{\gamma}$ 함수 대신에 $(1/\gamma)[(\cdot)^{\gamma}-1]$ 함수를 제안했다. Alexandre와 Lockwood[1]는 일반화된 켑스트럼(generalized cepstrum)[14]을 기반으로 루트 켑스트럼 영역을 도입하여 비모수 켑스트럼 분석 방법과 선형 예측 분석 방법을 통합하여 같은 해를 얻게 됨을 보였다.

2.2 PLP(Perceptually Linear Prediction) 계수

PLP 분석 방법[4][5]은 1982년 Hermansky에 의해 제안되었으며, 음성 신호의 전력 스펙트럼을 변화시켜 청각 특성이 고려된 전력 스펙트럼를 얻는다. 이를 얻기 위해 다음과 같은 과정을 거치게 된다.

1) 근사화된 임계대역 마스킹 패턴(critical-band masking pattern)을 전력 스펙트럼에 콘볼루션시킨다.

2) 임계대역 스펙트럼(critical-band spectrum)을 1-Bark 간격으로 재표본화(resampling)한다.

3) 근사화된 고정 균등 크기 곡선(fixed equal-loudness curve)로 강조(pre-emphasis)한다.

4) 음의 크기 및 강도 파워 규칙(intensity-loudness power law)을 근사화하여 나타내주는 큐빅 루트 비선형성(cubic-root nonlinearity)을 적용한다.

이러한 단계를 거쳐 얻어지는 낮은 차수($p=5, 6$)로 구한 스펙트럼은 인간이 실제 감지하는 소리와 유사한 특성을 갖게 되며, 음성 인식에 적용되어 좋은 성능을 보여주었다.

2.3 SMC(Short-time Modified Coherence)

All-pole 모델 파라메터는 신호로부터 직접 얻는 것보다 자기상관(autocorrelation) 함수로부터 얻는 것이 더 안전하다. 이는 신호는 잡음에 의해서 변형되었을 수 있기 때문이다. 자기상관 영역(autocorrelation domain)에서 잡음에 대한 강인함은 두개의 인접한 프레임사이에서의 결합의 긴밀성(coherence)으로부터 얻어지기 때문에 SMC (Short-time Modified Coherence)[3]라고 불리워진다. SMC를 계산하기 위한 과정은 다음과 같다.

1) 2N개의 데이터 샘플로 부터 N+1개의 상관 계수를 계산한다.

$$\rho_i = \frac{1}{N}\sum_{j=0}^{N} s_j s_{j+i}, \qquad i=0, 1, 2, ..., N \tag{1}$$

2) 상관계수의 열에 해밍(Hamming) 윈도우를 적용한다.

$$\rho_i^h = \rho_i(0.54-0.46\cos\frac{2\pi i}{N+1}), \qquad i=0, 1, 2, ..., N \tag{2}$$

3) ρ_i^h의 열의 이산 푸리어 변환을 수행한다.

$$R_i = \sum_{j=0}^{N} \rho_j^h W^{ij}, \qquad i=0, 1, 2, ..., N \tag{3}$$

4) R_i의 절대값을 역 푸리어 변환한다.

$$\tilde{\rho}_i = \sum_{j=0}^{N} |R_j| W^{-ij} \tag{4}$$

5) $\tilde{\rho}_i$, $i=0, 1, 2, ..., p$로부터 기존의 LPC 방법을 이용하여 AR 모델 $A(z)$를 계산한다.

2.4 LDA(Linear Discriminant Analysis)

LDA(linear discriminant analysis)[6]는 벡터 공간에서 각 클래스 사이의 구분성을 높이려는 목적으로 사용된다. LDA에서는 D차원의 벡터를 d차원의 벡터로 대응시켜주는 선형 변환(linear transformation)을 찾는데, D차원

으로부터 D보다 작은 차원인 d차원($d \le D$)의 공간으로 차수를 줄여주는 것은 선택적으로 행할 수 있다. X를 D 차원의 벡터라 하고 U를 $D \times d$의 선형 변환 행렬이라고 할 때, d차원으로 변환된 벡터는 $U^t X$로 표현된다. 선형 변환은 $tr(W^{-1}B)$를 최대로 하는 일반적인 제한조건(criterion)에 따라 정의된다. 여기서 $tr(m)$는 행렬 m의 트레이스(trace)를 가리킨다. W와 B는 클래스 공분산(class covariance) 행렬에 의해 다음과 같이 정의된다.

