β-Divergence Two-Dimensional Nonnegative Matrix Factorization with Sparseness Constraints for Biomedical Signal Separation

A. M. Darsono[1], C. C. Toh[1], S. Saat[1], N. A. Manap[1], M. M. Ibrahim[1] and M. I. Ahmad[2]
[1]*Machine Learning & Signal Processing Research Group (MLSP), Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka Durian Tunggal, Melaka, Malaysia.*
[2]*School of Computer and Communication Engineering, Universiti Malaysia Perlis, Perlis, Malaysia.*
abdmajid@utem.edu.my

***Abstract*—A novel of β-Divergence for nonnegative matrix factorization two-dimensional (NMF2D) with sparseness constraints is proposed in this paper. This research focuses on biomedical signal separation, which denotes a separation on the mixture of heart sound and lung sound. Initially, a mixture of heart sound and lung sound has been decomposed into an independent signal, which is an estimated heart sound signal and estimated lung sound signal. The spectrum of independent signal is modelled based on 2 dimensions, which are the temporal code and the spectral basis by using β-Divergence NMF2D algorithm with sparseness constraints. The algorithm has been updated multiplicative and iteratively via multiplicative update rules (MU rules). β-Divergence with sparseness constraints allows minimization on the vagueness of source model to be completed and oneness has been applied to it. Then, estimation of each separated audio has been analyzed via comparison with the original heart sound and lung sound signal in term of Signal-to-Distortion Ratio (SDR).**

***Index Terms*—Nonnegative Matrix Factorization; Sparseness Constraints; Β-Divergence; Multiplicative Rules.**

I. INTRODUCTION

Nonnegative Matrix Factorization (NMF) [1,2] plays an important role in various applications, such as automatic music transcription [3], cryptography [4], pattern recognition [5], biomedical field [6] and etc. We are more concerned on the heart and lung sound separation in biomedical field. There are several types of information used to analyze the lung condition: One of the non-invasive diagnoses is based on the lung sound information, which is a valuable indicator of respiratory health and disease [7]. To ease in the diagnosis of lung condition, it is better to have a clear sound of the lung. However, the lung sound will be interfered by the normal heart sound, which leads to the difficulty of diagnosis. In order to solve this issue, the β-Divergence NMF with sparseness constraints is imposed in the heart and lung sound separation.

Recently, advancement of NMF is extended to NMF2D model [8] to provide separation that can efficiently capture the temporal dependency of the frequency patterns within the source. The time-frequency (TF) has been modeled as temporal code and spectral basis for each source in NMF2D, in which the temporal code and spectral basis is known as two-dimension. Furthermore, adding sparseness constraint to NMF2D is more effective compared to the normal usage of NMF2D. This is due to the constraints of sparseness that allow to reduce the vagueness of source model and to complete the estimation of decomposition in low distortion or noise. One of the advantages of the properties of NMF is it usually produces a sparse representation of the data. This sparse representation or better known as sparse coding has a representation encodes much of the data using few 'active' components, which eases the construction of the encoding [9].

II. MATHEMATICAL FORMULATION

A. General Framework of β-Divergence

For NMF, a series of famous cost function has been utilized for years such as the Least Square (LS) divergence, the Kullback-Leibler (KL) divergence and the Itakura-Saito (IS) divergence. All of these selected divergences are categorized into a singleness framework, which is β-Divergence. The framework of β-Divergence is shown in the following equation:

$$d_\beta(V|\Lambda) = \begin{cases} \dfrac{V^\beta}{\beta(\beta-1)} + \dfrac{\Lambda^\beta}{\beta} - \dfrac{V\Lambda^{\beta-1}}{\beta-1} & \beta \in \Re\ \{0,1\} \\ \dfrac{1}{2}(V-\Lambda)^2 & \beta = 2 \\ V log\left(\dfrac{V}{\Lambda}\right) + \Lambda - V & \beta = 1 \\ \dfrac{V}{\Lambda} - \log\left(\dfrac{V}{\Lambda}\right) - 1 & \beta = 0 \end{cases} \quad (1)$$

where $d_\beta(y|x)$ is the scalar cost function. The LS divergence, the KL divergence and the IS divergence represent β=2, β=1, β=0 respectively.

