

# On Sparse Compression Complexity of Speech Signals

A. N. Omara<sup>\*1</sup>, A. A. Hefnawy<sup>1</sup>, and Abdelhalim Zekry<sup>2</sup>

<sup>1,2</sup>Computers and Systems Department, Electronics Research Institute, Giza, Egypt

<sup>3</sup>Communications and Electronics Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt

<sup>\*</sup>Corresponding author, e-mail: ahmed.omara@eri.sci.eg

## Abstract

In this paper, we have addressed the issue of the sparse compression complexity for the speech signals. First of all, this work illustrated the effect of the signal length on the complexity levels of Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP) algorithms. Also, this paper introduced a study of possibility to reduce that complexity by exploiting the shared atoms among the contiguous speech compressions. By comparing the shared atoms levels and a threshold level induced by an analytic model based on the both the central and non-central hyper-geometric distributions, we proved the ability of the shared atoms criterion to detect if there is biasing towards a subspace of atoms or not, and to decide if the biasing occurs due to the redundancy in the dictionary of atoms, or due to the redundancy in the signal itself. Moreover, we suggested a subspace bias-based approaches for complexity reduction called "Atoms Reuse" and "Active Cluster". Both methods exploits the higher levels of the shared atoms to reduce the compression complexity by reducing the search space during the pursuit iterations.

**Keywords:** Sparse Compression, Speech Signal, Complexity, Shared Atoms, Central and Non-Central Hyper-Geometric Distributions

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

## 1. Introduction

Nowadays, one of the efficient signal representations is the sparse modeling. This type of signal decomposition has recently received extensive research interest across several communities including signal processing, information theory, and optimization [1, 2, 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 atoms are larger than bases are needed to build sparse representations of complex signals [7]. But choosing is difficult and requires more complex algorithms.

Letting  $\Phi$  denotes a dictionary matrix of size  $M \times N$  (where typically  $M < N$ ) and  $y$  denotes a signal vector in  $R^M$ . The goal of sparse decomposition algorithms such as Matching Pursuit (MP) [8], Orthogonal Matching Pursuit (OMP) [9], Optimized Orthogonal Matching Pursuit (OOMP) [10], Backward-Optimized Orthogonal Matching Pursuit BOOMP [11], and others is to recover a coefficient vector  $x \in R^N$  with roughly  $k < M$  nonzero terms so that  $\Phi x$  equals  $y$  exactly or approximately.

$$y \simeq \Phi x \quad (1)$$

Actually, the aforementioned greedy algorithms and others are mainly concerned with decomposing a single vector sparsely regardless of the signal is a unique vector or splitted to many vectors. Naturally, the long signals such as speech signals should be splitted to  $F$  frames or vectors indexed by  $Y_j$  before the coding process. So, the sparse approximation of a signal  $Y \in R^{M \times F}$  will be  $\Phi X$  such that  $X \in R^{N \times F}$  and  $X_j$  represents the sparse decomposition of vector  $Y_j$ . This can be written in the following form

$$[Y_1 \ Y_2 \ \dots \ Y_F] \simeq \Phi [X_1 \ X_2 \ \dots \ X_F] \quad (2)$$

It is intuitively obvious that, the computational complexity of (2) is larger than that of (1) due to the signal length. So, in this research, we initiate a new trend in the complexity reduction, whose main idea is to resize the dictionary of atoms during the pursuit iterations. The intended sub-dictionary is the subspace of atoms at which the sparse compressions biases towards it. So, in this work, we try to exploit some of the  $F$  vectors to detect that if the sparse compressions biases to a subspace of atoms or not. In other words, the signal under consideration may live in the span of a subspace of  $\Phi$ . This subspace biasing may occur due to the redundancy nature of the speech signal, or due to the redundancy nature of  $\Phi$ .

The main contribution of this paper is introducing a new criterion so-called the "Shared Atoms" that can be monitored during the successive sparse compressions and then we can decide if there is subspace biasing or not. Finally, this paper is organized as follows. Section 2. studies the effect of the signal length on the complexity levels of MP and OMP algorithms. Section 3. reviews the related works on enhancing the pursuit algorithms complexity. Section 4. will study the shared atoms criterion from a probability standpoint illustrating its expected levels and bounds. Section 5. will illustrate the indications of the shared atoms and how we can benefit from it in achieving a satisfied complexity levels. Section 6. contains experimental results. Finally, conclusions are provided in Section 7..

## 2. Sparse Compression Complexity

Since the pursuit algorithms don't consider the splitting process, it will handle each vector independently. So, it is logic to say that there are three complexity levels. The first level is called the "iteration-based complexity" and depends on the atom choice methodology. The second level is called the "sparsity-based complexity" and depends on the sparsity level  $k$  or the number of nonzero elements. Both complexity levels are considered fixed per each independent decomposition if and only if the sparse modeling arguments are identical such as the sparsity level  $k$  and the dictionary size  $N$ . Finally, The third complexity level is due to the overall decomposition of the  $F$  vectors, and in this case the computational complexity depends on  $F$ .

Generally, the time complexity  $T$  of any pursuit algorithm could be denoted as  $O(\hat{G})$  where the function  $\hat{G}$  represents the fastest growing term in another function  $G(F, M, N, k)$ . According to the rule of the big  $O$  notation and for a positive constant  $\varepsilon$  the upper bound of  $T$  can be obtained as follows [12]:

$$T \leq \varepsilon \hat{G} \quad (3)$$

Usually, the function  $G$  represents the number of the elementary arithmetic operations in the algorithm such as the multiplications  $G_{\Pi}$  and the additions  $G_{\Sigma}$ . According to this definition the function  $G$  could be represented as follows:

$$G(F, M, N, k) = \sum_{j=1}^F \sum_{i=1}^k (G_{\Pi}^{(i)} + G_{\Sigma}^{(i)})_j \quad (4)$$