$$B = \frac{1}{N} \sum_{k=1}^{K} n_k (\mu_k - \mu)(\mu_k - \mu)^t \tag{5}$$

$$W = \frac{1}{N} \sum_{k=1}^{K} \sum_{n=1}^{n_k} (x_{kn} - \mu_k)(x_{kn} - \mu_k)^t \tag{6}$$

여기서 N은 학습 패턴의 수이며, K는 집단(class)의 수, n_k는 k번째 집단(class)의 학습 패턴의 수이다. k번째 집단(class)의 평균 μ_k와 전체 평균 μ는 다음과 같이 주어진다.

$$\mu_k = \frac{1}{n_k} \sum_{n=1}^{n_k} x_{kn} \tag{7}$$

$$\mu = \frac{1}{n_k} \sum_{k=1}^{K} n_k \mu_k \tag{8}$$

여기서 x_{kn}은 k번째 집단의 n번째 학습 패턴이다.

2.5 RASTA(RelAtive SpecTrAl) 처리

천천히 변화하는 잡음을 제거하면서 음성 신호의 특징을 추출하는 방법으로 RASTA[7][8][9]분석 방법이 제안되었다. RASTA 분석 방법에서는 일반적인 단구간 스펙트럼(short-term absolute spectrum)을 사용하는 대신 스펙트럼 성분 중 시간에 따라 천천히 변화하는 성분을 배제하는 대역 통과 스펙트럼(band-pass filtered spectrum)을 사용한다. 각 주파수 대역을 IIR 필터를 사용하여 대역 통과 필터링(bandpass filtering)하는 것과 같다. 이 대역 통과 필터의 전달 함수는 다음과 같다.

$$H(z) = 0.1 \times \frac{2 + z^{-1} - z^{-3} - 2z^{-4}}{z^{-4}(1 - 0.98z^{-1})} \tag{9}$$

이 필터의 저역 차단 주파수(low cut-off frequency)는 RASTA 분석에서 고려할 가장 느린 로그 스펙트럼(log spectrum)의 변화를 결정하고, 고역 차단 주파수(high cut-off frequency)는 로그 스펙트럼(log spectrum)의 가장 빠른 스펙트럼 변화를 결정하게 된다. 위의 전달함수는 0.26 Hz의 저역 차단 주파수를 갖고, 12.8 Hz에서 부터 전달 함수의 기울기가 감소하기 시작하여 28.9 Hz와 50 Hz에서 0이 되는 모양을 갖고 있다.

2.6 마스킹(masking) 효과

인간의 청각 특성중 중요한 현상의 하나로 마스킹(masking)[10] 효과가 있는데, 이 마스킹은 하나의 소리에 다른 소리가 가려져서 듣지 못하게 되는 현상을 말한다. 마스킹에는 크게 두가지 유형이 있다. 동시에 들려오는 소리에 의한 주파수 영역에서의 마스킹(frequency masking or simultaneous masking)과 약간의 시간적인 차이를 두고 들려오는 소이에 의한 시간 영역에서의 마스킹(temporal masking)이 있다. 마스킹을 실제로 고려하기 위해서는 다른 소리에 의해 마스킹된 소리가 들릴 수 있는 음압을 결정하는 마스킹 문턱값(masking threshold)을 이용한다. 음성 신호를 분석할 때에는 일정 시간 구간으로 나누어 분석을 하게된다. 한 구간에서의 마스킹 문턱값(masking threshold)은 그 분석 구간의 스펙트럼에 의한 주파수 영역에서의 마스킹과 이전의 분석 구간의 스펙트럼에 의한 시간 영역에서의 마스킹을 함께 고려하여 결정된다.

2.7 잡음 제거 기술

잡음 환경에서 음성 인식 시스템의 성능을 향상시키는 방법중의 하나로, 음성 인식 시스템의 전단에서 음성에 섞인 소음을 제거하는 방법은 기존의 인식 시스템의 구조를 변화시키지 않아도 되는 장점이 있다.

