Making Sense of Randomness: Fast Signal Recovery from Compressive Samples

V. Abrol, P. Sharma, Anil K. Sao
School of Computing and Electrical Engineering, Indian Institute of Technology, Mandi
Email: {vinayak_abrol, pulkit_s}@students.iitmandi.ac.in, anil@iitmandi.ac.in.

***Abstract*—In compressed sensing (CS) framework, a signal is sampled below Nyquist rate, and the acquired samples are generally random in nature. Thus, for efficient estimation of the actual signal, the sensing matrix must preserve the relative distances among the underlying sparse vectors. Provided this condition is fulfilled, we show that CS samples will also preserve the envelope of the actual signal. Exploiting this envelope preserving property of CS samples, we propose a new fast method which is able to extract prototype signals from compressive samples for efficient sparse representation and recovery of signals. These prototype signals are orthogonal intrinsic mode functions (IMFs) extracted from CS samples using empirical mode decomposition (EMD), which is one of the popular methods to capture the envelope of a signal. The extracted IMFs are used to seed the dictionary without even comprehending the original signal or the sensing matrix. Moreover, one can update the dictionary on-line as new CS samples are available. In particularly, to recover first L signals ($\in \mathbb{R}^n$) at the decoder, one can seed the dictionary in just $\mathcal{O}(nL\log n)$ operations, that is far less as compared to existing approaches. The efficiency of the proposed approach is demonstrated experimentally for recovery of speech signals.**

***Index Terms*—Compressed sensing, dictionary learning, empirical mode decomposition, speech processing.**

I. Introduction

Compressed sensing (CS) or sparse signal representations have recently drawn much interest in the field of speech and image processing [1], [2], [3], [4]. In particular, CS takes advantage of the sparsity of a signal $\mathbf{x} \in \mathbb{R}^n$ in an overcomplete dictionary $\mathbf{\Psi} \in \mathbb{R}^{n\times d}$ (d=n for complete dictionary). One can efficiently recover the signal from compressed measurements $\mathbf{y} \in \mathbb{R}^m$, sampled using a measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ with $m \ll n$, via recovery of its sparse representation $\mathbf{a} \in \mathbb{R}^d$ [5], [6], [7].

CS measurements are robust to degradations such as random perturbations or noise and does not require much memory for storage or transmission [5]. The estimation of sparse vector (or equivalently the original signal) using compressed samples is very much influenced by the choice of dictionary [5]. It has been shown that a sparse representation, estimated using a learned dictionary (e.g., KSVD) as compared to an analytic dictionary (e.g., DCT), results in better recovery of the signal [6]. The dictionary learning (DL) problem aims to find a dictionary $\mathbf{\Psi}$ such that the error, $\|\mathbf{x}_i - \mathbf{\Psi}\mathbf{a}_i\|_2^2 \ \forall_i$ is minimized and $\mathbf{a}_i$ is sparsest [8]. Typically this is achieved by alternating minimization over $\mathbf{a}_i$'s and $\mathbf{\Psi}$, i.e., the optimization is realized over one, keeping the other fixed [9]. However, when only CS samples are available, recovering the actual signal while simultaneously learning a dictionary is a difficult task. To address this issue, recent works in [10], [11], [12], have proposed methods where the dictionary is learned from CS samples by minimizing the objective function $\|\mathbf{y}_i - \mathbf{\Phi}\mathbf{\Psi}\mathbf{a}_i\|_2^2 \ \forall_i$. However, such DL methods are computationally expensive, and for efficient recovery rely on the assumption e.g., knowledge of signal support set [11]. Alternatively, one can use recovery based DL methods, that are mathematically tractable compared to conventional methods [13], [14]. Here, with an initial dictionary, the current estimate of the recovered signal from compressive samples is used to update the dictionary, and this procedure is performed iteratively till convergence is achieved. Recovery based DL methods are essentially based on the concepts of blind compressed sensing (BCS) [15]. As an alternative, we propose an optimization free method which exploits the properties of compressed signals.

