

다채널 주파수영역 독립성분분석에서 분리된 신호 전력비의 공분산을 이용한 주파수 빈 정렬

Frequency Bin Alignment Using Covariance of Power Ratio of Separated Signals in Multi-channel FD-ICA

전 성 일¹⁾ · 배 건 성²⁾

Quan, Xingri · Bae, Keunsung

ABSTRACT

In frequency domain ICA, the frequency bin permutation problem falls off the quality of separated signals. In this paper, we propose a new algorithm to solve the frequency bin permutation problem using the covariance of power ratio of separated signals in multi-channel FD-ICA. It makes use of the continuity of the spectrum of speech signals to check if frequency bin permutation occurs in the separated signal using the power ratio of adjacent frequency bins. Experimental results have shown that the proposed method could fix the frequency bin permutation problem in the multi-channel FD-ICA.

Keywords: BSS, independent component analysis, frequency bin permutation, power ratio

1. 서론

암묵적 신호분리 (BSS: Blind Source Separation)는 음원과 채널에 대한 사전정보 없이 혼합된 신호로부터 원래의 신호를 분리해 내는 기법이다. ICA(Independent Component Analysis)는 BSS의 한 가지 접근방식으로서, 음원이 서로 확률통계적으로 독립이라는 가정하에 확률통계모델을 이용하여 혼합된 신호로부터 원 음원을 분리해내는 기법이다. ICA 순시혼합모델(instantaneous mixing model)에 대해서는 2000년 초반까지 많은 연구가 수행 되었으며, minimum Kullback-Leibler divergence, maximum entropy, maximum negentropy 등 정보이론에 기반한 ICA 기본 알고리즘들이 정립되었다[1,2]. ICA는 본질적으로 스케일 모호성(scale ambiguity)과 치환 모호성(permutation ambiguity) 문제를 수반하게 되는데, 스케일 모호

성 문제는 minimal distortion principle[3]로 일정하게 보상 가능하지만 치환 모호성 문제로 인한 분리된 신호의 순서 뒤섞임 문제는 여전히 많은 연구가 계속되고 있다.

마이크로폰으로 수집된 음성신호는 음원과 채널의 컨볼루션 형태로 표현되며, 컨볼루션 형태의 혼합 모델(convolutive mixing model)을 적용하게 된다[4]. 일반적으로 시간 영역에서의 컨볼루션 혼합 모델은 주파수 영역으로 변환하여 곱의 형태로 표현되는 FD-ICA(Frequency Domain ICA) 모델을 구성한다[5]. FD-ICA는 각 주파수 빈(frequency bin) 별로 순시혼합모델 ICA를 수행하는데, ICA의 치환 모호성 문제로 하여 주파수 빈 뒤섞임 문제가 발생하며 따라서 원래의 신호를 제대로 분리해 낼 수 없게 된다[6].

FD-ICA 주파수 빈 뒤섞임 문제를 해결하고자 여러 가지 접근방식이 연구되고 있는데[6~10], IVA (Independent Vector Analysis)[7]는 전체 주파수 빈을 하나의 벡터로 간주하여 주파수 빈 상호간의 의존성을 고려하는 방식으로 접근하여 비교적 만족스러운 음원 분리 성능을 얻고 있지만 여러 경우에서 주파수 빈 뒤섞임 문제가 발생하여 분리된 신호의 음질 저하를 수반하게 된다[8]. 주파수 빈 뒤섞임을 정정하는 후처리 방식으로는 TDOA(Time Difference of Arrival) 정보를 이용하는 방법[6], 스펙트럼 상관도를 이용하는 방법[9], 전력비(power

1) 경북대학교, comet263@hotmail.com

2) 경북대학교, ksbae@mir.knu.ac.kr, 교신저자

이 논문은 방위사업청과 국방과학연구소의 지원으로 수행되었습니다.(계약번호:UD130007DD)

접수일자: 2014년 5월 19일

수정일자: 2014년 8월 25일

게재결정: 2014년 9월 11일

ratio) 상관도를 이용하는 방법[10] 등이 있는데, TDOA 정보를 이용하는 방법은 음원이 같은 방위각에 있을 경우 주파수 빈 뒤섞임을 정정하지 못하는 문제점이 있으며, 스펙트럼 상관도를 이용하는 방법보다 전력비 상관도를 이용하는 것이 주파수 빈 정렬에서 장점이 있다고 보고된 바 있다[10].

