Online Gaussian Process for Nonstationary Speech Separation

Hsin-Lung Hsieh and Jen-Tzung Chien

Department of Computer Science and Information Engineering
National Cheng Kung University, Tainan, Taiwan 70101, ROC
{hlhsieh,chien}@chien.csie.ncku.edu.tw

Abstract

In a practical speech enhancement system, it is required to enhance speech signals from the mixed signals, which were corrupted due to the nonstationary source signals and mixing conditions. The source voices may be from different moving speakers. The speakers may abruptly appear or disappear and may be permuted continuously. To deal with these scenarios with a varying number of sources, we present a new method for nonstationary speech separation. An online Gaussian process independent component analysis (OLGP-ICA) is developed to characterize the real-time temporal structure in time-varying mixing system and to capture the evolved statistics of independent sources from online observed signals. A variational Bayes algorithm is established to estimate the evolved parameters for dynamic source separation. In the experiments, the proposed OLGP-ICA is compared with other ICA methods and is illustrated to be effective in recovering speech and music signals in a nonstationary speaking environment.

Index Terms: speech enhancement, speech separation, Gaussian process, online learning, variational Bayes

10.21437/Interspeech.2010-168

1. Introduction

Blind source separation (BSS) aims to solve the cocktail party problem by means of recovering the source signals from the mixed signals. Suppose that there are M people speaking simultaneously in a room with N microphones in different locations. The recorded signals are the weighted sum of speech signals from individual speakers. Let $\mathbf{x}=[x_1,\cdots,x_N]^T$ denote the observations which are linear mixtures of M statistically independent sources $\mathbf{s}=[s_1,\cdots,s_M]^T$ by $\mathbf{x}=A\mathbf{s}$. The original speech signals $\mathbf{s}$ can be estimated from the observed signals $\mathbf{x}$ via a demixing matrix W. The independent component analysis (ICA) is a popular mechanism for solving BSS problem. In traditional ICA methods, a demixing matrix W is found by optimizing the metric of independence among the demixed signals $\mathbf{s}=W\mathbf{x}$ where W or A is assumed to be stationary. However, in real world, the cocktail-party problem is especially severe in presence of *nonstationary* sources. For example, a variety of sources including noise, music and multiple speakers may be acquired by microphones at the same time. During the conversation, we may need to detect the speakers even though they are moving or hesitating. For this nonstationary BSS problem, the mixing matrix is time-varying since the mixing coefficients depend on the distances between the sources and sensors. Pretending a static mixing matrix A in ICA methods shall not work well for these circumstances.

There have been several methods proposed to cope with the situations of dynamic sources or compensate the variations of mixing system in nonstationary environments. In general, two scenarios were considered in dynamic source separation [7]. First, the sources or the sensors may be moving. For this case, we need to compensate the variations of mixing matrix by an adaptive ICA algorithm. In [4], a state transition matrix was adopted to capture the variations in mixing matrix when the speaker or microphone was moving but the number of sources was fixed. In [11], a 3D tracker was utilized to detect the status of sources. If the sources are moving, a beamforming algorithm was applied for BSS. These algorithms assumed a fixed number of source signals. In the second scenario, the source signals may suddenly appear, disappear or permute at different time moments. For this case, we need to continuously estimate the time-varying source distributions or the number of source signals. The number of sources was selected by unsupervised learning [1] using the automatic relevance determination (ARD) [10]. In [2], a hidden Markov model (HMM) was presented to extract the temporal information and merge it with the prior density of mixture coefficients for Bayesian ICA. The state-dependent source distributions were derived. In [5], an indicator variable was used to indicate the status of sources either active or inactive while the movement of source or sensor was not considered. In [6], an online variational Bayes (VB) learning was performed in ICA procedure where the source signal was seen as a time-dependent parameter and the mixing matrix was assumed to be time-independent. The source distributions were updated incrementally while the number of sources was fixed. Such an approach could not tackle the case of abruptly appearing or disappearing sources. All of these methods do not simultaneously consider two scenarios for dynamic BSS.