This representation of  $G$  means that, during the decomposition of vector  $Y_j$ , both  $G_{\Pi}^{(i)}$  and  $G_{\Sigma}^{(i)}$  represent the number of the elementary operations at the  $i$ th iteration. Table 1 summarizes the key procedures of the two most common algorithms {MP, OMP} and their parameters  $\{G_{\Pi}^{(i)}, G_{\Sigma}^{(i)}\}$ . As shown in the table, for MP algorithm, the computation complexity at the iteration  $i$  focuses on the procedure  $\Phi^T r_{i-1}$ . This procedure calculates the correlation between the residual error  $r_{i-1}$  of the previous iteration and the  $N$  elements of  $\Phi$ . For  $r_{i-1} \in R^M$  and  $\Phi \in R^{M \times N}$ , the correlation process requires  $N$  vector multiplications and each vector multiplication consists of  $M$  elementary products and  $M - 1$  additions. As illustrated in the Table 1, both  $G_{\Pi}^{(i)}$  and  $G_{\Sigma}^{(i)}$  can be written as  $M(N + 1) + N$  and  $M(N + 1) - N$  respectively. By substituting into (4) we obtain  $G_{MP} = \sum_{j=1}^F \sum_{i=1}^k [2M(N + 1)]_j$ . Finally, the function  $G$  can be written in the form

$$G_{MP} = 2FkMN \left[ 1 + \frac{1}{N} \right] \quad (5)$$

Procedure	Routine	Subroutine	MP		OMP	
			$G_{\Pi}^{(i)}$	$G_{\Sigma}^{(i)}$	$G_{\Pi}^{(i)}$	$G_{\Sigma}^{(i)}$
Atom Sel.	$\Theta^{N \times 1} = \Phi^T r_{i-1}$		$MN$	$(M-1)N$	–	–
	or $\Theta_{(i)} = \Phi_{(i)}^T r_{i-1}$		–	–	$\frac{M}{(N-i+1)^{-1}}$	$\frac{M-1}{(N-i+1)^{-1}}$
	$\varphi_i = \text{argmax} \Theta $		$N$	$0$	$N-i+1$	$0$
$\hat{r}_{i-1}$	$\langle r_{i-1}, \varphi_i \rangle \varphi_i$		$M$	$0$	–	–
	or					
	$\mathfrak{G} = \Phi_i^T \Phi_i$		–	–	$Mi^2$	$(M-1)i^2$
	$A = \mathfrak{G}^{-1}$		–	–	$i^3$	$i^3 - 2i^2 + i$
	$B = \Phi_i A$		–	–	$Mi^2$	$M(i^2 - i)$
	$C = B \Phi_i^T$		–	–	$M^2 i$	$M^2(i-1)$
	$D = C r_{i-1}$		–	–	$M^2$	$M^2 - M$
Err Update	$r_i = r_{i-1} - \hat{r}_{i-1}$		$0$	$M$	$0$	$M$

Table 1.  $\{G_{\Pi}^{(i)}, G_{\Sigma}^{(i)}\}$  for MP and OMP

As shown in (5), the fastest growing term consists of the multiplication  $FkMN$ . So, the time complexity of MP can be represented in terms of the big  $O$  notation by  $T_{MP} = O(FkMN)$ . If we keep  $F$  out of  $T_{MP}$  the result is similar to the proved complexity expressions for the MP algorithm in [13] and [14].

Unlike MP, the computation complexity of OMP at the iteration  $i$  is distributed among the atom selection and the coefficients update procedures. Note that, we got  $G_{\Pi}$  and  $G_{\Sigma}$  for the Gram matrix inverse  $\mathfrak{G}^{-1}$  from [15]. As shown in the Table 1, at the  $i$ th iteration, the algorithm updates two sets simultaneously  $\Phi_{(i)}$  and  $\Phi_i$  such that  $\Phi_{(i)} = \Phi_{(i-1)} - \{\varphi_{i-1}\}$ ,  $\Phi_i = \Phi_{i-1} \cup \{\varphi_i\}$  and the initial sets are  $\Phi_{(1)} = \Phi$  and  $\Phi_0 = \emptyset$ , where  $\emptyset$  refers here to the empty set and  $\varphi_i$  refers to the selected atom at the iteration  $i$ . By substituting the values of  $G_{\Pi}^{(i)}$  and  $G_{\Sigma}^{(i)}$  into (4) we obtain  $G_{OMP}$  approximately in the form

$$G_{OMP} \approx 2FkMN \left[ 1 + \frac{M}{N} \right] \quad (6)$$

As shown in (6), the fastest growing term consists of the multiplication  $FkMN$ . So, the time complexity of OMP can be represented in terms of the big  $O$  notation by  $T_{OMP} = O(FkMN)$ .

### 3. Related Work

Over the last years, many methods being made in regards to reducing the complexity levels of the sparse pursuit algorithms. The majority of these approaches can be categorized into four groups. There are fast transformation-based methods, clustering-based methods, matrix factorization-based methods and optimization-based methods.

In fast transformation-based methods, the pursuit algorithm exploits the fast computations property of the Fast Fourier Transforms (FFT)[16], or the Fast Wavelet Transforms (FWT)[17]. The basic idea behind this trend is to use pre-structured dictionaries whose atoms are theoretically based, such as Fourier bases and wavelets. The fast transform algorithms reduces the number of elementary multiplications in matrix-vector product of  $\Phi^T r_{i-1}$  from  $MN$  to  $N \log M$ . Although pre-structured dictionaries lead to fast sparse compressions, they are limited in their ability to sparsify the signals they are designed to handle. Furthermore, most of those dictionaries are restricted to signals of a certain type, and cannot be used for a new and arbitrary family of signals of interest.

In clustering-based methods, the approaches exploit the correlation property among the atoms of  $\Phi$ . Due to the over-completeness of  $\Phi$ , there are highly correlated atoms that have similar properties. So, by means of clustering the similar atoms be grouped in a cluster and this procedure reduces the search time in those clusters. For example, the author in [18] proposed an efficient dictionary organization technique. This technique groups similar atoms together, and represent them by a unique element called molecule. Applying clustering recursively on atoms and molecules yields a hierarchical tree structure, which can be exploited to design a search algorithm with greatly reduced complexity.

To speed up the processing of matrix operations such as the inverse of gram matrix  $\mathbb{G} = \Phi_i^T \Phi_i$  (see Table 1), there are different matrix factorization-based methods can be used for that purpose. The Cholesky-based OMP [19] and QR-based OMP [20] use the Cholesky and QR factorization methods respectively to reduce the complexity of  $\mathbb{G}^{-1}$ . The basic idea behind Cholesky factorization is to decompose a Hermitian, positive-definite matrix  $\mathbb{G}$  into the product of a lower triangular matrix  $\mathbb{G}_L$  and its conjugate transpose  $\mathbb{G}_L^T$ . While the basic idea behind QR factorization is to decompose a real, square matrix into the product of an orthogonal matrix Q and an upper triangular matrix R.