일반적으로 잡음 제거(noise subtraction)와 음질 향상(speech enhancement)과는 사용되는 분야가 달라서 전자는 음성에 섞인 잡음을 제거하여 원래의 신호와 유사하게 하는 것에 목표를 두고 있고 후자는 신호 그 자체보다는 청각적으로 보다 우수한 음질로 들리도록 음질을 개선하는 것을 목표로 하는 것이 그 차이라고 하겠다.

따라서 잡음 환경에서 음성 인식을 위하여 사용되는 전처리 과정은 음질 향상보다는 잡음 제거가 주 목적이 되는 것이다. 만약, 소음 환경에서 사용하려는 기존 음성 인식 시스템이 소음에 대한 대책이 없이 설계되었다면 음성 인식 시스템의 성능에 전처리 과정의 성능이 중요한 변수가 된다.

잡음 제거를 위하여 사용되는 방법들은 대부분 음성에 첨가된 잡음의 형태를 미리 알거나 또는 음성이 없는 묵음 구간에서 잡음의 통계적 특성을 측정하여 음성에 섞인 소음을 제거하는 방법들을 사용하고 있다. 그러한 방법으로는 스펙트럼 차감법(spectral subtraction method)[10][11]이 있다.

Ⅲ. 실험 및 결과

3.1 데이터베이스의 구성

잡음 데이터베이스는 전자통신연구소에서 제공한 JEIDA (Japan Electronic Industry Development Association)의 잡음 데이터베이스를 이용하였다. 음성 인식 실험에 사용된 데이터 베이스는 11개 숫자음(0, 1, …, 9, 공)과 3개 명령어(걸어, 취소, 다음)의 14개로 구성되었다. 학습 데이터는 20-30대 남성 화자 50명이 각 단어를 2회씩 발음한(14단어 * 50명 * 2회 = 1400개) 음성으로 구성되었고, 테스트 데이터는 학습 데이터에 포함되지 않은 20-30대 남

성 화자 20명이 각 단어를 2회씩 발음한(14단어*20명*2회=560개) 음성으로 구성되었다. 각 음성은 비교적 조용한 연구실에서 지향성 마이크(AT831b)를 사용하여 DAT (Digital Audio Tape)에 녹음되었다.

