

새로운 서브밴드 추정-스펙트럼 차감법에 기반한 음성향상방법에 관한 연구

정회원 주상현*, 김수남**, 김기두**

A Study on Speech Enhancement Method Based on the New Spectral Subtraction with Subband Estimation

Sang-Hyun Joo*, Su-Nam Kim**, Ki-Doo Kim** *Regular Members*

요 약

이 논문에서는, 잡음환경에서의 음성 향상을 위해서 일반적인 주파수 차감법에 기반한 새로운 형태의 방법을 제안한다. 기존의 방법들이 각각의 주파수 성분에 대해 잡음 및 음성스펙트럼을 추정하는데 비해, 본 논문에서는 주파수 대역을 여러 개의 서브밴드로 대역을 나누어 각각의 서브밴드에 대해서 잡음 및 음성의 스펙트럼을 추정한다. 본 논문에서는 잡음 스펙트럼을 추정하기 위하여 최소추적(Minima Tracking) 방법을 선택하였고, 필터링 방법으로는 스펙트럼 차감법에 기반한 Mel-Scaled 필터뱅크를 이용한 새로운 방법을 제안하였다. 모의실험결과, 기존의 방법들에 비해 음성구간에서의 SNR의 향상 정도는 입력 SNR이 -10~4dB의 범위에서 향상된 결과를 얻었다. 또한 전 구간에 대해서도 다른 알고리즘들 보다 향상된 결과를 얻었다.

ABSTRACT

In this paper, we suggest a new type of speech enhancement method based on the general spectral subtraction in noise environments. The conventional spectral subtraction based algorithm are mostly related to the estimates noise or speech spectrum in each frequency bin, but the purpose of the proposed method is to estimates noise or speech spectrum for each subband. When it comes to determining subbands, it can be easily changed into the preprocessing of speech recognition in that Mel Scaled Filter Bank is used, which is modeled on human auditory nerve. In this paper, we propose a new method using the Mel Scaled Filter Bank based on the spectral subtraction for filtering and we chose to use a minima tracking algorithm for estimating noise spectrum. From the simulation, we obtained -10~4dB SNR performance improvement compared with general method and it is also shown that better result is achieved than other noise cancelling algorithms through the entire region.

1. 서론

헬즈프리, 이동전화에서의 음질향상, 그리고 음성 인식시스템에서의 음성 특징추출 등에서 잡음환경의 음성향상은 중요한 요소로 자리잡고 있다. 잡음환경에서 음성향상을 하는 방법은 일반적으로 크게 두 가지 부류로 나눌 수 있다. 70년대 후반 Steven F. Boll에 의해 처음 소개된 스펙트럼 차감법^[1]과 적응

신호처리 알고리즘에 기반한 방법이다^[2]. 이들은 구 현방법은 다르지만 두 방법 모두 처리하고자 하는 잡음의 특성이 음성에 대해 부가적이면서도 독립적이고 상관성이 없다는 가정에서 출발한다. 그리고, 잡음의 스펙트럼이나 파형을 얼마만큼 정확하게 추정해 내는가에 따라 잡음환경에서 음성만을 정확하게 추정할 수 있다.

스펙트럼 차감법의 경우, 구조가 간단하고 다른

* (주) 보나테크 (dugaegol@bonatech.co.kr)
논문번호 : 010140-0611 접수일자 : 2001년 6월 11일

** 국민대학교 전자공학부 (kdk@dsp.kookmin.ac.kr)

알고리즘과의 연계에 대해 유연하다. 그런 이유로 Boll의 의해 처음 발표된 스펙트럼 차감법에 기반한 많은 방법들이 소개되었다^{[3][4][5][6]}. 이 스펙트럼 차감법은 단시간 파워 스펙트럼의 추정기술에 기반한 것으로써 음성향상을 위한 알고리즘으로 많이 사용되고 있다.

