MULTI-SPEAKER VOICE ACTIVITY DETECTION BY AN IMPROVED MULTIPLICATIVE NON-NEGATIVE INDEPENDENT COMPONENT ANALYSIS WITH SPARSENESS CONSTRAINTS

L. Khadidja Hamaidi[1,2], *Michael Muma*[1], *Abdelhak M. Zoubir*[1,2]

[1]Signal Processing Group
Technische Universität Darmstadt
Merckstraße 25, 64283 Darmstadt, Germany
Email: {lhamaidi, muma, zoubir}@spg.tu-darmstadt.de

[2]Graduate School CE
Technische Universität Darmstadt
Dolivostraße. 15, 64293 Darmstadt Germany
Email: {hamaidi, zoubir}@gsc.tu-darmstadt.de

ABSTRACT

We propose an improved version of the non-negative independent component analysis algorithm that uses a multiplicative update rule (M-NICA). We examine a challenging NICA application in a noise-embedded multi-speaker voice activity detection (VAD) setup. We present a novel approach that includes sparsity constraints to solve the energy separation problem with independent source signals. A sparse feature extraction step is performed to project the non-negative signals onto a dimension-reduced subspace and identify sparse principal components. Then, we maximize the signal decorrelation by employing a median measure of central tendency in the computation of the covariance matrix that contributes in robustness against outliers. Moreover, our approach supplies a straightforward multi-speaker VAD, for which no empirical thresholding or other ad-hoc decision rule is required. Instead, an active voice frame simply corresponds to a non-zero value of the separated energy signal. Numerical experiments using real data validate the superior performance of the proposed technique.

Index Terms— Multiplicative Non-negative independent component analysis (M-NICA), ℓ_1-norm regularization, Multi-speaker voice activity detection (VAD), Wireless acoustic sensor networks

1. INTRODUCTION

Independent component analysis (ICA) is a well established technique that is capable of separating independent sources that are linearly mixed, e.g., in a wireless sensor network (WSN). Given a multivariate observed data, ICA characterizes the model for which some unmixed latent variables and a mixing system are unknown and subject to estimation. The latent variables are the source signals that determine the independent components of the observed data. A vast amount of research explores ICA, see, e.g. [1, 2], particularly its powerful performance compared to other methods such as principal component analysis (PCA) [3, 4]. Many applications require ICA for their data analysis, including image denoising [5] and face recognition [6] in digital images or speech enhancement [7] and voice activity detection (VAD) [3, 4].

In this work, we consider the application of multi-speaker VAD for a wireless acoustic sensor networks (WASN). This involves dealing with mixtures of simultaneously recorded speech signals at spatially distributed microphones. ICA is used to extract unmixed (non-negative) energy signals, based on which speaker specific VAD is performed. Non-negative ICA algorithms (NICA) are presented in [8, 9]. Similar representations that are tailored to the statistics of non-negative data exist in the literature, such as non-negative matrix factorization (NMF) in [10, 11]. In noisy environments, the lack of robustness is very problematic for NICA. The majority of the NICA methods assume a noise-free model in order to keep the problem tractable. However, this assumption is unrealistic in real world scenarios. Consequently, we assume an embedded-noise NICA model. A variety of non-negative data representation problems take advantage of the ℓ_1-norm regularization in order to obtain a sparse representation of the solution [12]. This is known as non-negative sparse coding and is proposed in [13–18].

Contributions: We provide an improvement of the M-NICA algorithm in [3, 4] by integrating sparsity constraints in the embedded-noise NICA model. Sparsity is introduced by using a sparse singular value decomposition (SSVD) as an initial step for the multiplicative update. We initialize our algorithm by projecting the non-negative data onto the right rotation matrix subspace on which we impose sparsity. The sparse right singular vectors are a low-dimensional representation of the independent components. We show that using these sparse features as input for the multiplicative update rule maximizes the decorrelation between the components, and yet ensures non-negativity and makes unnecessary the

The work of L. Khadidja Hamaidi is supported by the 'Excellence Initiative' of the German Federal and State Governments and the Graduate School of Computational Engineering at Technische Universität Darmstadt. The work of Michael Muma is supported by the project HANDiCAMS which acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944.

projection step in M-NICA. Our approach supplies a straightforward multi-speaker VAD, for which no empirical thresholding or other ad-hoc decision rule is required, since voice activity directly correponds to a non-zero energy signal value.

