

논문 2011-48CI-1-13

멀티밴드 스펙트럼 차감법과 엔트로피 하모닉을 이용한 잡음환경에 강인한 분산음성인식

(Robust Distributed Speech Recognition under noise environment using
MBSS and EH-VAD)

최 갑 근*, 김 순 협**

(Gab-Keun Choi and Soon-Hyob Kim)

요 약

음성인식의 실용화에 가장 저해되는 요소는 배경잡음과 채널에 의한 왜곡이다. 일반적으로 잡음은 음성인식 시스템의 성능을 저하시키고 이로 인해 사용 장소의 제약을 많이 받고 있다. DSR(Distributed Speech Recognition) 기반의 음성인식 역시 이와 같은 문제로 성능 향상에 어려움을 겪고 있다. 이 논문은 잡음환경에서 DSR기반의 음성인식을 향상 위해 정확한 음성 구간을 검출하고, 잡음을 제거하여 잡음에 강인한 특징추출을 하도록 설계하였다. 제안된 방법은 엔트로피와 음성의 하모닉을 이용해 음성구간을 검출하며 멀티밴드 스펙트럼 차감법을 이용하여 잡음을 제거한다. 음성의 스펙트럴 에너지에 대한 엔트로피를 사용하여 음성검출을 하게 되면 비교적 높은 SNR 환경(SNR 15dB)에서는 성능이 우수하나 잡음환경의 변화에 따라 음성과 비음성의 문턱 값이 변화하여 낮은 SNR환경(SNR 0dB)에서는 정확한 음성 검출이 어렵다. 이 논문은 낮은 SNR 환경(0dB)에서도 정확한 음성을 검출할 수 있도록 음성의 스펙트럴 엔트로피와 하모닉 성분을 이용하였으며 정확한 음성 구간 검출에 따라 잡음을 제거하여 잡음에 강인한 특징을 추출하도록 하였다. 실험결과 잡음환경에 따른 인식조건에서 개선된 인식성능을 보였다.

Abstract

The background noises and distortions by channel are major factors that disturb the practical use of speech recognition. Usually, noise reduce the performance of speech recognition system. DSR(Distributed Speech Recognition) based speech recognition also has difficulty of improving performance for this reason. Therefore, to improve DSR-based speech recognition under noisy environment, this paper proposes a method which detects accurate speech region to extract accurate features. The proposed method distinguish speech and noise by using entropy and detection of spectral energy of speech. The speech detection by the spectral energy of speech shows good performance under relatively high SNR(SNR 15dB). But when the noise environment varies, the threshold between speech and noise also varies, and speech detection performance reduces under low SNR(SNR 0dB) environment. The proposed method uses the spectral entropy and harmonics of speech for better speech detection. Also, the performance of AFE is increased by precise speech detections. According to the result of experiment, the proposed method shows better recognition performance under noise environment.

Keywords : Noisy speech recognition, Distributed Speech Recognition, speech recognition system

I. 서 론

정보통신산업의 비약적인 발전은 컴퓨팅 환경의 극단적 변화를 가져왔다.

현재의 컴퓨팅환경은 불과 몇 년 전 까지만 해도 음성통신이 위주였던 모바일 단말기에 대해 네트워크 기능과 다양한 응용소프트웨어를 사용할 수 있도록 성능이 강화된 스마트폰이 대중적으로 보급되기 시작하였다.

* 학생회원, ** 정회원, 광운대학교 컴퓨터공학과
(Computer Engineering Department, Kwangwoon University)

※ 본 연구는 광운대학교 2010년 교내연구비로 수행되었습니다.

