

정규논문 (Regular Paper)

방송공학회논문지 제18권 제2호, 2013년 3월 (JBE Vol. 18, No. 2, March 2013)

<http://dx.doi.org/10.5909/JBE.2013.18.2.311>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

잡음환경 음성명료도 향상을 위한 이진 마스크 추정 후처리 알고리즘

김 기 백^{a)‡}

A Post-processing for Binary Mask Estimation Toward Improving Speech Intelligibility in Noise

Gibak Kim^{a)‡}

요 약

시간-주파수 영역에서의 이진 마스크를 이용하여 잡음환경에서 잡음을 제거하여 음질을 향상하는 방법에 대해 논하고자 한다. 잡음이 섞여 있는 음성신호를 시간-주파수 영역으로 분해하여, 상대적으로 잡음이 많이 섞여 있는 시간-주파수 영역 (시간-주파수 유닛의 신호 대 잡음 비 (Signal-to-Noise Ratio: SNR)가 낮은 영역)의 신호에 마스크 “0”을 할당하여 제거함으로써 음성명료도를 향상시킬 수 있다. 이전의 연구에서는 가우시안 혼합 모델을 이용하여 마스크 “0”과 마스크 “1”을 분류하는 방법을 사용하였다. 각 주파수 밴드 별로 수집된 데이터를 이용하여 가우시안 혼합 모델을 학습하고 테스트 데이터가 들어오면 현재의 시간-주파수 마스크가 “0”인지 “1”인지 판별하게 된다. 본 논문에서는 이러한 알고리즘에 주파수 영역에서의 종속성을 고려하여 추정된 마스크에 대해 후처리를 수행하는 알고리즘을 제안한다. 주파수 영역에서의 종속성에 관한 후처리는 비터비 (Viterbi) 알고리즘을 이용하며, 제안된 후처리 알고리즘을 적용하여 이진 마스크 추정 오차를 줄여 음성 명료도 향상을 기대할 수 있다.

Abstract

This paper deals with a noise reduction algorithm which uses the binary masking in the time-frequency domain. To improve speech intelligibility in noise, noise-masked speech is decomposed into time-frequency units and mask “0” is assigned to masker-dominant region removing time-frequency units where noise is dominant compared to speech. In the previous research, Gaussian mixture models were used to classify the speech-dominant region and noise-dominant region which correspond to mask “1” and mask “0”, respectively. In each frequency band, data were collected and trained to build the Gaussian mixture models and detection procedure is performed to the test data where each time-frequency unit belongs to speech-dominant region or noise-dominant region. In this paper, we consider the correlation of masks in the frequency domain and propose a post-processing method which exploits the Viterbi algorithm.

Keyword : Binary mask, noise reduction, Viterbi algorithm

a) 숭실대학교 전기공학부 (School of Electrical Engineering, Soongsil University)

‡ Corresponding Author : 김기백(Gibak Kim)

E-mail: imkgb27@ssu.ac.kr

Tel: +82-2-828-7266

※ 이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행되었습니다. (2012-0003455).

· Manuscript received January 29, 2013 Revised February 25, 2013 Accepted February 25, 2013