2. PROBLEM FORMULATION

We examine an audio scenario that consists of a known number N of mutually independent speech sources $[\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_N]^T$ impinging on J sensors. The sources are considered to be uniquely labeled, which can be done, e.g. by applying the algorithm of [19] or [20]. These sensors form an ad-hoc WASN. Fig. 1 depicts our exemplary use-case. The N speakers generate signals $\tilde{s}_n[t]$, $n = 1, \ldots N$, where t denotes the sample time index. We assume statistical second order stationarity for blocks of length L and define the instantaneous power of a signal $\tilde{s}_n[t]$, $n = 1, \ldots N$ at each block as

$$s_n[k] = \frac{1}{L} \sum_{l=0}^{L-1} \tilde{s}_n[kL + l]^2, \tag{1}$$

where $k = 1, \ldots, K$ is the frame index. The $s_n[k]$ are stacked in an N dimensional vector $\mathbf{s}[k]$. The instantaneous noisy signal power in the j-th microphone is

$$y_j[k] = \frac{1}{L} \sum_{l=0}^{L-1} \tilde{y}_j[kL + l]^2, \quad j \in \{1, \ldots, J\}, \tag{2}$$

where $\tilde{y}_j$ denotes the noisy signal observed at the j-th microphone. The non-negative $y_j[k]$ are stacked in a J-dimensional vector $\mathbf{y}[k]$. We can then write

$$\mathbf{y}[k] \approx \mathbf{A}\mathbf{s}[k] + \boldsymbol{\omega}[k], \quad \forall k \in \mathbb{N}, \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{J \times N}$ is an unknown matrix of damping coefficients $[\mathbf{A}]_{jn}$ that describe the power attenuation between speaker n and microphone j. The J-dimensional vector $\boldsymbol{\omega}[k]$ represents the additive white noise at frame k that is constructed following Eqs. (1)-(2). Based on the observed set of instantaneous linear mixtures $\mathbf{y}[k]$ of mutually independent non-negative energy signals, our goal is to estimate a sparse representation of the unknown vectors $\mathbf{s}[k]$ and consequently to determine the voice activity of the speakers.

3. MEDIAN-BASED M-NICA WITH SPARSITY CONSTRAINTS (SMM-NICA)

In the following subsections, we explain how sparse coding is integrated into the M-NICA algorithm for the sake of better signal estimation and an enhanced VAD procedure in a noisy environment.

3.1. Sparse singular features

We define $\mathbf{Y} \in \mathbb{R}_+^{J \times K}$ as the matrix containing all vectors $\mathbf{y}[k]$ with $k = 1 \ldots K$. The standard M-NICA algorithm preprocesses the data using a singular value decomposition step (SVD). The latter can be seen as a PCA technique in itself that

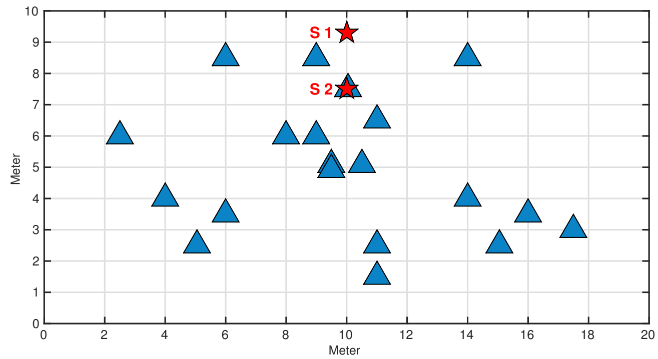

Fig. 1. Acoustic scenario containing $N = 2$ speech sources and $J = 20$ microphones in a 20×10 meter room with a reverberation time of T60 = 0.3 seconds. The sampling frequency of the microphone signals is fs = 16 kHz.