Obviously, LS divergence substituted in the general family of β-divergence will be transposed into the equation below:

$$\mathrm{C_{LS}} = ||\Lambda - \mathrm{V}||_\mathrm{f}^2 \quad (2)$$

$$= \sum_\mathrm{i}\sum_\mathrm{j}(\Lambda_\mathrm{ij} - V_\mathrm{ij})^2 \quad (3)$$

In contrast, the equation will be changed when the KL divergence has been applied,

$$C_{KL} = \sum_i\sum_j \Lambda\, log\frac{\Lambda_{i,j}}{V_{i,j}} - \Lambda_{i,j} + V_{i,j} \quad (4)$$

Meanwhile, it then turned over into equation below when IS divergence has been applied,

$$C_{IS} = \sum_i \sum_j \frac{\Lambda_{i,j}}{V_{i,j}} - log \frac{\Lambda_{i,j}}{V_{i,j}} - 1 \quad (5)$$

The BSS in this paper is classified into single channel source separation (SCSS). In time domain, the model of SCSS is defined as:

$$V(t) = \sum_{j=1}^{J} \Lambda_j(t) + e(t) \quad (6)$$

It then change into time-frequency domain via Short Time Fourier Transform (STFT),

$$V(t) = \sum_{j=1}^{J} \Lambda_{j.f,n} + e_{f,n} \quad (7)$$

where j=1,2,3,…,J denotes the amount of source, e(t) denotes the additional interference, f=1,2,3,…,F denotes the frequency bin and n=1,2,3,…,N denotes the time frame index.

$$\left|X_j\right|^2 = \sum_{\tau=0}^{\tau_{max}} \sum_{\phi=0}^{\phi_{max}} \overset{\downarrow \phi}{W_j^{\tau}} \overset{\rightarrow \tau}{H_j^{\phi}} \quad (8)$$

The matrix W shows the τ^{th} slice spectral basis and H shows the ϕ^{th} slice of temporal code for each spectral basis element. Arrow of $\overset{\downarrow \phi}{W_j^{\tau}}$ shows the shifting in each element by ϕ row down and arrow of $\overset{\rightarrow \tau}{H_j^{\phi}}$ shows the shifting in each element by τ column right [10].

B. Multiplicative Rules with sparseness constraints

Sparseness constraints denote a matrix in which most of the elements are zero. In other words, the matrix is considered dense if most of the elements are nonzero and vice versa. In this paper, we deployed multiplicative update (MU) [11] rules on the β-divergence using adding multiplicative gradient descent method, separating positive and negative terms and observing the fixed point of the update rules, which is reached when the sparse cost function reaches the minimum [12].

Now, we associated the β-divergence as defined in (1) with the sparsity constrained so that it will minimize the cost function as follow:

$$C\left(|\Lambda| \| \tilde{V}\right) = \sum_{f,n}\left(\frac{\left(|\Lambda|_{f,n}^2\right)^{\beta}}{\beta(\beta-1)} + \frac{(\tilde{V})^{\beta}}{\beta} - \frac{|\Lambda|_{f,n}^2 (\tilde{V})^{\beta-1}}{\beta-1}\right) + \lambda f(H) \quad (9)$$

where$\tilde{V}$ = $\sum_{j,\tau,\phi} \overset{\downarrow \phi}{W_j^{\tau}} \overset{\rightarrow \tau}{H_j^{\phi}}$ with $\widetilde{W}_{f,j}^{\tau} = \frac{W_{f,j}^{\tau}}{\sqrt{\sum_{\tau,f}\left(W_{f,j}^{\tau}\right)^2}}$ in $f = 1, \dots, F$, $n = 1, \dots, N$ and parameter λ is the sparsity constraint.

