

독립 성분 분석과 지각 필터를 이용한 음질 개선

Performance Improvement of Speech Enhancement Using Independent Component Analysis and Perceptual Filtering

구 교 식*, 차 형 태*
(Kyo-Sik Koo*, Hyung-Tai Cha*)

*송실대학교 전자공학과
(접수일자: 2010년 3월 15일; 채택일자: 2010년 4월 17일)

본 논문에서는 독립 성분 분석 기법과 적응 지각 필터를 이용하여 잡음에 오염된 신호에서 잡음을 추정 및 제거함으로써 보다 효율적으로 오디오 신호를 개선하는 방법에 대해 제안한다. 기존의 적응 지각 필터는 묵음 구간에서 추정된 고정적인 추정 잡음을 사용하기 때문에 잡음이 시간에 따라 가변적일 경우 성능이 상당히 저하된다. 그러나 제안된 알고리즘은 독립 성분 분석 기법을 이용하여 신호 구간마다 달라지는 잡음을 추정한 후 이를 적응 지각 필터에 적용시켜 준다. 제안된 알고리즘의 성능 분석을 위하여 기존 및 제안된 알고리즘을 이용한 시뮬레이션을 수행하였으며, 그 결과로 제안된 알고리즘은 기존의 알고리즘에 비해 향상된 결과를 생성하였다. 더불어 음질의 열화도 적응을 확인할 수 있었다.

핵심어: 잡음 제거, 지각필터, 심리음향, 독립성분분석

투고분야: 음향 신호처리 분야 (1.4)

In this paper, we proposed an algorithm that improves tone quality of noisy audio signals by using ICA(Independent Component Analysis) algorithm and perceptual filters. Many algorithms have been proposed to eliminate the noise from the audio signals, such as spectral subtraction method, perceptual filter, etc. The perceptual filter uses a noise that is acquired from silent ranges in the input signal. In this case, the improvement rate of tone quality decreases if the noise energy is changed by the environmental variation in a signal frame. But the proposed method estimates a noise that is changed at each frame using ICA algorithm. The estimated noise is applied to perceptual filter. To show the performance of the proposed algorithm, several tests are performed to various input signals. With the proposed algorithm, we could confirm the enhancement of tone quality in terms of segmental SNR (SSNR), noise-to-mask ratio (NMR) and Degradation Category Rating (DCR) test.

Keywords: Noise Cancellation, Perceptual Filter, Psychoacoustics Theory, Independent Component Analysis

ASK subject classification: Acoustic Signal Processing (1.4)

I. 서론

통신, 나날이 발전하는 기술로 인하여 디지털 시대가 도래한 현재, 멀티미디어 기술은 우리 사회의 커다란 이슈로 자리매김 하였다. 이 중에서도 저장용량의 확대, 전송망의 개선, 처리속도의 증가로 인하여 고음질에 대한 청자들의 요구가 급증하게 되었다. 그 중 무선 통신 기술의 발달로 이동전화기가 널리 보급되고 원격 화상 회의 등 스피커-폰 시스템의 사용이 증가함에 따라 잡

음제거와 관련된 기술이 더욱더 중요한 문제로 대두되게 되었다.

잡음은 음성 신호의 명료성을 떨어뜨리고 청자에게 불쾌감이나 귀의 피로감을 안겨준다. 그 결과로 정상적인 통화나 청취를 방해하고 음성인식 시스템의 성능을 떨어뜨리기 때문에 잡음은 원활한 의사전달을 위하여 반드시 제거 되어야 한다. 이에 잡음 제거를 통한 높은 SNR 개선과 더불어 신호의 지각적 측면에서도 음성 신호를 개선하기 위한 다양한 방법이 제안되어 왔다 [1]. 또한 인간의 청각 특성을 모델링하여 시간-주파수 영역의 특성을 이용한 방법 [2][3], 청각 모델 기반의 마스킹 특성을 이용한 방법 [4] 등이 있다. 하지만 위의 방법들은 잡음을 제거

하기 위해서 잡음에 대한 통계적인 정보가 필요하고, 만일 그 정보를 알 수 없을 때는 보통 묵음 구간을 통해 잡음을 추정한다 [1]. 하지만 불특정한 상황에서 잡음 신호의 특성은 항상 다르며, 시간에 따라 변화한다. 따라서 효과적으로 잡음을 추정하여 제거하는 연구가 필요하다. 이런 가변적인 상황에서 인간의 지각적인 청각 특성을 이용하여 적응적으로 잡음을 제거하는 연구는 계산량이 적으면서 성능이 우수하여 지속적인 관심의 대상이 되어 왔다. 본 논문에서는 ICA 알고리즘을 이용하여 미지의 잡음 환경 하에서 효율적으로 잡음을 추정한 후 적응지각필터를 이용하여 제거하는 방법을 제안한다. 논문의 내용은 2장에서는 ICA 알고리즘과 지각필터에 대해 설명하고, 3장 및 4장에서 제안된 알고리즘에 대한 설명 및 시뮬레이션에 관한 결과를 제시한다. 마지막으로 5장에서 결론을 맺는다.

