16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SUBSET SELECTION FROM BIASED DICTIONARIES FOR IMPACT ACOUSTIC CLASSIFICATION

Nuri Firat Ince, Fikri Goksu, and Ahmed H. Tewfik

Department of Electrical and Computer Engineering, University of Minnesota
200 Union St. SE, 55455, Minneapolis, USA
phone: + (1) 612 625 6024, fax: + (1) 612 6254583, email: firat@umn.edu

ABSTRACT

In this paper we study a sparse signal representation approach for the classification of impact acoustic signals obtained from empty and full hazelnuts. In particular, two custom dictionaries are designed for each class with a vector quantization algorithm by using the training data for each. In the following step each individual dictionary or their combination is used for representing the test acoustic signals with the belief that the representation will be biased. Two different subset selection (SS) techniques, matching pursuit (MP) and a bounded error subset selection algorithm (BESS) were investigated to approximate given signals by using the code vectors in these biased dictionaries. The approximation error and the number of code vectors selected from dictionaries were used as input features for the final classification step. We observe that this biased dictionary design allows one to distinguish between classes while representing them with a small number of codevectors. In particular the classification performance of the BESS algorithm by using the codevectors constructed from empty hazelnut acoustics was the best and outperforms that of using the MP algorithm. The combination of BESS algorithm, the dictionary derived from empty hazelnut acoustics and a decision tree yields classification accuracies of 94.8% and 100% for empty and full hazelnuts respectively. Our results indicate that subset selection from biased dictionaries can be used as a new approach for classification.

1. GENERAL INFORMATION

In the last few decades there is a growing interest in the use of adaptive signal representation techniques for compression and communication [1, 2 and 3]. The main idea behind this trend is, for a given signal $\boldsymbol{b}$ of length N, one wants to represent the signal within a tolerable error limit, ε_{th}, by using a small number of coefficients k, where $k<<N$. Recent research has shown that this goal can be achieved in a more effective manner when redundant dictionaries are preferred to complete bases such as Fourier and Cosine [1, 2, 3 and 4]. In this scheme wavelet and cosine packets are very effective examples [1, 2 and 3]. Both WP and CP first represent the signal using an overcomplete dictionary over a pyramidal tree structure. Within this pyramid structured dictionary, a mother node covers a space which is the union of its children. For a sparse and complete representation, this relationship allows one to prune the tree from bottom to top by so called best basis algorithm which implements a cost function minimization procedure, such as entropy over the expansion coefficients in the nodes of the pyramidal tree [1]. The richness and structural relationship in the dictionary design and the selection of signal adapted bases from this dictionary resulted in better compression rates. Although the structural relationship in dictionary design allows fast algorithms to be implemented for subset selection, it is more desirable to use more sophisticated and flexible dictionaries to represent the signal without any geometrical limitation. However, finding the optimal approximation to the signal $\boldsymbol{b}$, in an overcomplete dictionary represented by $\boldsymbol{N}$ dimensional vectors spanning the column space of the matrix $\boldsymbol{D}$ is ***NP*** hard. ($\boldsymbol{D} \in \Re^{(NxM)}$, and $M>>N$) [2, 4 and 5].

In order to tackle this problem, so called matching pursuit (MP), an iterative subset selection procedure, was proposed in [2, 4]. The MP uses a greedy search strategy to represent the signal b, using an overcomplete dictionary by minimizing the approximation error $\boldsymbol{e}$ in each step. Although the solution is not optimal, MP provides satisfactory results most of the time. Several other strategies including the basis pursuit and the orthogonal MP were proposed with an additional cost. The reader is referred to [2, 4 and 5] for further information. The recent progress in sparse representation algorithms shows that the subset selection procedure and dictionary design are key elements in signal approximation [5, 6, 7 and 8].