extracts the first principal components [21]. Transforming $\mathbf{Y}$ using the SVD projects the signal onto the sub-spaces

$$\text{SVD}(\mathbf{Y}) = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top, \tag{4}$$

where the left orthogonal matrix $\mathbf{U} \in \mathbb{R}^{J \times J}$ represents the principal directions, $\boldsymbol{\Sigma} \in \mathbb{R}^{J \times K}$ is the scaling matrix of singular values, and $\mathbf{V}^\top \in \mathbb{R}^{K \times K}$ is the right rotation orthogonal matrix of singular vectors. The matrix product $\boldsymbol{\Sigma}\boldsymbol{V}^\top \in \mathbb{R}^{J \times K}$ embodies the principal components. The linear transformation $\boldsymbol{\Sigma}$ controls the speech energies by a scale factor that is the same in all directions. Omitting this factor does not deteriorate the signal shape. In addition, based on [22], the criterion of orthogonality for the vectors in $\mathbf{U}$ forces the right vectors in $\mathbf{V}$ to be a mixture of sources. We suggest using the matrix of right singular vectors as a feature for the subsequent energy separation step. We employ a sparse decomposition (SSVD) in lieu of an SVD to extract sparse features. We impose sparsity on the right rotation matrix $\mathbf{V}$. We seek a lower rank representation of the matrix $\mathbf{Y}$ with the requirement that the right singular vectors $\mathbf{v}_n, n = 1, \ldots, N$ are sparse.

3.2. Sparse right singular vectors subspace projection

First, a rank-one SVD layer $(\sigma, \mathbf{u}, \mathbf{v})^\top$ is the best approximation of $\mathbf{Y}$ if it solves

$$\underset{\sigma, \mathbf{u}, \mathbf{v}}{\text{argmin}} \|\mathbf{Y} - \sigma\mathbf{u}\mathbf{v}^\top\|^2, \tag{5}$$

where $\mathbf{u}$ is a unit J-vector and $\mathbf{v}$ is a unit K-vector. In order to obtain a sparse vector $\mathbf{v}$, we add sparsity-inducing penalties on $\mathbf{v}$ in the optimization objective in Eq. (5). We thus can expand Eq. (5) with an ℓ_1 regularization penalty term to formulate a sparsity promoting problem. Specifically, we minimize with respect to the triplet $(\sigma, \mathbf{u}, \mathbf{v})$ the following penalized sum-of-squares criterion

$$\|\mathbf{Y} - \sigma\mathbf{u}\mathbf{v}^\top\|^2 + \lambda_{\mathbf{v}}\Phi(\sigma\mathbf{v}), \tag{6}$$

with $\Phi(\sigma\mathbf{v})$ being the ℓ_1 regularization function

$$\Phi(\sigma\mathbf{v}) = \sum_{k=1}^{K} |\sigma v_k| \tag{7}$$

and $\lambda_\mathbf{v}$ being the non-negative penalty parameter. The selection of $\lambda_\mathbf{v}$ corresponds to selecting the degree of sparsity of $\mathbf{v}$. The latter is the number of zero components in $\mathbf{v}$ or, based on [17], the number of k elements that satisfy

$$g(\lambda_\mathbf{v}) = \#\left(\left\{k \in \{1, \cdots, K\} : [\mathbf{Y}^\top \mathbf{u}]_k \sigma > \frac{\lambda_\mathbf{v}}{2}\right\}\right) \tag{8}$$

for a fixed $\mathbf{u}$. Where $g(\lambda_\mathbf{v})$ is the degree of sparsity function and $\#(\cdot)$ represents the cardinality symbol. Moreover, [17] and [23] suggest the use of the BIC, from [24], to estimate the optimal number of non-zero coefficients

$$\mathrm{BIC}(\lambda_\mathbf{v}) = \frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2}{jk\hat{\varrho}^2} + \frac{\log(jk)}{jk} g(\lambda_\mathbf{v}) \tag{9}$$

with $\hat{\varrho}^2$ denoting the ordinary least squares of the error variance in Eq. (6). In order to reach a sparse $\mathbf{v}$, the minimization of Eq. (6) with respect to $\sigma\mathbf{v}$ is iterated until convergence. A closed form solution for minimizing $\sigma\mathbf{v}_k$ in Eq. (6) is proposed in [17]. Consequently, it follows that the sparse representation of the vector $\mathbf{v}$ is obtained using

$$\mathbf{v}_k = \frac{1}{\sigma}\left[\mathrm{sgn}\left\{[\mathbf{Y}^\top \mathbf{u}]_k\right\}\left(|[\mathbf{Y}^\top \mathbf{u}]_k| - \frac{\lambda_\mathbf{v}}{2}\right)\right], \tag{10}$$

with $\lambda_\mathbf{v}$ being the minimizer of Eq. (9).

