퍼지 및 지능시스템학회 논문지 2007, Vol. 17, No. 7, pp. 945-951

이동형 단말기를 위한 다채널 입력 기반 비정상성 잡음 제거기

Multi-channel input-based non-stationary noise cenceller for mobile devices

정상배 · 이성독
Sang-bae Jeong · Sung-doke Lee

한국정보통신대학교 공학부

요 약

잡음의 제거는 음성을 인터페이스로 하는 기기들에 필수적이라고 할 수 있다. 실질적으로 통화 품질이나 음성 인식률은 음성 입력부의 주변에서 들어오는 원치 않는 가산성 잡음에 의해서 크게 열화된다. 본 논문에서는 기본적으로 두 개의 마이크로폰을 이용한 잡음제거 방법을 제안한다. 마이크를 여러 개 사용했을 때의 장점은 방향 정보를 이용할 수 있다는 것인데, 이는 사람 목소리, 음악 소리 등의 비정상성 잡음을 제거하는 데에 유용하다. 제안된 잡음제거 알고리즘은 위너필터에 기반 한다고 볼 수 있다. 위너필터에 의한 잡음제거를 위해서는 검출하고자 하는 음성과 제거하고자 하는 잡음의 주파수 응답이 동시에 추정 가능해야 한다. 이를 위해서 주파수 영역에서 스펙트럼 분류를 시행하여 위너필터 기반의 잡음제거에 필요한 정보를 얻는다. 제안된 알고리즘을 이용한 성능은 잘 알려진 프로스트 (Frost) 알고리즘 및 적응 모드 컨트롤러를 갖는 generalized sidelobe canceller (GSC)와 비교하였다. 성능의 지표로는 객관적 음질 평가의 방법 중에서 널리 쓰이고 있는 perceptual evaluation of speech quality (PESQ) 및 음성 인식률이 사용되었다.

키워드 : 잡음제거, 빔포밍, 음질향상, 음성인식

Abstract

Noise cancellation is essential for the devices which use speech as an interface. In real environments, speech quality and recognition rates are degraded by the additive noises coming near the microphone. In this paper, we propose a noise cancellation algorithm using stereo microphones basically. The advantage of the use of multiple microphones is that the direction information of the target source could be applied. The proposed noise canceller is based on the Wiener filter. To estimate the filter, noise and target speech frequency responses should be known and they are estimated by the spectral classification in the frequency domain. The performance of the proposed algorithm is compared with that of the well-known Frost algorithm and the generalized sidelobe canceller (GSC) with an adaptation mode controller (AMC). As performance measures, the perceptual evaluation of speech quality (PESQ), which is the most widely used among various objective speech quality methods, and speech recognition rates are adopted.

Key Words : Noise cancellation, Beamforming, Speech enhancement, Speech recognition

1. 서 론

음성은 여러 가지 기기들을 제어할 수 있는 수단 중에 가장 편리한 인터페이스임에 틀림이 없다. 근래에 PDA (personal digital assistant) 및 휴대폰 같은 이동형 기기의 성능이 급속도로 좋아지고 있으며 그러한 기기들을 사용하는 사람들의 수도 꾸준히 증가하고 있다. 이에 따라서 기기의 부가가치를 높이기 위하여 다양한 기능들이 이동형 기기에 집적되고 있다. 여러 가지 기능 중에서 음성 통신시의 통화 품질 및 음성 인식률의 성능은 가산성 잡음에 매우 민감하다고 볼 수 있으며 향후 음성인터페이스 시장의 확대를 위해서는 실제 상황에서 안정적으로 동작할 수 있는 잡음 제거기의 개발이 필수적일 것이다. 음성통신 시스템에서의 잡음제거는 전송 음질의 향상을 통해서 수화자의 통화 만족도를 높이는 데에 그 목적이 있다. 음성인식기를 채택하는 기기 측면에서는 음성 검출의 효율성 향상 및 검출된 음성 구간에서의 신호를 향상시켜서 음성 인식률의 향상을 가져오게 한다.

음성인터페이스의 성능을 열화시키는 가산성 잡음은 자동차 엔진 소음, 사무실의 펜 잡음 등의 정상성(stationary)인 것과 사람의 목소리 음악 소리 등의 비정상성(non-stationary)인 것으로 크게 나눌 수 있다. 정적인 잡음의 제거는 단일 마이크로폰 입력과 위너필터 및 칼만 필터를 비롯한 다

접수일자 : 2007년 11월 22일
완료일자 : 2007년 12월 3일

양한 방식으로 수행이 가능하다[1-2]. 일반적으로는 신호의 초기 200 ms 구간 또는 실시간 음성검출기를 이용하여 비음성 구간에서 잡음의 통계치를 추정하고 음성 구간에서 잡음의 통계량이 차감되도록 필터를 설계한다. 이때에 추정된 잡음의 통계량이 음성 구간에서 크게 바뀌지 않아야 한다는 것과 잡음과 음성의 특성이 통계적으로 독립이라는 가정이 선행적으로 부가된다. 그렇지만, 일상생활에서 음성인터페이스의 성능을 열화시키는 잡음은 정상성인 것 보다는 비정상성인 것이 더 많다. 비정상성 잡음의 특징은 신호의 크기 및 주파수 응답이 지속적으로 변하여 신호 자체만으로는 목적 음성 신호와의 구분이 어렵다는 것이다. 따라서 기존의 단일 마이크로폰 기반의 잡음 제거를 위하여 실시간 음성검출을 수행한다 하더라도 검출 성능이 현저하게 떨어질 수 있으며, 설령 정밀한 검출이 가능하다 하더라도 추정된 잡음의 통계량이 음성 구간의 통계량과 다를 수 있어서 성능을 보장할 수 없게 된다.