In optimization-based methods, the approaches try to speed up the orthogonal projection process needed by the OMP algorithm to update the coefficients. As depicted in Section 2., the difference in complexity between MP and OMP algorithms is concentrated in the coefficients update procedure that requires to solve  $r_{i-1} = \Phi_i b$ , where  $b$  is the vector of unknown coefficients. The original OMP solves this problem using the least squares method  $b = (\Phi_i^T \Phi_i)^{-1} \Phi_i^T r_{i-1}$ . Fast solvers for the linear equations had been exploited to approximate the orthogonal projection of the least squares method, for example, the author in [21] replaced the least squares method by another fast approaches such as the gradient, the conjugate gradient [22] and the approximate conjugate gradient methods.

#### 4. Subspace Bias-based Efficient Sparse Compression

As shown in the literature review, all efforts that had been made to reduce the computational complexity ignored the nature of the signal under consideration. In this paper, we will study the possibility to exploit the redundancy nature of the speech signal and the dictionary to make an efficient sparse compression. We seek to use a criterion called the "Shared Atoms" to work as a biasing monitor that detect if there is biasing towards a subspace of atoms or not, and decide if the biasing occurs due to the redundancy in the dictionary of atoms, or due to the redundancy in the signal itself or due to both.

##### 4.1. Shared atoms

Assume that  $\Phi = \{\varphi_n; n \in \Gamma\}$  is a dictionary of atoms,  $\Gamma$  is a set of indices. Let  $\Gamma^{(j)}$  and  $\Gamma^{(j+\psi)}$  be two support sets or subsets of  $\Gamma$  consisting of the nonzero elements indices in  $X_j$  and  $X_{j+\psi}$  respectively such that  $\psi$  is the neighborhood degree to the index  $j$ . Then, the indices of the shared atoms among  $\Phi X_j$  and  $\Phi X_{j+\psi}$  can be described as

$$\Gamma^{(j)} \cap \Gamma^{(j+\psi)} = \{\varphi_n; n \in \Gamma^{(j)} \text{ and } n \in \Gamma^{(j+\psi)}\} \quad (7)$$

Let  $C(\psi)$  be the cardinality of the intersection set in (7) or the number of the shared atoms, and given by

$$C(\psi) = |\Gamma^{(j)} \cap \Gamma^{(j+\psi)}| \quad (8)$$

Also, this cardinality can be expressed as  $\|X_j \odot X_{j+\psi}\|_0$  where the notation  $\odot$  stands for the Hadamard (or elementwise) product of two vectors. Let  $k_j$  and  $k_{j+\psi}$  be the number of the non zero elements in  $X_j$  and  $X_{j+\psi}$  respectively or the number of the elements in  $\Gamma^{(j)}$  and  $\Gamma^{(j+\psi)}$  respectively, then the maximum value of  $C(\psi)$  can be obtained when  $\Gamma^{(j)} \subseteq \Gamma^{(j+\psi)}$  or  $\Gamma^{(j+\psi)} \subseteq \Gamma^{(j)}$ . And the minimum value can be obtained when  $\Gamma^{(j)} \cap \Gamma^{(j+\psi)} = \emptyset$ . Mathematically, this can be by  $0 \leq C(\psi) \leq \inf\{k_j, k_{j+\psi}\}$ . If  $k_j = k_{j+\psi} = k$ , then we have  $0 \leq C(\psi) \leq k$ . To estimate the value of  $C(\psi)$ , let  $P(C(\psi) = \theta | \Gamma^{(j)})$  denotes the probability of  $\theta$  shared atoms among  $\Phi X_j$  and

$\Phi X_{j+\psi}$ , given the subset  $\Gamma^{(j)}$  with cardinality  $k_j$ . Then the expected value of  $C(\psi)$  is given by  $\mathbb{E}_{C(\psi)} = \sum_{\theta=0}^{k_j+\psi} \theta P(C(\psi) = \theta | \Gamma^{(j)})$ .

#### 4.1.1. Unbiased case

For a redundant  $\Phi$  consisting of  $N$  atoms, assume that all  $N$  atoms have the same chance during the selection procedure, i.e. they are chosen uniformly and somewhat randomly without replacement. Then the expected value of  $C(\psi)$  can be expressed in terms of the central hyper-geometric distribution

$$\mathbb{E}_{C(\psi)} = \sum_{\theta=0}^{k_j+\psi} \theta \frac{\binom{k_j}{\theta} \binom{N-k_j}{k_j+\psi-\theta}}{\binom{N}{k_j+\psi}} \quad (9)$$

So, the expected value is expressed by  $\mathbb{E}_{C(\psi)} = \frac{k_j+\psi k_j}{N}$ . For the case  $k_j = k_{j+\psi} = k$ , we have  $\mathbb{E}_{C(\psi)} = \frac{k^2}{N}$ , and the probability to obtain the upper bound of  $C(\psi)$  is given by  $P(C(\psi) = k | \Gamma^{(j)}) = 1/\binom{N}{k}$ . This probability reaches its maximum value when  $N = k$  but this condition is not practical. For any sparse compression, the number of atoms  $N$  must be greater than the dimension of the signal  $M$  which is naturally greater than  $k$ . On the other side, the probability of zero shared atoms is given by  $P(C(\psi) = 0 | \Gamma^{(j)}) = \binom{N-k}{k}/\binom{N}{k}$ , and this probability reaches its maximum value for the condition  $N \gg k$ .

#### 4.1.2. Biased case

Assume each atom in  $\Gamma^{(j)}$  has the weight  $\omega_1$ , and each atom in  $\Gamma - \Gamma^{(j)}$  has the weight  $\omega_2$ . Then, the expected value of  $C(\psi)$  can be expressed in terms of the Wallenius non-central hyper-geometric distribution [23]

$$\mathbb{E}_{C(\psi)} = \sum_{\theta=0}^{k_j+\psi} \theta \binom{k_j}{\theta} \binom{N-k_j}{k_j+\psi-\theta} \int_0^1 [f(t)]^\theta dt \quad (10)$$

where,

$$f(t) = \frac{1 - t^{\omega/\mu}}{(1 - t^{1/\mu})^{-k_j+\psi}}, \quad \omega = \frac{\omega_1}{\omega_2}, \text{ "Odds Ratio"} \quad (11)$$

$$\mu = \omega(k_j - \theta) + (N - k_j - k_{j+\psi} + \theta) \quad (12)$$

As illustrated in [24], the expected value of the Wallenius distribution is approximated by solution  $\mathbb{E}_{C(\psi)}$  to

$$\left(1 - \frac{\mathbb{E}_{C(\psi)}}{k_j}\right) = \left(1 - \frac{k_j+\psi - \mathbb{E}_{C(\psi)}}{N - k_j}\right)^\omega \quad (13)$$