Then, the derivatives of (9) are given by:

$$C\left(|\Lambda| \| \tilde{V}\right) = \sum_{f,n}\left(\frac{\left(|\Lambda|_{f,n}^2\right)^{\beta}}{\beta(\beta-1)} + \frac{(\tilde{V})^{\beta}}{\beta} - \frac{|\Lambda|_{f,n}^2 (\tilde{V})^{\beta-1}}{\beta-1}\right) + \lambda f(H) \quad (9)$$

where $\tilde{V} = \sum_{j,\tau,\phi} \overset{\downarrow \phi}{W_j^{\tau}} \overset{\rightarrow \tau}{H_j^{\phi}}$ with $\widetilde{W}_{f,j}^{\tau} = \frac{W_{f,j}^{\tau}}{\sqrt{\sum_{\tau,f}\left(W_{f,j}^{\tau}\right)^2}}$ in $f = 1, \dots, F$, $n = 1, \dots, N$ and parameter λ is the sparsity constraint.

Then, the derivatives of (9) are given by:

$$\frac{\delta C_{\beta}}{\delta W_{f',j'}^{\tau'}} = \frac{\delta}{\delta W_{f',j'}^{\tau'}}\left(\sum_{f,n}\left(\frac{\left(|\Lambda|_{f,n}^2\right)^{\beta}}{\beta(\beta-1)} + \frac{\left(\tilde{V}_{f,n}\right)^{\beta}}{\beta} - \frac{|\Lambda|_{f,n}^2\left(\tilde{V}_{f,n}\right)^{\beta-1}}{\beta-1}\right) + \lambda f(H)\right) \quad (10)$$
$$= \sum_{\phi,n}\left(\left(\tilde{V}_{f'+\phi,n}\right)^{\beta-1} - |\Lambda|_{f'+\phi,n}^2\left(\tilde{V}_{f'+\phi,n}\right)^{\beta-2}\right) H_{j',n-\tau'}^{\phi}$$

$$\frac{\delta C_{\beta}}{\delta H_{j',n'}^{\phi'}} = \frac{\delta}{\delta H_{j',n'}^{\phi'}}\left(\sum_{f,n}\left(\frac{\left(|\Lambda|_{f,n}^2\right)^{\beta}}{\beta(\beta-1)} + \frac{\left(\tilde{V}_{f,n}\right)^{\beta}}{\beta} - \frac{|\Lambda|_{f,n}^2\left(\tilde{V}_{f,n}\right)^{\beta-1}}{\beta-1}\right) + \lambda f(H)\right) \quad (11)$$
$$= \sum_{\tau,f} \widetilde{W}_{f-\phi',j'}^{\tau}\left(\left(\tilde{V}_{f,n'+\tau}\right)^{\beta-1} - |\Lambda|_{f,n'+\tau}^2\left(\tilde{V}_{f,n'+\tau}\right)^{\beta-2}\right) + \lambda \frac{\delta f(H)}{\delta H_{j',n'}^{\phi'}}$$

It is a method to update the parameters iteratively and formula for gradient descent method as shown below [13], [14]:

$$W_{f',j'}^{\tau'} \leftarrow \widetilde{W}_{f',j'}^{\tau'} - \eta_W \frac{\delta C_{\beta}}{\delta W_{f',j'}^{\tau'}} \quad (12)$$

$$H_{j',n'}^{\phi'} \leftarrow \widetilde{H}_{j',n'}^{\phi'} - \eta_H \frac{\delta C_{\beta}}{\delta H_{j',n'}^{\phi'}} \quad (13)$$