그러한 비정상성 잡음을 제거하기 위한 최근의 연구는 마이크로폰 어레이를 이용한 기법들에 집중되고 있다. 마이크로폰 어레이를 이용한 잡음제거 기법은 기본적으로 목적 신호의 방향정보 혹은 목적 신호 혹은 잡음 신호의 섞임 정보를 이용할 수 있으므로 단일 마이크로폰 기법보다는 더 좋은 성능을 기대할 수 있다. 마이크로폰 어레이 기반 잡음제거 알고리즘은 크게 맹목 신호 분리(blind source separation: BSS) 와 빔포밍(beamforming) 알고리즘으로 나눌 수 있다. 일반적으로 실제 환경에서는 목적 신호의 위치 정보를 선행정보로 사용하는 빔포밍 알고리즘이 BSS 보다 더 좋은 성능을 보이고 있다. 빔포밍 알고리즘 중에서는 프로스트(Frost) 알고리즘과 GSC(generalized sidelobe canceller) 알고리즘이 신뢰성있는 성능을 보이는 것으로 알려져 있다[3-6]. 그렇지만, 이동형 기기처럼 두 개 이상의 마이크를 사용할 수 없는 상황에서는 그 성능이 만족스럽지 못하다. 본 연구에서는 이동형 기기에 적합한 잡음 제거 알고리즘을 제안한다. 기본적으로 두 개의 마이크로폰을 이용한 빔포밍에 기반하고 있으며 목적 신호원의 위치 정보 및 시간-주파수 마스킹(masking) 기법을 사용하여 잡음과 목적신호의 스펙트럼을 주파수 영역에서 분리한다. 그런 후에 주파수 영역에서 신호대 잡음비(SNR: signal-to-noise-ratio)를 추정하고 위너필터를 통한 최종적인 잡음제거를 수행한다.

본 논문의 구성은 제 2장에서 관련 연구에 관한 내용을, 제 3장에서 제안된 방식의 잡음제거 기법을 논한다. 제 4장에서는 실험 및 결과에 대해서 논하고 제 5장에서는 결론을 기술한다.

2. 관련연구

2.1 프로스트 알고리즘

그림 1은 프로스트 알고리즘의 개요를 나타낸 것이다[6]. 그림 1에서 K는 마이크로폰의 개수를, J는 각 입력 신호에 사용되는 적응 필터의 길이를 나타낸다. 개념상으로 프로스트 알고리즘은 목적 신호원의 위치를 마이크로폰 어레이의 정면으로 가정하는데, 실제 구현 측면에서는 선행 정보로 입력되는 목적 신호원의 위치 정보를 바탕으로 각 채널 입력에 적당한 시간 지연 효과를 부가해 줌으로써 가정을 충족시킬 수 있다. 그림 1의 (b)에서는 정면의 목적 음성 신호 방향에 대한 주파수 특성을 구하기 위한 등가 구조도를 나타낸 것인데, 그림 1의 (a)에서 점선으로 연결된 필터의 값들을 모두 합한 것을 등가 구조도에서의 필터 값으로 할당한다. 목적 음성 신호 방향으로의 주파수 응답 왜곡을 일으키지 않으면서 잡음 제거가 가능하려면, 그림 1의 (b)에서 등가 필터의 단위 충격 응답이 $\delta(n)$을 유지하도록 하면서 전체적인 빔포밍 출력 $y(n)$의 전력을 최소화 시킴으로서 가능하다. 프로스트 알고리즘에 의한 빔포밍 기법을 수식 1에 정리하였다.

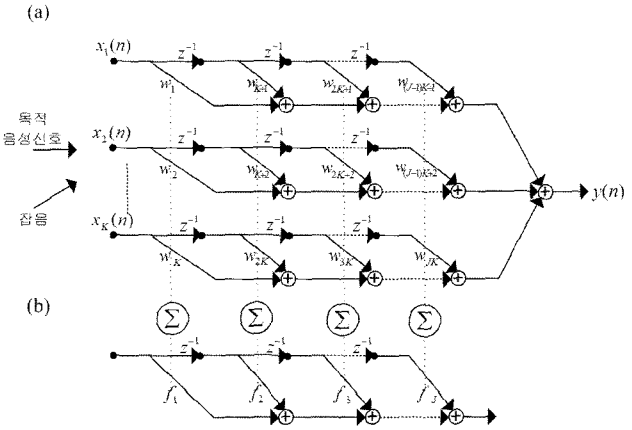


그림 1. 프로스트 알고리즘의 구조((a) 기본 구조도, (b)목적 음성 신호 방향의 등가 구조도)

Fig. 1. Structure of the Frost algorithm ((a) basic structure, (b) equivalent structure for the target speech source direction)

$$\min_{\vec{w}} E[y^2(n)], \text{ subject to } f(n)=\delta(n) \tag{1}$$

여기서 $\vec{w}=[w_1,...,w_{JK}]^T$, $E[\bullet]$는 통계적 평균치 연산자이다. 수식 (1)의 최적화는 Lagrange multiplyer를 이용해서 수행할 수 있다. 수식 (2)에 그 결과를 나타내었다.

$$\vec{w_{opt}}=R_{xx}^{-1}C[C^TR_{xx}^{-1}C]^{-1}F \tag{2}$$

R_{xx}는 입력을 $\vec{x}=[\vec{x_1^T}(n),\vec{x_2^T}(n),...,\vec{x_{J-1}^T}(n)]^T$ 로 두었을 때의 자기 상관도 행렬이다. 여기서, $\vec{x_{n_0}}(n)=[x_1(n-n_0),...,x_K(n-n_0)$이다. 행렬 C는 그 열벡터가 그림 (1)의 (a)에서 점선으로 표기한 각 필터의 값들을 합하는 역할을 할 수 있도록 구성된 것이다. 예를 들어, 행렬 C의 i 번째 열벡터는 그림 (1)의 (a)에서 i 번째 필터 계수들을 합하는 연산을 시행하도록 1 과 0으로 구성된다. 마지막으로 $F=[1,0,...,0]^T$이다. 수식 (2)의 폐형 공식으로부터 빔포머의 최적 필터 계수를 추정할 수 있겠으나, 행렬의 차수가 너무 클 수 있으므로 계산적인 측면에서 부적절하다. 따라서 프로스트 알고리즘을 적용할 때에는 신호의 표본 영역에서 동작하는 적응 필터 기법을 사용하는 것이 좋다.