3.3. SMM-NICA algorithm

The following algorithm summarizes the steps of our method. Given $\mathbf{Y}$, we iterate Eq. (5)-(13) to build a matrix of sparse singular vectors $\mathbf{V}^\mathcal{S}$. We use the sparse $\mathbf{V}^\mathcal{S}$ to initialize SMM-NICA as shown in Eq. (14). Then, Eqs. (15)-(19) are reiterated to retrieve an invariant estimate of $\mathbf{S}$. According to [3, 4], the nature of the multiplicative update introduced in Eq. (19) conserves the non-negativity of the matrix $\mathbf{S}$. The function $D\{\cdot\}$ in Eqs. (17)-(18) sets all off-diagonal elements to null. In Eq. (15), we use the median central measure instead of the mean suggested in [3, 4]. A precise descriptive measure depends highly on the shape of the data distribution. The median mid-point outperforms the mean in terms of accuracy for heavy tailed distributions since the mean can strongly be influenced by a small number of outliers [25]. Fig. 2 shows the right-skewed histogram for the energy of source $S1$ considered in Fig 1. Obviously, the mean characterizes the relatively high but infrequent values. For our purpose, the median is a better summary of the typical value.

4. EXPERIMENTAL RESULTS

In this section, we provide simulation results for the multi-speaker energy separation based on our proposed SMM-NICA technique. We consider the scenario depicted in Fig. 1

Algorithm SMM-NICA

Input

$\mathbf{Y} = (\mathbf{y}[1], \cdots, \mathbf{y}[K]) \in \mathbb{R}_+^{J \times K}$ based on Eq. (3)

$\mathbf{V}^\mathcal{S} \triangleq \varnothing$

Initialization

for $n = 1, \ldots, N$ **do**

  Extract rank-one SVD layer $(\sigma, \mathbf{u}, \mathbf{v})^\top$ from $\mathbf{Y}$ that solves Eq. (5)

  Minimize Eq. (6) with respect to $\mathbf{v}$

  Update the sparse right singular vector $\mathbf{v}$ using Eq. (10)

  Construct the sparse matrix $\mathbf{V}^\mathcal{S} \triangleq \mathbf{V}^\mathcal{S} \parallel \mathbf{v}^\top$, with $\parallel$ being the concatenation symbol.

  Set

$$\sigma = \mathbf{u}^\top \mathbf{Y} \mathbf{v} \tag{11}$$

  Compose a sparse lower-rank matrix

$$\mathbf{Y}_{\mathrm{SSVD}} = \sigma \mathbf{u} \mathbf{v}^\top \tag{12}$$

  Matrix subtraction

$$\mathbf{Y} = \mathbf{Y} - \mathbf{Y}_{\mathrm{SSVD}} \tag{13}$$

end for

Define

$$[\mathbf{S}]_{n,k} \leftarrow |[\mathbf{V}^\mathcal{S}]_{n,k}|, \forall n = 1, \ldots, N, \forall k = 1, \ldots, K. \tag{14}$$

repeat

$$\ddot{\mathbf{S}} = \underset{(n)\in N}{\mathrm{median}}\{\mathbf{S}_n\}, \forall n = 1, \ldots, N \tag{15}$$

$$\mathbf{C}_S = (\mathbf{S} - \ddot{\mathbf{S}})(\mathbf{S} - \ddot{\mathbf{S}})^\top \tag{16}$$

$$\mathbf{\Lambda}_1 = D\{\mathbf{C}_S\} \tag{17}$$

$$\mathbf{\Lambda}_2 = D\{(\mathbf{\Lambda}_1^{-1} \mathbf{C}_S)^2\} \tag{18}$$

  Minimize the correlation in $[\mathbf{S}]_{n,k}$

$$[\mathbf{S}]_{n,k} \leftarrow [\mathbf{S}]_{n,k} \left[\frac{\ddot{\mathbf{S}}\mathbf{S}^\top \mathbf{\Lambda}_1^{-1} \mathbf{S} + \mathbf{S}\mathbf{S}^\top \mathbf{\Lambda}_1^{-1} \ddot{\mathbf{S}} + \mathbf{\Lambda}_2 \mathbf{S}}{\ddot{\mathbf{S}}\mathbf{S}^\top \mathbf{\Lambda}_1^{-1} \ddot{\mathbf{S}} + \mathbf{S}\mathbf{S}^\top \mathbf{\Lambda}_1^{-1} \mathbf{S} + \mathbf{\Lambda}_2 \ddot{\mathbf{S}}}\right]_{n,k} \tag{19}$$

until reaching a fixed-point of Eqs. (15)-(19)