Therefore, the multiplicative rules become:

$$W^{\tau} \leftarrow \widetilde{W}^{\tau} \cdot \frac{\sum_{\phi}\left[\left(\overset{\uparrow \phi}{\tilde{V}}\right)^{\beta-2} \cdot \overset{\uparrow \phi}{|\Lambda|^2}\right] \overset{\rightarrow \tau}{H^{\phi}}^{T}}{\sum_{\phi}\left(\overset{\uparrow \phi}{\tilde{V}}\right)^{\beta-1} \overset{\rightarrow \tau}{H^{\phi}}^{T}} \quad (14)$$

$$H^{\phi} \leftarrow H^{\phi} \cdot \frac{\sum_{\tau} \overset{\downarrow \phi}{W^{\tau}}^{T}\left[\left(\overset{\leftarrow \tau}{\tilde{V}}\right)^{\beta-2} \cdot \overset{\leftarrow \tau}{|\Lambda|^2}\right]}{\sum_{\tau} \overset{\downarrow \phi}{W^{\tau}}^{T}\left(\overset{\leftarrow \tau}{\tilde{V}}\right)^{\beta-1} + \lambda \frac{\delta f(H)}{\delta H^{\phi}}} \quad (15)$$

where β ∈ R {0,1} from framework β-divergence. Figure 1 shows the procedure of updating the proposed β–divergence.

Initialization on W and H
Loop on divergence
for updating on β-divergence with sparseness constraints

$$W^{\tau} \leftarrow \widetilde{W}^{\tau} \cdot \frac{\sum_{\phi}\left[\left(\overset{\uparrow \phi}{\tilde{V}}\right)^{\beta-2} \cdot \overset{\uparrow \phi}{|\Lambda|^2}\right] \overset{\rightarrow \tau}{H^{\phi}}^{T}}{\sum_{\phi}\left(\overset{\uparrow \phi}{\tilde{V}}\right)^{\beta-1} \overset{\rightarrow \tau}{H^{\phi}}^{T}}$$

$$H^{\phi} \leftarrow H^{\phi} \cdot \frac{\sum_{\tau} \overset{\downarrow \phi}{W^{\tau}}^{T}\left[\left(\overset{\leftarrow \tau}{\tilde{V}}\right)^{\beta-2} \cdot \overset{\leftarrow \tau}{|\Lambda|^2}\right]}{\sum_{\tau} \overset{\downarrow \phi}{W^{\tau}}^{T}\left(\overset{\leftarrow \tau}{\tilde{V}}\right)^{\beta-1} + \lambda \frac{\delta f(H)}{\delta H^{\phi}}}$$

Cost function for β=0 to 1 & 2 of β-divergence

$$d_{\beta}(V|\Lambda) = \begin{cases} \frac{V^{\beta}}{\beta(\beta-1)} + \frac{\Lambda^{\beta}}{\beta} - \frac{V\Lambda^{\beta-1}}{\beta-1} & \beta \in \Re\ \{0,1\} \\ \frac{1}{2}(V-\Lambda)^2 & \beta = 2 \\ Vlog\left(\frac{V}{\Lambda}\right) + \Lambda - V & \beta = 1 \\ \frac{V}{\Lambda} - \log\left(\frac{V}{\Lambda}\right) - 1 & \beta = 0 \end{cases}$$

Reconstruct V=WH
if not meet the stopping criterion
return repeat divergence
end if
end for
until reconstruct Λ

Figure 1: Procedure on updating of β–divergence iteratively

III. Experimental Results

A. Experiment Setup

All simulations and analyses were run via PC with Intel Core 2 Duo CPU 6750 at 2.66 GHz and 4GB RAM as well as laptop with Intel Core i5 CPU 5200 at 2.2GHz and 4GB RAM. MATLAB 2010 was used as the programming platform to run the algorithm. The mixed signal which mingled with the heart sound and lung sound was sampled at 44.1 kHz sample rates. The time-frequency (TF) domain was computed by using short-time Fourier transform (STFT) [15] via 2048 point Hanning window FFT [16] and the frequency domain which was then logarithmically scaled. The convolutive components in time and frequency were selected to be $\tau_{max} = 3$ and $\phi_{max} = 31$ for every case after conducted Monte-Carlo experiments with 50 independent realization of mixture. Performance of results was processed through comparison of estimated to original audio signal in term of Signal-to-Distortion ratio (SDR) [17].

Firstly, by using NMF2D which is the proposed approach, we conducted several experiments that NMF2D without the sparseness constraints to investigate the result of different *β* value of *β*-divergence framework. Secondly, we repeated the experiments using the same approach, but we included sparseness constraints.

