1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 2(s) 호

잔류잡음 감소를 위한 백색화 스펙트럼 차감법

Spectral Subtraction Using Whitening Filter for Reducing Residual Noise

오태호[0], 전범기, 성굉모
서울대학교 전기공학부

Tae-ho Oh[0], Bumki Jeon, and Koeng-Mo Sung
School of Electrical Engineering, Seoul National University

E-Mail : klavier5@acoustics.snu.ac.kr

요 약

음성의 음질 향상(Speech Enhancement)을 위한 여러 가지 방법 중에서 주파수 차감법(Spectral Subtraction)은 계산량이 적기 때문에 현재 실시간으로 Speech Enhancement를 할 수 있는 가장 적절한 방법이다. 그러나, 이 방법은 원래의 입력음성에 없던 새로운 잡음을 만들어내는 큰 단점이 있는데, 이를 제거하기 위해 많은 연구가 되어오고 있다. 이러한 연구의 방향은 대부분 주변 프레임 또는 주변의 주파수 성분과의 평균을 통해 피크값을 무디게 해 줌으로써 새로 생긴 튀는 잡음을 감소시키는 것이다. 이런 방법은 음성자체의 정보 또한 평균이 되어버리게 하는 새로운 단점을 낳는데, 이런 현상은 무성음구간에서 특히 심각해진다.

본 논문에서는 입력음성의 LPC 분석으로 백색필터(Whitening Filter)를 구성하여 이를 통과시킨 잔류신호(Residual)를 주파수 차감하여 얻은 새로운 잔류신호를 인 필터링하여(Synthesis Filter) 개선된 음성을 얻는 방법을 제안하였다. 제안된 알고리듬은, 주파수 차감시 포만트(Formant)의 정보가 더 유지 될 수 있기 때문에 잔류잡음을 줄일 수 있다. 청취 테스트 결과 제안한 방법이 기존의 방법보다 잔류잡음을 더 줄이는 사실을 확인할 수 있었다.

I. 서 론

주파수 차감법(Spectral Subtraction)에 의한 음성의 음질 향상(Speech Enhancement)은 계산량이 적기 때문에 실시간으로 음성을 Enhancement하는 가장 적절한 방법이다. 그러나 이 방법은 처리 후 새로운 잡음(흔히 Musical Noise라 부른다)을 만들어 내는 단점이 있다[1]. 이 새로운 잔류잡음은 입력음성신호의 스펙트럼에서 잡음의 스펙트럼의 크기를 빼는 과정에서 원래의 스펙트럼에는 없는 새로운 큰 값들이 생긴 것인데, 이를 제거하기 위한 연구를 살펴보면 다음과 같다. 우선 입력음성신호의 스펙트럼의 크기보다 미리 계산된 잡음의 스펙트럼의 크기가 더 클 경우에 처리후 신호의 크기를 0으로 잡는 경우(Half Wave Rectifier)가 보통이다.[2]. 또한, 새로 생긴 잔류잡음을 제거하기 위해 근처 프레임값의 평균을 구하거나 최소값을 잡아 신호를 부드럽게 만드는(smothing) 경우가 많다[2][9]. 이 외에도 시간-주파수의 윈도우를 만들어 주변과의 에너지의 차가 너무 큰 피크값을 제거하는 기법[3]과 심리음향의 측면을 고려한 논문[4]등도 나오고 있다. 이중 Ephraim의 Least Mean Square Error의 최소값을 이용한 논문[6][7]이 현재 가장 좋은 효과를 가지고 있음이 알려져 있다.

이런 방법은 잔류잡음을 어느 정도 줄여주는 효과를 가진다. 그러나, 이런 방법은 피크값을 없애서 신호를 부드럽게 만드는 것이기 때문에 원래 음성의 명료도를 크게 떨어뜨린다는 단점이 있다. 또한, 처리 과정에 필요한 몇가지 상수를 경험적으로 정해주기 때문에 상수에 따라 결과가 달라지게 되어[5] 일관되게 모든 경우에 적용할 수 없다.