II. 잡음제거를 위한 독립성분분석 및 지각필터

2.1. 독립성분분석

(Independent Component Analysis)

독립성분분석은 통계적인 기법으로 독립된 정보와 다른 정보들 사이의 상관관계를 변환하는 기술이다 [5]. 즉 특징이 상이한 둘 이상의 신호들이 선형적으로 혼합되어 구성된 확률 변수 (random variable)들을 통계적인 방법에 따라 서로 독립적인 신호들로 분리하는 것을 통칭한다.

독립성분분석에서 신호원 벡터 (source signal vector)는 dimension이 M인 서로 독립적인 신호, s 로 가정하자. 이 때 우리가 측정하는 신호는 식 (2)와 같이 dimension이 N인 mixed signal, x 로 주어진다 할 수 있다.

$$s = [s_1, s_2, \dots, s_M]^T \tag{1}$$

$$x = As \tag{2}$$

이때 A는 혼합행렬 (mixing matrix)이라고 한다. 즉 여러 신호원들이 합쳐져서 x 로 측정이 되는데, 그 합쳐지는 특성이 혼합행렬 A에 나타난 형태이다. 이런 가정 하에 신호원 s 에 관한 정보가 전혀 없는 상황에서 측정된 신호 x 로부터 다시 신호원 벡터 s 를 추출할 수 있다. 이때 측정된 벡터에 대한 상호정보량은 식 (3)과 같이 KL

(Kullback-Leibler) 조건에 의해서 계산되어진다 [6].

$$I(x) = \int p(x) \log \frac{p(x)}{\prod_{i=1}^N p_i(x_i)} dx \tag{3}$$

여기서 $p(x)$ 와 $p_i(x_i)$ 는 각각 다변량 확률밀도함수 (probability density function)와 주변 독립 분포 (marginal independent distribution)이다.

상호정보량은 항상 양의 값을 가지며 각 신호의 구성요소들이 독립적일 때 0과 같다. 즉 독립성분분석은 출력되는 (추정된 원신호) 신호의 상호정보량이 가능하면 독립적으로 출력되는 선형 변환 W 를 구하는 것이다.

$$u = Wx = WAs \tag{4}$$

여기서 u 는 원신호 s 에 대한 추정된 신호이며 독립성분 분석의 출력이다. 그리고 W 는 혼합행렬 A와 역행렬 관계가 된다. 그림 1은 독립성분분석을 이용하여 혼합된 신호 원으로부터 독립된 신호원을 구하는 간단한 블록도이다.

2.2. 지각필터 (Perceptual filter)

잡음 제거는 신호처리 분야에서 가장 기본적인 목적 중의 하나로 음향 신호가 수신될 때 원하지 않는 신호가 같이 혼합되어 수신될 경우 이를 제거하고 원 신호를 추출해 내는 것을 말한다. 이 경우 관측된 신호에 어떠한 처리를 하여 그 중에서 필요한 성분만 추출해 내어야 하는데 일반적으로 사람의 청각적 특성 및 신호들의 주파수적인 특성을 이용하는 주파수 차감 기법을 기반으로 하는 연구가 활발히 이루어지고 있다.

인간의 청각적인 특성을 이용하여 각 임계 대역 (Critical Band)에서 주파수 변별력을 고려하여 각 대역에 해당하는 신호들에 대한 임계 대역 에너지와 마스크 특성을 고려한 청각 자극 에너지를 계산할 수 있다. 이를

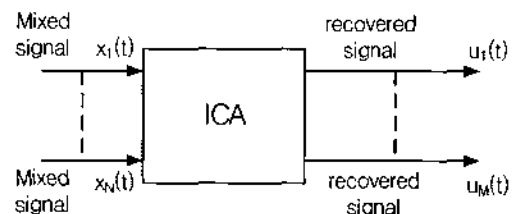


그림 1. ICA의 Block diagram
Fig. 1. Block diagram of ICA.

바탕으로 잡음의 열화도가 심하게 나타나는 고주파 영역에서는 신호의 에너지보다 잡음의 에너지가 더 크게 나타난다는 특성을 이용하여 잡음을 추정하는 필터를 설계할 수 있다 [7].

원래의 오디오 신호와 잡음 신호를 $s(n)$, $d(n)$ 이라 할 때, 잡음에 열화 된 신호 $x(n)$ 의 표현은 식 (5)와 같다.

$$x(n) = s(n) + d(n) \quad (5)$$

이 때 어떤 임계 대역에 대한 $s(n)$ 과 $x(n)$ 의 전력 스펙트럼을 $S(\omega, i)$, $X(\omega, i)$ 라고 정의할 때, 신호와 잡음이 서로 상관관계가 없다고 한다면 신호와 잡음의 자극 에너지는 식 (6)과 같은 관계를 가진다.

$$X(\omega, i) = S(\omega, i) + D(\omega, i) \quad (6)$$

이 때 i 는 시간 프레임에서의 인덱스를 의미하며 $S(\omega, i)$ 로 인한 기저막 (basilar membrane)에서의 임계 대역 에너지는 식 (7)로 표현 가능하다.