본 논문에서는 전력비의 상관도를 이용하는 새로운 방법의 주파수 빈 정렬 알고리즘을 제안하고, 실험을 통해 제안한 방법의 타당성을 검증하였다. 제안한 방법에서는 분리된 신호의 주파수 빈 전력비의 공분산 행렬을 구하여 상관도를 분석함으로써 주파수 빈 정렬을 위한 알고리즘 및 로컬 치환행렬의 구성이 단순화되는 장점을 가지게 된다.

본 논문의 구성은 다음과 같다. 먼저 제 2장에서 FD-ICA와 분리된 신호의 주파수 빈에 대한 전력비 개념에 대해 간단히 설명하고, 전력비 상관도를 이용하여 주파수 빈 뒤섞임 문제를 해결할 수 있는 새로운 방법을 제안한다. 제 3장에서는 실험조건 및 결과를 이용하여 제안한 방법의 타당성을 검증하고, 제 4장에서 결론을 맺는다.

2. 다채널 FD-ICA에서의 주파수 빈 정렬

2.1 FD-ICA 알고리즘 및 분리된 신호의 전력비

FD-ICA 모델은 식 (1)과 같으며 $\mathbf{S}(\omega, \tau)$, $\mathbf{X}(\omega, \tau)$, $\mathbf{Y}(\omega, \tau)$ 는 각각 원 음원, 혼합 신호 및 분리신호의 STFT (Short Time Fourier Transform)이고, τ 는 프레임 번호($1 \leq \tau \leq L$), ω 는 주파수 빈을 나타낸다. $\mathbf{H}(\omega)$, $\mathbf{W}(\omega)$ 는 각 주파수 빈에서의 혼합행렬과 분리행렬을 의미하며, 음원 수와 혼합신호의 수가 각각 N 이라고 가정하면 $N \times N$ 행렬이 된다.

$$\begin{aligned} \mathbf{X}(\omega, \tau) &= \mathbf{H}(\omega) \mathbf{S}(\omega, \tau) \\ \mathbf{Y}(\omega, \tau) &= \mathbf{W}(\omega) \mathbf{X}(\omega, \tau) \end{aligned} \quad (1)$$

분리행렬을 추정하는 FD-ICA 알고리즘은 식 (2)와 같다[5].

$$\begin{aligned} \Delta \mathbf{W}(\omega) &= [\mathbf{I} - E\{\Phi(\mathbf{Y}(\omega, \tau)) \mathbf{Y}^H(\omega, \tau)\}] \mathbf{W}(\omega) \\ \mathbf{W}(\omega) &\leftarrow \mathbf{W}(\omega) + \eta \Delta \mathbf{W}(\omega) \end{aligned} \quad (2)$$

식 (2)에서 \mathbf{I} 는 $N \times N$ 단위행렬이고 Φ 는 활성화 함수 (activation function)로서, $\Phi(\xi) = \tanh[\gamma|\xi|] \exp[j \text{Arg}(\xi)]$, γ 는 shape 파라미터, $\text{Arg}(\cdot)$ 는 복소수 argument이며, η 는 학습률(learning rate) 파라미터이다. FD-ICA는 각 주파수 빈에 대해 순시혼합모델 ICA를 수행함으로써 킨블루션으로 혼합된 신호를 분리할 수 있지만 ICA가 본질적으로 가지고 있는 출력신호의 뒤섞임 문제가 분리된 신호 사이의 주파수 빈의 뒤섞임 형태로 나타나므로 원 음원을 추정하기 위해서는 이 문제가 해결되어야 한다.

분리된 신호의 주파수 빈의 전력비는 식 (3)과 같이 정의된다[10].

$$\begin{aligned} \mathbf{B}(\omega) &= \mathbf{W}^{-1}(\omega) = [\mathbf{B}_1(\omega), \mathbf{B}_2(\omega) \cdots \mathbf{B}_N(\omega)] \\ P_i(\omega, \tau) &= \frac{\|\mathbf{B}_i(\omega) Y_i(\omega, \tau)\|^2}{\sum_{j=1}^N \|\mathbf{B}_j(\omega) Y_j(\omega, \tau)\|^2}, \quad (i = 1, \dots, N) \end{aligned} \quad (3)$$

여기서 $\mathbf{B}_i(\omega)$ 는 $\mathbf{B}(\omega)$ 의 i 번째 열이고, $Y_i(\omega, \tau)$ 는 FD-ICA의 i 번째 출력이다 ($\mathbf{Y}(\omega, \tau) = [Y_1(\omega, \tau), \dots, Y_N(\omega, \tau)]^T$).

