

차량내 잡음 환경에서 적응적 경계값을 이용한 가중치 주파수 차감에 관한 연구

A Study on Weighted Spectral Subtraction Using Adaptive Threshold in Car Noise Environment

전 선 도*, 강 철 호*, 김 중 찬**, 김 순 협**

(Sun Do June*, Chul Ho Kang*, Jong Chan Kim**, Soon Hyob Kim**)

요 약

본 연구는 자동차내의 배경 잡음에 의해 손상된 음성에서 배경 잡음을 주파수 차감에 의하여 제거시킨다. 특히 음성 정보의 손실이 적은 잡음 추정 방법으로 가중치를 이용하여 잡음을 추정하는 가중치 주파수 차감법을 이용하였다. 이러한 가중치 주파수 차감법은 잡음의 변화가 완만한 경우에 적당하다, 그러나 실제적인 상황에서 배경잡음신호의 변화가 큰 경우가 존재한다.

이러한 이유에서 본 연구는 잡음 추정시 잡음 추정값을 이용하여 추정 잡음 경계값을 적응적으로 변화시키는 차감법을 제안한다. 이 방법은 추정된 잡음 신호의 변화율을 이용하여 경계값을 상황에 따라 적응적으로 변화시키는 방법이다. 모의 실험에 의하여 고정적인 경계값을 갖는 가중치 주파수 차감법에 비해 제안한 적응적 경계값을 갖는 가중치 주파수 차감법의 출력 SNR이 증가함을 확인하였고, 음성 인식 시스템에 적용한 인식 실험에서도 성능이 향상됨을 확인하였다.

ABSTRACT

In this study, speech corrupted by car noise is enhanced using weighted spectral subtraction. Especially, it is proposed that weighted spectral subtraction is modified by adaptive threshold parameter at noise estimation. This threshold parameter is adapted to the car noise variation.

The criterion is SNR(Signal-to-Noise Ratio) and speech recognition rate. Experimental results show that the performance of proposed spectral subtraction with adaptive threshold parameter is better than spectral subtraction with fixed threshold parameter.

I. 서 론

현대의 정보화 사회에서 음성 서비스의 역할은 점점 확대되어 나간다. 특히 인간의 가장 자연스러운 도구인 음성을 이용하는 것이므로 더욱 개발 필요성이 요구되는 것이다. 이러한 음성 서비스는 구체적으로 음성 통신, 음성 자동 응답 장치, 음성 인식, 음성 합성 등 다양하다. 그러나 이러한 음성 서비스는 실질적인 환경에서 사용시 배경 잡음의 존재로 그 활용이 제한되어 있다. 특히 차량 내에서 음성을 이용한 음성 서비스의 운용시에 엔진 잡음, 차 바퀴와 지면과의 마찰에 의한 잡음 등에 의한 극심한 배경 잡음이 존재한다. 그리고 이러한 잡음은 액셀을 밟아서 배경 잡음의 에너지가 증가하는 경우 다양한 변화가 존재한다.

본 연구에서 사용하는 주파수 차감법은 가중치 주파수

차감법을 이용한다. 이 방법은 다른 주파수 차감법 즉, 프레임 평균 방법 및 히스토그램 방법에 비하여 SNR-(Signal-to-Noise Ratio)은 낮은 반면 음성 정보의 손실이 적어 음성 인식에 적용시 다른 방법에 비해 인식률이 높다[1-2]. 특히, 본 연구는 이러한 가중치 주파수 차감법의 잡음 추정 방법에서 경계값 파라미터를 추정하여 잡음의 변화 정도에 따라 경계값을 가변시키는 방법을 제안한다. 모의 실험에 의하여 잡음이 극심한 차내 환경의 배경 잡음 신호를 이용하여 고정적인 경계값을 갖는 가중치 주파수 차감법과 제안한 주파수 차감법의 출력 SNR과 음성 인식 시스템에 적용한 인식율으로써 그 성능을 비교한다.

II. 가중치 주파수 차감법

2.1 주파수 차감법

잡음 섞인 음성 신호의 모델은 다음과 같다.

$$s(t) + n(t) = x(t) \quad (1)$$

* 광운대학교 전자통신공학과

** 한국생산기술연구원

*** 광운대학교 컴퓨터공학과

접수일자: 1998년 9월 12일

여기서 $s(t)$ 는 순수 음성 신호의 모델이고, $n(t)$ 는 순수 배경 잡음을 의미한다. 결국 입력되는 음성 신호는 잡음과 혼합된 $x(t)$ 신호이다.