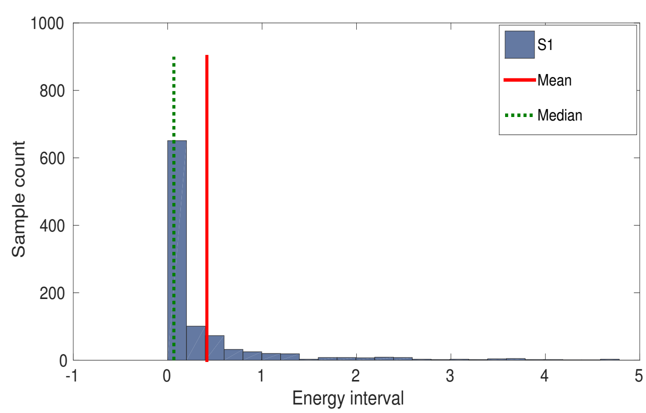

Fig. 2. Right-skewed histogram for the ground truth energies of $S1$ with the mean (red line) and median (dashed green) speech energy central values.

with two speech sources $S1$ and $S2$ affected by a reverberant environment. We compare the performance of the proposed algorithm with the original M-NICA based on diverse performance metrics in different noise variance environments. Table 1 outlines the overall separation results when a mixture of two active speech sources ($S1$, $S2$) is considered. These mixtures are corrupted with noise of two variance levels, i.e. $\sigma_{\omega}^2 = \{0.1, 0.5\}$. In a first experiment, an additive white Gaussian noise (WGN) is considered. The proposed method reduces the RMSE considerably. We further asses our results in terms of the signal correlation ρ. The proposed SMM-NICA is capable of enhancing the signal correlation for $S1$, as shown in Tab. 1. The distance between the estimated energies and the ground truth is evaluated through l_1 and l_2 norms, respectively. The developed SMM-NICA technique displays remarkably small distances outperforming M-NICA in all cases. Moreover, we analyse the normalized RMSE that omits the energy scaling in the performance assessment. Fig. 3-a and 3-b illustrate the ground truth energies for $S1$ and $S2$, respectively. The corresponding unmixed energies produced by M-NICA are depicted in Fig. 3-c and 3-d. It can be seen that some erroneous energy spikes, appear in the M-NICA result. For example, the energies in Fig. 3-c experience a cross-talk in the frame interval around $k = [450, \cdots, 550]$, which obviously belongs to the alternative source $S2$. On the other hand, Fig. 3-e shows a high accuracy in the sense that the cross-talk is attenuated and most of the supposedly zero-energies are indeed attenuated to zero and thus properly unmixed. We further compare the performance of M-NICA to our proposed technique for $S2$ where the SMM-NICA in Fig. 3-f precisely tunes the energies describing the pause regions to zero. As a second case, we also study the performance of the proposed method with an additive babble noise. The results are summarized in the bottom part of Tab. 1. Again, both SMeM-NICA and SMM-NICA outperform M-NICA. Regarding the VAD performance, we exploit the sparse estimated energies in the VAD procedure. Hence, a simple detector that does not require a threshold is implemented. Our VAD step simply assigns the estimated zero-energies to the non-active speech region and vice versa. Higher detection is obtained, as shown in Tab. 1. SMM-NICA achieves a significant 99.4% correct decision for $S2$ in the babble noise case with variance $\sigma_{\omega}^2 = 0.5$.

5. CONCLUSION

We examined multi-speaker VAD as a non-negative energy separation problem for a mixture of speech signals. Our proposed technique improves the M-NICA algorithm by integrating sparse SVD features. The decorrelation of the sparse feature mixture is maximized with a more robust median-based multiplicative update that retains non-negativity. Since the subspace spanned by the rows of the well separated energies does not change after the initialization, our technique does not require a subsequent subspace projection correction step. VAD reduces to determining the non-zero energies.