$\mathbf{X}(\omega, \tau) = \mathbf{W}^{-1}(\omega) \mathbf{Y}(\omega, \tau) = \sum_i \mathbf{B}_i(\omega) Y_i(\omega, \tau)$ 이므로,

$\|\mathbf{B}_i(\omega) Y_i(\omega, \tau)\|^2$ 는 $\mathbf{X}(\omega, \tau)$ 에서 i 번째 출력 $Y_i(\omega, \tau)$ 의 파워로 된다. 전력비 $P_i(\omega, \tau)$ 는 실수 값을 가지며, $0 \leq P_i(\omega, \tau) \leq 1$, $\sum_i P_i(\omega, \tau) = 1$ 인 특성을 가지고 있다.

그림 1은 $N=2$, $\mathbf{B}(\omega) = \mathbf{H}(\omega)$, $\mathbf{Y}(\omega, \tau) = \mathbf{S}(\omega, \tau)$ 로 가정한 이상적인 경우의 인접한 두 주파수 빈의 전력비 $P_i(\omega-1)$ 과 $P_j(\omega)$ 에 대한 산포도(scatter plot)를 보인 것이다. 동일한 음원의 인접한 두 주파수 빈 사이에는 상관도가 높으므로 $P_1(\omega-1)$ 대 $P_1(\omega)$ 는 그림 1(a)와 같이 양의 방향으로 주대각선이 놓이게 되며, $P_1(\omega, \tau) + P_2(\omega, \tau) = 1$ 이므로 다른 음원의 인접한 두 주파수 빈, 즉 $P_1(\omega-1)$ 대 $P_2(\omega)$ 는 그림 1(b)에서와 같이 반대방향으로 주대각선이 놓이게 됨을 볼 수 있다.

2.2 다채널 FD-ICA에서의 주파수 빈 정렬

본 논문에서는 동일한 음원의 두 인접한 주파수 빈 전력비의 상관계수는 큰 값을 가지고, 서로 다른 음원의 두 인접한 주파수 빈 전력비의 상관계수는 작은 값 혹은 음의 값을 가지게 되는 특성을 이용하여 전력비의 공분산 행렬을 이용하여 FD-ICA에서 주파수 빈의 뒤섞임 문제를 해결할 수 있는 알고

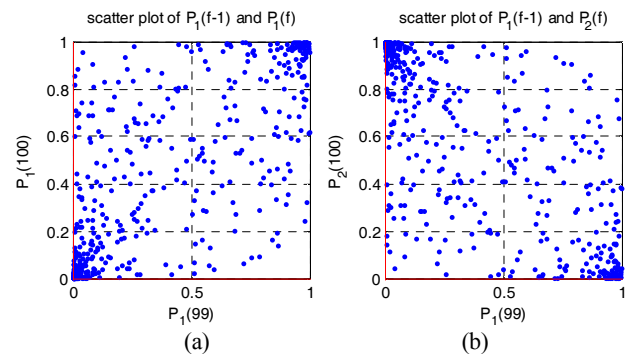


그림 1. 이상적으로 분리된 음원의 인접한 주파수 빈 사이의 전력비 산포도

(a) $P_1(\omega-1)$ 대 $P_1(\omega)$, (b) $P_1(\omega-1)$ 대 $P_2(\omega)$

Figure 1. Scatter plot of power ratio of adjacent frequency bin in an ideally separated case

(a) $P_1(\omega-1)$ vs. $P_1(\omega)$, (b) $P_1(\omega-1)$ vs. $P_2(\omega)$

리즘을 제안하고 실험을 통해 검증하였다. 제안한 알고리즘을 요약하면 다음과 같다.

1) 전력비 정규화

$$\widehat{P}_i(\omega) = [P_i(\omega) - \mu_i(\omega)] / \sigma_i(\omega) \quad (4)$$

여기서 $P_i(\omega) = [P_i(\omega, 1), \dots, P_i(\omega, L)]$ 는 행 벡터로서 L 개 관측샘플(프레임)을 가지는 랜덤 변수로 볼 수 있으며, $\mu_i(\omega), \sigma_i(\omega)$ 는 L 개 관측샘플의 평균과 표준편차를 뜻한다.