본 시스템은 전력 스펙트럼에 의하여 잡음을 추정 한다. 다음식은 전력 스펙트럼 변환식이다[3].

$$X_i(k) = \left| \sum_{t=0}^N x_i(t) e^{-j2\pi kt} \right|^2 \quad (2)$$

$$N_i(k) = \left| \sum_{t=0}^N n_i(t) e^{-j2\pi kt} \right|^2 \quad (3)$$

윗 식은 DFT(Discrete Fourier Transform)으로서 실질적인 구현에서는 FFT(Fast Fourier Transform) 으로 구현했으며, N 은 256이다.

잡음 추정 알고리즘에 의하여 잡음 전력 스펙트럼 $\overline{\mu}_i(k)$ 을 추정하며 이렇게 추정한 잡음을 입력 신호의 스펙트럼에서 차감하여 음성 신호를 추출할 수 있다. 다음 식은 차감식을 보여주고 있다.

$$\overline{S}(k) = X(k) - \overline{\mu}_i(k) \quad (4)$$

이렇게 추정한 잡음으로 음성 신호를 추출할 수 있으나 차감한 추정 잡음에 의해 오히려 음성 신호에 반전된 잡음이 더해진 경우가 존재한다. 이러한 이유에서 반파 정류기를 사용하고, 문턱값을 주어 이를 기준으로 신호를 압축시킨다.

가) 반파 정류기

$X(k)$ 에서 $\overline{\mu}_i(k)$ 를 차감했을 때 0 이하가 되면 음성 보다 잡음 $\overline{\mu}_i(k)$ 이 더 큰 경우이므로 반파 정류기를 사용한다.

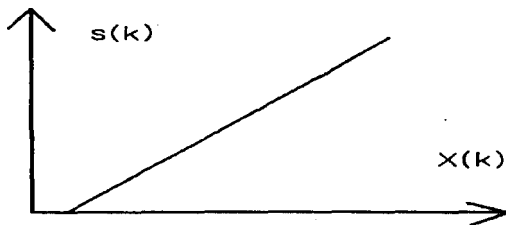


그림 2-1. 반파 정류기
Fig. 2-1. Half-Rectifier.

나) 문턱값

다음은 더해진 잡음신호를 문턱값을 기준으로 감쇄시키는 방법으로 그 기준에 도달하지 못했을 경우에 감쇄시키는 방법이다. 다음 식은 문턱값이다.

$$T = 20 \log_{10} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\overline{S}(e^{j\omega})}{\overline{\mu}(e^{j\omega})} d\omega \right] \quad (5)$$

이 식은 추정된 음성 전력 스펙트럼 $\overline{S}(e^{j\omega})$ 와 추정한 잡음 전력 스펙트럼 $\overline{\mu}(e^{j\omega})$ 에 의해 문턱값을 결정한다. 즉, 경험적으로 T 가 -12dB 보다 클 때 음성 구간으로 보며, T 가 -12dB 보다 작을 때는 다음과 같이 감쇄시킨다[2].

$$\overline{S}_i(e^{j\omega}) = \begin{cases} \overline{S}_i(e^{j\omega}), & T \geq -12dB \\ cX(e^{j\omega}), & T < -12dB \end{cases} \quad (6)$$

단 $20 \log_{10} c = 30 \text{ dB}$

2.2 가중치를 이용한 잡음 추정 방법

이 방법은 잡음 전력 스펙트럼과 입력 전력 스펙트럼과의 관계를 적용하여 적응적으로 추정해 나간다. 다음 식은 이러한 Recursive 한 방법에 의한 전 프레임의 추정 잡음 스펙트럼에 의하여 현재 프레임의 잡음을 추정하는 알고리즘식이다.

$$\overline{\mu}_i(k) = \beta((1-a)S_i(k) + a\overline{\mu}_{i-1}(k)) \quad (7)$$

이 식은 i 프레임의 입력의 스펙트럼 $X_i(k)$ 와 $i-1$ 프레임의 잡음을 추정한 스펙트럼 $\overline{\mu}_{i-1}(k)$ 에 의하여 i 프레임의 잡음 추정값 $\overline{\mu}_i(k)$ 을 결정한다.

이 식에서 a 는 실험에 의해 0.9를 취하였고 경계 파라미터 β 는 1.0를 취하였는데, 이것은 여러 실험에 의하여 얻은 최적 수치값이다.

III. 제안한 적응적인 경계치 알고리즘

다음 그림은 고속도로에서 80km/h 정도로 달리는 차량 내에서의 배경 잡음 신호이다.