In CS, although the signal acquisition is random, the obtained linear projections or measurements still approximately preserve the signal properties such as mean, variance, as well as the relative distance between two sparse vectors [5]. For instance in [16], it was shown that principal component analysis (PCA) on low-dimensional random projections of data produces the same result as performing PCA on the original data. This was supported by our observation that the compressive samples indeed preserves the envelope of the actual signal. It has been conjectured in the literature that in case of speech signals, the signal envelope is very important in perception, e.g., the words are identified according to their envelope [17]. Thus, our illustrations and experiments are shown using speech signals, but the approach can also be extended to other types of signals. Exploiting the envelope preserving property of CS measurements, we propose a novel method where the aim is to express a signal as a sparse linear combination of prototype signals extracted from CS samples. These prototype signals, can be intrinsic mode functions (IMFs) extracted using empirical mode decomposition (EMD), which is one of the popular methods to capture the envelope of a signal. We show that the IMFs extracted from compressed signal show similar behavior to the ones extracted from the signal itself. Hence, the extracted IMFs can be used to seed a dictionary, using which one can recover the original speech signal from CS samples.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

II. Related Works and Paper Contribution

Given the dictionary, one can efficiently recover a signal from its compressed samples via recovery of its sparse representation [18]. For instance, approaches in [19] and [20], recover a speech signal using a dictionary build from the pre-estimated vocal tract filter coefficients or line spectral frequency (LSF) code book, derived from the training data. Similarly, one can also learn the dictionary directly from the compressed signals. Approaches in [10], [11], [15], [21] and [12] are geared towards image processing applications. While, works in [13] and [22] proposed DL approach for recovery of speech signals. However, in CS domain existing DL algorithms have large computational complexity due to non-convex optimization involved. Moreover, for speech signals (which has lot of variations due to speaker, speaking style or spoken language) the dictionary should preferably be trained on speaker specific training data, which might not be available in each scenario and requires a huge amount of storage.

We propose a novel fast, unsupervised, and optimization free DL method for compressively sensed signals. In this work, we considered the case where only compressed measurements of the actual signal are available without any prior knowledge of signal's support set. We show that it is indeed possible to build a dictionary from CS samples, by bypassing the reconstruction of actual signal i.e., eliminating the abundant cost of recovering the irrelevant data. To this aim, we seed the dictionary using IMFs extracted directly from CS samples, without even comprehending the original signal or the sensing matrix used to acquire the signal. It is worth emphasizing that the goal of the paper is not to outperform a state-of-the-art CS recovery method, but is to propose an approach which can perform with an acceptable level of accuracy in heavily resource-constrained environments, both in terms of storage and computation.

The rest of the paper is organized as follows: In Section III, we briefly explains the modeling of speech signals using CS framework, and how envelope of a speech signal is preserved in compressive samples. In Section IV, we propose an efficient recovery method for compressive speech signals using EMD, and the experimental results are shown in Section V. The summary of paper is given in Section VI.

III. Modeling Signals using CS

In CS framework, given a matrix $\mathbf{Y} \in \mathbb{R}^{m \times l}$ consisting of l compressive signal frames $\{\mathbf{y}_i\}_{i=1}^{l}$ as columns, the recovery of the corresponding signal set $\mathbf{X} \in \mathbb{R}^{n \times l}$ is formulated as [18], [23]:

$$\begin{gathered}\hat{\mathbf{X}} \approx \mathbf{\Psi}\hat{\mathbf{A}} \;\; \text{where } \hat{\mathbf{A}} \text{ is computed as,} \\ \hat{\mathbf{A}} = \underset{\mathbf{A}}{\operatorname{argmin}}\, f(\mathbf{A}) \;\; \text{s.t.} \;\; \|\mathbf{Y} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{A}\|_F^2 = \|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2 < \epsilon,\end{gathered} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{d \times l}$ is the sparse coefficient matrix corresponding to $\mathbf{X}$, ϵ is the error tolerance, $f()$ is a function (e.g., l_1-norm) that promotes sparsity and $\mathbf{D} \in \mathbb{R}^{m \times d}$ is the overall effective dictionary. Provided $\mathbf{\Phi}$ satisfies restricted isometry property (RIP), and is incoherent with $\mathbf{\Psi}$, the signal can be recovered with very high probability by linear programming methods [5].

