Indonesian Journal of Electrical Engineering and Computer Science
Vol. 2, No. 1, April 2016, pp. 161 ~ 167
DOI: 10.11591/ijeecs.v2.i1.pp161-167

Sparse Modeling with Applications to Speech Processing: A Survey

AN Omara*[1], AA Hefnawy[1], Abdelhalim Zekry[2]
[1]Computers and Systems Department, Electronics Research Institute, Giza, Egypt
[2]Communications and Electronics Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt
*Corresponding author, e-mail: ahmed_omara@eri.sci.eg

Abstract

Nowadays, there has been a growing interest in the study of sparse approximation of signals. Using an over-complete dictionary consisting of prototype signals or atoms, signals are described by sparse linear combinations of these atoms. Applications that use sparse representation are many and include compression, source separation, enhancement, and regularization in inverse problems, feature extraction, and more. This article introduces a literature review of sparse coding applications in the field of speech processing.

Keywords: *Sparse modeling, signal representations, speech processing*

Copyright © 2016 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Natural and artificial sensors are the only tools we have for sensing the world and gathering signals of physical processes. These sensors are usually not aware of the physical process underlying the phenomena they “see,” hence they often sample the signal with a higher rate than the effective dimension of the process. To represent the sampled data efficiently, we have to reduce its dimension to be effective. In other words, the signal has to be linearly represented with a few parameters. Such representations often yield superior signal processing algorithms. Recent theory informs us that, with high probability, a relatively small number of random projections of a signal can contain most of its relevant information.

One of the efficient signal representations is the sparse decomposition. This type of signal decomposition has recently received extensive research interest across several communities including signal processing, information theory, and optimization [1-3]. Also, these representations have found successful applications in data interpretation, source separation, signal de-noising, coding, classification, recognition, and many more [4].

In sparse representation, the signal can be constructed by elementary waveforms chosen in a family called a dictionary [5]. The dictionary elements are called atoms that may be orthogonal or non-orthogonal [6].The over-completed dictionaries whose vectors are larger than bases are needed to build sparse representations of complex signals [7]. But choosing is difficult and requires more complex algorithms.

This article aims at presenting an overview of research efforts on sparse decompositions of speech signals. So, the structure of the article is as follows. In Section II, we review the basic definitions of the sparse coding. And we illustrate the methods of sparse optimization problem. In Section III, we show the aspect of over-complete dictionaries and its approaches. In Section IV, we illustrate the importance of sparse coding in different speech processing applications.

2. Sparse Modeling

It was first introduced in [8], [9] as a method to find sparse linear combinations of basis functions to encode natural images. Sparse representation of signals is a growing field of research which aims at finding a set of prototype signals called atoms $a_i \in \mathbb{R}^n$ which forms a dictionary $\mathcal{D} \in \mathbb{R}^{n \times k}$ that can be used to represent a particular set of given signals $x \in \mathbb{R}^n$ by some sparse linear combination of the atoms in the dictionary. Mathematically, for a given set of signals represented by X, we need to find a suitable dictionary $\mathcal{D}$ such that $x_i \approx \mathcal{D}\varphi_i$ where

Received December 22, 2015; Revised February 27, 2016; Accepted March 10, 2016

$\varphi_i \in \mathbb{R}^k$ is a sparse vector which contains the coefficients for the linear combination and $x_i \in$ X.

2.1. Non-Convex Sparse Optimization Problem

The problem of sparse representation can thus be formulated as a non-convex or (ℓ_0) optimization problem of finding $\mathcal{D}$, φ_i which satisfies

$$\{\widehat{\mathcal{D}}, \widehat{\varphi}_i\} := \underset{\mathcal{D},\varphi_i}{\text{Argmin}} \|x_i - \mathcal{D}\varphi_i\|_2 \text{ Subject to } \|\varphi_i\|_0 < S \quad (1)$$