일반적으로 프로스트 기법의 단점은 너무나 강력한 제약조건으로 알고리즘의 수렴 속도가 빠르지 않다는 것이다. 그래서 참고 문헌 [8]에서와 같이 수렴 속도를 높이기 위해서 여러 가지 알고리즘들이 제안되고 있으나 음성신호처리 영역에서는 유용성이 검증된바가 없다. 늦은 수렴 속도는 잡음제거의 양에 한계를 가져올 수밖에 없다.

2.2 GSC 알고리즘

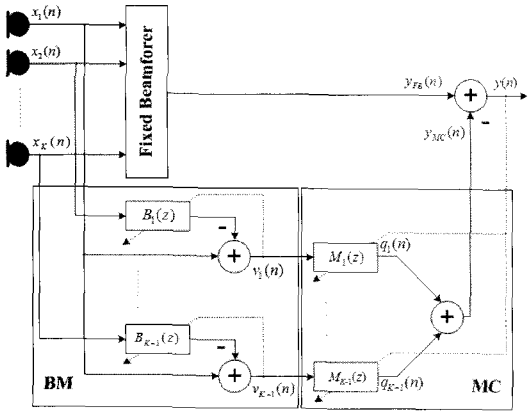


그림 2. GSC의 구조도 (BM: blocking matrix, MC: multiple input canceller)

Fig. 2. Structure of the GSC (BM: blocking matrix, MC: multiple input canceller)

그림 2에 GSC 알고리즘의 구조도를 나타내었다[4-5]. GSC에 의한 빔포밍은 고정 빔포밍으로부터 시작한다. 고정 빔포밍은 검출할 목적 음성 신호원의 위치를 바탕으로 적절한 시간 지연과 마이크로폰의 이득을 조정하여 목적 신호원의 전력이 최대화 되도록 한다. 목적 신호원의 전력은 앞서 언급한 다채널 입력의 위상을 정합시킴에 의해서 이루어지지만 잡음의 크기는 크게 줄어들지 않는다. 따라서 고정 빔포밍 후에 존재하는 잡음은 BM 및 MC 블록에 의해서 제거된다. 먼저 BM 블록은 목적 신호원의 위치 정보를 바탕으로 잡음 신호만을 추출해 내는 역할을 수행한다. 즉, 고정 빔포밍에 존재하는 잡음을 제거하기 위한 참조 잡음 신호를 만들어준다. 실질적인 잡음제거는 MC 블록에서 수행된다. BM에서 구한 참조 잡음 신호를 바탕으로 고정 빔포밍에 존재하는 잡음을 다수의 적응 필터를 이용하여 제거한다. 그렇지만, BM에서는 목적 신호 성분을 완벽하게 제거할 수 없으며 설령 제거한다 하더라도 MC 블록에서의 적응 필터링은 음성 신호를 심하게 왜곡시킬 수 있다. 따라서, GSC의 필터는 적응 모드 제어기(AMC: adaptation mode controller)에 의해서 그 동작이 조절된다. 적응 모드 제어기의 역할은 목적 음성 신호 구간과 순수 잡음 구간에서 필터의 동작을 다르게 하는 것이다. 목적 음성 구간에서는 BM 블록에서 목적 신호의 소거가 충분히 이루어지게 하기 위해서 적응 과 필터링을 동시에 수행하게 되고 MC 블록에서는 목적 음성 신호를 왜곡시키지 않기 위해서 필터링만 수행하게 된다. 순수 잡음 구간에서는 BM 블록에서는 필터링만 수행하게 되고 MC 블록에서는 적응과 필터링을 동시에 수행한다. BM 및 MC의 적응 필터의 학습은 수식 (3), (4)의 LMS (least mean square) 기반에 의한 알고리즘에 의해 이루어진다.

$$b_{i-1,k}^{(n+1)} = b_{i-1,k}^{(n)} + \mu(x_1(n) - v_{i-1}(n))x_i(n-k), \quad 2 \le i \le K-1 \tag{3}$$

$$m_{i,k}^{(n+1)} = m_{i,k}^{(n)} + \mu(y_{FB}(n) - y_{MC}(n))v_i(n-k), \quad 1 \le i \le K-1 \tag{4}$$

여기서, $b_{i,k}^{(n)}$는 시간 n에서 BM의 i번째 적응 필터의 k번째 계수를 나타내며, $m_{i,k}^{(n)}$은 시간 n에서 MC의 i번째 적응 필터의 k번째 계수를 나타낸다. μ는 학습률을 나타내는 상수이며, 수렴 속도의 향상을 위해서 필터 입력의 제곱 합으로써 정규화한 값을 사용할 수도 있다.

