

새로운 음성/비음성 분류함수에 기반한 스펙트럼 차감법에 의한 차량잡음제거

正會員 朴 英 植* 正會員 李 駿 宰* 正會員 李 應 柱* 正會員 河 永 浩*

Car Noise Cancellation by Using Spectral Subtraction Method Based on a New Speech/nonspeech Classification Function

Young Sik Park*, Jun Jae Lee*, Eung Joo Lee*,
Yeong Ho Ha* *Regular Members*

요 약

본 논문에서는 차량 잡음 환경하에서 하나의 마이크로폰 구조를 사용하여 스펙트럼 차감법을 이용한 잡음 제거 방법을 제안하였다. 변화하는 잡음에 의하여 손상된 음성신호에서 잡음의 성분을 제거하기 위하여 여러 상황에서의 차량 잡음을 분석하고 특성을 알아보았다. 음성/비음성의 분류와 잡음의 스펙트럼을 추정하기 위하여 잡음 분석을 바탕으로 음성/비음성 분류함수를 제안하였다. 이 분류함수에 의하여 적은 계산량으로 간단하게 정확한 음성/비음성의 분류가 가능하다. 또한 정확한 잡음의 스펙트럼 추정이 가능하다. 제안된 음성/비음성 분류함수에 의한 잡음의 스펙트럼 추정으로 인하여 왜곡이 거의 없는 깨끗한 음성신호를 추출할 수 있었다.

Abstract

In this paper, a scheme of noise cancellation using spectral subtraction method with single input in an automobile noise environment is proposed. In order to remove the changing automobile noise components from the noisy speech signal, the noise of various states is analyzed and its characteristics are presented. For the decision of speech/nonspeech and the estimation of noise spectrum, a classification function is proposed on the basis of noise analysis. This function presents the precise decision of speech/nonspeech and the optimal estimation of noise spectrum with less computation. As the result of the estimation of noise spectrum by the proposed classification function, the clean speech signal is extracted from the noisy speech signal with high signal-to-ratio.

*慶北大學校 電子工學科
Dept. of Electronics, Kyungpook Nat'l Univ.
論文番號 : 93197
接受日字 : 1993年 10月 18日

I. 서 론

자동차 문화의 급진적인 발전과 이동통신의 발달로 카폰이나 휴대폰의 보급이 날로 확장되고 있는 가운데 차량 잡음에 의해 손상된 음성신호 개선의 중요성은 더욱 커지고 있다. 특히 달리는 자동차내에서 전화기의 성능은 엔진, 바람, 환기팬등의 주위 잡음과 이동 터미날을 네트워크에 연결시켜 주는 라디오 채널의 전달함수가 계속적으로 변화하기 때문에 음성신호의 심한 손상을 가져온다. 또한 핸들을 조정하는 운전자가 주행중에 전화기를 직접 사용하기에는 사고의 위험이 있으므로 hand-free 전화기가 필요하게 되었다. 이로 인하여 차량잡음의 영향이 증가하므로 효율적인 잡음개선 방법이 더욱 요구된다. 유럽등의 선진국에서는 교통사고 방지를 위하여 차량내에서 hand-on 전화기 사용을 규제하고 차량에 장착된 hand-free 전화기 사용을 유도하고 있다. 그러므로 적응적으로 자동차 잡음을 제거하는 전처리 과정이 반드시 필요하다[1].

지금까지 전화상에서 음성신호의 음질을 개선하기 위한 노력으로 음성신호의 부호화와 채널의 특성문제 등에 많은 연구가 되어 왔으나 이들만으로는 음질을 개선하기에 부족하다. 그러므로 전처리 과정으로 적응필터(adaptive filter)와 같은 방법으로 주위 잡음에 의해 손상된 음성신호를 개선하고자 노력해 왔으며 최근에 DSP, 마이크로프로세서의 발달과 실시간 처리를 위한 신호처리 알고리즘의 꾸준한 연구로 말미암아 음성신호를 향상시키는 연구가 현재 활발하게 진행되고 있다. 더 나아가 음성 다이얼링 시스템과 같은 음성 인식, 음성제어 시스템 구현에 있어 잡음제거는 필수적으로 따라야할 과정이므로 이의 중요성은 크다.

주 마이크로폰과 기준 마이크로폰, 두개의 마이크로폰으로 잡음의 공간적 상관성을 이용하는 적응필터 잡음제거 방법은 두개의 마이크로폰에 입력되는 잡음의 상호 상관관이 클때 잡음제거의 성능은 좋아진다. 따라서 두개의 마이크로폰의 거리가 가까우면 상호상관이 커지지만 잡음만 입력되어야 하는 기준 마이크로폰에 음성신호가 포함되어 출력 음성신호의 에너지가 줄어들고 왜곡이 발생하므로 잡음의 개선에는 한계가 있다[1]-[4]. 또한 시간적 상관성을 이용하는 방법은 음성신호도 준주기적(semi-periodic)인 성질이 있으며 차량 잡음신호도 엔진 잡음과 같은

주기적인 성분을 포함하고 있으므로 효과적인 잡음제거 효과를 가지지 못한다[4].

