# β-Divergence Nonnegative Matrix Factorization on Biomedical Blind Source Separation

A. M. Darsono, C. C. Toh, M. S. Md Saat, A. A. M. Isa, N.A. Manap, M.M. Ibrahim

*Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka,*
*Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia*
*abdmajid@utem.edu.my*

*Abstract*—**β-divergence has been studied for years, but it is yet to be discovered thoroughly. In this paper, we proposed the nonnegative matrix factorization (NMF) by using β-divergence in blind source separation (BSS) on biomedical field. The proposed idea is basically aimed at the separation of normal heart sound with normal lung sound. Temporal codes and spectral basis were modelled into a separated source, which is applied to the synthesis and real life data using multiplicative update rules. In the experiment, estimated and original source were compared to evaluate the performance of various source separation algorithms within a general framework, where the original sources and the noise that perturbed the mixture were included.**

*Index Terms*—**Blind Source Separation; Nonnegative Matrix Factorization; β-Divergence; KL Divergence; LSE Divergence.**

## I. INTRODUCTION

In the past decades, many approaches or models have been utilized in the blind source separation (BSS) [1-2] techniques, such as the nonnegative matrix factorization (NMF) [3-6], independent component analysis (ICA) [7-9], sparse decompositions (SD) [10] and computational auditory scene analysis (CASA) [11-13].

NMF is one of the most promising ways and has been applied in different fields, such as speech enhancement, biomedical image processing, biomedical signal processing, remote sensing, communication system and neural networks. In the biomedical field, it is difficult to achieve a fully clear sound of a lung due to the interference of the heart sound in term of the time domain and the spectral content during recording [14]. A normal lung sound is the breathing-related sound heard from the chest of a healthy person, which mingles with the muscle and the cardiovascular sounds. The frequency of the lung sound is hard to differentiate with the frequency of the heart sound as it is dependent to the frequency of the lung sound, which is below 100 Hz, whereby the frequency of the heart sound is normally within the range of 24 to 104 Hz for the first ventricle heart sound and 24 to 144 Hz for the second ventricle heart sound [14, 15].

## II. PROPOSED METHOD

### A. Source Model

In NMF, the matrix notation $V = [v_1 \dots v_n]$, in which matrix V will be factorized into low rank matrices $V = WH$. It does make sense to assume W and H to be nonnegative while observation of V is nonnegative in many applications [16,17]. Different cost functions are used in

different application, and the functions normally used to add to the NMF are the Least Square distance (LS), Kullback-Leibler (KL) and Itakura-Saito (IS). These divergences can be generalized into β-divergence framework [18, 19] as shown in Equation (1),

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

where $d_\beta(y|x)$ is the scalar cost function. The IS divergence, KL divergence and LS divergence represent the limit cases of $\beta$ =0,1,2 which are the underlying multiplicative Gamma observation noise, Poisson noise and Gaussian additive observation noise respectively. Obviously, when $\beta = 2$, the general family of β-divergence will be transposed into Equation (2) and (3).

$$C_{LS} = \left\| \Lambda - V \right\|_f^2 \quad (2)$$

$$= \sum_i \sum_j \left( \Lambda_{ij} - V_{ij} \right)^2 \quad (3)$$

In contrast, the equation will be changed when $\beta = 1$.

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

Meanwhile, it then turns over into Equation (5) when $\beta = 0$.

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

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

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

It then changed 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.

$$|X_j|^2 = \sum_{\tau=0}^{\tau_{max}} \sum_{\phi=0}^{\phi_{max}} \overset{\downarrow \phi}{W_j^\tau} \overset{\rightarrow \tau}{H_j^\phi} \tag{8}$$

The matrix W shows the $\tau^{th}$ slice spectral basis, and H shows the $\phi^{th}$ slice of temporal code for each spectral basis element. The arrow of $\overset{\downarrow \phi}{W_j^\tau}$ show the shifting of each element by $\phi$ row down, while the arrow of $\overset{\rightarrow \tau}{H_j^\phi}$ show the shifting of each element by $\tau$ column right [21].