For the case of equal iterations, the Equation (13) can be rewritten as follows

$$\left(1 - \frac{\mathbb{E}_{C(\psi)}}{k}\right) = \left(1 - \frac{k - \mathbb{E}_{C(\psi)}}{N - k}\right)^\omega \quad (14)$$

From (14) we can obtain the bounds of  $\mathbb{E}_{C(\psi)}$  at the bounds of  $\omega$ . When  $\omega$  tends to infinity, i.e.,  $\omega_1 \gg \omega_2$  then the right-hand side (RHS) term equals 0 because the variable within the brackets is less than 1, and then  $\mathbb{E}_{C(\psi)}$  equals  $k$ . But for  $\omega \approx 0$ , i.e.,  $\omega_2 \gg \omega_1$ , the RHS term equals 1, and then  $\mathbb{E}_{C(\psi)}$  equals 0.

## 5. The Indications of $C(\psi)$ And The Benefits

One of the main conclusions of Section 4.1. was that, as long as the redundant dictionary  $\Phi$  consisting of finite number of atoms  $N$ , there is inherent shared atoms among the successive compressions  $(\Phi X_j, \Phi X_{j+1})$  regardless of the redundancy in the signal and the dictionary. Before experiments are conducted to measure  $C(\psi)$ , this section will illustrate the indications of the shared atoms according to its level and its dependency on  $\psi$ .

### 5.1. Probable levels of $C(\psi)$

From Equation (13), we can conclude the probable levels of  $C(\psi)$  according to the odds ratio  $\omega$  as follows

$$\mathbb{E}_{C(\psi)} = \frac{k_{j+\psi}k_j}{N}, \quad \omega = 1 \mathbb{E}_{C(\psi)} > \frac{k_{j+\psi}k_j}{N}, \quad \omega > 1 \mathbb{E}_{C(\psi)} < \frac{k_{j+\psi}k_j}{N}, \quad \omega < 1 \quad (15)$$

The proof of (15) is given in (A). As shown, one of the probable levels of  $C(\psi)$  is the hypergeometric mean  $k_j k_{j+\psi} / N$ . This level is considered a special case and indicates that all  $N$  atoms of  $\Phi$  are in use and there is no bias to any subset of atoms. The most important level of  $C(\psi)$  is that at which the shared atoms are greater than  $k_j k_{j+\psi} / N$ . In this case, the shared atoms indicates that there is somewhat bias to a subset of atoms in the space  $\Phi$ .

### 5.2. Compression complexity and $C(\psi)$

The bias property of  $C(\psi)$  may lead us to two different trends in the compression complexity reduction. The first trend is called "Atoms Reuse", and the second is called "Active Cluster". To justify the importance of measuring  $C(\psi)$ , let us illustrate the enhancement levels in complexity which may get them if we have benefited from  $C(\psi)$ .

#### 5.2.1. Atoms Reuse

Like the contiguous pixels in image signal, the contiguous speech frames may have common features. If the atoms in  $\Phi$  have the ability to describe different signal features, then we can expect that the level of  $C(\psi)$  are greater than the threshold level  $k_j k_{j+\psi} / N$ , and there are  $\delta$  atoms can be reused among the contiguous speech compressions.

*Proposition 1:* Assume  $k_j = k_{j+\psi} = k$ , if the MP algorithm is forced to select  $\delta$  atoms from the current support set  $\Gamma^{(j)}$  in the next compression, then the relative enhancement in complexity reaches up to  $\frac{\delta}{k}(1 - \frac{k}{N})$ . As for the OMP algorithm, the relative enhancement in  $G_{OMP}$  is slightly less than that in  $G_{MP}$  and depends on the signal dimension  $M$ .

The proof is given in (B) and (C).

#### 5.2.2. Active Cluster

If  $C(\psi)$  is greater than  $k_j k_{j+\psi} / N$  regardless of  $\psi$ , then we can say that there is biasing towards different subsets of atoms, and there is an active subspace  $\Phi' \subset \Phi$ .

*Definition:* In this paper, we will define the efficiency of the space  $\Phi$  by  $\eta = \frac{k_j k_{j+\psi} / C(\psi)}{N}$

*Proposition 2:* Assume  $k_j = k_{j+\psi} = k$ , if there are  $F'$  frames must be compressed before detecting the active cluster  $\Phi'$  with cardinality  $|\Phi'| = N'$ , such that the remainder frames  $(F - F')$  are compressed using the subspace  $\Phi'$ . Then the relative enhancement in the complexity reaches to  $(1 - \alpha)(1 - \lambda_1 \beta)$ , and  $(1 - \alpha)(1 - \lambda_2 \beta)$  for MP and OMP respectively, where  $\alpha$ ,  $\beta$ ,  $\lambda_1$  and  $\lambda_2$  represent the ratios  $\frac{F'}{F}$ ,  $\frac{N'}{N}$ ,  $\frac{1+1/N'}{1+1/N}$  and  $\frac{1+M/N'}{1+M/N}$  respectively. The proof is given in (D) and (E). Note that, in this paper we do not care about how we select the  $N'$  atoms from the dictionary, but we consider only its effect on the relative complexity.

