유사-가능도 최대화를 통한 가우시안 프로세스 기반 음원분리

(Gaussian Processes for Source Separation: Pseudo-likelihood Maximization)

박 선 호 † 최 승 진 ††

(Sunho Park) (Seungjin Choi)

요 약 본 논문에서는 각 음원이 시간적 구조를 가졌을 경우 음원들을 분리해내는 확률적 음원분리 방법을 제안한다. 이를 위해 각 음원의 시간적 구조를 가우시안 프로세스(Gaussian process)로 모델링하고 기존의 음원분리 문제를 유사-가능도 최대화 문제(pseudo-likelihood maximization)로 공식화한다. 본 알고리즘을 통해 얻어진 데이타의 유사-가능도는 정규 분포이며 이는 가우시안 프로세스 회귀방법(Gaussian process regression)을 통해 쉽게 계산이 가능하다. 음원분리의 역혼합 행렬은 경도(gradient) 기반 최적화 기법을 통해 데이타의 유사-가능도를 최대화 하는 해를 찾음으로써 구해진다. 여러 실험을 통하여 제안 알고리듬이 몇 가지 특정 상황에서 기존의 분리 알고리듬들에 비해 우수한 성능을 보임을 확인 할 수 있다.

키워드 : 음원분리, 음원의 시간적 구조, 가우시안 프로세스, 최대 유사-가능도

Abstract In this paper we present a probabilistic method for source separation in the case where each source has a certain temporal structure. We tackle the problem of source separation by maximum pseudo-likelihood estimation, representing the latent function which characterizes the temporal structure of each source by a random process with a Gaussian prior. The resulting pseudo-likelihood of the data is Gaussian, determined by a mixing matrix as well as by the predictive mean and covariance matrix that can easily be computed by Gaussian process (GP) regression. Gradient-based optimization is applied to estimate the demixing matrix through maximizing the log-pseudo-likelihood of the data. umerical experiments confirm the useful behavior of our method, compared to existing source separation methods.

Key words : source separation, temporal structure of source, Gaussian process, maximum pseudo-likelihood

1. Introduction

음원분리는 신호처리와 기계학습 분야의 근원적인 문제들 중 하나이다. 음원분리에서 관측 데이타 $\boldsymbol{x}_t = [x_{1,t}, \cdots, x_{n,t}]^\top$ ($x_{i,t}$는 $\boldsymbol{x}_t \in R^n$의 i번째 원소를 의미함)는 모델 (1)에 의해 생성된다고 가정한다.

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t, \tag{1}$$

여기서 $\boldsymbol{A} \in R^{n \times n}$는 혼합행렬이고 서로 통계적으로 독립적인 음원 벡터는 $\boldsymbol{s}_t = [s_{1,t}, \ldots, s_{n,t}]^\top$로 표현된다. 음원분리의 목적은 혼합행렬 $\boldsymbol{A}$에 대한 어떠한 정보도 없이 혼합행렬의 역변환인 역혼합행렬 $\boldsymbol{W} = \boldsymbol{A}^{-1}$을 예측하여 원래의 음원신호 $\boldsymbol{s}_t$를 복원하는 것이다. 그러나 크기와 교환 모호성에 의해 원 음원신호의 정확한 크기와

· 본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원 사업의 연구결과로 수행되었음(IITA-2008-C1090-0801-0045)
· 이 논문은 제34회 추계학술대회에서 '유사-가능도 최대화를 통한 가우시안 프로세스 기반 음원분리'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 포항공과대학 컴퓨터공학과
titan@postech.ac.kr
†† 종신회원 : 포항공과대학 컴퓨터공학과 교수
seungjin@postech.ac.kr

논문접수 : 2007년 12월 3일
심사완료 : 2008년 6월 2일

Copyright@2008 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제35권 제7호(2008.7)

순서를 알 수 없으므로 혼합행렬의 정확한 역변환이 아닌 보다 완화된 조건들을 만족하는 역혼합행렬을 구하게 된다. 즉, $\boldsymbol{A}=\boldsymbol{P\Lambda}$, 여기서 $\boldsymbol{P}$는 교환행렬이고 $\boldsymbol{\Lambda}$는 임의의 역변환이 가능한 대각행렬이다. 본 논문에서는 각 음원신호가 시간적 구조를 가졌을 경우 역혼합행렬 $\boldsymbol{W}$을 구하는 효과적인 알고리듬을 제안한다.

음원분리를 수행하기 위해 다양한 방법들이 제안되어 왔다([1]과 수록된 참고문헌을 참고). 가장 기본적인 예들로 독립성분분석(ICA)에 속하는 Infomax[2]와 FastICA [3]을 들 수 있는데, 이 방법들은 각 음원이 비정규분포를 따른다고 가정하고 음원의 공간적 독립성만을 이용하여 음원들을 분리하는 방법들이다. 사실 Infomax는 각 음원신호를 잠재변수로 가정하고 이를 marginalization out한 뒤 최대 가능도(maximum likelihood)를 이용하여 분리하는 방법으로 이해할 수 있다[4]. 또한 음원들이 시간적 구조를 가졌을 경우 2차 통계량을 통해 분리하는 방법들이 제안되었는데, SOBI가 이에 해당하는 대표적인 방법이다. SOBI[5]는 일련의 시간-지연 상관 행렬들(time-delayed correlation matrices)을 동시에 대각화함으로써 역혼합행렬을 구하는 방법을 취한다. 또한 음원의 시간적 구조를 모수 형태의 잠재함수로 모델링하여 음원분리를 수행하는 음원생성모델 기반 분리 방법들이 있다. 많은 방법들이 잠재함수로 선형 자기회귀(auto-regressive, 앞으로 AR표기)모델을 사용하여 음원의 시간적 구조를 표현하고 있다[6-8]. 이 경우 음원분리를 위해 혼합행렬이나 역혼합행렬을 계산 할 때 AR 음원생성모델의 계수들도 동시에 계산되어야 한다.