The recent works on NMF is a nonnegative matrix factorization 2-dimensional (NMF2D) model, which is the extension of the NMF in order to supply the decomposition that can capture temporal dependency of the frequency pattern. NMF2D relies on temporal code and spectral basis, which are known as 2-dimensional (2D) or time-frequency domain, and it allows several components to be reduced dramatically to fasten the progress [20].

### B. Multiplicative Update Rules

In this paper, we deployed the multiplicative update (MU) rules on the β-divergence by preliminary adding multiplicative gradient descent method. It is a method that updates parameters iteratively and a formula for gradient descent method, as shown in Equation (9) and (10) [21]. Furthermore, the MU rules have been concluded in Table 1. The steps of separation of mingle audio source are shown in Table 2.

$$W_{f',j'}^{\tau'} \quad \leftarrow \quad \widetilde{W_{f',j'}^{\tau'}} - \eta_W \frac{\delta C_\beta}{\delta W_{f',j'}^{\tau'}} \tag{9}$$

$$H_{f',j'}^{\phi'} \quad \leftarrow \quad \widetilde{H_{f',j'}^{\phi'}} - \eta_H \frac{\delta C_\beta}{\delta H_{f',j'}^{\phi'}} \tag{10}$$

Table 1
The algorithm of β-divergence with different value of β [21]

| β | Divergence | W | H |
|---|---|---|---|
| $\beta \in \Re \{0,1\}$ | - | $W^\tau \leftarrow \widetilde{W^\tau} \cdot \dfrac{\Sigma_\phi \left[ \left( \overset{\uparrow \phi}{\tilde{V}} \right)^{\beta-2} \cdot \dfrac{\overset{\uparrow \phi}{\vphantom{V}}}{\|\Lambda\|^2} \right]^{\rightarrow \tau T} H^\phi}{\Sigma_\phi \left( \overset{\uparrow \phi}{\tilde{V}} \right)^{\beta-1 \rightarrow \tau T} H^\phi}$ | $H^\phi \leftarrow H^\phi \cdot \dfrac{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T \left[ \left( \overset{\leftarrow \tau}{\tilde{V}} \right)^{\beta-2} \cdot \dfrac{\overset{\leftarrow \tau}{\vphantom{V}}}{\|\Lambda\|^2} \right]}{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T \left( \overset{\leftarrow \tau}{\tilde{V}} \right)^{\beta-1}}$ |
| 2 | LS | $W^\tau \leftarrow \widetilde{W^\tau} \cdot \dfrac{\Sigma_\phi \left[ \dfrac{\overset{\uparrow \phi}{\vphantom{V}}}{\|\Lambda\|^2} \right]^{\rightarrow \tau T} H^\phi}{\Sigma_\phi \left( \overset{\uparrow \phi}{\tilde{V}} \right)^{1 \rightarrow \tau T} H^\phi}$ | $H^\phi \leftarrow H^\phi \cdot \dfrac{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T \left[ \dfrac{\overset{\leftarrow \tau}{\vphantom{V}}}{\|\Lambda\|^2} \right]}{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T \left( \overset{\leftarrow \tau}{\tilde{V}} \right)^1}$ |
| 1 | KL | $W^\tau \leftarrow \widetilde{W^\tau} \cdot \dfrac{\Sigma_\phi \left[ \left( \overset{\uparrow \phi}{\tilde{V}} \right)^{-1} \cdot \dfrac{\overset{\uparrow \phi}{\vphantom{V}}}{\|\Lambda\|^2} \right]^{\rightarrow \tau T} H^\phi}{\Sigma_\phi \overset{\rightarrow \tau T}{H^\phi}}$ | $H^\phi \leftarrow H^\phi \cdot \dfrac{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T \left[ \left( \overset{\leftarrow \tau}{\tilde{V}} \right)^{-1} \cdot \dfrac{\overset{\leftarrow \tau}{\vphantom{V}}}{\|\Lambda\|^2} \right]}{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T}$ |
| 0 | IS | $W^\tau \leftarrow \widetilde{W^\tau} \cdot \dfrac{\Sigma_\phi \left[ \left( \overset{\uparrow \phi}{\tilde{V}} \right)^{-2} \cdot \dfrac{\overset{\uparrow \phi}{\vphantom{V}}}{\|\Lambda\|^2} \right]^{\rightarrow \tau T} H^\phi}{\Sigma_\phi \left( \overset{\uparrow \phi}{\tilde{V}} \right)^{-1 \rightarrow \tau T} H^\phi}$ | $H^\phi \leftarrow H^\phi \cdot \dfrac{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T \left[ \left( \overset{\leftarrow \tau}{\tilde{V}} \right)^{-2} \cdot \dfrac{\overset{\leftarrow \tau}{\vphantom{V}}}{\|\Lambda\|^2} \right]}{\Sigma_\tau \overset{\downarrow \phi}{W^\tau}{}^T \left( \overset{\leftarrow \tau}{\tilde{V}} \right)^{-1}}$ |