I. 서론

잡음은 음성 신호 처리 분야에서 성능을 저해하는 가장 주요한 요소로 볼 수 있으며, 잡음을 제거하고자 하는 노력은 음성 통신, 자동 음성 인식, 음성 부호화 등의 분야에서 오랫동안 여러 연구자에 의해 연구되어 왔다^[1-6]. 그럼에도 불구하고 정상 잡음 (stationary noise) 환경을 제외하고는 뚜렷한 성과를 보여주지 못하고 있다. 그 이유는 음성 및 음향 신호에서 다루는 잡음이 통신 등 다른 신호 처리에서 다루는 잡음과 달리 매우 다양한 잡음의 형태가 존재하기 때문이라고 볼 수 있다. 최근 들어 하드웨어 성능의 발달과 더불어 고속연산이 가능해짐에 따라 하나 이상의 마이크를 이용하여 신호의 공간 정보를 사용함으로써 보다 효과적인 잡음 제거를 시도하는 연구들도 활성화되고 있다^[7]. 그러나 여전히 비용 및 하드웨어의 제약으로 인해 하나의 마이크를 이용하는 시스템들이 대다수를 차지하고 있으며 이에 따라 하나의 마이크를 이용하는 잡음 제거 알고리즘이 절실히 요구되고 있는 상황이다. 하나의 마이크를 이용하여 잡음을 제거하는 기존의 많은 연구들은 음질 개선에 초점이 맞춰져 있는 경우가 많았다. 최근의 연구에 따르면 이러한 음질 개선에서는 어느 정도의 성과를 보이고 있음을 알 수 있지만, 음성의 명료도 개선의 측면에서 볼 때는 거의 성과를 보여주지 못하고 있다^[8-10].

최근 제안된 이진 마스크를 적용하는 방법에서는 신호를 시간-주파수 영역으로 분해한 후, 잡음이 많은 영역은 제거하고 그렇지 않은 부분은 어떤 이득을 곱하지 않고 그대로 두는 방식을 취하고 있다^[11,12]. 이러한 이진 마스크는 CASA^[13,14] (Computational Auditory Scene Analysis)로부터 연구되었던 것으로서 기존의 많은 잡음 제거 알고리즘들이 시간-주파수 영역에서 잡음 성분을 추정하여 필터링하는 것과는 달리 잡음 성분이 적은 유닛은 아무것도 하지 않고 그대로 두고, 잡음이 많은 유닛은 제거하는 방법을 사용하고 있다. 기존의 필터링 방식에서는 적절한 이득을 곱하게 되는데 이득을 구하는 과정에서 추정오차로 인해 음성에 왜곡이 발생하게 된다. 이진 마스크 방식에서는 잡음이 많은 부분을 제거하는 것을 제외하고는 어떠한 처리도 가하지 않음으로써 음성 왜곡을 최소화하고자 하였다. 그

러나 이진 마스크 추정 방법은 학습된 잡음 환경에서만 동작하는 단점이 있는데 이러한 단점을 극복하고자 새로운 잡음 환경에 대해 집진적으로 학습하는 방법이 제안되었다^[12].

이진 마스크를 추정하는 기존의 방식에서는 각 시간 프레임에 대해 몇 개의 주파수 대역으로 나눈 다음, 각 주파수 영역별 이진 마스크가 서로 독립이라고 가정하고 각 주파수 밴드별로 데이터를 수집하여 훈련하여 분류 모델을 구축하였다. 그러나 이진 마스크는 주파수 영역에서 독립적이라고 볼 수 없으며 특히 음성 신호의 경우는 주파수 영역에서의 종속성이 더 강하게 나타난다. 본 논문에서는 이러한 주파수 영역에서의 종속성을 고려하여 각 시간 프레임에서 나타난 주파수 영역의 이산 프레임 시퀀스를 구하기 위하여 비터비(Viterbi) 알고리즘^[15,16]을 적용하였다.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 학습 기반의 이진 마스크 추정에 대해 설명하고, 3장에서는 비터비 알고리즘을 이용한 이진 마스크 후처리 과정을 제안한다. 4장에서는 제안하는 알고리즘의 성능 검증을 위한 실험 결과에 대해 기술한다.

II. 학습 기반의 이진 마스크 추정

본 장에서는 잡음 환경에서 녹음된 음성 신호를 시간-주파수 영역으로 분해하여 각 유닛의 이진 마스크를 추정하는 과정에 대해 설명한다. 이진 마스크의 추정은 특징 벡터 추출과 학습 및 판별 과정을 거치게 된다. 특징 벡터로 사용하는 진폭 변조 스펙트로그램과 분류 모델 학습 과정에 대해 설명한다.