## 6. Experimental Results

After declaring the importance of  $C(\psi)$  in the Subsections 5.2.1. and 5.2.2., in this section we will present a set of experiment results. The intention of these experiments is to illustrate the

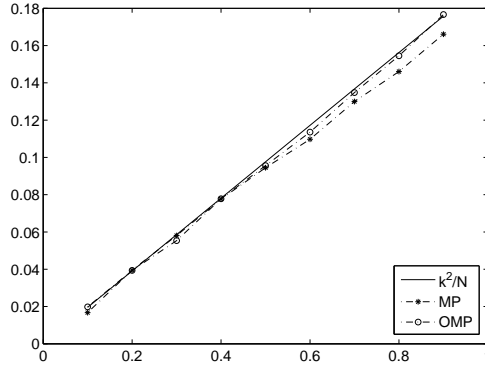


Figure 1. Exp.(1).  $\frac{C}{k}$  vs.  $\frac{k}{M} = 0.1, \dots, 0.9$

Table 2. Exp.(1).  $\omega$  vs.  $\frac{k}{M}$

$\frac{k}{M}$	No bias level	MP	OMP
0.1	1	0.853	1.015
0.2	1	1.013	1.009
0.3	1	0.991	0.94
0.4	1	0.993	0.995
0.5	1	0.962	0.975
0.6	1	0.924	0.963
0.7	1	0.939	0.982
0.8	1	0.917	0.986
0.9	1	0.927	1.007

ability of the shared atoms criterion to detect the existence of the subspace biasing and to show the impact of the signal redundancy and the dictionary redundancy on the biasing levels. All experiments use the most common sparse algorithms MP and OMP and were designed in MATLAB environment using the sparse compression toolbox. Also, all simulations use the ISOLET speech database [25], and the main parameters of the sparse compressions are  $M = 100$ , pre-structured dictionary with  $N = 512$  and finally the sparsity level  $k$  ranges from 10 to 90 nonzero elements. To facilitate comparison, we have normalized  $k$  and  $C(\psi)$  by dividing all them by their maximum level  $M$  and  $k$  respectively.

### 6.1. Experiment 1: Bias of $\Gamma^{(j+\psi)}$ to a random set $\Gamma^{(r)}$

In this experiment, the set  $\Gamma^{(j)}$  is replaced with another set  $\Gamma^{(r)}$  whose  $k$  elements are uniformly selected from the whole elements of the dictionary  $\Phi$ . So, the function  $C$  will represent the cardinality of  $\Gamma^{(r)} \cap \Gamma^{(j+\psi)}$ . Figure (1) illustrates that, the average values of  $C(\psi)$  are approximately identical to the expected values determined by the central hyper-geometric distribution, and this result indicates that there is no bias towards the randomly selected subsets  $\Gamma^{(r)}$ . We used the  $t$ -test for comparing the means and the result showed that the values of  $sig$  parameter are greater than 0.05 and equal 0.886, 0.975 for MP and OMP respectively which scientifically means that the variability is not significantly different. Note that, we can obtain the bias level according to the value of the odds ratios " $\omega$ " that can be obtained directly from Equation (14) to be  $\omega = \frac{\log(1-\frac{C}{k})}{\log(1-\frac{k-C}{N-k})}$ . As evidenced by Table 2, the odds ratio converges to the "No bias" level or 1 for all  $k/M$  values. This result illustrated that, there is no bias to any randomly selected  $k$  atoms.

### 6.2. Experiment 2: Bias of $\Gamma^{(j+\psi)}$ to $\Gamma^{(j)}$

In the second experiment we will try to measure  $C(\psi)$  in two cases. In first case, we will measure it when the positions of  $Y_j$  in the matrix  $Y$  are left unchanged. In this case, the function  $C(\psi)$  takes into considerations the chronological order of frames. On the contrary, in the second case, we will measure the shared atoms when the positions of  $Y_j$  in  $Y$  are changed to ignore the effect of the chronological order of speech frames, and to check the efficiency of the space of atoms. In both cases, we chose the fifth order polynomial for fitting the measurements of  $C(\psi)$  for  $\psi = 1$  to 50.

Figure (3) illustrates that, the chronological order of  $Y_j$  has a great effect on  $C(\psi)$ . For the first case,  $C(\psi)$  degrades as  $\psi$  increases. Also, it is noted that  $C(1)$  is the maximum value which means that the largest shared atoms occurs among the adjacent decompositions. This case is in stark contrast to the second case, as shown in Figure (4), the values of  $C(\psi)$  tends to constant levels for all values of  $\psi$ . This result illustrates the absence of the chronological order of  $Y_j$  and its effect on  $C(\psi)$ . It is also interesting to note that all values of  $C(\psi)$  are larger than the central hyper-geometric threshold level  $k^2/N$  which means that the odds ratios of the Wallenius

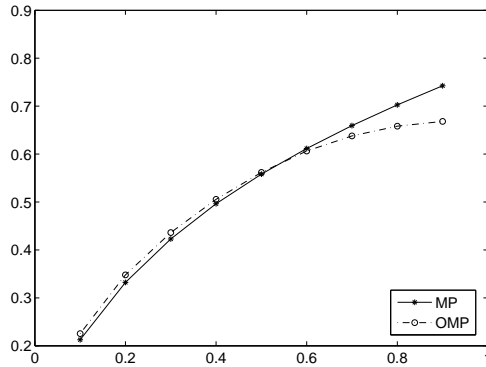


Figure 2. Exp.(2), case 2.  $\eta$  vs.  $\frac{k}{M} = 0.1, \dots, 0.9$

Table 3. Exp.(2), case 2.  $\omega$  vs.  $\frac{k}{M}$

$\frac{k}{M}$	No bias level	MP	OMP
0.1	1	5.278	4.936
0.2	1	3.425	3.239
0.3	1	2.707	2.604
0.4	1	2.312	2.258
0.5	1	2.057	2.039
0.6	1	1.874	1.896
0.7	1	1.732	1.816
0.8	1	1.618	1.78
0.9	1	1.521	1.786

distribution are greater than 1. Table 3 shows the odds ratio levels for the second case results. The values of the odds ratios proved that, the bias property is clearly visible at the low levels of  $k/M$ , but the odds ratios are slightly more than the "No bias" level at the higher sparsity levels.

