

독립성분분석에 기반한 음성신호분리

박형민 (서강대학교)

I. 서론

다양한 신호처리 문제에 있어서 잡음에 대한 강인성의 확보는 실세계 적용에 있어서 가장 중요하게 다루어야 할 부분이다^[1]. 신호나 잡음의 특성이 비교적 간단하며 잘 알려져 있는 경우에는 잡음을 제거하거나 줄이는 잡음 처리가 상당한 수준에 이르렀다고 볼 수 있으나, 처리하고자 하는 신호나 잡음의 특성이 매우 복잡하고 잘 알려져 있지 않은 신호의 경우에는 아직까지 만족할 만한 성능을 얻기 힘들다. 특히 음성에 섞여 있는 잡음을 처리함에 있어서 기존 대부분의 연구는 잡음의 특성과 음성에서의 영향을 모델링하거나 잡음의 정상성(stationarity)을 가정하는 접근 방법을 취해왔으며, 이러한 방법들의 성능은 보통 실제 잡음의 특성이 사용한 모델 및 가정과 얼마나 부합하는지에 따라 크게 좌우된다. 따라서 실세계에서 자주 접하게 되는 비정상 잡음과 같이 모델링이 어려운 음성신호처리에는 매우 어려운 문제로 인식되어 왔다^[1~2].

복잡한 파티 장소에서 친구들과 이야기하고 있는 자신을 상상해 보자. 우리는 주위 다른 사람들의 말 또는 음악 소리와 같은 많은 소리가 있

음에도 불구하고 이야기하고 있는 친구의 말을 어렵지 않게 알아들을 수 있다. 이렇게 여러 가지 음원이 섞여있는 신호로부터 원래의 음원 신호들을 분리해 내는 문제를 ‘각테일파티 문제’라고 한다^[3~4]. 사람은 이러한 각테일파티 문제를 효과적으로 해결할 수 있는 능력을 가지고 있지만, 고전적인 신호처리 기법은 아직 이러한 문제를 잘 해결하지 못하고 있기 때문에 통제된 실험실 환경에서와 달리 주위환경이 시끄럽고 변화가 심한 실제적인 환경에서의 성능은 제한적이다^[2].

이러한 문제를 해결하기 위해서 독립성분분석(ICA: independent component analysis)은 새로운 대안적 방법을 제시해 줄 수 있다. 독립성분분석은 다차원의 데이터를 통계적으로 서로 독립인 확률 변수로 표현하는 신호처리 기법을 말한다^[3~4]. 2차 통계적 특성을 이용하던 기존 일반적 신호처리 기법들과 달리 독립성분분석은 고차 통계적 특성을 이용함으로써 다양한 신호처리 분야에서 폭넓게 응용되어 왔으며, 음성 향상 및 특징 추출 분야에서도 활용되고 있다^[3~5]. 하지만, 실세계에서 접하는 음향 혼합 신호는 단순 가중치 합으로 이루어진 경우를 좀처럼 찾기 어려우며, 전파에 따른 시간 지연과 반향을 고려

한 convolution의 형태로 섞이게 된다. 특히 음향 신호에 있어서 그 혼합 환경은 일반적으로 매우 복잡하기 때문에 음향 혼합 신호에 대한 독립성분 분석은 현재까지 매우 어려운 문제로 남아있다.

음향 혼합 신호를 다루기 위해서 단순 가중치 혼합 신호에 적용해오던 기존의 독립성분분석 알고리즘을 시간 영역 또는 주파수 영역으로 확장하는 방법을 생각해 볼 수 있다^[6~8]. 시간 영역 접근 방법은 반향이 매우 긴 일반적인 음향 혼합 환경에서 매우 많은 계산이 필요하고 음성신호에 대하여 느린 수렴 속도를 보인다. 반면, 주파수 영역 접근 방법은 convolution 연산이 각 주파수에서의 곱셈으로 대체가 가능하므로 계산량은 적다. 하지만 긴 반향을 다루기 위해 푸리에 변환을 위한 프레임의 길이를 길게 해야 하지만 이 때 각 주파수별 입력 데이터의 개수가 줄어드는 문제점으로 인해서 분리 성능에 한계를 가진다^[9]. 이러한 시간 영역 및 주파수 영역 접근 방법의 문제점을 해결하고자 convolution 혼합 신호 분리를 위한 필터뱅크 접근 방법이 제시되었다^[10]. 그리고 구조 및 수식화가 간단한 일반적인 등간격 필터뱅크와 달리 실세계에서 존재하는 다양한 음향신호의 분리에 적합한 Bark 척도 필터뱅크를 도입하여 독립성분분석에 적용하는 방법을 소개한다^[11]. 여기에서는 독립성분분석 접근 방법들의 특성을 파악하고 음향 혼합 신호에 대한 분리 실험을 통하여 성능을 비교하고자 한다.

II. 독립성분분석

미지의 N 개 상호 독립 음원 신호 $s_i(n)$ 이 음향 채널을 통해서 다수의 마이크로폰에서 convolution 혼합 신호를 형성한다고 하면 그 혼

합 신호는 다음과 같이 표현할 수 있다^[3]

$$x_i(n) = \sum_{j=1}^N \sum_{m=0}^{L_m-1} a_{ij}(m) s_j(n-m). \quad (1)$$

여기서 $a_{ij}(m)$ 는 혼합 필터 계수를 나타낸다. 이 혼합 신호로부터 음원 신호를 분리하기 위해서 다음과 같은 피드백 구조 분리 네트워크를 구성해 볼 수 있다^[6].