Variance	Source	Method	NRMSE	RMSE	ρ	l_1-norm	l_2-norm	VAD (%)
Case 1: Additive white Gaussian noise								
$\sigma_{\omega}^2 = 0.1$	$S1$	M-NICA	0.974	97.1	0.78	4.6×10^4	3.1×10^3	63.3
		SMeM-NICA	**0.972**	**0.97**	0.77	403.45	30.79	**92.8**
		SMM-NICA	**0.972**	**0.97**	**0.83**	**401.89**	**30.76**	**92.8**
	$S2$	M-NICA	0.97	1.7×10^3	0.8	6×10^5	5.2×10^4	26.1
		SMeM-NICA	0.97	**0.97**	0.8	**321.46**	**30.8**	**82**
		SMM-NICA	0.97	**0.97**	0.8	321.78	**30.8**	**82**
$\sigma_{\omega}^2 = 0.5$	$S1$	M-NICA	0.97	180.3	0.78	9.22×10^4	5.7×10^3	62.9
		SMeM-NICA	0.97	0.973	0.78	403.32	30.79	**89.6**
		SMM-NICA	0.97	**0.97**	**0.83**	**401.51**	**30.76**	**89.6**
	$S2$	M-NICA	0.97	1.7×10^3	0.8	6.49×10^5	5.34×10^4	26.1
		SMeM-NICA	0.97	0.974	0.8	**321.47**	**30.8**	**81.4**
		SMM-NICA	0.97	**0.97**	0.8	321.79	**30.8**	**81.4**
Case 2: Babble noise								
$\sigma_{\omega}^2 = 0.1$	$S1$	M-NICA	0.974	10	0.78	4.7×10^3	316	63.8
		SMeM-NICA	**0.972**	0.974	**0.83**	**401.3**	**30.74**	92.9
		SMM-NICA	**0.972**	**0.973**	**0.83**	402.6	30.76	**98.1**
	$S2$	M-NICA	0.974	1.6×10^3	0.8	5.5×10^5	5.2×10^4	26.1
		SMeM-NICA	**0.973**	**0.97**	**0.81**	**321.4**	**30.78**	84.7
		SMM-NICA	**0.973**	**0.97**	0.8	321.7	30.79	**99.3**
$\sigma_{\omega}^2 = 0.5$	$S1$	M-NICA	0.973	78.1	0.78	4×10^4	2.5×10^3	62.7
		SMeM-NICA	**0.972**	**0.97**	**0.84**	**401**	**30.74**	88.2
		SMM-NICA	**0.972**	**0.97**	0.83	401.9	30.76	**99.3**
	$S2$	M-NICA	0.97	1.7×10^3	0.8	6×10^5	5.3×10^4	26.1
		SMeM-NICA	0.97	**0.973**	0.8	**321.5**	30.8	85.7
		SMM-NICA	0.97	0.974	0.8	321.7	**30.79**	**99.4**

Table 1. Comparison of the energy separation performance of the original M-NICA algorithm and the proposed approaches: the sparse mean-based M-NICA (SMeM-NICA), and the median-based M-NICA (SMM-NICA) for two sources ($S1$ and $S2$). Case 1: additive white Gaussian noise and Case 2: Babble noise of variance $\sigma_{\omega}^2 = 0.1$ and $\sigma_{\omega}^2 = 0.5$.

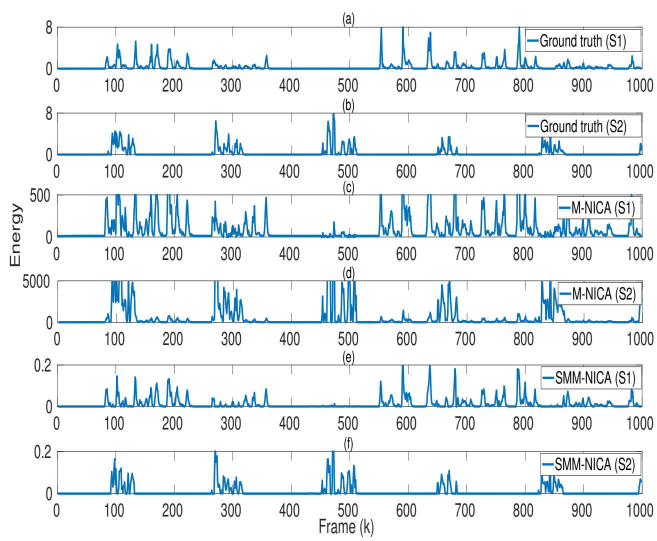

Fig. 3. (a)-(b) Energy ground truth for the speech sources in Fig. 1. The corresponding energy estimates using the M-NICA algorithm (c)-(d), and (e)-(f) the energy estimates using the proposed sparse and median based multiplicative non-negative component analysis (SMM-NICA) approach, under additive white Gaussian noise with variance $\sigma^2 = 0.5$.