Recently, we presented an online Bayesian learning for dynamic source separation where two scenarios were considered at the same time [7]. The mixing matrix and the distribution of source signals were jointly estimated at each time frame via an incremental VB estimation where the robustness of speech separation was assured. In this study, we further extend the online or real-time learning strategy for BSS via incorporating a Gaussian process [14] to capture the temporal structure of the nonstationary mixing coefficients. An online Gaussian process ICA (OLGP-ICA) is developed to establish a precise and robust model for dynamic source separation. The online learning of OLGP parameters is exploited according to a recursive Bayes method for regression and classification which is different from incremental training of GP based on sparse representation in [3][13]. The proposed OLGP-ICA is evaluated and compared by the experiments considering different scenarios where the dynamic sources of music and multiple speakers are present.

2. Background Survey

2.1. Nonstationary blind source separation

Considering the realistic circumstances in a cocktail party, we need to enhance the mixed signals where multiple people are chatting, laughing and moving with background noise and music. The microphones may acquire dynamic source signals including music, babble and noises. Source signals may suddenly appear or disappear. The location of speakers may be changed frequently. Figure 1 illustrates the scenarios in a nonstationary cocktail party. There are three microphones $\mathbf{x}=[x_1,x_2,x_3]^T$ and three source signals including two speech signals s_1 and s_2 from speakers 1 and 2, respectively, and one

Copyright © 2010 ISCA

background music signal s_3. As the time evolves from t to $t+1$, the mixing matrix $A=\{a_{nm}\}$ should be changed as $A^{(t)} \to A^{(t+1)}$ to reflect the nonstationary speaking environment. For example, the *first* speaker s_1 may move and his distance to the *second* sensor is changed from d_1 to d_1'. The resulting change occurs in mixing coefficient $a_{21}^{(t)} \to a_{21}^{(t+1)}$ which depends on a time-varying function of the distance between sensor and source. The second speaker s_2 and the music s_3 have no change, then $a_{22}^{(t)} = a_{22}^{(t+1)}$ and $a_{23}^{(t)} = a_{23}^{(t+1)}$. At the next time frame $t+2$, the first speaker may stop talking and simultaneously a new speaker comes in the party to replace the second speaker. For this scenario, the mixed signals are obtained due to the changes of coefficients $a_{21}^{(t+2)} = 0$ and $a_{22}^{(t+1)} \to a_{22}^{(t+2)}$. In this study, we deal with a nonstationary cocktail-party problem. The temporal structure of time-varying mixing matrix $A^{(t)}$ is characterized by a Gaussian process (GP).

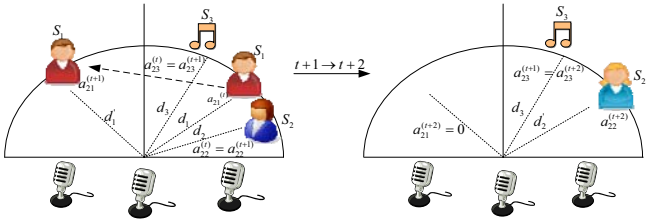

Figure 1: Scenarios in a nonstationary cocktail-party problem.

2.2. Speech separation using Gaussian process

Speech signals are temporally correlated. It is important to incorporate temporal information in a speech separation system. Gaussian process is known as a mechanism of characterizing the temporal structure of observation data. The probability distributions of GP are latent and can be expressed as the space of functions $f: S \to \Re$ where any subset of the variables has a joint Gaussian distribution. Different from the parametric model using autoregressive model, GP is a *nonparametric* latent model which is specified by a mean function $m(\mathbf{s}_m^{(t)})$ and a covariance function $\kappa(\mathbf{s}_m^{(t)}, \mathbf{s}_m^{(\tau)})$, i.e.

$$f(\mathbf{s}_m^{(t)}) \sim \mathrm{GP}(m(\mathbf{s}_m^{(t)}), \kappa(\mathbf{s}_m^{(t)}, \mathbf{s}_m^{(\tau)})) \tag{1}$$

where

$$m(\mathbf{s}_m^{(t)}) = E[f(\mathbf{s}_m^{(t)})] \tag{2}$$

$$\kappa(\mathbf{s}_m^{(t)}, \mathbf{s}_m^{(\tau)}) = E[(f(\mathbf{s}_m^{(t)}) - m(\mathbf{s}_m^{(t)}))(f(\mathbf{s}_m^{(\tau)}) - m(\mathbf{s}_m^{(\tau)}))]. \tag{3}$$