스펙트럼 차감법에 의한 잡음제거 효과는 “잡음의 랜덤한 성질에 기인하는 평균에 대한 변동정도를 어느 정도 줄일 수 있는가” 와 “차감의 결과로 생기는 처리왜곡을 어느 정도 줄일 수 있는가”에 달려 있다.^{[7][9]}

잡음의 스펙트럼을 추정하기 위해서는 일반적으로 음성/비음성 구간을 판단할 수 있는 도구인 VAD(Voice Activity Detector)^{[10][11]}를 사용한다. VAD의 경우 프레임 에너지와 영 교차율에 기반을 둔 방법들이 구현상의 간단함을 이유로 많이 사용되어 왔으나, VAD를 사용할 경우에 음성/비음성구간을 판단하는데 오차가 생기게 되면 심각한 결과를 낼 수도 있으므로, 본 논문에서는 이를 배제하기 위해 최소추정(minima tracking) 방법^{[4][5]}을 사용하였다. 이 방법은 잡음의 시변 특성이 음성에 비해 천천히 나타난다는 것을 이용한 것이다. 또한, 음소의 조합으로 되어있는 언어에서 항상 비음성 구간이 존재하게 된다는 가정에도 기인한다.

기존의 개선된 형태의 스펙트럼 차감법에서는 잡음 성분을 충분히 차감하기 위해 사용하는 과 차감 인자(over subtraction factor)의 크기를 추정 SNR에 의해 결정하기 위해서 각각의 주파수 성분들을 사용한다. 본 논문에서는 추정 SNR을 각각의 주파수 성분에 의해 구하지 않고 Mel Scale^{[3][10]}에 의한 서브밴드 단위로 추정 SNR을 구하는 방법을 제안한다. 이것은 각각의 주파수 성분에 의해 추정 SNR을 결정하는 방법에 비해 추정오차를 둔감하게 하고, 차감의 결과로 생기는 음성의 왜곡으로 인하여 듣기 거부해 지는 것을 완화시킬 수 있다.

알고리즘의 수행결과에서 “SNR의 향상정도”와 “처리왜곡의 정도를 어느 정도 줄일 수 있는가”는 서로 상충관계이다. 처리왜곡이란 원래의 음성과 필터링후의 음성이 주파수 영역에서의 특성이 조금씩 달라지는 것을 의미한다. 그러므로, 처리왜곡이 많은 필터링후의 음성은 사람의 귀로 들었을 때 부자연스러운 음성으로 들리게 된다. 반대로 처리왜곡을 줄이기 위해 주파수 영역에서의 잡음의 크기를 작게 설정하면, 잔여잡음이 많이 남게 되어서 SNR의 향상정도가 저게 된다.

SNR의 향상정도는 프레임단위로 계산하게 되며, 음성구간과 비음성구간을 모두 합친 전체 구간과 음성이 존재하는 구간으로 나뉘어서 평가하게 된다. SNR의 향상정도로 잡음 섞인 신호로부터 잡음을 얼마만큼 제거했는지를 평가하게 된다.

본 논문의 II장에서는 본 논문에서 제안한 서브밴드 추정 알고리즘의 개요와 시스템의 구현, III장에서는 기존의 방법들과 제안된 방법과의 성능비교를 하고, 마지막으로 IV장에서 결론으로 마무리를 짓는다.

II. 서브밴드 신호대 잡음비 추정에 의한 스펙트럼 차감법의 구현

2.1 시스템의 구조

본 논문에서 제안한 내용에 의해 구현된 시스템의 블록도를 그림 1에 나타내었다. 시스템의 입력으로 들어오는 신호는 8 kHz, 16 비트, 선형 양자화된 잡음 섞인 음성신호이다. 본 논문에서 구현된 시스템은 음성/비음성구간을 분리해 낼 필요가 없기 때문에 입력단에 VAD가 필요 없다. 따라서, 입력 신호는 잡음 신호일 수도, 잡음 섞인 신호일수도 있다. 이렇게 입력된 신호는 FFT에 의해 4 kHz 까지의 스펙트럼을 추정하게 되고 추정된 스펙트럼에서 위상을 저장하게 된다. 여기서 저장된 위상은 IFFT에 이용된다.

단구간 스펙트럼을 추정하기 위해서 입력신호는 프레임별로 나뉘어진다. 프레임별로 나뉘어진 후에는 해밍(Hamming) 윈도우를 거쳐 FFT에 의해 스펙트럼을 추정하게 된다. FFT에 의해 추정된 스펙트럼은 표 1의 Mel scale에 의해 각 서브밴드로 나누어지고, 각 서브밴드에 대해 스무딩 인자 ρ 를 곱하게 된다. 스무딩 인자 ρ 는 서브밴드에 상관없이 모두 같은 값이며, 잡음스펙트럼을 추정하기 위해 사용된다. 시스템의 입력부에서 음성/비음성 구간을 분리해 내는 부분이 없으므로, ρ 는 입력 신호의 파워 스펙트럼의 스무딩과 잡음 스펙트럼 추정을 위한 스무딩에 같이 사용되며 실험에 의해 값을 결정하였다. 잡음 스펙트럼의 추정은 일정시간동안 저장된 순환버퍼에 의해 행해지게 되며, 이 버퍼에서 각각의 주파수 성분들 중 가장 작은 값들이 각 서브밴드에서의 추정 잡음 스펙트럼이 된다.

