2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Nonnegative Shifted Tensor Factorization in Time Frequency Domain

Qiang Wu, Ju Liu, Fengrong Sun, Jie Li and Andrzej Cichocki

***Abstract*— In this paper, we proposed a Nonnegative Shifted Tensor Factorization (NSTF) model considering multiple component delays by time frequency analysis. Explicit mathematical representation for the delays is presented to recover the patterns from the original data. In order to explore multilinear shifted component in different modes, we use fast fourier transform (FFT) to transform the non-integer delays into frequency domain by gradients search. The ALS algorithm for NSTF is developed by alternating least square procedure to estimate the nonnegative factor matrices in each mode and enforce the sparsity of model. Simulation results indicate that ALS-NSTF algorithm can extract the shift-invariance sparse features and improve the recognition performance of robust speaker identification and structural magnetic resonance imaging (sMRI) diagnosis for Alzheimer's Disease.**

I. Introduction

In many cases sequential data exists component shifts/delays, for example, speech signal in cocktail party problem [1]. Various methods are proposed to estimate the delays and recover the original patterns. Convolutive blind source separation methods such as spatio-temporal fast ICA [2] have often been proposed as possible solutions to estimate the time delays between different microphones. Shifted nonnegative matrix factorization (Shifted NMF) [3] is developed to extend NMF naturally to handle the potential delays of biomedical data. Shifted independent component analysis (Shifted ICA) [4] is a kind of subspace analysis method considering non-integer shifts by information maximization in the complex domain. In order to capture shifts in the features, sparse shift-invariant nonnegative matrix factorization (ssiNMF) algorithm is given in [5] with efficient computation by FFT.

For tensor structure data, above subspace methods are no longer suitable. After matricization or vectorization of data, the spatial structure of higher order data is destroyed. How to extract latent component from higher order tensor structure data has potential applications in neuroscience, signal processing and machine learning fields. As stated in [6], [7], fMRI data often contain dimension of trial, subject, group, time and spatial etc. Multiple-modality patterns from EEG data [8] are also efficient for brain computer interface and ERP analysis. Various tensor decomposition approach have been developed for multilinear analysis, for example Nonnegative Tensor Factorization (NTF), PARAFAC, Tucker model and its extension HOSVD and HOOI algorithms [9].

However latent component delays exist across multilinear data [7], [10], [11] such as time shifts in fMRI data because of hemodynamic delay, shifts across trials in EEG data when onsets changes were not locked to the event, these component delays will cause degeneracy problem of tensor decomposition [7], [12]. Several algorithms are proposed to solve the tensor decomposition considering component delays. Mørup presents the Shifted CP algorithm [7] to estimate time or trial delays and extract latent components in fMRI and ERP experiment. N-way shifted factor analysis model including PARAFAC and Tucker is investigated in [12]. Convolutive Nonnegative Tucker Decomposition model is given in [13] to explore the patterns across columns of different modes.

In this paper, we investigate the nonnegative tensor factorization model considering the delays across different modes to avoid the degeneracy problem. The multilinear shifted component in different modes are extracted by alternating least square procedure. In order to estimate the non-integer shifts, FFT is employed to transform the component delays into time frequency domain and Newton-raphson method is used to find optimal solution. Finally, ALS-NSTF algorithm is developed to calculate the factor matrices and component delays. Experimental results confirm the validity of ALS-NSTF algorithm and show that it can extract efficient shift-invariance feature for robust speaker identification task and sMRI-based AD diagnosis.

II. Multilinear Algebra

Multilinear algebra is the theory of higher order tensor which is an extension of matrix. Let $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \cdots \times I_N}$ denote a tensor. The order of $\underline{\mathbf{X}}$ is N. The elements of $\underline{\mathbf{X}}$ is defined as $\mathbf{X}_{i_1,\ldots,i_N}$. The survey of multilinear algebra in detail can be found in [9]. Some basic notations of multilinear algebra are described in Table I.

The n-mode matricized of an N order tensor $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \ldots \times I_N}$ is a matrix $\mathbf{X}_{(n)} \in R^{K \times L}$, where $K = I_n$ and $L = \prod_{i \neq n} I_i$. We denote the n-mode matricizing of $\underline{\mathbf{X}}$ as $\mathbf{X}_{(n)}$.

The mode-n matrix product defines multiplication of a tensor with a matrix in mode n. Let $\underline{\mathbf{X}} \in R^{I_1 \times \ldots \times I_N}$

Qiang Wu, Ju Liu and Fengrong Sun are with the School of Information Science and Engineering, Shandong University, Jinan, Shandong, China (email: {wuqiang, juliu, sunfr}@sdu.edu.cn).