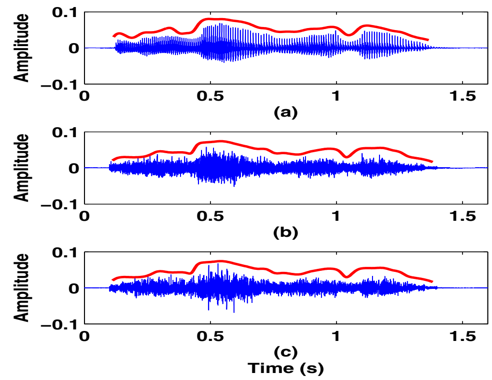


Fig. 1. Comparison of the envelopes (manually marked red) of (a) original speech signal, (b) and (c) interpolated compressive speech signal orignally sampled at compression ratio (m/n) of 0.6 and 0.4 respectively.

A. Randomness Do Make Sense: Properties of Compressive Samples

CS acquires random signal measurements[1] and hence do not preserve any structures of signal in their raw form. However, random projections approximately preserves the properties such as sample mean, variance, as well as the relative distance between sparse vectors of two signals [5], [16], i.e.,

$$\begin{aligned}&\|\mathbf{D}(\mathbf{a}_1 - \mathbf{a}_2)\|_2^2 \approx \|\mathbf{a}_1 - \mathbf{a}_2\|_2^2 \;\forall\, \mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^l \\ &\hat{\mathbf{\Sigma}} = \mathbf{Y}\mathbf{Y}^{\mathbf{T}} = \mathbf{\Phi}\mathbf{X}\mathbf{X}^{\mathbf{T}}\mathbf{\Phi}^{\mathbf{T}} = \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}^{\mathbf{T}} \text{and} \\ &\mathbf{E}\left[\|\mathbf{\Phi}\mathbf{x}\|_2^2\right] = \|\mathbf{x}\|_2^2\end{aligned} \tag{2}$$

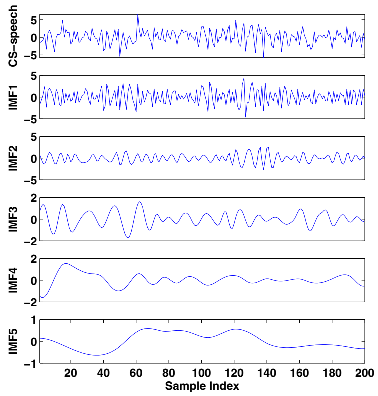
As an illustration, Fig. 1, shows a example of the original and compressively sensed speech signal. Note that the sampling rate of a compressed signal is less than that of the original signal, and for a fair comparison, the interpolated compressed signal, computed using cosine interpolation is plotted in the figure. It can be observed that though the measurement vector exhibits some random noise-like nature, envelopes[2] of both the original and the compressive speech signal are approximately similar, even at different compression ratios. One way to exploit this preserved envelope is to do EMD on compressed signal in order to extract prototype signals for building the dictionary. EMD exploits the signal envelope or evolution of a signal between two consecutive local extrema to decompose a signal into orthogonal modes or IMFs, which can be used as dictionary atoms.

IV. CS-EMD: Fast Signal Recovery from Compressive Samples

The proposed approach sparsely represents a signal frame as a linear combination of few IMFs from the dictionary, selected

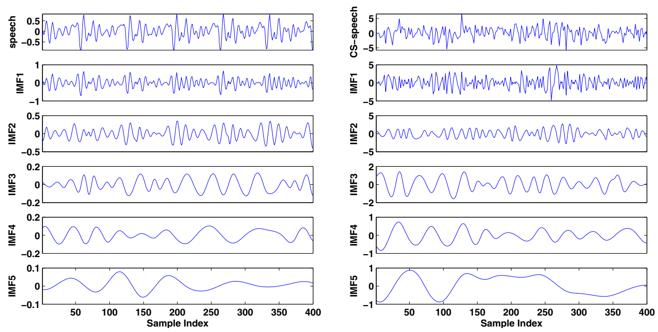
[1]The elements of the sensing matrix are assumed to be i.i.d. random variables