$$S_a(z, i) = a_0(z)S_i(z, i) = a_0(z) \sum_{\omega_i}^{\omega_{2i}} S(\omega, i) \quad (7)$$

where, $0 \leq z \leq Z-1$

여기서 ω_{2i} 와 ω_{2i} 는 Z 개의 전체 임계 대역 중 임의의 z 번째 임계대역의 주파수 경계를 나타낸다. $a_0(z)$ 는 외이 (외이)에서부터 중이 (중이)까지의 다양한 전달 특성 요소 (transmission factor)들에 의한 주파수 의존 감쇠 특성을 나타내는 감쇠 성분이다. 단, 이때 $a_0(z)$ 는 외이부터 중이까지의 감쇠 특성을 나타낸 것이므로, 신호를 직접 내이 (내이)로 전달하는 경우 또는 난청 환경과 같이 이러한 특성의 변화가 있을 경우를 제외하고는 $a_0(z)$ 를 1로 값을 설정하여 감쇠에 따른 영향을 무시할 수 있다.

더불어 임계 대역에 대한 청각 자극 에너지는 $S_a(z, i)$ 과 기저막에서의 신호 에너지의 확산 현상을 나타내는 지각적인 에너지 확산 함수 (spreading function)와의 컨볼루션 (convolution)을 통하여 기저막에서의 신호 에너지에 대한 응답을 나타낼 수 있다. 이는 식 (8)로서 나타낼 수 있다 [1].

$$S_c(z, i) = \sum_{v=0}^{Z-1} [SF(v, z)S_a(v, i)] \quad (8)$$

where, $0 \leq z \leq Z-1$

여기서, 확산 함수 $SF(v, z)$ 는 다음과 같이 표현할 수 있다.

$$SF(v, z) = \frac{15.81 + 7.5(\Delta z + 0.474)}{-17.5\sqrt{1 + (\Delta z + 0.474)^2}} \quad (9)$$

여기서, $\Delta z = z - v$ 는 바크 인덱스 (Bark index)로 표현된 임계 대역 간 차이 값이며 청각 자극에너지는 임계 대역 에너지보다 변화가 급격하지 않다.

이런 특성을 이용하여 잡음에 열화 된 신호에서 잡음을 추정할 후 이를 통해 신호를 개선시키는 방법들이 많이 제안되고 있는데, 특히 신호의 자극 에너지를 이용하는 주파수 차감 방법이 그 대표적인 예이다. 주파수 차감 방식은 추정된 잡음의 주파수 스펙트럼을 잡음에 열화 된 주파수 스펙트럼에서 차감하는 방식으로 음질 개선 지각 필터 $H_c(z, i)$ 를 이용하여 식 (10)과 같이 나타낼 수 있다 [4].

$$\begin{aligned} \hat{S}_c(z, i) &= H_c(z, i)X_i(z, i) \\ &= \left(1 - \frac{\hat{D}_c(z, i)}{X_c(z, i)}\right)X_i(z, i) \\ &\text{for } 0 \leq z \leq Z-1 \end{aligned} \quad (10)$$

여기서 $\hat{D}_c(z, i)$ 는 추정된 잡음의 청각 자극 에너지로서 추정된 잡음이 잡음에 의해 열화 된 신호보다 크지 않은 경우에만 고려하기로 한다. 식 (10)을 볼 때, 추정 잡음의 오차에 따라서 필터의 성능에 많은 차이가 생김을 알 수 있으며 만약에 인지 가능한 잡음의 에너지를 $H_c(z, i)$ 를 이용하여 $X(z, i)$ 로부터 제거했다면 잡음은 더 이상 청각의 청각기관에 자극을 줄 수 없게 된다. 따라서 식 (11)과 같이 개선된 신호의 자극 에너지는 순수 신호의 자극 에너지와 같게 될 것이다 [1].

$$\hat{S}_c(z, i) = S_c(z, i) \quad (11)$$

III. 제안된 알고리즘

서론에서 설명한 바와 같이 잡음을 제거하기 위해 상당히 많은 알고리즘이 제안되었는데 그 중 일반적인 방법으로 지각필터를 이용한 방법을 들 수 있다. 지각필터 알고리즘은 배경 잡음에 의해 손상된 음성의 명료도 향상을 위한 잡음제거 기법으로 입력되는 신호에 따라 잡음을

추정하고 그에 따라 적절히 필터가 동작되도록 하는 처리 기법이다.

그림 2는 기존의 지각 필터의 모델링을 보여 주고 있다. $s[n]$ 은 사운드 신호이며 마이크를 통해 수신되는 신호인 $x[n]$ 은 $s[n]$ 및 $s[n]$ 과 함께 수신되는 잡음 신호인 $d[n]$ 의 합으로 정의된다. $d_{est}[n]$ 은 묵음 구간에서 추정된 잡음 신호로서 지각 필터에 적용된다. 최종적으로 잡음이 개선된 신호는 $y[n]$ 으로 나타낸다.