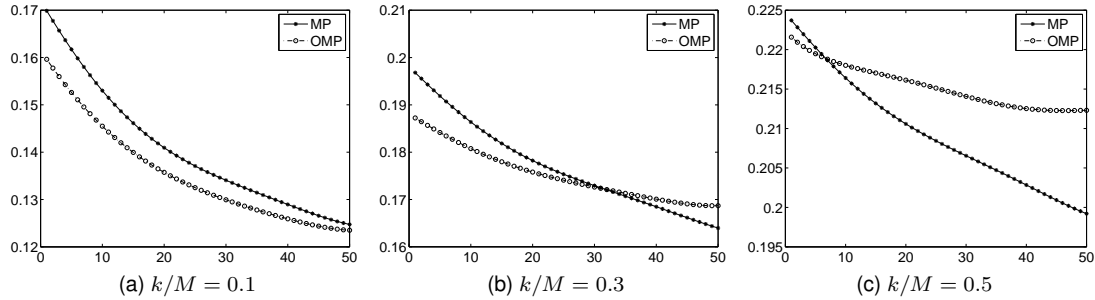


Figure 3. Exp.(2), case 1:  $C/k$  vs.  $\psi = 1, \dots, 50$

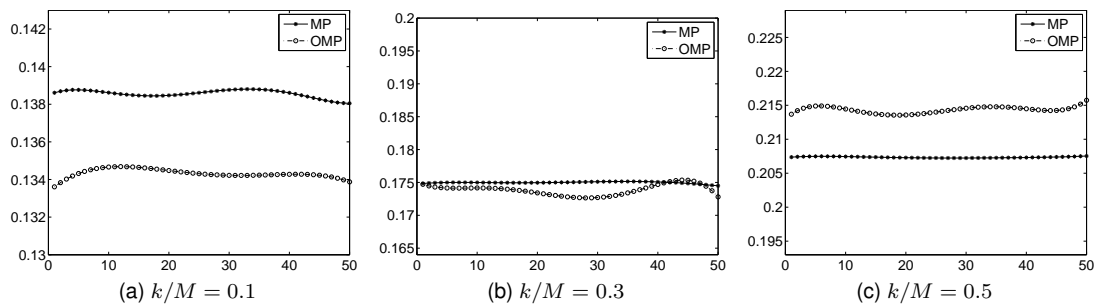


Figure 4. Exp.(2), case 2:  $C/k$  vs.  $\psi = 1, \dots, 50$

The results of both cases confirmed that, there are two types of subspace biasing. First type of bias is called " $\psi$ -based bias", and this bias makes  $C(\psi)$  to change slightly with respect to  $\psi$ . While the other type is called " $\Phi'$ -based bias", and this bias makes  $C(\psi)$  to have significant values greater than  $k^2/N$  regardless of  $\psi$ . Finally, Figure (2) shows the dictionary efficiency  $\eta$  for different values of  $k/M$ . As shown in the figure, the pre-structured dictionary have efficiencies ranges approximately from 20% to 70%.

### 6.3. Experiment 3: Speech redundancy and shared atoms

As illustrated in the results of the second experiment, the suggested criterion  $C(\psi)$  refers to a little bias to the contiguous support set, and this result is logic because the pre-structured



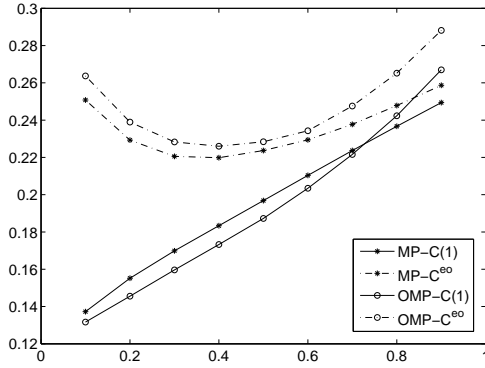


Figure 5.  $\frac{C(1)}{k}$  &  $\frac{C^{eo}}{k}$  vs.  $\frac{k}{M} = 0.1, \dots, 0.9$

Table 4. Exp.(3).  $\omega$  vs.  $k/M$

$\frac{k}{M}$	No bias level	MP		OMP	
		$C(1)$	$C^{eo}$	$C(1)$	$C^{eo}$
0.1	1	8.52	19.2	8.09	20.72
0.2	1	4.83	8.18	4.45	8.69
0.3	1	3.51	5.01	3.24	5.27
0.4	1	2.82	3.63	2.62	3.78
0.5	1	2.41	2.89	2.25	2.97
0.6	1	2.13	2.42	2.04	2.49
0.7	1	1.93	2.11	1.9	2.24
0.8	1	1.77	1.9	1.84	2.11
0.9	1	1.65	1.74	1.83	2.06

atoms "wavelet basis" are local descriptors. So, in this experiment we will try to remeasure the shared atoms among the highly correlated frames. We will use a channel splitting technique in [26] to extract two highly correlated frames instead of the sequential splitting procedure used in the previous experiment. To extract two highly correlated frames based on the idea [26], we will split the signal  $Y \in R^{M \times F}$  to two signals, even samples signal  $Y^e \in R^{M \times \lceil \frac{F}{2} \rceil}$ , and odd samples signal  $Y^o \in R^{M \times \lceil \frac{F}{2} \rceil}$ . The corresponding sparse coefficients matrices will be  $X^e \in R^{N \times \lceil \frac{F}{2} \rceil}$  and  $X^o \in R^{N \times \lceil \frac{F}{2} \rceil}$  respectively. In this experiment we will measure the average values of  $C_i^{eo}$  such that  $C_i^{eo} = \|X_i^e \odot X_i^o\|_0$ . Note that,  $X_i^e$  and  $X_i^o$  are the column vectors in  $X^e$  and  $X^o$  respectively.

Figure (5) illustrates the effect of speech splitting on the level of the shared atoms. As depicted in figure, the even-odd splitting method increases the shared atoms significantly at the low sparsity level. But, at the higher levels of  $k/M$  the values of  $C/k$  converges slightly to that levels of the second experiment.

As for the effect of the even-odd splitting on the biasing levels, Table 4 illustrates that the redundancy among the contiguous speech samples increased the odds ratios significantly at the lower sparsity levels.

## 7. Conclusions and Future works

In this research, a particular attention was paid to the sparse compression complexity of the speech signal. In the first part of this paper, we illustrated the effect of the signal length  $F$  on the computational complexity  $G$ . As shown in Section 2., the complexity levels increased linearly from  $O(kMN)$  to  $O(FkMN)$ . If we look deeply into the multiplication term  $FkMN$ , we can see that  $F$ ,  $k$  and  $M$  are unchangeable parameters. For this reason, we have sought to exploit the redundancy of the dictionary and the redundancy of the signal itself to resize  $\Phi$  according to the biasing of the sparse compressions towards a subspace of atoms.

In this paper we have suggested two subspace bias-based approaches that resize the dictionary during the iterations either by forcing the algorithm to select some atoms from the last support set such as in the "Atoms Reuse" approach, or by ignoring some atoms from the whole dictionary  $\Phi$  such as in the so-called "Active Cluster" approach. Since both approaches are applicable if there is somewhat biasing towards subspace of atoms, in this research, we do not care about applying the approaches, but we considered only how to detect the biasing. So, we have suggested the "Shared Atoms" criterion that can be measured through the successive compressions and then we can decide if there is biasing towards a subspace of atoms or not according to the gap between the measured level and the analytic level  $k^2/N$ .

