

A SIGNAL REPRESENTATION APPROACH FOR DISCRIMINATION BETWEEN FULL AND EMPTY HAZELNUTS

Ibrahim Onaran¹, Nuri F. Ince², Ahmed H. Tewfik², A. Enis Cetin¹

Department of Electrical and Electronics Engineering, University of Bilkent, Turkey¹
Department of Electrical and Computer Engineering, University of Minnesota, USA²

ABSTRACT

We apply a sparse signal representation approach to impact acoustic signals to discriminate between empty and full hazelnuts. The impact acoustic signals are recorded by dropping the hazelnut shells on a metal plate. The impact signal is then approximated within a given error limit by choosing codevectors from a special dictionary. This dictionary was generated from sub-dictionaries that are individually generated for the impact signals corresponding to empty and full hazelnut. The number of codevectors selected from each sub-dictionary and the approximation error within initial codevectors are used as classification features and fed to a Linear Discriminant Analysis (LDA). We also compare this algorithm with a baseline approach. This baseline approach uses features which describe the time and frequency characteristics of the given signal that were previously used for empty and full hazelnut separation. Classification accuracies of 98.3% and 96.8% were achieved by the proposed approach and base algorithm respectively. The results we obtained show that sparse signal representation strategy can be used as an alternative classification method for undeveloped hazelnut separation with higher accuracies.

1. INTRODUCTION

Hazelnuts (*Corylus avellana*) are one of the main ingredients used in the chocolate and flavored coffee industries. One of the main quality attributes of raw bulk hazelnuts is the ratio of kernel weight to shell weight. Empty hazelnuts and hazelnuts containing undeveloped kernels negatively affect this ratio. If the ratio of kernel weight to gross weight is less than 0.5 then some buyers reject the produce. Occasionally, a physiological disorder such as plant stress from dehydration or lack of nutrients causes a hazelnut shell to develop without a kernel. A nut with undeveloped kernel appears like a normal hazelnut from outside. Currently, raw hazelnuts are processed by an "airleg" which is a pneumatic device to separate empty hazelnuts from fully developed ones. However, these devices have high classification error rates. There remains a need for more advanced systems to improve upon the segregation of empty and full hazelnuts. In addition, empty

hazelnuts and hazelnuts containing undeveloped kernels may also contain the mold *Aspergillus flavus* that produces aflatoxin, a cancer causing material [1]. Therefore, a more accurate classification of hazelnuts will enhance food safety.

In order to detect fully developed hazelnuts from empty hazelnuts, one can weigh them one by one or shell them. Obviously, this is not an economically viable practice. Recently in another area, a high-throughput, low cost acoustical system has been developed to separate pistachio nuts with closed shells from those with cracked shells [2-4]. In this system, pistachio nuts were dropped onto a steel plate and the sound of the impact 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. It works reliably in a food processing environment with little maintenance or skill required to operate. A similar prototype system based on impact acoustics is extended to hazelnuts in [5]. It was experimentally observed that the algorithms described in [2-4] did not produce high classification rates in empty-full hazelnut separation. Recently the algorithm introduced in [5] combined Line Spectral Frequencies (LSFs); discrete Fourier transforms (DFT) and some time domain feature parameters for accurate classification of full and empty hazelnuts. However, this algorithm uses a large number of features to achieve a lower error rate. Also, the time domain modeling method which is used by the algorithm has high computational complexity.

The purpose of this study is to explore the effectiveness of a sparse signal representation approach, based on the Bounded Error Subset selection (BESS) algorithm for the classification of empty and full hazelnuts. In particular, we are interested in the classification accuracy of the method and the number of features selected by it. The BESS algorithm aims to describe the signal with the minimal number of codevectors selected from a dictionary which is specifically designed for classification.

The paper is organized as follows in the next section we describe the data acquisition system to record the impact acoustics signals. In sections 3 and 4, we describe the BESS and base algorithms used in this area for classification.

Finally, in section 4 we give experimental results to show the efficiency of the proposed approach.

