

음성코덱에서의 잡음제거 방식 비교

이진걸

배재대학교 컴퓨터전자정보통신공학부

Comparison of Noise Suppression Methods in Voice CODEC

Jin Geol Lee

Division of Computer, Electronic, and Information Engineering, Paichai University

지난 30년간 부가 잡음에 의해 열화된 음성신호의 개선에 관해 많은 연구가 진행되어 왔다. 잡음제거를 위한 고전적인 방법인 spectral subtraction, Wiener filter와 최근에 제안된 심리음향모델에 근거한 perceptual filter, EVRC의 잡음제거단을 성능과 구현의 복잡도 측면에서 비교하였다.

Considerable research in the last three decades has examined the problem of enhancement of speech degraded by additive background noise. We compare traditional methods such as spectral subtraction and Wiener filter, recently proposed psychoacoustic model based methods such as perceptual filter and noise suppression in EVRC in terms of performance and complexity.

Key words : speech enhancement, noise suppression

I. 서 론

최근 디지털 이동통신이 보급됨에 따라 음성 부호화 알고리즘이나 음성인식과 같은 신호처리에 잡음제거는 중요성이 증대하는 추세이다. 이동통신 환경 하에서의 잡음제거는 음성의 품질향상보다는 부가잡음에 의해 열화된 음성의 명료도 향상이 더 요구된다. 이러한 이동통신 환경 하에서 잡음제거의 성능목표를 구현하기 위한 연구가 수년간 수행되어 왔으며 EVRC는 음성부호화기 전처리 단계에 잡음제거단을 도입하였다.

본 논문에서는 고전적인 음성개선방식인 Spectral Subtraction 및 Wiener Filter와 최근에 제안된 심리음향모델을 근간으로 한 Perceptual Filter, EVRC의 전처리단계에 도입된 잡음제거방식을 성능과 구현의 복잡도 측면에서 비교함으로써 이동통신 환경 하에서의 최적의 잡음제거방식을 제안하고자 한다.

본 논문의 구성은 2장에서 Spectral

Subtraction, Wiener Filter, Perceptual Filter, EVRC의 잡음제거단에 대한 이론을 요약하였으며 3장에서는 이러한 잡음제거 방식을 동일한 시험음성 및 부가잡음에 대해 주관적 및 개관적 시험결과에 대해 기술하였다. 4장에서는 결론 및 앞으로의 연구방향에 대해 서술하였다.

II. 잡음제거 방식

1. Spectral Subtraction

음성신호 $s(n)$ 이 상관성이 없는 부가잡음 $d(n)$ 에 의해 열화된 신호 $y(n)$ 과 이들의 전력밀도는 다음의 식으로 표현된다.[1, 2]

$$y(n) = s(n) + d(n) \quad (1)$$

$$p_y(\omega) = p_s(\omega) + p_d(\omega) \quad (2)$$

부가잡음에 대한 사전정보가 주어져 있는 경우 STSA (Short Time Spectral Analysis)에서 위상정보가 인간의 지각에 크게 기여하지 못하는 점을 이용하여 크기 성분만 다음의 식과 같이 필터링하여 개선된 음성 $\hat{s}_w(n)$ 을 구할 수 있다.

$$\hat{S}_w(\omega) = Y_w(\omega)H(\omega) \quad (3)$$

$$H(\omega) = \left(\frac{p_s(\omega)}{p_s(\omega) + p_d(\omega)} \right)^{1/2} \quad (4)$$

위 식에서 아래 첨자 w 는 윈도우를 취했음을 나타낸다.

2. Wiener Filter

부가잡음에 의해 열화된 신호 $y_w(n)$ 는 필터링에 의해 개선된 음성신호 $\hat{s}_w(n)$ 로 변환된다.[1]

$$\hat{s}_w(n) = x_w(n) * h(n)$$

원래 음성신호와 개선된 음성신호의 차가 LMS 기준에서 최소화되도록 하여 필터의 전달함수를 구한다.

$$\text{Minimize } E[|s_w(n) - \hat{s}_w(n)|^2] \quad (5)$$

$$H(\omega) = \frac{p_s(\omega)}{p_s(\omega) + p_d(\omega)} \quad (6)$$

3. Perceptual filter

소리의 감성적 요인인 시간 및 주파수 영역의 은폐효과를 이용한 심리음향모델에 근거하여 음성신호를 개선한다.[3, 5, 6] 부가잡음에 의해 열화된 음성신호 $y(n)$ 은 다음 식과 같이 STFFT (Short Time FFT)에 의해 주파수 영역의 신호로 변환된다.

$$Y_w(k, i) = \sum_{n=0}^{K-1} y(n + \text{off}_i) w(n) e^{-j2\pi kn/K} \quad (7)$$

$w(n)$ 은 윈도우 함수이며 아래 식에 의해 전력밀도로 변환된다.