잡음 스펙트럼의 추정은 그림 1의 블록도에서 볼 수 있듯이 Mel scale로 나뉘어진 각 서브밴드에 대해서 수행되며 이때, 과 차감 인자 a 는 각 서브밴

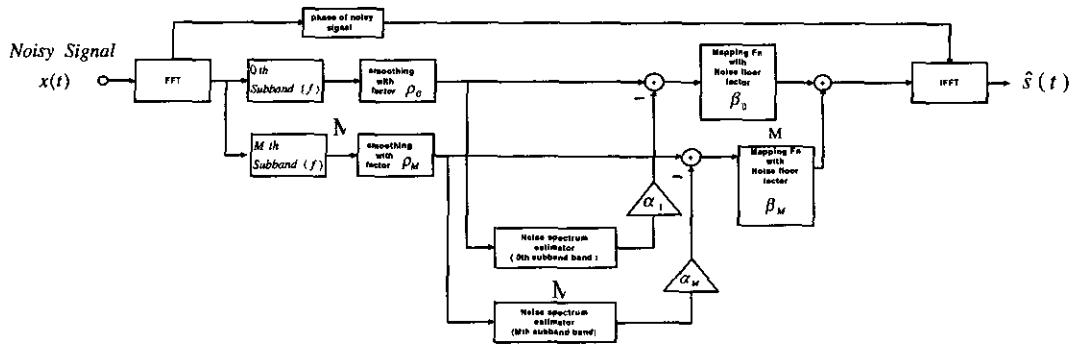


그림 1. 서브밴드 추정 스펙트럼 차감법의 구조

표 1. 구현된 Mel Scaled Filter Bank

Critical band index	FFT bins		Real frequency [Hz]
	간격	bin의 갯수	
1	1-3	3	0-94
2	4-6	3	94-187
3	7-10	4	187-312
4	11-13	3	312-406
5	14-16	3	406-500
6	17-20	4	500-625
7	21-25	5	625-781
8	26-29	4	781-906
9	30-35	6	906-1094
10	36-41	6	1094-1281
11	42-47	6	1281-1469
12	48-55	8	1469-1719
13	56-64	9	1719-2000
14	65-74	10	2000-2312
15	75-86	12	2312-2687
16	87-100	14	2687-3125
17	101-118	18	3125-3687
18	119-128	9	3687-4000

드의 추정 SNR에 의해 설정되게 된다. 이렇게 과 차감 인자 a 가 곱해진 추정된 잡음 스펙트럼을 스 무딩 인자 ρ 가 곱해진 입력 신호의 스펙트럼에서 차감 한 후, 식 (9)의 매핑 함수에 의해 차감의 결과 가 음의 값이 되는 것을 양의 값으로 조정 한 후 에 IFFT에 의해 시간 영역으로의 변환이 이루어진다.

2.2 단구간 파워 스펙트럼의 추정

한 프레임의 길이는 32 msec로, 256 샘플이고, 3/4 겹침을 하였다. 즉, 프레임 이동의 길이는 8 msec로 64 샘플씩 이동한다. 또한, 각 프레임은 해밍 윈도우에 의해 프레임의 양끝에서 생기는 블록

속점에 의한 누수현상을 줄였다.

본 논문에서의 윈도우 함수가 취해진 한 프레임의 스펙트럼은 식 (1)과 같이 표현된다. 이때 윈도우의 길이는 프레임의 길이와 동일한 256 샘플을 사용한다.

$$S_x(f) = \sum_{k=0}^{255} w(i-k)s(k)e^{-j2\pi f k/256} \quad (1)$$

여기서,

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{256-1}\right)$$

2.3 Mel Scale 서브밴드

사람은 소리를 들을 때 필터뱅크(filter bank)와 같은 특성을 갖게 된다^{[10][14]}. 이것은 일정 대역폭에 대하여 주파수 성분을 달리해도 사람은 그 차이를 명확하게 느끼지 못하는 것을 말하며, 이런 특성에 대해 공학적으로 사용되는 것이 Mel Scaled Filter Bank이다^[3].