Through the experimental results of Section 6., we have concluded that the suggested criterion have the ability to detect the subspace biasing. Also, it was evident that the biasing appears significantly at the lower sparsity levels and partially disappear at the higher levels. Moreover, the odds ratios illustrated that the biasing levels due to the dictionary redundancy ranges approximately from 2 to 5 for  $\frac{k}{M} \leq 60\%$ . But, for the signal redundancy, it ranges from 2 to 20 for

$\frac{k}{M} \leq 60\%$ . These results encouraged us to extend the research in the future to implement the approaches and to study the impact of the dictionary resizing on the final approximation error and on the quality of the speech signal. Furthermore, we seek to join between the unknown elements of the active subspace  $\Phi'$  and the union of different shared atoms, and to join between the number of those elements and the measured efficiency  $\eta$  of the dictionary.

## References

- [1] P. Fildik and P. Fiildihk, "Forming sparse representations by local anti-hebbian learning," *Biological Cybernetics*, vol. 64, pp. 165–170, 1990.
- [2] M. S. Lewicki, T. J. Sejnowski, and H. Hughes, "Learning overcomplete representations," *Neural Computation*, vol. 12, pp. 337–365, 1998.
- [3] J. A. Tropp, "Topics in sparse approximation," Ph.D. dissertation, The University of Texas at Austin, August 2004.
- [4] W. Dai, T. Xu, and W. Wang, "Simultaneous codeword optimization (simco) for dictionary update and learning," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6340–6353, Dec 2012.
- [5] S. Mallat, *A wavelet tour of signal processing*, 3rd ed. Academic Press, 2009.
- [6] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via  $l_1$  minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [7] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *Information Theory, IEEE Transactions on*, vol. 50, no. 6, pp. 1341–1344, June 2004.
- [8] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, 1993.
- [9] Y. C. Pati, R. Rezaifar, Y. C. P. R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, 1993, pp. 40–44.
- [10] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *Signal Processing Letters, IEEE*, vol. 9, no. 4, pp. 137–140, April 2002.
- [11] M. Andrieu, L. Rebollo-Neira, and E. Sagianos, "Backward-optimized orthogonal matching pursuit approach," *Signal Processing Letters, IEEE*, vol. 11, no. 9, pp. 705–708, Sept 2004.
- [12] V. V. Muniswamy, *Design and Analysis of Algorithms*. New Delhi, India, India: I K International Publishing House, 2009.
- [13] B. Mailh, R. Gribonval, P. Vandergheynst, and F. Bimbot, "Fast orthogonal sparse approximation algorithms over local dictionaries," *Signal Processing*, vol. 91, no. 12, pp. 2822 – 2835, 2011.
- [14] T. Blumensath and M. Davies, "Gradient pursuits," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2370–2382, June 2008.
- [15] H. Anton and C. Rorres, *Elementary Linear Algebra: Applications Version*. John Wiley & Sons, 2010.
- [16] J. Cooley and J. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.
- [17] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms I," *Comm. Pure Appl. Math.*, vol. 44, pp. 141–183, 1991.
- [18] P. Jost, P. Vandergheynst, and P. Frossard, "Tree-based pursuit: Algorithm and properties," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4685–4697, Dec 2006.
- [19] S. Cotter, R. Adler, R. Rao, and K. Kreutz-Delgado, "Forward sequential algorithms for best basis selection," *Vision, Image and Signal Processing, IEE Proceedings*, vol. 146, no. 5, pp. 235–244, Oct 1999.
- [20] S. Mallat and Z. Zhang, "Adaptive time-frequency decomposition with matching pursuits," in *Time-Frequency and Time-Scale Analysis, 1992., Proceedings of the IEEE-SP International Symposium*, Oct 1992, pp. 7–10.

- [21] T. Blumensath and M. Davies, "Gradient pursuits," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2370–2382, June 2008.
- [22] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*. Upper Saddle River, NJ: Prentice Hall, 2000. [Online]. Available: <http://opac.inria.fr/record=b1096461>
- [23] K. T. Wallenius, "Biased sampling: The non-central hypergeometric probability distribution," Ph.D. dissertation, Stanford University, Department of Statistics, 1963.
- [24] A. Fog, "Sampling methods for wallenius' and fisher's noncentral hypergeometric distributions," *Communications in Statistics - Simulation and Computation*, vol. 37, no. 2, pp. 241–257, 2008.
- [25] Y. M. Cole, R. A. and M. A. Fauty, "The isolet spoken letter database," Computer Science Department, Oregon Graduate Institute, Tech. Rep., 1990.
- [26] N. S. Jayant, "Subsampling of a DPCM speech channel to provide two 'self-contained' half-rate channels," vol. 60, no. 4, pp. 501–509, Apr. 1981.

### A Proof of $\mathbb{E}_{C(\psi)}$ Levels

For  $\omega = 1$ , we can rewrite (13) as follows

$$\begin{aligned} \frac{\mathbb{E}_{C(\psi)}}{k_j} &= \frac{k_{j+\psi} - \mathbb{E}_{C(\psi)}}{N - k_j} \Rightarrow \mathbb{E}_{C(\psi)} \left( 1 + \frac{k_j}{N - k_j} \right) = \frac{k_j k_{j+\psi}}{N - k_j} \\ \therefore \mathbb{E}_{C(\psi)} &= \frac{k_j k_{j+\psi}}{N} \end{aligned} \quad (16)$$

For  $\omega > 1$ , and  $(N - k_j) > (k_{j+\psi} - \mathbb{E}_{C(\psi)})$ , we have

$$\left( 1 - \frac{k_{j+\psi} - \mathbb{E}_{C(\psi)}}{N - k_j} \right)^\omega < \left( 1 - \frac{k_{j+\psi} - \mathbb{E}_{C(\psi)}}{N - k_j} \right) \quad (17)$$

The LHS in (17) equals  $1 - \frac{\mathbb{E}_{C(\psi)}}{k_j}$ . So,  $\mathbb{E}_{C(\psi)} > \frac{k_j k_{j+\psi}}{N}$ . As for  $\omega < 1$ , and  $(N - k_j) > (k_{j+\psi} - \mathbb{E}_{C(\psi)})$ , we have  $\mathbb{E}_{C(\psi)} < \frac{k_j k_{j+\psi}}{N}$