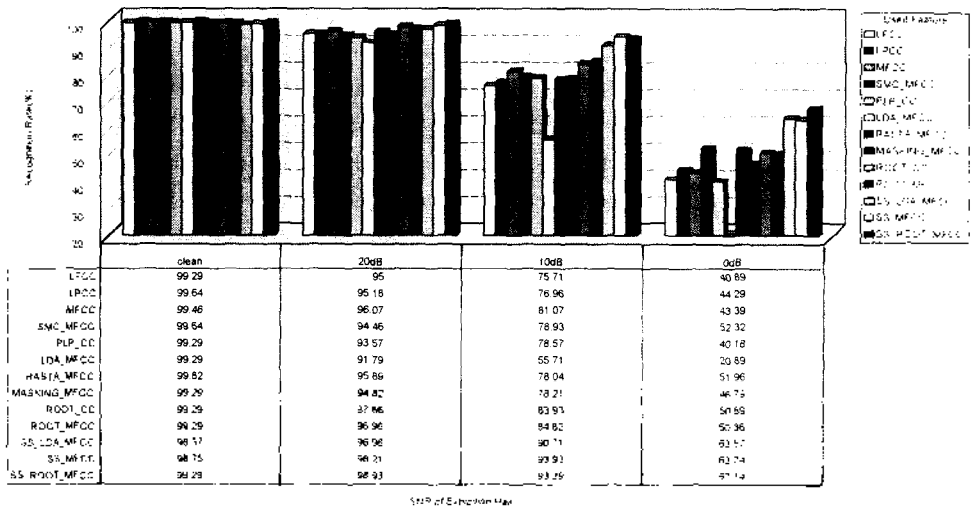
3.2 음성 신호 분석 및 인식시스템의 구성

14개의 단독음 데이터베이스의 경우 4.5kHz의 차단 주파수(cutoff frequency)를 갖는 저역 통과 필터(low pass filter)를 통과한 음성 신호는 10kHz, 16비트로 표본화된다. 표본화된 음성 신호는 $1-0.95z^{-1}$의 전달 함수를 갖는 프리엠퍼시스(pre-emphasis) 필터를 사용하여 고주파 성분을 강조한다. 이러한 음성 신호는 끝점 검출 과정에서 묵음(silence)과 음성으로 구분된다. 실험에 사용한 특징 벡터로 LPC 캡스트럼을 사용할 경우 검출된 음성 신호는 20ms(200샘플)의 크기를 갖는 해밍 창을 사용하여 10ms씩 이동하면서 p개의 차수를 갖는 선형 예측 계수를 구하는 LPC 분석 과정을 거친다. 이러한 LPC 계수로부터 인식 과정에 사용될 LPC 캡스트럼 계수를 LPC 계수와 동일한 차수까지 구한다. 마찬가지로 FFT 캡스트럼의 경우에는 LPC 분석 과정대신 FFT 분석 과정을 통하여 LPC 캡스트럼과 동일한 차수를 갖는 FFT 캡스트럼을 추출한다. 이때 이를 선형 주파수 또는 멜 스케일의 주파수 영역으로부터 역 DFT 변환을 통하여 캡스트럼을 생성하게되는데 주파수 변환 방법에 따라 LFCC(Linear Frequency Cepstral Coefficient), MFCC(Mel Frequency Cepstral Coefficient)로 구분한다. 생성된 특징 벡터는 LBG 알고리즘[15]을 사용하여 128개 코드를 갖는 코드북을 구한다. 구하여진 코드북을 반연속 HMM 모델이 공유하는 128(M=128, 특징 벡터로부터 구한 코드북의 크기와 일치하게 된다)개의 가우시안 분포의 초기 평균치로 사용하였다. 이 때 상태 j로부터 관찰열 O_t의 확률은 M개의 가우시안 분포로부터 F개($F<M$)를 적절히 조합하여 구할 수 있는 데, 사용하는 가우시안 분포의 수(F)에 따라 인식 성능에 영향을 미치게 된다. 특징 벡터로는 캡스트럼, 차등 캡스트럼 및, 에너지와 차등에너지를 이용하는데 각각 4, 4, 2개의 가우시안 분포 갯수를 할당하였다. 학습 과정은 Baum-Welch 알고리즘[16]가을 이용하였고 테스트 시에는 Viterbi 디코딩 방법[16]을 이용하였다.

3.3 각 처리 방법의 인식 성능 비교

각 처리 방법간의 비교를 위하여 특징 벡터의 차수를 12로 통일하였다. 그러나 LDA 및 PLP의 경우에는 차수를 줄이는데 그 잇점이 있으므로 6차 및 5차의 차수를 사용하였다. 거리 측정 방법은 유클리디안 거리 측정 방법을 이용하였다. 각 처리 방법에 있어서도 파라메터 값 및 세부 처리 내용에 따라 여러 가지 방법을 고려할 수 있으나 비교적 좋은 성능을 나타내는 한가지 방법만을 선택하였다. 실험에서는 캡스트럼만 이용한 경우와 차등캡스트럼을 함께 이용한 경우에 대하여도 실험을 수행하였으나 3가지의 특징 벡터를 모두 이용한 경우 가장 좋은 성능을 나타내었으므로 이를 이용한 결과만을 나타내었다. 그림 1, 2의 결과에 나타낸 특징 벡터는 다음과 같다.