본 논문에서는 Mel Scaled Filter Bank를 서브밴드의 SNR을 추정할 때에 사용하였는데, 그 이유는 과 차감 인자 a 의 값을 잡음환경에 적응시키기 위해서이다. 이와 같이 했을 때, 차감의 결과에 대한 왜곡이 Mel Scaled Filter Bank에 의해 생기기 때문에 각각의 주파수 성분에 의해 적응시킬 때 보다 SNR은 다소 줄어들더라도 스펙트럼의 왜곡을 서브밴드 단위로 둔감하게 해주어서 거친 소리를 줄여 주기 때문이다.

각 서브밴드에서의 에너지를 계산하기 위한 블록도를 그림 2에 나타내었고 각 서브밴드의 전달함수 $G_k(f)$ 는 다음과 같이 정의된다.

$$G_k(f) = \frac{1}{M_k} \sum_{i=0}^{M_k} F(i)W_k(i) \quad (2)$$

여기서, M_k 는 각 서브밴드에 포함되어 있는 주파수 성분의 개수이고 k 는 각 서브밴드의 순서를 나타낸다. $P(i)$ 는 0에서 127 까지의 주파수 성분들이고 $W_k(i)$ 는 주파수 영역에서 각 서브밴드로 분할하기 위한 윈도우 함수이다. 또한, L 개로 분할된 각 서브밴드의 평균 에너지를 $Sub(k)$ 로 표현했다. 본 논문에서 구현된 Mel scaled 필터뱅크를 표1에 나타내었다.

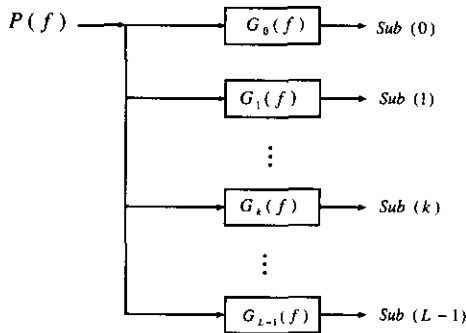


그림 2. L개의 채널에 의한 서브밴드 에너지평균의 계산

2.3.1 스무딩 인자 ρ 의 설정

스무딩 인자 ρ 는 입력신호의 스펙트럼 추정과 잡음 스펙트럼의 추정에 동시에 쓰이는 인자이다. ρ 는 입력신호의 주파수 성분의 크기가 시간에 따라 급격하게 변하는 것을 막아 차감의 결과로 특정 주파수 성분에서 큰 왜곡이 생기는 것을 막아준다.

$$\overline{P_X(i, f)} = \rho \overline{P_X(i-1, f)} + (1-\rho) P_X(i, f) \quad (3)$$

여기서, i 는 프레임 인덱스이다. 잡음 스펙트럼 추정 면에서 보면, 스무딩 인자 ρ 는 각각의 주파수 성분의 시간 평균을 구하기 위해 사용된다. 스무딩 인자 ρ 는 반복실험을 통해 가장 결과가 좋은 0.8의 값을 사용하였다.

2.3.2 단구간 잡음 파워 스펙트럼의 추정

단구간 잡음 파워 스펙트럼의 추정은 Martin의 최소추정 방법을 사용하며 M 개의 FIFO 버퍼를 사용한다.

$$\widehat{P}_N(k, f) = \min_{k=0}^{M-1} \overline{P_X(k-1, f)} \quad (4)$$

여기서, k 는 M 개의 버퍼에 대한 인덱스, f 는 각각의 주파수 성분을 가리킨다. $\overline{P_X}$ 는 M 개의 버퍼에 저장된 M 개의 시간 평균으로 식 (3)에 의해 계산되

며, 매 프레임마다 FIFO 버퍼링된다. 그러므로, M 개의 버퍼안에는 항상 최근 M 프레임에 대한 스펙트럼의 주파수 성분들이 존재하게 되며, 식 (4)에 의해 M 개의 버퍼에 들어있는 스펙트럼들 중에서 가장 작은 주파수 성분들만을 선택하여 추정 잡음 스펙트럼 \widehat{P}_N 을 구하게 된다.