임의의 시간 프레임 인덱스 i 에서 $x[n]$ 과 $d_{est}[n]$ 의 전력 스펙트럼을 $X(\omega, i)$, $D_{est}(\omega, i)$ 이라 할 때, 일반적으로 기존 지각 필터에 적용되는 잡음은 식 (12)와 같이 잡음이 첨가된 신호의 묵음 구간에서 새 프레임의 평균값을 잡음으로 추정하였다.

$$D_{est}(\omega, i) = \frac{1}{3} \sum_{k=-1}^1 X(\omega, i+k) \quad (12)$$

그러나 식 (12)에서 구한 잡음은 모두 다른 SNR 값을 가지고 있는 각 프레임마다 동일한 값으로 적용되므로 최적화된 성능을 내기는 어렵다. 이에 [1]에서는 각 프레임의 상황에 따라 개별적으로 추정된 잡음을 보정할 수 있는 알고리즘을 제안하였다. 그러나 제안된 알고리즘도 잡음이 시간에 따라서 거의 변함이 없는 stationary적인 특성을 가지고 있을 때 제대로 된 효과를 발휘할 수 있다. 따라서 배경 잡음의 변화 정도가 클 경우 지각 필터의 성능에 영향을 끼쳐 효율적인 지각 필터를 구현할 수 없다. 그 결과로 잡음이 완전히 제거되지 못하고 잔여 잡음이 남게 되어 음질이 훼손되는 결과를 가져온다.

이에 본 논문에서는 이러한 고정적인 추정 잡음을 사용하지 않고 독립성분분석 알고리즘을 적용하여 효율적으로 잡음을 추정하여 지각 필터의 성능을 개선시키는 알고리즘을 제안한다. 독립성분분석은 두 개 이상의 다수의 센서로부터 얻어진 혼합 신호 (mixture)를 독립적인 (independent) 성분들로 분리해 내는 통계적인 방법이다. 이는 고정적인 신호뿐만 아니라 변화하는 신호들에 대해서도 적용할 수 있어 가변적인 잡음 환경에서 효과적인

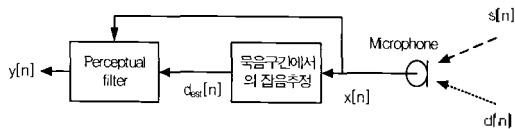


그림 2. 일반적인 지각 필터의 모델링
Fig. 2. Modeling of exited perceptual filter.

으로 잡음을 분리해 낼 수 있다는 장점을 가지고 있다. 따라서 원 잡음에 최대한 근접하게 잡음을 추정할 수 있다면 지각 필터를 이용하여 최적화된 원 신호를 추출해 낼 수 있다는 결론에 도달하게 된다.

그림 3은 본 논문에서 사용된 지각 필터의 모델링을 보여준다.

사용된 지각 필터는 그림 2의 모델과는 다르게 두 개의 마이크로폰을 사용하고 있다. 두 개의 마이크로폰은 각각 선형으로 섞여서 유입되는 사운드 신호와 잡음을 받아 들이게 되며 이를 $x_1[n]$, $x_2[n]$ 이라 정의한다. 그리고 받아들인 신호를 독립 성분 분석 알고리즘을 이용하여 잡음 $d_{ICA}[n]$ 을 추정한 후 지각 필터에 적용하게 된다. 추정된 잡음은 원 잡음 신호에 거의 근접하므로 여러 다양한 잡음 하에서도 쉽게 지각 필터의 성능을 향상시킬 수 있다.

그림 3의 모델에서 신호와 잡음은 식 (13)과 같이 계수 (rank)가 2인 미지의 혼합행렬, A 를 정의할 때, 그림 4와 같이 혼합행렬 A 에 의해 선형으로 섞이는 것으로 가정할 수 있다.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (13)$$

제안된 알고리즘에서는 독립성분분석 알고리즘을 이용하여 마이크로폰에 수신된 두 신호에서 잡음만을 추정하도록 시도하게 된다. 이를 위해서 독립성분모델을 이용하여 분리행렬 W 를 추정하는 작업이 필요하다.

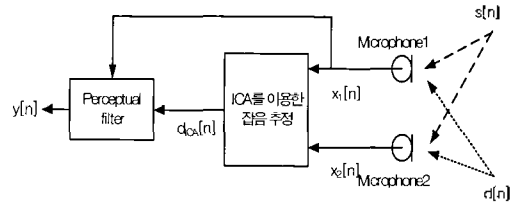


그림 3. 제안된 지각 필터의 모델링
Fig. 3. Modeling of proposed perceptual filter.

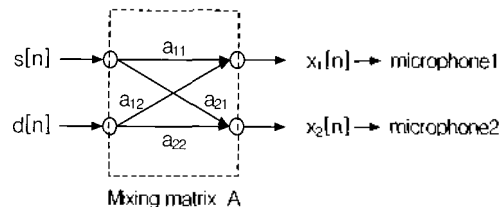


그림 4. 신호 혼합 모델
Fig. 4. Signal mixing model.

분리 행렬 W 를 결정하는 독립성분분석 알고리즘으로 는 여러 가지 알고리즘들이 있으나 본 연구에서는 $u[n]$ 의 비가우시안성 (nongaussianity)이 최대로 되도록 하는 고속독립성분분석 (Fast ICA) 알고리즘을 사용하였다.