Here, $\mathbf{s}_m^{(t)}$ and $\mathbf{s}_m^{(\tau)}$ denote the speech samples at the mth source in time frames t and τ. For simplicity, the mean function is usually assumed to be zero. Some covariance functions or kernels are available in [14]. In [9], Gaussian process priors were combined into the factor analysis model for representation of spatio-temporal data. In [12], GP model was used to characterize the temporal structure of source signal via the distribution of latent functions $f(\cdot)$ which related the current source sample to its past samples. A Gaussian process prior was merged and the pseudo-likelihood estimation was performed. Compared with the autoregressive process, GP is more flexible and allows the nonlinear time series as source signals. However, these methods were designed to explore the temporal structure of source signals rather than that of the nonstationary mixing system which is more realistic in nonstationary BSS. We are motivated to conduct an online Gaussian process for mixing coefficients and perform the dynamic source separation.

3. Online Gaussian Process

3.1. Online Bayesian learning

Typically, the environmental conditions of the mixed signals are changed naturally as time moves on. A flexible ICA procedure should be established by relaxing the assumptions in standard ICA where the number and the distribution of sources are varied with time. It is preferable to perform *online speech separation* by Bayesian framework. This paper presents a new online Gaussian process ICA (OLGP-ICA) algorithm to incrementally characterize the temporal structure of the time-varying mixing matrix and to estimate the GP parameters to identify dynamic sources. The sufficient statistics from previous frames is merged with the likelihood of current frame and is propagated for adaptive speech separation. The mixing matrix and the distributions of source signals are incrementally estimated instead of performing Markov chain for each source signal in switching ICA [2][5]. The computational load can be reduced accordingly. Our idea is to incrementally detect the source signals and estimate the corresponding distributions from online observation data $\chi^t = \{X^{(1)}, X^{(2)}, \cdots, X^{(t)}\}$. Since different scenarios of the nonstationary conditions are considered, the mixing coefficients have prominent temporal structure in essence. The GP is applied to express the nonlinear time series appearing in the mixing matrix. The time-dependent parameters are incrementally estimated instead of running time-consuming Markov process for each source signal [5]. At each time frame t, we estimate ICA parameter $\theta^{(t)}$ by maximizing the posterior distribution using χ^t. The posterior distribution $p(\theta^{(t)} \mid \chi^t)$ is calculated through the stages of prediction and correction. Given the historical data χ^{t-1}, the prediction of the parameters $\theta^{(t)}$ is obtained according to the probability $p(\theta^{(t)} \mid \chi^{t-1})$ calculated by

$$\int p(\theta^{(t)} \mid \theta^{(t-1)}) p(\theta^{(t-1)} \mid \chi^{t-1}) d\theta^{(t-1)}. \tag{4}$$

When a new frame $X^{(t)}$ is observed, the prediction is corrected by

$$p(\theta^{(t)} \mid \chi^t) = \frac{p(X^{(t)} \mid \theta^{(t)}) p(\theta^{(t)} \mid \chi^{t-1})}{\int p(X^{(t)} \mid \theta^{(t)}) p(\theta^{(t)} \mid \chi^{t-1}) d\theta^{(t)}}. \tag{5}$$

Different from the batch training [2][5], we present an efficient approach to detect the status of each source signal and to estimate the distributions of reconstructed source signals at each frame. This online learning works for dynamic source separation under different scenarios [7].

3.2. OLGP model construction

In model construction using OLGP, the noisy ICA model $X^{(t)} = A^{(t)} S^{(t)} + E^{(t)}$ with a noise term $E^{(t)} = \{\boldsymbol{\varepsilon}^{(t,i)}\}_{i=1}^{L}$ is adopted for data generation. Let $X^{(t)} = \{\mathbf{x}^{(t,i)}\}_{i=1}^{L}$ denote a set of L mixed samples acquired from N microphones. These mixed samples are modeled by a linear mixture of M unknown independent source signals $S^{(t)} = \{\mathbf{s}^{(t,i)}\}_{i=1}^{L}$. The graphical representation of OLGP model is depicted in Figure 2. Assuming that each entry of the $N \times M$ time-varying mixing

matrix $A^{(t)}$ is generated by a latent function $f_{nm}(\cdot)$ which connects the relation between the current mixing coefficient $a_{nm}^{(t,i)}$ and the past p mixing coefficients by

$$a_{nm}^{(t,i)} = f_{nm}(\mathbf{a}_{nm}^{(t,i-1,i-p)}) \tag{6}$$

where $\mathbf{a}_{nm}^{(t,i-1,i-p)}$ represents a set of past p samples

$$\mathbf{a}_{nm}^{(t,i-1,i-p)} = [a_{nm}^{(t,i-1)}\ a_{nm}^{(t,i-2)}\ \cdots\ a_{nm}^{(t,i-p)}]^T. \tag{7}$$