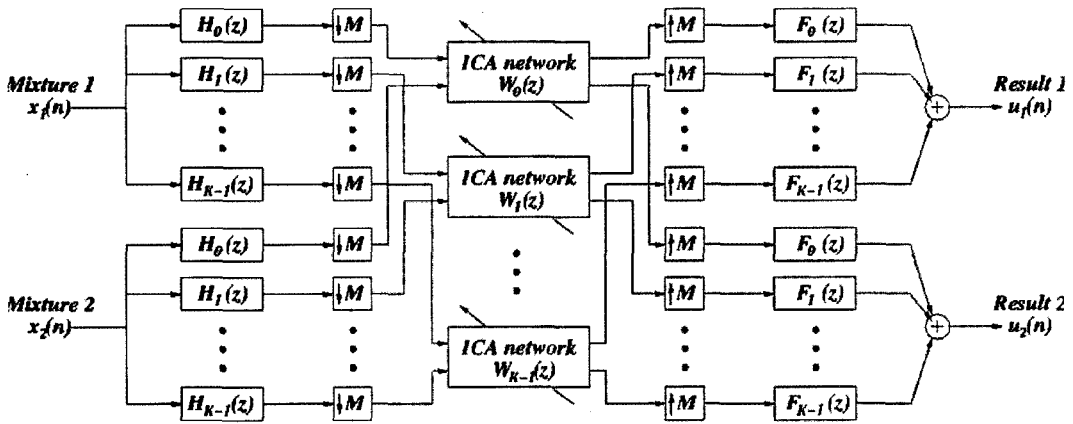
$$u_i(n) = \sum_{m=0}^{L_a} w_{ii}(m) x_i(n-m) + \sum_{j=1, j \neq i}^N \sum_{m=1}^{L_a} w_{ij}(m) u_j(n-m). \quad (2)$$

여기서 적응 필터 $w_{ij}(m)$ 은 출력 신호 $u_i(n)$ 이 원 음원 신호 $s_i(n)$ 를 복원하도록 추정되어야 한다. 여러 가지 독립성분분석 알고리즘 중에서 엔트로피 최대화 방법이 간단하면서도 생물학적 개연성으로 인해서 각광받고 있으며, 이를 수식 (2)에 적용하면 다음과 같다^[6].

$$\begin{aligned} \Delta w_{ii}(0) &\propto 1/w_{ii}(0) - \varphi(u_i(n)) x_i(n), \\ \Delta w_{ii}(m) &\propto -\varphi(u_i(n)) x_i(n-m), \quad m \neq 0, \\ \Delta w_{ij}(m) &\propto -\varphi(u_i(n)) u_j(n-m), \quad i \neq j, \\ \varphi(u_i(n)) &= -\frac{\partial(p(u_i(n)))/\partial u_i(n)}{p(u_i(n))}. \end{aligned} \quad (3)$$

여기서 $p(\cdot)$ 는 확률밀도함수를 나타낸다. 시간 영역 접근 방법은 반향이 긴 일반적인 음향 혼합 환경에서 긴 필터의 convolution 계산 및 필터 계수의 적응을 위해서 매우 많은 계산량이 필요하며, 음성과 같이 백색 신호와 특성이 다른 신호에 대하여 느린 수렴 속도를 보인다.

반면에 주파수 영역 접근 방법은 convolution 혼합 신호를 다음과 같이 나타낸다.



(그림 1) 2개의 혼합 및 음원 신호에 대한 과표본 등간격 필터뱅크에 기반한 독립성분분석 구조

$$x(f,n) = A(f)s(f,n), \quad \forall f. \quad (4)$$

여기서 $x(f,n)$ 과 $s(f,n)$ 은 각각 마이크로폰 과 음원 신호의 푸리에 변환을 통해 주파수 f 에서 얻은 값으로 이루어진 벡터를 나타내며, $A(f)$ 는 혼합 필터의 푸리에 변환을 통해 주파수 f 에서 얻은 값을 원소로 갖는 행렬을 나타낸다. 식 (4)로부터 convolution 혼합은 주파수 영역에서 단순 가중치 혼합의 집합으로 표현이 가능하기 때문에, 모든 주파수에서 단순 가중치 혼합 신호를 분리하는 독립성분분석을 적용함으로써 convolution 혼합 신호를 분리할 수 있다. 유일한 차이는 입력 신호가 실수가 아닌 복소수의 형태를 취하고 있다는 점이다. Natural gradient^[12]를 이용하여 각 주파수에서 엔트로피 최대화 방법은 다음과 같다.

$$\Delta W(f) \propto [I - \varphi(u(f,n))] u^H(f,n) W(f),$$

$$\varphi(u_i) = -\frac{\partial p(|u_i|) / \partial |u_i|}{p(|u_i|)} \exp(j \cdot \angle u_i). \quad (5)$$