M 을 64로 사용하면 계산량도 많아지고 차지하는 메모리의 양도 커지므로, 실제로 구현 할 때는 다음과 같은 방법을 사용하였다. 식 (5)와 식(6)은 추정 잡음스펙트럼을 구하기 위한 수식이다.

$$\overline{P_b(0, f)} = \rho \overline{P_N(k-1, f)} + (1-\rho) P_N(k, f) \quad (5)$$

$$\widehat{P}_N(k, f) = \min \left\{ \begin{matrix} M_1-1 \\ \min_{l=0} P_b(l, f), P_a(f) \end{matrix} \right\} \quad (6)$$

여기서, l 은 버퍼의 인덱스로 0에서 M_1-1 까지의 범위를 갖는 정수이다. M_1 은 M 보다 작은 값이 될 수 있다. P_a 는 보조적인 버퍼로 $M_1 M_2 = M$ 을 만족하는 M_2 개의 프레임마다 P_X 의 최소값을 저장해둔다. P_b 는 M_1 개의 버퍼로서 매 프레임 진행 할 때마다 저장되고 갱신되며, 이렇게 갱신된 스펙트럼들 중에서 최소가 되는 주파수 성분과 P_a 에 저장되어 있는 주파수 성분을 비교하여 최소가 되는 주파수 성분들을 구해서 스펙트럼을 추정하게 된다. 본 논문에서는 계산량을 줄이기 위하여 $M_1 = 8$, $M_2 = 8$ 로 설정하였다.

2.3.3 과 차감 인자 α 의 설정

과 차감 인자 α 는 다음과 같이 각 서브밴드의 추정 SNR을 기반으로 얻어지게 된다.

$$\alpha(k) = 1 + \widehat{SNR}_{Sub}(k) \quad (7)$$

$$\widehat{SNR}_{Sub}(k) = 10 \log_{10} \left\{ \frac{Sub_X(k) - \widehat{Sub}_N(k)}{\widehat{Sub}_N(k)} \right\} \quad (8)$$

여기서, $\alpha(k)$ 는 k 번째 서브밴드에 대한 과 차감 인자, $\widehat{SNR}_{Sub}(k)$ 은 k 번째 서브밴드에 대한 SNR 추정치, $Sub_X(k)$ 는 해당 서브밴드가 포함하는 주파수 성분들의 평균, $\widehat{Sub}_N(k)$ 는 해당 서브밴드가 포함하는 추정된 잡음 스펙트럼에 대한 주파수 성분들의 평균이다. 식 (7)과 식 (8)의 각 서브밴드에 대한 정보는 표 1에 따른다.

2.3.4 잡음 최저 인자 β 의 설정

잡음 최저 인자는 차감의 결과로 각각의 주파수 성분의 값이 음의 값이 되는 경우에 양의 값으로 만들어 주는 역할을 한다.

$$\hat{F}_s(f) = \begin{cases} \hat{F}_s(f), & \hat{F}_s(f) > \beta \hat{F}_x(f) \\ \beta \hat{F}_x(f), & \text{그 외} \end{cases} \quad (9)$$

여기서 $\hat{F}_s(f)$ 는 추정 음성신호의 스펙트럼, $\hat{F}_x(f)$ 는 잡음 섞인 신호의 스펙트럼을 나타낸다. 비 선형적인 매핑함수를 사용하는 이유는 “musical tone”이 생겼을 경우에 그 크기를 줄일 수 있기 때문이다. 본 논문에서 다루는 것이 파워 스펙트럼인 경우이므로, $\beta = 0.01$ 의 값을 선택했다.