2. MATERIALS AND METHODS

In order to inspect nuts at high throughput rates, a prototype system was set up to drop nuts onto a steel plate and process the resulting impact acoustic signal.

An experimental apparatus was fabricated to slide hazelnuts down a chute and project them onto an impact plate, then collecting the acoustic emissions from the impact. The impact plate was a polished block of stainless steel with dimensions 7.5 x 15 cm and a depth of 2 cm. The mass of the impact plate was chosen to be much larger than the hazelnuts in order to minimize vibrations from the plate interfering with acoustic emissions from kernels. A microphone, that is sensitive to frequencies up to 20 KHz, was used to capture impact sounds. The sound card in a typical personal computer was used to digitize and store the microphone signals for analysis. For each type of hazelnut 230 recordings were obtained. Figure 1 shows two representative records and available dataset.

2.1 Bounded Error Subset Selection (BESS)

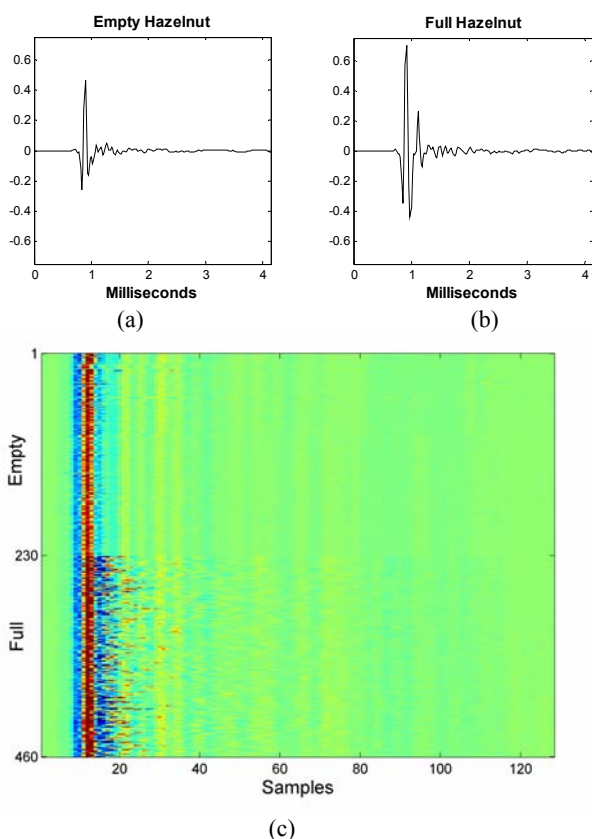


Figure 1 – (a) Sample hazelnut sound from each class. (b) Empty (top 230 rows) and full (bottom 230 rows) hazelnut records. Each row represents a record.

Sparse signal representations find applications in many signal processing areas such as coding, signal restoration, direction finding, source localization, and linear inverse problems, to name a few. In the subset selection (SS) problem, it is required to find the best signal representation for a signal vector \mathbf{b} using an over complete dictionary represented by N -dimensional vectors spanning the column space of matrix \mathbf{A} . By construction, the number of vectors M in the dictionary is such that the matrix \mathbf{A} has a dimension of $M \times N$. Thus, it is required to find the sparsest vector \mathbf{x} (with the minimum number of non-zero solution) such that $\mathbf{Ax} = \mathbf{b}$. It is known that the SS is NP-hard [6]. Sparseness is imposed explicitly by minimizing the number of non-zero coefficients in the solution vector. The Bounded Error Subset Selection (BESS) has been introduced by the authors of [7, 8] as a reformulation of the classical subset selection problem. It has been shown that by introducing a perturbation vector $\vec{\epsilon}$ to the signal \mathbf{b} , under investigation, such that $\|\mathbf{Ax} - \mathbf{b}\| \leq \epsilon$, one can obtain a maximally sparse representation of the signal from the over complete dictionary \mathbf{A} . Rather than using a greedy approximation as in Matching Pursuit algorithm, in BESS, the sparseness is achieved by keeping other alternative approximations to the signal. A pseudo code of BESS is given in Box.1

However, in this study we reformulate the SS problem for classification. In particular we use BESS algorithm to represent a given signal with minimum number of code vectors which are selected from a dictionary that is specifically constructed for classification.