하나의 마이크로폰을 이용하는 스펙트럼 차감법은 잡음신호만 존재하는 비음성 구간에서 잡음의 스펙트럼을 추정하여 잡음이 포함된 음성신호의 스펙트럼에서 차감하여 잡음을 제거하는 방법이다[5]. F-15/F-16 비행기의 환경하에서 상호상관에 의한 피치 정보로써 음성/비음성을 분리하여 추정된 잡음의 스펙트럼을 의사 캡스트럼 영역에서 차감하는 방법이 있다[6]. 이 방법은 캡스트럼 영역으로의 변환에 따른 계산량이 많고 차감인자에 따라 음성신호의 왜곡이 많이 유발되며 입력신호에 비해 출력 신호가 감쇄되어 출력단에서 auto-scaling 과정이 필요한 문제점들이 있다.

본 논문에서는 주기적인 성분이 지배적인 차량잡음의 환경하에서는 상호상관에 의한 음성신호의 피치 정보로써 음성/비음성을 분류하는 방법은 적합하지 않으므로 음성신호와 잡음신호간에 구별되는 두가지 특징을 이용한 음성/비음성 분류함수를 제안하였다. 이를 음성/비음성의 분류와 잡음의 스펙트럼 추정에 적용하여 음성신호의 왜곡이 없고 구조적으로 간단한 새로운 방법의 적응 스펙트럼 차감법을 제시하였다. 제안된 분류함수는 시간영역에서 음성신호의 변화분과 잡음신호의 변화분이 차이가 많다는 것과 주파수영역에서 음성신호와 잡음신호의 스펙트럼 분포가 다른 성질을 이용하였다. 시간영역에서 음성신호는 조밀하게 발진하는 것으로 보아 변화분이 잡음신호의 변화분 보다 대체적으로 큰 성질이 있으며 주파수영역에서 음성신호의 스펙트럼 분포가 잡음신호 보다 높은 주파수 대역에 대체적으로 많이 분포하는 것을 특징값으로 이용하여 연속적으로 변화되는 환경하에서 잡음의 스펙트럼과 잡음에 대한 특징 값들을 다시 개선할 것인지를 매 프레임마다 결정하고 음성신호에 대해서는 추정된 잡음의 스펙트럼을 차감해서 음성신호에 포함된 잡음을 제거하는 방법이다.

본 알고리즘의 성능을 평가하기 위해 먼저 깨끗한 원 음성신호와 자동차의 잡음신호를 섞은 신호에 대해 신호대 잡음비(SNR) 개선, 음성신호의 왜곡등을 평가하고 실제의 자동차 주행중에서 획득한, 잡음에 의하여 손상된 음성신호에 대하여 실험하였다. 제안된 알고리즘의 실험결과 음성신호를 손상하지 않으면서 신호대 잡음비가 상당히 향상되었음을 알 수 있었고 잡음이 제거된 깨끗한 음성신호를 들을 수가 있었다.

II. 차량 잡음 분석

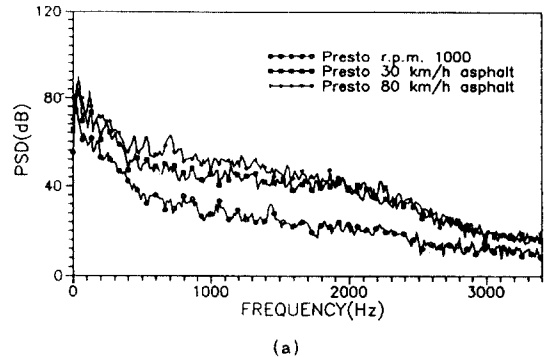
1. 잡음 획득 방법

실험에 사용된 자동차 잡음을 의미있는 실제적인 환경하에서 획득하였다. 차종에 따라 잡음의 정도가 다르므로 잡음의 정도가 큰 프레스토와 비교적 작은 소나타에 대하여 분석하였으며 각각에 대하여 정지한 상태에서는 엔진회전수를 가변하면서 환기 팬을 고/저로 놓고 각각 실험하였다. 자동차의 주행상태에서는 프레스토의 경우 30km/h, 60km/h, 80km/h로, 소나타의 경우 60km/h, 80km/h, 100km/h로 달리면서 실험하였다. 방향성을 가진 마이크로폰을 자동차 계기판에 고정시켜 methal tape에 녹음하여 12kHz로 샘플링 하였다.

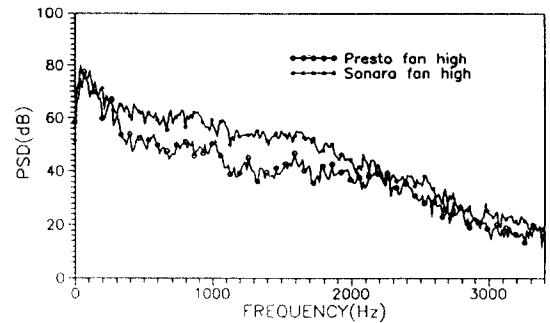
2. 잡음 특성 분석

자동차내의 잡음은 비정적(nonstationary)이며 발생원은 엔진, 공기흐름, 도로와의 마찰, 팬 등 다양하며 자동차내로 들어오는 경로도 창문, 바퀴, 계기판 등 다양하다. 잡음신호는 12kHz로 샘플링하였고 주파수 영역에서의 스펙트럼 분포는 시간영역의 잡음신호에 Hamming 창을 곱하고 FFT(fast Fourier transform)를 취해서 얻은 스펙트럼 밀도의 평균값을 나타내고 있다. 여기서 창의 길이는 512이며 50% 중첩하면서 진행하고 10프레임을 평균하였다. Rectangular 창은 주엽(main lobe)의 폭이 좁아 스펙트럼 해상도는 좋지만 고주파 누설(leakage) 성분이 많고, Hamming 창은 main lobe의 폭이 넓어서 스펙트럼 해상도가 좋지 않은 반면 고주파 누설(leakage) 성분이 적다. 따라서 50% 중첩을 감수하면서 Hamming 창을 이용한다.