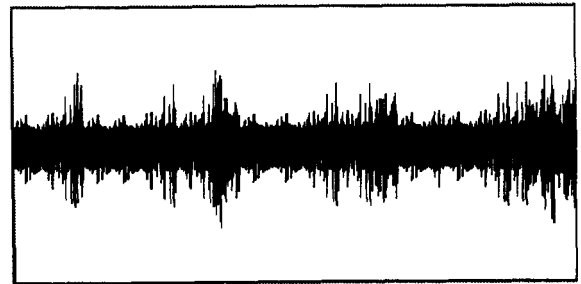


그림 3-1. 차량내 배경 잡음
Fig. 3-1. Car Noise Environment.

위와 같은 배경 잡음은 시간에 따른 변화가 많음을 확인할 수 있다. 본 연구에서 제안하는 것은 이러한 배경 잡음의 변화를 잡음 추정 알고리즘에 적용하여 경계 파라미터 β 를 잡음 추정에 적용시켜 나간다. 다음식은 전 프레임에서의 추정 잡음치의 변화 정도를 보여주는 식이다.

$$\gamma_i(k) = \frac{\overline{\mu_{i-2}}(k) - \overline{\mu_{i-1}}(k)}{\overline{\mu_{i-1}}(k)} \quad (8)$$

식 (8)에 의하여 식 (7)에서 β 를 추정 배경 잡음의 변화에 따라 적응시킨다.

$$\beta_i(k) = 1 - \left(\frac{0.14}{1 + \exp(-\alpha \gamma_i(k))} - 0.03 \right) \quad (9)$$

단, $-0.11 \leq \gamma_i(k) \leq 0.03$ 이다. 또 이외의 범위에서는

$$\begin{aligned} -0.11 > \gamma_i(k), \quad \beta_i(k) &= 1.11 \\ 0.03 < \gamma_i(k), \quad \beta_i(k) &= 0.97 \end{aligned} \quad (10)$$

그리고 α 는 다음과 같이 결정하였다.

$$\alpha = \frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{\gamma_i(k)} \quad (11)$$

이러한 $\beta_i(k)$ 의 변화는 다음 그림과 같다.

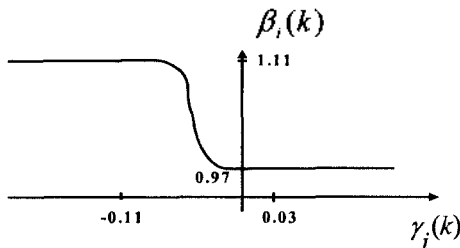


그림 3-2. 적응적 경계값
Fig.3-2. Adaptive Threshold.

결국 이렇게 변화하는 $\beta_i(k)$ 는 추정 잡음의 변화 정도에 의하여 변함이 없는 β 의 최적값인 1을 기준으로 추정 잡음치의 크기를 결정한다. 다음식은 이에 따른 잡음 추정식이다.

$$\overline{\mu_i}(k) = \beta_i(k) \left((1-a)S_i(k) + a\overline{\mu_{i-1}}(k) \right) \quad (12)$$

다음그림은 적응적 경계값을 갖는 가중치 주파수 차의 전체 블록도이다.

IV. 모의 실험

고속도로에서 80km/h 속도로 주행하는 차량내에서 배경 잡음 신호를 녹음하여 잡음 신호를 얻었다. 그리고 실험실내에서 차내에서 사용하는 22개의 명령 단어를 선택

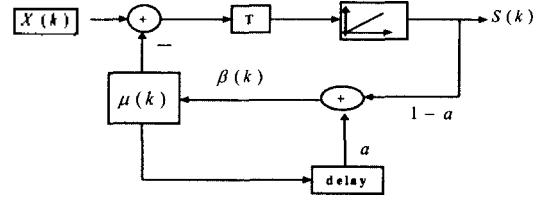


그림 3-3. 적응적 경계값을 갖는 주파수 차감 블록도
Fig. 3-3. Spectral Subtraction using Adaptive Threshold.

하여 잡음 신호와 랜덤하게 혼합하여 22개의 SNR 입력 신호를 만들었다. 그리고 식(7)에의해 고정적인 값을 갖는 β 를 최적값인 1로 한 가중치 주파수 차감법과 제안한 적응적인 $\beta(k)$ 를 갖는 적응적 경계값에 의한 가중치 주파수 차감법에 의한 잡음 제거한 SNR을 표 4-1과 표 4-2에 표기하였다. 이 때 표 4-1은 SNR이 -50 dB에서 -25 dB 사이의 값을 갖는 혼합 신호의 차감 실험이고, 표 4-2는 SNR이 -15 dB에서 10 dB 사이에서의 혼합 신호를 사용한 경우이다.