B. Original Signal

As shown in Figure 2 and Figure 3, it can be noted that the portion of waveform of lung sound was momentously larger than the heart sound when the heart and lung sound convolutive were mixed. This is due to the reason that the time interval for each period of heartbeat is significantly short compared to the period of respiration.

Therefore, overlapping of heart and lung sound occurred, which causes a challenge for BSS. The shade of color in spectrogram is changing upon the intensity of the sound or audio. The deep black color indicates the highest intensity of sound and vice versa, the light grey indicates that the sound intensity is fading out. As a result, the deeper the color representation, the higher the intensity of sound; hence, the signal received by receiver becomes more significant.

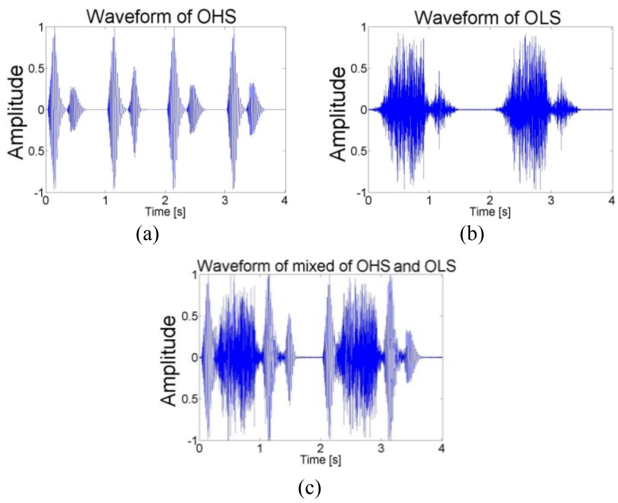

Figure 2: Time domain representation of (a) original heart sound, (b) original lung sound, (c) mixed of original heart sound and original lung sound

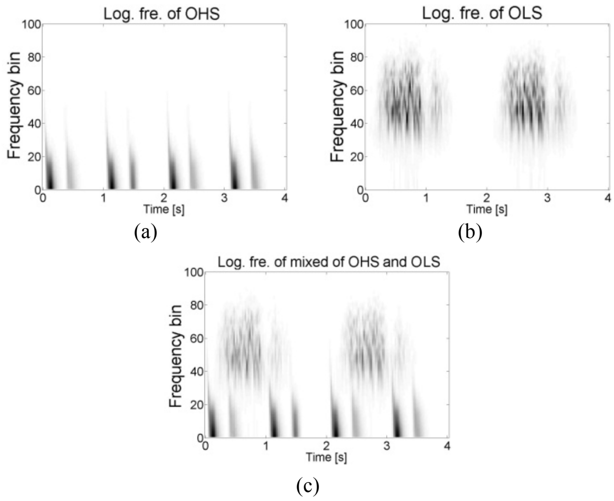

Figure 3: TF representation of (a) log. frequency of original heart sound, (b) log. frequency of original lung sound, (a) log. frequency of mixed of original heart sound and original lung sound

C. Initialization of *β*

We deployed the algorithm to examine the optimal *β* value by using MU rules. Table 1 shows the graph of *β* of *β*-divergence against SDR value with the step size of 0.1 in between *β*=0 until *β*=1. It should be embodied the general framework of *β*-divergence, which includes IS divergence, Kl divergence and LS divergence. We found that SDR at β=0.8 is considered high compared to the other β values as it was over 15 dB on its average. Therefore, it is recommended to use *β*=0.8 as the following phase to inspect the influence of *λ(H)* or sparseness constraints of NMF2D to algorithm.