2) 공분산 행렬

정규화된 전력비의 인접한 주파수 빈의 공분산 행렬은 식 (5)와 같이 계산한다.

$$C(\omega-1, \omega) = \widehat{P}(\omega-1) \widehat{P}^T(\omega) / (L-1) \quad (5)$$

여기서 $\widehat{P}(\omega) = [\widehat{P}_1^T(\omega), \dots, \widehat{P}_N^T(\omega)]^T$, $N \times L$ 행렬이다.

3) 로컬 치환행렬(local permutation matrix)구성

공분산 행렬의 각 행에서, 가장 큰 값을 가지는 열, 즉 $(\omega-1)$ 주파수 빈과 상관도가 가장 큰 ω 주파수 빈은 동일한 음원에서 분리된 것으로 볼 수 있으므로 그 위치에 1을 놓고 행의 기타 열에는 0을 놓는다. 이렇게 0과 1로 구성된 행렬은 로컬 치환행렬 $\mathbf{u}(\omega)$ 로 된다. $N=3$ 의 경우를 예로 들면 아래와 같다.

$$C(\omega-1, \omega) = \begin{pmatrix} -0.46 & 0.70 & -0.29 \\ 0.68 & -0.52 & -0.20 \\ -0.25 & -0.24 & 0.59 \end{pmatrix} \Rightarrow \mathbf{u}(\omega) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$\mathbf{u}(\omega)$ 에 분리행렬을 곱하면, 즉, $\widetilde{W}(\omega) = \mathbf{u}(\omega) W(\omega)$ 구하면 FD-ICA 출력신호에서 $(\omega-1)$ 주파수 빈과 ω 주파수 빈이 동일한 음원에서 얻어진 것들로 정렬할 수 있게 된다.

동일한 음원 일지라도 FD-ICA 알고리즘이 혼합신호를 잘 분리하지 못할 때 공분산 행렬 각 행의 최대값이 작은 값을 가지는 경우도 발생할 수 있는데, 이때 아래 경우와 같이 한 열에 1이 여러 개 나타나게 되어 치환행렬로 사용할 수 없게 된다.

$$C(\omega-1, \omega) = \begin{pmatrix} 0.21 & -0.21 & -0.04 \\ -0.09 & 0.10 & -0.01 \\ -0.20 & 0.19 & 0.05 \end{pmatrix} \Rightarrow \mathbf{u}(\omega) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

이 경우, 신뢰도를 따져서 가장 가능한 로컬 치환행렬을 구성한다. 본 논문에서는 각 행의 최대값과 두 번째로 큰 값의 차로 신뢰도를 정의한다. 신뢰도가 큰 순서로 1의 위치를 확정하여 각 행과 열에 1이 하나만 놓이도록 로컬 치환행렬을 만든다. 위 예에서 신뢰도 $r(\omega)$ 와 최종적으로 얻어진 로컬 치환행렬은 다음과 같다.

$$r(\omega) = \begin{pmatrix} 0.21+0.04 \\ 0.10+0.01 \\ 0.19-0.05 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.11 \\ 0.14 \end{pmatrix} \Rightarrow \mathbf{u}(\omega) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

4) 주파수 빈 정렬

$\mathbf{u}(\omega)$ 는 $(\omega-1)$ 주파수 빈에 대응된 ω 주파수 빈의 로컬 치환행렬이므로, 첫 번째 주파수 빈에 대응된 ω 주파수 빈의 치환행렬 $U(\omega)$ 는 식 (6)과 같이 표현되고, 최종적으로 주파수 빈 뒤섞임 정정은 식 (7)을 통해 이루어진다.

$$U(\omega) = \prod_{\omega_k=1}^{\omega-1} \mathbf{u}(\omega_k) \quad (6)$$

$$\widetilde{W}(\omega) = U(\omega) W(\omega) \quad (7)$$

$$Y(\omega, \tau) = \widetilde{W}(\omega) X(\omega, \tau)$$

3. 실험 및 검토

제안한 방법의 성능을 검증하기 위하여 image method toolbox[11]로 룸 임펄스 응답을 생성하고 음성신호와 컨볼루션 하여 마이크로폰 신호를 생성하였다. 잔향이 없는 환경에서 그림 2와 같은 음원 및 마이크로폰의 위치를 가정하고, 마이크로폰 간격은 0.1m, 음원은 마이크로폰 배열에 평행으로 2m 거리를 두고 혼합신호를 생성하였다.