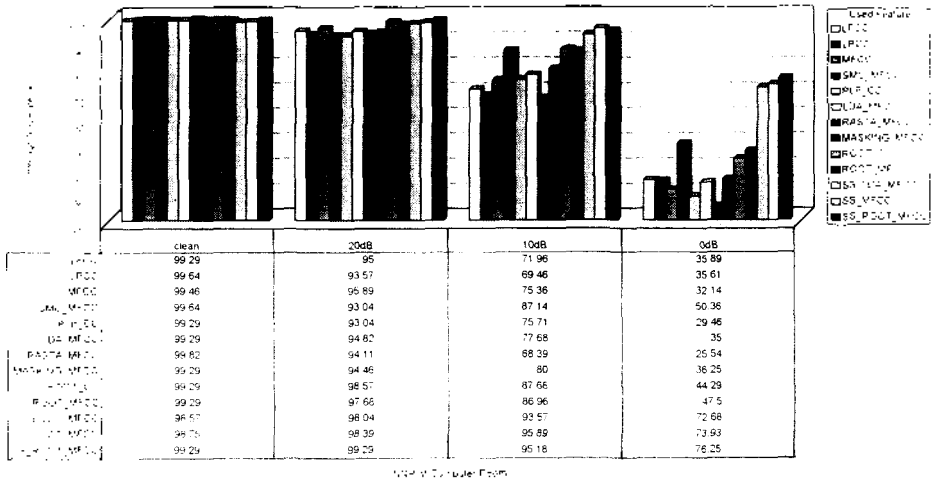
- LFCC : 선형 주파수 영역 캡스트럼 계수
- LPCC : 선형 예측 캡스트럼 계수
- MFCC : 멜 주파수 영역 FFT 캡스트럼 계수
- SMC_MFCC : Short time Modified Coherence를 이용한 멜 캡스트럼 계수
- PLP_CC : PLP 분석을 이용한 캡스트럼 계수
- LDA_MFCC : LDA 분석을 이용한 멜 캡스트럼 계수
- RASTA_MFCC : RASTA 처리를 이용한 멜 캡스트럼 계수
- MASKING_MFCC : 마스킹 특성을 이용한 멜 캡스트럼 계수



	clean	20dB	10dB	0dB
LFCC	99.29	95	75.71	40.89
LPCC	99.64	95.18	76.96	44.29
MFCC	99.46	96.07	81.07	43.39
SMC_MFCC	99.64	94.46	78.93	52.32
PLP_CC	99.29	93.57	78.57	40.18
LDA_MFCC	99.29	91.79	55.71	20.89
RASTA_MFCC	99.82	95.89	78.04	51.96
MASKING_MFCC	99.29	94.82	78.21	46.79
ROOT_CC	99.29	97.86	83.93	50.89
ROOT_MFCC	99.29	96.96	84.82	50.36
SS_LDA_MFCC	98.57	96.96	90.71	63.57
SS_MFCC	98.75	98.21	93.93	63.74
SS_ROOT_MFCC	99.29	98.93	93.39	67.14

그림 1. 각 처리 방법의 성능 비교(전시장 잡음)

Fig 1. Performance comparisson of each processing (exhibition hall noise)



	clean	20dB	10dB	0dB
	99.29	95	71.96	35.89
LPCC	99.64	93.57	69.46	35.61
MFCC	99.46	95.89	75.36	32.14
	99.64	93.04	87.14	50.36
	99.29	93.04	75.71	29.46
	99.29	94.82	77.68	35
	99.82	94.11	68.39	25.54
	99.29	94.46	80	36.25
	99.29	98.57	87.66	44.29
	99.29	97.66	86.96	47.5
	96.57	98.04	93.57	72.68
	98.75	98.39	95.89	73.93
	99.29	99.29	95.18	76.25

그림 2. 각 처리 방법의 성능 비교(컴퓨터실 잡음)
Fig 2. Performance comparisson of each processing (computer room noise)

- ROOT_CC : 루트 켑스트럼 계수
- ROOT_MFCC : 루트 처리를 이용한 멜 켑스트럼 계수
- SS_LDA_MFCC : 스펙트럼 차감법을 이용한 LDA
- SS_MFCC : 스펙트럼 차감법을 이용한 멜 켑스트럼 계수
- SS_ROOT_MFCC : 스펙트럼 차감법을 이용한 ROOT MFCC 켑스트럼 계수