가우시안 프로세스(Gaussian process, 앞으로 GP로 표기) 모델은 최근 기계학습의 여러 분야에서 널리 쓰여지기 시작한 방법으로 회귀와 분류 문제 등에서 내재된 비선형 함수를 예측하는 데 사용된다([9,10]와 수록된 참고문헌 참고). GP 모델은 대표적인 비모수 방법으로, 비모수 방법의 유연성을 가지면서 결과 도출을 위한 계산이 단순하고 동시에 확률적 해석이 가능하다는 장점들을 가진다. 본 논문에서는 GP 모델을 이용하여 음원의 시간적 구조를 표현하는 잠재함수를 정의한다(이에 대한 최근의 참고문헌[11,12]). 이 잠재함수는 현재 음원신호와 과거 음원신호들 사이에 정의되는 매우 일반적인(비모수 형태의) 비선형함수이다. 그러나 주지해야 할 사실은 음원분리의 최종 목표는 정확한 잠재함수의 예측이 아닌 혼합행렬(또는 역혼합행렬)의 계산이라는 점이다. 그러므로 확률적 해석이 가능한 GP 모델을 이용하여 잠재함수의 사후확률분포를 구한 뒤, margianlization을 통해 잠재함수를 제거하고 혼합행렬(또는 역혼합행렬)만으로 정의되는 목적식을 얻는다. 이를 위해서는 잠재함수의 사후확률분포에 대한 적분계산이 필요한데 GP 모델을 이용할 경우 이 계산이 해석적으로 가능하다는 이점이 있다. 위의 방법을 통하여 최종적으로 유사-가능도(pseudo-likelihood)를 얻을 수 있다. 이와 같은 유사-가능도는 GP 회귀의 결과물인 예측 평균(predictive mean)과 예측 상관행렬(predictive covariance matrix)에 의해서 쉽게 구해지며, 최종 목표인 역혼합행렬은 전체 데이타의 로그-유사-가능도(log-pseudo-likelihood)에 최적화 기법을 적용함으로써 예측이 가능하다. 본 제안 알고리듬이 가지는 장점으로는 비모수 접근방법의 유연성에 의해 비선형 시계열 데이타를 쉽게 처리할 수 있고 분리성능이 알고리듬의 내부 변수인 모델 차수의 영향을 적게 받는다는 점이다. 또한 제안 알고리듬은 음원들이 비슷한 파워 스펙트라를 가질 경우에도 불안정한 특성을 보이는 SOBI와는 달리 좋은 성능을 보인다.

본 논문의 순서는 다음과 같다. Sec. 2는 GP 모델을 이용한 음원생성모델을 소개하고, Sec. 3은 데이타의 로그-유사-가능도의 최적화를 통하여 역혼합행렬를 구하는 본 알고리듬을 설명한다. 각 알고리듬의 음원 분리 성능평가를 위한 실험 수행은 Sec. 4에서 제시된다. 마지막으로 Sec. 5에서 결과를 도출하고 제안 알고리듬의 유용한 특성들에 대해 강조한다.

2. GP Source Generative Model

각 음원의 시간적 구조를 이용하기 위해 음원신호 $s_{i,t}$를 (2)와 같이 모델링한다.

$$s_{i,t}=f_i(\vec{\boldsymbol{s}}_{i,t-1})+\epsilon_{i,t}, \tag{2}$$

여기서 $\vec{\boldsymbol{s}}_{i,t-1}\in R^p$는 과거 p(모델 차수)개의 음원신호로 만들어진 벡터이다,

$$\vec{\boldsymbol{s}}_{i,t-1}=[s_{i,t-1},s_{i,t-2},...,s_{i,t-p}]^\top.$$

모델에 대한 확률적 불확실도는 $\varepsilon_{i,t}$에 의해 표현되며 이는 제로 평균과 단일 분산을 가지는 백색정규분포로 정의 된다. 즉,

$$\varepsilon_{i,t}\sim N(0,1). \tag{3}$$

여기서 함수 $f_i(\cdot)$는 잠재함수로 현재 음원신호 $s_{i,t}$와 과거 신호 값들 $\vec{\boldsymbol{s}}_{i,t-1}$사이에 정의된 함수이다. 예를 들어 선형 AR 모델의 경우, 잠재함수는 다음과 같이 표현된다.

$$f_i(\vec{\boldsymbol{s}}_{i,t-1})=\sum_{\tau=1}^{p}h_{i,\tau}s_{i,t-\tau}, \tag{4}$$

여기서 $\{h_{i,\tau}\}$은 AR 계수들이다.

GP 모델은 대표적인 비모수 방법으로 잠재함수 $f_i(\cdot)$는 GP 사전확률을 가지는 랜덤 프로세스로 표현된다. GP는 모수형태를 띄는 (4)와는 달리 어떠한 특정

형태도 가정하지 않고 함수 값 $f_i(\vec{s}_{i,t})$ 자체를 확률변수로 설정한다. 이는 잠재함수 $f_i(\cdot)$에 GP 사전확률을 가정하는 것으로 표현 될 수 있는데 [10]. 즉,

$$f_i \sim GP(0, k(\vec{s}_{i,t}, \vec{s}_{i,\tau})), \tag{5}$$

여기서 $k(\vec{s}_{i,t}, \vec{s}_{i,\tau})$는 공분산함수로 (covariance function) 보통 커널(kernel) 함수가 사용된다[10]. 본 논문에서는 가우시안 커널 함수 (6)를 공분산 함수로 사용한다.

$$k(\vec{s}_{i,t}, \vec{s}_{i,\tau}) = \exp(-\lambda_i |\vec{\mathbf{s}}_{\mathrm{i,t}} - \vec{\mathbf{s}}_{\mathrm{i,\tau}}|^2), \tag{6}$$

여기서 λ_i는 길이조절 초매개변수(length-scale hyper-parameter)이다.

앞서 정의된 GP 기반 음원생성모델은 표준적인 GP 회귀모델로 이해될 수 있다. 즉, $\mathbf{s}_{i,1:N}^{\top} = [s_{i,1}, \cdots, \mathrm{s}_{\mathrm{i,N}}]^{\top}$는 응답변수들의 모음이며 $S_i = \{\vec{\mathbf{s}}_{\mathrm{i,t-1}}\}_{\mathrm{t=1}}^{\mathrm{N}}$는 입력변수들의 집합에 해당한다. 표준적인 GP 회귀모델으로부터 새로운 입력 값 $\vec{s}_{i,*}$에 대한 예측분포($f_i^* = f_i(\vec{s}_{i,*})$)는 (7)과 같이 구해진다.

$$p(f_i^*, |\mathbf{s}_{\mathrm{i,1:N}}^{\top}, \mathrm{S_i}, \vec{\mathbf{s}}_{\mathrm{i,*}}) = \mathrm{N}(\bar{\mathrm{f}}_{\mathrm{i}}^*, \mathrm{var}(\mathrm{f}_{\mathrm{i}}^*)), \tag{7}$$

여기서 평균 $\bar{f}_i^*$과 분산 $var(f_i^*)$은 다음과 같이 결정되며

$$\bar{f}_i^* = [\boldsymbol{k}_i(\vec{s}_{i,*})]^{\top} \boldsymbol{K}_i^{-1} \mathbf{s}_{i,1:N}^{\top}, \tag{8,9}$$
$$var(f_i^*) = k(\vec{s}_{i,*}, \vec{s}_{i,*}) - [\boldsymbol{k}_i(\vec{s}_{i,*})]^{\top} \boldsymbol{K}_i^{-1} \boldsymbol{k}_i(\vec{s}_{i,*}),$$

$\boldsymbol{I}$는 적절한 차원의 단위행렬이고 $\boldsymbol{K}_i$는 $N \times N$ 행렬로써 (j,k)원소는 아래와 같은 값을 갖는다.

$$[\boldsymbol{K}_i]_{u,v} = k(\vec{s}_{i,u-1}, \vec{s}_{i,v-1}) + \delta_{u,v}, \tag{10}$$

여기서 $\delta_{u,v}$는 크로네커 델타(Kronecker delta)이고 $\boldsymbol{k}_i(\vec{s}_{i,*})$는 N차원의 벡터이다:

$$\boldsymbol{k}_i(\vec{s}_{i,*}) = [k(\vec{s}_{i,0}, \vec{s}_{i,*}), \cdots, k(\vec{s}_{i,N-1}, \vec{s}_{i,*})]^{\top}. \tag{11}$$

n개의 GP 사전 확률 (5)에 기반 한 음원생성모델(2)들을 종합하여 아래와 같은 관측모델 $\boldsymbol{x}_t$를 얻을 수 있다.