Let us here briefly explain our strategy for constructing a discriminant dictionary. From a signal representation perspective the sparse signal representation provides a solution vector \mathbf{x} which gives good compression. Here, we organize our dictionary \mathbf{A} in such a way where the selection of \mathbf{x} is biased within \mathbf{A} . In particular, we generate \mathbf{A} from

Box:1- Pseudo Code of BESS

Step-1: Set the number of alternative approximations, k .

Step-2: Select the best codevector from \mathbf{A} to represent the signal

Step-3: Remove that index from the dictionary \mathbf{A} and find the best alternative codevector to represent the signal. Go to *Step-2* until desired number of alternatives found.

Step-4: For each alternative codevector index find the best combination from the dictionary.

Step-5: List, L_i , all subsets of dictionary vectors to produce approximations. Keep the best k subsets that provide lower approximation error.

Step-6: Goto *Step-2*, add new codevectors until the approximation error is above a given threshold.

Step-7: Return the best subset from L_i and corresponding of approximation.

two different sub-dictionaries which are individually formed from signals of each class. By using half of the available dataset, we constructed the sub-dictionaries D_E for empty and D_F for full hazelnut classes respectively. These dictionaries were estimated from the training data with the LBG-Vector Quantization algorithm [9]. These sub-dictionaries from empty and full classes are merged in \mathbf{A} ,

$$A = D_E \cup D_F \quad (1)$$

to form a united dictionary. Then, a given test signal that is not used in the dictionary generation, is represented by BESS by using the code-vectors in \mathbf{A} .

Let s_i be the test signal. For each signal, we calculate the number of code vectors, V_{ij} in \mathbf{x}_i which came from D_E and D_F . We have

$$s_i = Ax_i \quad (2)$$

$$V_{i,j} = \sum_{x_i \in D_j} (x_i \in D_j), \quad (3)$$

where $V_{i,E}$ and $V_{i,F}$ are the number of codevectors selected from sub dictionaries D_E and D_F in \mathbf{A} while representing the signal s_i . Here we expect that the BESS will be biased towards a subdictionary while selecting the codevectors from the union dictionary. This bias will be represented by $V_{i,E}-V_{i,F}$. In our experimental studies we observed that the approximation error during the selection of the initial code vectors is different for empty and full class signals. In particular, the approximation error for full hazelnut acoustics was larger than for the empty class. We used this biased behavior as another feature. Since the number of codevectors selected by BESS is signal dependant we calculated the approximation error, ε_i , using the first best 8 code vectors. Finally we input this three dimensional feature vector, which was formed by $V_{i,j}$ and ε_i to a LDA for classification.

2.2. Sound Processing

In order to compare the efficiency of BESS algorithm we implemented the base line approach introduced in [2]. This approach uses time domain modeling, short time windows extremum and variance estimation and spectral analysis of the signal.

2.2.1 Weibull Parameters

The Weibull function is used for modeling the many widely used curves such as Gaussian, log-normal curve and exponential decay. It has the following form

$$y(t) = \frac{cb}{a} \left(\frac{t-t_0}{a} \right)^{b-1} e^{-\left(\frac{t-t_0}{a}\right)^b} \text{ for } t < t_0 \quad (4)$$

otherwise $y(t) = 0$

The hazelnut sound signal is first rectified by taking its absolute value. Next it is nonlinearly filtered to smooth the curve before the Weibull parameters a , b , c , t_0 and R^2 (mean-square error value) computation. Weibull equation parameters are estimated as in [10].

2.2.2 Short Time Variance Window Processing

In addition to the time domain processing by modeling the signals with a Weibull function, variances of the signals are also computed in short time windows. The Weibull function captures the shape of the recorded signal globally and the short-time variance information models the local time domain variations in the signal. The short time windows were 50 points in duration and incremented in steps of 30 points so that each window overlapped by 20 points. The first window began 50 points in front of the impact. A total of eight short time windows were computed to cover the entire duration of all signals. After all variances were computed, they were normalized by the sum of all eight variances as follows:

$$\sigma_m = \frac{\sigma_i}{\sum_k \sigma_k}, \quad (5)$$

where σ_m^2 and σ_i^2 are the normalized and computed variances from window i with $i=1$ being the first window and $i=8$ being the last.