[2]With a slight abuse of definition, envelope here denotes evolution of signal over time/samples



(a)

Fig. 2. EMD decomposition of a voiced frame of compressive speech sampled at compression ratio (m/n) of 0.4



(a) (b)

Fig. 3. EMD decomposition of a voiced frame of (a) orignal speech signal and (b) interpolated compressive speech signal

optimally using sparsity constraints. However, the IMFs used to build the dictionary are extracted from CS samples directly. Using the EMD method, a given compressed signal frame $\mathbf{y}$ is expressed as

$$\mathbf{y} = \sum_{q=1}^{J} \mathbf{m}_q + \mathbf{r} \tag{3}$$

i.e., a sum of J orthogonal modes $\mathbf{m}_q \in \mathbf{R}^m$ and a residual $\mathbf{r} \in \mathbf{R}^m$ [24]. For efficient decomposition, we used the modified EMD algorithms called the Ensemble Empirical Mode Decomposition (EEMD) as proposed in [25]. Figs. 2 and 3 (a) shows an example of compressive and corresponding original voiced speech frame along with the first 5 extracted IMFs, respectively. One can observe that most of the IMFs extracted using CS samples (Fig. 2) show similar behavior as in case of the IMFs extracted using raw speech samples (Fig. 3 (a)). This motivated us to use these IMFs as dictionary atoms. The biggest advantage of this method is its time complexity, which follows from the fact that building the dictionary does not require the sensing matrix to be known. Further, the extracted IMFs being orthogonal results in a dictionary having good coherence bounds. However, there are still two major issues in building the dictionary in order to recover the signal: (1) dimensionality of dictionary atoms, and (2) building a

Algorithm 1 CS-EMD algorithm

Inputs: Compressive signal matrix $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_L]$, and sensing matrix $\mathbf{\Phi}$
Outputs: Recovered signal matrix $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_L]$

Initialization: $\mathbf{\Psi} = [\cdot]$, J, ϵ, β and K_q $\forall_q$ s.t. $d = \sum_{q=1}^{J} K_q$

Preprocessing Stage
Compute $\mathbf{Y}' = [\mathbf{y}_1' \dots \mathbf{y}_L']$, using cosine interpolation on $\mathbf{Y}$
Dictionary Seeding Stage
for: $i = 1$ to L
Compute J IMFs $\mathbf{m}_{qi}, q = 1 \dots J$ from $\mathbf{y}_i'$ using EMD
end for
for: $q = 1$ to J
Collect q^{th} IMFs $\mathbf{m}_{qi}, i = 1 \dots L$ as a column of matrix $\mathbf{M}_q$
Cluster columns of matrix $\mathbf{M}_q$ into K_q clusters
Collect cluster centroids as columns of matrix $\mathbf{C}_K$
Update Dictionary using cluster centroids as $\mathbf{\Psi} = [\mathbf{\Psi} \mid \mathbf{C}_K]$
end for
Sparse Coding and Signal Recovery stage
Solve $\hat{\mathbf{A}} = \text{argmin } \|\mathbf{A}\|_1$ s.t. $\|\mathbf{Y} - \mathbf{\Phi\Psi A}\|_F^2 = \|\mathbf{Y} - \mathbf{DA}\|_F^2 < \epsilon$
Recover signal matrix as $\hat{\mathbf{X}} \approx \mathbf{\Psi}\hat{\mathbf{A}}$

dictionary of appropriate size.

Dimensionality of dictionary atoms: In order to recover the original speech frame, the dimensionality of each dictionary atom and the signal frame should be equal. However, any extracted IMFs from the CS measurement vector will have low dimensionality. It has been shown that EMD can still be effective (within tolerable limits) if the signal is interpolated such as by Fourier and cosine interpolation methods. Hence, we used the raised cosine EEMD method [26] (with roll-off factor $\beta = 1$) to extract IMFs of appropriate dimensions. As an illustration, we have plotted the extracted IMFs of the compressive speech frame considered in Fig. 2 after interpolation using EEMD in Fig. 3 (b). It can be observed that the IMFs are now more similar to the case of original speech signal, and can help in building a better dictionary.