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{f}_{t-1} + \boldsymbol{A}\boldsymbol{\varepsilon}_t. \tag{12}$$

여기서 잠재벡터(latent vector) $\boldsymbol{f}_{t-1} \in \mathrm{R}^{\mathrm{n}}$는 아래와 같이 주어진다,

$$\boldsymbol{f}_{t-1} = [f_{1,t-1}, \cdots, \mathrm{f}_{\mathrm{n,t-1}}]^{\top}. \tag{13}$$

위 식에서 $f_{i,t-1} = f_i(\vec{s}_{i,t-1})$이고 $\boldsymbol{\varepsilon}_t \in \mathrm{R}^{\mathrm{n}}$는 독립적인 정규분포를 따른다,

$$p(\boldsymbol{\varepsilon}_t) = N(0, I).$$

관측모델 (12)는 본 알고리듬의 중요한 부분으로써, 다음 장에서 전체 데이타의 로그-유사-가능도를 유도하는 핵심식이다.

3. Algorithm: Pseudo-likelihood Maximization

최대 가능도 음원분리에서 음원신호는 잠재변수로써 결국에는 marginalized out된다. 이와 유사하게 잠재벡터 $\boldsymbol{f}_{t-1}$는 marginalized out되어 제거지고 오직 혼합행렬(또는 역혼합행력)에 의해 결정되는 최대 유사-가능도 식을 얻을 수 있다. 이를 위하여 데이타의 단일단위 유사-가능도는 (14)와 같이 정의되며,

$$p(\boldsymbol{x}_t | \boldsymbol{X}^{(-t)}) = \int p(\boldsymbol{x}_t | \boldsymbol{f}_{t-1}, \boldsymbol{X}^{(-t)}) p(\boldsymbol{f}_{t-1} | \mathbf{X}^{(-\mathrm{t})}) \mathrm{d}\mathbf{f}_{\mathrm{t-1}}, \tag{14}$$

$\boldsymbol{X}^{(-t)} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t+1}, \ldots, \boldsymbol{x}_N\}$는 $\boldsymbol{x}_t$를 제외한 모든 관측데이타를 의미한다:

$$p(\boldsymbol{x}_t | \boldsymbol{f}_{t-1}, \boldsymbol{X}^{(-t)}) = \mathrm{N}(\mathbf{A}\mathbf{f}_{\mathrm{t-1}}, \mathbf{A}\mathbf{A}^{\top}), \tag{15}$$

$$p(\boldsymbol{f}_{t-1} | \mathbf{X}^{(-\mathrm{t})}) = \mathrm{N}(\boldsymbol{\mu}_{\mathrm{t}}, \boldsymbol{\Sigma}_{\mathrm{t}}). \tag{16}$$

또한 예측 평균 벡터($\boldsymbol{\mu}_t \in \mathrm{R}^{\mathrm{n}}$)와 대각 공분산 행렬 $\boldsymbol{\Sigma}_t \in \mathrm{R}^{\mathrm{n} \times \mathrm{n}}$는 (8)과 (9)를 이용하여 계산되는데,

$$\mu_{i,t} = \left[\boldsymbol{k}_i^{(-t)}(\vec{s}_{i,t-1})\right]^{\top} \left[\boldsymbol{K}_i^{(-t)}\right]^{-1} [\mathbf{s}_{i,1:N}^{(-t)}]^{\top}, \tag{17-19}$$
$$\boldsymbol{\Sigma}_t = diag(\sigma_{1,t}^2, \ldots, \sigma_{\mathrm{n},t}^2),$$
$$\sigma_{i,t}^2 = k(\vec{s}_{i,t-1}, \vec{s}_{i,t-1}) - \left[\boldsymbol{k}_i^{(-t)}(\vec{s}_{i,t-1})\right]^{\top} \left[\mathbf{K}_{\mathrm{i}}^{(-\mathrm{t})}\right]^{-1} \mathbf{k}_{\mathrm{i}}^{(-\mathrm{t})}(\vec{\mathbf{s}}_{\mathrm{i,t-1}}),$$

여기서 어깨글자(super-script) (-t)는 원래의 벡터에서 t번째 원소를 제거한 것을 의미한다. 즉,

$$\boldsymbol{k}_i^{(-t)}(\vec{s}_{i,t-1}) = [k(\vec{s}_{i,0}, \vec{s}_{i,t-1}), \cdots, k(\vec{s}_{i,t-2}, \vec{s}_{i,t-1}), k(\vec{s}_{i,t}, \vec{s}_{i,t-1}), \cdots, k(\vec{s}_{i,N-1}, \vec{s}_{i,t-1})]^{\top},$$
$$\mathbf{s}_{i,1:N}^{(-t)} = [s_{i,1}, \cdots, s_{i,t-1}, s_{i,t+1}, \cdots, s_{i,N}].$$

그리고 이와 유사하게 $\boldsymbol{K}_i^{(-t)}$는 행렬 $\boldsymbol{K}_i$에서 t번째 열과 행이 제거된 부행렬(sub-matrix, 즉 $\boldsymbol{K}_i^{(-t)}$는 $(N-1) \times (N-1)$ 행렬)로 정의되는데, $\mu_{i,t}$와 $\sigma_{i,t}^2$들의 계산은 완전한 공분산 행렬의 분할 역변환(inversion by partitioning) 방법에 의해 효과적으로 구해진다(Ch. 5 in [10] 또는 [13] 참고). 앞의 결과를 종합하여 얻어지는 결과는 다음과 같다.

$$\mu_{i,t} = s_{i,t} - \frac{\left[\boldsymbol{K}_i \mathbf{s}_{i,1:N}^{\top}\right]_t}{\left[\boldsymbol{K}_i^{-1}\right]_{t,t}}, \tag{20,21}$$
$$\sigma_{i,t}^2 = \frac{1}{\left[\boldsymbol{K}_i^{-1}\right]_{t,t}} - var(\varepsilon_{i,t}) = \frac{1}{\left[\boldsymbol{K}_i^{-1}\right]_{\mathrm{t,t}}} - 1.$$

(14), (15)와 (16)부터 단일단위 유사-가능도는 아래의 정규 분포를 따른다.

$$p(\boldsymbol{x}_t | \boldsymbol{X}^{(-t)}) = N(\boldsymbol{A}\boldsymbol{\mu}_t, \boldsymbol{\Gamma}_t), \tag{22}$$

여기서 $\boldsymbol{\Gamma}_t = \boldsymbol{A}(\boldsymbol{\Sigma}_t + \boldsymbol{I})\boldsymbol{A}^{\top}$. 그러므로 우리는 전체 데이타의 원-가능도를 근사화하는 유사-가능도(23)를 얻을 수 있다[14].

