

ROEX 청각 필터를 이용한 단일채널 Speech Enhancement

김학윤*

*청주대학교 전자공학과

1 Channel Speech Enhancement using ROEX Auditory Filter

Kim, Hack-Yoon*

*Dept. Electronic Eng., Chongju Univ.

hykim@alpha94.chongju.ac.kr

요 약

배경 잡음에 의해 저하된 음성을 복원하는 기술은 이미 오래 전부터 여러 가지 기법들이 연구되어왔다. 이들 기법 중, Spectral Subtraction 기법은 단일 채널에 의한 Speech Enhancement의 대표적인 방법이다. 그러나, 기존의 단일 채널 Speech Enhancement 기법의 중요한 단점은 Musical Noise라 불리는 잔존 Noise의 발생 및 목적신호가 왜곡된다는 것이다. 이 잔존 Noise에 의해 지금까지 연구 보고된 단일 채널 Speech Enhancement 기법들은 거의 대부분 SNR은 향상되었지만 평균도의 향상이 곤란하였다고 보고되어왔다.

그러므로, 본 연구에서는 인간의 청각기구의 지각과정을 충실히 모방한 ROEX(Rounded Exponential) 청각 Filter를 이용하여 잔존 Noise인 Musical Noise를 억제시키는 기법을 제안하고자 한다.

1. 서 론

단구간 진폭 스펙트럼 추정 기법[1-4]에 의한 음성 강조 기법(Speech Enhancement)은 소규모의 하드웨어 및 적은 계산 량으로 실현 가능하므로, 가까운 장래 디지털 보청기등에 쉽게 도입할 수 있는 좋은 기법이라 할 수 있다. 그러나, 종래의 방법의 커다란 단점은, 잡음의 진폭 스펙트럼의 추정치가 실제값을 상회할 경우에는 이득이 필요이상으로 저하되고, 그것으로 인하여 처리후에 음색이 변형되게된다. 특히 이것이 현저할 경우에는 Musical Noise 형태의 새로운 잡음이 발생한다[3].

인간의 청각기구에서는, 임계영역 또는 청각

필터라고 불리는 통상 단구간 진폭 스펙트럼 추정에 이용되어지는 것보다도 넓은 폭을 가지는 영역 필터에 의해 주파수 분석이 수행되고 있다. 즉, 이 주파수 분석 기구에서는 필터의 대역폭 내에서 우세한 성분이 다른 것을 마스크(mask)하는 마스크(masking) 현상이 있다는 사실이 알려졌다[5]. 이런 관점을 토대로, 잡음 성분을 완전히 제거하지 않고서도 음성성분에 의해 마스크되는 레벨까지 감쇠 시킨다면 충분하다고 생각된다. 즉, 이와같은 마스크 개념을 도입하여 추정한 잡음의 진폭이 실제값을 상회할 경우가 발생하는 과도의 이득의 저하를 보정하고, 음성의 왜곡 현상을 보정 할 수 있다.

그러므로, 본연구에서는 인간의 청각구조와 비슷한 특성을 갖는 청각 필터 및 마스크(masking)의 개념을 도입하고, 이득이 필요 이상으로 저하되는 것을 방지하는 기법을 제안한다.

II. 청각 필터의 모델링

Patterson[6]은 Zwicker[7]가 수행한 임계대역을 구하는 실험에서는 청각 필터의 정확한 대역폭이 구해지지 않았을 가능성을 지적했다. 바꿔 말하면, 신호음이 필터의 중심주파수로부터 멀어짐에 따라, 신호음을 마스크하기 위해서 보다 넓은 대역폭의 잡음이 필요하게 되고, 임계대역폭이 실제의 폭보다 크게 평가되었을 가능성이 있다는 것을 보였다. 따라서, 그는 노치 노이즈를 이용해 청각 필터의 주파수 폭을 측정하는 방법을 제안하였고, 이렇게 구한 청각 필터를 정확하게 적용하기 위하여, 식(1)에 나타난 ROEX(ROunded EXponential)라고 불리는 함수를 이용해, 필터의 주파수 특성을 근사화 하였다.

$$W(g) = (1 - \gamma)(1 + \beta g) \exp(-\beta g) + \gamma \quad (1)$$

여기서, g 는 필터의 중심주파수로부터의 주파

수 차를, 중심 주파수로 정규화한 것이며, β 는 필터의 대역폭을 제어하는 파라미터이다. 또한, γ 는 필터의 경사선 부분의 형태를 제어하는 파라미터이다. 그림 1에 ROEX 필터의 주파수 특성을 나타냈다.