Dictionary size: In general, a signal is processed on short frame basis, and thus a dictionary build using all extracted IMFs from each compressed frame will make it highly overcomplete. Since each level of EMD decomposition has different structural information, to restrict the dictionary atoms to a desired number, the extracted IMFs from the J^{th} level across all signal frames are clustered using K-means algorithm. Now the cluster centers are used as dictionary atoms and the number of clusters depends on number of atoms one wish the dictionary to have from each level. To have a sparser representation, more atoms should come from initial levels which contains more structures/patterns. Algorithm 1 shows the pseudo-code of the proposed approach[3].

A. Computational Complexity

The time complexity of EMD for extracting all possible IMFs from L n-dimensional signal frames approximately scales to $\mathcal{O}(nL \log n)$, that is equal to that of Fast Fourier

[3]Note that apart from the presented approach, one is free to explore any variation of EMD algorithm, clustering approach or other optimal way to build the dictionary from the extracted IMFs. Also, apart from batch processing on all compressive frames, the dictionary can be updated on-line as soon as a new frame is available for processing.

TABLE I
COMPARITIVE ANALYSIS OF DIFFERENT METHODS FOR SIGNAL RECOVERY AVERAGED FOR 20 UTTERANCES OVER 10 TRIALS.

Method	CS Matrix	DL Iterations	PESQ	Runtime
CS-EMD	Sparse-Gaussian	N.A	2.92	0.83 min
	SRM [28]		2.91	
	Gaussian		2.90	
	Bernoulli [29]		2.84	
CS+DCT	Gaussian	N.A	2.30	0.3 min
Blind CS	Gaussian	20	2.97	5 min
IHT	Gaussian	20	3.10	3 min

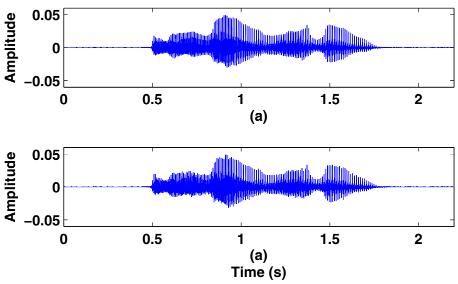


Fig. 4. (a) Original speech signal. (b) Recovered speech signal from compressed measurements at compression ratio m/n of 0.4.

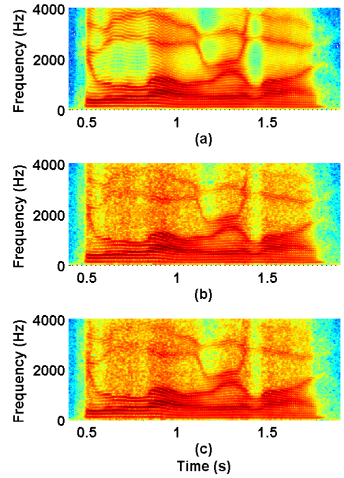


Fig. 5. Spectrogram of (a) original speech signal; (b) and (c) recovered speech signal from compressed measurements at compression ratio m/n of 0.4 and 0.6, respectively.

transform. Further, the complexity of clustering using K-means algorithm is approximately $\mathcal{O}(nLKi)$, where K is the number of clusters and i the number of iterations until convergence. Thus, the overall complexity of the proposed non-iterative approach is less as compared to conventional DL methods, for which the time complexity per iteration scales to $\mathcal{O}(n^2L)$, and in some cases to $\mathcal{O}(n^3L)$ [8].

V. Experimental Results

In our experiments, we have considered speech signals (sampled at 8 KHz) taken from KED TIMIT corpus [27]. Speech is processed on a short time frame basis, where framing is achieved by applying a 50 ms long Hanning window with the frame overlap set to 50%. The sensing matrix $\mathbf{\Phi}$ is chosen to be a random Gaussian matrix with a compression ratio $m/n = 0.4$ unless otherwise stated and the number of dictionary atoms to be learned is set to 600. The maximum number of IMFs extracted using EEMD (with noise realizations $N_e = 50$) for each compressive speech frame is set to 5. As initial IMF levels contribute more towards the overall signal approximation, the number of dictionary atoms chosen empirically from each IMF level across all frames after clustering are 140, 140, 110, 110, and 100, respectively. For reasons of brevity, we shall focus on speech signal recovery, but the proposed approach can also be readily applied to other signal processing applications.