$$pseudo-likelihood \doteq \prod_{t=1}^{N} p(\boldsymbol{x}_t|\boldsymbol{X}^{(-t)}). \quad (23)$$

이에 따라 전체 데이타의 로그-유산-가능도는 아래와 같다.

$$L = \sum_{t=1}^{N} \log p(\boldsymbol{x}_t|X_{-t}) \quad (24)$$
$$= -\frac{1}{2}\sum_{t=1}^{N}(\boldsymbol{x}_t - \mathbf{A}\boldsymbol{\mu}_t)^{\top}\boldsymbol{\Gamma}_t^{-1}(\mathbf{x}_t - \mathbf{A}\boldsymbol{\mu}_t)$$
$$-\frac{1}{2}\sum_{t=1}^{N}\log|\boldsymbol{\Gamma}_t| - \frac{1}{2}N\log 2\pi.$$

다음 정의를 통해

$$[\boldsymbol{\alpha}_t]_i = \left[\boldsymbol{K}_i^{-1}\boldsymbol{s}_{i,1:N}^{\top}\right]_t \left[\boldsymbol{K}_i^{-1}\right]_{t,t},$$
$$[\boldsymbol{\beta}_t]_i = \left[\mathbf{K}_i^{-1}\mathbf{s}_{i,1:N}^{\top}\right]_t.$$

로그-유산-가능도(24)는 보다 간단히 표현된다.

$$L = -\frac{1}{2}\sum_{t=1}^{N}\left\{\log 2\pi + \log|\boldsymbol{\Gamma}_t| + \boldsymbol{\alpha}_t^{\top}(\boldsymbol{\Sigma}_t + \boldsymbol{I})^{-1}\boldsymbol{\alpha}_t\right\} \quad (25)$$
$$= -\frac{1}{2}\sum_{t=1}^{N}\left\{\log 2\pi + \log|\boldsymbol{\Gamma}_t| + \boldsymbol{\beta}_t{}^{\top}(\boldsymbol{\Sigma}_t + \mathbf{I})\boldsymbol{\beta}_t\right\}.$$

혼합행렬 $\boldsymbol{A}$나 역혼합행렬 $\boldsymbol{W}$는 로그-유사-가능도(25)를 최대화 하는 해를 구함으로써 구해진다. (25)의 해를 찾기 위해 경도(gradient)기반의 최적화 기법이 적용되며, 매트랩(Matlab)의 fminunc이 사용된다. 이와 같은 경도를 계산하기 위해, 먼저 $\boldsymbol{Z}_i$를 정의하고

$$\boldsymbol{Z}_i{}^{kl} = \boldsymbol{K}_i^{-1}\frac{\partial \boldsymbol{K}_i}{\partial w_{k,l}}\boldsymbol{K}_i^{-1}, \quad (26)$$

여기서 공분산행렬 $\boldsymbol{K}_i$의 $w_{k,l}$(역혼합행렬 $\boldsymbol{W}$의 (k,l) 원소)에 대한 경도는 다음과 같이 계산된다.

$$\left[\frac{\partial \boldsymbol{K}_i}{\partial w_{k,l}}\right]_{u,v} =$$
$$-2\lambda_i k(\vec{\boldsymbol{s}}_{i,u-1}, \vec{\boldsymbol{s}}_{i,v-1})\left[\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\boldsymbol{w}_{i,:}^{\top}\right]_l \delta_{i,k},$$
$$\boldsymbol{\Delta} = \left[(\boldsymbol{x}_{u-1} - \boldsymbol{x}_{v-1}), \ldots, (\boldsymbol{x}_{u-p} - \boldsymbol{x}_{v-p})\right].$$

앞의 결과를 이용하여 (26)의 $w_{k,l}$에 대한 경도는 아래와 같이 구해진다.

$$\frac{\partial L}{\partial w_{k,l}} =$$
$$-\frac{1}{2}\sum_{t=1}^{N} tr\left\{\boldsymbol{\Gamma}_t^{-1}\frac{\partial}{\partial w_{k,l}}\left[\mathbf{W}^{-1}(\boldsymbol{\Sigma}_t + \mathbf{I})\mathbf{W}^{-\top}\right]\right.$$
$$+\left(\frac{\partial}{\partial \boldsymbol{\beta}_t}\left[\boldsymbol{\beta}_t^{\top}(\boldsymbol{\Sigma}_t + \mathbf{I})\boldsymbol{\beta}_t\right]\right)^{\top}\frac{\partial \boldsymbol{\beta}_t}{\partial w_{k,l}}$$
$$\left.+\left(\frac{\partial}{\partial \boldsymbol{\Sigma}_t}\left[\boldsymbol{\beta}_t^{\top}(\boldsymbol{\Sigma}_t + \boldsymbol{I})\boldsymbol{\beta}_t\right]\right)^{\top}\frac{\partial \boldsymbol{\Sigma}_t}{\partial w_{k,l}}\right\}.$$

앞의 식은 아래와 같이 간단하게 표현될 수 있다.

$$\frac{\partial L}{\partial w_{k,l}}$$
$$= \frac{1}{2}\sum_{t=1}^{N} tr\left\{2\boldsymbol{W}^{-1}\frac{\partial \boldsymbol{W}}{\partial w_{k,l}} - \boldsymbol{\Gamma}_t^{-1}\boldsymbol{W}^{-1}\frac{\partial \boldsymbol{\Sigma}_t}{\partial w_{k,l}}\boldsymbol{W}^{-\top}\right.$$
$$\left.-2\boldsymbol{\beta}_t^{\top}(\boldsymbol{\Sigma}_t + \mathbf{I})\frac{\partial \boldsymbol{\beta}_t}{\partial w_{k,l}} - \boldsymbol{\beta}_t^{\top}\frac{\partial \boldsymbol{\Sigma}_t}{\partial w_{k,l}}\boldsymbol{\beta}_t\right\},$$

여기서

$$\frac{\partial \boldsymbol{\beta}_t}{\partial w_{k,l}} = \left[-\boldsymbol{Z}_i^{kl}\boldsymbol{s}_{i,1:N}^{\top} + \boldsymbol{K}_i^{-1}\boldsymbol{x}_{l,1:N}^{\top}\right]_t \delta_{i,k},$$

그리고

$$\left[\frac{\partial \boldsymbol{\Sigma}_t}{\partial w_{k,l}}\right]_{i,i} = \frac{\left[\boldsymbol{Z}_i^{kl}\right]_{t,t}}{\left[\boldsymbol{K}_i^{-1}\right]_{t,t}^2}\delta_{i,k}.$$

원리적으로 초매개변수 λ_i 역시 최대 유사-가능도 방법으로 계산 할 수 있다. 그러나 초-매개변수의 값이 음원의 분리결과에 많은 영향을 끼치지 않는다는 실험결과를 통하여(Fig. 1(a) 참고), 본 논문에서는 λ_i들을 고정된 상수 값으로 가정한다. 이러한 접 방식은 음원분리가 역혼합행렬을 학습만을 통해 이루어질 수 있도록 하여 효과적인 알고리듬 설계를 가능케 한다.

