CLASSIFICATION OF CLOSED AND OPEN SHELL PISTACHIO NUTS USING PRINCIPAL COMPONENT ANALYSIS OF IMPACT ACOUSTICS

A. Enis Cetin[1], Tom C. Pearson[2], and Ahmed H. Tewfik[3]

[1]Dept. of Electrical and Electronics Engineering Bilkent University, Ankara, 06800, Turkey
E-mail: cetin@ee.bilkent.edu.tr
[2]United States Department of Agriculture, USDA–ARS–GMPRC, Manhattan, Kansas
E-mail: tpearson@gmprc.ksu.edu
[3]Dept. of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455

ABSTRACT

An algorithm was developed to separate pistachio nuts with closed-shells from those with open-shells. It was observed that upon impact on a steel plate, nuts with closed-shells emit different sounds than nuts with open-shells. Two feature vectors extracted from the sound signals were melcepstrum coefficients and eigenvalues obtained from the principle component analysis of the autocorrelation matrix of the signals. Classification of a sound signal was done by linearly combining feature vectors from both mel-cepstrum and PCA feature vectors. An important property of the algorithm is that it is easily trainable. During the training phase, sounds of the nuts with closed-shells and open-shells were used to obtain a representative vector of each class. The accuracy of closed-shell nuts was more than 99% on the test set.

1. INTRODUCTION

An acoustical sorting system was developed by Pearson [1] to separate pistachio nuts with closed-shells from those with open-shells. Several different acoustic methods to evaluate food quality was developed to detect firmness in fruits [8]-[9]. Most of the acoustical systems developed thus far involve tapping the food with a plunger, recording the resulting sound to extract dominant frequency bands or other signal features correlated with firmness [9]. This technique eliminated some error caused by size variations in the fruits; however, they are not readily adaptable to high speed inspection. Pistachio shells that are open have much more freedom to vibrate about the hinge point near the stem, where the two halve are connected. In contrast, nuts with closed shells are much more ridged since the two shell halves are sealed at the suture. This physical property causes the sound of open shell and closed shell nuts to be quite different upon impact with a hard surface. However, no prior research before [1] was performed to investigate the feasibility of separating nuts with closed shellsbased on impact sounds. The sorting system included a microphone, DSP hardware, material handling equipment, and an air reject mechanism. Upon impact with a steel plate, nuts with closed-shells emitted different sounds than nuts with

open-shells. In [1], linear discriminant analysis was used to classify nuts using three features extracted from the recorded signal during the first 1.4 ms after impact. One of the discriminant features was the integrated absolute value of the signal during the first 0.11 ms after impact. The other two features were the number of data points in the recorded signal, between 0.6 and 1.4 ms after impact, having a slope and signal magnitudes below preset threshold levels. Classification accuracy of this system was approximately 97 %, with a throughput rate of about 40 nuts/s. Currently, closed-shell nuts are removed by mechanical devices, which have a lower accuracy (95 %) and can damage kernels in open-shell nuts by pricking them with a needle. The needle hole can give the appearance of an insect tunnel and cause rejection by the consumer. The acoustic based system does not cause such damage. Increased sorting accuracy of the acoustic sorter, coupled with low cost of the hardware, enables a payback period of less than one year.

Here, a new classification algorithm, based on speech recognition technology, for acoustical impact emissions is proposed for detecting pistachio nuts with closed-shells. This has the potential to broaden the number of applications currently possessed by the acoustic sorter as the algorithms are easily trainable [5]. This allows the system to sort nuts from different regions and climates.

2. CLASSIFICATION METHOD

The main discriminating features used in our classification algorithm are mel-cepstrum and the eigenvalues obtained from Principle Component Analysis (PCA) [7] are discussed together with basic principles of automatic speech recognition.

2.1. Mel-cepstrum Computation: The most common features used in modern speech recognition systems are mel-cepstrum coefficients. These features are extracted from the speech signal with the help of short-time Fourier transform analysis or wavelet analysis [2]-[5]. Mel-cepstrum coefficients are widely used in speech recognition problems as feature vectors [2]-[6]. The speech data is first divided into vectors. Let x be a vector

containing N sound samples. Mel-cepstrum vector is obtained as follow:

* Discrete Fourier Transform (DFT) $\hat{x}$ of the Hanning windowed data vector x is computed.

* The DFT, $\hat{x}$, is divided into M non-uniform sub-bands, and the energy e_i, $i=1,2,...,M$, of each sub-band is estimated. The subbands are distributed across the frequency domain according to a "mel-scale", which is linear at low frequencies and logarithmic thereafter. Most of impact sound energy lies below 15 kHz as seen in Figure 1. Thus, the DFT is divided linearly into 12 bands, in our case, and at higher frequency bands, covering 10 to 44 kHz, the sub-bands are divided using a logarithmic scale into 12 sections so that more emphasis is given to low frequency bands compared to high frequency data. In other words the DFT coefficients are grouped into M=24 sub-bands in a non-uniform manner.

* Mel-cepstrum coefficients, c_k, are computed using the discrete cosine transform (DCT)

$$c_k = \sum_{i=1}^{M} \log(e_i)\cos(k(i-0.5)\pi/M), \quad k=1, 2, ..., K \quad (1)$$

where the size of the mel-cepstrum vector K is much smaller than the data size N. The mel-cepstrum sequence is a decaying sequence for sound signals. Here, K is chosen as 20, as coefficients with an index greater than $K=20$ were observed to be negligible. The DCT has the effect of compressing the log-spectrum, thereby providing a small set of coefficients representing most of the variance of the original data set. It is observed that mel-cepstrum, c_k, give better recognition rates than both subband energies e_i, or log-subband-energies, $\log(e_i)$ [5].