그림 1에서 지속인 경우 자동차 잡음은 엔진의 회전수에 의한 500Hz 이하 저주파 영역의 잡음이 지배적이며 주행속도가 증가할수록 고주파 성분이 증가하였다. 팬을 동작시킨 경우를 제외하고는 소나타와 프레스토의 잡음의 분포 형태가 비슷하였으며 같은 환경에서는 소나타가 전 대역에 걸쳐 잡음전력이 낮았다. 팬의 일반적인 잡음의 분포는 바람에 의한 고주파 성분이 많이 포함되어 있으며 특히 팬의 성능이 좋은 소나타의 경우의 스펙트럼 분포는 거의 평탄하였다. 그리고 도로의 상황(아스팔트, 콘크리트)에 따라 잡음의 분포는 큰 차이가 없었다.



(a)



(b)

그림 1. 잡음의 전력 스펙트럼 밀도

(a) 엔진회전수와 속도 (b) 차량에 따른 팬

Fig 1. The power spectral densities of noise

(a) engine rate and speed (b) fan of vehicles

III. 적응 스펙트럼 차감법

하나의 마이크로폰을 이용하는 스펙트럼 차감법은 잡음신호만 존재하는 비음성 구간에서 잡음의 스펙트럼을 추정하여 다음에 입력되는 잡음이 포함된 음성신호의 스펙트럼에서 추정 잡음의 스펙트럼을 차감함으로써 잡음을 제거한 음성신호를 추출하는 방법이다. 따라서 정확한 음성/비음성 분류와 잡음의 스펙트럼 추정은 잡음 제거의 성능을 좌우하기 때문에 중요한 사항이다.

입력신호의 신호대 잡음비가 낮은 경우 상호상관 함수를 이용하여 추출된 피치 정보에 의하여 음성/비음성을 분류하는 방법은 계산량도 많으며 엔진 잡음과 같은 시간적인 상호 상관성이 큰 주기적인 성분에

의하여 정확한 음성/비음성의 분류가 어렵다. 따라서 비정적인 차량잡음의 환경하에서 간단하면서 정확한 음성/비음성의 연속적인 분류가 필요하고 분류된 결과에 따라 잡음의 스펙트럼도 변화된 환경하에서 연속적으로 개선되어야 잉여 잡음이 감소된 깨끗한 음성신호를 추출할 수 있다.

1. 부가적인 잡음의 모델

잡음에 의하여 손상된 음성신호 $ns_l[i]$ 는 깨끗한 음성신호 $s_l[i]$ 와 부가적인 비정적 차량잡음 $n_l[i]$ 의 합으로 구성 되어진다.

$$ns_l[i] = s_l[i] + n_l[i] \quad i = 0, 1, \dots, N-1 \quad (1)$$

여기서 i 는 프레임 지수(frame index) l 에서의 이산 시간 지수(discrete-time index)이다. Hamming창은 50% 중첩하면서 진행하고 FFT(fast Fourier transform)의 크기는 창 의 크기 N 과 같다. (1)식에 FFT를 취하면 다음과 같다.

$$NS_l(k) = S_l(k) + N_l(k) \quad k = 0, 1, \dots, N-1 \quad (2)$$

여기서 k 는 프레임 지수 l 에서의 이산주파수 지수(discrete-frequency index)이다. $NS_l[k]$ 의 스펙트럼은 다음과 같이 근사화 되어질 수 있다.

$$|NS_l(k)| = |S_l(k)| + |N_l(k)| \quad k = 0, 1, \dots, N-1 \quad (3)$$

여기서 입력신호의 스펙트럼 $|NS_l[k]|$ 만을 알 수 있고 잡음과 음성신호의 스펙트럼은 알 수 없다. 따라서 미지의 잡음 스펙트럼 $|N_l[k]|$ 을 추정하여 잡음에 의하여 손상된 음성신호의 스펙트럼 $|N_l[k]|$ 에서 잡음의 스펙트럼만 추출하였을 때 깨끗한 음성신호 $|S_l[k]|$ 를 얻기 때문에 잡음의 스펙트럼 $|N_l[k]|$ 를 정확히 추정하는 것이 스펙트럼 차감법의 성능을 좌우한다. 위상정보는 출력단에서 IFFT할 때 이용된다.