AMC 수행을 위한 목적 음성 신호 및 잡음 구간의 판별은 고정 빔포밍의 에너지 궤적 혹은 고정 빔포밍 출력과 MC 블록 출력 사이의 상관계수를 사용하고 있다. GSC는 앞서 언급한 프로스트 기법의 단점인 적응 필터의 수렴 속도를 획기적으로 빠르게 할 수 있는 장점이 있다. 그렇지만, 음성 신호를 위한 빔포밍 기법에서는 개선되어야 할 부분이 많다. 가장 큰 문제점은 AMC에 있다고 볼 수 있다. 안정적인 AMC를 수행하기 위해서는 음성 및 잡음 구간을 정확히 검출해줘야 하는데, 비정상성 잡음 환경에서는 그 신뢰도가 높지 못한 편이다. AMC의 구간 추정 오류가 발생할 경우에는 잡음 구간 신호의 감쇄를 크게 이루지 못하게 되며 음성 구간에서는 최악의 경우에 구간의 일부를 소거할 수도 있다. AMC의 성능이 어느 정도 높다 하더라도 MC 블록은 음성 구간에서 단순히 필터링만 수행하므로 잡음 제거의 효과를 크게 기대할 수는 없다.

3. 제안된 알고리즘

본 연구의 목적은 이동형 단말기에 적합한 잡음제기기의 구현에 있다. 제 2장에서 언급한 프로스트 기법이나 GSC 기반의 기법들은 주로 6개 이상의 마이크로폰을 장착할 수 있는 홈로봇용 음성 인터페이스에 적합하다고 볼 수 있다. 일단, 마이크로폰의 개수가 증가될 때 기존 빔포밍의 장점은 고정 빔포밍 후에 신호대 잡음비를 어느 정도 높일 수 있다는 것이다. 그런 후에 적응 필터에 의한 실질적인 잡음 제거를 이룰 수 있다. 여기서, 일반적인 잡음 제거 알고리즘의 특징 중의 하나는 입력 신호의 신호대 잡음비가 낮아질수록 잡음 제거 후의 음질 저하가 크게 발생한다는 것이다. 즉, 이동형 단말기에서처럼 마이크로폰의 개수가 적을 때는 기존의 방식을 따를 경우에 고정 빔포밍 후의 신호대 잡음비를 높이는 데에 한계가 따르므로 효율적인 음성 개선을 얻을 수 없다고 말 할 수 있다. 이에 본 연구에서는 기존의 빔포밍 방식의 개념에서 벗어난 새로운 알고리즘을 제안한다. 제안된 알고리즘은 기본적으로 참고 문헌 [7]에서 제안된 시간-주파수 영역의 마스킹 기법을 응용한다. 시간-주파수 영역에서의 마스킹은 선행 정보로서 입력되는 검출할 목적 음성 신호원의 위치 정보를 바탕으로 이루어지며 최종적으로는 잡음과 목적 음성신호의 주파수 응답을 구하는 역할을 수행한다. 이를 근간으로 하는 제안된 방식의 잡음제거를 그림 3에 나타내었고 3.1 절부터 3.5 절에서 이에 대한 자세한 설명을 하였다.

3.1 주파수 영역에서의 신호 분석

본 연구에서 제안한 알고리즘은 주파수 영역에서의 신호 분석으로부터 시작한다. 스테레오 마이크로폰으로부터 입력된 신호는 각각 단구간 주파수 분석으로써 이산 푸리에 변환 (DFT: discrete Fourier transform)을 거치게 된다. 그림 3에서 $x_1(n)$ 및 $x_2(n)$은 각 채널별로 입력된 신호, $X_1(k)$ 및 $X_2(k)$는 채널별 이산 푸리에 변환 후에 얻어지는 k번째 주파수 응답을 나타낸다.

3.2 채널 등화기의 설계

실질적인 잡음제거를 위한 과정에 앞서서 채널 등화기를 설계해 주는 것이 바람직하다. 일반적으로 마이크로폰을 포함한 A/D(analog-to-digital) 변환기의 특성은 채널에 따라서 같지 않다고 볼 수 있다. 두 채널간에 존재하는 부정합을 제거하기 위해서 수식 (5)의 비용함수를 고려한다.

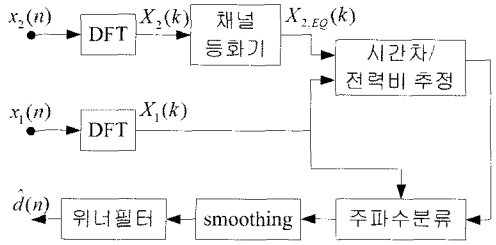


그림 3. 제안된 알고리즘의 구조도
Fig. 3. Structure of the proposed algorithm

$$J_k = \sum_{t=0}^{T-1} |D_1^{(t)}(k) - \alpha(k) D_2^{(t)}(k)|^2 \tag{5}$$

프레임 개수를 나타낸다. $D_1^{(t)}(k)$ 및 $D_2^{(t)}(k)$는 잡음이 없는 조건하에서 마이크로폰의 정면에서 수신되는 t번째 목적 음성 신호 프레임의 각 채널별 주파수 응답이다. $\alpha(k)$는 이산 주파수 영역에서 k번째 주파수 응답의 등화를 위한 복소 이득이다. 수식 (5)는 $\alpha(k)$에 대한 2차 함수이므로 단일의 최적해를 가진다. 따라서, $\alpha(k)$의 켤레값으로 수식 (5)의 양변을 미분한 후에 0으로 놓고 최적해를 구할 수 있다. 수식 (6)에 그 결과를 나타내었다.

$$\alpha_k^{(opt)} = \frac{\sum_{t=0}^{T-1} D_1^{(t)}(k)(D_2^{(t)}(k))^*}{\sum_{t=0}^{T-1} |D_2^{(t)}(k)|^2} \tag{6}$$

$(\bullet)^*$는 복소 켤레 연산자이다.