where $\mathcal{S}$ is some predefined threshold which controls the sparseness of the representation and $\|\varphi_i\|_0$ denotes the ℓ_0 pseudo norm which counts the number of non-zero elements of the vector φ_i. This problem can alternately be formulated as

$$\{\widehat{\mathcal{D}}, \widehat{\varphi}_i\} := \underset{\mathcal{D},\varphi_i}{\text{Argmin}} \|\varphi_i\|_0 \text{ Subject to} \|x_i - \mathcal{D}\varphi_i\|_2 < \varepsilon \quad (2)$$

where ε is the tolerable limit of error in reconstruction. Though the solutions to Eq.1 and Eq.2 need not be the same mathematically, they are similar in essence to what the sparse representation problem aims at achieving. This problem is thus involves a choice of the dictionary and a sparse linear combination of the atoms in the dictionary to represent each desired signal.

2.2. Non-Convex Sparse Optimization Problem

Using the ℓ_0 norm in the sparse approximation problem makes it a NP-Hard with a reduction to NP-complete subset selection problems in combinatorial optimization. A convex relaxation of the problem can instead be obtained by taking the ℓ_1 norm instead of the ℓ_0 norm, where $\|\varphi_i\|_1 = \sum_{j=1}^{k}|\varphi_{j,i}|$. The ℓ_1 norm induces sparsity under certain conditions [10]. The solution of the convex optimization problem will be in the form of

$$\{\widehat{\mathcal{D}}, \widehat{\varphi}_i\} := \underset{\mathcal{D},\varphi_i}{\text{Argmin}} \|x_i - \mathcal{D}\varphi_i\|_2 \text{ Subject to} \|\varphi_i\|_1 < S \quad (3)$$

Or

$$\{\widehat{\mathcal{D}}, \widehat{\varphi}_i\} := \underset{\mathcal{D},\varphi_i}{\text{Argmin}} \|\varphi_i\|_1 \text{ Subject to} \|x_i - \mathcal{D}\varphi_i\|_2 < \varepsilon \quad (4)$$

Efforts devoted to this problem have resulted in the creation of a number of algorithms including basis pursuit (BP) [11], matching pursuit (MP) [12], orthogonal matching pursuit (OMP) [22], subspace pursuit (SP) [13], [14], regression shrinkage and selection (LASSO) [15], focal under-determined system solver (FOCUSS) [16], and gradient pursuit (GP) [17]. Sparse decompositions of a signal, however, rely greatly on the degree of fitting between the data and the dictionary, which leads to the second problem, i.e., the issue of dictionary design.

3. Over-Complete Dictionaries

An over-complete dictionary, one in which the number of atoms is greater than the dimension of the signal, can be obtained by either an analytical or a learning-based approach. The analytical approach generates the dictionary based on a predefined mathematical transform, such as discrete Fourier transform (DFT), discrete cosine transform (DCT), wavelets [18], curvelets [19], contourlets [20], and bandelets [21]. Such dictionaries are relatively easy to obtain and more suitable for generic signals. In learning-based approaches, however, the dictionaries are adapted from a set of training data [8], [9], and [22]-[27]. Although this may involve higher computational complexity, learned dictionaries have the potential to offer improved performance as compared with predefined dictionaries, since the atoms are derived to capture the salient information directly from the signals.

Optimizing the dictionary $\mathcal{D}$ is a challenging problem, and the numerical strategy commonly employed consists in iterative algorithms that start from an initial dictionary and alternate between the following steps [28]:

- Sparse coding: given a fixed dictionary $\mathcal{D}$, the matrix $\emptyset$ of sparse approximation coefficients is calculated using any suitable algorithm for sparse approximation.
- Dictionary update: given a fixed approximation matrix $\emptyset$, the dictionary $\mathcal{D}$ is updated in order to minimize the residual cost function $\|X - \mathcal{D}\emptyset\|_F$.