Hyvarinen과 Oja에 의해 제안된 Fast ICA는 kurtosis의 극점을 찾는 것이 비가우시안 (non-Gaussian) 독립 요소를 찾아내는 것과 동일하다는 사실에 근거하여 출력신호 $u=wx$ 의 kurtosis를 극대 혹은 극소화 시키는 방법이다 [11][12]. 여기서 w 는 분리행렬 W 의 한 열 벡터 (row vector)를 나타낸다. 이 방법에서 학습의 기준으로 사용한 오차함수는 식 (14)와 같으며 F 는 제약조건 $\|w\|^2 = 1$ 을 위한 항이다.

$$J(w) = \langle (wx)^4 \rangle - 3\|w\|^4 + F(\|w\|^2) \quad (14)$$

$$w^+ = \langle x(w(p)x^3 \rangle - 3w(p) \quad (15)$$

오차함수로부터 유도된 간단한 학습 방법은 p 가 학습 횟수 (iteration)를 나타낸다고 할 때, 식 (15)를 계산한 후 $w(p+1) = w^+ / \|w^+\|$ 에 따라 가중치 벡터를 변경하는 것이다. 위와 같이 w 의 학습에 의해 신호원에 해당하는 신호가 찾아지면, 이 신호를 x 에서 제거한 후 또 다른 신호원에 해당하는 신호를 같은 방식으로 찾아낸다.

이어서 독립성분분석을 사용하여 2개로 분리된 신호에서 잡음을 선택하기 위한 과정을 거친다. 음성인식과 관련된 기기에서 일반적으로 사람의 목소리는 배경 잡음보다 가까운 거리에서 존재한다. 따라서 gun 잡음과 같은 특정한 성격의 잡음이 존재하지 않는다고 가정할 때, 일반 생활 잡음 환경에서 배경 잡음보다 사람의 목소리가 크기가 큰 특성을 이용하여 2개의 신호 중에서 작은 신호를 잡음으로 선택하게 된다. 식 (16)에서 $\max(\cdot)$ 은 각각의 분리된 신호에서 크기가 가장 큰 sample 값을 나타낸다.

$$d_{ICA} = \begin{cases} u_1, & o_1 \leq o_2 \\ u_2, & o_1 > o_2 \end{cases} \quad (16)$$

where, $o_j = \max(u_j), j = 1, 2$

기존의 연구 결과에서 확인할 수 있듯이 ICA 알고리즘 만으로도 상당량의 잡음을 제거할 수 있다 [8][9]. 그러나 실제로 완벽하게 독립적인 신호들은 존재하지 않으며 인간의 신체 특성이나 장애물 등에 의해 신호가 변형될 수도 있다. 따라서 실제 환경에서는 ICA에 의해 분리된 사운드 신호는 일정량의 잡음을 포함하게 되므로 지각필터

를 이용하여 잡음을 제거해 주는 과정이 필요하다.

그림 3의 모델에서 확인할 수 있듯이 독립성분분석 방법으로 분리된 잡음은 지각필터에 추정된 잡음으로 적용된다. 먼저 식 (7)과 (8)을 이용하여 추정 잡음 $d_{ICA}[n]$ 과 잡음에 의해 열화된 신호, $x_i[n]$ 의 자극 에너지를 구하고, 이를 $D_{ICA,i}(z, i), X_{i,ICA,i}(z, i)$ 로 정의한다.

이런 자극 에너지를 식 (10)에서 주어진 지각 필터에 적용함으로써, 그림 5와 같은 필터 응답 및 개선된 신호의 자극 에너지를 구할 수 있다.

그림 5는 추정된 잡음을 적응지각필터에 적용 하였을 때의 결과 그림이다. 13 bark 대역까지는 신호의 주파수 에너지가 잡음보다 더 크기 때문에 인간의 청각 기관에 영향을 미치지 못한다. 따라서 해당 대역의 청각자극 에너지는 원 신호의 청각자극 에너지라 할 수 있기 때문에 필터 응답은 13 bark 대역까지는 1과 가까운 값으로 나타내게 된다. 그러나 14 이상의 bark 대역에서는 잡음의 청각자극 에너지가 원 신호보다 더 크기 때문에 각 bark 대역의 잡음 에너지 정도에 따라서 필터 응답이 달라진다. 이렇게 구현된 필터를 잡음에 열화 된 신호에 적용시킴으로서 잡음이 청각기관에 미치는 영향을 조절한다. 그 결과로 잡음의 지각적인 영향이 제거됨으로서 청자는 잡음을 느끼지 못하게 되는 것이다. 이렇게 개선된 결과는 첨가된 잡음을 그대로 사용한 최적화 필터 (optimum)와 비교하여 거의 같은 결과를 나타냄을 알 수 있다.

더불어 개선된 신호의 자극 에너지를 주파수 영역에서의 전력 스펙트럼 결과로 원 신호의 전력 스펙트럼과 유사함을 그림 6에서 다시 확인할 수 있다.