The Gaussian process is adopted to describe the distribution of latent function by

$$f_{nm}(\mathbf{a}_{nm}^{(t,i-1,i-p)}) \sim \mathrm{GP}(0, \kappa(\mathbf{a}_{nm}^{(t,i-1,i-p)}, \mathbf{a}_{nm}^{(t,\tau-1,\tau-p)})) \tag{8}$$

where $\kappa(\mathbf{a}_{nm}^{(t,i-1,i-p)}, \mathbf{a}_{nm}^{(t,\tau-1,\tau-p)})$ is a covariance matrix based on an exponential-quadratic kernel function with the kernel hyperparameters $\{\lambda_{nm}^{(t,i)}, \rho_{nm}^{(t,i)}\}$

$$\rho_{nm}^{(t,i)} \exp\left\{-\frac{\lambda_{nm}^{(t,i)}}{2} \| \mathbf{a}_{nm}^{(t,i-1,i-p)} - \mathbf{a}_{nm}^{(t,\tau-1,\tau-p)} \|^2\right\} \tag{9}$$

The GP prior over the latent function $f_{nm}(\cdot)$ for a set of mixing coefficients $\mathbf{a}_{nm}^{(t)} = [a_{nm}^{(t,i-1)}\ a_{nm}^{(t,i-2)}\ \cdots\ a_{nm}^{(t,i-L)}]^T$ in frame t is given by a Gaussian distribution with zero mean and covariance matrix $\mathbf{K}_{nm}^{(t)}$

$$p(f_{nm} \mid \{\mathbf{a}_{nm}^{(t,i-1,i-p)}\}_{i=1}^{L}) = N(0, \mathbf{K}_{nm}^{(t)}) \tag{10}$$

The (l,k) element of the covariance matrix is given by

$$[\mathbf{K}_{nm}^{(t)}]_{lk} = \kappa(\mathbf{a}_{nm}^{(t,l-1,l-p)}, \mathbf{a}_{nm}^{(t,k-1,k-p)}) \tag{11}$$

In addition, the $N \times 1$ noise vector $\boldsymbol{\varepsilon}^{(t,i)}$ is assumed to have an isotropic Gaussian distribution $N(0, \beta^{(t,i)^{-1}} I_N)$ with zero mean and precision $\beta^{(t,i)}$. The prior density of $\beta^{(t,i)}$ is assumed as a Gamma distribution $p(\beta^{(t,i)}) = \mathrm{Gamma}(\mathrm{u}_\beta^{(t,i)}, w_\beta^{(t,i)})$ with hyperparameters $\{\mathrm{u}_\beta^{(t,i)}, w_\beta^{(t,i)}\}$. The likelihood function using OLGP-ICA is yielded by

$$p(\mathbf{x}^{(t,i)} \mid \mathbf{s}^{(t,i)}, A^{(t,i)}, \beta^{(t,i)}) = N(A^{(t,i)}\mathbf{s}^{(t,i)}, \beta^{(t,i)^{-1}} I_N). \tag{12}$$

The source signals $\mathbf{s}^{(t,i)} = [s_1^{(t,i)} \cdots s_M^{(t,i)}]^T$ are modeled by a mixture of K Gaussians

$$p(\mathbf{s}^{(t,i)} \mid \boldsymbol{\pi}^{(t,i)}, \boldsymbol{\mu}^{(t,i)}, \boldsymbol{\gamma}^{(t,i)}) = \prod_{m=1}^{M}\left[\sum_{k=1}^{K} \pi_{mk}^{(t,i)} N(\mu_{mk}^{(t,i)}, \gamma_{mk}^{(t,i)^{-1}})\right] \tag{13}$$