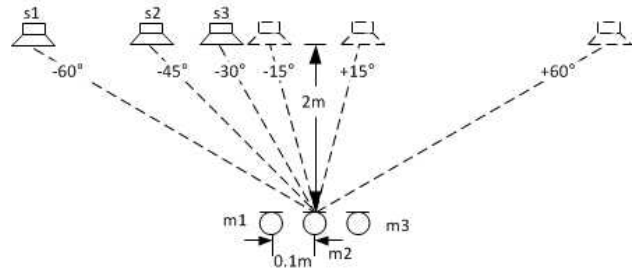


그림 2. 시뮬레이션 환경의 마이크로폰 배열
Figure 2. Microphone array for simulation

표 1. 음원 위치

Table 1. Source position

	-60°	-45°	-30°	-15°	+15°	+60°
위치 1	S1	S2	S3			
위치 2	S1	S2			S3	
위치 3			S1	S2		S3
위치 4			S1	S2	S3	

표 2. 음원 조합

Table 2. Sound source combination

음원 조합 A	female1, female2, female3
음원 조합 B	male1, male2, male3
음원 조합 C	female1, female2, male1
음원 조합 D	female3, male1, male2
음원 조합 E	female2, female3, male3
음원 조합 F	female2, male2, male3

음원으로 사용된 음성신호는 16kHz 샘플링률로 수집한 각각 8초 길이의 여성화자/남성화자의 음성신호를 표 1의 위치에서 표 2와 같이 6가지로 조합하여 실험하였다. FD-ICA 실행 시 STFT 분석 프레임의 길이는 512샘플로 하고 75% 중첩하면서 이동함에 따라 $L=1032$ 프레임이 생성되며, 분리행렬 $W(\omega)$ 의 추정 및 갱신 과정에서의 $\eta=0.1$, 반복(iteration) 차수는 1000으로 설정하고, 스케일링 모호성은 minimal distortion principle을 적용하였다.

그림 3에 FD-ICA로 분리된 각 신호에 대해 인접한 주파수 빈 전력비의 공분산 행렬 값을 STFT의 주파수 빈에 대해 보인다. 분리된 신호의 공분산 행렬 각 행에서 대체로 C11, C22, C33에 해당되는 한 열만 큰 값을 가지고 나머지 두 열은 작

은 값을 가지거나 음의 값을 가짐을 볼 수 있다. 그런데 그림 3(a)에서는 C11, C22, C33의 값이 항상 크지 않고 많은 변화를 보이고 있는데, 이것은 FD-ICA의 주파수 빈 뒤섞임 문제로 인해 인접한 주파수 빈과의 상관관계가 낮음을 나타내는 것이다. 반면에 그림 3(b) 본 논문에서 제안한 주파수 빈 정렬 알고리즘을 적용하여 분리된 신호의 뒤섞임 주파수 빈에서 동일한 음원의 주파수 빈을 갖도록 수정한 결과를 보인 것으로, C11, C22, C33의 값이 나머지 두 열의 값에 비해 안정적으로 큰 값을 가지면서 바르게 수정되었음을 알 수 있다.

분리된 신호의 음질을 객관적 음질평가 척도인 PESQ[12]를 이용하여 분석하였다. 그림 4에 다양한 음원 위치 및 신호 조합에 따른 24개 경우의 FD-ICA 수행 결과에 대한 PESQ 값을