접수일자: 2010년7월26일, 수정완료일: 2010년12월30일

하지만 휴대가 간편하고 다양한 응용소프트웨어를 사용할 수 있는 스마트폰 역시 입력장치로 터치패드와 키패드를 사용하는 과거의 방식을 채택하고 있다. 이와 같은 입력장치의 불편함을 극복하고자 근래 들어 음성을 이용하여 웹을 검색하는 응용소프트웨어가 개발되고 상당한 성과를 거두면서 스마트폰 등에 대한 음성인식 기술의 관심이 나날이 확대되고 있다. 특히 스마트폰 등에 적용할 수 있는 음성인식 기술로 음성의 특징 추출은 단말기가 담당하고 연산량과 메모리용량을 많이 요구하는 인식부분은 원격서버가 담당하도록 하는 분산 음성인식이 연구되고 있다. 분산 음성인식은 배경잡음을 효과적으로 제거하고 정확한 음성 특징 추출을 위한 다양한 연구가 시도 되고 있으며 이를 위해 ETSI (European Telecommunication Standard Institute)에서는 이에 관한 표준을 제정하고 있고 특징추출과 관련하여 FE(Front End)가 제정된 이후 잡음제거와 모델 보상 기법 등을 보강한 AFE(Advanced Front End)가 제정되었고 기존의 방법들과 비교하여 우수한 성능을 보이며 현재까지 가장 높은 인식률을 보이나 아직까지 완벽한 성능을 보이지 않아 개선의 여지를 남겨두고 있다. [그림 1]은 ETSI에서 제정한 AFE(Advanced Front End) 표준안이다.^[1~2]

음성인식 성능의 가장 큰 저해요소는 배경잡음으로 인한 인식환경과 훈련환경의 불일치가 가장 크다. 일반적으로 음성인식을 위한 훈련용 음성은 조용한 환경에서 수집된 음성을 사용한다. 하지만 실생활에서 발생하는 다양한 배경잡음은 인식환경이 훈련환경과 차이를 보이게 하는 가장 큰 요소이다. 분산음성인식 역시 사용 환경에 따라 발생하는 배경잡음의 처리는 인식을 개선에 중요한 요소가 된다. 이와 같이 배경잡음에 의한 인식을 저하를 막기 위해 잡음을 제거하는 다양한 방법들이 소개되고 있는데 음성향상기술이 그 대표적인 기술이다.

음성향상 알고리즘 중에서 일반적으로 널리 사용되고 있는 알고리즘은 Wiener 필터 방식이며, 음성신호의

스펙트럼에 대한 MMSE(Minimum Mean Square Error) 추정 기반의 필터가 사용된다. 이보다 개선된 것은 음성신호와 잡음신호의 스펙트럼에 대한 사전 확률 분포를 가정하고 통계모델에 근거하여 음성신호의 스펙트럼의 크기를 추정하는 MMSE-STSA(Short Time Spectral Amplitude)방법이 좋은 성능을 보이나 음성통신에서 음성품질향상을 목표로 연구된 알고리즘으로 인식상황에 적합하지 않을 수 있으며 특히 잡음제거 후 음악잡음과 잔존잡음이 남아 인식률을 저하시키게 되어 음성인식을 위해서는 적절한 알고리즘의 선택이 필요하다.^[3~4]

낮은 SNR환경에서 효과적인 잡음제거와 음성인식률 향상을 위해서는 정확한 음성검출이 필요하다. 비교적 적은 연산량과 메모리 용량을 갖는 ZCR(Zero Crossing Rate), logEnergy등의 방법과 다소 연산량은 많지만 우수한 성능을 보이는 통계적 확률에 기반한 LRT(Likelihood Ratio Test)등이 있다.

본 논문에서는 SNR의 변화에 비교적 안정적인 문턱값 성능을 보이는 EH-VAD(Entropy Harmonic Voice Activity Detection)를 사용하였다.^[5~6]

II. 엔트로피 하모닉 음성검출과 멀티밴드 스펙트럼 차감법(Entropy Harmonic VAD and Multi-Band Spectral Subtraction)

2.1. Entropy of Energy Spectrum

낮은 SNR 에너지 스펙트럼에서 음성영역은 비음성 영역에 비해 상대적으로 높은 에너지 스펙트럼을 나타낸다. 따라서 음성에너지 스펙트럼은 비음성 에너지 스펙트럼에 비해 상대적으로 높은 에너지 스펙트럼을 갖고 있다고 가정 할 수 있다. 이와 같은 에너지 스펙트럼은 섀넌(Shannon)에 의해 소개된 정보 엔트로피와 유사하게 표현할 수 있다.^[7]

섀넌의 엔트로피는 식 (1)과 같이 정의된다.