In this paper we use the achievements in this area for classification purposes. In particular, we use the flexibility in dictionary design and efficiency in subset selection for the recognition of impact acoustic waveforms obtained from full and empty hazelnuts. As a first step, we design custom dictionaries with a vector quantization algorithm where each dictionary is specifically generated from the available acoustic training data in each class. In the following step, we use a subset selection algorithm to approximate the acoustic signals by forcing the algorithms to use the codevectors in these biased dictionaries. The subset selection module returns an approximation error and the number of codevectors used to achieve that error selected from a partial dictionary. To evaluate the efficiency of subset selection module here we consider matching pursuit and bounded error subset selection algorithms. As a last step, we feed these features to a classifier module for the final decision. The classifiers con-

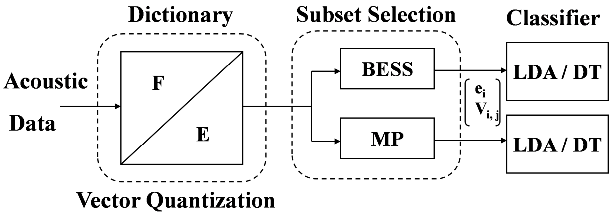


Fig.1. The block diagram of the proposed system.

sidered in this work are a linear discriminant analysis and a decision tree. A schematic diagram of the proposed approach is presented in Fig.1. In the rest of this paper we will investigate the effect of:

- Dictionary design in final accuracy.
- Subset selection approach in final accuracy.
- Different classifiers such as linear discriminant and a decision tree.

This paper is organized as follows. In the next section we explain the dictionary design and the BESS procedure in detail. In section 3 we describe the acoustic data acquisition system and provide sample waveforms. In section 4 we provide experimental results. We conclude with discussion and future work in section 5.

2. DICTIONARY DESIGN AND SUBSET SELECTION

As described above we are interested in developing a signal representation approach that is dictionary biased. In this scheme using the deterministic dictionaries constructed from such as sine, cosine or Gabor functions are not suitable. For this reason, we utilize the LBG- vector quantization algorithm of [9] for dictionary generation. Here, we organize our dictionary $\boldsymbol{D}$ in such a way that the selection of the codevectors, ψ_i, will be biased within $\boldsymbol{D}$. In particular, two sub-dictionaries $\boldsymbol{D_{Empty}}$ for empty and $\boldsymbol{D_{Full}}$ for full hazelnut classes were constructed by using half of the available dataset. These dictionaries were estimated from the acoustic data with the LBG-Vector Quantization algorithm [6]. In the next step they are used by the subset selection algorithms for the representation of the impact acoustic signal of empty and full hazelnuts. In addition we also investigate the use of a combined dictionary $\boldsymbol{D_{Combined}}$ computed by merging empty and full sub dictionaries**,**

$$D_{Combined} = D_{Empty} \cup D_{Full} \tag{1}$$

Then, a given test signal that is not used in the dictionary generation, is approximated by one of the subset selection algorithms which will be described in the following section.

2.1 Matching Pursuit for Subset Selection

In our solution strategy, selection of codevectors from the biased dictionaries in a sparse manner is expected to emphasize the differences between classes. In this respect we evaluate two different subset selection procedures to observe their effect in classification. It is known that the subset selection problem is NP-hard [2, 4]. Sparseness is imposed explicitly by minimizing the number of non-zero coefficients in the solution vector. One of the solutions to the problem is the Matching Pursuit algorithm that uses a greedy search strategy [4]. We will use the MP as a subset selection approach. The pseudo code of MP is summarized in Box-1. Here f is the residual, ψ_i is the i^{th} column vector (code vector) of $\boldsymbol{D}$, and x_i is the inner product between the input signal and ψ_i. The signal is iteratively decorrelated from the basis vector which has maximum correlation (x_k) with the residual. This greedy approximation strategy offers a solution in a polynomial time in the presence of an overcomplete dictionary.

As an alternative method we also considered the bounded error subset selection (BESS) algorithm of [6, 7], which has higher computational complexity but provides a more optimal solution.