Table 2
The Algorithm of β-divergence with Different Value of β [21].

| Steps | Description |
|---|---|
| 1. | Initialization on W and H |
| 2. | Normalization on $\widetilde{W_{f,J}^T} = \dfrac{w_{f,j}^T}{\sqrt{\Sigma_{\tau,f}\left(w_{f,j}^T\right)^2}}$ |
| 3. | Compute V=WH |
| 4. | Updating on β-Divergence |
| 5. a. | $W^\tau \leftarrow \widetilde{W^\tau} \cdot \dfrac{\Sigma_\phi \left[ \left( \overset{\uparrow\phi}{\tilde{V}} \right)^{\beta-2} \cdot \dfrac{\overset{\uparrow\phi}{\vphantom{V}}}{\|\Lambda\|^2} \right]^{\rightarrow\tau T} H^\phi}{\Sigma_\phi \left( \overset{\uparrow\phi}{\tilde{V}} \right)^{\beta-1 \rightarrow\tau T} H^\phi}$ |
| b. | $H^\phi \leftarrow H^\phi \cdot \dfrac{\Sigma_\tau \overset{\downarrow\phi}{W^\tau}{}^T \left[ \left( \overset{\leftarrow\tau}{\tilde{V}} \right)^{\beta-2} \cdot \dfrac{\overset{\leftarrow\tau}{\vphantom{V}}}{\|\Lambda\|^2} \right]}{\Sigma_\tau \overset{\downarrow\phi}{W^\tau}{}^T \left( \overset{\leftarrow\tau}{\tilde{V}} \right)^{\beta-1}}$ |
| 6. | Retry the β-Divergence by replacing β=0, 0.1, 0.2, … , 2 |
| 7. | Repeat from steps 2 to 5 until convergence |

## III. Experiments and Analysis

In this paper, we worked the proposed NMF2D based method, which is directly applied into the biomedical field to demonstrate the method of separation of the heart sound and the lung sound. The divergence implemented in NMF2D is the β-Divergence. In addition, the performance on efficiency between the original audio sources and the estimated audio sources in terms of signal-to-distortion ratio (SDR) was evaluated.

## A. Experiment Setup

All simulations and analysis were run on a PC with Intel Core 2 Duo CPU 6750 at 2.66 GHz and 4GB RAM as well as a laptop with Intel Core i5 CPU 5200 at 2.2GHz and 4GB RAM. The software utilized to run this experiment was MATLAB 2010 used as the programming platform. The mixed signal was sampled at 44.1 kHz sample rates. All cases were mixed with equal average power over the duration of the signals, which normalized the time domain in the same decibel for all sources. This was done to obtain better performance during the separation process. The time-frequency (TF) domain was computed using STFT via 2048- point Hamming window FFT, and the frequency domain was then logarithmically scaled. The convolutive components in time and frequency domain were selected to be $\tau = \{0,\ldots,3\}$ and $\phi = \{0,\ldots,31\}$ for every cases.