More specifically, several methods have been proposed to formalize the notion of the suitability of a dictionary for sparse approximation. These include the mutual coherence [29], the cumulative coherence [30], the exact recovery coefficient (ERC) [30], the spark [31], and the restricted isometry constants (RICs) [32], [33]. Except for the mutual coherence and cumulative coherence, none of these measures can be efficiently calculated for an arbitrary given dictionary.

3.1. Mutual Coherence of a Dictionary

The performance of sparse approximation algorithms depends on the mutual coherence of the dictionary μdefined as the maximum absolute inner product between any two different atoms.

$$\mu(\mathcal{D}) \stackrel{\text{def}}{=} \max_{i \neq j} \left| \langle a_i, a_j \rangle \right| \tag{5}$$

The mutual coherence of a dictionary measures the similarity between the dictionary's atoms. For an orthogonal matrix $\mathcal{D}$, $\mu(\mathcal{D}) = 0$. For an over-complete matrix $(K > n)$ we necessarily have $\mu(\mathcal{D}) > 0$. There is an interest in dictionaries with $\mu(\mathcal{D})$ as small as possible for sparse representation purposes. If $\mu(\mathcal{D}) = 1$, it implies the existence of two parallel atoms, and this causes ambiguity in the construction of sparse atom compositions. In [34] it was shown that for a full rank dictionary of size $n \times k$

$$\mu(\mathcal{D}) \geq \sqrt{\frac{k-n}{n(k-1)}} \tag{6}$$

and equality is obtained for a family of dictionaries called Grassmannian frames. For $k \gg n$ the mutual coherence we can expect to have is thus of the order of $1/\sqrt{n}$.

3.2. Cumulative Coherence of a Dictionary

A refinement of the coherence parameter is the cumulative coherence function [35], [36]. It measures how much a collection of m atoms can resemble a fixed, distinct atom. Formally [1]

$$\mu_1(m) \stackrel{\text{def}}{=} \max_{|\Lambda|=m} \max_{w \notin \Lambda} \sum\nolimits_{\lambda \in \Lambda} |\langle a_w, a_\lambda \rangle| \tag{7}$$

We place the convention that $\mu_1(0) = 0$. The subscript on μ_1 serves as a mnemonic that the cumulative coherence is an absolute sum, and it distinguishes the function μ_1 from the number μ. When the cumulative coherence grows slowly, we say informally that the dictionary is incoherent or quasi-incoherent.

3.3. Spark of a Dictionary

The spark of a dictionary $\mathcal{D}$ is the smallest number of columns that form a linearly dependent set [37]. In-spite the similar definition, note that spark is markedly different from the matrix rank, being the greatest number of linearly independent columns. A trivial relation between the spark $\sigma(\mathcal{D})$ and the mutual coherence $\mu(\mathcal{D})$ is [37].

$$\sigma(\mathcal{D}) \geq 1 + \frac{1}{\mu(\mathcal{D})} \tag{8}$$

4. Speech Processing Based on Sparse Modeling

Sparse modeling has ubiquitous applications in speech and audio processing areas, including dimensionality reduction, model regularization, speech compression and reconstruction, acoustic/audio feature selection, acoustic modeling, speech recognition, blind source separation, and many others. This section presents some efforts on speech processing using sparse modeling.

4.1. Speaker Identification

The authors in [38] introduced a novel method for speaker identification or determining an unknown speaker's identity based on a sparse signal model and the use of Compressed Sensing (CS). The use of CS permits the use of less transmission power for the sensor recording the voice. Additionally, this method had been shown to be robust to noise in the recorded speech signal. This is encouraging and warrants further investigation.

4.2. Speech Compression

In [39], the author presented the Molecular Matching Pursuit (MMP) algorithm that is suitable for speech coding. The main goal of MMP is to make a practical decomposition such that at every iteration the algorithm identifies and removes a whole cluster of (orthogonal) atoms. At the cost of a slight sub optimality in the approximation error rate, this offers a number of advantages, most notably it is significantly faster since the inner products update step is made for a large number of atoms at every iteration. Also, the use of a Modified Discrete Cosine Transform (MDCT) for speech coding was investigated in [40]. This approach produces a sparser decomposition than the traditional MDCT-based orthogonal transform and allows better coding efficiency at low bitrates. Contrary to state-of-the-art low bitrate coders, which are based on pure parametric or hybrid representations, the approach is able to provide transparency.