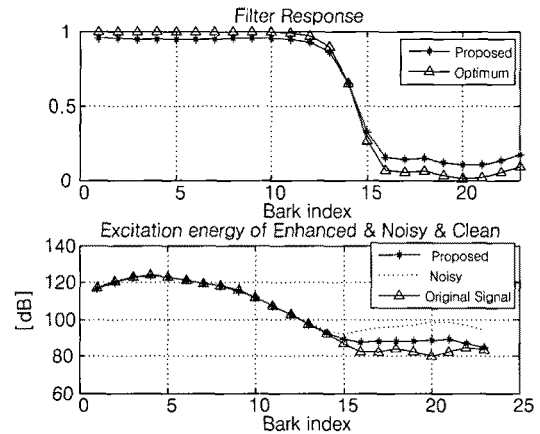


그림 5. 개선 신호의 필터 응답(상) 및 자극 에너지(하)
Fig. 5. Filter response (top) and excitation energy (bottom) of enhanced signal.

IV. 실험 및 결과

본 논문은 독립성분분석과 지각필터를 이용하여 보다 잡음에 열화된 신호에서 효율적으로 잡음을 제거하기 위한 알고리즘을 제안한다. 제안된 알고리즘의 성능을 확인하기 위하여 주관적 및 객관적인 테스트를 수행하였다. 테스트에 사용된 사운드 및 잡음 신호로는 표 1과 같이 오디오 CD 및 영화 DVD에서 추출된 Vocal 신호 및 비 정적인 특성을 갖는 Rock, Water, Siren 등의 3가지 음원을 사용하였다. 그리고 해닝 윈도우 (Hanning window) 함수와 중첩 가산 (Overlap add) 방식을 주파수 축 변환에 사용하였으며 실험은 Matlab을 이용하였다. 입력되는 열화된 신호의 음질 개선 결과를 비교하기 위하여 기존의 ICA만을 사용하여 잡음을 분리하는 경우 및 주파수 차감 방식을 추가한 경우와의 결과를 비교하였다.

4.1. 객관적 시뮬레이션 및 테스트

임의의 혼합행렬, A를 임의로 식 (17)과 같이 정의할 때, 첫 번째 시뮬레이션으로 원 신호와 잡음에 대하여 A를 이용하여 약 선형적으로 혼합함으로써 사운드 신호를

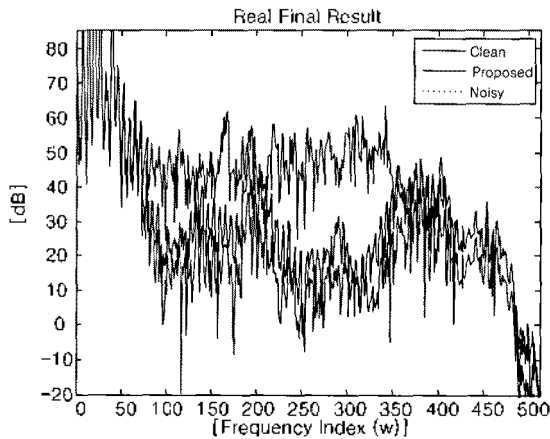


그림 6. 개선된 신호의 전력 스펙트럼
Fig. 6. Power spectrum of enhanced signal.

표 1. 시뮬레이션에 사용된 사운드 및 잡음 신호
Table. 1. Sound and noise signals for simulations.

Signal no.	Sound	Noise		
		1	2	3
이름	Vocal	Rock	Water	Siren
샘플 수	80000			
샘플링 주파수	22 kHz			
샘플당 비트수	16 bits/sample			

약 8 dB의 SNR 정도로 열화시킨 후 잡음제거 성능을 확인하였다.

$$A = \begin{bmatrix} 0.6 & 0.6 \\ 0.5 & 0.9 \end{bmatrix} \tag{17}$$

그림 7은 실험에 사용된 원 신호 (상) 및 마이크를 통해 수신된 신호 (하)를 표현하며, 그림 8은 각각의 방법을 통하여 개선된 신호를 나타낸 것이다. ICA만을 적용하거나 ICA에 주파수 차감법을 추가한 경우를 보면 신호 구간에서는 SNR이 크게 향상되었다. 그러나 비신호 구간에서는 아직도 잡음의 잔재가 남아있다. 그러나 제안된 알고리즘을 이용한 경우에는 그림의 파형으로부터 예측할 수 있듯이 상당한 양의 인식이 개선되었다. 더불어 원

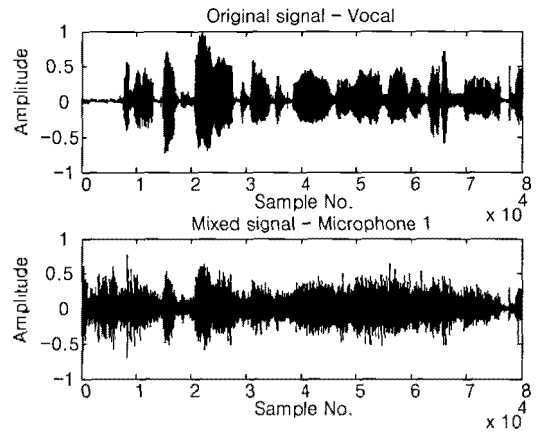


그림 7. 원 신호 - Vocal (상) 및 마이크론 1을 통해 수신된 신호 (하)
Fig. 7. Original signal - Vocal (top) and Mixed signal - Microphone 1 (bottom).