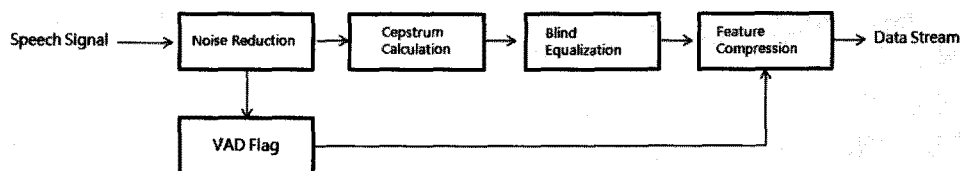


그림 1. 분산 음성인식의 AFE(Advanced Front End)

Fig. 1. Advanced Front End of Distributed Speech Recognition.

$$H(S) = - \sum_{i=1}^N P(s(i)) \cdot \log_2(P(s(i))) \quad (1)$$

여기서 N 은 심볼의 수, $s(i)$ 는 심볼 i , 그리고 $P(i)$ 는 심볼 i 에 대한 사후 확률이다. 엔트로피는 스펙트럴 에너지 영역에서 식(2)와 같이 정의 할 수 있다.

$$H(|Y(k,l)|^2) = - \sum_{k=1}^{N/2} \{ P(|Y(k,l)|^2) \cdot \log_2(P(|Y(k,l)|^2)) \} \quad (2)$$

엔트로피의 계산을 위해 먼저 DFT(Discrete Fourier Transform)를 이용하여 이산 스펙트럴 파워를 계산한다. 여기서 k 는 주파수 빈(Frequency bin)인덱스 이고 l 은 프레임 인덱스이다. 주어진 프레임 l 에서 주파수빈 k 에 대한 스펙트럴 에너지 확률은 식 (3)과 같이 계산한다.

$$P(|Y(k,l)|^2) = \frac{|Y(k,l)|^2}{\sum_{k=1}^{N/2} |Y(k,l)|^2} \quad (3)$$

구해진 각 주파수 빈의 확률은 식(2)에 의해 엔트로피로 계산되어진다.^[7]

2.2 Harmonic Scoring

음성의 하모닉(Harmonic)은 모음에서 주로 발견된다. 인간의 음성은 모음에서 성대의 공명을 통해 발생된 주기성이 강한 기본주파수(Fundamental Frequency)를 중심으로 정수배에 해당하는 주파수 영역에 하모닉 성분이 발견되며 이러한 특징은 음성과 비음성을 구분하는 중요한 요소이다. 또한 기본주파수는 피치(Pitch)라고도 한다. 본 논문에서는 낮은 SNR에서 비교적 잡음과 음성의 구분이 뚜렷한 하모닉 성분을 이용하여 낮은 SNR에서 문턱 값의 성능이 떨어지는 스펙트럴 에너지 엔트로피에 보조적 알고리즘으로 사용한다. 하모닉 성분은 [그림 2]에서 보여지는 바와 같이 특히 낮은 SNR 환경에서 모음구간을 검출하는데 유리하기 때문에 주어진 프레임 내에서 하모닉 성분을 검출하여 스코어를 주고 검출된 스코어를 이용하여 엔트로피에 의해 정해진 문턱 값에 가중치를 주게 된다.^[5,8]

알고리즘을 처리하기 위해 입력신호에 대한 단구간 푸리에 분석은 음성신호 $x(t)$ 를 식 (4)와 같이 정의하고 주파수 영역에서 처리하기 위하여 식(6)을 이용하여

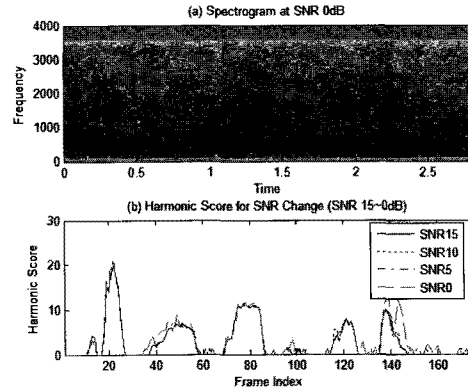


그림 2. SNR 변화에 대한 하모닉 점수
Fig. 2. Harmonic score for SNR change.