2.2.3 Short Time Extrema

The first 165 samples from the 50th sample before the impact sound was divided into 11 non overlapping time domain windows and the extremum value of each window was selected as a feature value. Extrema in short-time windows also capture the envelope of the impact sound similar to the variances in short-time windows.

2.2.4 Frequency Domain Processing

A 256 point DFT was computed from each signal using a Hamming window. The 256-point window covers the impact sound of the hazelnut. The magnitude of each spectra was computed and then low pass filtered using a 20 tap Finite Impulse Response (FIR) filter was applied to remove jagged spikes in the spectra. The low pass filter has a cutoff frequency of $\pi/4$ in the normalized DFT domain. The frequency corresponding to the peak magnitude in the frequency spectra was saved as a potential discriminating feature. In addition, the 15 magnitude values before the

peak and 15 points after the peak were saved and normalized by the peak magnitude.

2.2.5 Line Spectral Frequencies

Linear predictive modeling techniques are widely used in various speech coding, synthesis and recognition applications [11]. Linear Minimum Mean Square Error (LMMSE) prediction based data analysis is equivalent to Auto-Regressive (AR) modeling of the data. Line Spectral Frequency (LSF) representation of the Linear Prediction (LP) filter was introduced by [12] and used in common cell phone communication systems including the GSM and Mixed Excitation Linear Prediction (MELP) speech coding systems, [11]. Here 20 order LSFs are used as feature parameters to represent impact sounds.

3. RESULTS

3.1 BESS Results

In our experimental studies, we used 2 times 2 fold cross validation to estimate classification error. Half of the dataset is used to calculate 32 code-vectors for each sub-dictionary via LBG-VQ [9] algorithm for each class. All records were 256 samples long.

Figure 2.a shows the selected codevectors by BESS from the dictionary. The first 32 columns of the matrix represent the code vectors of D_E and the remaining 32 the codevectors of D_F . The first 230 records correspond to the empty hazelnuts and the remaining records correspond to the full hazelnuts. As expected, the BESS algorithm selected mostly those indices that correspond to a single class. In Fig.2.b the scatter plot of the number of codevector and approximation error is visualized for both classes. A linear decision line was able to discriminate between empty and hazelnuts. The described BESS method achieved a classification accuracy of 98.3% with only 3 features. In particular individual classification accuracies for empty and full hazelnuts were 98% and 98.7%.

3.2 Base Approach Results

Features based on Weibull parameters, short time variances, short time extremum, frequency domain processing and LSFs are examined. When all 77 features are combined with a liner discriminant a classification accuracy of 96.8% is obtained.

4. DISCUSSION

We used two different types of signal representation and feature extraction strategies. Both algorithms resulted in low error rates. The best performance and complexity were obtained with the BESS approach. The bias in selecting the codevectors from the dictionary and the energy of the residual within initial approximations provided very good information for classification. In addition by using SS