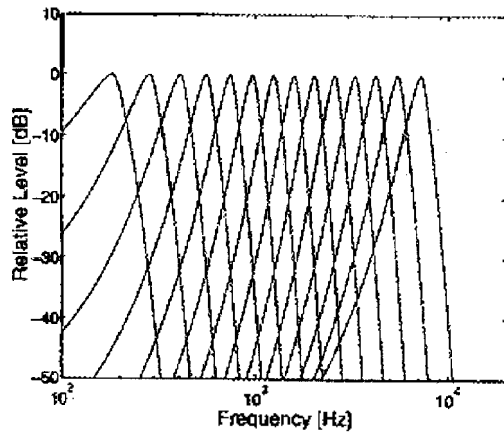


그림 1. ROEX 필터의 주파수 특성

Moore등은[5] 청각 필터의 주파수 응답으로부터 입력 대역폭에 상응하는 값을 구하기 위해서 청각 필터의 통과대역의 면적과 같은 면적을 가지는 구형의 폭(ERB: Equivalent Rectangular Bandwidth)을 계산하고, 또한, 청각 필터의 중심 주파수와 ERB의 관계식 및 ERB를 기준으로 하는 주파수 척도인 ERB-rate를 다음식과 같이 제안하였다.

$$ERB = 24.7(4.37f + 1) \quad (2)$$

$$ERB-rate = 21.4 \log_{10}(43.7f + 1) \quad (3)$$

여기서, f_c : 청각 필터의 중심 주파수(Hz)를 나타낸다. 즉, Moore등은[5], 식(2), 식(3)에 의해 주어지는 주파수 축상의 등간격으로 분포하는 ROEX 필터를 이용해, 이것들의 필터의 출력에 의한, 어떤 음에 대한 흥분패턴을 나타내는 것을 제안했다.

한편, Patterson[6]등은 청각 필터의 주파수 응답을 가상 모델로 한, Gammatone 필터를 제안했다. 이 필터의 임펄스 응답은 Gamma 함수를 포락선으로 한 정현파이나, 원래는 청신경의 임펄스 응답을 기술한 것이다.

반면, 청각 계의 필터링처리를 가상 모델로 한, 기저막의 진동계 서역 필터 특성을 모델화한 아날로그 달팽이모형과 필터의 주파수 특성의 입력 음압에 의한 변화도 고려한 적응 Q형 비선형 달팽이 필터 모델등이 제안되었다.

이상과 같이, 청각 필터에 관하여 여러 가지 방법이 연구되었으나, 어떤 방법이 최선이라는

논의는 아직 결론지어지지 않았다. 그러나, 휴대용 디지털보청기라는 계산능력이 제한되어진 상태에 적용을 고려한 음성강조를 목적으로 하는 경우, 이 주파수 특성이 식(1)과 같이 비교적 간단한 식으로 주어지면서 청각 필터의 형상을 한 눈사하고 있다고 생각되는 ROEX 필터를 이용하는 것이 유리하다고 생각된다.

III. 청각필터를 이용한 단일 채널 음성강조법

III.1 청각필터를 이용한 이득의 결정

우선 제 1단계로 현재 프레임의 음성신호를 이미 알고있다고 가정하고, 청각 필터를 이용하여 이득을 결정하는 기법에 관하여 서술한다.