DFT(Discrete fourier Transform) 처리하여 식(5)와 같이 주파수 성분을 $[X_k]$ 로 표시한다.

$$x(t) = [x_k] = [x_0, x_1, x_2, \dots, x_{N-1}] \quad (4)$$

$$[X_k] = DFT[x_k] = [X_0, X_1, X_2, \dots, X_{N/2}] \quad (5)$$

$$X_k = \left[\sum_{n=0}^{N-1} x_n e^{-j(2\pi nk/N)} \right] \quad (6)$$

여기서 식(6)은 우측항이 복소수이므로 절대치를 X_k 로 하며 k 는 주파수 빈(frequency bin)의 인덱스(index)이다.

$$\begin{aligned} &\text{if } X(k,l) > X(k-1,l), X(k,l) > X(k+1,l) \\ &\text{then } H_{peak}(k,l) = X(k,l) \\ &\text{else } H_{peak}(k,l) = 0 \end{aligned} \quad (7)$$

식 (7)은 주어진 프레임 l 에서 스펙트럴 피크를 계산하기 위한 것으로 각각의 주파수 빈에 대해 인접한 $k-1$ 주파수 빈과 $k+1$ 주파수 빈을 비교하여 k 번째 주파수 빈이 크면 스펙트럴 피크로 본다. 그렇지 않은 경우에는 에너지를 0으로 주어 하모닉을 조사하기에 용이하도록 한다.^[5] 주어진 프레임 내에서 스펙트럴 피크 성분이 등 간격을 이루지 않고 불연속 될 수 있다. 이러한 문제를 피하기 위해 기본주파수를 중심으로 등 간격을 이루지 않는 주파수 빈에 대해서도 에너지를 0으로 주어 보다 정확한 하모닉 성분을 추출할 수 있도록 한다.^[5]

$$H_{score}(l) = \sum_{k=1}^{N/2} H_{peak}(k, l) \quad (8)$$

식(7)에서 H_{peak} 는 주어진 프레임 내에서 하모닉 성분을 갖는 peak값을 갖게 되며 그 외의 주파수 빈은 에너지가 0이 되도록 한다. 이 값들을 모두 더한 값을 하모닉 스코어로 결정한다.

2.3 MBSS(Multi-Band Spectral Subtraction)

순수한 음성신호에 잡음신호를 가산하면 다음과 같이 표현할 수 있다.

$$y(n) = s(n) + d(n) \quad (9)$$

여기서 $y(n)$ 은 잡음에 의해 손상된 신호를 의미하고, $s(n)$ 은 순수한 음성신호, 그리고 $d(n)$ 은 잡음을 표현한다. 손상된 음성신호의 파워 스펙트럼은 다음과 같이 표현될 수 있다.

$$|Y(k)|^2 \approx |S(k)|^2 + |D(k)|^2 \quad (10)$$

여기서 $|S(k)|$ 과 $|D(k)|$ 는 음성과 잡음의 매그니튜드 스펙트럼이다. 잡음 스펙트럼은 바로 얻을 수 없고 음성이 없는 묵음구간을 계산해서 잡음예측신호 $\hat{D}(k)$ 를 구한다. Ball등에 의해 소개된 스펙트럼 차감법은 스펙트럼 상에서 직관적인 방법을 사용하여 잡음을 제거하는 알고리즘으로 다양한 방법들이 소개되었으나, 대역별로 SNR이 서로 달라 음악잡음과 잔존잡음을 제거하는데 효과적이지 못하였다. Kamarth와 Loizou는 이와 같은 문제를 해결하기 위해 Berouti가 제안한 방법으로 깨끗한 음성 스펙트럼을 예측하며 식(11)과 같이 잡음에 오염된 신호에서 깨끗한 음성신호를 얻는다.^[9-11]

$$|\hat{S}(k)|^2 = |Y(k)|^2 - \alpha |D(k)|^2 \quad \alpha \geq 1 \quad (11)$$

여기서 α 는 밴드별 세그멘탈 SNR을 위한 Over subtraction factor 이다.^[11]

$$\alpha = \begin{cases} \alpha_0 + \frac{3}{4} & SNR_i \leq -5db \\ \alpha_0 - \frac{3}{20} SNR_i & -5db \leq SNR_i \leq 20db \\ \alpha_0 - 3 & SNR_i \leq 20db \end{cases} \quad (12)$$

i^{th} 주파수 밴드의 세그멘탈 SNR_i 는 식 (12)와 같이

계산되어 진다.

$$SNR_i = 10 \log \frac{\sum_{k=b_i}^{e_i} |Y_i(k)|^2}{\sum_{k=b_i}^{e_i} |D_i(k)|^2} \quad (13)$$

여기서 b_i 와 e_i 는 주파수 밴드의 i^{th} 시작 주파수 빈과 끝 주파수 빈이다.