Jie Li is with the College of Electronics and Information Engineering, Tongji University, Shanghai, China (email: nijanice@gmail.com).

Andrzej Cichocki is with the Laboratory for Advanced Brain Signal Processing,BSI RIKEN, Wakoshi, Saitama, Japan and Warsaw University of Technology Department of EE, Poland (email: cia@brain.riken.jp).

The work was supported by the National Natural Science Foundation of China (Grant nos. 61305060 and 61105122), Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20130131120025), the Excellent Youth and Middle Age Scientists Fund of Shandong Province(Grant no. BS2012DX020) and the Fundamental Research Funds for the Central Universities (Grant No. 0800219202).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

TABLE I

NOTATIONS IN MULTILINEAR ALGEBRA

Notation	Description
$\odot$	Khatri-Rao product
$\circledast$	Hadamard product
$\mathbf{X}$	matrix
$[\mathbf{U}]_i$	ith column of matrix $\mathbf{U}$
$[\mathbf{U}]_{j,:}$	jth row of matrix $\mathbf{U}$
$\mathbf{U}^{(n)}$	the nth factor
$\tilde{\mathbf{U}}$	matrix $\mathbf{U}$ after FFT
$(\cdot)^T$	transpose
$(\cdot)^H$	conjugate transpose
$\underline{\mathbf{X}}$	tensor
$\mathbf{X}_{(n)}$	n-mode matricized of tensor $\underline{\mathbf{X}}$
$\times_n$	mode-n matrix product of tensor and matrix
$\mathbf{U}^{\odot}$	$U^{(N)} \odot \cdots \odot U^{(1)}$
$\mathbf{U}^{\odot -n}$	$U^{(N)} \odot \cdots \odot U^{(n+1)} \odot U^{(n-1)} \odot \cdots \odot U^{(1)}$

and $\mathbf{A} \in R^{J \times I_n}$. Then the elements of $\underline{\mathbf{Y}} \in R^{I_1 \times \ldots \times I_{n-1} \times J \times I_{n+1} \times \ldots \times I_N}$ tensor defined by

$$\begin{aligned} \underline{\mathbf{Y}} &= (\underline{\mathbf{X}} \times_n \mathbf{A})_{i_1,\ldots,i_{n-1},j,i_{n+1}\ldots,i_N} \\ &= \sum_{I_n} (\mathbf{X}_{i_1,\ldots,i_n,\ldots,i_N} \mathbf{A}_{j,i_n}) \end{aligned} \tag{1}$$

In this paper we simplify the notation as

$$\underline{\mathbf{X}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times \ldots \times \mathbf{A}_N = \underline{\mathbf{X}} \prod_{n=1}^{N} \times_n A_n \tag{2}$$

and

$$\underline{\mathbf{X}} \prod_{n=1, n\neq i}^{N} \times_n A_n = \underline{\mathbf{X}} \overline{\times}_i A_i \tag{3}$$

Considering a tensor $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \times \cdots \times I_N} \geq 0$, the Nonnegative Tensor Factorization model can be described as

$$\underline{\mathbf{X}} = \underline{\mathbf{\Lambda}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} \tag{4}$$

where $\underline{\mathbf{\Lambda}}$ is a diagonal tensor with 1 on the main diagonal and $\mathbf{U}^{(n)}|_{n=1}^{N} \in R^{I_n \times K}$ are the factor matrices. $\underline{\mathbf{X}}$,$\underline{\mathbf{\Lambda}}$ and $\mathbf{U}^{(n)}|_{n=1}^{N}$ are restricted to have only nonnegative elements in the factorization.

After unfolding we can obtain matrix forms of equation (4) expressed by the Khatri-Rao product:

$$\mathbf{X}_{(n)} = \mathbf{U}^{(n)} \mathbf{Z}^{(n)} \tag{5}$$

where $\mathbf{X}_{(n)} \in R^{I_n \times L_n}$, $\mathbf{Z}^{(n)} \in R^{K \times L_n}$, $\mathbf{Z}^{(n)} = \mathbf{U}^{\odot -n T}$, $\mathbf{U}^{\odot -n T}$ is transpose of $\mathbf{U}^{\odot -n}$, $L_n = \Pi_{i\neq n}^{N} I_i$.