2. 음성/비음성 분류 함수

하나의 마이크로폰을 이용하는 스펙트럼 차감법에 의한 성능은 잡음의 스펙트럼을 추정하는 방법에 의하여 좌우되기 때문에 정확한 음성/비음성의 구분이 필요하다. 상호상관에 의한 피치 정보로써 음성/비음성을 분류하는 방법은 계산량이 많고 주기적인 성분

이 지배적인 차량잡음의 환경에서는 적합하지 못하다. 따라서 제안된 방법에서는 계산량이 적고 간단한 음성/비음성의 분류 함수를 제안하여 정확한 음성/비음성의 분류와 적응적 잡음의 스펙트럼 추정에 이용하여 잡음제거의 성능을 향상시켰다.

1) 변화분에 의한 특징추출

음성 구간과 비음성 구간이 구분되는 특징중의 하나는 시간영역에서 음성신호의 시간에 대한 변화분이 잡음신호의 변화분 보다 크다는 것이다. 한 프레임내에서 이웃하는 샘플간에 변화분 에너지의 평균값을 표현하면 다음과 같다.

$$V_l = \frac{1}{N-1} \sum_{i=1}^{N-1} (ns_l[i] - ns_l[i-1])^2 \quad (4)$$

그림 2의 잡음 섞인 음성신호에 대하여 한 프레임내에서 이웃하는 샘플간의 변화분 에너지의 평균값은 그림 3과 같이 음성 프레임에서는 큰 값의 V_l 을 가지는 반면 비음성, 즉 잡음의 프레임에서는 작은 값의 V_l 을 가지게 되므로 음성/비음성 분류를 위한 특징으로 이용할 수 있다. 잡음의 정도에 대한 기준이 설정되어야 음성과 비음성 프레임의 분류가 가능하므로 잡음에 대한 기준값으로써 잡음 프레임의 V_l 만을 평균한것을 이용한다.

$$\bar{V} = \frac{1}{M} \sum_{l=0}^{M-1} V_l \quad (5)$$

(5)식은 처음 짧은 시간에 입력되는 신호는 잡음만으로 구성되어 있다는 가정하에서 처음 M 개 프레임의 V 값으로 초기 \bar{V} 를 구하는 것이다. 임의의 입력 프레임의 V 값이 \bar{V} 보다 많이 크면 그 프레임은 음성 프레임일 확률이 크고 \bar{V} 보다 작거나 비슷하면 비음성 프레임일 확률이 크다. 그리고 음성 프레임인지 비음성 프레임인지 구분이 모호한 값을 가지는 프레임도 있다. 이와 같은 성질을 이용하여 분류 함수로써 표현하며 다음과 같다.

$$J_{Vl} = \begin{cases} 1, & V_l \geq C_H \cdot \bar{V} \\ 2 \cdot \left\{ \left[\frac{\log((V_l - \bar{V})/\bar{V}) - \log(C_L - 1)}{\log(C_H - 1) - \log(C_L - 1)} \right] - \frac{1}{2} \right\}, & C_L \cdot \bar{V} < V_l < C_H \cdot \bar{V} \\ -1, & V_l \leq C_L \cdot \bar{V} \end{cases} \quad (6)$$

여기서 C_H, C_L 는 입력 프레임이 음성 프레임, 비음성 프레임 혹은 모호한 프레임인지를 구별하는 문턱치를 결정하기 위한 상수이며 실험적으로 구하였다. 상한 문턱치 보다 큰 V 의 프레임의 분류함수는 1로 두고 하한 문턱치 보다 작은 V 의 프레임의 분류함수는 -1로 둔다. 상한 문턱치와 하한 문턱치의 사이에 존재하는 V 에 대해서는 V 의 값이 하한 문턱치 근처에 많이 분포하기 때문에 \log 함수를 이용하여 분류함수 값이 식(6)에서와 같이 -1~1 사이에 선형적인 분포를 가지도록 하였다.

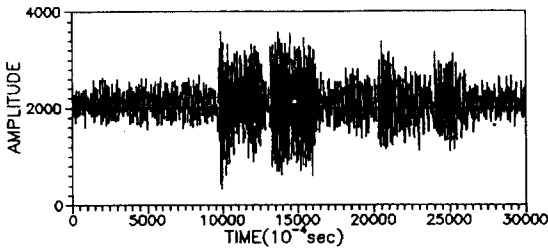


그림 2. 잡음섞인 음성신호
Fig 2. Noisy speech signal

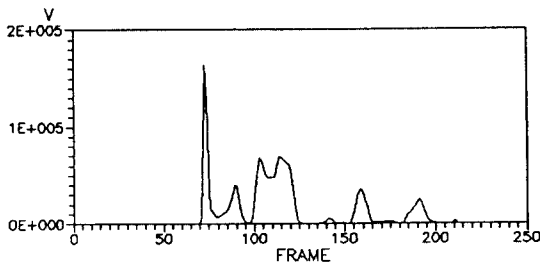


그림 3. 잡음섞인 음성신호의 변화분에 의한 특징추출
Fig 3. Feature extraction by variation of noisy speech signal

2) 스펙트럼 분포에 의한 특징추출

음성 구간과 비음성 구간의 구분되는 다른 특징은 주파수 영역에서 음성 프레임의 스펙트럼 분포가 비음성 즉 잡음 프레임의 스펙트럼 분포와 다르다는 것이다. 자동차 잡음의 스펙트럼은 500Hz 이하에 대부분이 분포하며 음성은 그 보다 높은 대역에도 많이 분포한다. 입력 프레임의 스펙트럼이 추정한 잡음의 스펙트럼 보다 클때 각 주파수 성분의 평균 차의 에너지를 식(4)와 같이 임의의 프레임에 대한 스칼라 값으로 다음과 같이 표현하였다.