> **<u>Box-1: MP Algorithm</u>**
>
> **start with** $f = b$
> **repeat**
> **for** i=1 to N
> $x_i = \langle \psi_i, f \rangle$.
> **end**
> $k = \arg\max_i (x)$
> $f = f - \psi_k x_k \,,\; e = \dfrac{\|f\|_2^2}{\|b\|_2^2}$
> **until** $e \le \varepsilon_{th}$ **then**
> **return** e

2.2 Bounded Error Subset Selection (BESS)

The BESS algorithm has been introduced by the authors of [6, 7] as a reformulation of the classical subset selection problem. It has been shown that by introducing a perturbation vector $\vec{\varepsilon}$ to the signal $\boldsymbol{b}$ under investigation, one can obtain a maximally sparse representation of the signal from the over complete dictionary $\boldsymbol{D}$**,** such that $\|Dx - b\| \le \varepsilon_{th}$. The solution for the BESS problem can be achieved via exhaustive enumeration procedure with an exponential complexity. By utilizing a trimming procedure the authors have reduced the computational complexity to polynomial time. Furthermore, by utilizing a stack decoding algorithm, the memory requirements of the algorithm is reduced. In BESS, the sparseness is achieved by keeping other alternative approximations to the signal in a tree based search structure. In particular the algorithm keeps most promising approximations in each step and removes remaining ones for memory savings and to decrease the computational cost. Furthermore, BESS uses the Gramm-Schmidt orthogonalization procedure in each step to decorrelate the codevectors in the

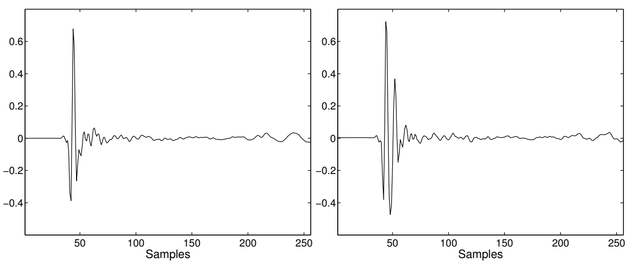
<u>Box:2- Pseudo Code of BESS</u>

Step-1: Set the number of alternative approximations, *k*.
Step-2: Compute the inner product of the signal with all codevectors in the dictionary
Step-3: Select the codevector from ***D*** with maximum absolute correlation.
Step-4: Remove that index from the dictionary ***D*** and find the second alternative codevector with max coefficient. Go to *Step-2* until desired number of alternatives found.
Step-5: For each alternative generate new dictionaries by orthogonalizing the codevectors with respect to the selected codevector.
Step-6: For each alternative codevector compute the residual and approximate it with the rest of the codevectors in the dictionary.
Step-7: List, $L_{i,}$ all subsets of dictionary vectors to produce approximations. Keep the best *k* subsets that provide lower approximation error.
Step-8: Goto *Step-2*, add new codevectors until the approximation error is below a given threshold.
Step-9: Return the best subset from L_i and the corresponding approximation.

dictionary with respect to the selected index. Therefore the BESS can be seen as an extended version of orthogonal least squares (OLS) solution [10]. We also note that the combination of trimming and stack decoding algorithms have an interesting connection to tree based A* (*A star*) and branch-bound search algorithms which are widely used in artificial intelligence since the 1960's [11, 12]. To be more specific the search strategy of BESS can be seen as a special case of memory bounded A* search. A pseudo code of BESS is given in Box 2. Now let us describe the hazelnut acoustic signals we recorded to test our algorithm.

3. HAZELNUT IMPACT ACOUSTIC RECORDINGS

Hazelnuts are widely used in the chocolate and flavoured coffee industries. Empty hazelnuts and hazelnuts containing undeveloped kernels are one of the main causes reducing the quality attribute. Moreover, empty hazelnuts and the ones containing undeveloped kernels may also contain the mould, "Asperguillus flavus" that produces aflatoxin, a cancer causing material [13]. Therefore, separation of empty and undeveloped hazelnuts from developed ones carries significance for food quality and human health. Currently, empty hazelnuts are separated by pneumatic devices that use weight differences with an airleg. However, the false positive rate of these devices is high. There remains a need for more advanced systems to improve the classification accuracies.



(a) (b)

Fig.2. Empty (a) and full (b) hazelnut acoustics.

Table 1. Classification accuracies (%) obtained with $\boldsymbol{D_{Empty}}$ dictionary which is constructed from the impact acoustic signals of empty hazelnuts. The columns E and F stand for the classification accuracies of empty and full hazelnuts respectively.