III. Nonnegative Shifted Tensor Factorization

In order to investigate the component delays model of tensor factorization, we introduce shifted operation in each factor matrices. The tensor factorization model (4) can be extended into following model

$$\underline{\mathbf{X}} = \underline{\mathbf{\Lambda}} \times_1 \hat{\mathbf{U}}^{(1)} \times_2 \hat{\mathbf{U}}^{(2)} \cdots \times_N \hat{\mathbf{U}}^{(N)} \tag{6}$$

where

$$\hat{\mathbf{U}}^{(n)} = [\mathbf{U}^{(n)}]_{i_n, j_n + \tau^n_{i_n,j_n}} \tag{7}$$

and $\tau^n_{i_n,j_n}$ is the shift step of element of matrix $\mathbf{U}^{(n)}$ at (i_n, j_n). In real application, the shift step can be the time delay of ERP across different trials. We rewrite equation (6) in element-wise form

$$\begin{aligned} [\mathbf{X}_{(n)}]_{i_n,p_n} &= \sum_{j_n=1}^{J_n} [\mathbf{U}^{(n)}]_{i_n, j_n + \tau^n_{i_n,j_n}} [\mathbf{Z}^{(n)}]_{j_n,p_n} \\ &= \sum_{j_n=1}^{J_n} [\mathbf{U}^{(n)}]_{i_n,j_n} [\mathbf{Z}^{(n)}]_{j_n, p_n - \tau^n_{i_n,j_n}} \end{aligned} \tag{8}$$

where $p_n = 1, \cdots, L_n$. In equation (8), the model considers component delays in tensor structure and it can be seen as an extension of Shifted NMF proposed in [3]. We can give following equivalent model in frequency domain using FFT

$$[\tilde{\mathbf{X}}_{(n)}]_{i_n,f} = \sum_{j_n=1}^{J_n} [\mathbf{U}^{(n)}]_{i_n,j_n} [\tilde{\mathbf{Z}}^{(n)}]_{j_n,f} e^{-i2\pi \frac{f-1}{L_n} \tau^n_{i_n,j_n}} \tag{9}$$

Above equation (9) can be rewritten in matrix form as

$$[\tilde{\mathbf{X}}_{(n)}]_f = \tilde{\mathbf{U}}^{(n)}_{(f)} [\tilde{\mathbf{Z}}^{(n)}]_f \tag{10}$$

where $\omega_n = 2\pi \frac{f-1}{L_n}$, $\tilde{\mathbf{U}}^{(n)}_{(f)} = \mathbf{U}^{(n)} \circledast e^{-i\omega_n \tau^n}$, $[e^{-i\omega_n \tau^n}]_{i_n,j_n} = e^{-i\omega_n \tau^n_{i_n,j_n}}$, i.e. $\tau^n_{i_n,j_n}$ is the element of τ^n.

In order to find the approximate tensor factorization, we construct following cost function for NSTF model based on least square error

$$\begin{aligned} & \mathbf{J}_{LS}(\mathbf{U}^{(n)}|_{n=1}^{N}, \tau^n|_{n=1}^{N}) \\ = \; & \frac{1}{2L_n} \| [\tilde{\mathbf{X}}_{(n)}]_f - \tilde{\mathbf{U}}^{(n)}_{(f)} [\tilde{\mathbf{Z}}^{(n)}]_f \|_F^2 \end{aligned} \tag{11}$$

We estimate the factor matrices $\mathbf{U}^{(n)}|_{n=1}^{N}$ and delay $\tau^n|_{n=1}^{N}$ by alternating least square method. Multiplicative learning algorithm for factor matrices can be derived by exponential gradient similar to NMF. The monotonic convergence analysis in [14] can be applied to our case as well. The optimization procedure and update rules derivation are described as following in detail.

A. Update $\mathbf{U}^{(n)}|_{n=1}^{N}$

Let $\mathbf{Z}^{(n)}_{(t)}$ denote the shifted version of matrix $\mathbf{Z}^{(n)}$, and corresponding to frequency domain $\tilde{\mathbf{Z}}^{(n)}_{(f)} = \tilde{\mathbf{Z}}^{(n)} \circledast e^{-i\omega_n \tau^n}$. According to equation (8), we have

$$[\mathbf{X}_{(n)}]_{i_n,:} = [\mathbf{U}^{(n)}]_{i_n,:} \mathbf{Z}^{(n)}_{(t)} \tag{12}$$

Because $\mathbf{U}^{(n)}|_{n=1}^{N}$ and $\mathbf{Z}^{(n)}$ are still nonnegative after shifting a given amount, we can use the regular NMF update rules given in [14].