본 논문에서는 주변 프레임과의 평균을 통한 Smoothing을 하지 않고도 새로이 생기는 피크값을 줄여주기 위해서, 잡음이 더해진 입력 음성을 LPC 분석으로 구성된 백색 필터(Whitening Filter)를 통과시킨 잔류신호(Residual)상에서 주파수 차감법을 쓰는 알고리듬을 제안했다. 이렇게 되면 포만트 정보는 차감되지 않기 때문에 Smoothing에 의한 음성정보의 훼손을 줄이면서도 주파수 차감시 새로 생기는 잔류잡음을 줄일 수 있다.

이 논문의 구성은 다음과 같다. 2장에서는 주파수 차감법의 개요와 현재 가장 좋은 성능을 가지고 있는 것으로 알려져 있는 Ephraim의 방법에 대해 알아본다. 3장에서는 본 논문에서 제안된 방법을 서술하고, 4장에서는 실험환경과 청취 테스트 결과를 기술하였다. 마지막으로 5장에서 결론을 내렸다.

Ⅱ. 주파수 차감법에 의한 Speech Enhancement

A. 주파수 차감법

깨끗한 음성을 $s(n)$, 더해진 잡음을 $d(n)$이라 하면 잡음이 더해진 음성 $y(n)$은 식(1)과 같이 나타내 진다.

$$y(n) = s(n) + d(n) \tag{1}$$

이 때, 이 음성의 한 구간을 푸리에 변환하면,

$$Y(\omega) = S(\omega) + D(\omega) \tag{2}$$

와 같이 된다. 양변에 절대값을 제곱하면

$$\begin{aligned} |Y(\omega)|^2 &= |S(\omega)|^2 + |D(\omega)|^2 \\ &+ S(\omega)\cdot D^*(\omega) + D(\omega)\cdot S^*(\omega) \end{aligned} \tag{3}$$

와 같이 되는데, 실제환경에서 잡음의 성분은 알 수 없으므로 잡음의 스펙트럼의 크기 $|D(\omega)|$는 음성이 없는 구간에서 잡음의 평균을 구해서 쓰게된다. 또, 잡음과 음성간에 Correlation이 없다고 가정하면 통계적으로 $E[S(\omega)\cdot D^*(\omega)]$와 $E[S^*(\omega)\cdot D(\omega)]$은 0이 되어 추출된 음성 스펙트럼의 크기를 구하는 식은 다음과 같이 간략하게 할 수 있다.

$$|S(\omega)|^2 = |Y(\omega)|^2 - E[|D(\omega)|^2] \tag{4}$$

그런데, 실제 실험에서는 다음과 같이 k와 a를 적당히 조정하면 더 좋은 결과가 주어진다는 것이 경험적으로 알려지 있다[5].

$$|S(\omega)|^a = |Y(\omega)|^a - kE[|D(\omega)|^a] \tag{5}$$

개선된 음성 스펙트럼의 크기는 식 5와같이 구하는 반면에 스펙트럼의 위상은 사람의 귀기 위상에 둔감하다는 성질을 이용해서 입력 음성의 위상을 그대로 쓰게 된다. 그러므로, 최종적으로 추출된 음성스펙트럼은 다음과 같이 된다

$$S(\omega) = |S(\omega)|\cdot \exp[j\angle Y(\omega)] \tag{6}$$

B. Ephraim-Malah의 방법

지금까지 언급했듯이 주파수 차감법은 흔히 Musical Noise라고 불리는 새로운 잔류잡음을 만들어낸다. 이 잔류잡음을 줄이기 위해 여러 가지 방법이 제시 되었는데, 이중 현재 가장 성능이 좋다고 알려진 방법이 Ephraim-Malah Noise Supression Rule(앞으로 EMSR로 줄여 부르겠다.)이다. 이것은 다음과 같이 음성의 스펙트럼의 절대값을 필터 $G(\omega)$를 통과시켜서 구하는 것이다.