여기서 $W(f)$ 는 주파수 f 에서 음원 신호 복원

을 위한 적응 필터의 푸리에 변환 값으로 이루어진 행렬을 나타낸다. 주파수 영역 접근 방법은 시간 영역의 convolution 연산이 각 주파수별 곱셈으로 변환되어 계산량을 크게 줄일 수 있다. 하지만, 긴 반향으로 혼합된 신호를 분리하기 위해서 푸리에 변환을 위한 프레임의 길이가 충분히 길어야 한다. 프레임의 길이가 길어질 때, 계산량의 효율성을 유지하면서 이웃한 프레임 데이터 간의 차이를 확보하기 위해서 다음 프레임의 위치는 현재 프레임으로부터 시간 영역 데이터 표본을 그만큼 많이 건너뛰어야 한다. 이 때, 각 주파수별 데이터의 개수는 줄어들게 되므로 분리 행렬을 추정하는데 그 데이터의 개수가 충분하지 않을 수 있다. 따라서 주파수 영역 접근 방법은 반향에 따른 프레임의 길이와 데이터 개수 사이에서 적절한 절충점을 찾아야 하므로 분리 성능에 한계가 있을 수밖에 없다.^[9]

III. 독립성분분석을 위한 과표본 필터뱅크 접근 방법

본 접근 방법에 적용할 필터뱅크로 기존 최소

평균 제곱법에 기반한 대역별 적응 필터 방법에서 효과적으로 사용해진 과표본 필터뱅크를 사용한다^[13]. 필터 대역의 개수 K 보다 적은 decimation 값 M 을 사용하여 aliasing을 없앤 과표본 필터뱅크는 저대역 표본 필터를 일반화된 푸리에 변환으로 복소 변조하여 효율적으로 구성할 수 있다. 즉 필터뱅크의 해석 필터 $h(k, n)$ 를 다음과 같이 나타낼 수 있다.

$$h(k, n) = e^{j(2\pi/K)(k+1/2)(n-(L_q-1)/2)} \cdot q(n), \quad (6)$$

$$k=0, 1, \dots, K-1, \quad n=0, 1, \dots, L_q-1.$$

여기서 L_q 는 표본 필터 $q(n)$ 의 길이를 나타낸다. 합성 필터로는 다음과 같이 복소켈레와 시간 반전된 해석 필터를 사용한다.

$$f(k, n) = h^*(k, L_q - n - 1). \quad (7)$$

표본 필터 $q(n)$ 은 원 신호의 복원 성능과 고대역 신호의 제거 성능을 최대화 하도록 최소 제곱 반복법(iterative least-squares algorithm)을 적용하여 구하도록 한다^[13]. 또한, 해석 및 합성 필터의 다상 표현을 이용하여 필터뱅크를 효율적으로 구현할 수 있다^[14].

Aliasing이 없는 과표본 필터뱅크에서 독립성분분석을 수행할 때, 각 대역별 적응 필터는 다른 대역과 관계없이 독립적으로 추정이 가능하므로 병렬 처리에 적합한 구조를 얻을 수 있다. <그림 1>은 2개의 혼합 및 음원 신호에 대한 과표본 필터뱅크에 기반한 독립성분분석 구조를 나타낸다^[10]. 입력 신호는 해석 필터를 통해서 대역별 신호로 나뉘고, M 표본만큼 decimation된다. 따라서 반향 길이가 M 배로 줄어든 혼합 환경을 고려하여 시간 영역 접근 방법보다 M 배 짧은 분리

적응 필터를 사용하고, 이를 시간 영역 접근 방법과 유사한 방법으로 적응시킬 수 있다. 만약 식 (2)와 같은 피드백 구조 분리 네트워크를 사용한다면 분리 네트워크의 출력 신호는 다음과 같다.

$$u_i(k, n) = \sum_{m=0}^{L_q} w_{ii}(k, m)x_i(k, n-m) + \sum_{j=1, j \neq i}^N \sum_{m=1}^{L_q} w_{ij}(k, m)u_j(k, n-m). \quad (8)$$

이 때, 대역 신호가 복소수이므로 다음과 같이 필터 계수를 적응시킨다.

$$\begin{aligned} \Delta w_{ii}(k, 0) &\propto 1/w_{ii}^*(k, 0) - \varphi(u_i(k, n))x_i^*(k, n), \\ \Delta w_{ii}(k, m) &\propto -\varphi(u_i(k, n))x_i^*(k, n-m), \quad m \neq 0, \\ \Delta w_{ij}(k, m) &\propto -\varphi(u_i(k, n))u_j^*(k, n-m), \quad i \neq j. \end{aligned} \quad (9)$$

여기서, $w_{ij}(k, m)$, $u_i(k, n)$, $x_i(k, n)$ 은 각각 k 번째 대역의 적응 필터 계수, 분리 신호, 입력 신호를 나타내며, $\varphi(\cdot)$ 의 형태로는 식 (5)를 사용한다. 마지막으로 분리된 신호로부터 순서와 크기를 맞춘 후 M 표본만큼 expansion하고 합성 필터를 통과하여 음원 신호를 복원할 수 있다.