그리고 차량내 배경 잡음과 22개의 명령 단어를 5번씩 하여 전체 110개의 단어로써 -50 dB에서 -15dB 사이

표 4-1. SNR (-50 dB ~ -25 dB)
Table 4-1. SNR (-50 dB ~ -25 dB).

차감전 SNR <dB>	차감후SNR <dB> 고정적인 β	차감후SNR <dB> 적응적인 $\beta(k)$
-40.354931	-31.486883	-29.565319
-46.911957	-37.648109	-35.519875
-32.378555	-23.453810	-20.967293
-40.360458	-31.515253	-28.662167
-38.887115	-29.696491	-27.093920
-40.036423	-32.555668	-31.109026
-40.387257	-31.339870	-29.252016
-35.655712	-23.732441	-19.755905
-49.967636	-40.886612	-38.565483
-26.557444	-17.097679	-14.889548
-26.318062	-17.187555	-14.509309
-35.748638	-27.097839	-25.197517
-29.096539	-19.539892	-17.276833
-35.148834	-26.705042	-24.266422
-34.082134	-24.550970	-22.341120
-25.371143	-16.620245	-14.928670
-35.383678	-26.927732	-24.593948
-32.700947	-24.513805	-22.065628
-33.803165	-23.968710	-21.569664
-35.505993	-27.750727	-25.520632
-32.579247	-24.025984	-22.159899
-44.523464	-32.442192	-30.695887

표 4-2. SNR (-15 dB~10 dB)
Table 4-2. SNR (-15 dB~10 dB).

차감전SNR<dB>	차감후SNR<dB> 고정적인 β	차감후SNR<dB> 적응적인 $\beta(k)$
-4.519758	2.055017	4.321944
-11.076790	-2.499295	-0.694732
3.456586	6.879951	7.932575
-4.525304	2.357409	4.194456
-3.051928	2.800077	3.629798
-4.201285	2.093297	2.676441
-4.552104	1.586089	2.055827
0.179433	6.638091	8.586070
-14.132468	-5.460260	-3.708472
9.277699	10.280668	11.403491
9.517139	10.586631	11.576250
0.086531	5.437360	5.666948
6.738589	8.610320	9.077520
0.686341	5.453649	6.327345
1.753039	6.153047	6.412012
10.464036	10.637793	9.476174
0.451454	5.188625	6.240670
3.134222	6.888262	7.157604
2.031995	7.272416	8.321311
0.329209	4.697155	4.135390
3.255919	5.675492	5.139420
-8.688248	-1.291459	0.043283

의 SNR을 갖도록 한 후 인식 실험을 한 것을 표 4-3에, -25 dB에서 -5 dB 사이의 SNR의 혼합 신호인 경우를 표 4-4에, 10 dB와 5 dB 사이에서의 경우는 표 4-5에 잡음이 혼합된 상태의 인식율, 가중치 주파수 차감법(β)에 의한 인식율 및 적응적인 경계값을 갖는 가중치 주파수 차감법 $\beta(k)$ 에 의한 인식율을 각각 비교하였다.

이 인식 실험은 한 화자로서 화자 종속 실험을 하였고, One-Stage DP(Dynamic Programing) 알고리즘을 사용하여 인식 실험을 하였다[4]. 음성 특징 파라미터로는 잡음에 강한 PLP(Perceptual Linear Prediction)를 사용하였다[5][6].

표 4-3. 인식율 (-50dB~-15dB)
Table 4-3. recognition rate (-50dB~-15dB).

	혼합신호	차감신호 β	차감신호 $\beta(k)$
인식된단어 /총단어	55/110	59/110	63/110
인식률 (%)	50	53.6	57.2

표 4-4. 인식율 (-25dB~-5dB)
Table 4-4. recognition rate (-25dB~-5dB).

	혼합신호	차감신호 β	차감신호 $\beta(k)$
인식된단어 /총단어	86/110	85/110	91/110
인식률 (%)	78	77.2	82.7

표 4-5. 인식율 (-10 dB~5 dB)
Table 4-5. recognition rate (-10 dB~5 dB).