Table 1
The influences of sparseness constraints to the SDR of output with step size of 0.1

Sparseness constraints (λ)	Average of SDR of output 1 & 2 (dB)
1	8.4212
1.5	9.3625
2	16.4195
2.5	17.0557
3	12.1395
3.5	12.0823
4	12.0675
4.5	11.8028

D. Additional of sparseness constraints

In order to investigate the outcome of sparseness constraints against SDR, *β*=0.8 has been located as a constant variable. The step size of the *λ* equals to 0.1, which is used for inspection starts from 1 until 4.9. From Figure 4, it reveals the normal distribution graph or named as the bell-curve graph informally. According to Figure 4 and Table 1, the peak point is at *λ*=2.5 which contains SDR=17.0557dB. It started with an increment from *λ*=1 until *λ*=2. Then, it presents a high SDR with an approximate constant, which advanced to *λ*=2.6 and a tail-off occurred at *λ*=3.7 with 12.2421dB. Finally, it is constantly reduced until the end of *λ*. Normally, SDR over than 10dB is considered decent in the performance of decomposition, in which the SIR and SAR will be over 10dB incidentally as well. For without sparseness constraints in NMF2D, the SDR presents 16.6886dB. In comparison, NMF2D with sparseness constraints has high SDR to NMF2D without sparseness constraints with the increment of 0.3671dB or 2.2%.

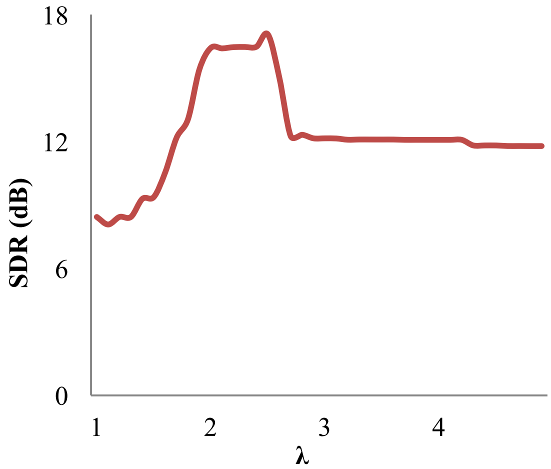

Figure 4: SDR (dB) against sparseness constraints

Figure 5 reveals heart sound and lung sound after decomposition with and without the aids of sparseness constraints. Figure 5(a) and Figure 5(c) are nearly the same, indicating that there is not much changes for the intensity via observation; hence, it does not need further discussion.

Figure 5(b) shows a poor result as the intensity of sound is shown blurrily (marked as red box) compared to Figure 5(d), which has strong and hue saturated (marked as blue box) of the intensity of sound. This is due to the vagueness arises, while there is a drought of sparseness constraints. In contrast, aids of sparseness constraints will enable it to qualitatively provides exceptional decomposition. Therefore, obviously, the sound that has been separated is having high intensity through divergence if it is being applied by the sparseness constraints. The higher the intensity of sound, which means the color in divergence is deeper, the better separation performance is produced. From the above statements, it can be deduced that the sparseness constraints support the ability of NMF2D to produce momentous effect on it.

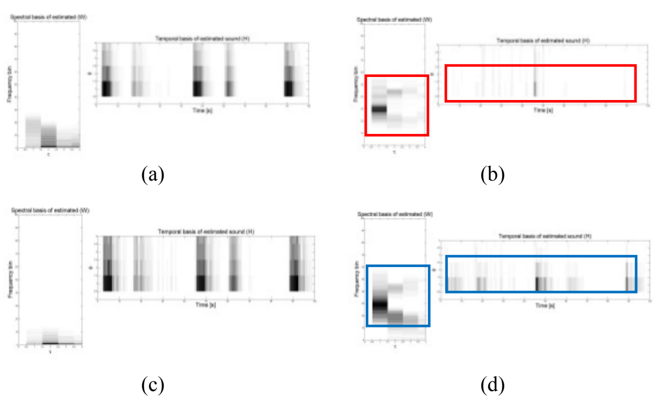

Figure 5: The estimated W and H without sparseness constraints for heart sound (a) and lung sound (b), and with sparseness constraints for heart sound (c) and lung sound (d)

IV. CONCLUSIONS

In conclusion, through this experiment, we validated that the β=0.8 is the optimal result from the family of β-divergence in term of SDR, SIR and SAR. In conjugate, the sparseness constraints added, which is λ=2.5 when the β-0.8 is fixed. The λ=2.5 is considered as the peak point among all λ values which enable it to eliminate the ambiguity and vagueness issue; hence increases the SDR. Therefore, the additional of sparseness constraints is certainly performing well in blind source separation with high SDR. In the future, we believe that these constraints will become helpful in various applications, which is addressed by NMF2D beyond audio source separation.