4. Numerical Experiments

본 실험에서는 제안 알고리듬인 GP 음원 분리방법을 기존의 FastICA[3], Infomax[2.15], SOBI[5], 쌍대 AR 모델 기반 음원분리방법(앞으로 AR-BSS 표기)[8] 분리방법들과 비교한다. FastICA와 Infomax는 음원의 시간적 구조에 대한 정보없이 오직 음원의 비가우시안 가정에 의해 분리를 하는 방식을 취한다. 또한 SOBI는 일련의 시간-지연 공분산 행렬들을 동시에 대각화함으로써 역혼합행렬을 예측한다. AR-BSS 방법은 음원의 시간적 구조를 표현하기 위해 모수방법인 선형 AR모델을 이용하고 분리를 하기 위해 혼합행렬과 선형 AR계수를 동시에 계산한다.

본 실험에서는 음원에 대한 두 가지 특수한 상황에서 분리실험을 수행한다:

(1) 음원들이 비슷한 파워 스펙트라를 가진 경우;
(2) 음원들이 복잡한 비선형 특성을 가진 시계열 데이타인 경우.

음원의 갯수에 따라 두 가지 혼합행렬이 사용되었다:

$$(n=2)\quad \boldsymbol{A}_2 = \begin{pmatrix} 1.1 & 0.7 \\ 1.2 & 1.5 \end{pmatrix}$$
$$(n=3)\quad \boldsymbol{A}_3 = \begin{pmatrix} -0.4667, & 2.0636, & -0.5136 \\ 0.0680, & 2.3982, & -0.1961 \\ -2.5108, & 0.3002, & 0.2247 \end{pmatrix}$$

각 알고리듬의 분리 성능평가를 하기위해 아래의 성능지수(performance index, PI)를 사용하였다,

$$PI = \frac{1}{n}\sum_{i=1}^{n}\{(\sum_{k=1}^{n}\frac{|g_{i,k}|^2}{\max_j |g_{i,j}|^2} - 1)$$
$$+ (\sum_{k=1}^{n}\frac{|g_{k,i}|^2}{\max_j |g_{j,i}|^2} - 1)\},$$

여기서 $g_{i,j}$는 전역 변환(global transformation) $\boldsymbol{G} = \boldsymbol{WA}$의 (i,j)원소이다. 만약 모든 음원들이 완벽하게 분리된다면 PI=0 된다. 보통의 경우 PI < 0.005는 뛰어난 분

리 성능을 의미하고, PI < 0.05 일반적인 분리 성능을 나타낸다. 본 실험에서는 각각의 알고리듬에 대해 20번의 독립적인 실험을 수행하고 그 평균을 계산한다.

본 실험에서는 GP 모델의 초매개변수를 고정된 값으로 사용한다(커널 함수의 길이 조절 초매개변수 $\{\lambda_i\}$). 사실 $\{\lambda_i\}$역시 최대 유사-가능도 최적화 방법으로 학습이 가능하지만, 미리 설정한 값을 사용하더라도 분리 성능에 큰 영향을 주지 않는다. 이를 확인하기위해 두 비선형 시계열 데이타를 분리하는 실험은 다음과 같다. 한 음원은 Mackey-Glass MG_{30}이고 다른 음원은 Santa Fe competition Laser인(참고 Fig. 1) 혼합신호의 분리 실험이다. 그림 1(a)는 $\lambda=\lambda_1=\cdots=\lambda_n$가 $[2^2,12^2]$로 변할 때의 안 방법의 분리결과(PI)인데, $\{\lambda_i\}$의 영향이 매우 적음을 알 수 있다. 또한 모델 차수 p 는 현재 신호를 결정하는데 얼마나 많은 과거 신호 값들이 필요하는지를 나타내는 변수인데, 그림 1(b)를 통해 $p \geq 5$일 때 좋은 성능을 보임을 알 수 있다. 앞으로의 모든 실험에서는 $\lambda=50$, p=5로 고정된 값을 사용하였다.

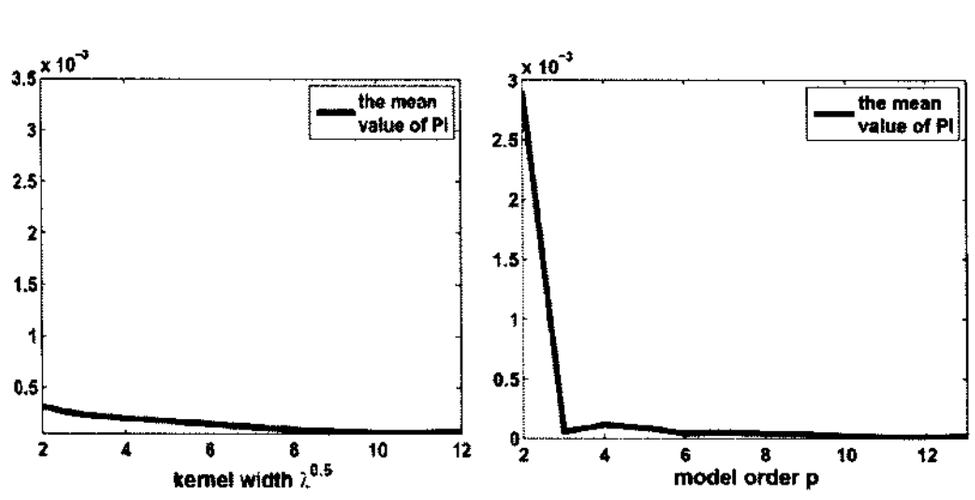

그림 1 본 그림은 λ와 모델 차수 p의 변화에 따른 제안 알고리듬의 분리 성능을 보인다. 그림 (a): 길이-조절 초매개변수의 제곱근 $\sqrt{\lambda}$ 변화에 따른 제안 알고리즘의 출력 PI값 그림(p=5), 그림 (b): 모델 차수 p가 변함 변화에 따른 제안 알고리즘의 출력 PI값 그림 ($\lambda=50$).

Experiment 1

이 예제는 세 음원들에 대한 분리실험이다. 두 음원은 서로 독립적인 왜색가우시안음원이고 나머지 음원은 정규 분포에 가까운 음악 신호이다. 앞의 두 왜색 가운시안 음원은 서로 다른 두개의 AR 모델에 의해 생성되었는데 각각의 AR 계수 $\boldsymbol{h}_i=\{h_{i,\tau}\}$들은 다음과 같다,

$$\boldsymbol{h}_1=\{1.3117,-0.8664,0.5166,-0.2534\}$$
$$\boldsymbol{h}_2=\{0.7838,0.3988,-0.4334,-0.1792\}.$$

이러한 가우시안 음원들에 대해서 FastICA와 Infomax가 제대로 동작하지 못할 것이라는 것은 명확하다. 또한 랜덤하게 생성되어 신호들의 파워 스펙트라가 비슷한 경우 SOBI는 좋은 성능을 보이지 못한다. 그림 2