A. Speech recovery from compressive measurements

In this experiment we considered multiple long compressed speech utterances, and for each one a dictionary is learned. The learned dictionary is then applied in the CS framework to obtain the sparse representation of each speech frame using l_1-minimization, solved using YALL1 package [30]. The speech utterance was then reconstructed using the standard overlap and add method. Fig. 4 shows an example of the original and the reconstructed speech waveform, along with spectrogram plots shown in Fig. 5. It can be observed that the proposed method is able to recover the speech signal well. However, as observed in Fig. 4 (b), the first few extracted IMFs are generally corrupted, and as a results the higher frequency bands of the recovered signal are also distorted. This is also supported by a lower perceptual evaluation of speech quality (PESQ) score for the recovered speech using the proposed approach, compared to other recovery based DL methods as shown in Table I. Note that sensing matrices do preserve the envelope, but fails to preserve the pitch related speech variations, and hence they result in poor recovery as compared to other methods. However, some reduction in speech quality is acceptable, considering the time complexity gain achieved via the proposed approach. To illustrate this, Table I also show the average CPU run times to recover a speech utterance of approximately 3 sec (including the time for DL), and the results confirms that the proposed approach is indeed fast compared to existing approaches. Note that for the proposed approach, run time is dominated by sparse coding stage.

1) Discussion: Our experiments shows that one can approximately recover a signal directly from its compressed samples, provided the CS samples preserve statistical properties of the signal. The sampling is done in time domain, and the proposed approach works well for signals that are rich in terms of spectrum and are often highly correlated. Further, the choice of sensing matrix is crucial and if a sensing matrix is carefully chosen or designed one can improve the performance of the proposed approach by learning a better dictionary.

As shown in Table I, compared to random matrices such as Gaussian/Bernoulli matrices, the reconstruction performance increases, in case of carefully designed matrices such as sparse Gaussian and structurally random matrices (SRM)[4]. Note that our goal is to recover a signal from CS samples at the decoder having limited resources both in terms of storage and computation. The extracted IMFs can reveal important properties about signal segments without knowing anything about the sensing matrix used to acquire the signal. For instance, this has promising applications in various inference problems where actual signal recovery is not required, and only CS samples (which require less memory) are available e.g., voiced/nonvoiced speech detection [14] and speech unit classification [3], [31]. However, we defer this or any other extensions to future work.

VI. Summary

In this paper, we have proposed a fast approach for recovery of signals from compressive samples. We show that it is indeed possible to build a dictionary using only compressive samples, and hence the proposed approach is promising in resource-constrained environments. EMD decompositions of compressive samples are used to form the atoms of the dictionary, and is motivated by the fact that CS samples have envelop similar to the envelop of original samples. Preliminary result on speech signal recovery experiment, show that the proposed approach can be an alternative to the existing explicit and implicit CS recovery methods. The full potential of this new approach is yet to be realized, and additional work is required to establish the gains. In our future works, we wish to extend this approach to some inference problems, where actual signal recovery is not required. One possible extension is to incorporate the proposed approach in various speech applications such as voice activity detection, speaker identification or speech recognition.

References

[1] Li Zeng, Xiongwei Zhang, Liang Chen, Zhangjun Fan, and Yonggang Wang, "Scrambling-based speech encryption via compressed sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–12, December 2012.

[2] A. Asaei, H. Bourlard, and V. Cevher, "Model-based compressive sensing for multi-party distant speech recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 4600–4603.

[3] P. Sharma, V. Abrol, A. D Dileep, and A. K. Sao, "Sparse coding based features for speech units classification," in *16th INTERSPEECH*. September 2015, ISCA.