$$Q_l = \frac{1}{N} \sum_{k=0}^{N-1} (|NS_l(k)| - |\tilde{N}(k)|)^2, |NS_l(k)| > |\tilde{N}(k)| \quad (7)$$

$$\bar{Q} = \frac{2}{M+1} \sum_{l=(M-1)/2}^{M-1} Q_l \quad (8)$$

식 (7)과 (8)에서 잡음의 초기 스펙트럼 $|N[k]|$ 는 식(4)와 (5)에서와 같이 처음 M 개의 프레임이 비음성 프레임이라는 가정하에서 기수 M 프레임중 앞의 $(M-1)/2$ 개 프레임의 평균 스펙트럼으로 구하고 잡음의 기준값으로써 Q 의 평균값 \bar{Q} 는 M 개중 뒤의 $(M+1)/2$ 개의 프레임에서 구하였다.

3) 분류함수

Q 에 대한 분류함수는 V 에 대한 분류함수와 같은 의미와 형태를 가진다. 따라서 최종적인 분류함수는 각각의 함으로 다음과 같다.

$$J_l = J_{Vl} + J_{Ql} \quad (9)$$

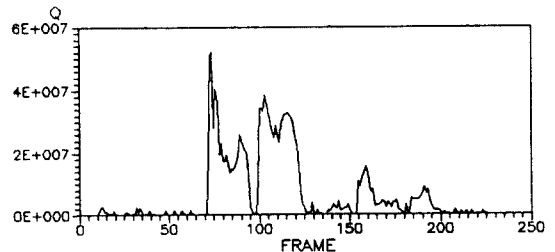


그림 4. 잡음섞인 음성신호의 스펙트럼 분포에 의한 특징추출
Fig 4. Feature extraction by spectral distribution of noisy speech signal

여기서 분류함수 J 는 변화하는 잡음의 환경하에서 음성/비음성의 분류에 이용되며 또한 음성 신호에 포함된 잡음의 스펙트럼을 추정하는데 이용된다.

3. 스펙트럼 차감법

음성신호에 포함된 잡음의 스펙트럼은 많은 프레임을 평균할수록 추정 잡음의 스펙트럼에 근접할 것이다. 그러나 음성신호는 비정적이기 때문에 제한된 수의 프레임의 평균이 가능하다. $l=0$ 인 처리하고자 하는 프레임과 전후 프레임을 평균하면 음성신호의 왜곡이 거의 없고, 잡음신호의 스펙트럼은 추정잡음 스펙트럼의 분포와 같아질 확률이 커진다. 따라서 음성신호의 왜곡없이 잡음신호를 잘 제거할 수 있다.

$$|\overline{NS}_l(\mathbf{k})| = \frac{1}{3} \sum_{i=-1}^1 |NS_i(\mathbf{k})| \quad (10)$$

여기서 위상 성분은 $l=0$ 인 프레임의 값을 출력단에서 IFFT할때 이용한다. 입력신호의 스펙트럼을 변형하였고 잡음의 스펙트럼을 추정하였으므로 음성신호의 스펙트럼은 차감법에 의하여 다음과 같이 추정할 수 있다.

$$|S_l(\mathbf{k})| = \begin{cases} |\overline{NS}_l(\mathbf{k})| - |\tilde{N}(\mathbf{k})|, & |\overline{NS}_l(\mathbf{k})| > |\tilde{N}(\mathbf{k})| \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

추정한 음성신호의 스펙트럼과 앞에서 저장된 위상 정보를 이용하여 시간영역의 신호로 IFFT 한다. 비음성 구간의 잉여 잡음신호의 크기를 축소시키면 좀더 깨끗한 음성 신호를 들을 수 있다. 그러나 너무 많이 줄이면 음성신호에 포함된 잉여 잡음의 효과가 강조되는 효과가 있어 적절히 조정하여야 한다. 최종적인 음성/비음성 프레임의 결정은 처리하고자 하는 프레임과 전후 프레임의 분류함수의 합으로 결정한다.

$$J_T = \sum_{l=-\lfloor M-1 \rfloor/2}^{\lfloor M-1 \rfloor/2} J_l \quad (12)$$

M 개의 분류함수의 합을 구하여 문턱치 보다 크거나 같으면 음성 프레임, 작으며 비음성 프레임으로 결정하여 IFFT 할때 비음성 구간에서는 잉여 잡음의 크기를 줄인다.

$$\hat{s}_l[i] = \begin{cases} IFFT[\hat{S}_l(\mathbf{k})], & J_T \geq \theta \\ C \cdot IFFT[\hat{S}_l(\mathbf{k})], & J_T < \theta \end{cases} \quad (13)$$

그림 5에서는 이상에서 설명한 내용을 개략적으로 나타낸 것으로 음성/비음성 분류함수값에 의하여 변화하는 잡음의 환경하에서 추정 잡음의 스펙트럼을 개선할 것인지를 결정하게 되며 다시 음성/비음성의 분류는 추정 잡음의 스펙트럼과 입력신호를 이용하여 분류함수 값을 구한다. 그리고 입력 스펙트럼에서 추정 잡음의 스펙트럼을 차감함으로써 잡음이 제거된 음성신호의 스펙트럼을 추정하며 시간영역의 신호로 변환할때 분류함수값에 의해 비음성 구간의 잉여 잡음의 크기를 줄인다. 이러한 과정을 입력되는 매 프레임에서 수행하여 비정적인 차량잡음을 적용적으로 제거하기 위해서는 초기값들을 계속해서 개선하여야 한다. D 는 지연을 의미한다.