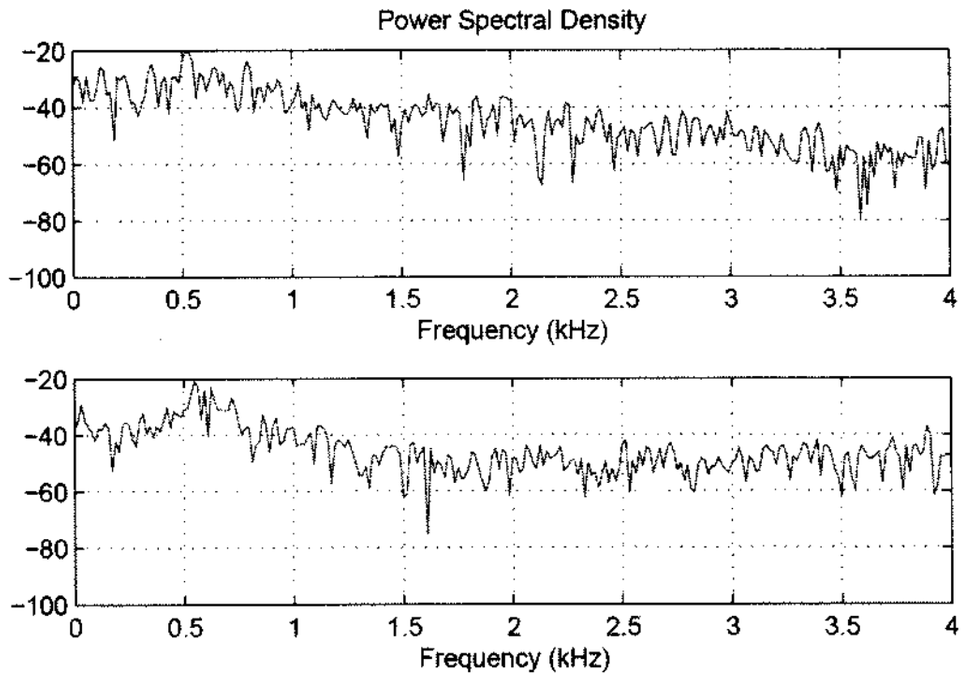

그림 2 실험 1: 두 왜색가우시안 음원의 파워 스펙트럼

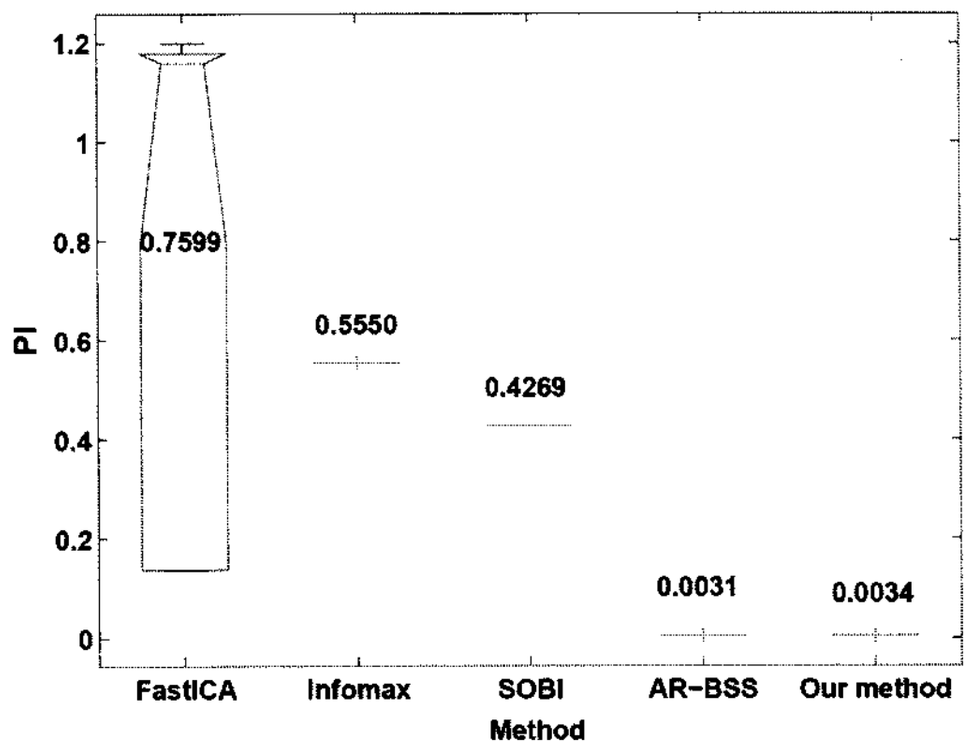

그림 3 실험 1: 각 알고리듬별 분리 결과의 상자그림 (boxplot). 그림 안에 표시된 숫자는 각 알고리즘별 출력 PI의 평균값을 의미하며, 붉은 선은 중간 값을 표시한다.

처럼 두 왜색가우시안 음원들의 파워 스펙트라가 비슷한 경우, 본 알고리듬과 AR-BSS는 만족할 수준의 분리 결과를 보이지만 SOBI는 분리에 실패하는 것을 볼 수 있다(참고 그림 3).

Experiment 2

이번 예는 음원의 시간적 구조를 표현하기 위해 모수형태의 음원생성모델에 기반 한 음원분리 방법들의 한계를 지적한다. 선형 AR모델 (4)가 비선형 시계열 음원의 비선형성을 설명하는데 한계가 있음은 자명하다. 이에 반해 제안 알고리듬의 경우, 비모수방법인 GP 모델이 가지는 유연성에 의해 비선형 시계열 데이타의 시간적 구조도 쉽게 모델링 할 수 있다. 이 본 알고리듬이 가지는 장점은 다음의 분리실험에서 확인 가능하다.

• 비선형 시계열 데이타:

세 가지 비선형 시계열 음원들이 사용되었다(그림 4

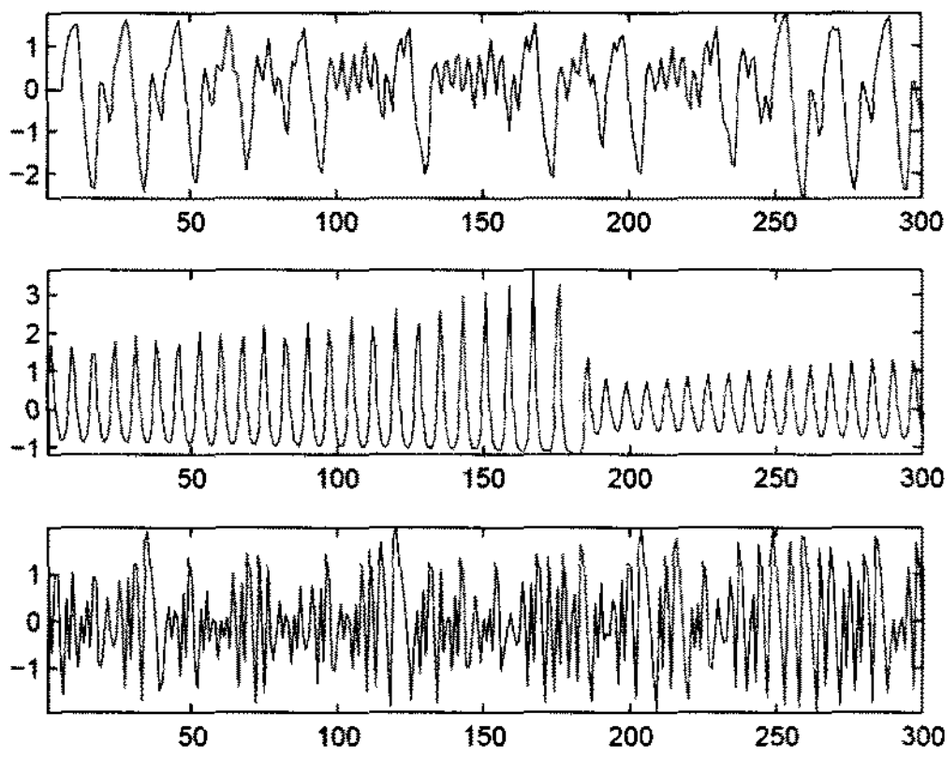

그림 4 실험 2의 음원: 위 - Mackey-Glass MG_{30}, 중간 - Santa Fe competition laser 아래 - Ikeda map의 첫 번째 변수

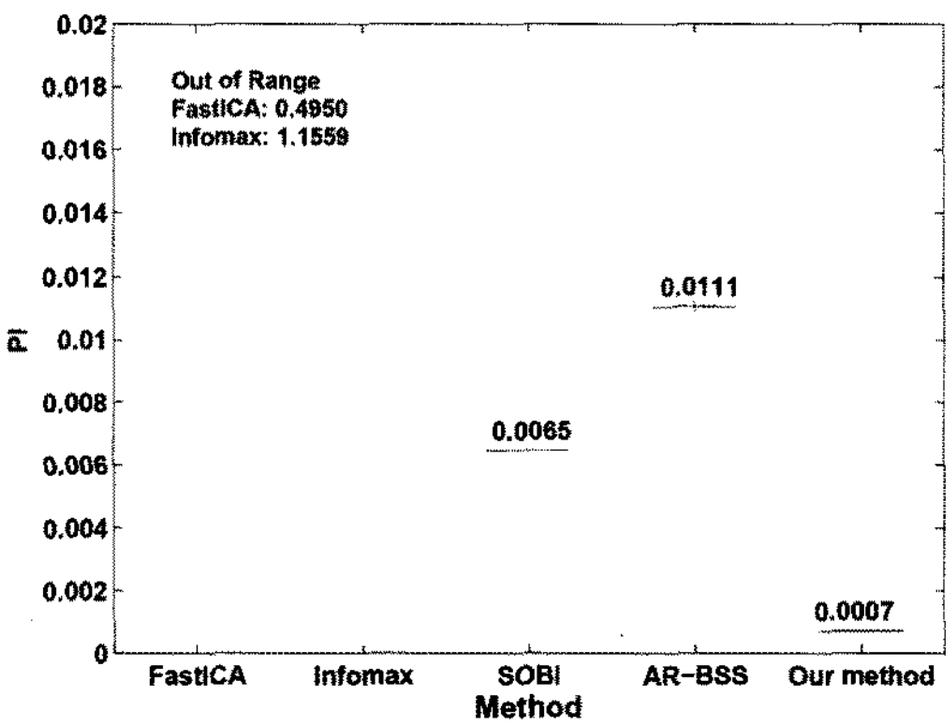

그림 5 실험 2에서의 각 알고리듬별 분리 결과의 상자그림. FastICA and Informax의 결과는 PI값이 0.4보다 크므로 제외되었다.

참고). 즉, Mackey-Glass MG_{30}, Santa Fe competition laser, 카오스 데이타인 Ikeda map의 첫 번째 변수. 선형 AR 모델은 이와 같은 비선형 시계열 데이타가 가지는 비선형성을 정확하게 표현하기 어려우며 그에 따라 AR-BSS 또한 좋은 성능을 보여주지 못한다(see 그림 5).

Experiment 3

마지막으로 실제 음성신호에 대한 분리 성능평가이다. 두 가지 음성신호 데이타 모음(data set)으로, 각각 세 명의 화자의 (1)"one" 과 (2)"two" 발음이다. 각 음성은 사무실 환경에서 잡음 없이 녹음되었으며, 8Khz로 재샘플링(resampled)되어 혼합되었다. 각각의 음성 데이타에 대한 분리결과는 그림 6과 그림 7제시되었는데 SOBI, AR-BSS와 제안 알고리듬의 성능만을 수록하였다. 본