where $\pi_{mk}^{(t,i)}$ is the mixture weight, $\mu_{mk}^{(t,i)}$ is the mean vector, and $\gamma_{mk}^{(t,i)}$ is the precision matrix. The non-Gaussianity of source signals is considered. Hence, the parameter set is formed by $\theta^{(t,i)} = \{A^{(t,i)}, \mathbf{s}^{(t,i)}, \beta^{(t,i)}\}$. The marginal likelihood using OLGP-ICA model is yielded by

$$p(X^{(t)}) = \prod_{i=1}^{L} \int p(\mathbf{x}^{(t,i)} \mid A^{(t,i)}, \mathbf{s}^{(t,i)}, \beta^{(t,i)}) p(A^{(t,i)} \mid \lambda_{nm}^{(t,i)}, \rho_{nm}^{(t,i)}) \times p(\mathbf{s}^{(t,i)} \mid \boldsymbol{\pi}^{(t,i)}, \boldsymbol{\mu}^{(t,i)}, \boldsymbol{\gamma}^{(t,i)}) p(\beta^{(t,i)} \mid \mathrm{u}_\beta^{(t,i)}, w_\beta^{(t,i)}) dA^{(t,i)} d\mathbf{s}^{(t,i)} d\beta^{(t,i)}. \tag{14}$$

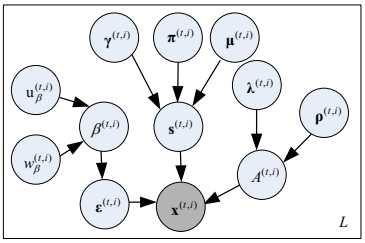

Figure 2: Graphical model of OLGP-ICA for speech separation.

Importantly, the temporal structure of the time-varying mixing matrix $A^{(t)}$ is described by a Gaussian process with the scaling hyperparameters $\boldsymbol{\lambda}^{(t)} = \{\lambda_{nm}^{(t)}\}$. If the estimated $\lambda_{nm}^{(t)}$ is very small, it implies that the associated $\mathbf{a}_{nm}^{(t)}$ have little effect on the prediction. Estimating the parameter $\boldsymbol{\lambda}^{(t)}$ is equivalent to performing the *automatic relevance determination* (ARD) which reflects the activities or the dynamics of source signals. For example, when the source signal is abruptly inactive, the corresponding mixing coefficients shall become zero. The previous mixing coefficients have little impact on prediction of the current mixing coefficients. By tracking and updating the scaling parameter, we determine the status of source signals for different time moments.

3.3. OLGP model inference

Since the exact posterior distribution $p(\theta \mid X^{(t)})$ does not exist, we employ VB inference to approximate the posterior by a variational distribution $q(\theta)$. In VB formulation, we maximize the lower bound of logarithm of marginal distribution or equivalently minimize the Kullback-Leibler distance between true posterior and approximate posterior. Considering the factorized variational inference

$$q(\theta^{(t)}) = q(A^{(t)}) q(\mathbf{s}^{(t)}) q(\beta^{(t)}) \tag{15}$$

, the lower bound or the negative free energy is obtained by

$$\begin{aligned} \log p(X^{(t)}) \geq & < \log p(X^{(t)} \mid A^{(t)}, \mathbf{s}^{(t)}, \beta^{(t)}) >_{q_A q_s q_\beta} \\ & + S(q_A) + < \log p(A^{(t)} \mid \boldsymbol{\lambda}^{(t)}, \boldsymbol{\rho}^{(t)}) >_{q_A} + S(q_s) \\ & + < \log p(\mathbf{s}^{(t)}) >_{q_s} + S(q_\beta) + < \log p(\beta^{(t)}) >_{q_\beta} \end{aligned} \tag{16}$$

where $<\cdot>$ denotes the expectation function over variational distribution $q(\cdot)$ and $S(q)$ is the entropy of $q(\cdot)$. By taking the derivative of lower bound with respect to $q(\theta^{(t)})$, we can find the solutions to variational parameters of $q(A^{(t)})$, $q(\mathbf{s}^{(t)})$ and $q(\beta^{(t)})$ which are similar to those in [8]. Also, the updated formulas of the mixing coefficients in $N(\mathbf{a}_{nm}^{(t)} \mid \hat{\boldsymbol{\mu}}_{a_{nm}}^{(t)}, \hat{\mathrm{K}}_{a_{nm}}^{(t)})$ is given by