1. 특징 벡터: 진폭 변조 스펙트로그램

본 논문에서 사용하는 특징 벡터는 Tchorz와 Kollmeier가 사용했던 진폭 변조 스펙트로그램 방법^[17-19]을 기초로 하고 있으며 특징 벡터 추출과정은 아래와 같이 정리할 수 있다.

마이크로 입력되는 신호에 대해 4ms 크기의 Hann 윈도우

우를 적용한다¹⁾. 본 논문의 실험에서 사용한 데이터의 표본 주파수는 12kHz이고, 이 때 4ms는 48샘플에 해당한다. 80개의 제로를 추가하여 128 샘플을 만든 후, 128 포인트 FFT (Fast Fourier Transform)를 적용한다. 그 다음 FFT 계수의 절대값에 제곱을 취하여 전력 스펙트럼 (power spectrum)을 얻는다. 0.25ms가 진행될 때마다 위에서 언급한 Hann 윈도우를 적용하여 FFT를 수행하게 되며, 이렇게 되면 전력 스펙트럼의 크기에 대한 표본 주파수는 4kHz (1/0.25ms)가 된다.

인간의 청각 특성을 반영하기 위해 mel 스케일 주파수^[20]

를 기반으로 한 25개의 필터로 구성된 필터뱅크를 전력 스펙트럼에 적용한다. 이렇게 해서 25개의 mel 스케일 주파수 채널에 대해 4kHz로 표본화된 전력 스펙트럼을 얻게 된다.

각 주파수 채널에 대해 32ms (128샘플) 동안의 전력 스펙트럼들을 모아서 다시 Hann 윈도우를 적용하고, 이러한 과정을 16ms마다 진행하여 수행하게 된다. 이제 Hann 윈도우를 적용한 128개의 데이터에 128개의 제로를 추가하여 256 샘플을 만든 후, 256 포인트 FFT를 적용한다. 이후, 절대값을 취하여 진폭 변조 스펙트럼을 얻는다.

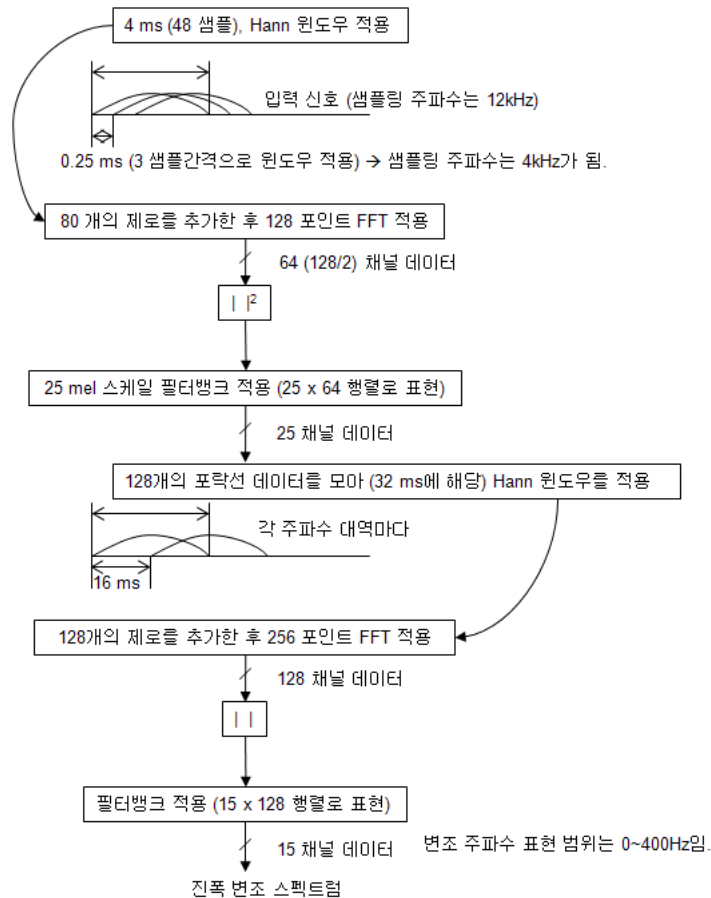


그림 1. 진폭 변조 스펙트로그램의 추출 과정
 Fig. 1. Extraction of amplitude modulation spectrogram

1) Hamming 윈도우를 적용해도 유사한 결과를 얻을 수 있다.