$$|S(\omega)| = G(\omega)\cdot|Y(\omega)| \tag{8}$$

여기서 $G(\omega)$는 통계적으로 추정된 $|S(\omega)|$와 원래잡음이 없는 깨끗한 음성의 $|S(\omega)|$과의 Mean-square Error가 최소가 되게하도록 정해지는데, p번째 프레임의 주파수에 따른 G는 다음과같이 주어진다.

$$\begin{aligned} G(p,\omega) = \frac{\sqrt{\pi}}{2} &\sqrt{\left(\frac{1}{1+R_{post}}\right)\left(\frac{R_{prio}}{1+R_{prio}}\right)} \\ &* M[(1+R_{post})\left(\frac{R_{prio}}{1+R_{prio}}\right)] \end{aligned} \tag{9}$$

여기서 M은

$$M[\theta] = \exp(-\frac{\theta}{2})[(1+\theta)I_0(\frac{\theta}{2}) + \theta I_1(\frac{\theta}{2})] \tag{10}$$

과 같이 주어지는 함수이고, I_0와 I_1 각각 0차와 1차의 Modified 베셀 함수이다. 또, p번째 프레임의 a posteriori SNR과 a prioi SNR은 다음과 같이 주어진다.

$$R_{post}(p,\omega) = \frac{|Y(p,\omega)|^2}{E[|D(\omega)|^2]} - 1 \tag{11}$$

$$\begin{aligned} R_{prio}(p,\omega) = (1-\alpha)\ & R_{post}(p,\omega) \\ &+ \alpha\frac{|G(p,\omega)Y(p,\omega)|^2}{E[|D(\omega)|^2]} \end{aligned} \tag{12}$$

이 EMSR역시 앞 뒤 프레임간의 평균적인 스무딩 효과를 가지고 있음이 나중에 증명되었다[8].

Ⅲ. 잔류신호 상에서의 주파수 차감

주파수 차감법에서 새로 생기는 잔류잡음을 제거하기 위한 기존의 방법들은 주로 앞 뒤 프레임간의 스펙트럼의 평균을 이용한 스무딩 효과로 잔류 잡음을 제거한다. 이런 방법은 귀에 거슬리는 잔류잡음 제거에는 어느정도 효과를 가지지만, 스무딩으로 인해 음성의 명료도가 떨어지는 단점이 있다. 또한, 처리에 필요한 몇몇 상수를 경험적으로 정해주기 때문에 이 상수에 따라 결과가 달라져서 모든 경우에 일반적으로 적용할 수 없는 단점을 지닌다.

본 논문에서는 주변 프레임의 평균을 통하지 않고도

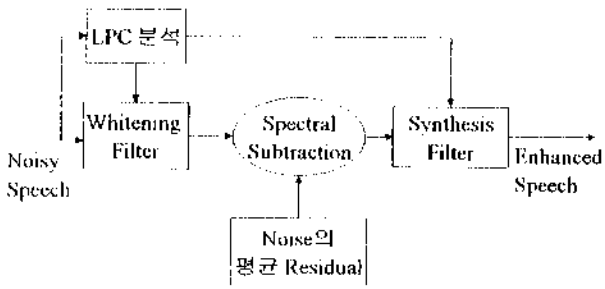

그림 1. 잔류신호상에서 주파수차감법의 흐름도

잔류잡음을 제거하기 위해서, 입력 신호의 포만트 정보는 유지하고 잔류신호만 주파수 차감하는 방법을 제안했다.

그림 1과 같이 입력신호의 LPC 분석으로 구성한 백색 필터를 통과한 잔류신호상의 스펙트럼에서 미리 구해둔 잡음의 평균 잔류신호의 스펙트럼을 차감해서 다시 합성 필터를 통과시키는 방법이다.

이 방법은 포만트 정보는 차감하지 않고, 또한, 상대적으로 크기가 작은 잔류신호상에서 주파수를 차감하는 것이기 때문에 새로이 생기는 잔류잡음의 크기가 상대적으로 작게된다. 그리고, 이 방법은 조절해야 할 상수가 없기 때문에 일반적인 경우에도 적용이 가능한 장점이 있다.

그림 2에 제안된 방법에 따른 Speech Enhancement의 과정을 나타냈다.