$$\hat{m}(\mathbf{a}_{nm}^{(t)}) = (\hat{\mathrm{K}}_{nm}^{(t)})^{-1} \mathbf{b}_{nm}^{(t)} \tag{17}$$

$$\hat{\mathrm{K}}_{nm}^{(t)} = \left[C + < \mathrm{K}_{nm}^{(t)} >^{-1}\right]^{-1} \tag{18}$$

where $\mathbf{b}_{nm}^{(t)}$ is a $L \times 1$ vector and the ith element is given by

$$b_{nm}^{(t,i)} = < \beta_n^{(t,i)} >< s_m^{(t,i)} > (x_n^{(t,i)} - \sum_{k \neq m}^{M} < a_{nm}^{(t,i)} >< s_k^{(t,i)} >) \tag{19}$$

, $\mathbf{\beta}_n^{(t)}$ and C are the $L \times L$ diagonal matrices with ith diagonal elements $\beta_n^{(t,i)}$ and $[C]_{ii} = < \beta_n^{(t,i)} >< s_m^{(t,i)^2} >$, respectively. The updating of hyperparameters $\boldsymbol{\lambda}^{(t)}$ and $\boldsymbol{\rho}^{(t)}$ is similar to [9]. In this study, the *converged variational parameters* at frame t are *propagated* as the *initial variational parameters* at frame $t+1$.

4. Experiments

In the experiments, we evaluated the proposed OLGP-ICA for dynamic source separation. We simulated the scenarios with dynamic mixing matrices and dynamic source signals. The nonstationary mixing matrix was simulated by $A^{(t)} = [[\cos(2\pi f_1 t)\ -\sin 2\pi f_1 t]^T\ [\sin(2\pi f_2 t)\ \cos(2\pi f_2 t)]^T]$. The

entries of mixing matrix represent the relative position and distance between sources and sensors. The first and the second sources relative to the sensors had a changing rate of $f_1 = 1/20$ Hz and $f_2 = 1/10$ Hz, respectively. Figure 3 displays four nonstationary mixing coefficients $\{a_{11}^{(t)}, a_{12}^{(t)}, a_{21}^{(t)}, a_{22}^{(t)}\}$. Figure 4 shows the waveforms of two channels which contained two speakers and one background music which were sampled from http://www.kecl.ntt.co.jp/icl/signal/. In the first channel, a male speaker was inactive at 2 sec and the music signal was active at 3 sec, and so the distribution of the source signal was changed. The scenario of switching and moving sources was considered. In the second channel, the female speaker was inactive at 3 sec and active at 4 sec. The scenario of presence and absence of the same speaker was examined. In the implementation, the online signal separation was performed with a time window of 0.25 sec. For comparison, the variational Bayesian ICA (VB-ICA) [8], the switching ICA [5], the Bayesian ICA with hidden Markov sources (BICA-HMM) [2] and the online VB-ICA [6] was implemented. The demixed signals of OLGP-ICA were displayed in Figure 4. We provided the waveforms of original signals, mixed signals and demixed signals at http://chien.csie.ncku.edu.tw/~olgp. The signal-to-interference ratios (SIRs) of two demixed signals were reported in Table 1. The proposed OLGP-ICA achieved the highest SIR among different methods. In our investigation, OLGP-ICA significantly saves 73% of running time compared to the switching ICA [5] when implementing on a personal computer with a core 2 quad 2.4GHz CPU and a memory of 8 GB.

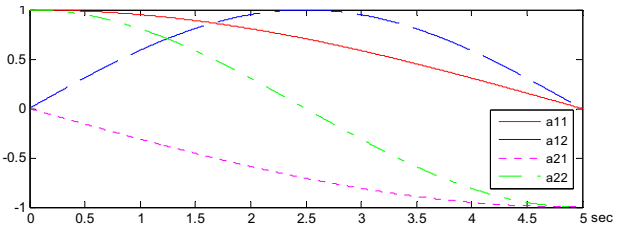

Figure 3: Four nonstationary mixing coefficients.