## B. Result Performances

Firstly, we obtained the original heart and lung sound signal in time-frequency (TF) domain. Secondly, we separated the mixed heart and lung sound via different β value from β-divergence. A spectrogram is a visual representation of the frequency content of a signal, which shows how the quantity of energy in different frequency regions varies as a function of time. From Figure 1, we noted that the shade of color changes upon the intensity of the sound or audio. The aterrimus, also known as deep black color, delegated the highest intensity of sound. It then faded out to become charcoal grey and turned into light grey of the aterrimus. This means that the intensity of sound reduces over a certain period, but it increases again over another period. The deeper the color representation, the higher is the intensity of sound, which means that the amplitude of sound signal is high. In simple term, the sound becomes louder.
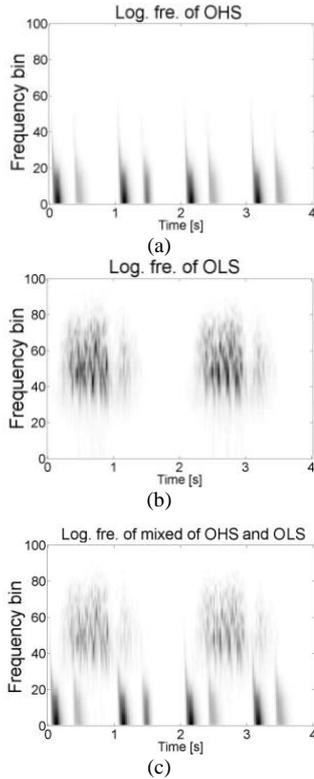


(a)



(b)



(c)

Figure 1: TF representation of (a) log. frequency of OHS, (b) log. frequency of OLS, (c) log. frequency of mixed of OHS and OLS

## C. Relationship of β and SDR

Figure 2. (a) and (b), show the estimated lung sound (ELS) and the estimated heart sound (EHS) respectively after the separation. In Figure 2 (a), the product of spectral basis estimated (W) and temporal basis estimated (H) became EHS. Meanwhile, in Figure 2 (b), the product of spectral basis estimated (W) and temporal basis estimated (H) become ELS. The blackest portion of Figure 2(a) indicates the strongest intensity of sound, which means the amplitude of signal is high. However, the color of Figure 2(b) faded out representing the intensity of sound as well as the amplitude of signal decreases. This is due to the sound intensity in term decibel, which is significantly higher in (a) which is assigned to an apparent audibility.
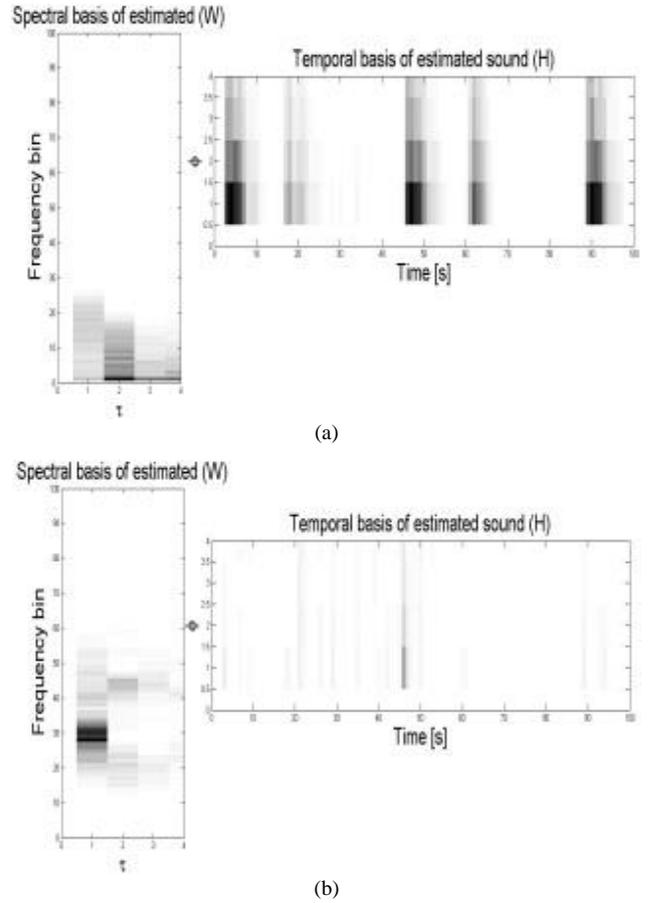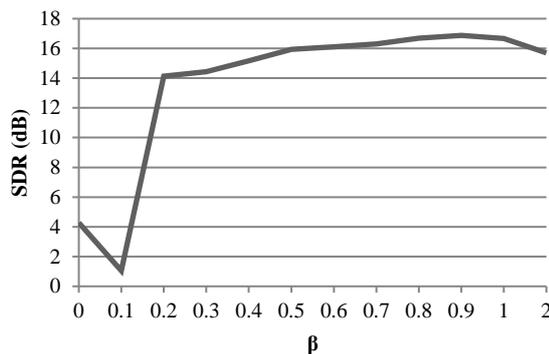