실험 결과에서 특징 벡터로 많이 이용되는 LFCC나 LPCC, MFCC를 이용한 비교에 있어서는 MFCC를 이용한 경우가 잡음이 첨가된 경우 좀더 나은 성능을 나타내었다. 전체적인 비교에서 좋은 성능을 나타낸 것은 SMC_MFCC나 ROOT_CC 또는 ROOT_MFCC를 이용한 경우로 10dB SNR을 기준으로 볼 때 LPCC에 비하여 약 9.86%, 12.68%의 성능이 향상되었다. SMC_MFCC는 특히 컴퓨터실의 잡음 환경에서 좋은 성능을 나타내었는데 10dB에서 17.68% 향상되었다. ROOT_MFCC은 ROOT_CC를 멜 영역에서 구한 것으로 이들은 대체로 유사한 성능을 나타내었다. 그외에 PLP는 화자 독립 및 잡음 환경에서 좋은 성능을 나타내는 것으로 알려져 있으나 본 실험에서는 좋은 결과를 나타내지 못하였다. RASTA 처리를 이용한 결과를 보면 잡음이 섞이지 않은 경우 가장 좋은 성능을 나타내었으나 잡음이 첨가됨에 따라 성능이 많이 저하되었다. LDA나 마스킹 특성을 이용한 인식 결과에서도 잡음이 있는 환경에서 성능이 전반적으로 저하되었다. 이처럼 여러 가지 특징 벡터를 이용한 인식 성능이 많이 향상되지 않은 것은 실험에 사용된 잡음의 특성이 안정적이지 못하고 유색 잡음의 특성을 갖기 때문으로 보여진다. 이러한 특성이 두드러진 전시장 잡음을 이용한 경우 성능이 더 많이 저하되었다. 이러한 잡음의 영향을 제거하기 위하여 스펙트럼 차감법을 통한 인식 실험을 수행하였다. 본 실험에서는 MFCC과 ROOT_MFCC를 이용하여 잡음 환경에서 가장 좋은 인식률을 얻을 수 있었다. SNR 10dB에서 94.91%와 94.28%로 스펙트럼 차감법을 이용하지 않은 경우에 비하여 16.7%, 8.4%의 인식률 향상을 거두었다. 스펙트럼 차감법을 적용하지 않은 경우에는 ROOT_MFCC의 성능이 MFCC의 성능 보다 우수하였으나 스펙트럼 차감법을 적용한 경우에는 그 성능이 유사하였다. 스펙트럼 차감법은 제거하고자 하는 잡음의 특성이 안정되어야 좋은 성능을 기대할 수 있는데 실험 결과에서도 전시장 잡음보다는 컴퓨터실 잡음에 대하여 더 좋은 성능 향상이 이루어 졌다. 그러나 스펙트럼 차감법을 이용하지 않은 것보다는 모두 인식 성능이 향상되었으므로 잡음 환경에서 유용하게 이용될 수 있을 것이다.

Ⅳ. 결 론

본 논문에서는 잡음에 강인한 것으로 알려져 있는 특징 벡터들을 이용하여 반역속 HMM을 기반으로 구성한 고립단어 인식 시스템의 성능을 비교하였다. 성능 평가를 위하여 일반적으로 인식시스템이 많이 이용될 수 있는 전시장과 컴퓨터실의 부가 잡음을 0, 10, 20dB의 SNR에 적용하였다. 인식 실험 결과 다음과 같은 특성을 관측하였다.

1) LFCC, LPCC 및 MFCC의 비교에 있어서는 MFCC를 이용한 경우의 인식 성능이 가장 좋음을 확인하였다.

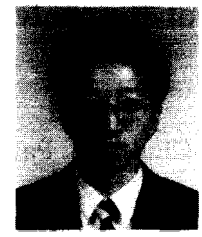
2) SMC_MFCC나 ROOT_CC, ROOT_MFCC를 이용한 경우 사용한 부가 잡음에 강인한 특성을 볼 수 있었다.

3) ROOT_CC와 ROOT_MFCC에 스펙트럼 차감법을 이용하여 음질을 향상시킨 경우 SNR 20dB 및 10dB에서 모두 90%이상으로 실제 인식기 사용에 적합한 인식 성능을 거둘 수 있었다.