$$[\mathbf{U}^{(n)}]_{i_n,j_n} \leftarrow [\mathbf{U}^{(n)}]_{i_n,j_n} \frac{[\mathbf{X}_{(n)}]_{i_n,:} [\mathbf{Z}^{(n)}_{(t)}]^T_{j_n,:}}{[\mathbf{U}^{(n)}]_{i_n,:} \mathbf{Z}^{(n)}_{(t)} [\mathbf{Z}^{(n)}_{(t)}]^T_{j_n,:}} \tag{13}$$

B. Update $Z^{(n)}$

Considering the model given in equation (10), we can calculate the gradient of the cost function in frequency domain:

$$\begin{aligned}\mathbf{G}_f &= \frac{\partial \mathbf{J}_{LS}}{\partial[\tilde{\mathbf{Z}}^{(n)}]_f} \\ &= -\frac{1}{2L_n}[\tilde{\mathbf{U}}^{(n)}_{(f)}]^H\left([\tilde{\mathbf{X}}_{(n)}]_f - \tilde{\mathbf{U}}^{(n)}_{(f)}[\tilde{\mathbf{Z}}^{(n)}]_f\right) \quad (14)\end{aligned}$$

Similar to [3], we separate the gradient in the frequency domain into two parts which relate to positive or negative component of corresponding gradient. The update rules are given as

$$\tilde{\mathbf{G}}^+_f = \frac{1}{L_n}[\tilde{\mathbf{U}}^{(n)}_{(f)}]^H\tilde{\mathbf{U}}^{(n)}_{(f)}[\tilde{\mathbf{Z}}^{(n)}]_f \quad (16)$$

$$\tilde{\mathbf{G}}^-_f = \frac{1}{L_n}[\tilde{\mathbf{U}}^{(n)}_{(f)}]^H[\tilde{\mathbf{X}}_{(n)}]_f \quad (17)$$

Then we can calculate the inverse FFT of $\tilde{\mathbf{G}}^+_f$ and $\tilde{\mathbf{G}}^-_f$ as $\mathbf{G}^+$ and $\mathbf{G}^-$. As a result, we can update $\mathbf{Z}^{(n)}$ as

$$[\mathbf{Z}^{(n)}]_{j_n,p_n} = [\mathbf{Z}^{(n)}]_{j_n,p_n}\left(\frac{\mathbf{G}^-_{j_n,p_n}}{\mathbf{G}^+_{j_n,p_n}}\right)^\alpha \quad (18)$$

where α is the step size and tuned to keep cost function decreasing.

C. Update $\tau^n|_{n=1}^N$

For the uncertainty of shift factor $\tau^n|_{n=1}^N$, we estimate these shift steps by the Newton-Rhapson method as stated in [4]. As described in equation (11), the least square cost function for NSTF model can be rewritten as equation (15).

Let $\mathbf{T}^n = vec(\tau^n) \in R^{I_nJ_n\times 1}$, i.e. $\mathbf{T}^n_{i_n+(j_n-1)*I_n} = \tau^n_{i_n,j_n}$. Further we define

$$[\tilde{\mathbf{Q}}^n]_{i_n,j_n,f} = [\tilde{\mathbf{U}}^{(n)}_{(f)}]_{i_n,j_n}[\tilde{\mathbf{Z}}^{(n)}]_{j_n,f} \quad (19)$$

$$[\tilde{\mathbf{E}}^n]_f = [\tilde{\mathbf{X}}_{(n)}]_f - \tilde{\mathbf{U}}^{(n)}_{(f)}[\tilde{\mathbf{Z}}^{(n)}]_f \quad (20)$$

Then the gradient of $\mathbf{J}_{LS}$ with respect to $\tau^n_{i_n,j_n}$ is derived as

$$\begin{aligned}\mathbf{g}^n_{i_n+(j_n-1)*I_n} &= \frac{\partial \mathbf{J}_{LS}}{\partial \mathbf{T}^n_{i_n+(j_n-1)*I_n}} \\ &= \frac{\partial \mathbf{J}_{LS}}{\partial \tau^n_{i_n,j_n}} \\ &= -\frac{1}{L_n}\sum_f \omega_n \Im\left([\tilde{\mathbf{Q}}^n]_{i_n,j_n,f}[\tilde{\mathbf{E}}^n]_{i_n,f}\right) \quad (21)\end{aligned}$$

The Hessian matrix is described in equation (26), where $s = i_n + (j_n-1)I_n$ and $s' = i'_n + (j'_n-1)I_n$. Then the shifted steps τ^n can be estimated by the Newton-Raphson method as

$$\mathbf{T}^n \leftarrow \mathbf{T}^n - \eta(\mathbf{H}^n)^{-1}\mathbf{g}^n \quad (22)$$

where the step size parameter η is tuned to guarantee the cost function deceases.