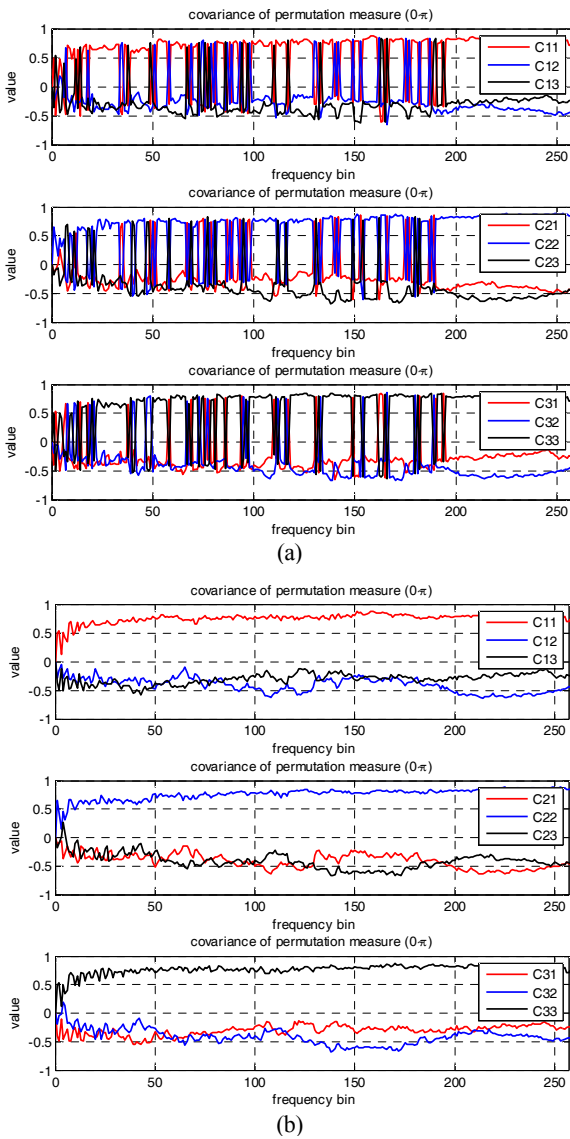


그림 3. 분리된 신호의 인접한 주파수 빈의 공분산 행렬 값 (a) 주파수 빈 정렬 전, (b) 주파수 빈 정렬 후
Figure 3. Covariance matrix of adjacent frequency bins for separated signals (a) before frequency bin alignment, (b) after frequency bin alignment

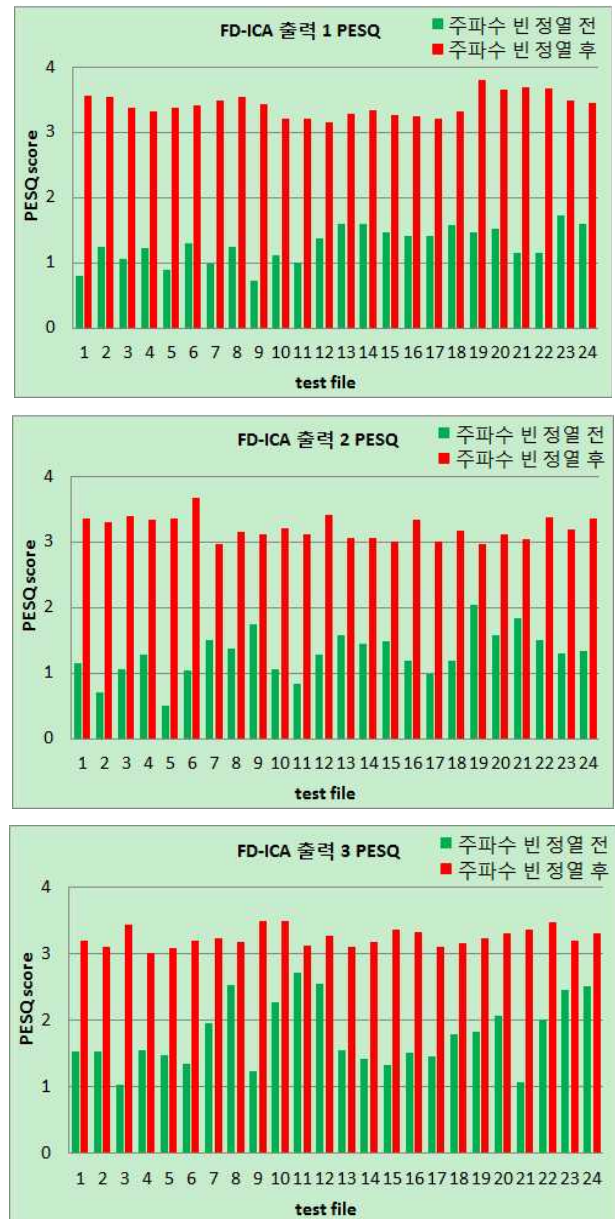


그림 4. 분리된 신호의 PESQ 값
Figure 4. PESQ of separated signals

보인 것이며, 각 분리된 신호의 PESQ 평균값을 표 3에 나타내었다. 주파수 빈을 정렬하기 전에는 FD-ICA의 주파수 빈 뒤섞임 문제로 인해 평균 PESQ 값이 모두 1.8보다 작는데, 본 논문에서 제안한 방법으로 주파수 빈을 정렬 한 후에는 3.2 이상으로 혼합신호가 바르게 분리되었음을 알 수 있다.