LDA		**MP**			**BESS**	
$\boldsymbol{\varepsilon_{th}}$	**E**	**F**	**Avg.**	**E**	**F**	**Avg.**
0.1	97.2	91.1	94.1	97.8	88.3	93
0.05	95.5	98.3	96.9	98.2	90.2	94.2
0.01	97	96.5	96.7	94.6	99	96.8

DT		**MP**			**BESS**	
$\boldsymbol{\varepsilon_{th}}$	**E**	**F**	**Avg.**	**E**	**F**	**Avg.**
0.1	92.4	98.9	95.7	**94.8**	**100**	**97.4**
0.05	92.2	98.9	95.6	94.8	99.8	97.3
0.01	92	99.6	95.8	94.1	99.1	96.6

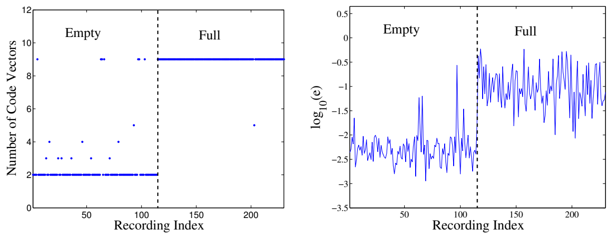
Recently in another area, a high-throughput, low cost acoustical system has been developed to separate pistachio nuts with closed shells from those with open shells [14]. In this system, pistachio nuts were dropped onto a steel plate and the sound of the impact is analyzed in real time. Pistachio nuts with closed shells produce a different sound than those with cracked shells, as expected. The classification accuracy of this system is approximately 97%, with a throughput rate of approximately 20-40 nuts/second.

In this study we also use impact acoustic signals of empty and full hazelnuts which are recorded by dropping nut kernels onto a steel plate. The impact plate was a polished block of stainless steel with dimensions 7.5 x 15 cm and a depth of 2 cm. A microphone, that is sensitive to frequencies up to 20 KHz, was used to capture the impact sounds. The sound card in a typical personal computer was used to digitize acoustic signals with 44.1 kHz. For each type of hazelnut, 230 recordings were obtained. Figure 2 shows two representative records from the available dataset.

4. RESULTS

In our experimental studies, we used 2 times 2 fold cross validation to estimate the classification error. Half of the dataset is used to calculate 32 code-vectors for each subdictionary via LBG-VQ [9] algorithm for each class. All impact acoustics records were 256 samples long. For the BESS algorithm we keep 3 alternative code vectors in each step

In order to study the effect of the dictionary on final classification, we employ three different strategies. We separately tested individual dictionaries and their combination to be used by the subset selection algorithm. For each dictionary formulation, we use the MP and BESS for subset selection. For each signal the subset selection procedures returned e_i the approximation error and $V_{i,j}$ the number of codevectors selected from each dictionary where *i* keeps the index of the sample acoustics and *j* keeps the index of the dictionary.



(a) (b)

Fig.3. The number of code vectors selected from the dictionary $\boldsymbol{D}_{Empty}$ by the BESS algorithm for empty and full hazelnuts is given in (a). The approximation error, achieved with these code vectors is given in (b). Note that although a higher number of codevectors are selected for full hazelnuts their corresponding approximation error is higher than empty ones.

On the last step e_i and $V_{i,j}$ were used as input features to a linear discriminant analysis and decision tree for final classification. We also study the effect of the approximation threshold, $\varepsilon_{th} \in \{0.1, 0.05, 0.01\}$.

During our initial studies we observed that when the number of codevectors to approximate the target signal is limited, then the classification accuracies were better than the unlimited case. This indicates that the approximation with the initially selected codevectors is more important for discrimination. Therefore the maximum number of codevectors that can be selected from the dictionary is limited to nine such that the bias is emphasized in signal approximation.

Table 1, 2 and 3 show the classification accuracies obtained with the proposed system for different ε_{th}, and sub-module designs.