Ⅳ. 실험 환경 및 결과

미리 녹음된 깨끗한 음성신호에 백색잡음을 첨가해서 표 1에 제시된 Segmental SNR로 각각 잡음이 더해진 음성신호를 만들었다. 각각의 음성신호에 대해 기본적인 주파수 차감법 (SS), 잔류신호상에서의 기본적인 주파수 차감법 (LPC-SS), EMSR, 잔류신호 상에서의 EMSR (LPC-EMSR)을 구현하였다.

청취 테스트는 20명의 임의의 사람에게 같은 SegSNR의 SS와 LPC-SS중 어느것이 더 나은지를 고르게 해서 LPC-SS가 SS보다 나아진 것을 1점, 차이를 구별하기 힘든 것을 0점, 나빠진 것을 -1점을 준 뒤 모든 사람의 평균 점수를 표 2에 실었다. EMSR과 LPC-EMSR의 비교도 같은 방법으로 수행했다.

청취 테스트를 한 사람들은 대체로 본 논문에서 제안된 알고리즘에 의해 처리된 음성을 '부드럽다', '피크를 형성하는 잔류잡음이 많이 사라졌다'고 느꼈으며, 반면에 기존의 방법보다 배경잡음은 조금 더 많이 있다고 느꼈다.

표 1. 실험 사양

환경	사양
Sampling	16kHz, 16Bit
Window	512 pt. Hamming
Frame overlap	1/2
SNR(seg)	-5, 0, 5, 10 dB

표 2 제안된 잔류신호 상에서의 주파수 차감법과 보통의 차감법 사이의 청취테스트 결과 (1: 개선됨, 0: 비슷함, -1: 나빠짐)

SegSNR	비교 알고리즘	평가 점수
10dB	LPC-SS : SS	0.27
	LPC-EMSR : EMSR	0.40
5dB	LPC-SS : SS	0.13
	LPC-EMSR : EMSR	0.27
0dB	LPC-SS : SS	0.13
	LPC-EMSR : EMSR	0.07
-5dB	LPC-SS : SS	-0.20
	LPC-EMSR : EMSR	0.13

청취 테스트의 값을 평균한 결과, 제안된 알고리즘에 의해 구한 음성을 대체로 선호하는 것을 알 수 있었다. 제안된 방법은 SNR이 높은 경우에 더 큰 효과를 가짐을 알 수 있었고, 특히 흔히 Musical Noise라 불리는 잔류잡음 제거에 큰 효과를 기대할 수 있었다.

Ⅴ. 결론

본 논문에서는 잡음이 더해진 입력 음성의 LPC 백색 필터를 통한 잔류신호 스펙트럼 상에서 잡음의 스펙트럼을 주파수 차감하는 방법을 제시하였다. 잔류신호상에서 기본적인 주파수 차감법과 EMSR을 수행해서 청취 테스트를 한 결과 입력음성을 그냥 차감하는 것 보다 더 좋은 결과가 나옴을 알 수 있었다. 특히, 제시한 알고리즘은 Musical Noise라 부르는 잔류잡음 제거에 큰 효과가 있었다.

< 참고문헌 >

[1] J. S. Lim and V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586-1604, Dec. 1979.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 113-120, Apr. 1979.

[3] G. Whipple, "Low residual noise speech enhancement utilizing time-frequency filtering," *proc. ICASSP*, pp. I-5-I-8, 1992.

[4] T. Haulick, K. Linhard, and P. Schrogmeier,

1997.

[5] W. M. Kushner, V. Goncharoff, and Chung Wu, "The effect of subtractive - type speech enhancement/noise reduction algorithms on parameter estimation for improved recognition and coding in high noise environments," *proc. ICASSP*, pp. 211-214, 1989.

[6] Y. Ephraim and David Malah, "Speech Enhancement using a minimum mean-square err short time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1109-1121, Dec. 1984.

[7] Y. Ephraim and David Malah, "Speech Enhancement using a minimum mean-square err log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP 33, pp. 443-445, Apl. 1985.