그림 2는 모음을 입력한 경우의 단구간 파워 스펙트럼과 같은 신호의 청각필터의 출력을 비교한 것이다. 단, ROEX 필터의 대역폭을 1ERB로 하였고, 0.5 ERB-rate마다 중심주파수를 설정하였다. 그림에서 실선은 음성의 단구간 파워 스펙트럼을 나타내고, 점선은 음성신호를 청각 필터를 통과시킨 후의 각 필터의 출력파워를 필터의 중심 주파수의 함수로 나타낸 것이다.

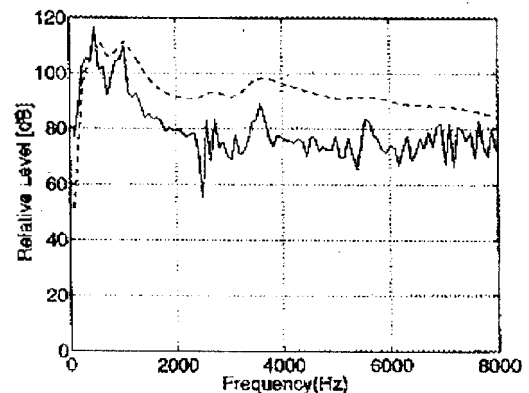


그림 2. 음성신호의 단구간 스펙트럼(실선)과 청각필터의 출력파워(점선)

앞 절에서 본한바와 같이, 청각필터 출력 레벨 이하에 존재하는 잡음성분은 음성에 의해 마스킹되어 들리지 않게 된다. 따라서, 음성신호의 파워에서 청각필터의 출력을 넘는 잡음성분만 억압하면 된다.

그림 3은 그림 2와 같은 분석을 잡음이 중첩된 음성에 수행한 것이다.

마스킹 현상에 따라, 잡음에 대한 청각 필터의 출력 파워가 음성 신호의 청각 필터 출력 파워보다 작은 경우에는 음성에 의한 잡음이 거의 마스킹되므로 이 레벨까지 잡음을 억압한다면

충분하다고 생각할 수 있다.

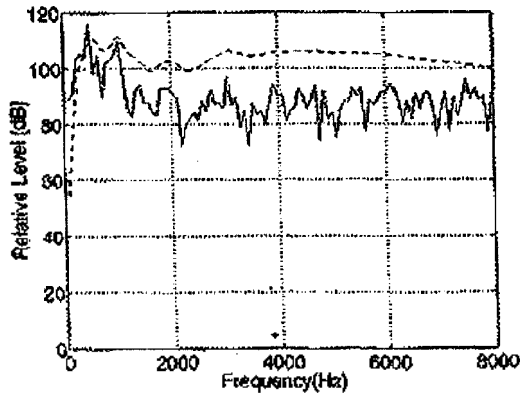


그림 3. 잡음이 중첩된 음성신호의 단구간 스펙트럼(실선)과 청각필터의 출력파워(점선)

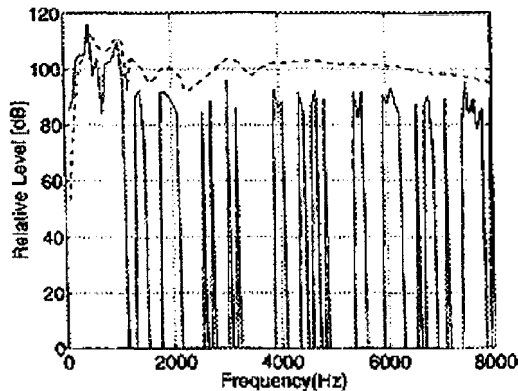


그림 4. Spectral Subtraction에 의해 추정된 스펙트럼(실선)과 청각필터의 출력파워(점선)

즉, 잡음이 중첩된 음성 $y_m(n)$ 에 대한 청각필터의 출력 파워와 원음성 신호 $s_m(n)$ 에 대한 청각필터의 출력파워를 각각 $A_{y,m}(k)$, $A_{s,m}(k)$ 이라고 한다면, 이득은 다음 식과 같이 된다.

$$H_{A,m}(k) = \left[\frac{A_{s,m}(k)}{A_{y,m}(k)} \right]^{\frac{1}{2}} \quad (4)$$

여기서, m 은 분석 프레임의 번호를, k 는 청각필터인 ROEX 필터의 ERB번호를 각각 나타낸다.