We observed that the best classification accuracy is obtained from $\boldsymbol{D}_{Empty}$, BESS and decision tree combination for $\varepsilon_{th} = 0.1$ as shown in Table 1. The individual classification accuracies for empty and full hazelnuts in this setup are 94.8% and 100%, respectively. Fig. 3 shows the number of codevectors by BESS from the dictionary $\boldsymbol{D}_{Empty}$ for each class and related approximation error. When the subset selection module is replaced with the MP algorithm, the individual classification accuracies drop to 92% and 99.6% with $\varepsilon_{th} = 0.01$, respectively. Moreover, when the classifier is replaced with an LDA, interestingly the average classification accuracy drops to 96.8% for BESS and increases to 96.9% for MP setups. For the MP algorithm the best classification accuracy is obtained at $\varepsilon_{th} = 0.05$ level, whereas it is at $\varepsilon_{th} = 0.01$ level for the BESS algorithm.

When the dictionary is replaced with $\boldsymbol{D}_{Full}$, the average classification accuracies drop radically for all designs (See Table 2). The best result with this dictionary is 79.2% with BESS and LDA setup using an error threshold of $\varepsilon_{th} = 0.1$ among all designs.

For combined dictionary, the classification accuracies are comparable to those obtained with $\boldsymbol{D}_{Empty}$ as shown in Table 3. Once again, the best result is achieved when BESS is used

Table 2. Classification accuracies (%) obtained with $\boldsymbol{D}_{Full}$ dictionary which is constructed from the impact acoustic signals of full hazelnuts. The columns E and F stand for the classification accuracies of empty and full hazelnuts respectively.

LDA	**MP**			**BESS**		
$\boldsymbol{\varepsilon_{th}}$	**E**	**F**	**Avg.**	**E**	**F**	**Avg.**
0.1	75.3	48.1	61.7	**71.4**	**87**	**79.2**
0.05	50.5	48.7	49.6	79.6	77.4	78.5
0.01	63.5	43.3	53.4	52.9	88.3	70.6

DT	**MP**			**BESS**		
$\boldsymbol{\varepsilon_{th}}$	**E**	**F**	**Avg.**	**E**	**F**	**Avg.**
0.1	69.9	47.9	58.1	64.6	69.1	66.8
0.05	67.6	49.3	58.5	65.7	74	69.8
0.01	66.8	54.6	60.7	55.2	71	63.0

as the subset selection algorithm. In particular, the BESS and LDA combination achieved an average classification accuracy of 97.1% at $\varepsilon_{th} = 0.01$ level. For the MP, the average classification accuracy at this level is 90.4%.

In order to assess the overall efficiency of the proposed approach, we compare it to a baseline design. In particular, we extract time and frequency domain features and use them for classification. The used features are the absolute values of the discrete Fourier transform of the entire signal, the maximum and minimum values and the standard deviation of the raw data. The classification accuracies with these features and LDA combination were 98.5% and 83.1% for empty and full hazelnuts respectively. Our proposed approach not only outperformed in average classification accuracy but also provided higher true positive rates for full hazelnuts.

5. CONCLUSION

In this paper we proposed a subset selection approach from class dependent dictionaries for impact acoustic classification. In this scheme, two different subset selection methods such as MP and BESS are evaluated. Furthermore, we inspected the performance of the proposed system for different classifiers and approximation error thresholds. We noticed that the BESS, and the dictionary constructed from the acoustic waveforms of empty hazelnuts, and a decision tree provided the best average classification accuracy of 97.4% with only two features. Interestingly, the results obtained with the dictionary constructed from full hazelnut acoustic

Table 3. Classification accuracies obtained with combined dictionary.

LDA	**MP**			**BESS**		
$\boldsymbol{\varepsilon_{th}}$	**E**	**F**	**Avg**	**E**	**F**	**Avg**
0.1	97.8	84.6	91.2	95.2	96.5	95.8
0.05	97.2	86.8	92	94.8	98.7	96.7
0.01	87.9	92.9	90.4	**95.3**	**99**	**97.1**

DT	**MP**			**BESS**		
$\boldsymbol{\varepsilon_{th}}$	**E**	**F**	**Avg**	**E**	**F**	**Avg**
0.1	93.1	98.9	96	89.3	100	94.6
0.05	93.5	97.4	95.4	89.3	100	94.6
0.01	82.4	95	88.7	93.7	100	96.8

waveforms were radically poor. When two sub dictionaries are combined comparable results were obtained at expense of higher computational complexity.