### B Atoms Reuse And $G_{MP}$ Enhancement

Let  $\Phi_1^{M \times k} = \{\varphi_n; n \in \Gamma^{(j)}\}$  denotes the first search space for  $Y_{j+1}$  decomposition and the MP algorithm will select  $\delta$  atoms from that space. Also, let  $\Phi_2^{M \times N} = \{\varphi_n; n \in \Gamma\}$  denotes the second search space required to get the remainder  $k - \delta$  atoms of  $Y_{j+1}$  decomposition. Recall (4) and let  $F = 1$  then function  $G$  can be written as follows:

$$G_{MP} = \sum_{i=1}^{\delta} \left[ 2M(k+1) \right] + \sum_{i=1}^{k-\delta} \left[ 2M(N+1) \right] \quad (18)$$

So,  $G_{MP} = 2kM(\delta + N) + 2M(k - N\delta)$ . Assume  $G_e$  denotes the enhancement in complexity, then  $G_e$  can be written as the following ratio:

$$G_e = \frac{G(M, N, k, \delta = 0) - G(M, N, k, \delta)}{G(M, N, k, \delta = 0)} \quad (19)$$

By substitution, we obtain  $G_e = \frac{\left(\frac{\delta}{k}\right)\left(1 - \frac{k}{N}\right)}{\left(1 + \frac{1}{N}\right)}$ , and for large  $N$ , we obtain  $G_e \approx \frac{\delta}{k} \left(1 - \frac{k}{N}\right)$ .

### C Atoms Reuse And $G_{OMP}$ Enhancement

Let  $\Phi_1^{M \times k} = \{\varphi_n; n \in \Gamma^{(j)}\}$  denotes the first search space for  $Y_{j+1}$  decomposition and the OMP algorithm will select  $\delta$  atoms from that space. Also, let  $\Phi_2^{M \times (N-\delta)} = \{\varphi_n; n \in \Gamma \setminus \Gamma^{(j)}\}$ ,

where  $\Gamma_{(\delta)}$  refers to the indices of the  $\delta$  atoms, denotes the second search space required to get the remainder  $k - \delta$  atoms of  $Y_{j+1}$  decomposition. Unlike MP algorithm, forcing  $\delta$  atoms in OMP compression affects only on the atom selection procedure. Assume that  $G_{OMP}^{(AP)}$  denotes the complexity of **A**tom selection **P**rocedure, and  $G_{OMP}^{(OP)}$  denotes the complexity of **O**ther **P**rocedures. So, we can rewrite (19) for OMP as follows

$$G_e = \frac{G_{OMP}^{(AP)}(\delta = 0) - G_{OMP}^{(AP)}(\delta \neq 0)}{G_{OMP}^{(AP)}(\delta = 0) + G_{OMP}^{(OP)}(\delta = 0)} \quad (20)$$

The denominator in (20) represents the overall complexity of the OMP that is stated in (6). The value of  $G_{OMP}^{(AP)}(\delta)$  can be obtained from the following summation

$$\begin{aligned} G_{OMP}^{(AP)}(\delta) &= 2M \left[ \sum_{i=1}^{\delta} (k - i + 1) + \sum_{i=\delta+1}^k ((N - \delta) - i + 1) \right] \\ G_{OMP}^{(AP)}(\delta) &= 2M \left[ \delta^2 + N(k - \delta) - \frac{k(k-1)}{2} \right] \end{aligned} \quad (21)$$

For  $\delta = 0$ , we have

$$G_{OMP}^{(AP)}(\delta = 0) = 2M \left[ Nk - \frac{k(k-1)}{2} \right] \quad (22)$$

By substituting (21), (22) and (6) in (20), we obtain  $G_e = \frac{\left(\frac{\delta}{k}\right)\left(1 - \frac{\delta}{N}\right)}{\left(1 + \frac{M}{N}\right)}$ .

#### D Active Cluster and $G_{MP}$ Enhancement

Let  $F' = \alpha F$  denotes the required frames that should be compressed to detect the active atoms  $N'$ , where  $\alpha$  represents the ratio  $\frac{F'}{F}$ . So, The MP algorithm will use the space  $\Phi$  for compressing the frames  $Y_1$  till  $Y_{F'}$ . But, for the other frames  $Y_{F'+1}$  till  $Y_F$ , it will use the shrink space  $\Phi'$ . In this case, we can rewrite (5) as follows

$$\begin{aligned} G_{MP} &= 2kM \left[ \sum_{j=1}^{F'} N \left(1 + \frac{1}{N}\right) + \sum_{j=F'+1}^F N' \left(1 + \frac{1}{N'}\right) \right] \\ G_{MP} &= 2FkMN \left[ 1 + \frac{1}{N} \right] \left[ \alpha + \lambda_1 \beta (1 - \alpha) \right] \end{aligned} \quad (23)$$

where  $\beta = \frac{N'}{N}$ ,  $\lambda_1 = \frac{1+1/N'}{1+1/N}$ . The enhancement in  $G_{MP}$  can be obtained as follows

$$G_e = \frac{G_{MP}(\alpha = 1) - G_{MP}(0 \leq \alpha < 1)}{G_{MP}(\alpha = 1)} = (1 - \alpha)(1 - \lambda_1 \beta) \quad (24)$$

#### E Active Cluster and $G_{OMP}$ Enhancement

Like the MP algorithm, the OMP will use the space  $\Phi$  for compressing the frames  $Y_1$  till  $Y_{F'}$ . But, for the other frames  $Y_{F'+1}$  till  $Y_F$ , it will use the shrink space  $\Phi'$ . In this case, we can rewrite (6) as follows

$$G_{OMP} = 2kM \left[ \sum_{j=1}^{F'} N \left(1 + \frac{M}{N}\right) + \sum_{j=F'+1}^F N' \left(1 + \frac{M}{N'}\right) \right] = 2FkMN \left(1 + \frac{M}{N}\right) \left[ \alpha + \lambda_2 \beta (1 - \alpha) \right] \quad (25)$$

where  $\lambda_2 = \frac{1+M/N'}{1+M/N}$ . The enhancement in  $G_{OMP}$  can be obtained as follows

$$G_e = \frac{G_{OMP}(\alpha = 1) - G_{OMP}(0 \leq \alpha < 1)}{G_{OMP}(\alpha = 1)} = (1 - \alpha)(1 - \lambda_2\beta) \quad (26)$$