필터뱅크 접근 방법은 대역별로 여전히 convolution 혼합 신호를 분리하게 되고 충분한 데이터를 확보하고 있으므로 주파수 영역 접근 방법과 달리 성능이 저하되지 않는다. 게다가, 시간 영역 접근 방법에 비하여 decimation에 의해 각 대역에서 분리 네트워크가 M 배 짧고, 입력 신호도 M 배 느린 속도로 입력되므로 길이가 긴 적응 필터에 대해서 계산량을 크게 줄일 수 있다^[10]. 또한, 반향과 관계없이 대역의 개수를 임의로 선택할 수 있으며, decimation이 입력 신호를 더욱 백색화하고 대역별 적응 필터의 길이를 줄이므로 적응 필터 계수의 수렴 속도가 빨라진다.

독립성분분석은 서로 독립인 음원 신호를 복원하기 때문에, 복원된 신호의 크기와 순서는 임의로 변경될 수 있다^[3~4]. 특히 엔트로피 최대화 방법은 분리 신호가 시간적으로 백색화되도록 필터를 추정하므로 이를 방지하기 위해서 입출력간 직결 필터 $w_{ii}(k, m)$ 을 단순 가중치로 변경한다^[6,10]. 과표본 필터뱅크를 이용한 독립성분분석에서 각 대역별 적응 필터는 다른 대역과 관계없이 독립적으로 추정되므로 주파수 영역 접근 방법과 유사하게 대역별로 분리 신호의 크기와 순서가 다를 수 있다. 실제적인 음성 및 음향신호는 일반적으로 시간에 따라서 변하는 통계적 특성을 나타내므로 Murata의 방법^[15]을 필터뱅크 접근 방법에 맞게 변형하여 사용할 수 있다^[10]. 주파수 영역 접근 방법에서는 긴 반향을 고려하여 매우 많은 주파수에서 독립적으로 분리 행렬을 추정하고 주파수별 데이터의 개수도 제한적이기 때문에 주파수별 크기와 순서를 보정하는 것이 쉽지 않지만, 필터뱅크 접근 방법에서는 반향과 관계없이 대역별로 충분한 대역폭 및 데이터를 확보할 수 있기 때문에 대역별 크기와 순서를 정확히 보정하는 것이 쉽다^[10].

IV. Bark 척도 필터뱅크를 이용한 독립성분분석

실제적 음향신호는 일반적으로 대부분의 에너지가 저주파 영역에 집중되어 있고, 주파수가 높아짐에 따라서 고주파 영역보다 저주파 영역에서 에너지가 더 급격하게 감소하는 경향이 있다. 이러한 음향신호에 등간격 필터뱅크 접근 방법을 사용하면 고주파 대역보다 저주파 대역에서 입력 신호의 특성이 백색 신호와 더 많이 달라서

분리 필터 추정에 있어서 상대적으로 느린 수렴 속도를 보인다. 하지만, 일반적인 음향신호는 저주파 영역에 대부분의 에너지가 집중되어 있기 때문에 전체적인 분리 성능이 저주파 느린 수렴 속도에 큰 영향을 받게 되며, 고주파 영역에서 정확한 적응 필터 추정에 필요한 데이터가 충분히 확보되지 못할 수도 있다. 따라서 독립성분분석에 사용한 등간격 필터뱅크는 일반적인 음향신호의 특성을 적절히 고려하기 어려우며, 이러한 문제점을 해결하기 위해서는 필터뱅크의 중심 주파수가 올라감에 따라서 대역폭도 넓어져야 한다. 대역폭이 넓어지면 aliasing을 막기 위해서 그만큼 decimation 값도 감소해야 하는데, 이는 동일한 시간 구간을 담당하기 위한 적응 필터의 길이를 증가시키게 된다. 일반적인 최대 경사법 기반 알고리즘들은 적응 필터의 입력 신호들로 구성된 상관행렬의 condition 값에 따라 그 수렴 속도가 결정된다^[16]. 그런데 Bordering 정리^[17]에 의하면, 그 condition 값은 필터의 길이가 증가함에 따라 단조증가만 가능하므로 적응 필터의 수렴 속도는 느려질 뿐이다.

위의 추론을 종합해보면, 음향신호에 대한 독립성분분석으로 적합한 필터뱅크의 대역 분할 방법은 일정한 대역별 수렴 속도 및 고주파 영역에서의 정확한 적응 필터 추정과 동시에 적응 필터의 길이가 길어짐에 따른 수렴 속도 저하 문제 해결을 고려하여 적절한 절충점을 찾아야 하며, 여기에서는 Bark 척도 필터뱅크^[18]를 하나의 대안으로 제시한다. Bark 척도 필터뱅크에서 단위가 rad/s인 선형 주파수 ω 는 Bark 주파수 Ω 로 다음과 같이 변환된다^[19].

$$\Omega(\omega) = 6 \log \left\{ \frac{\omega}{1200\pi} + \left[\left(\frac{\omega}{1200\pi} \right)^2 + 1 \right]^{0.5} \right\}. \quad (10)$$

이 Bark 척도 필터뱅크 접근 방법을 사용하게 되면, 저주파 대역에서 좁은 대역폭에 해당하는 큰 decimation 값을 사용함으로써 백색 신호와 더 유사한 특성을 갖는 입력 신호를 사용하여 적응 필터의 수렴 속도를 개선할 수 있다. 또 고주파 대역에서는 대역폭이 넓어져 적응 필터의 정확한 추정에 충분한 데이터를 확보할 수 있게 되므로 분리 성능을 개선할 수 있다.