III. 분산 음성인식을 위한 잡음환경에 강인한 음성추출 (Robust Speech Feature Extraction under Noise Environment for Distributed Speech Recognition)

본 논문에서 제안하는 분산 음성인식 시스템의 성능 향상 방법은 인식환경과 훈련환경의 불일치를 줄이고자 음성신호의 매그니튜드 스펙트럼에 대해 엔트로피와 하모닉스를 구하여 잡음환경에 강인한 음성검출을 하였고, Kamarth와 Loizou가 제안한 멀티밴드 스펙트럼 차감법을 이용하여 잡음신호를 제거하였다. 음성을 검출하고 잡음을 제거한 후 알고리즘은 MFCC(Mel Frequency Cepstral Coefficient)특징 추출 절차를 따르며 잡음이 제거된 음성 특징을 추출하게 된다. 이 과정을 상세하게 보면, 알고리즘은 잡음이 부가된 음성신호에서 엔트로피와 하모닉스를 이용하여 음성구간을 검출한다. 하모닉스는 음성의 모음에서 추출할 수 있는 성분으로 기본 주파수의 정수배를 갖는 고조파 성분이다. 스펙트럴 에너지 엔트로피는 다른 음성검출 알고리즘에 비해 비교적 높은 SNR에서는 문턱 값 설정이 용이하지만 낮은 SNR에서 성능이 급격히 나빠지는 특징을 갖고 있다. 이와 같은 문제를 해결하고 문턱 값을 안정하게 하기 위해 프레임 내에 하모닉스 존재 여부를 점수화하여 엔트로피의 문턱 값을 보정하게 된다. 음성검출이 이루어진 후 멀티밴드 스펙트럼 차감법에 의해 잡음을 제거하며 그 과정은 [그림 3]과 같다.

Wiener 필터, MMSE-STSA(Minimum Mean Square Error-Short Time Spectral Amplitude)등 대부분의 음성향상 알고리즘들이 갖고 있는 잔존잡음과 음악잡음은 음성 향상에서 음성의 품질도 나쁘게 하지만 음성인식에서 역시 잘못된 특징을 추출하게 하여 인식률의 저하를 가져다주게 된다.

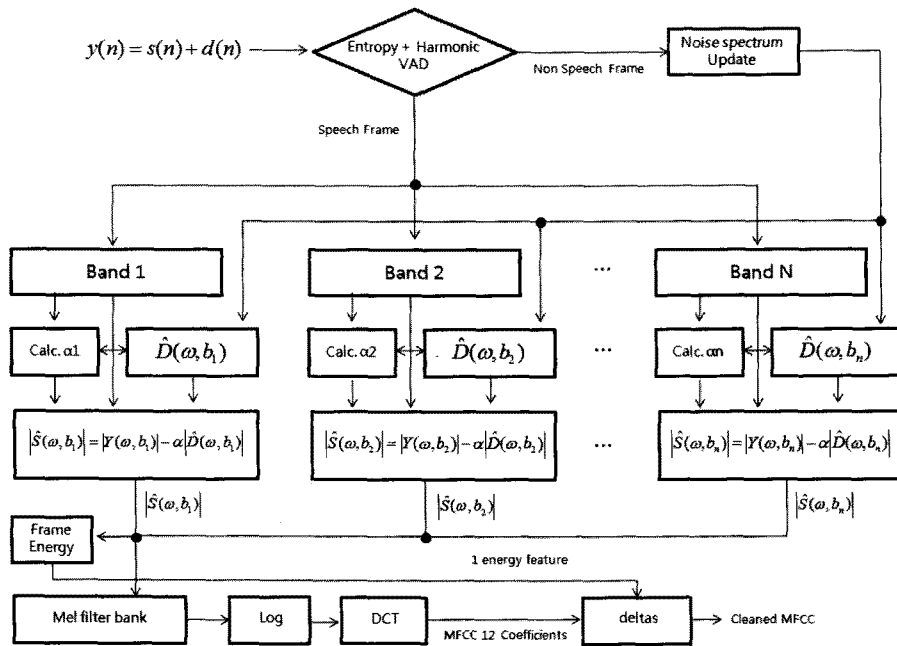


그림 3. 잡음환경에서 분산 음성인식의 강인한 음성 특징추출을 위한 전처리 방법
 Fig. 3. Preprocessing Method for Robust DSR speech feature extraction under Noise Environment.