특징벡터의 크기를 줄이기 위해 15개의 삼각필터로 이루어진 필터뱅크를 적용하여 15차원의 진폭 변조 스펙트럼을 얻는다. 주파수 채널이 모두 25개이므로 전체 주파수 대역에 대한 진폭 변조 스펙트럼 특징벡터의 크기는 25×15 가 된다.

특징벡터를 추출하는 이상의 과정을 그림 1에 정리하였다.

위와 같이 얻어진 진폭 변조 스펙트로그램에 대해 시간 영역과 주파수 영역에서의 차분값도 특징 벡터로 고려하였다^[11].

2. 이진 마스크 분류 모델

본 연구에서는 이진 마스크의 분류를 위해서 기존에 제안되었던 베이지안 (Bayesian) 분류기를 사용하였다^[11]. 즉, 마스크 "0"에 해당하는 확률 모델과 마스크 "1"에 해당하는 확률 모델을 생성하고, 테스트 음성의 특징 벡터가 주어졌을 때, 마스크 "0"에 해당할 사후 (a posteriori) 확률과 마스크 "1"에 해당할 사후 확률을 각각 구하여 서로 비교하여 이진 마스크를 추정한다. 각 이진 마스크의 확률 모델로는 가우시안 혼합 모델 (Gaussian Mixture Model: GMM)을 사용하였다. 마스크 "0"과 "1"은 시간-주파수 영역의 로컬 SNR과 문턱값을 비교하여 정하게 되는데, 25개의 주파수 대역 중 처음 15개의 주파수 대역에 대해서는 -8dB의

문턱값을 사용하고 이후 10개의 고주파 대역에 대해서는 -16dB의 문턱값을 사용하였다. 고주파 대역에 위치하는 자음 성분들이 제거되지 않고 남아 있도록 하기 위해 고주파 대역에 보다 낮은 문턱값을 사용하였다. 기존의 모델 학습[11]에서와 같이 256개의 가우시안 분포를 사용하였다. 이진 마스크 분류에 대한 수식 및 보다 구체적인 내용에 대해서는 기존에 발표되었던 논문^[11]을 참고하기 바란다.

III. 주파수 종속성을 이용한 후처리 과정

주파수 영역에서의 음성신호는 잡음신호와 어느 정도 구별되는 패턴을 갖고 있다고 볼 수 있다. 이는 음성신호가 어느 정도 주파수 대역 간에 종속성을 띠고 있기 때문인데, 이러한 현상으로 볼 때, 각 시간 프레임에서의 이진 마스크의 패턴 역시 주파수 종속성으로 인하여 자주 관찰되는 패턴과 상대적으로 잘 나타나지 않는 패턴이 있을 것으로 추정할 수 있다. 그러나 이전 연구에서의 이진 마스크 추정방법에서는 각 시간 프레임에서 각 주파수 대역의 이진 마스크를 독립적으로 추정하므로 각 시간 프레임에서 추정된 이진 마스크의 패턴이 제한없이 나타나게 된다. 본 논문에서는 이진 마스크의 주파수 종속성을 이용하기 위해 각 주파수 대역간의 이진 마스크 상태 천이 확률을 추정한 후, 비터비(Viterbi) 알고리즘을 이용하여 추정된 이진 마스크