과표본 Bark 척도 필터뱅크를 설계하기 위해서 앞 절에서 소개한 등간격 필터뱅크 설계 방법을 적절히 수정하여야 한다. 주파수별 대역폭이 다르기 때문에 표본 필터도 대역마다 달라야 하므로 k 번째 대역의 해석 필터 $h(k, n)$ 는 다음과 같이 나타낼 수 있다.

$$h(k, n) = e^{j2\pi(F_k/F_s)(n-(L_{q_k}-1)/2)} \cdot q(k, n), \quad (11)$$

$$k = 1, 2, \dots, K, n = 0, 1, \dots, L_{q_k} - 1.$$

여기서 $q(k, n)$ 은 k 번째 대역에 해당하는 표본 필터이고, F_k, F_s, L_{q_k} 는 각각 대역 중심주파수, 표본화 주파수, $q(k, n)$ 의 길이를 나타낸다. 모든 대역별 표본 필터는 해당 대역폭을 유지하면서 원 신호의 복원 성능과 고대역 신호의 제거

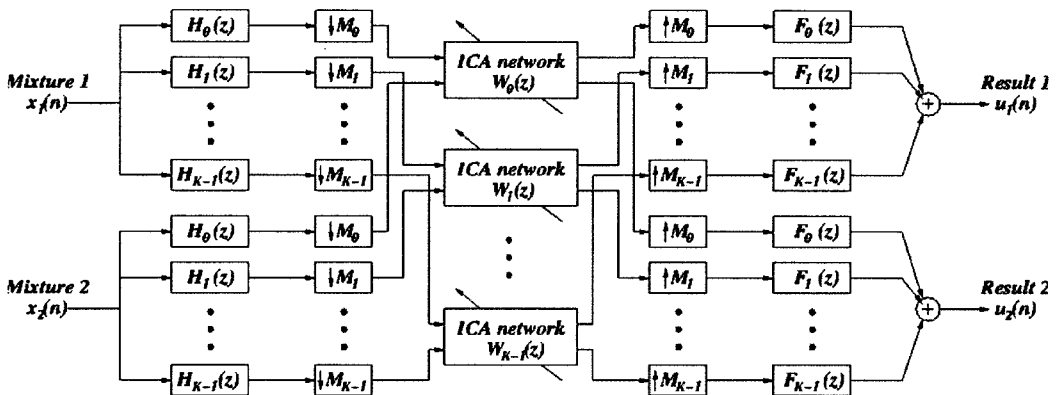
성능을 최대화 하도록 최소 제곱 반복법을 적용하여 구하도록 한다^[11].

2개의 혼합 및 음원 신호에 대하여 이 필터뱅크를 적용하여 구성한 독립성분분석 구조는 <그림 2>와 같다^[11]. 전체적인 구조는 등간격 필터뱅크를 사용한 독립성분분석과 유사하지만, 등간격 필터뱅크가 아니기 때문에 대역별 decimation 값이 다르다. 따라서 대역별 분리 네트워크의 출력 신호는 다음과 같다.

$$u_i(k, n_k) = \sum_{m=0}^{L_{q_i}} w_{ii}(k, m_k) x_i(k, n - m_k) + \sum_{j=1, j \neq i}^N \sum_{m=1}^{L_{q_i}} w_{ij}(k, m_k) u_j(k, n - m_k). \quad (12)$$

이 때, 대역 신호가 복소수이므로 등간격 필터뱅크의 경우와 유사하게 $\varphi(\cdot)$ 의 형태로 식 (5)를 사용하여 다음과 같이 필터 계수를 적용시킨다.

$$\begin{aligned} \Delta w_{ii}(k, 0) &\propto 1/w_{ii}^*(k, 0) - \varphi(u_i(k, n_k))x_i^*(k, n_k), \\ \Delta w_{ii}(k, m_k) &\propto -\varphi(u_i(k, n_k))x_i^*(k, n - m_k), \quad m \neq 0, \\ \Delta w_{ij}(k, m_k) &\propto -\varphi(u_i(k, n_k))u_j^*(k, n - m_k), \quad i \neq j. \end{aligned} \quad (13)$$