$$Y_p(k, i) = |Y_w(k, i)|^2 \quad (8)$$

전력밀도를 이용하여 임계주파수대역(Critical band)별로 에너지를 구한다.

$$Y_a(b, i) = \sum_{k=k_{lb}}^{k=k_{hb}} Y_p(b, i), \quad 0 \leq b \leq B-1 \quad (9)$$

b 는 주파수대역 지수이며 k_{lb} , k_{hb} 는 각각 주파수대역 b 의 하한 및 상한이다. B 는 주파수대역의 개수이다. 시간영역의 은폐는 다음의 식과 같이 모델링 된다.

$$Y_i(b, i) = (1-c) Y_i(b, i-1) + c Y_a(b, i) \quad (10)$$

c 는 시간영역에서의 은폐효과 크기와 관련된 상수이다. 주파수영역의 은폐효과를 고려한 필터의 전달함수는 다음의 식과 같다.

$$H(b, i) = 1 - \frac{D_f(b, i)}{Y_i(b, i)} \quad (11)$$

$$\hat{S}_p(k, i) = H(b, i) Y_p(k, i), \quad k_{lb} \leq k \leq k_{hb} \\ 0 \leq b \leq B-1 \quad (12)$$

$D_f(b, i)$ 는 부가잡음에 대해 시간 및 주파수 영역의 은폐효과를 모델링한 신호이다. 식(11)의 전달함수는 심리음향모델이 도입되었을 뿐 형태는 Spectral Subtraction의 전달함수와 동일함을 알 수 있다.

4. EVRC에서의 잡음제거

음성부호화 규격인 EVRC에서는 잡음제거가 전처리단계에서 수행되며 Perceptual Filter와 같이 심리음향모델을 근간으로 하여 구현되었다. 임계주파수 대역(ch)에 대한 시간영역의 은폐효과를 고려한 입력 음성신호의 주파수영역에서의 표현은 아래 식과 같다.[4]

$$E_{ch}(m, i) = \max\{E_{\min}, \alpha_{ch}(m)\}$$

$$E_{ch}(m-1, i) + (1 - \alpha_{ch}(m)) \frac{1}{f_H(i) - f_L(i) + 1} \sum_{k=f_L(i)}^{f_H(i)} |G(k)|^2 \quad (13)$$

$G(k)$ 는 입력 음성신호에 대한 DFT이며, $f_L(i)$, $f_H(i)$ 는 각각 주파수대역(ch)의 하한 및 상한이다. $\alpha_{ch}(m)$ 은 시간대역의 은폐효과 크기에 관련된 상수이며 총 16개의 대역(ch)에 대해 계산하였다. 필터의 전달함수는 부가잡음과 입력 음성신호의 주파수영역에서의 비, $D_i(b, i) / Y_i(b, i)$ 와 관련되어 다른 방법과 동일하게 부가잡음의 크기가 큰 영역에서는 전달함수의 크기가 작아진다. EVRC에서는 시간영역에서의 은폐효과만을 고려하였으나 Perceptual Filter에서는 시간 및 주파수영역에서의 은폐효과를 동시에 고려하였음을 알 수 있다.

III. 비교시험 결과

동일한 구현조건에서의 비교시험을 위하여 Spectral Subtraction, Wiener Filter, Perceptual Filter는 길이가 256인 Hann 윈도우를 사용하고 128 샘플씩 이동하여(이웃하는 분석프레임과 128 샘플 겹침) Overlap-add에 의한 주파수 영역분석 및 시간영역 합성을 하였다. FFT의 크기는 윈도우 크기와 동일한 256, 필터의 임펄스 반응의 길이를 고려한 512를 적용해본 결과 합성음의 차이가 거의 없음을 확인하고 FFT의 크기를 256으로 하였다. EVRC의 잡음제거단은 길이가 104인 사다리꼴형의 윈도우를 사용하며 이웃하는 분석프레임과 24 샘플 겹친다. FFT의 크기는 128이다.

Spectral Subtraction, Wiener Filter, Perceptual Filter의 구현에서 부가잡음의 전력밀도는 음성이 없는 구간을 다음의 식으로 정의된 에너지의 비를 이용하여 찾아내어 평균을 구하였다.

$$T = \int_{-\pi}^{\pi} Y(\omega) d\omega / \int_{-\pi}^{\pi} \mu(\omega) d\omega \quad (14)$$

$\mu(\omega)$ 는 부가잡음의 평균 전력밀도이고 아래 식에 의해 계산된다.

$$\mu(\omega) = .9\mu(\omega) + .1Y(\omega), \quad T < 2.3 \quad (15)$$

평균부가잡음의 전력밀도가 입력음성신호의 전력밀도보다 큰 경우에는 아래 식에 의해 처리함으로써 합성음의 전력밀도를 0으로 하는 방법(반파정류)보다 음질이 개선됨을 확인하였다.

$$\hat{S}(\omega) = .3Y(\omega), \quad T < 2.3 \quad (16)$$

Spectral Subtraction, Wiener Filter, Perceptual Filter, EVRC의 잡음제거단의 성능을 비교시험하기 위하여 다음의 3개의 문장에 대해 시험하였다.