(a)



(b)

Figure 2: Estimated W and H for (a) EHS and (b) ELS after separated via β = 0.9.

Table 3 and Figure 3 show the result of the test conducted to β from 0 to 1 with the step size of 0.1 and bypass to 2. It literally covered the Least Square distance (LS), Kullback-Leibler (KL) and Itakura-Saito (IS) divergence of NMF2D. According to Table 3 and Figure 3, the SDR was augmented in term of dB as the value of β was increased and a distinct increment right at the point of β = 0.1. However, there was a slight shrink of SDR when it reached the peak point of β = 0.9. The lowest SDR was β = 0, which was the IS divergence. In other words, SDR of β = 0.9 was higher than IS. Despite of only two β values just passed 1dB, the overall result is considered as satisfactory, in which all have positive values and overwhelming majority were above 10dB Therefore, this can be concluded that the β = 0.9 is the optimal value for audio separation among β-divergence.

Table 3
Performance Measurement of Average of Output A* and B* with Respect to β value (*A: The comparison of OHS and EHS; *B: The comparison of OLS and ELS).

| Divergence | Average SDR of output A and output B (dB) |
|---|---|
| β = 0 (IS) | 4.27115 |
| β = 0.1 | 1.0513 |
| β = 0.2 | 14.1242 |
| β = 0.3 | 14.42455 |
| β = 0.4 | 15.16105 |
| β = 0.5 | 15.9412 |
| β = 0.6 | 16.1097 |
| β = 0.7 | 16.3045 |
| β = 0.8 | 16.68855 |
| β = 0.9 | 16.8751 |
| β = 1 (KL) | 16.66285 |
| β = 2 (LS) | 15.6942 |



Figure 3: Performance of the proposed β-divergence algorithm

## IV. CONCLUSION

We proposed a novel method on BSS using NMF2D that was implemented by β-divergence in the biomedical field. Regardless that some of the β values were under satisfactory, β-divergence still managed to be implemented in the heart and lung sound separation with the highest value of SDR at the spot of β = 0.9. The β = 0.9 is known as the optimal value underneath the probabilistic framework. Therefore, throughout this experiment, we had achieved our target by decomposing the heart and lung sound using β-divergence NMF2D with affirmative outstanding results.