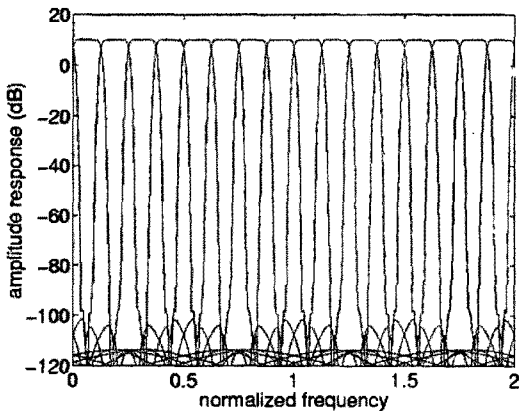


<그림 2> 2개의 혼합 및 음원 신호에 대한 과표본 Bark 척도 필터뱅크에 기반한 독립성분분석 구조

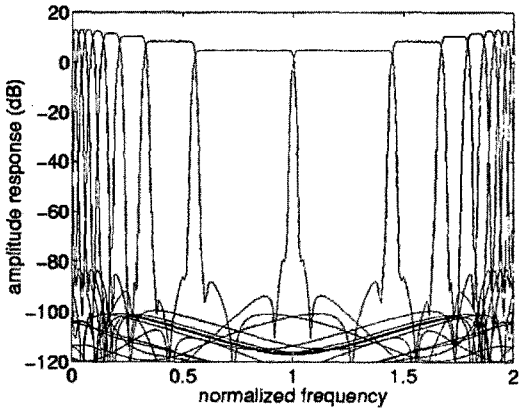
V. 실험 결과

1. 필터뱅크 설계

〈그림 3〉은 16개의 대역을 갖는 과표본 등간격 필터뱅크와 Bark 척도 필터뱅크의 해석 필터 주파수 응답을 나타낸다. 16kHz 표본화 주파수를 고려하여 등간격 필터뱅크에서는 대역별 중심 주파수를 일정한 간격으로 배치하고 decimation 값으로 10을 사용한 반면에 Bark 척도 필터뱅크에서는 대역별 중심주파수를 [109 328 578

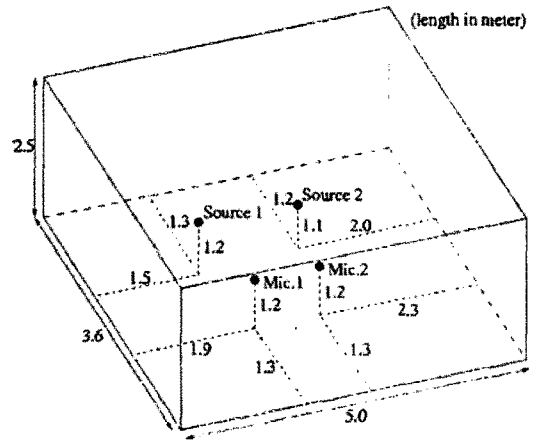


(a) 등간격 필터뱅크



(b) Bark 척도 필터뱅크

〈그림 3〉 16개의 대역을 갖는 과표본 필터뱅크의 해석 필터 주파수 응답



〈그림 4〉 음향 필터 추정을 위한 2개의 음원 위치와 2개의 마이크로폰 위치가 설정된 가상의 직육면체 공간

906 1391 2156 6203]Hz에 배치하고 [22 22 20 18 14 11 7 3]의 decimation 값을 사용하였다. 이 decimation 값은 Bark 척도에 따라 구성된 중심주파수에 대하여 계산량을 줄이기 위해서 aliasing이 생기지 않는 가장 큰 자연수 값을 사용하였다. 16개의 대역을 사용함으로써 저주파 대역 신호에 대해서는 백색화에 충분한 decimation 값을 확보할 수 있었고, 고주파 대역 신호에 대해서는 저주파 대역보다 적은 decimation 값을 사용함으로써 분리 필터 추정에 충분한 데이터를 확보할 수 있었다. 모든 표본 필터는 220개의 계수로 이루어져 있었다. 필터뱅크 통과 전의 광대역 혼합 신호의 표본 개수에 대한 필터뱅크 통과 후 대역별 신호 총 표본 개수의 비인 과표본율은 등간격 및 Bark 척도 필터뱅크 각각 160%와 167%였다.

2. 독립성분분석 실험

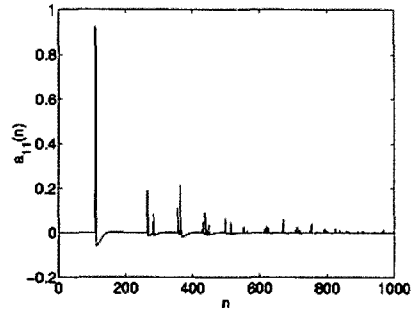
음향 혼합 신호를 분리하기 위한 독립성분분석 실험을 진행하기 위하여 독립 음원 신호로서

5초의 길이에 해당하는 2개의 음성신호를 사용하였다. 혼합 환경을 구성하기 위하여 <그림 4>와 같이 가상의 직육면체 공간에서 2개의 음원 위치와 2개의 마이크로폰 위치를 설정한 후 각 음원과 마이크로폰 사이의 음향 필터를 이미지 방법^[20]으로 추정하여 <그림 5>에 나타내었다. 여기서 벽의 반사 계수는 모두 0.4, 천정과 바닥의 반사 계수는 각각 0.5와 0.25이다. 독립성분석 성능은 신호 대 간섭 비 (SIR: signal-to-interference ratio)로 비교하며, 2개의 혼합 및 음원 신호의 경우에 다음과 같이 나타낼 수 있다^[21].

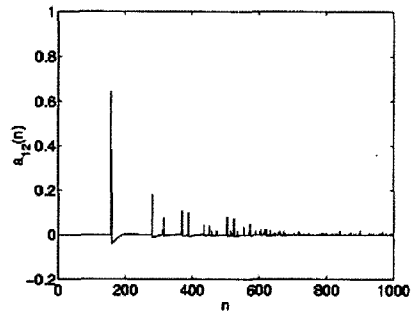
$$SIR(dB) = \frac{1}{2} \left| 10 \log \left(\frac{\langle (u_{1,s_1}(n))^2 \rangle}{\langle (u_{1,s_2}(n))^2 \rangle} \cdot \frac{\langle (u_{2,s_2}(n))^2 \rangle}{\langle (u_{2,s_1}(n))^2 \rangle} \right) \right| \quad (14)$$

여기서 $u_{j,s_i}(n)$ 는 i 번째 음원 신호 $s_i(n)$ 만 활성화되었을 때 분리 네트워크의 j 번째 출력 신호를 나타낸다.