4.3. Blind Source Separation

Underdetermined speech separation is a challenging problem that has been studied extensively in recent years. The author in [56] presented a promising method to the Blind Source Separation (BSS) for speech signals based on sparse representation with adaptive dictionary learning. In another work [58], the author showed that the use of sparse decomposition in a proper signal dictionary provides high-quality blind source separation. Moreover, he proved that the maximum a posteriori framework gives the most general approach, which includes the situation of more sources than sensors. In [41], the author addressed the convolutive BSS issue and suggested a solution using sparse Independent Component Analysis (ICA).

4.4. Speech Enhancement

Recently, sparse representation is widely used for speech processing in noisy environments; however, many problems need to be solved because of the particularity of speech. In [42], a novel view for the enhancement of signals was applied successfully to speech using the K-Singular Value Decomposition Algorithm (K-SVD) [22]. The K-SVD algorithm is designed for training an over-complete dictionary that best suits a set of given signals. Another speech enhancement technique was suggested in [57] when the author proposed an exemplar-based technique for the noisy speech. The technique works by finding a sparse representation of the noisy speech in a dictionary containing both speech and noise exemplars, and uses the activated dictionary atoms to create a time-varying filter to enhance the noisy speech.

A good effort was done in [43]; the author proposed an effective dual-channel noise reduction algorithm based on sparse representations. The algorithm is composed of four steps. Firstly, overlapping patches sampled from two channels together instead of each channel one by one are trained to be a dictionary via K-SVD. Secondly, OMP reconstruction algorithm is applied to obtain the sparse coefficients of patches using the dictionary. Thirdly, the denoising speech can be obtained by the updated coefficients. Lastly, the above three steps are iterated to get clearer speech until some conditions are reached. Experimental results show that this algorithm performs better than that with single channel.

Another speech denoising method based on greedy orthogonal adaptive dictionary learning was proposed in [25]. The algorithm constructs a user-defined complete dictionary, whose atoms clearly encode local properties of the signal. The performance of the algorithm was compared to that of the Principal Component Analysis (PCA) method, and it was found to give good signal approximations, even as the number of atoms in the reconstructions decreases considerably; it was also observed that the algorithm has good tolerance to noise, comparable to that afforded by PCA.

The enhancement of speech degraded by non-stationary interferers was addressed in [55]. The author presented a monaural speech enhancement method based on sparse coding of noisy speech signals in a composite dictionary, consisting of the concatenation of a speech and interferer dictionary, both being possibly over-complete. The speech dictionary is learned off-line on a training corpus, while an environment specific interferer dictionary is learned on-line during speech pauses.

4.5. Speech Recognition

Most of automatic speech recognition (ASR) technologies are based on hidden Markov models (HMMs), which model a time-varying speech signal using a sequence of states, each of which is associated with a distribution of acoustic features. While HMMs reach a relatively high performance in good conditions, they have problems in modeling wide variances in natural speech signals, such as speech in natural environments which is often interfered by environmental noises.

Recently, some studies [44-45], and [51-54] have aimed at ASR using sparse representations of speech. In them, a time-frequency representation of speech is as a weighted linear combination of speech atoms. Benefits of the existing systems range from improved recognition accuracy to an easy incorporation of robustness to additive noises. Some of these systems construct the dictionary of atoms to be used in the sparse representation from exemplars of speech, which are realizations of speech in the training data, spanning multiple time frames [54].

When the weights of the sparse representation are used directly in the recognition, a fundamental problem is the association of higher-level information with the atoms in the dictionary to enable the recognition. In [45], the author trained a neural network to map the weights of the atoms directly to phoneme classes. Whereas in [53], the author associated each atom with one phonetic class, and recognition was done by finding the phoneme class with the highest sum of weights. Also in [52], the author used a dictionary consisting of both acoustic information and higher-level phonetic information. But in [51], the author used the index of the speech atom with the highest weight as an additional feature for their Dynamic Bayesian Network recognizer.