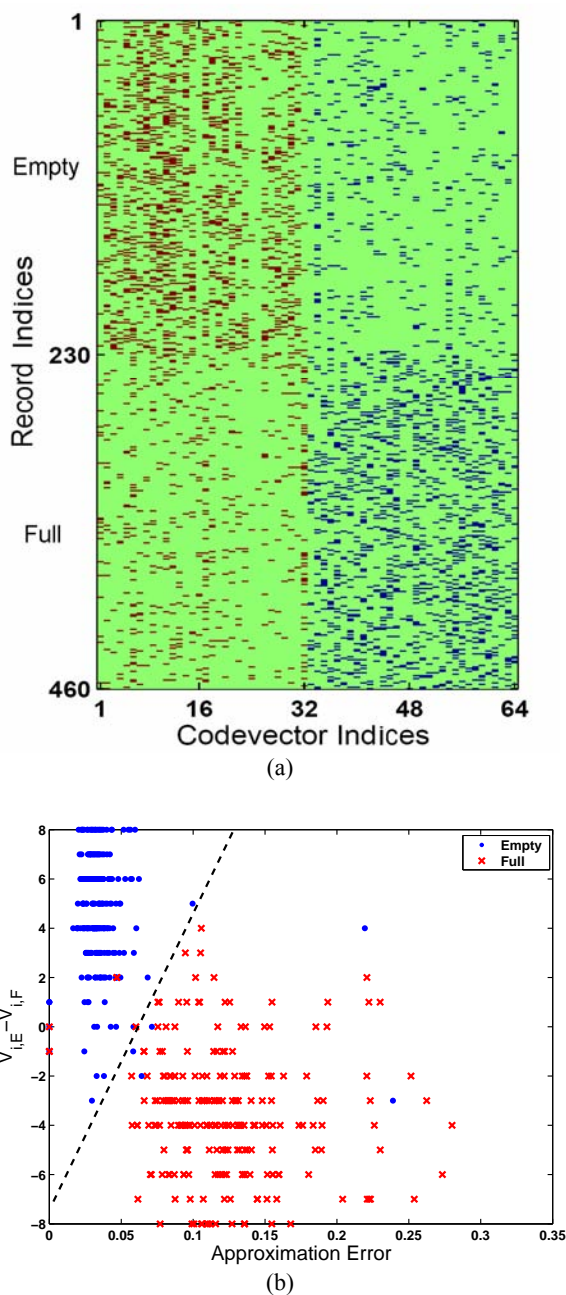


Fig.2. (a) The code vectors selected for each signal. First 230 rows belong to empty and the rest belongs to full hazelnut acoustics. The first and second 32 codevectors are generated from empty and full hazelnut training sets. (b) The scatter plot of approximation error versus to the difference between the numbers of selected vectors from each sub dictionary.

approach the classification is implemented with very small number of features, compared to the base line method. Here the capability of approximating the signal with minimal number of error plays an important role. For this problem the energy is closely related to the classification

performance. However one should also notice that the energy may not always represent discriminant information.

5. REFERENCES

- [1] Marklinder, I., M. Lindblad, A. Gidlund, M. Olsen. "Consumers' ability to discriminate aflatoxin-contaminated Brazil nuts", *Food Additives and Contaminants*, 22(1): 56–64, January 2005.
- [2] Pearson, T.C., Detection of pistachio nuts with closed shells using impact acoustics. *Applied Engineering in Agriculture*, 17(2):249-253, 2001.
- [3] Cetin, A.E., T.C. Pearson, A. H. Tewfik, "Classification of closed- and open-shell pistachio nuts using voice-recognition technology", *Transactions of the ASAE*, 47(2), 659-664, 2004.
- [4] Cetin, A.E., T.C. Pearson, A. H. Tewfik. 2004. Classification of closed- and open-shell pistachio nuts using impact acoustical analysis, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2004.
- [5] Cetin A.E, T.C Pearson, Y. Yardimci, B. Dulek, I. Onaran, "Detection of Empty Hazelnut from Fully Developed Nuts by Impact Acoustics", *Proceedings of EUSIPCO 2005*.
- [6] B. Natarajan, "Sparse Approximate Solutions to Linear Systems," *SIAM J. Comp.*, Vol. 24, pp. 227-234, Apr. 1995.
- [7] Masoud Alghoniemy and Ahmed 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.
- [8] Massoud Alghoniemy and Ahmed H. Tewfik, "Bounded Subset Selection with Noninteger Coefficients," *Proceedings of the Eusipco-2004 Conference.*, pp. 317-320, Sept. 2004.
- [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] Nonlinear regression dynamic link library (DLL) (NLREG Phil Sherod, Brentwood, TN).
- [11] Quatieri, T., *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice-Hall, 2001.
- [12] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", *J. Acoust. Soc. Amer.*, p. 535a., 1975.