3.3 시간차 및 전력비의 추정 및 주파수 분류

본 연구의 잡음 제거는 기본적으로 음성 및 잡음의 스펙트럼 추정과 위너필터링에 의해서 이루어진다. 따라서 시간차 및 전력비의 추정은 가장 중요한 부분이라고 할 수 있다. 제 1 장 및 2 장에서 언급한 바와 같이 마이크로폰을 2개 이상 사용할 경우에 신호의 시간차를 통한 위치 정보의 파악이 가능하다. 일반적인 이동형 기기의 음성 입력은 기기의 정면에서 행하는 경우가 가장 자연스럽다고 볼 수 있다. 정면에서 음성 입력을 할 경우에 채널간의 시간차 및 전력비는 이상적일 경우에 각각 0 및 1이 될 것이다. 따라서 그 이외의 값을 갖는 신호 성분은 잡음원에서 나온 것으로 간주할 수 있다. 만약, 주파수 응답 측면에서 신호원과 잡음원의 성분 분리가 가능하다면 위너필터에 의한 잡음제거가 성공적으로 동작할 수 있다. 주파수 영역에서의 채널 간 전력차 $m(k)$ 및 시간차 $\delta(k)$ 및 를 수식 (7), (8)에 나타내었다.

$$m(k) = \frac{|X_{2,EQ}(k)|}{|X_1(k)|} \tag{7}$$

$$\delta(k) = -\frac{N}{2\pi k} \angle \frac{X_{2,EQ}(k)}{X_1(k)} \tag{8}$$

N은 단구간 주파수 분석을 위한 입력 신호 프레임의 길이이며, $\angle z$는 $-\pi \sim \pi$ 범위의 값을 갖는 복소수 z의 위상을 나타낸다. $X_{2,EQ}(k) = \alpha(k) X_2(k)$이다. 따라서 신호원의 위치가 선행정보로 주어졌을 때의 잡음 및 목적 음성 신호의 주파수 분류는 쉽게 이루어진다. 수식 (9)에 목적 음성 신호원의 위치가 마이크로폰 어레이의 정면에 존재할 경우의 주파수 분류 방법을 나타내었다.

$$X_1(k) = \begin{cases} target, & \text{if } (m(k)-1)^2 - \delta^2(k) < r_{TH}^2 \\ noise, & otherwise \end{cases} \tag{9}$$

여기서, r_{TH}는 주파수 분류를 위해서 선행 입력된 임계치를 나타낸다.

3.4 주파수 응답 스무딩

최종적인 잡음 제거의 수행을 위한 위너필터의 설계는 모든 주파수 성분에서의 신호대 잡음비를 필요로 한다. 그렇지만, 수식 (9)의 과정을 거쳐서 잡음 및 목적 음성 신호의 주파수 분류를 시행한다 하더라도 그 분류 방식이 이산적으로 수행되므로 분류 후에 각 주파수 응답에는 정의되지 부분이 존재할 수밖에 없다. 정의되지 않는 성분은 0으로 볼 수는 없다. 왜냐하면 일반적인 음향 신호의 주파수 응답은 인접한 것과 연관성이 크기 때문이다. 따라서, 본 연구에서는 정의되지 않는 부분의 주파수 응답 추정을 위하여 정규화된 해밍창 함수를 초기 추정된 주파수 응답과 복적분을 시행한다. 수식 (10)에서 정규화된 창 함수를, 수식 (11)에서 복적분에 의한 스무딩 방법을 나타내었다.

$$h_{norm}(k) = \frac{h(k)}{\sum_{n=0}^{L-1} h(n)} \tag{10}$$

$$|Q'(k)|^2 = \sum_{m=0}^{L-1} h_{norm}(m) |Q(k-m)|^2 \tag{11}$$

수식 (11)에서 L은 창 함수의 길이를, $Q(k)$는 임의의 복소 주파수 응답을 나타낸다. 이 때, 전력 스펙트럼 영역에서 스무딩을 시행하는 이유는, 위너필터의 설계에는 주파수 영역에서 신호대 잡음비 만을 필요로 하기 때문이다. 다시 말하면, 위상 응답에 대한 것은 고려할 필요가 없기 때문이다. 그에 대한 자세한 내용은 3.5절에서 다룬다.

3.5 위너필터

수식 (12)에서 위너필터에 대한 짧은 소개를 위한 비용함수를 나타내었다[10].

$$J_{Wiener} = E\left[\left(d(n) - \sum_{k=-\infty}^{\infty} w(k) x(n-k)\right)^2\right] \tag{12}$$

검출하고자 하는 원본 신호를 $d(n)$으로 두고 수신된 신호 $x(n)$이 $d(n)$과 잡음 신호 $u(n)$의 합으로 이루어진다고 가정한다. 수식 (12)의 최소화는 위너필터 계수 $w(n)$에 의해서 입력 신호 $x(n)$이 원본 신호 $d(n)$에 매우 유사해짐을 의미한다. 수식 (12)는 $w(n)$에 대한 2차 함수이므로 유일한 최적치가 존재한다. 수식 (13)에서 최적의 위너필터 계수를

주파수 영역의 응답으로 나타내었다.

$$W^{(opt)}(\omega) = \frac{P_d(\omega)}{P_d(\omega)+P_u(\omega)} = \frac{SNR(\omega)}{1+SNR(\omega)} \quad (13)$$

$P_d(\omega)$ 및 $P_u(\omega)$ 는 $-\pi \sim \pi$로 주어지는 디지틀 주파수 ω에서 $d(n)$ 및 $u(n)$ 의 전력 스펙트럼을 나타낸다. 즉, 3.4 절에서 언급한 것과 마찬가지로 잡음과 목적 음성 신호의 전력 스펙트럼만 알면 위너필터에 의한 잡음 제거를 수행할 수 있다. 수식 (13)에서 알 수 있듯이 위너필터는 신호대 잡음비가 큰 주파수 성분은 그대로 유지 시켜주고 반대로 낮은 주파수 성분은 감쇄를 시켜주는 특성을 보인다. 본 연구에서의 위너필터의 설계는 수식 (9)~(11) 과정에서 얻어지는 목적 음성 및 잡음의 전력 스펙트럼을 이용하여 수식 (13)에 대입하여 이루어진다. 그런 후에 역 이산 푸리에 변환을 취하여 시간 영역에서의 필터의 계수와 입력 채널간의 복적분을 수행하여 최종적으로 잡음이 제거된 목적 음성 신호를 추정한다.