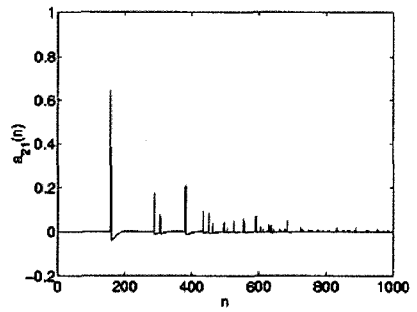
<그림 6>은 2개의 음성 혼합 신호에 대하여 독립성분석을 위한 시간 영역, 주파수 영역, 등간격 과표본 필터뱅크 접근 방법을 적용한 신호 분리 결과의 신호 대 간섭 비를 나타내고 있다. 시간 영역 접근 방법에서는 각 분리 적응 필터에 2048개의 계수를 사용하였고, 주파수 영역 접근 방법에서는 프레임 크기를 2048 표본, 프레임 이동을 프레임 크기의 반으로 설정하였다. <그림 3>(a)를 사용한 필터뱅크 접근 방법에서 각 분리 적응 필터는 $\lceil 2048/M \rceil$ 개의 계수를 사용하였다. 음성신호의 확률밀도함수는 대략적으로 Laplace 분포를 따른다고 할 수 있으므로, 음성 음원 신호를 분리하기 위한 식 (9)에서 $\varphi(u_i)$ 로 $\text{sgn}(|u_i|)\exp(j \cdot \angle u_i)$ 를 사용하였다. <그림 6>에서 주파수 영역 접근 방법은 다른 두 접근



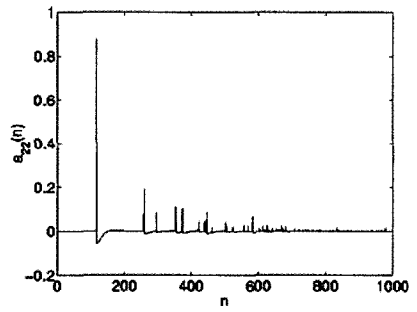
(a) a_{11}



(b) a_{12}

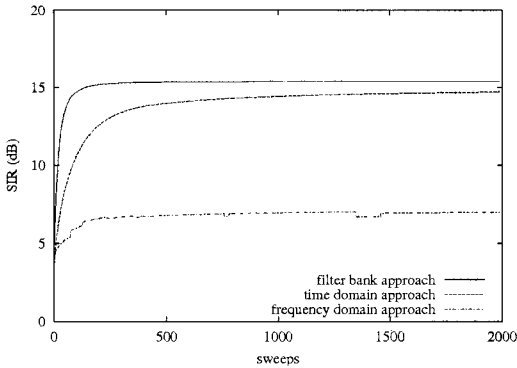


(c) a_{21}



(d) a_{22}

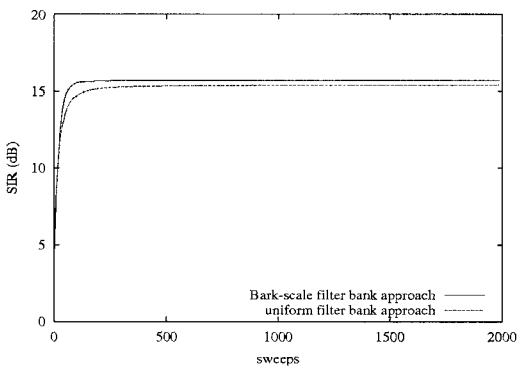
<그림 5> 혼합 신호 생성을 위한 음향 필터



〈그림 6〉 2개의 음성 혼합 신호에 대하여 독립성분분석을 위한 시간 영역, 주파수 영역, 등간격 과표본 필터뱅크 접근 방법을 적용한 신호 분리 결과의 신호 대 간섭 비

방법과 비교하여 분리 성능이 크게 떨어짐을 알 수 있다. 이는 주파수 영역 접근 방법이 가지고 있는 긴 반향 고려와 충분한 데이터 확보의 절충을 통한 성능의 한계에 기인한다. 그리고 필터뱅크 접근 방법은 대역별 신호가 decimation 과정을 거치므로 시간 영역 접근 방법보다 훨씬 빠른 수렴 속도를 나타내었다.

〈그림 7〉은 2개의 음성 혼합 신호에 대한 과표본 필터뱅크 기반 독립성분분석에서 〈그림 3〉



〈그림 7〉 2개의 음성 혼합 신호에 대한 과표본 필터뱅크 기반 독립성분분석에서 〈그림 3〉의 등간격 및 Bark 척도 필터뱅크를 사용할 때 신호 분리 결과의 신호 대 간섭 비

의 등간격 및 Bark 척도 필터뱅크를 사용할 때 신호 분리 결과의 신호 대 간섭 비를 나타낸다. Bark 척도 필터뱅크를 사용하였을 때, 등간격 필터뱅크를 사용하였을 때보다 더 빠른 수렴 속도와 더 높은 신호 대 간섭 비를 나타냄을 알 수 있다. 이는 Bark 척도 필터뱅크를 사용함으로써 저주파 대역에서 decimation 값을 크게 하여 수렴 속도를 상대적으로 빠르게 할 수 있고, 고주파 대역에서 decimation 값을 작게 하여 상대적으로 더 정확한 필터 계수 추정이 가능해지기 때문이다.