6. REFERENCES

[1] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural netw.*, vol. 13, no. 4, pp. 411–430, 2000.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, vol. 46, John Wiley & Sons, 2004.

[3] A. Bertrand and M. Moonen, "Blind separation of non-negative source signals using multiplicative updates and subspace projection," *Signal Process.*, vol. 90, no. 10, pp. 2877–2890, 2010.

[4] A. Bertrand and M. Moonen, "Energy-based multi-speaker voice activity detection with an ad hoc microphone array," in *IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP 2010*, 2010, pp. 85–88.

[5] H. Y. Li, G. L. Ren, and B. J. Xiao, "Image denoising algorithm based on independent component analysis," in *World Congr. Software Eng. WCSE '09.*, May 2009, vol. 4, pp. 465–469.

[6] Y. Huang, M. Li, C. Lin, and L. Tian, "Gabor-based kernel independent component analysis for face recognition," in *Sixth Int. Conf. Intell. Inf. Hiding and Multimedia Signal Process. (IIH-MSP)*, Oct 2010, pp. 376–379.

[7] H. Y. Li, Q. H. Zhao, G. L. Ren, and B. J. Xiao, "Speech enhancement algorithm based on independent component analysis," in *Fifth Int. Conf. Nat. Comput.*, Aug. 2009, vol. 2, pp. 598–602.

[8] M. Plumbley, "Conditions for nonnegative independent component analysis," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 177–180, 2002.

[9] M. D. Plumbley, "Algorithms for nonnegative independent component analysis," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 534–543, 2003.

[10] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[11] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Adv. Neural Inf. Process. syst.*, 2001, pp. 556–562.

[12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B*, pp. 267–288, 1996.

[13] P. O. Hoyer, "Non-negative sparse coding," in *Proc. 12th IEEE Workshop on Neural Netw. Signal Process.* IEEE, 2002, pp. 557–565.

[14] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, no. Nov, pp. 1457–1469, 2004.

[15] J. Eggert and E. Korner, "Sparse coding and NMF," in *Proc. IEEE Int. Joint Conf. Neural Netw.* IEEE, 2004, vol. 4, pp. 2529–2533.

[16] J. Kim and H. Park, "Sparse nonnegative matrix factorization for clustering," 2008.

[17] M. Lee, H. Shen, J. Z. Huang, and J. S. Marron, "Biclustering via sparse singular value decomposition," *Biometrics*, vol. 66, no. 4, pp. 1087–1095, 2010.

[18] J. Xi and A. Li, "Discovering recurrent copy number aberrations in complex patterns via non-negative sparse singular value decomposition," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 4, pp. 656–668, July/Aug. 2015.

[19] S. Chouvardas, M. Muma, K. Hamaidi, S. Theodoridis, and A. M. Zoubir, "Distributed robust labeling of audio sources in heterogeneous wireless sensor networks," in *IEEE 40th Int. Conf. Acoust., Speech, Signal Process. ICASSP'15*, 2015, pp. 5783–5787.

[20] M. H. Bahari, J. Plata-Chaves, A. Bertrand, and M. Moonen, "Distributed labelling of audio sources in wireless acoustic sensor networks using consensus and matching," in *24th Eur. Signal Process. Conf. EUSIPCO'16*, 2016, pp. 2345–2349.

[21] L. L. Scharf, *Stat. Signal Process.*, vol. 98, Addison-Wesley Reading, MA, 1991.

[22] V. D. Vrabie, J. I. Mars, and J. L. Lacoume, "Modified singular value decomposition by means of independent component analysis," *Signal Process.*, vol. 84, no. 3, pp. 645 – 652, 2004.

[23] H. Zou, T. Hastie, R. Tibshirani, et al., "On the degrees of freedom of the lasso," *Ann. Stat.*, vol. 35, no. 5, pp. 2173–2192, 2007.

[24] G. Schwarz et al., "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[25] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, July 2012.