[8] O. Cappe, "Elimination of the musical noise phenomenon with Ephraim and Malah noise supressor," *IEEE trans. Speech and Audio Processing*, vol. 2, no. 2, pp.345-349, Apl, 1994.

[9] B. L. Sim, Yit Chow Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE trans. Speech and Audio Processing*, vol. 6, no. 4, pp.328-337, July 1998.

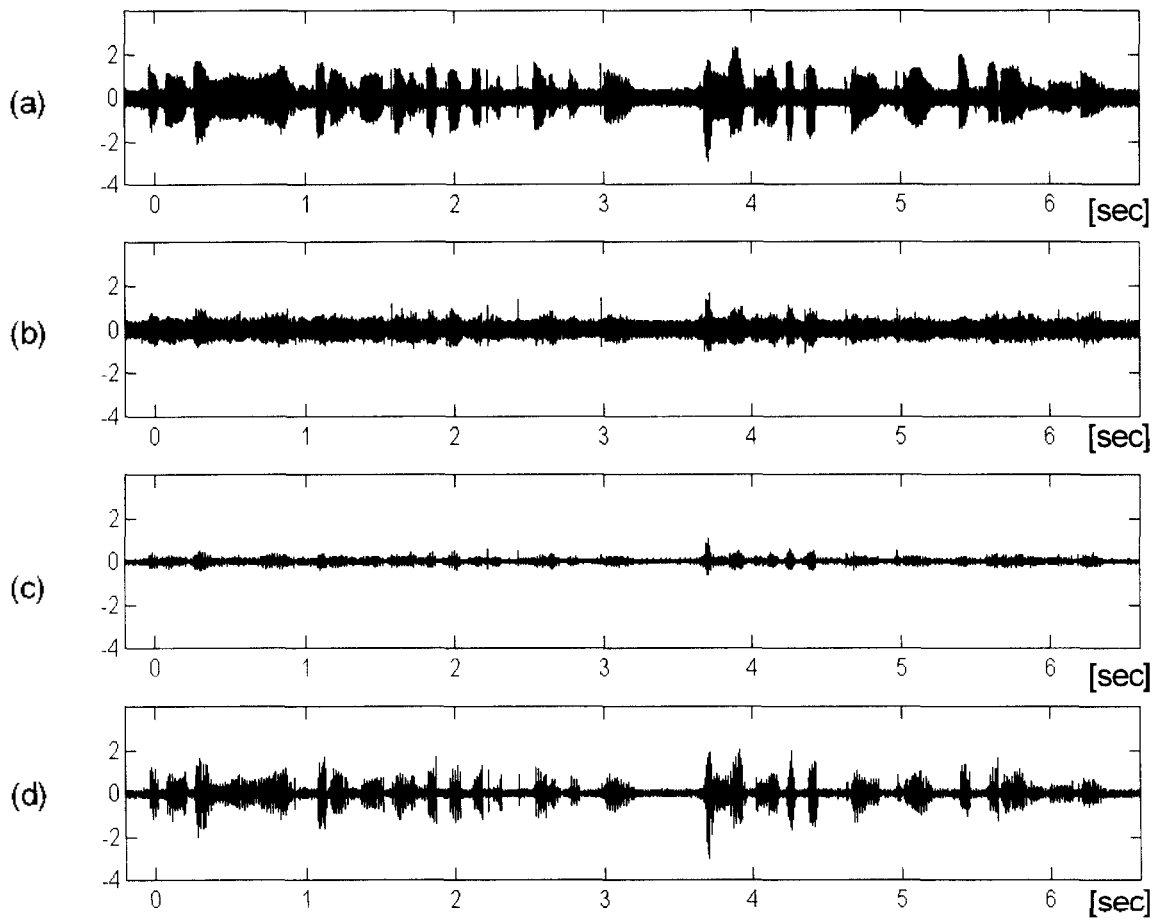

그림 2 제안된 알고리듬에 따른 speech enhancement 과정 (a)잡음이 더해진 원래 음성 (b)백색 필터를 통과한 잡음신호 (c)잡음신호에서 주파수 차감후의 신호 (d) 합성필터를 통과한 최종 신호