III. 알고리즘의 비교

음성샘플은 8 kHz, 16 비트, 선형적으로 양자화된 것을 사용하였다. 음성데이터는 잡음이 없는 조용한 환경에서 녹음된 대화체 문장데이터를 사용하였고 잡음의 경우엔 주행중인 차안에서 녹음된 잡음을 사용하였다. 알고리즘들의 모의 실험을 위하여 잡음과 음성을 인위적으로 섞을 때 음성구간에 대하여 프레임 SNR을 설정하여 잡음과 음성을 섞을 수 있게 하였다. 모의 실험에서 테스트한 환경은 프레임 SNR이 10dB 보다 낮은 환경에 대한 것들이며 음성에 대한 잡음의 프레임 SNR이 -10 ~ 10dB 까지의 환경에 대해서 모의 실험을 하였다. 또한, 실험결과와 비교를 위하여, 스펙트럼차감법 + 반파 정류방법, 그리고 Martin의 방법과 Doblinger의 방법이 구현되었다. 각 알고리즘들은 오프라인으로 구현되었으며 제안한 알고리즘 외의 파라미터들에 대한 최적화를 위해 여러 번 반복실험을 통해 가장 좋은 결과를 출력 할 때의 파라미터 값들을 적용하여 실험을 하였다. 음성구간의 입력 프레임 SNR을 설정하기 위해서 VAD를 사용하여 음성구간을 찾은 후에 잡음을 설정한 SNR에 맞게 섞었으며 그 오차는 $\pm 0.7\text{dB}$ 이다. 잡음섞인 음성과 여러 가지 알고리즘의 수행결과를 시간영역에서의 파형으로 그림 5에 나타내었고, 그에 해당하는 스펙트로그램을 그림 6에 나타내었다.

구현된 서브밴드 추정 스펙트럼 차감법에 의한 잡음제거 평가는 입력 SNR에 대한 출력 SNR에 의한 SNR의 향상정도와 듣기 평가를 실행하였다. 한

프레임의 음성구간의 에너지에 대한 입력 SNR과 알고리즘들의 수행결과에 의한 출력 SNR을 그림 7에 나타내었다. SNR의 향상정도는 본 논문에서 제안한 방법이 다른 방법들보다 전체 구간의 SNR에서는 전반적으로 높은 값을 보였다. 음성구간에 대한 출력 SNR은 입력 SNR이 4dB가 되는 지점까지 Martin의 방법보다 향상된 것을 볼 수 있다. 식 (10)은 알고리즘 수행결과를 확인하기 위해서 프레임 SNR을 구하기 위한 수식이다.

$$SNR_{Frame} = \frac{1}{N} \sum_{i=0}^{N-1} 10 \log_{10} \left(\frac{\sum_{t=0}^{25} s^2(t)}{\sum_{t=0}^{25} (s(t) - \hat{s}(t))^2} \right) \quad (10)$$

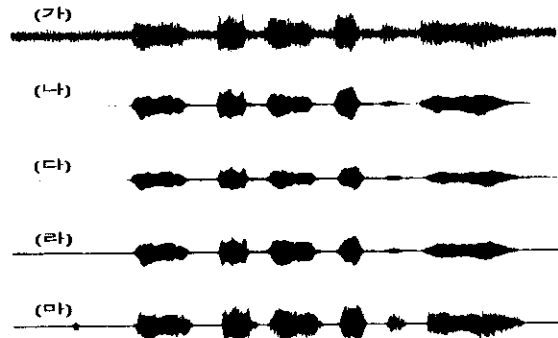


그림 5. 각 알고리즘별 수행결과, “총무과 좀 대주세요”에 대한 시간 영역에서의 파형 (가) 잡음섞인 신호 (SNR 0dB) (나) 제안된 알고리즘 (다) Martin의 알고리즘 (라) Doblinger의 알고리즘 (마) 반파 정류방법

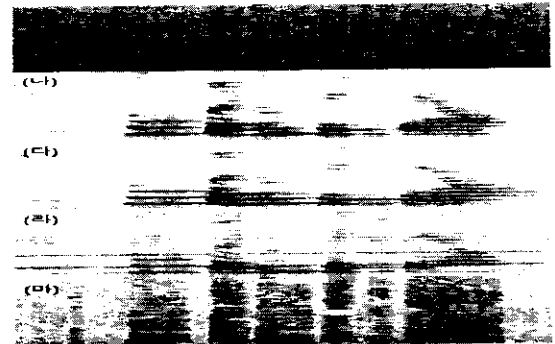


그림 6. 각 알고리즘별 수행결과, “총무과 좀 대주세요”에 대한 주파수 영역에서의 스펙트로그램 (가) 잡음섞인 신호(SNR 0dB) (나) 제안된 알고리즘 (다) Martin의 알고리즘 (라) Doblinger의 알고리즘 (마) 반파 정류방법

표 2는 음성구간에 대해 입력 SNR에 따른 각 알고리즘의 출력 SNR을 나타내며 식 (10)을 사용해 계산한 결과를 정리한 것이다. 그림 7은 음성구간에