표 3. 분리된 혼합신호의 평균 PESQ
Table 3. Average PESQ of separated signals

	주파수 빈 정렬 전			주파수 빈 정렬 후		
	출력1	출력2	출력3	출력1	출력2	출력3
평균 PESQ	1.28	1.29	1.77	3.42	3.21	3.25

4. 결론

다채널 주파수 영역 독립성분 분석 기법을 이용하여 혼합 신호를 분리하고자 할 때 치환 모호성으로 인한 주파수 빈 뒤섞임 현상으로 인해 분리된 신호의 음질이 떨어지게 된다. 본 논문에서는 음성신호에서 동일한 음원의 인접한 주파수 빈의 전력비는 상관도가 높고, 서로 다른 음원과의 인접한 주파수 빈의 전력비는 상관도가 낮거나 음의 값을 가지게 되는 특성을 이용하여, 인접한 주파수 빈의 전력비 공분산 행렬을 이용한 다채널 FD-ICA의 주파수 빈 정렬 방법을 제안하고, 실험을 통해 제안한 방법의 타당성을 검증하였다

참고문헌

[1] Haykin, S. (2009). *Neural networks and learning machines*, USA: Pearson Prentice Hall.

[2] Cichocki, A and Amari, S. (2002). *Adaptive blind signal and image processing (learning algorithms and applications)*, New York: John Wiley & Sons.

[3] Matsuoka, K (2002). Minimal distortion principle for blind source separation, *Proceedings of the 41st SICE Annual Conference*. Vol. 4, pp. 2138-2143.

[4] Lee, T. W. & Bell, A. J. & Lambert, R. H. (1997). Blind separation of delayed and convolved sources, *Advances in neural information processing systems*, pp. 758-764.

[5] Sawada, H. & Mukai, R. & Araki, S. & Makino, S. (2001). A polar-coordinate based activation function for frequency domain blind source separation, *Int. Conf. ICA and BSS*, pp. 663-668.

[6] Sawada, H. & Araki, S. & R. Mukai, R. & Makino, S. (2007). Grouping separated frequency components by estimating

propagation model parameters in frequency-domain blind source separation, *Audio, Speech, and Language Processing, IEEE Transactions on*. Vol. 15, No. 5, pp. 1592-1604.

[7] Kim, T. & Attias, H. T. & Lee, S. Y. & Lee, T. W. (2007). Blind source separation exploiting higher-order frequency dependencies, *Audio, Speech, and Language Processing, IEEE Transactions on*. Vol. 15, pp. 70-79.

[8] Chu, Z. H. & Bae, K. S. (2013). Post-processing of IVA-based 2-channel blind source separation for solving the frequency bin permutation problem. *Malsori*, Vol. 5, No. 4, 211-216 (추쓰하오, 배건성 (2013). IVA 기반의 2채널 암묵적신호분리에서 주파수빈 뒤섞임 문제 해결을 위한 후처리 과정. 말소리 제5권 제4호, 211-216)

[9] Murata, N. & Ikeda, S. & Ziehe, A. (2001). An approach to blind source separation based on temporal structure of speech signals, *Neurocomputing*. Vol. 41, pp. 1-24.

[10] Sawada, H. & Araki, S. & Makino, S. (2007). Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS, *ISCAS 2007. IEEE International Symposium on*. pp. 3247-3250.

[11] Donohue, K. D. (2009) Audio Array Toolbox. <http://www.engr.uky.edu/~donohue/audio/Arrays/MATtoolbox.htm>

[12] ITU-T. (2001) PESQ. <http://www.itu.int/rec/T-REC-P.862/en>

- **전성일 (Quan, Xingri)**
경북대학교 IT 대학 전자공학부
대구 북구 산격3동 1370번지
Tel: 053-940-8627
Email: comet263@hotmail.com
관심 분야: 마이크로폰 배열 음성신호처리, adaptive filter
2010~현재 경북대학교 IT 대학 전자공학부 박사과정
- **배건성 (Bae, Keunsung)**
경북대학교 IT 대학 전자공학부
대구 북구 산격3동 1370번지
Tel: 053-940-8627
Email: ksbae@mir.knu.ac.kr
관심 분야: 음성신호처리, 디지털신호처리
현재 경북대학교 IT 대학 전자공학부 교수