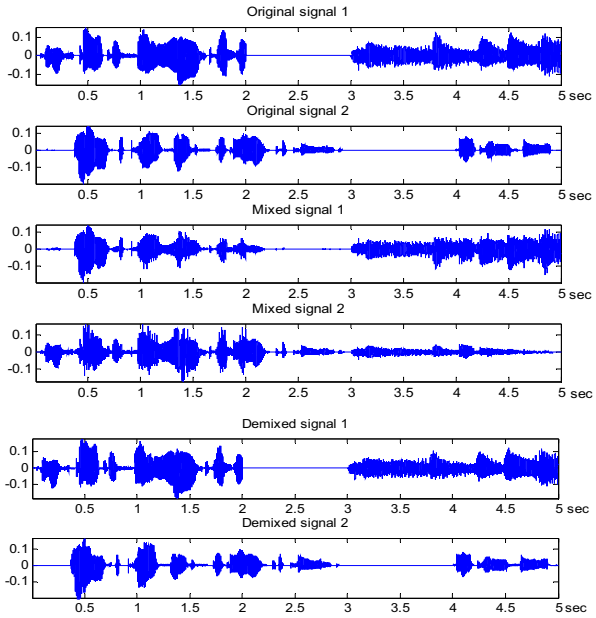

Figure 4: Waveforms of source signals, mixed signals and demixed signals using OLGP-ICA.

Table 1: Comparison of SIRs (in dB) for different methods

	VB-ICA	Switching ICA	BICA-HMM	Online VB-ICA	OLGP-ICA
Demixed signal 1	7.97	12.06	9.04	11.26	17.24
Demixed signal 2	-3.23	-4.82	-1.5	4.47	9.96

5. Conclusions

This paper presented a novel nonstationary ICA algorithm to cope with the problem of dynamic source separation. The online learning adaptively captured the statistics of the source signals and mixing coefficients. This learning scheme is based on the VB learning where the robust modeling was achievable. In the nonstationary mixing process, the mixing coefficients were time-variant and nonlinear. The comparative experimental results demonstrated the efficiency and effectiveness of modeling the temporal information of mixing coefficients by using OLGP-ICA model. Instead of incorporating Markov model for individual source signals using batch estimation, the proposed OLGP employed an ARD model and efficiently determined the number of the latent sources by incremental learning method. This method was practical for real-world speech enhancement and separation. In the future, we will investigate sparse learning for GP model so as to reduce the computation cost and enhance the model generalization for speech separation.

6. References

[1] K. Chan, T. W. Lee and T. J. Sejnowski, "Variational learning of clusters of undercomplete nonsymmetric independent components", *Journal of Machine Learning Research*, vol. 3, pp. 99-114, 2002.

[2] R. A. Choudrey and S. J. Roberts, "Bayesian ICA with hidden Markov sources", *Proc. of ICA*, pp. 809-814, 2003.

[3] L. Csato and M. Opper, "Sparse on-line Gaussian processes", *Neural Computation*, vol. 14, pp. 641-668, 2002.

[4] R. Everson and S. J. Roberts, "Non-stationary independent component analysis", *Proc. of ICANN*, pp. 503- 508, 1999.

[5] J. Hirayama, S. Maeda and S. Ishii, "Markov and semi-Markov switching of source appearances for nonstationary independent component analysis", *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1326-1342, 2007.

[6] A. Honkela and H. Valpola, "On-line variational Bayesian learning", *Proc. of ICA*, pp. 803-808, 2003.

[7] H.-L. Hsieh and J.-T. Chien, "Online Bayesian learning for dynamic source separation", *Proc. of ICASSP*, pp. 1950-1953, 2010.

[8] N. D. Lawrence and C. M. Bishop, "Variational Bayesian independent component analysis", *Technical Report*, University of Cambridge, 2000.

[9] J. Luttinen and A. Ilin, "Variational Gaussian-process factor analysis for modeling spatio-temporal data", *Advances in Neural Information Processing Systems*, pp. 1177-1185, 2009.

[10] D. J. C. MacKay, "Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks", *Network: Computation in Neural Systems*, vol. 6, pp. 469-505, 1995.

[11] S. M. Naqvi, Y. Zhang and J. A. Chambers, "Multimodal blind source separation for moving sources", *Proc. of ICASSP*, pp. 125-128, 2009.

[12] S. Park and S. Choi, "Gaussian processes for source separation", *Proc. of ICASSP*, pp. 1909-1912, 2008.

[13] J. Quinonero-Candela and O. Winther, "Incremental Gaussian processes", *Advances in Neural Information Processing Systems*, pp. 1001-1008, 2003.

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