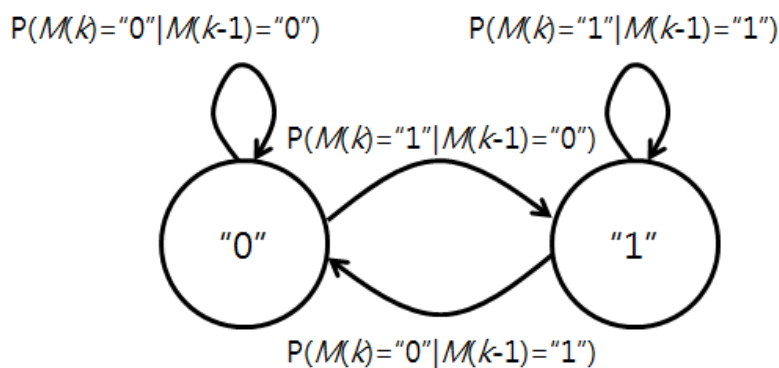


그림 2. 상태 천이
Fig. 2. State transition

를 수정하는 후처리 방법을 제안한다.

비터비 알고리즘은 상태 천이 확률 (state transition probability)과 관찰 확률 (observation probability)을 고려하여 가장 높은 likelihood를 갖는 상태 시퀀스 (state sequence)을 구하는 동적 알고리즘^[15,16]으로서, 학습 데이터로부터 구해진 상태 천이 확률과 이진 마스크 추정과정에서 얻어지는 관찰확률을 이용하여 각 프레임에서 추정된 이진 마스크 시퀀스에 대해 비터비 알고리즘을 적용하는 후처리 방법을 제안한다. II장에서 설명한 바와 같이 각 시간 프레임에 대해 주파수 영역은 25개의 대역으로 나누어진다. 이진 마스크를 적용하므로 각 주파수 대역은 “0” 또는 “1”의 마스크 값을 갖게 되고 이를 두 가지 “상태”(state)로 간주한다. 그림 2에 상태 천이에 대해 나타내었다. 그림에서 k는 주파수 대역을 나타내고 M(k)는 k번째 주파수 대역에서의 이진 마스크 값을 나타낸다. 이러한 두 가지 상태를 이동하는 상태 천이 확률과 각 데이터에 대해 마스크 값이 “0”이 될 확률과 마스크 값이 “1”이 될 확률인 관찰 확률을 이용하여 비터비 디코딩을 수행한다. 먼저, 상태 천이 확률을 구하기 위해서는 다음과 같은 일련의 과정들을 거치게 된다.

- a. 학습 데이터로부터 이상적인 (ideal) 이진 마스크를 구한다. 백색 잡음이 더해진 음성 데이터를 학습 데이터로 사용하며, 신호 대 잡음 비 (SNR)는 -15, -10, -5, 0, 5, 10, 15, 20, 25 dB를 사용한다. 이상적인 이진 마스크를 구하기 위해서는, 학습 데이터를 시간-주파수 분해하여 각 시간-주파수 영역의 신호 대 잡음 비가 문턱값보다 큰 경우는 이진 마스크 “1”을 부여하고 그렇지 않은 경우는 이진 마스크 “0”을 부여한다.
- b. 학습 데이터의 각 시간 프레임에서 25 개의 주파수 대역에 대해 구해진 이진 마스크를 하나의 벡터로 두어 K개의 클러스터로 분류한다. 이 때 클러스터링 방법으로는 k-means 클러스터링을 이용한다. 본 실험에서는 32개의 클러스터를 사용하였다.

c. 이진 마스크 벡터를 k-means 클러스터링 기법으로 분류하기 위해서는 이진 마스크 간의 거리를 측정해야 한다. 이진 마스크 벡터는 0과 1로 나타나는 이진 (binary) 벡터이다. 이진 벡터의 거리를 측정하는 방법은 여러 가지가 있는데^[21] 본 논문의 실험에서는 Hamming distance를 사용하였다²⁾. Hamming distance는 두 개의 스트링이 있을 때 같은 위치에 있는 심볼이 다른 것의 개수를 세어 계산할 수 있다.

d. 각 클러스터에 속한 학습 데이터의 이진 마스크 벡터들로부터 상태 이동 횟수를 계산하여 상태 천이 확률을 추정한다.