III.2 단구간 진폭 스펙트럼 추정을 이용한 이득 산출법

앞 실에서는 현재 프레임에서 음성신호를 이미 알고있다고 하고 이득을 산출하는 방법을 논하였다. 그러나, 실제의 음성신호는 미지이므로, 여기에서는 기존의 단구간 진폭스펙트럼 추

정법을 이용하여 현재 분석하고 있는 구간에서의 음성신호를 추정하고, 이것을 앞실에서 논한 이득 산출법에 적용하는 기법을 논한다.

단구간 진폭스펙트럼 추정법에 의하여 추정된 제 m 번째 구간에서의 추정값을 $\hat{s}_m(n)$ 라 하고, 이 추정값을 청각필터에 통과시킨 출력 파워를 $A_{\hat{s}_m}(k)$ 라 놓으면, 이득은 다음식과 같이 구할 수 있다.

$$\hat{H}_{A,m}(\omega) = \left[\frac{A_{\hat{s}_m}(k)}{A_{y,m}(k)} \right]^{\frac{1}{2}} \quad (5)$$

그림 4에 단시간 진폭스펙트럼 추정 방법으로 Spectral Subtraction을 이용한 경우의 예를 나타냈다. 그림에서 실선은 Spectral Subtraction에 의해 추정된 파워스펙트럼, 점선은 청각 필터의 출력파워를 나타낸다.

III.3 청각 필터를 이용한 음성강조법

그림 8에 앞 실에서 논한 이득을 이용하여 음성강조를 행하는 기법의 시스템 블록도를 나타냈다.

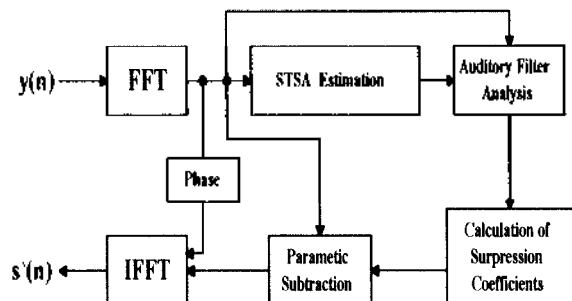


그림 5. 제안한 시스템 블록도(STSA: 단구간 진폭스펙트럼을 나타냄)

먼저, 현재 분석하고있는 프레임의 음성신호를 단구간 진폭스펙트럼법을 이용하여 가추정하고, 이 가추정된 음성신호를 청각 필터에 입력시켜, 출력 파워 $A_{\hat{s}_m}(k)$ 를 계산한다. 또한, 이 출력 파워를 이용하여 식(6)에서 이득을 계산한다. 이득을 이용하여 식(7)과 같이 필터량을 수행하고, 밴드 출력으로부터 각 이산주파수의 보강은 직선 보강을 이용한다.

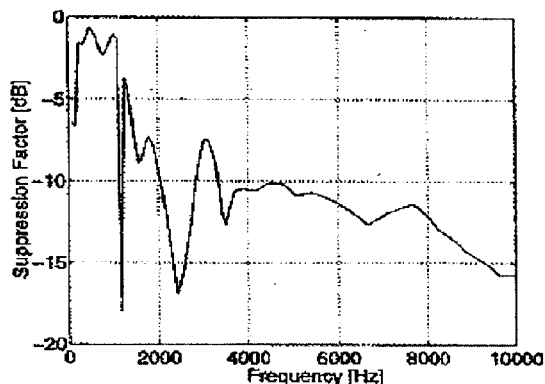
$$\hat{H}_{A,m}(k) = \left[\frac{A_{\hat{s}_m}(k)}{A_{y,m}(k)} \right]^{\frac{1}{2}} \quad (6)$$