ACKNOWLEDGMENTS

The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Education Malaysia for the research grant funding RAGS/2013/FKEKK/TK02/04/B0033 that makes this research work possible.

REFERENCES

[1] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5:111–126, 1994.
[2] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755):788–791, 1999.
[3] D. FitzGerald, "Automatic drum transcription and source separation," Ph.D. thesis, Dublin Institute of Technology, Dublin, Ireland, 2004.
[4] S. Xie, Z. Yang, and Y. Fu, "Nonnegative matrix factorization applied to nonlinear speech and image cryptosystems," IEEE Trans. on Circuits and Systems I, vol. 55, no. 8, pp. 2356-2367, Sep 2008.
[5] I. Biciu, N. Nikolaidis, and I. Pitas, "Nonnegative matrix factorization in polynomial feature space", IEEE Trans. On Neural Network, vol. 19, pp. 1090-1100, 2007.
[6] C. S. Lin and E. Hasting, "Blind Source Separation of Heart and Lung Sounds Based on Nonnegative Matrix Factorization," Intelligent Signal Processing and Communications Systems (ISPACS), 2013.
[7] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds: Advances beyond the stethoscope," Amer. J. Respir. Crit. Care Med., vol. 156, pp. 974-987, 1997.
[8] M. Morup and M. N. Schmidt, "Sparse nonnegative matrix factor 2-D deconvolution," Techical Report, Technical University of Denmark, Copenhagen, Denmark, 2006.
[9] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," Journal of the Optical Society of America, 4:2379–2394, 1987.
[10] M. Morup, M. N. Schmidt, "Sparse nonnegative matrix factor 2-D deconvolution, " Technical Report, Technical University of Denmark, Copenhagen, Denmark, 2006.
[11] S. K. Tjoa and K. J. Ray Liu, "Multiplicative Update Rules for Nonnegative Matrix Factorization with Co-Occurrence Constraints," IEEE International Conference on Acoustics, Speech and Signal Processing, 2010.
[12] L. Sui, X. Zhang, J. Huang, G. Zhao and Y. Yang, "Speech Enhancement Based on Sparse Nonnegative Matrix Factorization with Priors," Systems and Informatics (ICSAI), 2012.
[13] A.M. Darsono, N. Z. Haron, A. S. Jaafar and M. I. Ahmad, "β-Divergence Two-Dimensional Sparse Nonnegative Matrix Factorization for Audio Source Separation," Wireless Sensor (ICWISE), 2013.
[14] A. M. Darsono, N.Z. Haron, Shakir Saat, M.M. Ibrahim, and N.A. Manap, "Blind Audio Source Separation with Sparse Nonnegative Matrix Factorization", Research Journal of Applied Sciences, Engineering and Tchnology, vol. 7 No 23, 2014.
[15] A. Ozerov, and C. Févotte, "Multichannel Nonnegative Matrix Factorization in Convolutive Mixtures for Audio Source Separation," IEEE Transactions On Audio, Speech, And Language Processing, Vol. 18, No. 3, March 2010.
[16] S. R. Chintakindi, O. V. S. R. Varaprasad and D. V. S. S. Siva Sarma, "Improved Hanning window based interpolated FFT for power harmonic analysis," TENCON 2015, IEEE Region 10 Conference, 2015
[17] C. F´evotte, R. Gribonval and E. Vincent, "BSS EVAL Toolbox User Guide", IRISA Technical Report 1706, Rennes, France, April 2005. http://www.irisa.fr/metiss/bss eval/.