참 고 문 헌

1. P. Alexandre, P. Lockwood, "Root Cepstral Analysis: A Unified View. Application to Speech Processing in Car Noise Environments", *Speech Communication*, vol. 12, no. 3, pp. 277-288, 1993.
2. P. Lockwood, P. Alexandre, "Root Adaptive Homomorphic Deconvolution Schemes for Speech Recognition in Noise", *in Proc. ICASSP*, pp. 441-444, 1994.
3. D. Mansour, B. H. Juang, "The Short-Time Modified Coherence Representation and Noisy Speech Recognition", *IEEE Trans. on ASSP*, vol. 37, no. 6, pp. 795-804, June 1989.
4. H. Hermansky, B. A. Hanson and H. Wakita, "Perceptually Based Linear Predictive Analysis of Speech," *in Proc. ICASSP*, pp. 509-512, March 1985.
5. H. Hermansky, "Perceptual Linear Predictive(PLP) Analysis of Speech", *J. Acoust. Soc. Am*, vol. 87., no. 4, pp. 1738-1752, 1990.
6. O. Siohan, "On the Robustness of Linear Discriminant Analysis as a Preprocessing Step for Noisy Speech Recognition", *in Proc. ICASSP*, pp. 125-128, 1995.
7. H. Hermansky, N. Morgan, H. G. Hirsch, "Recognition of Speech in Additive and Convolutional Noise based RASTA Spectral Processing", *in Proc. ICASSP*, pp. 83-86, 1993.
8. J. Koehler, N. Morgan, H. Hermansky, H. G. Hirsch, G. Tong, "Integrating RASTA-PLP into Speech Recognition", *in Proc. ICASSP*, pp. 421-424, 1994.
9. H. Hermansky, N. Morgan, A. Bayya, P. Kohn, "Compensation for the Effect of the Communication Channel in Auditory-Like Analysis of Speech(RASTA-PLP)", *in Proc. EUROSPEECH*, vol. 3, pp. 1367-1370, Sep. 1991.
10. T. Usagawa, M. Iwata, M. Ebata, "Speech Parameter Extraction in Noisy Environment using A Masking Model", *in Proc. ICASSP*, pp. 81-84. 1994.
11. S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-27, No 2, pp. 113-120, April 1979.
12. J. A. N. Flores, S. J. Young, "Continuous Speech Recognition on Noise using Spectral Subtraction and HMM Adaptation", *in Proc. ICASSP*, pp. 409-412, 1994.
13. J. S. Lim, "Spectral Root Homomorphic Deconvolution System," *IEEE Trans. Acoust., Speech, Signal Processing.*, Vol ASSP-27, No. 3, Jun. 1979.
14. T. Kobayashi, S. Imai, "Spectral Analysis using Generalizes Cepstrum," *IEEE Trans. Acoust., Speech, Signal Processing.*, Vol ASSP-32, No. 5, Oct. 1984.
15. X. D. Huang, Y. Ariki and M. A. Jack, *Hidden Markov Models for Speech Recognition*, Edinburgh University Press, 1990.
16. Y. Linde, A. Buzo, R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Comm.*, Vol. 28, No. 1, pp. 84-95, Jan. 1980.
17. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, Vol. 77, No. 2, pp. 257-285, 1989.
18. L. R. Rabiner, B. H. Juang, "An Introduction to Hidden Markov Model," *IEEE ASSP Mag.*, Vol. 3, No. 1, pp. 4-16, 1986.

▲신 원 호(Won Ho Shin) 1967년 8월 24일생



1991년 2월 : 연세대학교 전자공학과 졸업(공학사)
1994년 2월 : 연세대학교 대학원 전자공학과 졸업(공학석사)
1994년 3월~현재 : 연세대학교 대학원 전자공학과 박사과정
※관심분야 : 음성인식, 잡음처리.

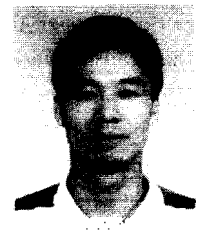
▲양 태 영(Tae Young Yang) 1970년 3월 12일생

1993년 2월 : 연세대학교 전자공학과 졸업(공학사)
1995년 8월 : 연세대학교 대학원 전자공학과 졸업(공학석사)
1995년 9월~현재 : 연세대학교 대학원 전자공학과 박사과정
※관심분야 : 음성인식, 화자적응.

▲김 원 구(Weon Goo Kim) : 1994년 13권 1호 참조.

▲윤 대 희(Dae Hee Youn) : 1994년 13권 1호 참조.

▲서 영 주(Young Joo Seo) 1969년 12월 16일생



1991년 2월 : 경북대학교 전자공학과 졸업(공학사)
1993년 2월 : 경북대학교 대학원 전자공학과 졸업(공학석사)
1993년 2월~현재 : 한국 전자통신연구소 근무
※관심분야 : 음성 인식 잡음 제거