4. 실험 및 결과

4.1 성능 지표 및 데이터 수집 조건

성능 지표로서 참고 문헌 [9]에서 제안한 객관적 음질 평가 도구의 일종인 PESQ (perceptual evaluation of speech quality) 와 음성 인식률을 채택하였다. 잡음 제거 후의 음질 향상 정도를 살펴보기 위하여 30개의 한국어 문장을 수집하여 목적 신호로 사용하였다. 각각의 길이는 10~12초 정도였다. 비정상성 잡음 신호로는 Beatles의 'Yesterday' 및 Westlife의 'Mandy'를 사용하였다. 모든 신호들은 고품질 오디오 시스템으로부터 재생되었다. 스테레오 녹음 장비로 SONY에서 제조한 Hi-MD(MZ-RH1)가 사용되었다. 신호의 이산화는 16 kHz 표본화율로 행해졌으며 각 샘플은 16 bit의 해상도를 가진다. 마이크간의 간격은 2 cm이었으며, 잡음원간의 각도는 90도로 두었다. 녹음은 일반적인 사무실에서 시행되었으며 음원 및 녹음 장비의 배치는 그림 4와 같다.

입력 신호대 잡음비에 따른 성능의 측정을 위해서 목적 음성 신호와 잡음원은 각기 녹음되었다. 잡음 신호는 무작위로 구간 선택이 되어 크기가 조절되고 목적 신호와 합해져서 원하는 신호대 잡음비를 갖는 테스트 데이터를 구성한다. 음성 인식률의 성능 측정을 위해서 연속 은닉 마르코프 모델 (continuous hidden Markov model) 기반의 인식기가 사용되었다. 음성 인식의 대상 어휘는 한국어 인명이며 어휘 크기는 500이었다. 1000개의 공유 상태를 근간으로 하는 트라이폰(triphone)기반의 HMM이 훈련되었으며 특징 벡터는 12차의 MFCC(mel-frequency cepstral coefficient) 과 단구간 로그 에너지 값 그리고 그것들의 미분 및 가속도 성분으로 구성된다. 특징벡터는 30 ms의 분석 프레임에서 10 ms 당 1회 추출된다. HMM 훈련 데이터 베이스의 신호대 잡음비는 25 dB 이상이었다. 목적 음성 신호로서 각 10명의 남녀가 2000개의 한국어 인명을 발성하였다. 잡음원 및 다른 녹음 조건은 음질 평가 실험의 것과 동일하였다.

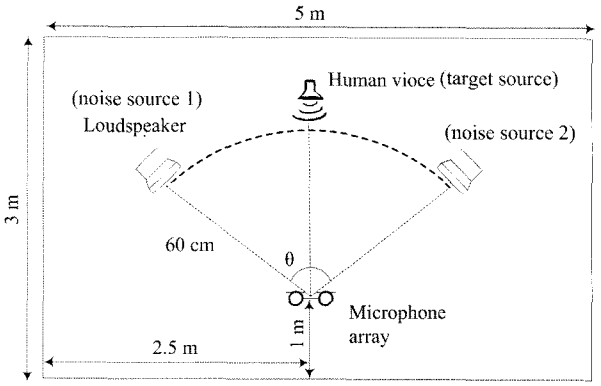


그림 4. 음원 및 녹음 장비의 배치
Fig. 4. Arrangement of sound sources and recording device

4.2 제안된 알고리즘 및 비교 대상 알고리즘의 사양

제안된 알고리즘의 성능은 참고 문헌 [6]에서 제안된 프로스트 기법과 참고 문헌 [4]에서 제안한 AMC-GSC 와 비교되었다. 제안된 알고리즘에서 스테레오 입력의 주파수 분석을 위해서 10ms 간격으로 각각 512-FFT 가 사용되었다. 분석 프레임의 길이 역시 512였다. 추정된 잡음 및 목적 음성 신호 스펙트럼의 스무딩을 위해서 사용된 정규화된 해밍 창의 길이는 10이 였으며, 시간 영역에서의 위너필터 길이는 129 였다. 수식 (9)의 스펙트럼 분류를 위한 r_{TH}는 실험적으로 최적화하였으며 0.3 두었다. AMC-GSC의 모든 적응 필터의 길이는 129이며 NMLS (normalized LMS)에 의해 학습이 이루어졌다. 적응 필터의 학습률은 0.1이었다. 목적 신호 차단 블록의 제약은 최대 목적 신호 추정 각도 오차가 20도인 조건에서 설계되었다. AMC의 수행은 참고 문헌 [1]의 VADNest 모듈이 사용되었으며 그것의 입력으로는 고정 빔포밍 후의 단구간 에너지 값이었다. 프로스트 알고리즘에서의 각 적응 필터의 길이는 129 였으며 필터 입력의 제곱 합으로 정규화되는 학습률은 0.01이었다.