During our experiments we observed that the signals from both classes were fairly represented by the codevectors in $\boldsymbol{D}_{Full}$ dictionary. However when we use $\boldsymbol{D}_{Empty}$ dictionary we note that the full hazelnut acoustics could not be represented efficiently by neither MP nor BESS algorithms (See Fig. 3(b)). We observed that not only the dictionary design but also the subset selection procedure has an important role on the final classification accuracy. As we expected the sparsest selection produced better classification accuracy which is validated by the results with the BESS algorithm. We believe the reason for this is that the BESS algorithm emphasizes the bias in dictionaries even more while implementing a computationally complex search within the dictionary. However one should note that the BESS procedure has much higher computational complexity with respect to MP and may provide slower rate processing speed in real-time applications.

We utilized a vector quantization algorithm for dictionary design. This approach is different from the general approach where the dictionaries are fixed in advance. A similar approach where the initial dictionary is iteratively updated using training signals was developed in the K-SVD.algorithm, [8]. Currently, the authors are investigating such custom dictionary designs and their effects on the classification performance.

6. ACKNOWLEDGEMENT

The authors are grateful to Enis Cetin and Ibrahim Onaran for providing the impact acoustics data.

REFERENCES

[1] K. Ramchandran, M. Vetterli and C. Herley, *Wavelets, subband coding, and best bases*, IEEE Proceedings, Special Issue on Wavelets, Vol. 84, Nr. 4, pp. 541-560, 1996.

[2] S. Mallat, “A Wavelet Tour of Signal Processing”, 2nd edn (New York: Academic), 2000.

[3] P.L. Dragotti and M. Vetterli, “Wavelet Footprints: Theory, Algorithms and Applications”, *IEEE Transactions on Signal Processing*, Vol. 51, Nr. 5, pp. 1306-1323, 2003.

[4] S. Mallat, Z. Zhang: “Matching pursuit with time-frequency dictionaries”. *IEEE Transactions on Signal Processing*, **41:**3397-3415, 1993.

[5] SS. Chen, DL. Donoho, Saunders MA, “Atomic decomposition by basis pursuit”. *SIAM Review*, 43(1):129-159, 2001.

[6] M. Alghoniemy and A. H. Tewfik, “A Sparse Solution to the Bounded Subset Selection Problem: A Network Flow Model Approach,” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, pp. 89-92, May 2004.

[7] M. Alghoniemy and Ahmed H. Tewfik, ”Bounded Subset Selection with Noninteger Coefficients,” *Proceedings of the Eusipco-2004 Conference*., pp. 317-320, Sept. 2004.

[8] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," IEEE Tans. Signal Processing, Vol. 54, No. 11, Nov 2006.

[9] Y. Linde, A. Buzo, and R. M. Gray, “An Algorithm for Vector Quantizer Design,*" IEEE Transactions on Communications*, pp. 702-710, January 1980.

[10] Chen, S., Cowan, C.F.N., and Grant, P.M., 1991, "Orthogonal least squares learning for radial basis function networks," IEEE Transactions on Neural Networks, 2, 302-309.

[11] S. J. Russell, P. Norvig, “Artificial Intelligence: A Modern Approach”, pp. 97-104. 2003.

[12] P. Narendra, K. Fukunaga, “A branch and bound algorithm for feature subset selection”. *IEEE Trans. Comput*. C-26, 917-922, 1977.

[13] I. Marklinder, M. Lindblad, A. Gidlund, M. Olsen. “Consumers’ ability to discriminate aflatoxin-contaminated Brazil nuts”, *Food Additives and Contaminants*, 22(1): 56–64, January 2005.

[14] T.C. Pearson, , Detection of pistachio nuts with closed shells using impact acoustics. Applied Engineering in Agriculture, 17(2):249-253, 2001.