VI. 결론

음성 및 음향신호분리를 위한 다양한 독립성분분석 알고리즘의 장단점을 분석하고 실험을 통해서 그 성능을 비교하였다. 과표본 필터뱅크 접근 방법은 주파수 영역 접근 방법보다 훨씬 더 좋은 분리 성능과 시간 영역 접근 방법보다 더 적은 계산량으로도 더 빠른 수렴속도를 나타내었다. 특히 필터뱅크 접근 방법에 Bark 척도 필터뱅크를 사용함으로써 등간격 필터뱅크보다 더 빠른 수렴속도와 더 좋은 분리 성능을 얻을 수 있다.

참고문헌

- [1] B.-H. Juang, Speech recognition in adverse environments, *Comput. Speech Lang.*, Vol.5, No.3, pp.275-294, 1991.
- [2] B. Raj, V. Parikh, R. M. Stern, The effects of background music on speech

- recognition accuracy, in Proc. IEEE ICASSP, pp.851-854, 1997.
- [3] A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, Wiley, New York, NY, 2001.
- [4] T.-W. Lee, Independent component analysis, Kluwer Academic Publisher, Boston, MA, 1998.
- [5] J.-H. Lee, T.-W. Lee, H.-Y. Jung, S.-Y. Lee, On the efficient speech feature extraction based on independent component analysis, Neural Process. Lett., Vol.15, No.3, pp.235-245, 2002.
- [6] K. Torkkola, Blind separation of convolved sources based on information maximization, in Proc. IEEE Int. Workshop on NNSP, pp.423-432, 1996.
- [7] H.-M. Park, H.-Y. Jung, T.-W. Lee, S.-Y. Lee, Subband-based blind signal separation for noisy speech recognition, Electron. Lett. Vol.35, No.23, pp.2011-2012, 1999.
- [8] P. Smaragdis, Blind separation of convolved mixtures in the frequency domain, Neurocomputing, Vol.22, pp.21-34, 1998.
- [9] S. Araki, S. Makino, R. Aichner, T. Nishikawa, H. Saruwatari, Subband based blind source separation with appropriate processing for each frequency band, in Proc. Int. Conf. ICA and BSS, pp.499-504, 2003.
- [10] H.-M. Park, C. S. Dhir, S.-H. Oh, S.-Y. Lee, A filter bank approach to independent component analysis for convolved mixtures, Neurocomputing, Vol.69, pp.2065-2077, 2006.
- [11] H.-M. Park, S.-H. Oh, S.-Y. Lee, A Bark-scale filter bank approach to independent component analysis for acoustic mixtures, Neurocomputing, doi: 10.1016/j.neucom.2009.08.009, accepted, 2009.
- [12] S. Amari, A. Cichocki, H. H. Yang, A new learning algorithm for blind source separation, in Advances in Neural Information Processing Systems, vol. 8, MIT Press, Cambridge, MA, pp.757-763, 1996.
- [13] M. Harteneck, S. Weiß, R. W. Stewart, Design of near perfect reconstruction oversampled filter banks for subband adaptive filters, IEEE Trans. Circuits Syst. II, Vol.46, pp.1081-1085, 1999.
- [14] S. Weiß, L. Lampe, R. W. Stewart, Efficient implementations of complex and real valued filter banks for comparative subband processing with an application to adaptive filtering, in Proc. First Int. Symposium on Communication Systems and Digital Signal Processing, pp.32-35, 1998.
- [15] N. Murata, S. Ikeda, A. Ziehe, An approach to blind source separation based on temporal structure of speech signals, Neurocomputing, Vol.41, pp.1-24, 2001.
- [16] S. Haykin, Adaptive Filter Theory, fourth ed., Prentice-Hall, Upper Saddle River, NJ, 2002.

[17] M. H. Hayes, Statistical Digital Signal Processing and Modeling, Wiley, New York, NY, 1996.

[18] L. Rabiner, B.-H. Huang, Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[19] H. Hermansky, Perceptual linear predictive (PLP) analysis of speech, J. Acoust. Soc. Am., Vol. 87, No.4, pp.1738-1752, 1990.

[20] J. B. Allen, D. A. Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Am., Vol.65, No.4, pp.943-950, 1979.

[21] D. Schobben, K. Torkkola, P. Smaragdis, Evaluation of blind signal separation methods, in Proc. Int. Conf. ICA and BSS, pp.261-266, 1999.

저자소개



박형민

1997년 2월 한국과학기술원 전기및전자공학과
 1999년 2월 한국과학기술원 전기및전자공학과
 2003년 8월 한국과학기술원 전자전산학과
 2003년 9월~2005년 2월 한국과학기술원 바이오시스템학과 Post-Doc
 2005년 3월~2007년 1월 Post.-Doc. at the Language Technologies Institute, Carnegie Mellon University
 2007년 3월~현재 서강대학교 전자공학과 조교수

주관심 분야 : 잡음하 음성인식을 위한 독립성분분석, 압축신호분리, 적응잡음제거, 두 귀 신호처리 기법 등의 이론 및 응용