4.3 실험 결과

그림 5에서 입력 신호대 잡음비가 5 dB일 때의 각 알고리즘에 의한 잡음 제거 결과를 나타내었다. 그림 (5)에서 알 수 있듯이 제안된 방식에 의해서 잡음이 획기적으로 감소함을 알 수 있다. AMC-GSC 방식은 프로스트 알고리즘보다 주변 잡음을 약간 더 제거할 수 있지만, 제안된 방식처럼 목적 음성 구간에서 그 궤적을 잘 살려줄 정도의 성능은 보이지 못하였다. 그 이유는 2.2 절에서 언급한 것 처럼 AMC에 의해서 음성 구간일 경우에는 MC 모듈의 적응 필터가 적응 과정 없이 단순이 필터링만 수행하기 때문이다.

그림 6에서 각 알고리즘 및 입력 신호대 잡음비에 따른 PESQ 스코어를 나타내었다. 제안한 알고리즘이 전반적으로 가장 좋은 성능을 보였다. 입력 신호대 잡음비가 10dB 일 때, PESQ 스코어는 제안된 알고리즘이 AMC-GSC 및 프로스트 알고리즘에 비해서 각각 0.32 및 0.23 높음을 알 수 있었다. 입력 신호대 잡음비가 낮아짐에 따라서 제안된 알고리즘에 의한 성능 향상은 더욱 눈에 띄었다. AMC-GSC 알고리즘은 가장 나쁜 성능을 보였다. 그 이유는 2.2에서 언급한 AMC의 불안정성 및 MC 블록의 부최적성에 있다고 볼 수 있다. 프로스트 알고리즘은 AMC-GSC에 비해서 약간 좋은 성능을 보였다. 그렇지만, 제안된 알고리즘에는 필적하지 못

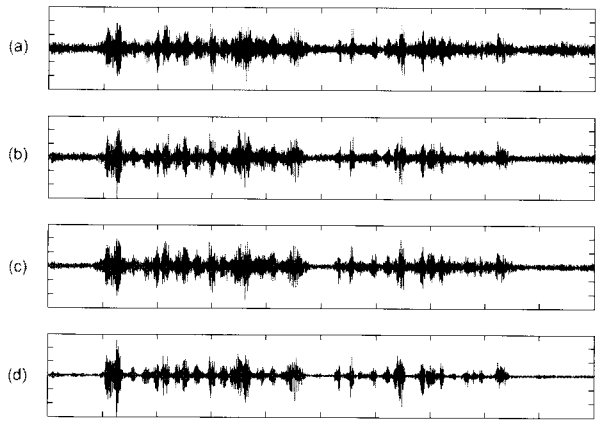


그림 5. 알고리즘별 잡음 제거 결과 (수평축 1칸이 1초를 나타냄. (a) 입력, (b) 프로스트 알고리즘, (c) AMC-GSC 알고리즘, (d) 제안한 알고리즘)

Fig. 5. Noie reduction results of each algorithm (one second per tick in the horizontal axis. (a) input, (b) Frost algorithm, (c)AMC-GSC algorithm, (d) proposed algorithm)

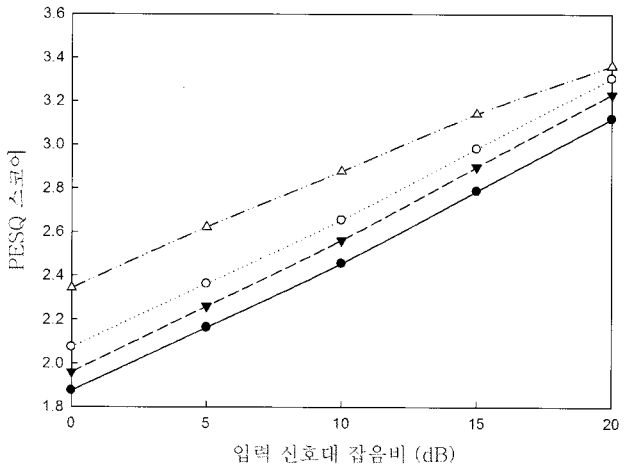


그림 6. 입력 신호대 잡음비에 따른 각 알고리즘의 PESQ 스코어 ('●': 입력, '▼': AMC-GSC, '○': 프로스트 알고리즘, '△': 제안된 알고리즘)

Fig. 6. PESQ score curves of each algorithm according to input SNRs ('●': input, '▼': AMC-GSC, '○': Frost algorithm, '△': proposed algorithm)

하였다. 프로스트 알고리즘은 마이크로폰의 정면에서 들어오는 신호원에 대해서는 절대로 왜곡을 시키지 않는다는 아주 강력한 제약 조건이 때문에 잡음의 제거에 효율적이지 못함을 추측할 수 있었다.

그림 7에서는 음성인식 실험 결과를 나타내었다. PESQ에 의한 음질 평가 실험에서와 마찬가지로 제안된 알고리즘이 모든 입력 신호대 잡음비에 대해서 가장 좋은 성능을 나타내었다. 그리고 각 알고리즘의 성능 추세 역시 비슷하였다. 잡음이 없는 상황에서 음성 인식률은 95.5 %였다. 입력 신호대 잡음비가 10dB일 때, 제안된 알고리즘은 AMC-GSC 보다 20.1 %, 프로스트 기법보다 18.4 % 높은 성능을 보였다. 음성 인식률의 측정에서는 목적 음성 신호의 검출 오류에 의한 성능 저하를 고려하지 않기 위해서 수작업으로 검출된 구간을 인식기에 입력하였다. 입력 신호대 잡음비가 10dB일 때에