Kamarth와 Loizou가 제안한 멀티밴드 스펙트럼 차감법은 주파수 차감법이 갖는 비교적 간단하고 직관적인 특징을 유지하면서 음성에 부가된 잡음을 제거하는데 효과적이다. 이 방법은 잡음이 부가된 음성신호를 여러 개의 밴드로 나누어 밴드마다 변하는 잡음의 SNR을 예측하여 제거하므로 식(12)와 같이 실험적으로 얻어진 적절한 α 값의 적용에 따라 비교적 효과적으로 잔존잡음과 음악잡음을 억제할 수 있다.^[11]

IV. 실험 및 결과

본 연구를 위하여 잡음 음성인식을 위한 음성 데이터 베이스를 위해 남, 여 각각 20명이 5회씩 발성한 200개의 음성에 NoiseX-92 데이터베이스에서 제공하는 잡음을 인공적으로 부가하여 사용하였다. 훈련은 깨끗한 음성 데이터만을 이용하여 HMM(Hidden Markov Model)을 훈련하는 방식이다. 인식을 위해서는 SNR (15dB, 10dB, 5dB, 0dB)에 따른 잡음이 부가된 음성을 이용하여 실시하였다.

음성특징을 위해서는 13차의 MFCC(Mel Frequency Cepstral Coefficient)를 사용하였고 delta와 acceleration 계수를 추가하여 39차 특징벡터를 사용하였다.

HMM은 6개의 가우시안 성분을 가지는 3개의 상태로 이루어지며 인식기는 본 연구를 위해 자체적으로 개

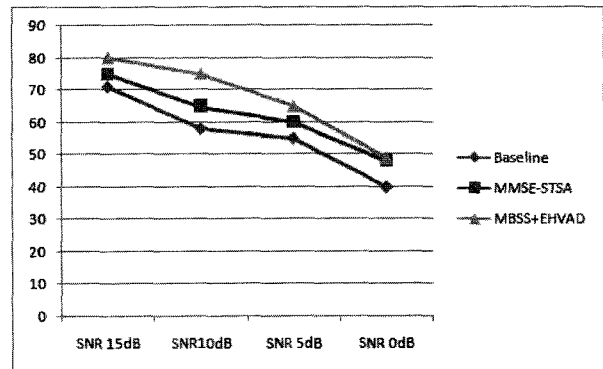


그림 4. 음질개선 방식의 인식성능 비교 (Word Accuracy(%))
 Fig. 4. Performance comparison between speech enhancement methods.

발된 인식기를 사용하였으며 Baum-Welch 기반의 훈련 과정을 통해서 HMM 모델을 구성하고 Viterbi 알고리즘을 이용하여 인식결과를 얻었다. 실험결과 [그림 4]에서 보여 지는 바와 같이 잡음제거가 이루어지지 않은 상황에서 잡음의 변화에 따라 SNR이 낮아지면 인식률도 급격히 나빠지며, 음성향상방법을 이용하여 잡음을 제거하게 되면 높은 SNR에서 비교적 좋은 인식률 개선을 확인할 수 있다.

MMSE-STSA방법과 비교해 MBSS(Multi-Band Spectral Subtraction)+EHVAD(Entropy Harmonic - Voice Activity Detection)의 성능이 비교적 우수한 것

은 잡음을 제거한 후에 생성되는 잔존잡음과 음악잡음이 인식모델을 손상시키는 것이 상대적으로 적게 생성되었고, 하모닉 점수를 이용한 음성검출이 음성신호의 존재에 대한 불확실성에 관한 정확한 정보가 상대적으로 많이 제공되어 인식률을 향상시키는 것을 알 수 있다. 하지만 음성검출성능이 뛰어나지 못한 낮은 SNR(0dB)에서는 성능의 차이가 크지 않은 것을 확인할 수 있으며 결국 음성신호의 존재에 대한 정확한 검출이 잡음을 추정하여 잡음을 제거하는 것에 영향을 주고 잡음으로 인한 왜곡된 음성 특징 추출을 방지하는 것을 알 수 있다.