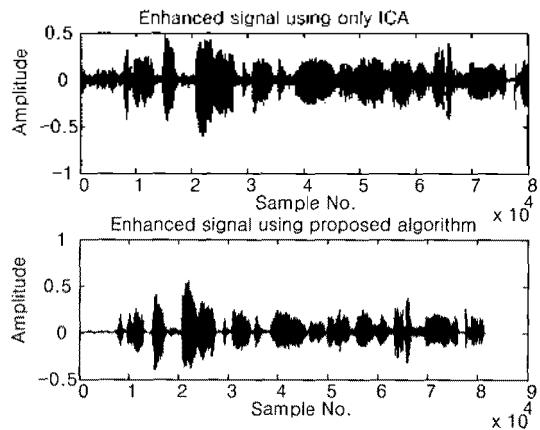


그림 8. 개선된 신호 - ICA만을 적용 (상) 및 제안된 알고리즘을 적용 (하)
Fig. 8. Enhanced signal - Only ICA (top) and Proposed algorithm (bottom).

음성 신호와 거의 유사하게 잡음이 제거되었음을 알 수 있었다.

더불어 각 방법으로 얻은 개선 결과에 대한 비교로 객관적 측정 (Objective Measurement) 결과로 사용되는 프레임 단위 신호 대 잡음 비 (Segmental Signal-to-Noise Ratio, SSNR)와 잡음 대 마스크 비 (Noise-to-Mask Ratio, NMR)를 측정하였다.

표 2와 3은 음성 신호를 여러 잡음으로 연화시킨 후 각 방법을 적용하여 얻은 개선량을 나타내고 있다. 결과에서 볼 수 있듯이 SSNR의 결과는 제안된 방식으로 개선된 신호는 기존의 방법의 경우보다 평균 2 dB 정도 향상되었다. 그러나 SSNR 개선량 중 사이렌 신호의 경우에는 개선량의 차가 매우 큰 결과를 얻었다. 이는 사이렌 신호가 다른 장르에 비해 거의 모든 주파수 대역에서 에너지가 존재하므로 원 신호와 잡음의 특징이 유사한 부분이 많이 존재하게 된다. 그 결과로 1차로 독립성분분석을 적용하는 과정에서 제대로 신호를 분리하지 못하고 원 신호의 에너지가 분리된 잡음 신호에 많이 포함되어 원 신호를 변화시키거나 잡음이 제대로 분리되지 않는 문제가 발생하였다. 즉 음원이 잡음과 유사한 특징을 지닌다면 필터의 성능을 비교하는데 부적합하게 된다. 하지만 제안된 알고리즘을 통해 이를 상당량 해소할 수 있었다. 더불어 NMR 개선량 측면에서는 모든 장르별로 개선된 신호에 존재하는 가청 가능한 잡음의 많은 양이 감소한 결과를 확인할 수 있다.

표 2. 세그멘탈 신호 대 잡음 비의 개선량 비교 (dB)

Table 2. Comparison of the proposed filter's performances with those of existed methods - segmental SNR (dB).

Input noise	Method (SSNR)		
	Only ICA	ICA + spectral subtraction	Proposed
Rock	3.7664	5.1574	6.983
Water	10.4737	11.235	11.173
Siren	-0.3827	3.951	6.8372

표 3. 잡음 대 마스크 비의 개선량 비교 (dB)

Table 3. Comparison of the proposed filter's performances with those of existed methods - NMR (dB).

Input noise	Method (NMR)		
	Only ICA	ICA + spectral subtraction	Proposed
Rock	13.5804	22.7453	27
Water	13.0839	26.4427	31.6009
Siren	13.0908	15.4591	16.0184

4.2. 주관적 청감 테스트

성능을 평가하기 위한 두 번째 테스트로는 제안된 알고리즘을 통해 개선된 사운드 신호의 음질 변화 정도를 판단하기 위해 청감 테스트를 실시하였다. 평가 기준으로는 CCIR Recommendation에서 음질 평가를 위해 채택한 방법인 Degradation Category Rating (DCR) Method를 사용하였다. DCR Method는 표 3과 같은 Degradation Category Scale를 통하여 수정된 사운드 신호의 음질이 원 사운드 신호의 음질과 관련하여 얼마나 충실한지를 비교한다 [10]. 테스트는 20세 이상 오디오 전문 분야에 종사하지 않는 남녀 10명을 대상으로 하였으며, 테스트 장비로는 청취자의 귀에 밀착되는 헤드폰 (audio-technica ATH-ES7), Pentium computer (Intel core2duo E8500)를 사용하였다. 그리고 재생시스템과 피 실험자 간의 거리를 두고 랜덤한 순서로 사운드 신호를 들려주어 청취자가 음원의 종류를 사전에 알 수 없도록 하였다.

표 5의 측정 결과에서 볼 수 있듯이 기존 필터의 경우에는 잡음에 열화된 신호에서 잡음을 완전하게 분리해 낼 수 없다. 이에 따라 잔여 잡음으로 인하여 DCR test 점수가 낮다. 그러나 제안된 알고리즘을 적용한 경우에는 지각필터를 이용하여 가능한 가청 잡음을 제거하였으므로 우수한 성능을 보이고 있음을 알 수 있다.

V. 결 론

본 논문에서는 잡음에 의해 열화된 오디오 신호의 청각 자극 에너지를 이용하여 지각 필터에 적용함으로써 신호

표 4. 청감 테스트 점수

Table 4. Degradation Category Scale.