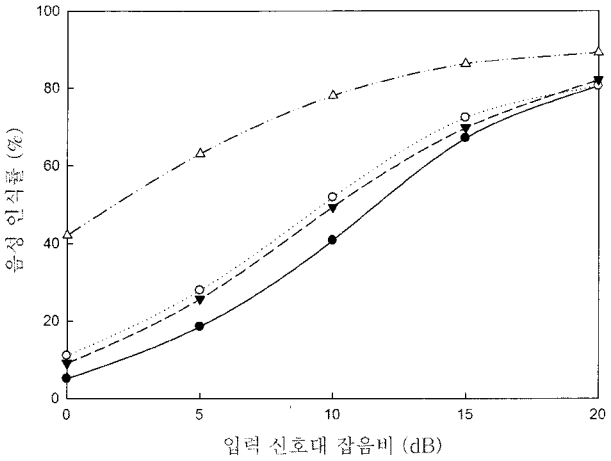


그림 7. 입력 신호대 잡음비에 따른 각 알고리즘의 음성 인식 성능

Fig. 7. Speech recognition performances of each algorithm according to input SNRs

제안된 방식에 의한 인식률은 84.7 % 인데, HMM 훈련 과정에 사용되었던 데이터와 테스트 과정에서 사용된 데이터 사이의 정합을 이루기 위한 노력이 부가 될 경우에는 90 % 이상의 인식률을 얻을 수 있을 것으로 판단된다.

5. 결 론

본 논문에서는 스테레오 마이크로폰을 이용한 비정상성 잡음 제거 방법에 대한 새로운 알고리즘을 제안하였다. 제안된 알고리즘은 주파수 분류를 통해서 잡음과 목적 음성 신호의 주파수 성분을 추정하고 시간 영역 위너필터를 이용하여 잡음 제거를 수행한다. 실험 결과의 출력 파형을 통해서 제안한 방식이 비음성 구간에서 잡음을 크게 감소시킬 뿐 아니라 목적 음성 구간에서도 효과적인 잡음제거를 수행하고 있음을 알 수 있었다. 객관적 음질 평가와 음성 인식률 실험 결과 기존의 빔포밍 기법과 비교하였을 때 제안된 방식이 가장 좋은 성능을 나타내었다. 제안된 알고리즘은 결과적인 측면에서는 마이크로폰을 두 개 이상 장착하기 어려운 이동형 단말에서 상용화가 가능할 정도의 성능을 보였다.

향후 연구로는 주행 차량 환경에서 네비게이션용 대어휘 음성인식기의 선처리기로 사용되었을 때의 인식 성능 측정을 생각할 수 있다. 또한 제안된 잡음 제거 알고리즘의 소프트웨어를 고정 소수점 연산 기반으로 구현한 후에 PDA 혹은 휴대 단말기에서의 계산량 검증이 필요하다고 판단된다.

참 고 문 헌

[1] ETSI Std. ES 202 212 V1.1.2, "Speech processing, transmission and quality aspects (STQ)", (ETSI, 2005)

[2] S. Jeong and M. Hahn, "Speech quality and recognition rate improvement in car noise environments", *Electronics Letters*, Vol. 37, No. 2, pp. 800-802, 2001.

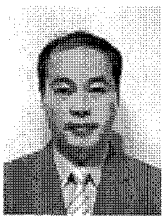
[3] B. Veen and K. Buckley, "Beamforming: a versa-

tile approach to spatial filtering", *IEEE ASSP Magazine*, pp. 4-24, Apr. 1988.
[4] Hoshuyama, O., Sugiyama. A., and Hirano. A., "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters", *IEEE Trans. Signal Proc.*, Vol. 47, No. 10, pp. 2677-2684, 1999.
[5] Y. Jung, H. Kang, C. Lee, D. Youn, C. Choi, and J. Kim, "Adaptive microphone array system with two-stage adaptation mode controller", *IEICE Trans. Fundamentals*, Vol. E88-A, No. 4, Apr. 2005.
[6] O. Frost, "An algorithm for linearly constrained adaptive array processing", *Proc. of the IEEE*, Vol. 60, No. 8, pp. 926-935, 1972.
[7] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking", *IEEE Trans. Signal. Proc.*, Vol. 52, No. 7, pp. 1830-1847, 2004.
[8] S. Werner, J. Apolinario, M. Campos, and P. Diniz, "Low-complexity constrained affine-projection algorithms", *IEEE Trans. Signal Proc.*, Vol. 53, No. 12, pp. 4545-4555, Dec. 2005.
[9] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech coders", Feb. 2001.
[10] M. Hayes, "*Statistical digital signal processing and modeling*", John Wiley & Sons, INC., 1996.

저 자 소 개

정상배(Sangbae Jeong)
1997년 : 부산대학교(공학사)
1999년 : 한국과학기술원(공학석사)
2002년 : 한국정보통신대학교 (공학박사)
2002년~2006년 : 삼성종합기술원 (전문연구원)
2006년~현재 : 한국정보통신대학교 공학부(연구조교수)

관심분야 : 음성개선, 빔포밍, 음성인식, 음성부호화
Phone : 042-866-6285
E-mail : sangbae@icu.ac.kr



이성독(Sungdoke Lee)
1988년 : 전북대학교(공학사)
1991년 : 전북대학교(공학석사)
2002년 : 일본 동북대학교 (정보과학박사)
1991.5~1993.7 : 군산대학교 전기공학과 조교

2002.4~2003.3 : 일본 동북대학교 전기통신연구소 연구원
2003.8~현재 : 한국정보통신대학교 공학부 연구조교수

관심분야 : 지식표현, 에이전트공학, 멀티미디어시스템, 음성신호처리
Phone : 042-866-6285
E-mail : sdlee@icu.ac.kr