Beside the foregoing efforts, there are more researches on speech recognition based on sparse representations [46-50]. In [46], the author enhanced the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm for improving the speech recognition rates. In [50], the author used the sparse representation to estimate the missing (unreliable) coefficients of the speech signal. In [47], the author had evaluated the sparsity assumptions incorporated in sparse component analysis in the framework of Degenerate Un-mixing Estimation Technique (DUET) for speech recognition in a multi-speaker environment. In [48], the author proposed a state-based labeling for acoustic patterns of speech and a method for using this labeling in noise robust automatic speech recognition. In [49], a framework for an exemplar-based, de-convolutive speech recognition system was presented.

5. Conclusion

This study sheds light on the applications of the sparse modeling in the field of speech processing. Although the sparse modeling is just a signal decomposition technique, this survey showed the importance of this strategy in the speech source separation, speech compression, speaker identification, speech recognition and noise reduction. Not only the sparseness of the representation plays a role in these wide applications, but also the choice of the dictionary plays an imoportant role in this variation of the applications.

References

[1] JA Tropp. Topics in sparse approximation. Ph.D. Dissertation, Computational and Applied Mathematics, The University of Texas at Austin. 2004.

[2] P Földiák. Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*. 1990; 64: 165-170.
[3] Lewicki MS, Sejnowski TJ. "Learning overcomplete representations". *Neural Comput*. 2000; 12(2): 337-65.
[4] W Dai, T Xu, and W Wang. "*Simultaneous Codeword Optimization (SimCO) for Dictionary Learning*". in Proc. 49th Annual Allerton Conference on Communication, Control, and Computing (ALLERTON 2011), Monticello, Illinois, USA. 2011.
[5] S Mallat. A Wavelet Tour of Signal Processing. 2nd ed. New York: Academic. 1998.
[6] David L Donoho and Michael Elad. *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization*. Proc. Natl. Acad. Sci. USA 100. 2003.
[7] JJ Fuchs. "On sparse representations in arbitrary redundant bases". *IEEE Trans. Inf. Theory*. 2004; 50: 1341–1344.
[8] Bruno A Olshausen, David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996; 381(6583): 607-9.
[9] Olshausen BA, Field DJ. "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?". *Vision Research*. 1997; 37: 3311-3325.
[10] David L Donoho. "For most large underdetermined systems of linear equations, the minimal 1-norm near-solution approximates the sparsest near-solution," Dept. of Statistics, Stanford Univ. 2004, Tech. Rep. 2004-11.
[11] Huggins, PS, Zucker, SW. "Greedy Basis Pursuit". *IEEE Transactions on Signal Processing*. 2007; 55(7): 3760-3772.