관찰 확률은 II장에서 소개한 베이지안 분류기를 적용할 때 구해진다. 즉, 특징 벡터를 GMM에 적용하여 likelihood를 구한 다음, 사전 확률을 곱하여 관찰 확률을 구한다. 이렇게 구해진 관찰 확률과 상태 천이 확률을 바탕으로 비터비 알고리즘을 수행하여 최적의 상태 시퀀스 즉, 현재 시간 프레임에서의 각 주파수 대역의 이진 마스크를 구하게 된다.

IV. 실험 결과

본 논문에서 제안한 알고리즘의 성능을 검증하기 위해 IEEE 문장 데이터 베이스^[22]를 이용하였다. 데이터 베이스에 있는 문장들은 영어 문장이며 하나의 문장에 작게는 5개, 많게는 12개의 단어들을 포함하고 있다. 문장들은 최초 25kHz 표본화주파수로 녹음된 것을 12kHz로 다운 샘플링하여 사용하였다. GMM 학습을 위해서는 390개의 문장을 -5, 0, 5 dB SNR로 잡음을 섞어서 사용하였는데, 본 실험에서는 NOISEX^[23]에 포함된 factory 잡음과 20명 음성을 섞어 만든 babble 잡음을 사용하였다. 테스트 데이터는 GMM 학습에 사용되지 않은 60개의 문장을 사용하였는데, -5 dB SNR³⁾로 잡음을 섞어 사용하였다.

2) 두 이진 벡터의 거리를 측정하는 다른 방법들도 적용하였을 때 대부분 미미한 차이를 나타내는데, 본 실험에서는 Hamming distance가 가장 좋은 결과를 보여주었다.

음성 명료도 평가를 위해서는 기존의 방법^{[11,12]}}에서와 같이 정상청력을 가진 피험자들에게 음성 파일을 들려주고 들은 것을 받아 적게 하여 얼마나 많은 단어들인 인식되었는지를 계산하는 실험을 진행하는 것이 바람직하다. 즉, 잡음 제거 전의 잡음 환경 데이터를 들려주어 인식률을 계산한 다음, 잡음 제거 알고리즘을 거치고 난 후의 데이터를 들려주어 인식률을 계산하여 결과를 비교할 수 있다. 그러나 결과 평가를 위해 피험자들을 모집하여 실험하는 것은 시간과 비용측면에서 부담이 될 뿐만 아니라 비교적 큰 분산 오차를 보이므로 큰 성능 향상이 있는 경우를 제외하고는 의미있는 실험결과를 도출하기 어려운 단점이 있다. 특히 본 연구와 같이 후처리를 추가하여 기존 알고리즘을 추가적으로 향상시키는 경우는 피험자들을 통한 청취 테스트로는 성능을 검증하기가 쉽지 않다.