$$\hat{S}_m(k) = \hat{H}_{A,m}(k) Y_m(k) \quad (7)$$

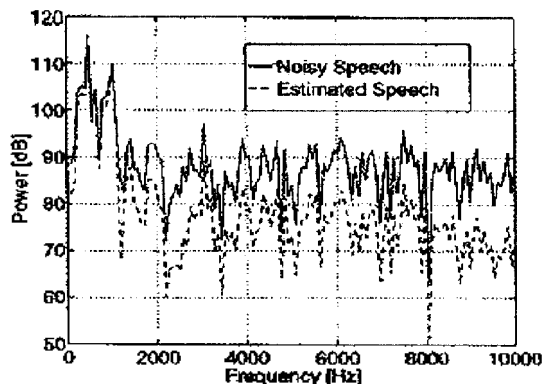
$$\hat{s}_m(n) = F^{-1}[\hat{S}_m(k)] \quad (8)$$

위 식에서 m 은 현재 처리하고있는 구간,

$\hat{H}_{A,m}(k)$ 은 구간 m 에서의 이득, $A_{s,m}(k)$ 은 추정된 음성신호에 대한 청각필터의 출력 파워, $A_{y,m}(k)$ 는 관측신호(잡음이 중첩된 음성)에 대한 청각필터의 출력과워를 나타낸다.



(a)



(b)

그림 6, 제안한 기법의 차리에: (a)이득, (b)처리 결과

제안한 기법을 그림 2와 같은 신호에 적용한 처리 예를 그림 6에 나타냈다.

그림 6(a)는 청각필터의 출력과워로부터 산출한 이득을 나타내고 있고, (b)는 처리의 결과를 나타내며, 실선은 잡음이 중첩된 음성의 파워스펙트럼, 점선은 제안한 기법에 의해 강조된 음성신호의 파워스펙트럼을 나타낸다.

본 논문에서는 복직음의 단구간 진폭스펙트럼을 기추정하는 방법으로 가장 간단한 방법인 Spectral Subtraction을 이용하였다. 이와같은 예비 추정 방법은, Spectral Subtraction 이외의 기법(예를 들면, Wiener Filter과 MMSE)을 적용하여 원 음성신호의 청각필터 출력 과워를 보다 정확한 예비추정이 가능하다면, Spectral Subtraction을 이용한 경우와 비교하면 보다 좋은 결과를 얻을 수 있다고 생각한다.

IV. 결 론

기존의 단구간 진폭스펙트럼 추정법에서는 잡음의 진폭스펙트럼의 추정치가 분석 프레임의 단구간 스펙트럼을 상회하는 경우, 이득이 필요이상으로 저하되고, 이것에 의해 처리 후의 음성 왜곡이 일어나는 것이 문제가 되고 있다. 이 왜곡이 현저한 경우는 musical noise가 발생한다.

그러므로, 본 논문에서는 청각 필터에서 유력한 성분이 다른 성분을 마스크하는 마스크의 원리를 토대로, 이득의 저하를 최소한으로 한정하고, 처리후 음성의 왜곡을 감소시키는 기법을 제안하였다.

시뮬레이션 결과 기존의 단구간 진폭스펙트럼 추정법의 하나인 Spectral Subtraction과 비교한 결과, 기존의 방법보다 제안된 기법이 극단적인 이득의 저하가 발생하지 않으며, 출력의 음성 왜곡 현상도 감소되었다.

참고문헌

1. S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," VOL. ASSP-27, APR., 1979
2. Yariv Epharim and David Malah, Speech Enhancement using a Minimum Mean-square Error Short-Time spectral Amplitude Estimator," VOL. ASSP-32, NO. 6, DEC., 1984
3. J. S. Lim and A. v. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," Pro., IEEE, VOL. 67, DEC 1979
4. R. J. McAulay, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter.," VOL. ASSP-28, NO. 2, APR., 1980
5. Moore, B. C. J. and Glasberg, B. R. "Suggested formulate for calculating auditory filter band-widths and excitation patterns," J. Acoust. Soc. Am. 74, pp. 750-753, 1983
6. Patterson, R., "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. 59, pp640-655, 1976
7. Zwicker, E. and Fastle, H., "Psychoacoustics," Springer-Verlag, befrin 1990