[4] S. Mandal and A. K. Sao, "Edge preserving single image super resolution in sparse environment," in *IEEE International Conference on Image Processing (ICIP)*, September 2013, pp. 967–971.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[6] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, March 2011.

[7] P. Sharma, V. Abrol, and A. K. Sao, "Compressed sensing for unit selection based speech synthesis system.," in *European Signal Processing Conference (EUSIPCO)*, August 2015.

[4]We observed only marginal improvement when sensing matrix other than Gaussian was employed with existing approaches.

[8] Michael Elad, *Sparse and Redundant Representations - From Theory to Applications in Signal and Image Processing.*, Springer, 2010.

[9] R. Rubinstein, A.M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, June 2010.

[10] C. Studer and R. G. Baraniuk, "Dictionary learning from sparsely corrupted or compressed signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 3341–3344.

[11] F. Pourkamali Anaraki and S. M. Hughes, "Compressive K-SVD," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 5469–5473.

[12] M. Aghagolzadeh and H. Radha, "Compressive dictionary learning for image recovery," in *IEEE International Conference on Image Processing (ICIP)*, September 2012, pp. 661–664.

[13] Ch. Srikanth Raj and T. V. Sreenivas, "Time-varying signal adaptive transform and IHT recovery of compressive sensed speech," in *12th INTERSPEECH*, August 2011, pp. 73–76.

[14] V. Abrol, P. Sharma, and A. K. Sao, "Voiced/nonvoiced detection in compressively sensed speech signals," *Speech Communication*, vol. 72, no. 0, pp. 194 – 207, 2015.

[15] S. Gleichman and Y.C. Eldar, "Blind compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6958–6975, October 2011.

[16] J. E. Fowler, "Compressive-projection principal component analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2230–2242, October 2009.

[17] MehmetCenk Sezgin, Bilge Gunsel, and GunesKarabulut Kurt, "Perceptual audio features for emotion detection," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, no. 1, 2012.

[18] E. J. Candés and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[19] T. V. Sreenivas and W. B. Kleijn, "Compressive sensing for sparsely excited speech signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009, pp. 4125–4128.

[20] D. Giacobello, M.G. Christensen, M.N. Murthi, S.H. Jensen, and M. Moonen, "Retrieving sparse patterns using a compressed sensing framework: Applications to speech coding based on sparse linear prediction," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 103–106, 2010.

[21] J. Wormann, S. Hawe, and M. Kleinsteuber, "Analysis based blind compressive sensing," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 491–494, May 2013.

[22] Ch. Srikanth Raj and T. V. Sreenivas, "Joint pitch-analysis formant-synthesis framework for cs recovery of speech," in *13th INTERSPEECH*, September 2012, pp. 947–950.

[23] V. Abrol, P. Sharma, and A. K. Sao, "Speech enhancement using compressed sensing," in *14th INTERSPEECH*. September 2013, pp. 3274–3278, ISCA.

[24] A. Bouzid and N. Ellouze, "Empirical mode decomposition of voiced speech signal," in *First International Symposium on Control, Communications and Signal Processing (ISCCSP).*, March 2004, pp. 603–606.

[25] M.E. Torres, M.A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 4144–4147.

[26] A. Roy and J.F. Doherty, "Raised cosine filter-based empirical mode decomposition," *IET Signal Processing*, vol. 5, no. 2, pp. 121–129, April 2011.

[27] "University of Edinburgh's KED TIMIT," http://festvox.org/.

[28] T.T. Do, Lu Gan, N.H. Nguyen, and T.D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 139–154, Jan 2012.

[29] Gesen Zhang, Shuhong Jiao, Xiaoli Xu, and Lan Wang, "Compressed sensing and reconstruction with bernoulli matrices," in *IEEE International Conference on Information and Automation (ICIA)*, June 2010, pp. 455–460.

[30] J. Yang Y. Zhang and W. Yin, "YALL1: Your algorithms for l1," http://www.yall1.blogs.rice.edu, 2011.

[31] V. Abrol, P. Sharma, and A. K. Sao, "Greedy dictionary learning for kernel sparse representation based classifier," *Pattern Recognition Letters*, vol. 78, pp. 64 – 69, 2016.