V. 결 론

본 논문에서는 잡음환경하의 분산음성인식에서 AFE (Advanced Front End)의 성능을 개선하기 위해 음성향상 방법을 이용하여 잡음을 제거한다. 잡음을 제거하기 위한 방법은 주파수 차감법을 개선한 MBSS (Multi-band Spectral Subtraction)방법을 사용하였고, 특별히 낮은 SNR에서 정확한 음성검출을 하도록 음성의 하모닉 성분을 검출하여 스펙트럴 에너지 엔트로피에 점수를 주어 비교적 안정적인 문턱 값을 유지할 수 있도록 돕는 보조적 알고리즘을 사용하였다. 이 같은 방법을 사용하여 일반적으로 많이 사용되는 MMSE-STSA방법과 비교 인식실험을 실시한 결과 개선된 성능을 보이는 것을 확인했다. 하지만 음성의 검출과 음성향상 방법만을 이용해서는 잡음환경에서 음성인식을 향상에 한계가 있는 것으로 보인다. 따라서 근래에 많이 제안되고 있는 특징보상 방식을 적용하면 보다 나은 음성인식을 향상을 기대 할 수 있을 것으로 본다.

VI. 감사의 글

본 연구는 광운대학교 2010년 교내연구비로 수행되었습니다.

참 고 문 헌

- [1] ETSI standard document, "Speech Processing, Transmission and Quality aspects(STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithms," *ETSI ES 201 108 v.1.1.1* (2000-02), Feb. 2002.
- [2] ETSI standard document, "Speech Processing, Transmission and Quality aspects(STQ); Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms," *ETSI ES 202 050 v.1.1.3* (2003-11), Nov. 2003.
- [3] Scart, P., Filho, J. "Speech enhancement based on a priori signal to noise estimation", *Proc IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 629-632, 2002.
- [4] Ephraim, Y., Malah, D. "Speech enhancement Using a minimum mean square error short-time spectral amplitude estimator", *IEEE Trans. Acoust., Speech Signal Process.*, Vol 32, pp. 1109-1121, 1984.
- [5] 최갑근, 김순협, "엔트로피와 하모닉 검출을 이용한 잡음환경에 강인한 음성검출", *전자공학회논문지*, 제 47권 SP편 1호, 169-174쪽, 1229-6384, 2010년.
- [6] Rabiner, L. R., M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances", *The Bell System Technical Journal*, Vol. 54, No. 2, pp. 297-315, 1975.
- [7] Abdallah I., Montresor S., Baudry M, "Robust speech/non-speech detection in adverse conditions using an entropy based estimator" *Digital Signal Processing Proceedings 1997*, pp. 752-760, Santorini Greece, Jul 1997.
- [8] Ramalho, M.A. Mammone, R.J. "New speech enhancement techniques using the pitch mode modulation model" *Circuits and Systems, 1993 Proceedings of the 36th Midwest Symposium*, pp. 1531-1534, Detroit, USA, Aug 1993.
- [9] Ball, S, F., "Supression of acoustic noise in speech using spectral subtraction", *IEEE Trans., Speech Signal Process.*, Vol. 27, pp 113-120, 1979.
- [10] Berouti, M., Schwartz, R., Makhoul, J. "Enhancement of speech corrupted by acoustic noise", *Proc. IEEE Int. Conf. Acoustic, Speech Signal Process.*, pp 208-211., 1979.
- [11] Kamarth, S., Loizou, P. "A multi-Band spectral subtraction method for enhancing speech corrupted by colored noise", *Proc. IEEE Int. Conf. Acoustic Speech Signal Process.*, pp 101-111, 2002.
- [12] A. Varga and H. J. M. Steeneken, "Assessmentfor automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol 12, no. 3, pp. 247-251, July 1993.

저 자 소 개



최 갑 근(학생회원)
 1999년 광운대학교 정보과학원
 학사 졸업.
 2002년 광운대학교 컴퓨터공학과
 석사 졸업.
 2006년 광운대학교 컴퓨터공학과
 박사수료.

2008년~2010년 에이펫 주식회사 IT사업부장
 2010년~현재 (주) 오성미디컴 부설연구소
 오디오 신호처리팀 팀장
 <주관심분야 : 오디오신호처리, 음성인식>



김 순 협(정회원)
 1974년 울산대학교 전자공학과
 학사 졸업.
 1976년 연세대학교 전자공학과
 석사 졸업.
 1983년 연세대학교 전자공학과
 박사 졸업.

1979년~현재 광운대학교 컴퓨터공학과 교수
 <주관심분야 : 음성인식, 신호처리>