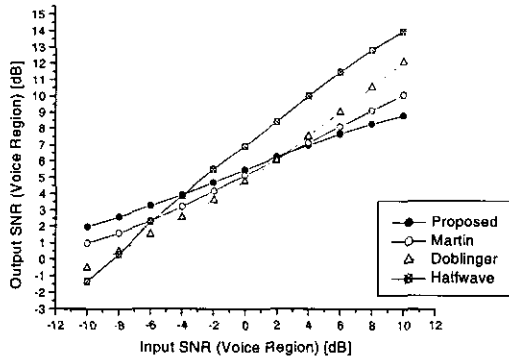


그림 7. 음성구간에서의 입력 SNR에 대한 출력 SNR

표 2. 음성 구간에 대한 입력 SNR에 따른 각 알고리즘의 출력 SNR

알고리즘 입력 SNR [dB]	제안 출력 SNR [dB]	Martin 출력 SNR [dB]	Dobligner 출력 SNR [dB]	Halfwave 출력 SNR [dB]
-10	1.95	0.97	-0.54	-1.36
-8	2.58	1.57	0.47	0.28
-6	3.25	2.31	1.49	2.25
-4	3.92	3.20	2.55	3.87
-2	4.67	4.16	3.59	5.51
0	5.47	5.12	4.76	6.91
2	6.23	6.11	6.02	8.40
4	6.97	7.11	7.51	10.00
6	7.66	8.10	9.01	11.48
8	8.28	9.10	9.10	12.83
10	8.81	10.08	12.09	13.96

표 3. 전 구간에 대한 입력 SNR에 따른 각 알고리즘의 출력 SNR

알고리즘 입력 SNR [dB]	제안 출력 SNR [dB]	Martin 출력 SNR [dB]	Dobligner 출력 SNR [dB]	Halfwave 출력 SNR [dB]
-10	-797.86	-820.48	-821.61	-822.20
-8	-797.41	-820.04	-820.87	-821.01
-6	-796.92	-819.50	-820.13	-819.57
-4	-796.43	-818.86	-819.35	-818.39
-2	-795.89	-818.16	-818.59	-817.28
0	-795.32	-817.47	-817.75	-816.18
2	-794.76	-816.75	-816.83	-815.10
4	-794.23	-816.03	-815.75	-813.94
6	-793.73	-793.42	-814.67	-812.87
8	-793.28	-792.69	-813.55	-811.89
10	-771.00	-791.98	-812.43	-811.07

대한 입력 SNR에 대한 출력 SNR을 보인다. 표 2의 결과는 제안된 알고리즘이 -10 ~ 4dB 범위에서 다른 알고리즘들 보다 프레임별 SNR이 향상되었음을 보인다. 4 ~ 10dB의 범위에서의 결과는 다른 알고리즘에 비해 프레임별 SNR 향상정도가 낮다.

표 3은 전 구간에 걸쳐 입력 SNR에 따른 각 알고리즘의 출력 SNR을 식 (10)에 의거하여 정리한 것이다. 이 결과는 제안된 알고리즘이 -10 ~ 10dB 범위에서 다른 알고리즘들 보다 프레임별 SNR이 향상되었음을 보인다. 그리고 비음성구간과 음성구간 전체에서의 잔여잡음을 다른 알고리즘들 보다 충분히 제거한다는 것을 보여준다.

VI. 결론

각각의 주파수 성분에 대해 추정 SNR을 구하여 과 차감 인자를 결정하게 되면, SNR이 낮은 경우에 원래의 음성의 스펙트럼에 대한 왜곡이 크게 일어나게 된다. 그러나, 제안된 알고리즘은 이러한 현상을 각 서브밴드에 대하여 행하기 때문에 특정 주파수 성분의 왜곡이 커지는 것을 감소시켜주는 효과가 있었다. 이로 인해 입력 SNR의 변화에 대해서 출력 SNR이 비교적 규칙적으로 변하면서도 입력 SNR이 4dB 보다 낮은 환경에서는 기존의 알고리즘들 보다 출력 SNR의 면에서 향상된 것을 볼 수 있었다.