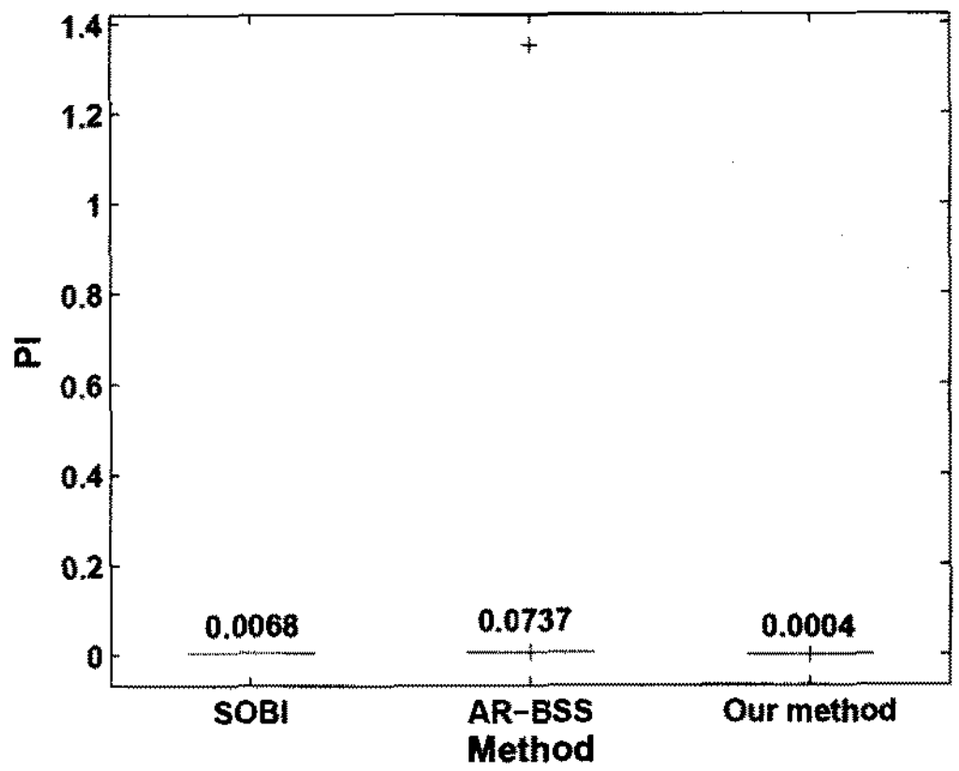

그림 6 실험 3-데이타 (1): 각 알고리듬별 분리 결과의 상자그림

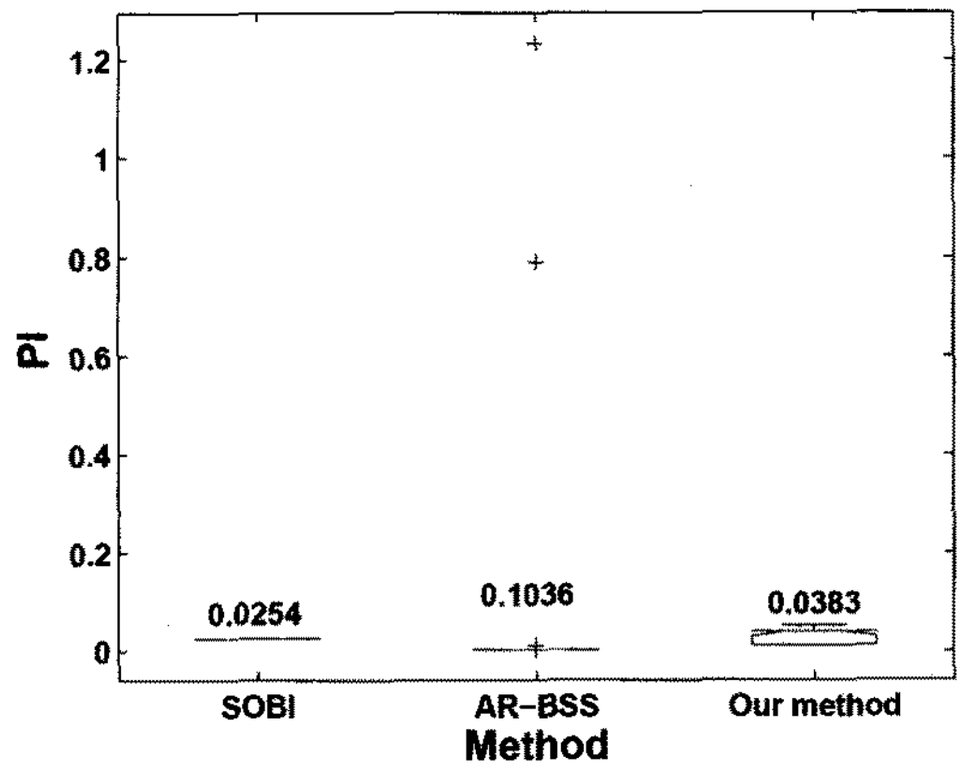

그림 7 실험 3-데이타 (2): 각 알고리듬별 분리 결과의 상자그림

제안 알고리듬은 두 데이타 모음에서 모두 안정적인 성능을 보임을 알 수 있다. 이에 반해 AR-BSS는 불안정적인 성능을 보이는데, 특히 그림 7의 상자그림을 보면 AR-BSS가 두 번의 분리실험에서 완전히 분리에 실패함을 알 수 있다.

5. Conclusion

음원의 시간적 구조를 GP로 모델링하고, 주어진 데이타의 로그-유사-가능도의 최대화 문제로 공식화하여 역혼합행렬을 구하는 음원분리 알고리듬을 제안하였다. 본 알고리듬은 기존의 모수형태(e.g., AR 모델) 기반의 음원분리 알고리듬에 비해 두 가지 측면의 장점들이 있다:

(1) 비선형 시계열음원을 쉽게 다룰 수 있다;

(2) 분리 결과가 음원생성모델의 모델 차수의 영향에 강인하다.

또한 음원들이 비슷한 주파수 스펙트럼을 가질 경우에도, SOBI와는 달리 본 알고리듬은 좋은 성능을 보여준다. 그러나 안 알고리듬은 다른 비모수 방법들의 문제점을 공유하고 있다. 즉, 많은 계산상의 노력이 필요로 한다는 약점이다. 그러므로 제안 알고리듬은 아직 많은 데이타 신호를 다루는 데 어려움이 있다. 이런 점은 앞으로 온라인 접근 방법이나 근사화 방법 - 특히 sparse approximations[16] - 등을 이용하여 해결할 것이다.

참 고 문 헌

[1] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, John Wiley & Sons, Inc., 2002.

[2] A. Bell and T. Sejnowski, An information maximisation approach to blind separation and blind deconvolution, Neural Computation, Vol.7, pp.1129-1159, 1995.

[3] A. Hyvarinen and E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation, Vol.9, pp. 1483-1492, 1997.

[4] S. Amari and J. F. Cardoso, Blind source separation: Semiparametric statistical approach, IEEE Trans. Signal Processing, Vol.45, pp. 2692-2700, 1997.

[5] A. Belouchrani, K. Abed-Merain, J. F. Cardoso, and E. Moulines, A blind source separation technique using second order statistics, IEEE Trans. Signal Processing}, Vol.45, pp.434-444, Feb. 1997.

[6] B. Pearlmutter and L. Parra, A context-sensitve generalization of ICA, in Proceedings of the International Conference on Neural Information Processing, pp.151-157, 1996.

[7] H. Attias and C. E. Schreiner, Blind source separation and deconvolution: The dynamic component analysis algorithms, Neural Computation, Vol.10, pp.1373-1424, 1998.

[8] Y. M. Cheung, Dual auto-regressive modelling approach to Gaussian process identification' in Proceedings of IEEE International Conference on Multimedia and Expo, 2001, pp.1256-1259.

[9] M. Seeger, Gaussian processes for machine learning, International Journal of Neural Systems, Vol. 14, No.2, pp.69-106, 2004.

[10] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. MIT Press, 2006.

[11] S. Park and S. Choi, Source separation with Gaussian process models, in Proceedings of the European Conference on Machine Learning. Warsaw, Poland: Springer, pp.262-273, 2007.

[12] S. Park and S. Choi, Gaussian processes for source separation, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, USA, 2008.

[13] S. Sundararajan and S. S. Keerthi, Predictive approaches for choosing hyperparameters in Gaussian processes, Neural Computation, Vol.13, pp. 1103-1118, 2001.

[14] J. Besag, Statistical analysis of non-lattice data, The Statistician, Vol.24, No.3, pp.179-195, 1975.

[15] S. Amari, A. Cichocki, and H. H. Yang, A new learning algorithm for blind signal separation, in Advances in Neural Information Processing Systems, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds., Vol.8. MIT Press, pp. 757-763, 1996.

[16] L. Csato and M. Opper, Sparse on-line Gaussian processes, Neural Computation, Vol.14, pp. 641-668, 2002.

박 선 호

2006년~현재 포항공과대학교 컴퓨터공학과 박사과정. 2006년 포항공과대학교 컴퓨터공학과 석사. 2004년 고려대학교 전자공학부 학사. 관심분야는 기계학습 (kernel, Gaussian process 등)

최 승 진

2001년~현재 포항공과대학교 컴퓨터공학과 교수. 1997년~2001년 충북대학교 전자공학과 교수. 1997년 일본 RIKEN Frontier Researcher. 1996년 University of Notre Dame 전자공학과 방문조교수 1996년 University of Notre Dame 전자공학과 박사. 1989년 서울대학교 전기공학과 석사. 1987년 서울대학교 전기공학과 학사. 관심분야는 통계적 기계학습, 확률 그래프 모델