[12] Mallat SG, Zhifeng Zhang. "Matching pursuits with time-frequency dictionaries". *IEEE Transactions on Signal Processing*. 1993; 41(12): 3397-3415.
[13] Wei Dai; Milenkovic O. "Subspace Pursuit for Compressive Sensing Signal Reconstruction". *IEEE Transactions on Information Theory*. 2009; 55(5): 2230-2249.
[14] D Needell and JA Tropp Co Sa MP. Iterative signal recovery from incomplete and inaccurate samples Applied and Computational Harmonic Analysis. 2009
[15] Tibshirani R. Regression shrinkage and selection via the Lasso. *J. Roy. Statist.Soc. Ser. B*. 1996; 58(1): 267-288.
[16] Gorodnitsky IF, Rao BD. "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm". *IEEE Transactions on Signal Processing*. 1997; 45(3): 600-616.
[17] Blumensath T, Davies ME. "Gradient Pursuits". *IEEE Transactions on Signal Processing*. 2008; 56(6): 2370-2382.
[18] Selesnick IW, Baraniuk RG, Kingsbury NC. "The dual-tree complex wavelet transform". *IEEE Signal Processing Magazine*. 2005; 22(6): 123- 151.
[19] EJ Candes, DL Donoho et al. Curvelets: *A surprisingly effective nonadaptive representation for objects with edges*. Department of Statistics, Stanford University. 1999.
[20] MN Do and M Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*. 2005; 14(12): 2091–2106.
[21] Le Pennec E, Mallat S. "Sparse geometric image representations with bandelets". *IEEE Transactions on Image Processing*. 2005; 14(4): 423-438.
[22] M Aharon, M Elad and A Bruckstein. "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation". *IEEE Transactions on Signal Processing*. 2006; 54(11): 4311-4322.
[23] Wipf DP, Rao BD. "Sparse Bayesian learning for basis selection". *IEEE Transactions on Signal Processing*. 2004; 52(8): 2153- 2164.
[24] MD Plumbley. *Dictionary Learning for L1-Exact Sparse Coding*. In ME Davies, CJ James, SA Abdallah and MD Plumbley (eds.), Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation, ICA 2007, London, UK, 9-12 Sept 2007, LNCS 4666, pp 406-413, Springer, Berlin.
[25] MG Jafari and MD Plumbley. *Speech denoising based on a greedy adaptive dictionary algorithm*. In Proc. 17th European Signal Processing Conference (EUSIPCO 2009) Glasgow, Scotland. 2009: 1423-1426.
[26] Rémi Gribonval, Karin Schnass: Dictionary identification: sparse matrix-factorization via l1-minimization. *IEEE Transactions on Information Theory*. 2010; 56(7): 3523-3539.
[27] Q Geng, H Wang, and J Wright. "*On the Local Correctness of L^1 Minimization for Dictionary Learning*". Presented at CoRR. 2011.
[28] Barchiesi D, Plumbley MD. "Learning incoherent dictionaries for sparse approximation using iterative projections and rotations". *IEEE Transactions on Signal Processing*. 2013; (99): 1, 0.
[29] Donoho DL, Huo X. "Uncertainty principles and ideal atomic decomposition". *IEEE Transactions on Information Theory*. 2001; 47(7): 2845-2862.
[30] Tropp JA. "Just relax: convex programming methods for identifying sparse signals in noise". *IEEE Transactions on Information Theory*. 2006; 52(3): 1030-1051.