However this updating rule is not stable and sensitive to local minima. So we employ cross-correlation procedure for estimation of shifted steps to reduce the effect of local minima as stated in [3], [4]. Let

$$\tilde{\mathbf{R}}^n_{i_n,f} = [\tilde{\mathbf{X}}_{(n)}]_{i_n,f} - \sum_{j_n=1,j_n\neq j'_n}^{\mathbf{J}_n}[\tilde{\mathbf{U}}^{(n)}_{(f)}]_{i_n,j_n}[\tilde{\mathbf{Z}}^{(n)}]_{j_n,f} \quad (23)$$

Further define $\tilde{c}^n_f = [\tilde{\mathbf{R}}^n_{i_n,f}]^*[\tilde{\mathbf{Z}}^{(n)}]_{j'_n,f}$ as cross-correlation between the j'_nth source and i_nth sensor and we can estimate $\tau^n_{i_n,j'_n}$ as the delay corresponding to maximum cross-correlation.

$$t^n = \arg\max_{p_n} \quad c^n_{p_n}, \quad \tau^n_{i_n,j'_n} = t^n - (L_n+1) \quad (24)$$

By this shifted step, we update $[\mathbf{U}^{(n)}]_{i_n,j'_n}$ by

$$[\mathbf{U}^{(n)}]_{i_n,j'_n} = \frac{c^n_{t^n}}{[\mathbf{Z}^{(n)}]_{j'_n,:}[\mathbf{Z}^{(n)}]^T_{j'_n,:}} \quad (25)$$

In order to ensure finding the optimal solution, we re-estimate $\tau^n|_{d=1}^N$ by the cross-correlation procedure after every 20^{th} iteration. The detailed description of ALS Non-negative Shifted Tensor Factorization Algorithm is presented in Algorithm1.

Algorithm 1: Algorithm of ALS Nonnegative Shifted Tensor Factorization

Data: Training data $\underline{\mathbf{X}} \in R^{I_1\times I_2\times\cdots I_N} \geq 0$, maximum iterations T, error threshold ε.

Result: The projection matrix $\mathbf{U}^{(n)} \geq 0 (l = 1, \cdots, N)$, the estimated shifts τ^n.

```
Initialization: Set U_(0)^(n) ≥ 0 (l = 1, ··· , r) randomly,
  iteration index t = 1, α, η;
repeat
  for n ← 1 to N do
    while t < T or update error e > ε do
      %Update U_(t-1)^(n);
      [U^(n)]_{i_n,j_n} ←
        ([X_(n)]_{i_n,:}[Z_(t)^(n)]^T_{j_n,:} / [U^(n)]_{i_n,:}Z_(t)^(n)[Z_(t)^(n)]^T_{j_n,:}) [U^(n)]_{i_n,j_n}
      %Update Z_(t-1)^(n);
      [Z^(n)]_{j_n,p_n} = [Z^(n)]_{j_n,p_n} (G^-_{j_n,p_n} / G^+_{j_n,p_n})^α,
      %Update τ^n;
      T^n ← T^n − η(H^n)^{-1} g^n,
      if mod(t,20) = 0 then
        Re-estimate τ^n by cross-correlation
        procedure.
until about 20 iterations;
```

IV. Experiments

In this section, we illustrate two experiment results to show the performance of ALS-NSTF algorithm compared with other baseline methods without considering component delays.

$$\mathbf{J}_{LS} = \frac{1}{2L_n}\sum_f \left([\tilde{\mathbf{X}}_{(n)}]_f - \tilde{\mathbf{U}}^{(n)}_{(f)}[\tilde{\mathbf{Z}}^{(n)}]_f\right)^H \left([\tilde{\mathbf{X}}_{(n)}]_f - \tilde{\mathbf{U}}^{(n)}_{(f)}[\tilde{\mathbf{Z}}^{(n)}]_f\right) \tag{15}$$

$$\mathbf{H}^n_{s,s'} = \begin{cases} \frac{1}{L_n}\sum_f \omega_n^2 \Re\left([\tilde{\mathbf{Q}}^n]_{i_n,j_n,f}[\tilde{\mathbf{Q}}^n]^*_{i'_n,j'_n,f}\right) & \text{if } i_n \neq i'_n \wedge j_n \neq j'_n \\ \frac{1}{L_n}\sum_f \omega_n^2 \Re\left([\tilde{\mathbf{Q}}^n_{i_n,j_n,f}\left([\tilde{\mathbf{Q}}^n]^*_{i'_n,j'_n,f} + [\tilde{\mathbf{E}}^n]^*_{i_n,f}\right)\right), & \text{if } i_n = i'_n \wedge j_n = j'_n \end{cases} \tag{26}$$