따라서 본 논문에서는 청취테스트 대신 추정된 이진 마스크와 이상적인 이진 마스크를 비교하여 그 오차로서 성능을 평가하였다. 이상적인 이진 마스크는 각 시간-주파수 영역의 로컬 SNR을 직접 측정⁴⁾하여 기준 SNR과 비교하여 구할 수 있는데, 로컬 SNR이 기준 SNR보다 큰 경우는 이진 마스크를 “1”로 두고, 그렇지 않은 경우는 이진 마스크를 “0”으로 둔다^[24]. 이진 마스크 추정 오차는 두 가지 타입이 존재하는데 첫 번째 타입은 이상적인 마스크가 “1”일 때 추정된 마스크가 “0”으로 나오는 경우이고 이를 miss라 부른다. 두 번째 오차 타입은 이상적인 마스크가 “0”일 때 추정된 마스크가 “1”인 경우인데 이를 false alarm이라 부른다. 반면, 이상적인 마스크가 “1”일 때 추정된 마스크가 “1”로 제대로 추정된 경우를 hit라 하는데 이진 마스크 추정 성능을 hit rate에서 false alarm rate를 뺀 값으로 가늠할 수 있다. 또한 hit rate - false alarm rate는 1-(miss rate + false alarm rate)와 같으며, 이 값이 클수록 이진 마스크 추정기가 좋은 성능을 보임을 알 수 있다. 이 측정값과 음성 명료도 간에는 높은 상관 관계가 있음이 알려져 있으므로 hit rate - false alarm rate가 향상된다면 음성 명료도가 향

상될 가능성이 높다고 할 수 있을 것이다^[11]. 표 1과 2에 babble 잡음과 factory 잡음 환경에서의 hit rate - false alarm rate를 후처리 적용 전후로 나누어 비교하였다. 표에서 나타난 바와 같이 두 가지 잡음 환경 모두 제안된 후처리를 거침으로써 hit rate를 높이면서 동시에 false alarm rate를 낮출 수 있었다. 제안된 후처리 과정을 통해 두 가지 잡음 환경에 대해 hit rate - false alarm rate를 약 4% 이상 향상시킬 수 있는 것으로 확인되었다.

표 1. Babble 잡음 환경 실험 결과
Table 1. Experimental result for babble

	hit rate	false alarm rate	Hit rate - false alarm rate
후처리 적용 전	86.27 %	11.97 %	74.30 %
후처리 적용 후	89.99 %	11.38 %	78.61 %

표 2. Factory 잡음 환경 실험 결과
Table 2. Experimental result for factory noise

	hit	false alarm rate	hit rate - false alarm rate
후처리 적용 전	75.14 %	10.61 %	64.53 %
후처리 적용 후	77.08 %	8.32 %	68.76 %

V. 결 론

본 논문에서는 잡음 환경에서의 음성 명료도 향상을 위한 이진 마스크 적용을 다루고 있다. 학습된 모델을 이용하여 추정된 이진 마스크의 오차를 줄이기 위한 후처리 방법을 제안하였다. 시간-주파수 영역 이진 마스크의 주파수 종속성을 고려하여 비터비 알고리즘으로 최적화된 이진 마스크 시퀀스를 추정하였다. 제안된 방법의 성능 검증을 위해 -5 dB SNR babble, factory 잡음 환경에서 이진 마스크 추정 오차를 측정된 결과, 제안된 후처리를 거치지 않은 경우에 비해 두 가지 타입의 오차인 miss와 false alarm을 모

3) 정상적인 청력을 가진 사람은 0dB 이상의 잡음 환경에서는 문장에 담긴 단어들 거의 이해하는 것으로 알려져 있다^[24]. 따라서 음성 명료도 향상을 논하는 실험에서는 0dB보다 낮은 SNR의 데이터를 이용하여 테스트하는 것이 바람직하다.
4) 테스트에 사용된 데이터는 잡음이 없는 상태에서 녹음한 음성에 원하는 SNR을 갖도록 잡음을 더하여 사용하여 생성되었으므로 각 시간-주파수 영역에서의 잡음의 양과 음성의 양을 정확하게 알 수 있다. 이를 이용하여 각 시간-주파수 영역에서의 로컬 SNR값을 정확하게 측정할 수 있다.

두 감소시키는 것으로 나타났다. 또한 음성 명료도와 높은 상관 관계를 갖는 hit rate - false alarm rate를 4% 가량 향상시키는 것으로 확인되었다.