[31] Donoho DL, Elad M, Temlyakov VN. "Stable recovery of sparse overcomplete representations in the presence of noise". *IEEE Transactions on Information Theory*. 2006; 52(1): 6- 18.
[32] Candes EJ and Tao T. The Dantzig selector: statistical estimation when p is much larger than n. *Annals of Statistics*. 2007; 35(6): 2313-2351.
[33] EJ Candes, JK Romberg and T Tao. "Stable Signal Recovery from Incomplete and Inaccurate Measurements". *Communications on Pure and Applied Mathematics*. 2006; LIX: 1207–1223.
[34] T Strohmer and RW Heath Jr. "*Grassmannian Frames with Applications to Coding and Communications*". in Applied and Computational Harmonic Analysis. 2003; 14(3): 257-275.
[35] Tropp JA. "Greed is good: algorithmic results for sparse approximation". *IEEE Transactions on Information Theory*. 2004; 50(10): 2231- 2242.
[36] D Donoho and M Elad. "*Maximal Sparsity Representation via l1 Minimization*". The Proc. Nat. Aca. Sci. 2003; 100: 2197-2202.
[37] M Aharon, M Elad, and AM Bruckstein. "On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them". *J. Linear Algebra Appl*. 2006; 416: 48-67.
[38] A Griffin, E Karamichali, and A Mouchtaris. "*Speaker Identification using Sparsely Excited Speech Signals and Compressed Sensing*". in Proc. 18th European Signal Processing Conference (EUSIPCO '10), Aalborg, Denmark. 2010.
[39] Daudet L. "Sparse and structured decompositions of signals with the molecular matching pursuit". *IEEE Transactions on Audio, Speech, and Language Processing*. 2006; 14(5): 1808-1816.
[40] Ravelli E, Richard G, Daudet L. "Union of MDCT Bases for Audio Coding". *IEEE Transactions on Audio, Speech, and Language Processing*. 2008; 16(8): 1361-1372.
[41] Jafari MG, Abdallah SA, Plumbley MD, Davies ME. *Sparse Coding for Convolutive Blind Audio Source Separation*. In ICA. 2006: 132-139.
[42] Zhao Nan, Xu Xin and Yang Yi. "Sparse Representations for Speech Enhancement". *Chinese Journal of Electronics*. 2011; 19(2).
[43] Jiancheng Zou, Jingsai Liang, Xin Yang. "*Dual-channel noise reduction via sprase representations*". 2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP). 2012: 221-225.
[44] B King and LE Atlas. "*Single-channel source separation using simplified-training complex matrix factorization*". in Proc. ICASSP. 2010: 4206-4209.
[45] GSVS Sivaram, SK Nemala, M Elhilali, TD Tran and H Hermansky. "*Sparse coding for speech recognition*". in Proc. ICASSP. 2010: 4346-4349.
[46] QF Tan, PG Georgiou, and SS Narayanan. "Enhanced sparse imputation techniques for a robust speech recognition front-end". *IEEE Trans. Audio, Speech, Language Process*. 2011; 19(8): 2418-2429.
[47] Afsaneh Asaei, Hervé Bourlard, and Philip N Garner. "*Sparse component analysis for speech recognition in multi-speaker environment*". INTERSPEECH. 2010: 1704-1707.
[48] T Virtanen, JF Gemmeke, and A Hurmalainen. "*State-based labelling for a sparse representation of speech and its application to robust speech recognition*". in Proc. INTERSPEECH. 2010: 893-896.
[49] Antti Hurmalainen, Jort F Gemmeke, Tuomas Virtanen: Non-negative matrix deconvolution in noise robust speech recognition. *ICASSP*. 2011: 4588-4591.
[50] J Gemmeke, B Cranen. "*Using sparse representations for missing data imputation in noise robust speech recognition*". Proc. of EUSIPCO. 2008
[51] Schuller B, Weninger F, Wöllmer M, Sun Y, Rigoll G. "Non-negative matrix factorization as noise-robust feature extractor for speech recognition". *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. 2010: 4562-4565.
[52] Maarten Van Segbroeck, Hugo Van Hamme. Unsupervised learning of time-frequency patches as a noise-robust representation of speech. *Speech Communication*. 2009; 51(11): 1124-1138.
[53] Sainath TN, Carmi A, Kanevsky D, Ramabhadran B. "Bayesian compressive sensing for phonetic classification". *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. 2010: 4370-4373.
[54] Gemmeke JF, Virtanen T. "Noise robust exemplar-based connected digit recognition". *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. 2010;: 4546-4549.
[55] Sigg CD, Dikk T, Buhmann JM. "Speech enhancement with sparse coding in learned dictionaries". in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. 2010: 4758-4761.
[56] Tao Xu, Wenwu Wang. "Methods for learning adaptive dictionary in underdetermined speech separation". 2011 *IEEE International Workshop on Machine Learning for Signal Processing (MLSP).* 2011: 1-6.
[57] Gemmeke JF, Virtanen T, Hurmalainen A. "Exemplar-Based Sparse Representations for Noise Robust Automatic Speech Recognition". in *IEEE Transactions on Audio, Speech, and Language Processing*. 2011; 19(7): 2067-2080.
[58] M Zibulevsky, BA Pearlmutter, P Bofill and P Kisilev. "*Blind Source Separation by Sparse Decomposition*". Chapter in the book: SJ Roberts, and RM Everson eds. Independent Component Analysis: Principles and Practice, Cambridge. 2001.