4. 추정값의 개선

실제적인 자동차 잡음의 환경은 연속적으로 변화하며 상황에 따라 고주파 대역의 잡음도 다소 존재한다. 따라서 깨끗한 음성신호를 얻기 위해서는 변화된 환경에서 추정 잡음 스펙트럼을 적용적으로 매 순간 개선하여야 한다. 위에서 설명한 분류함수를 이용한 추정 잡음의 스펙트럼 개선은 모든 l 에 대해 $J_l \leq 0$ 이면 초기 추정 잡음의 스펙트럼을 다음과 같이 변화된 환경의 잡음 스펙트럼으로 개선한다.

$$|\tilde{N}(\mathbf{k})| = \frac{1}{M} \sum_{l=-\lfloor M-1 \rfloor/2}^{\lfloor M-1 \rfloor/2} |NS_l(\mathbf{k})| \quad (14)$$

여기서 $l=0$ 는 처리하는 프레임이며 양의 값과 음의 값을 가지는 프레임은 각각 전후 프레임을 의미한다. 식(5)와 (8)에서의 잡음신호의 기준값의 초기값들도 식 (14)와 동일한 조건하에서 개선시킨다.

$$\bar{V} = \frac{1}{M} \sum_{l=-\lfloor M-1 \rfloor/2}^{\lfloor M-1 \rfloor/2} V_l \quad (15)$$

$$\bar{Q} = \frac{1}{M} \sum_{l=-\lfloor M-1 \rfloor/2}^{\lfloor M-1 \rfloor/2} Q_l \quad (16)$$

이상에서 전후 프레임을 고려하기 때문에 $(M-1)/2$ 개 프레임의 진행만큼의 시간적 지연이 필요하다.

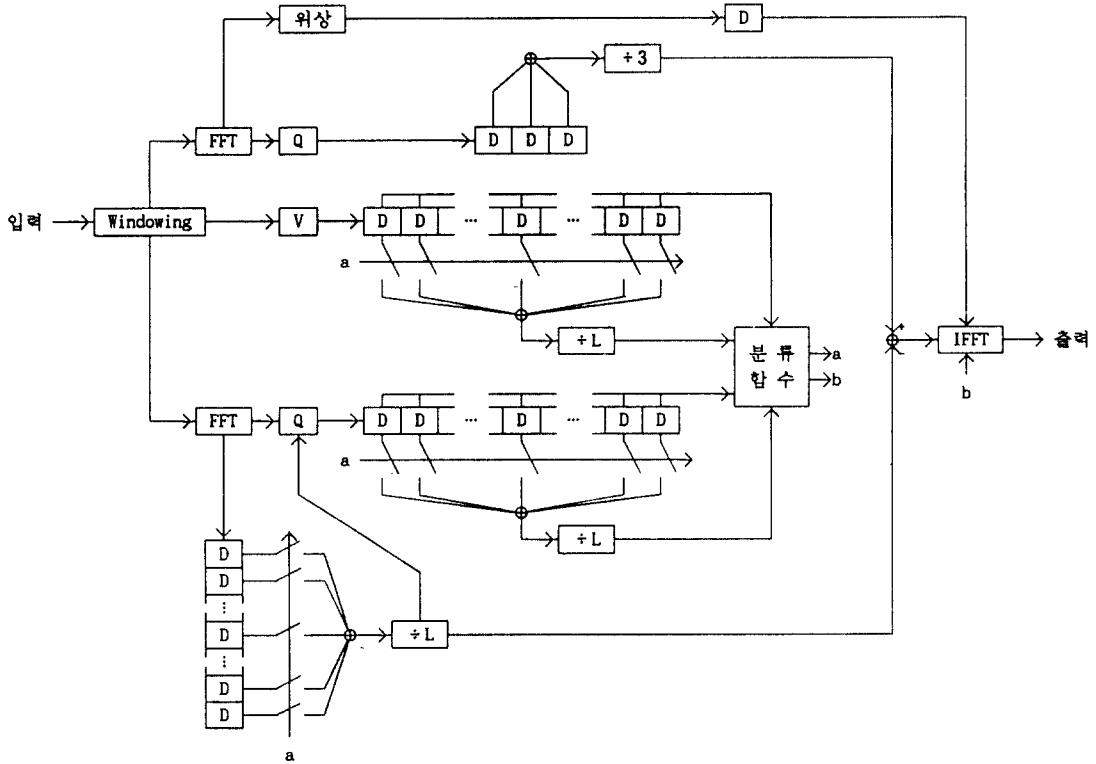


그림 5. 제안된 스펙트럼 차감법의 개략도
Fig. 5. The block diagram of the proposed spectral subtraction