A. Speaker Identification

Speaker identification is an important technology to find the difference among speakers. While its performance degenerate rapidly in noisy environments, how to extract robust speech features becomes the key to improve identification accuracy. In this experiment, we tested the identification performance on Grid Corpus which was designed for the speech separation and recognition. Speech sentences of 34 speakers were used as training data to build speaker models. In order to extract robust speech features, the cortical-based feature extraction framework proposed in [15], [16] was employed to explore the shift-invariance features for recognition.

The sample rate of speech signal $x(t)$ in Grid Corpus was 8kHz. We used hamming window of 25ms with 10ms shift over each sentence to segment the given speech signal. 256-points short time fourier transformation was employed to calculate power spectrum $\mathbf{P}(f,t)$. We filtered the power spectrum to extract the multiple resolution Gabor tensor features $\underline{\mathbf{G}}$ by four scales and four directions Gabor functions and preserve the efficient frequency component by Mel filter banks. Then we decomposed the higher order cortical representation $\underline{\mathbf{G}}$ by ALS-NSTF algorithm to obtain projection matrices $\mathbf{U}^{(n)}|_{n=1}^{N}$. The projection matrix in frequency mode $\mathbf{U}^{(f)}$ was used to project cortical representation $\underline{\mathbf{G}}$ into feature subspace and obtain the efficient Gabor sparse feature $\underline{\mathbf{F}}$ with shift invariance characteristic. Finally, we matricized $\underline{\mathbf{F}}$ as $\mathbf{F}_m$ and employed Discrete Cosine Transformation (DCT) on feature vectors to remove the correlation of components.

Based on extracted features, we built 34 speaker models by GMM with random selected 1700 sentences (50 sentences for each speaker) and tested the recognition accuracy with 2040 sentences (60 sentences for each speakers). In order to evaluate the robustness of our proposed feature extraction method, the testing speech signals were mixed with Babble, Destroyerops, F16 and Pink noises in SNR intensities of -5 dB, 0dB and 5dB respectively.

We also tested the performance of ALSNTF algorithm, MFCC and Spectral Substraction (SS) as baseline system. Baseline system of ALSNTF algorithm was achieved by similar feature extraction framework as ALS-NSTF algorithm.

The average identification accuracy of ALS-NSTF algorithm and baseline system in all noisy conditions was summarized in Fig.1. The identification performance were tested on four types noises and in three different SNR intensity (-5, 0, 5dB). The final accuracy in each SNR with different noises was averaged on 10 different testing sets. The overall average accuracy was across all the conditions (different noise type and SNR intensity).

As shown in the final results in Fig.1, the testing condition in SNR intensity from -5dB to 5dB was very noisy and the features extracted by ALS-NSTF algorithm provided significantly better performance for Pink and F16 noises. But the recognition accuracy for Babble and Destroyerops noises was improved slight better than MFCC and SS methods. The Babble and Destroyerops noises contained the same frequency characteristic as clean speech signal and contaminated the whole spectral domain component. This caused the identification accuracy in Babble and Destroyerops noises degraded rapidly compared with other noise such as F16. The ALS-NSTF algorithm assumed the processed data was sparse and shift-invariant and made energy of clean signal only concentrate on a few components. After sparse projection, the noisy components without strong energy were reduced. The final identification accuracy also indicated that our proposed method based on ALS-NSTF algorithm provided a better average accuracy than ALSNTF algorithm and other traditional speech feature extraction methods.

B. MRI-based AD Diagnosis

In this experiment, we employed ALS-NSTF algorithm to classify the Alzheimer's Disease (AD) and Health Control (HC) subjects by their sMRI data. It was essential to achieve accurate AD diagnosis for disease pathogenesis and prevention research. The free public brain imaging data from OASIS [17] was employed to test the performance of our proposed feature extraction method and baseline system. 100 AD subjects and 109 HC subjects were selected as the datasets for the classification task.