- 문장 1. 하나 둘 셋 넷 다섯 여섯 일곱 여덟 아홉 열(남자, 부가잡음: 지하철 소음, SNR: 9.41 dB)
- 문장 2. 이번 겨울은 예년과 달리 포근합니다. (남자, 부가잡음: 지하철 소음, SNR: 12.23 dB)
- 문장 3. 미는 피부 한 겹질 차이입니다. (여자, 부가잡음: 버스 소음, SNR: 13.84 dB)

이러한 시험문장에 대해 잡음제거 처리 후 SNR 및 SEGSNR 측정결과를 표 1에 나타내었다. SEGSNR 측정 시 측정샘플길이는 필터의 분석 및 합성길이가 동일하게 256으로 하였다. 주관적 시험방법으로 시험인원을 10 명으로 하여 비공식적 MOS 시험결과를 표 2에 나타내었다. (최고음질: 5, 최저음질: 1) 객관적 시험인 SNR/SEGSNR 측정에서 EVRC의 잡음제거단의 수치가 다른 방법에 비해 현저히 낮은 것은 잡음제거 신호처리에 따른 입출력 신호간의 위상차에 기인하는 것으로 추정된다. 구현의 복잡도 정도는 EVRC의 잡음제거단 이나 Perceptual Filter가 Spectral Subtraction이나 Wiener Filter보다 훨씬 복잡하나 주관적 음질시험 결과에서와 같이 Spectral Subtraction이나 Wiener Filter의 성능이 다소 우수함을 알 수 있다.

표 1. SNR/SEGSNR 측정결과

	문장 1	문장 2	문장 3
EVRC	-2.82/-1.15	-1.40/-0.73	-5.10/-2.42
Perceptual Filter	11.25/10.59	14.49/12.56	15.93/11.06
Spectral Subtraction	9.21/5.37	12.27/8.30	14.23/8.13
Wiener Filter	8.87/5.07	11.71/7.76	14.07/7.96

표 2. 주관적 음질시험 결과

	문장 1	문장 2	문장 3	평균
EVRC	3.2	3.0	3.2	3.1
Perceptual Filter	3.2	3.8	1.9	3.0
Spectral Subtraction	3.1	3.5	2.7	3.1
Wiener Filter	3.0	4.0	2.9	3.3

IV. 결 론

잡음제거 방식인 Spectral Subtraction, Wiener Filter, Perceptual Filter, EVRC에서의 잡음제거 단의 성능을 동일한 문장과 부가잡음에 대해 시험한 결과 고전적인 방식인 Spectral Subtraction이나 Wiener Filter가 최근에 제안된 심리음향모델을 기반으로 한 Perceptual Filter나 EVRC의 잡음제거단보다 구현의 복잡도에서 현저히 간단하며 성능면에서도 우수함을 확인하였다. Perceptual Filter가 EVRC에 비해 주파수영역의 은폐효과를 고려했음에도 불구하고 성능이 떨어지는 것은 EVRC에 비해 음성구간과 주위잡음구간을 구분하는 방법이 간단히 구현(에너지 비)되었기 때문인 것으로 추정되며 앞으로 이 부분의 연구가 진행될 예정이다. 또한 Perceptual Filter에 은폐 문턱치의 개념을 도입한 잡음제거방식의

연구도 계획하고 있으며 이동통신 환경 하에서 발생할 수 있는 모든 소음에 대해 다양한 시험이 병행될 예정이다.

감사의 글

이 연구의 일부는 배재대학교 1998년도 교내 학술연구비의 지원에 의하여 이루어진 것으로 이에 감사드립니다

참 고 문 헌

1. Jae S. Lim, Alan V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech", Proceedings of the IEEE, Vol. 67, No. 12, pp. 1586-1604, Dec. 1979
2. Steven F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 2, pp. 113-120, Apr. 1979
3. Dionysis E. Tsoukalas, John Mourjopoulos, and George Kokkinakis, "Perceptual Filters for Audio Signal Enhancement", J. Audio Eng. Soc., Vol. 45, No. 1/2, pp. 22-36, Jan./Feb. 1997
4. TIA/EIA/IS-127, Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems, Sep. 1996
5. James D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria", IEEE Journal on Selected Areas in Communications, Vol. 6, No. 2, pp. 314-323, Feb. 1988
6. 최민주, "소리의 감성적인 측면과 활용", 전자공학회지, 제24권 제11호, pp. 33-49, 1997