	혼합신호	차감신호 β	차감신호 $\beta(k)$
인식된단어 /총단어	99/110	98/110	99/110
인식률 (%)	90	89	90

V. 고찰 및 결론

표 4-1의 잡음이 극심한 경우에 적응적인 경계값을 적용하였을 때가 적용하지 않은 가중치 주파수 차감보다 2~3dB 정도의 SNR 증가를 보여 주고 있다. 또 표 4-2의 0 dB 이하의 경우에도 2 dB 이상의 적응적인 경계값의 주파수 차감이 더 우수한 성능을 보여준다. 이것은 적응적인 경계값인 $\beta(k)$ 가 잡음의 크기 변화에 적응하므로써 가중치 주파수 차감법의 잡음 추정을 더 정확하게 하는 역할을 함을 알 수 있다. 또한 표 4-3은 SNR이 50dB에서 -15dB 정도로 배경 잡음이 극심한 상태에서 인식율이 상당히 떨어지는 경우로써 적응적인 경계값의 가중치 주파수 차감법이 차감을 하지 않은 경우나, 고정된 경계값을 갖는 차감법보다 인식율이 향상됨을 알 수 있다. 또한 표 4-4 와 표 4-5에서 SNR 이 높을수록 고정적인 경계값의 주파수 차감은 오히려 차감을 하지 않을 때 보다 SNR이 떨어지는 경우가 발생한다. 이것은 차감시 음성 신호를 차감함으로써 음성 정보를 잃는 경우가 된다. 그러나 제한한 배경 잡음의 변화에 적응하는 적응적인 경계값의 주파수 차감은 음성 정보의 손실을 줄여주기 때문에 고정된 β 에 비해 인식율이 향상된다.

이상의 결과에 의해 실제 차량내의 음성 처리 시스템은 음성 신호와 배경 잡음의 신호가 여러 상황에 따라 다양하게 변화하는 경우 본 논문이 제안하는 적응적 경계값을 갖는 가중치 주파수 차감법은 이러한 시스템에 효율적으로 사용될 것으로 기대된다.

참고 문헌

1. H. G. Hirsch, C. Ehrlicher, "Noise Estimation for Speech Recognition," IEEE, Vol. ICASSP-95, pp.153-156, 1995.
2. Steven F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE, Vol. ASSP-27, No.2, pp.113-120, 1979.
3. J. C. Junqua, J. P. Haton, "Robustness in Automatic Speech Recognition," Kluwer Academic Publishers, 1996.
4. L. Rabiner, B. H. Juang, "Fundamental of Speech Recognition," Prentice Hall, 1993.
5. H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis for Speech," J. Acoust. Soc Am., pp.1738-1752, 1990.
6. Reiner Plomp, "Auditory Handicap of Hearing Impairment and the Limited Benefit of Hearing Aids," Acoustical Society of America, Vol. 63, No.2, pp. 533-549, 1978.
7. Jae S. Lim, "Evaluation of a Correlation Subtraction Method for Enhancing Degraded by Additive White Noise," IEEE,

Vol. ASSP-26, No.5, pp.471-472, October 1978.

8. Robert J. McAulay, "Speech Enhancement Using Soft-Decision Noise Suppression Filter," IEEE, Vol. ASSP-28, No.2, pp.137-145, April, 1980.

▲전 선 도 (June Sun Do) 1967년 5월 9일생



1993년 2월 : 광운대학교 전자통신공학과 공학사

1995년 2월 : 광운대학교 대학원 전자통신공학과 공학석사

1995년 3월~1998년 2월 : 광운대학교 대학원 전자통신공학과 박사수료

※주관심분야: 음성 인식, 잡음 제거, 통신 시스템 설계, 적응 필터, DSP 코딩

▲김 철 호 (Chul-Ho Kang) : 1952년 2월 5일생



1975년 2월 : 한양대학교 전자공학과 공학사

1979년 2월 : 서울대학교 대학원 전자공학과 공학석사

1977년~1981년 : 국방과학연구소(연구원)

1988년 2월 : 서울대학교 대학원 전자공학과 공학박사

1991년 1월~1992년 1월 : 미국 일리노이 대학 객원교수

1983년~현재 : 광운대학교 전자통신공학과 교수

※주관심분야: 음성 및 잡음 신호처리, 통신 신호처리, 적응 신호처리, DSP, 신경망, FPGA 설계

▲김 종 찬 (Jong Chan Kim) 1960년 2월 14일생



1983년 : 연세대학교 의용전자공학과 졸업

1989년 : 연세대학교 산업대학교 전자공학과 졸업

1988년~현재 : 한국생산기술연구원 선임연구원

▲김 순 협 (Soon Hyob Kim)

한국음향학회지 제14권 5호 참조