또한, 잡음환경을 의미하는 입력 SNR에 대한 알고리즘 수행결과인 출력 SNR의 비율이 다른 알고리즘에 비해 완만한 기울기를 갖는데, 이는 잡음환경에 대해 본 논문에서 제안한 알고리즘의 수행결과가 크게 변동이 없음을 보인다. 이는 음성인식과 같은 분야에서 본 논문에서 제안한 알고리즘이 유용하게 사용될 수 있음을 입증하는 것이다.

스펙트럼 차감법에 기반한 음성향상기술은 음성 인식기가 실용화되기 위하여 해결해야 되는 혼란환경과 테스트환경의 불일치를 극복할 수 있는 도구로 전환하기가 쉽다. 본 논문에서 Mel Scale에 따른 Filter Bank를 고려한 이유도 차후, 음성인식기에 적용하기 위해서이다. 음성인식기에서 혼란환경과 실제환경인 잡음환경과의 불일치로 인해 인식이 저하되는 문제를 해결하기 위해서는 본 논문에서 구현된 알고리즘이 사용되는 음성인식기에서, SNR에 상관없이 같은 포락을 갖는 음성특징 벡터열을 추출할 수 있어야 한다. 차후, 본 논문에서의 연구결과를 바탕으로 인식기에 적용될 수 있는 잡음제거 방법의 연구가 요망된다.

참 고 문 헌

[1] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. ASSP*, ASSP-27, No. 2, pp. 113-120, April 1979.

[2] Jan Kybic, *Kalman Filtering and Speech Enhancement*, Diploma Work of Czech Technical University in Prague, January 1998.

[3] Nathalie Virag, "Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System," *IEEE Trans on Speech and Audio Processing*, Vol. 7, No. 2, March 1999.

[4] R. Martin, "Spectral Subtraction Based on Minimum Statistics," *Proceedings of EUSIPCO '94*.

[5] G. Doblinger, "Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands," *Proceedings of EUROSPEECH'95*.

[6] M. A. Tuffy, D. I. Laurenson, "Estimating Clean Speech Thresholds for Perceptual Based Speech Enhancement," *IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, New Paltz, New York, October 1999.

[7] S. V. Vaseghi, *Advanced Signal Processing and Digital Noise reduction*, John Wiley & Sons, USA, 1996.

[8] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, John Wiley & Sons, USA, 1978.

[9] Thomas W. Parsons, *Voice and Speech Processing*, McGraw-Hill, USA, 1986.

[10] Petr Pollak, Pavel Sovka, and Jan Uhlir, "Cepstral SPEECH/PAUSE Detectors," *In proc. of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, pp. 388-391, Neos Marmaras, Halkidiki, Greece, June 20-22, 1995.

[11] P. Sovka and P. Pollak, "The Study of SPEECH/PAUSE Detector for Speech Enhancement Methods," *Proceedings of the 4th European Conference on Speech*

Communication and Technology, pp. 1575-1578, Madrid, Spain, September 1995.

[12] A. M. Kondoz, *Digital Speech(Coding for Low Bit Rate Communications Systems)*, John Wiley & Sons Ltd, USA, 1994.

[13] A. V. Oppenheim, Ronald W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, USA, 1998.

[14] H. Fletcher, 'Auditory patterns', *Rev. Mod. Phys.* 12, pp. 47-65. 1940.

주 상 현(Sang-Hyun Joo)

정회원



1998년 8월 : 국민대학교
전자공학과 졸업
2000년 8월 : 국민대학교
전자공학과 석사
~현재 : (주)보나테크
주임 연구원

<주관심 분야> 멀티미디어 통신 신호처리

김 수 남(Su-Nam Kim)

준회원



1998년 2월 : 국민대학교
전자공학과 졸업
2000년 3월~현재 : 국민대학교
전자 공학과 석사과정
1998년 1월~2000년 2월 :
삼성전자 시스템 LSI
사업부 연구원

<주관심 분야> 디지털 신호처리, 디지털 통신

김 기 두(Ki-Doo Kim)

정회원



1980년 2월 : 서강대학교
전자공학과 졸업
1988년 8월 : The
Pennsylvania State
University, MS
(Electrical Eng.)

1990년 12월 : The Pennsylvania State
University, Ph.D. (Electrical Eng.)

1980년 3월~1985년 12월 : 국방과학연구소 연구원
1998년 2월~1999년 2월 : UCSD, CA 연구교수
1991년 2월~현재 : 국민대학교 전자공학부 교수
<주관심 분야> 디지털 신호처리, 디지털 통신