IV. 실험 및 결과

1. 실험방법

실험에서 사용된 변수들은 다음과 같다: $N=256$; $M=9$; $C_H=3.5$; $C_L=1.5$; $\theta=0$; $C=0.05$. 실험실에서 획득한 깨끗한 여자의 음성신호에 자동차 잡음 신호를 더한 신호를 입력으로 이용하여 처리한 결과를 신호대 잡음비의 개선 및 왜곡의 개선 정도로서 성능을 평가한 후 실제의 자동차 환경하에서 얻은 남자 음성신호 데이터에 대해서 실험하였다. 방향성을 가지는 마이크로폰을 자동차의 계기판에 부착하고 녹음을 하여 12kHz로 샘플링 하였으며 발음한 음성은 “안녕하세요 감사합니다”이다. 실험한 자동차는 잡음의 정도가 심한 프레스토이며 아스팔트 주행시에 창문은 닫은 상태에서 실험하였다. 객관적인 개선 정도를 판단하기 위해 다음과 같이 신호대 잡음비와 왜곡을 계산하였다.

$$SNR = \frac{\sum_{i=0}^{N-1} |s[i]|^2}{\sum_{i=0}^{N-1} (s[i] - f[i])^2} \quad (17)$$

$$왜곡 = \frac{1}{N} \sum_{i=0}^{N-1} |s[i] - f[i]| \quad (18)$$

여기서 $s[i]$ 는 평가하고자 하는 신호이며 $f[i]$ 는 원 신호이다. 본 실험에서는 “안녕하세요”구간과 “감사합니다”구간으로 나누어 계산하였다.

2. 실험결과

그림 6은 음성신호를 나타내며 그림 7의(a)는 그림 6의 음성신호와 프레스토가 아스팔트를 30km/h로 주행할 때의 잡음신호를 더한, 잡음 섞인 음성신호이고 (b)는 제안된 스펙트럼 차감법을 이용하여 잡음을

제거한 음성신호이다. 그때의 신호대 잡음비와 왜곡은 표 1과 같다. 그림 8은 80km/h의 아스팔트 주행시 프레스토에서 즉 실제의 잡음이 존재하는 환경하에서 얻은 잡음섞인 음성신호이고 (b)는 처리한 결과이다.

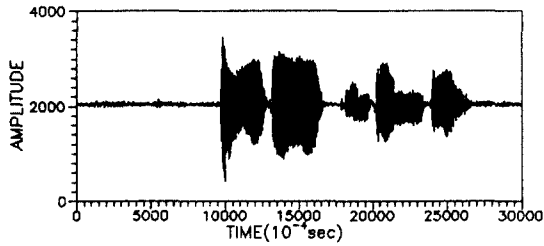
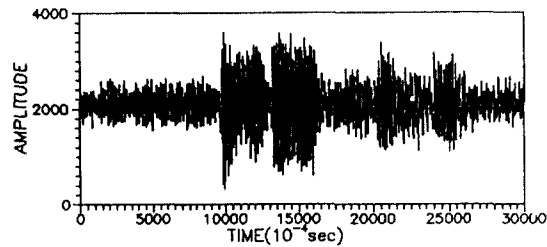
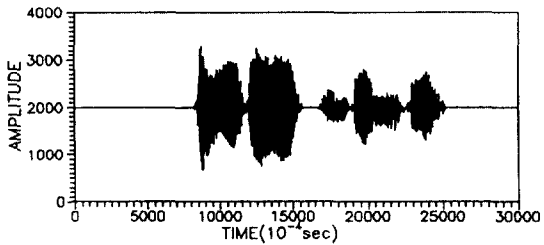


그림 6. 원 음성신호(여자)
Fig 6. Original speech signal(female)



(a)



(b)

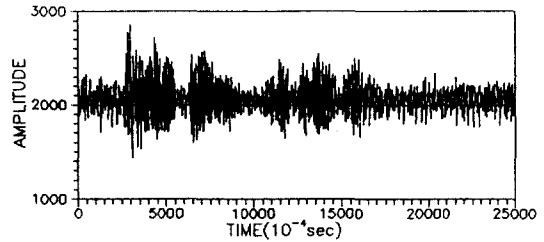
그림 7. 원 음성신호와 잡음신호(프레스토, 아스팔트, 30km/h)의 합
(a) 잡음섞인 음성신호 (b) 실험결과

Fig 7. The sum of original speech and noise signal (Presto, asphalt, 30km/h)
(a) Noisy speech signal (b) Processing result

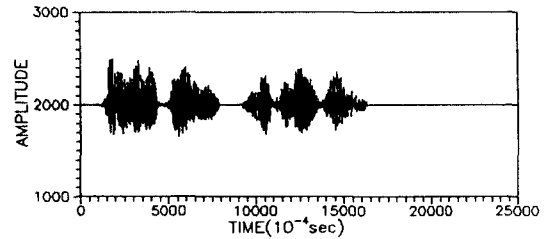
표 1. 신호대 잡음비와 왜곡 개선(프레스토, 아스팔트, 30km/h)

Table 1. The improvement of SNR and distortion (Presto, asphalt, 30km/h)

	안녕하세요		감사합니다	
	SNR	왜곡	SNR	왜곡
입력신호	4.536 dB	185.880	0.642 dB	191.651
출력신호	29.363 dB	69.649	7.484 dB	56.302
개선정도	24.827 dB	116.321	6.842 dB	135.349