## ACKNOWLEDGEMENT

## REFERENCES

[1] Hyvarian, J. Karhunen and E. Oja, "Independent component analysis and blind sources separation," *John Wiley and Sons*, 2001

[2] A. Cichocki and S.I.Amari, "Adaptive Blind Signal and Image Processing – Learning Algorithm and Applications," *John Wiley and Sons*, 2003.

[3] 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, pp. 550-563, March 2010.

[4] I. Biciu, N. Nikolaidis and I. Pitas, "Nonnegative matrix factorization in polynomial feature space", *IEEE Trans. Neural Network*, vol. 19, pp. 1090-1100, 2007.

[5] A.M. Darsono, Shakir Saat, N.M. Z. Hashim, A.A.M ISA "Unsupervised Single Channel Source Separation with Nonnegative Matrix Factorization," *ICIT 2015 The 7th International Conference on Information Technology*, Amman, Jordan, 2015.

[6] B. Gao, W.L. Woo and S.S. Dlay, "Single Channel Source Separation Using EMD-Subband Variable Regularised Sparse Features," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, pp. 961–976, 2011.

[7] W.L. Woo and S.S. Dlay, "Neural network approach to blind signal separation of mono-nonlinearly mixed sources," *IEEE Trans. Circuits and System I*, vol. 52, no. 6, pp. 1236-1247, 2005.

[8] J. Zhang, W.L. Woo and S.S. Dlay, "Blind Source Separation of Post-Nonlinear Convolutive Mixture," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2311-2330, 2007.

[9] A.M. Darsono, Bin Gao, W.L. Woo, S.S. Dlay, "Nonlinear single channel source separation", *International Symposium on communications systems, networks and digital signal processing (CSNDSP 2010)*, 2010, pp: 507-511

[10] P. Li, Y. Guan, B. Xu and W. Liu, "Monaural speech separation based on computational auditory scene analysis and objective quality assessment of speech," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, no. 6, pp. 2014–2023, Nov. 2006.

[11] G. Hu and D.L. Wang, "Monaural speech segregation based on pitch tracking and amplitude modulation," *IEEE Trans. Neural Networks*, vol. 15, no. 5, pp. 1135–1150, Sep. 2004.

[12] M.S. Pedersen, D.L. Wang, J. Larsen and U. Kjems, "Two-Microphone Separation of Speech Mixtures," *IEEE Trans. on Neural Networks*, vol. 19, no. 3, pp. 475–492, Mar. 2008.

[13] H. Pasterkamp, R. Fenton, A. Tal and V. Chernick, "Interference of cardiovascular sounds with phonopneumography in children," Am. Rev. Respir. Dis., vol. 131, no. 1, pp. 61–64, Jan. 1985.

[14] 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.

[15] P. J. Arnott, G. W. Pfeiffer and M. E. Tavel, "Spectral analysis of heart sounds: Relationships between some physical characteristics and frequency spectra of first and second heart sounds in normals and hypertensives," J Biomed. Eng., vol. 6, no. 2, pp. 121-128, Apr. 1984.

[16] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, pp. 791, 1999.

[17] D. D. Lee and H. S. Seung, "Unsupervised learning by convex and conic coding," *Proceedings of the Conference on Neural Information Processing Systems 9*, pp.515–521, 1997.

[18] R. Kompass, "A generalized divergence measure for non-negative matrix factorization". In *Neuroinformatics workshop*, Torun, Poland, Sept. 2005.

[19] A.M. Darsono, NZ Haron, Shakir Saat, MM Ibrahim, NA Manap "Blind Audio Source Separation with Sparse Nonnegative Matrix Factorization," *Research Journal of Applied Sciences, Engineering and Technology*, vol.7, issue 23, pp. 5015-5020, June 2014.

[20] M. Morup and M. N. Schmidt, "Sparse nonnegative matrix factor 2-D deconvolution," Technical Report, Technical University of Denmark, Copenhagen, Denmark, 2006.

[21] 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," *IEEE Conference on Wireless Sensors (ICWiSe2013)*, Dec. 2013