In pre-processing step, we use SPM8 toolbox to realign and normalize all the sMRI data. 2×2×2 mm^3 voxel-size images were obtained by re-slice processing. Based on all the pre-processed sMRI data, we constructed a 4-order tensor $\underline{\mathbf{X}} \in R^{34\times39\times34\times209}$ to explore the factor correlations of each coordinates x, y, z and subjects. The ALS-NSTF algorithm was employed to decompose higher order tensor $\underline{\mathbf{X}}$ and extract efficient shift-invariance features for diagnosis of AD. We extracted the features of sMRI data for AD and HC subjects using the similar framework as proposed in [13]. The row of factor matrix in subjects mode $\mathbf{U}^{subject} \in R^{209\times30}$ was regarded as feature vector for each subject.

The training feature sets for building AD diagnosis model was randomly selected from 75% feature vectors of AD and HC subjects to train the SVM classifier. The remaining 25% feature vectors of AD and HC subjects were used to test

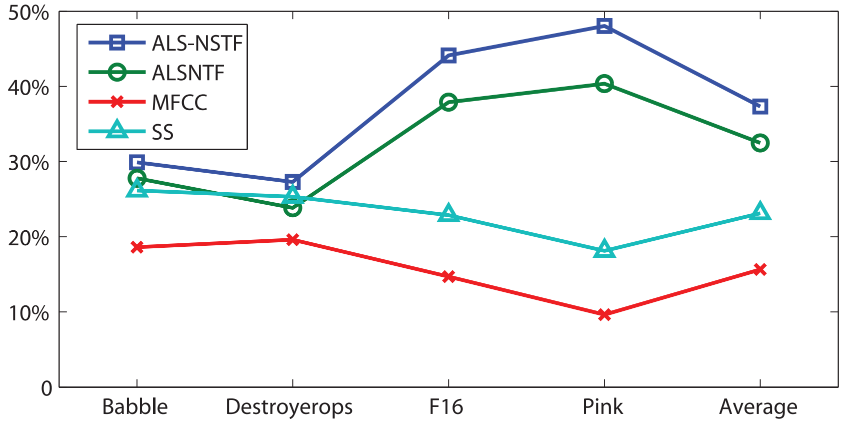

Fig. 1. Average identification accuracy of ALS-NSTF algorithm and baseline systems

the performance of classifier. We repeated the training and testing procedure over 100 times by randomly selecting the training and testing feature vectors to ensure the effective recognition results. Except for recognition accuracy, we also evaluate the sensitivity and specificity of recognition as defined in [13]. In order to compare the performance with other methods, we also provided the recognition accuracy of HALSNTF and Shifted NMF algorithm as baseline system for AD diagnosis. The 4-order tensor $\underline{\mathbf{X}}$ was matricized with two dimension $subjects$ and $samples$ and decomposed by Shifted NMF to extract sMRI features.

By tensor based methods, the essential spatial structure was preserved compared with Shifted NMF method by unfolding or vectorization. Also the component delay assumption of ALS-NSTF algorithm made sure that the higher order shifted projection recovered the localized, parted-based components.

The final recognition results of ALS-NSTF algorithm and baseline system were presented in Fig.2. ALS-NSTF algorithm based method provided better recognition performance than ALSNTF and Shifted NMF algorithm. From the accuracy and sensitivity results, shift-invariance characteristic was important for sMRI data feature extraction. It indicated that our proposed sparse and shift-invariance feature extraction method was potential for dealing with a wider variety sMRI data diagnosis.

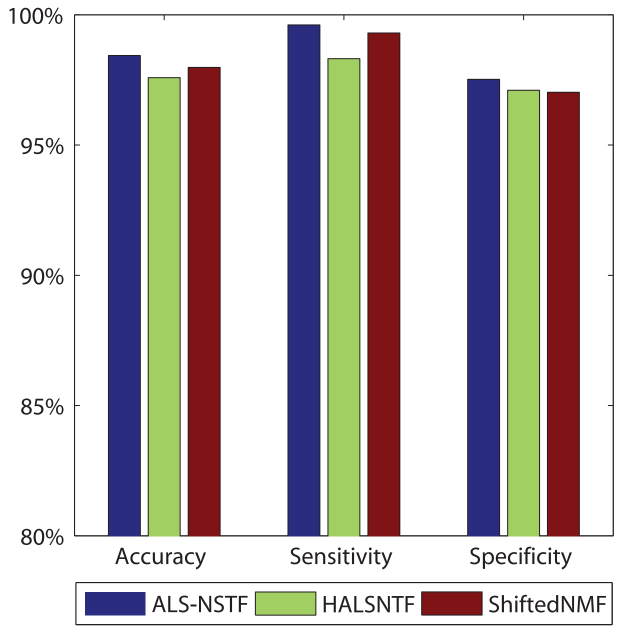

Fig. 2. Average accuracy, sensitivity, specificity results of AD diagnosis based on sMRI