(a)



(b)

그림 8. 자동차 환경하에서의 잡음섞인 음성신호(남자, 프레스토, 아스팔트, 80km/h)
(a) 잡음섞인 음성신호 (b) 실험결과

Fig 8. Noisy speech signal on the field(male, Presto, asphalt, 80km/h)
(a) Noisy speech signal (b) Processing result

V. 결 론

차량내의 각종 환경하에서 잡음 및 음성 데이터를 획득하고 이의 특성을 분석하였다. 두개의 마이크로폰의 구조의 공간적 상호상관을 이용하는 적응잡음 제거 방법은 두개의 마이크로폰의 거리가 가까우면 잡음제거의 효과가 증대되지만 주 마이크로폰에만

입력되어야 하는 음성신호가 기준 마이크로폰에도 입력되어 출력신호의 전력이 감소되는 문제점을 가지고 있다.

주기적인 성분이 지배적인 차량잡음 환경하에서는 시간적 상호상관에 의한 피치 정보로써 음성/비음성을 분류하는 방법은 적합하지 않아 본 논문에서는 계산량이 적고 구조적으로 간단하며 정확한 음성/비음성의 분류가 이루어지도록 하는 분류함수를 제안하였다. 이를 변화된 환경에서 정확한 음성/비음성 분류와 적응적으로 잡음의 스펙트럼을 추정하는데 이용하여 잉여잡음 성분이 적고 비교적 왜곡이 적은 음성신호를 얻을 수가 있었다. 깨끗한 음성신호와 차량잡음신호를 더한 음성신호에 대해서 실험하여 많은 신호대 잡음비와 왜곡의 개선이 있음을 확인하였고 실제의 차량환경에서 얻은 잡음 섞인 음성신호에 대하여 실험한 결과 잡음이 제거된 깨끗한 음성신호를 얻을 수 있었다. 프레스토와 소나타의 여러 조건하에서 전반적으로 잡음을 잘 제거하였으나 배경 잡음이 매우 급격하게 변화하는 경우에 대해서는 더욱 많은 연구가 필요하다.

참 고 문 헌

1. M. M. Goulding and J. S. Bird, "Speech enhancement for mobile telephony," *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, p.316, Nov. 1990.
2. B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692-1716, Dec. 1975.
3. R. B. Wallace and R. A. Goubran, "Improved tracking adaptive noise canceler for nonstationary environments," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp.700-703, March 1992.
4. N. Dal Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," *Signal processing*, vol. 15, no. 1, pp.43-56, July 1988.
5. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. ASSP*, Vol. ASSP-27, 1979, pp.113-120.
6. C. S. Wu, V. V. Nguyen, H. Sabrin, W. Kushner, and J. Demoulakis "Fast self-adapting broadband noise removal in the cepstral domain," *Proc. ICASSP*, 1991.
7. Haykin, S., *Adaptive filter theory*, Prentice-Hall Englewood Cliffs, New Jersey, 1986.
8. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, New Jersey, 1985.
9. M. R. Sambur, "Adaptive noises cancelling for speech signals," *IEEE Trans. ASSP*, pp.419-423, 1978.
10. I. Lecomte, M. Lever, J. Boudy, and A. Tassy, "Car noise processing for speech input," *Proc. ICASSP*, 1989.
11. C. Wheddon and R. Linggard, *Speech and Language Processing*, Chapman and Hall



朴 英 植 (Young Sik Park) 정회원
 1969년 4월 18일생
 1992년 2월 : 경북대학교 공과대학
 전자공학과 졸업(공학사)
 1994년 2월 : 경북대학교 대학원 전
 자공학과 졸업(공학석사)

1994년 3월~현재: 경북대학교 대학원 전자공학과 박사과정
 ※주관심분야: 신호처리 영상압축



李 駿 宰 (Jun Jae Lee) 정회원
 1962년 5월 31일생
 1986년 8월 : 경북대학교 공과대학
 전자공학과 졸업(공학사)
 1990년 8월 : 경북대학교 대학원 전
 자공학과 졸업(공학석사)

1991년 9월~1993년 7월 : 경북대학교 조교
 1993년 2월 : 경북대학교 대학원 전자공학과 박사과정 수료
 ※주관심분야: 신호처리 영상처리 및 컴퓨터비전



李應柱(Eung Joo Lee) 정회원
1965년 3월 20일생
1990년 2월 : 경북대학교 공과대학
전자공학과 졸업(공학
사)
1992년 2월 : 경북대학교 대학원 전
자공학과 졸업(공학석
사)

1993년 3월 ~ 현재 : 경북대학교 대학원 전자공학과 박사과정
※주관심분야 : 신호처리 영상처리 및 컴퓨터비전



河永浩(Yeong Ho Ha) 정회원
1953년 8월 10일생
1976년 2월 : 경북대학교 전자공학
과(공학사)
1978년 8월 : 경북대학교 전자공학
과(공학석사)
1985년 8월 : University of Texas
at Austin 공학박사

1985년 9월 ~ 현재 : 경북대학교 전자공학과 부교수
※주관심분야 : 컴퓨터비전, 영상처리, 신호처리, HDTV
등입