참 고 문 헌

- [1] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586 - 1604, 1979.
- [2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans.on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, no. 2, pp. 113 - 120, 1979.
- [3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans.on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109 - 1121, 1984.
- [4] Y. Ephraim and H. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans.on Speech and Audio Processing*, vol. 3, no. 4, pp. 251 - 266, 1995.
- [5] J. Huang and Y. Zhao, "An energy-constrained signal subspace method for speech enhancement and recognition in white and colored noises," *Speech Communication*, vol. 26, no. 3, pp. 165 - 181, Nov. 1998.
- [6] K. Hermus, P. Wambacq, and H. Hamme, "A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 045821, 2007.
- [7] M. Brandstein and D. Ward (Eds.), *Microphone Arrays*, Springer-Verlag, 2001.
- [8] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech communication*, vol. 49, no. 7, pp. 588 - 601, Jul. 2007.
- [9] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. on Speech and Audio Processing*, vol. 16, no. 1, pp. 229 - 238, 2008.
- [10] Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms." *J. Acoust. Soc. Am.*, vol. 122, no. 3, p. 1777, Sep. 2007.
- [11] G. Kim, Y. Lu, Y. Hu and P. C. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," *J. Acoust. Soc. Am.*, vol. 126, no. 3, pp. 1486-1494, September 2009.
- [12] G. Kim and P. C. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE trans. Audio, Speech and Language Processing*, vol. 18, no. 8, pp. 2080-2090, November 2010.
- [13] G. Brown and M. Cooke, "Computational auditory scene analysis," *Computer speech and language*, vol. 8, pp. 297 - 336, 1994.
- [14] D. Wang and G. Brown, *Computational Auditory Scene Analysis : Principles, Algorithms, and Applications*, Wiley, Hoboken, NJ, 2006.
- [15] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Information Theory*, vol. 13, Issue 2, pp. 260-269, April 1967.
- [16] G. Forney, Jr., "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, Issue 3, pp. 268-278, March 1973.
- [17] J. Tchorz and B. Kollmeier, "Estimation of the signal-to-noise ratio with amplitude modulation spectrograms," *Speech Communication*, vol. 38, no. 1 - 2, pp. 1 - 17, Sep. 2002.
- [18] J. Tchorz and B. Kollmeier, "SNR estimation based on amplitude modulation analysis with applications to noise suppression," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 3, pp. 184 - 192, May 2003.
- [19] M. Kleinschmidt and V. Hohmann, "Sub-band SNR estimation using auditory feature processing," *Speech Communication*, vol. 39, no. 1 - 2, pp. 47 - 63, Jan. 2003.
- [20] S. Stevens, J. Volkman, and E. Newman, "A scale for the measurement of the psychological magnitude pitch," *J. Acoust. Soc. Am.*, vol. 8, no. 3, pp. 185-190, 1937.
- [21] B. Zhang and S. N. Srihari, "Binary vector dissimilarity measure for handwriting identification," *Proceeding of SPIE*, pp. 155-166, 2003.
- [22] IEEE, "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.*, vol. 17, pp. 225-246, 1969.
- [23] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A databased and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, pp. 247-251, 1993.
- [24] N. Li and P. C. Loizou, "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," *J. Acoust. Soc. Am.*, vol. 123, no. 3, pp. 1673-1682, March 2008.

저 자 소 개



김 기 백

- 1994년 : 서울대학교 전자공학과 학사
- 1996년 : 서울대학교 전자공학과 석사
- 2007년 : 서울대학교 전기컴퓨터공학부 박사
- 1996년 ~ 2000년 : LG전자기술원 연구원
- 2000년 ~ 2003년 : (주)보이스웨어 선임연구원
- 2008년 ~ 2010년 : Univ. of Texas at Dallas, Research Associate
- 2010년 ~ 2011년 : 대구대학교 전자공학부 전임강사
- 2011년 ~ 현재 : 송실대학교 조교수
- 주관심분야 : 음성신호처리, 영상신호처리, 멀티모달신호처리, 어레이신호처리