V. Conclusion

In this paper, we presented a new nonnegative tensor factorization method considering component delay. This method was data driven and able to extract sparse and shift-invariance features for higher order complex data such as audio and brain imaging data. In order to estimate the component shift in different modes, we transform the non-integer delays to frequency domain by FFT and explore multiple linear patterns with shift-invariance characteristic. Based on the alternating least square procedure, the component delays/shifts and optimal solution of factorization matrices in each mode were estimated by Newton-Rhapson and exponential gradient method respectively. We evaluated the feature extraction performance of our proposed algorithm on noisy speech and AD sMRI data. Experimental results showed that NSTF model reduced the effect of component delays to some extent and improved the robustness and discriminant of multiple factors patterns for speaker identification in noisy condition and AD diagnosis.

References

[1] S.Choi, A.Cichocki, H.M.Park, et al. "Blind source separation and independent component analysis: A review," *Neural Information Processing-Letters and Reviews,* vol.6, no.1, pp.1-57, 2005.

[2] S.C.Douglas, H,Sawada, S.Makino, "A spatio-temporal fastICA algorithm for separating convolutive mixtures," *IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp.165-168, 2005.

[3] M.Mørup, K.H.Madsen, L.K.Hansen, "Shifted non-negative matrix factorization, " *In Machine Learning for Signal Processing, 2007 IEEE Workshop on*, pp.139-144, 2007.

[4] M.Mørup, K.H.Madsen, L.K.Hansen, "Shifted independent component analysis," *Independent Component Analysis and Signal Separation*, pp.89-96, 2007.

[5] V.K.Potluru, S.M.Plis, V.D.Calhoun, "Sparse shift-invariant NMF," *Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on*, pp.69-72, 2008.

[6] A.H. Andersen and W.S. Rayens, "Structure-seeking multilinear methods for the analysis of fMRI data," *Neuroimage,* vol. 22, no. 2, pp. 728-739, 2004.

[7] M. Mørup, L.K. Hansen, S.M. Arnfred, L.H. Lim, K.H. Madsen, "Shift-invariant multilinear decomposition of neuroimaging data," *NeuroImage,* vol. 42, no. 4, pp. 1439-1450, 2008.

[8] J. Li, L.Q. Zhang, D.C Tao, H. Sun, Q.B. Zhao, "A prior neurophysiologic knowledge free tensor-based scheme for single trial EEG classification," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 17, no. 2, pp. 107-115, 2009.

[9] A.Cichocki, R. Zdunek, A.H. Phan, S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation,* Wiley, 2009.

[10] M.I. Sereno, A.M. Dale, J.B. Reppas, K.K. Kwong, J.W. Belliveau, T.J. Brady, B.R. Rosen, R.B. Tootell, "Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging," *Science,* vol. 268, no. 5212, pp. 889-893, 1995.

[11] R.B. Buxton, E.C. Wong, L.R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model," *Magnetic Resonance in Medicine,* vol. 39, no. 6, pp. 855-864, 1998.

[12] R.A. Harshman, S. Hong, M.E. Lundy, "Shifted factor analysisłPart I: Models and properties, "*Journal of chemometrics*, vol. 17, no. 7, pp. 363-378, 2003.

[13] Q. Wu, L.Q. Zhang, A. Cichocki, "Multifactor sparse feature extraction using Convolutive Nonnegative Tucker Decomposition," *Neurocomputing,*, vol. 129, pp. 17-24, 2014.

[14] D.D. Lee, S.H. Seung, "Algorithms for Non-negative Matrix Factorization, " *Advances in Neural Information Processing Systems,* vol.13, pp.556-562, 2000.

[15] Q. Wu, L.Q. Zhang, G.C. Shi, "Robust feature extraction for speaker recognition based on constrained nonnegative tensor factorization," *Journal of Computer Science and Technology,* vol. 25, no. 4, pp. 783-792, 2010.

[16] Q. Wu, L.Q. Zhang, G.C. Shi, "Robust Multifactor Speech Feature Extraction Based on Gabor Analysis," *Audio, Speech, and Language Processing, IEEE Transactions on,* vo. 19, no.4, pp.927-936, 2011.

[17] D.S. Marcus, T.H. Wang, J.Parker, J.G. Csernansky, J.C. Morris and R.L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience,* vol. 19, no. 9, pp.1498-1507,2007.