5	Degradation is inaudible.
4	Degradation is audible but not annoying.
3	Degradation is slightly annoying.
2	Degradation is annoying.
1	Degradation is very annoying.

표 5. DCR test 결과

Table 5. DCR test results.

Input noise	DCR test		
	Only ICA	ICA + spectral subtraction	Proposed
Rock	3.2	3.6	4
Water	3.5	3.4	3.8
Siren	2.9	3.2	3.5

의 음질을 개선하는 과정에서 통계적 특성을 갖지 않는 non-stationary 잡음을 보다 정확하게 추정함으로써 지각 필터의 성능을 향상시키는 알고리즘을 제안하였다.

제안된 방식에서는 지각 필터에 적용되는 추정잡음은 원 신호와 최대한 유사하게 추정하기 위하여 독립성분분석 방법을 사용하며 그 결과로 원 신호와 가까운 효과적 인 개선 결과를 얻을 수 있었다.

또한 기존 방식과 제안된 방식의 성능 평가를 위하여 수행된 SNR과 NMR의 비교 및 피실험자들의 주관적인 평가를 알아볼 수 있는 청감 테스트 결과에서도 원 잡음을 적용시킨 경우와 비슷한 성능의 신호 개선을 확인할 수 있었다.

그러나 제안된 알고리즘으로는 고주파 대역에 에너지가 크거나 원 신호와 특성이 비슷한 잡음인 경우, 잡음 추정에 있어서 오차가 발생하였다. 이러한 잡음 추정의 오차로 인하여 필터 적용 후에도 잔여 잡음이 존재하게 되는 문제가 발생하기도 하였다.

이에 향후 진행될 연구로는 독립성분분석을 이용하여 잡음을 추정하는 과정에서 오차를 보다 줄임으로서 지각 필터의 성능을 더욱더 향상시킬 수 있는 방법에 대한 연구가 진행될 예정이다.

감사의 글

본 연구는 숭실대학교 교내 연구비 지원으로 이루어 졌음.

참고 문헌

1. 서정국, 차형태, "잡음 패턴의 지능적 추정을 통한 음질 개선 알고리즘," *한국지능시스템학회지* 15권, 2호, 230-235쪽, 2005.
2. D. Tsoukalas, J. Mourjopoulos, and G. Kokkinakis, "Perceptual filters for audio signal enhancement," *Journal of Audio Eng. Soc.*, vol. 45, no. 1/2, pp. 22-35, Jan/Feb. 1997
3. S. F. Boll, "Suppression of acoustic noise speech using spectral subtraction", *IEEE Trans. Acoustic, Speech, and Signal processing*, Vol. ASSP-27, 1997
4. 김현중, 차형태, "잡음 신호의 지각 패턴 제어를 통한 음질 개선 알고리즘 개발에 관한 연구," *한국음향학회 하계학술대회 논문집* 21권, 1(s)호, 199-202쪽, 2002.

5. 임형규, 이창기, "독립성분 분석을 이용한 강인한 음성인식," *컴퓨터산업교육학회지*, 5권, 2호, 269-274쪽, 2004.
6. 김응명, "독립성분분석에서 점수함수 생성을 이용한 임묵신호분리," *경희대학교*, 2006
7. 엄혜영, 한현수, 홍민철, 차형태, "적응 지각 필터를 이용한 오디오 음질 개선 알고리즘," *한국음향학회지* 22권, 8호, 687-693쪽, 2003.
8. 최승호, "독립 성분 분석과 스펙트럼 향상에 의한 잡음 환경에서의 음성인식," *말소리*, 제48호, 81-91쪽, 2003
9. Takahashi Y., Takatani T., Osako K., Saruwatari H, and Shikano K., "Blind Spatial Subtraction Array for Speech Enhancement in Noisy Environment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, Issue 4, pp. 650 - 664, 2009
10. ITU-T P.800, *Methods for subjective determination of transmission quality*, Aug. 1996
11. A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for Independent Component Analysis," *Neural Computation*, vol.9, pp. 1483-1492, 1997
12. A. Hyvarinen, "Fast and robust fixed-point algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol.10, no.3, pp. 626-634, May 1999.

저자 약력

• 구 교 식 (Kyo-Sik Koo)



2005년 2월 : 숭실대학교 정보통신전자공학부 (공학사)
 2007년 2월 : 숭실대학교 전자공학과 (공학석사)
 2007년 2월~현재 : 숭실대학교 전자공학과 박사과정 재학
 ※ 주관심분야 : 오디오 및 음성 신호처리, 통신 신호 처리

• 차 형 태 (Hyung-Tai Cha)



1993년 : The University of Pittsburgh (공학박사)
 1993년~1996년 : 삼성전자 신호처리 연구소 선임연구원
 1996년~ 현재 : 숭실대학교 정보통신전자공학부 부교수
 2001년~ 현재 : 한국 신호 및 시스템 학회 이사
 2009 ~ 현재 : 한국 방송공학회 이사
 ※ 주관심분야 : Multimedia Systems and Applications Audio and Video Signal Processing, Communication System, ASIC Implementation of Digital System