2.2. Principle Component Analysis (PCA)

In addition to mel-cepstrum coefficients, eigenvalues obtained from PCA of the autocorrelation matrix of a sound signal data are used as feature vectors for modeling the sound signals. PCA based systems also consist of two phases: a training and recognition phase. In the training phase, feature vectors representing each data class are estimated from the training data. In the recognition phase, the feature vector of the current data is compared with the representative feature vector of each class.

Turk *et.al* [7] used PCA projections as feature vectors to solve the problem of face recognition in images, using Euclidean distance as the similarity function. In this approach, the correlation matrix C of the training data is first obtained:

$$C=E[(x-x_m)(x-x_m)^T] \quad (2)$$

where x represents the random sound vector, and x_m is the mean of x. The matrix C is an N by N matrix, where N is the size of data vector x. The eigenvectors of this matrix represent the projection axes or eigen-sounds of the data, and the eigenvalues represent the projection variance of

the corresponding eigen-sound. Significant eigenvalues of C form feature vectors of the data. One can select large eigenvalues as a representative feature vector of each sound class. Also these eigenvalues, together with eigen-sounds, determine a representative sound of each class.

The correlation matrix is estimated from the training set of L sound vectors $x_1, x_2, ..., x_L$ as follows: Let $X = [(x_1 -x_m) (x_2 -x_m)... (x_L -x_m)]$ be the matrix of the training vectors obtained by concatenating the sound vectors. The mean vector, x_m, is the average vector of the data set. An estimate of C is given by $C_e= XX^T$. The rank of the matrix, C_e, is less than or equal to L. Usually the training vectors are linearly independent of each other; therefore, C_e has L non-zero eigenvalues

$$XX^T u_k = \lambda_k u_k, \ k=1,2, ..., L \quad (3)$$

where λ_k and u_k are the eigenvalues and eigenvectors of C_e, respectively. The largest L' out of L eigenvalues are usually selected as a representative set of data, and the corresponding eigenvectors are used in the PCA analysis based recognition systems.

Projections of a sound vector x onto L' eigenvectors define a feature vector representing the signal x:

$$\underline{\omega}_x = [\omega_{x,1}\omega_{x,2}...\omega_{x,L'}] \quad (4)$$

where $\omega_{x,k}=u_k\cdot(x-x_m)$.

3. MATERIALS AND METHODS

The experimental setup is described in detail in [1]. The system was designed to feed pistachio nuts to an impact surface, acquire the sound signal upon impact, process the data, and then divert the product into either an open-shell or closed-shell stream. The impact plate was made of a 50.8x50.8 mm polished stainless steel bar. The large thickness was required to minimize vibrations of the bar when impacted by a pistachio nut. A highly directional microphone was used to minimize the effect of ambient noise. The sound data was sampled at 250 KHz.

When a nut impacted the plate, the microphone output signal ranged from 0 to $\pm$ 0.7 V. Data acquisition began when the output rose above 0.085 V. This threshold level was sufficient to trigger acquisition on virtually all nuts, while preventing false triggering from ambient sound. Data acquisition continued for 1.4 ms after triggering, producing 350 data points. Impact sounds from 300 closed- and 300 open-shell nuts were collected and used for this study. The training and recognition of nut split types was carried out with different subsets as will be discussed later.

3.1. Feature Extraction: Two different sets of features for classifying pistachio sounds as open or closed-shell were used: (i) Eigenvalues from PCA of sound amplitudes alone. The sound data was normalized by the Euclidian norm of each sound vector and the absolute value of sound samples was used instead of actual sound samples.

(ii) Eigenvalues from PCA of both the normalized sound amplitudes linearly combined with eigenvalues from PCA of the Mel-cepstrum coefficients. The duration of the impact sound from pistachio nuts last much shorter than a typical word or a phoneme; therefore, only one short time window was used so only one set of mel-cepstral coefficients were computed for each impact sound.

In some practical situations, C_e is too large for eigenvalue and eigenvector estimation. This was the case with the pistachio data set used in this study, as x contains $N=350$ sound samples. This difficulty can be overcome by noting that the eigensystem of $X^T X$ has the same non-zero eigenvalues of C_e, since $X X^T X\ u_k = \lambda_k X u_k$, where λ_k and u_k are the eigenvalues and eigenvectors of C_e, respectively. As a result, the reduced eigensystem of $X^T X \in R^{L \times L}$ can be solved instead of C_e, as the size of the training set L is usually less than the number of samples N in each data vector x. The new eigenvalues are the same as eigenvalues of the original system, but eigenvectors are $w_k = X u_k$.

3.2. Training Phase

In the pistachio recognition case, there are two classes of data, open and closed-shell. Assume there are $L/2$ closed-shell nut sounds, $x_1, x_2, ..., x_{L/2}$, and $L/2$ open-shell nut sounds, $x_{L/2+1}, ..., x_L$.

In the training phase of the algorithm the data is projected onto the eigenvectors, and results are averaged to find a representative feature vector for each class. Let $\underline{\omega}_o$ be the representative feature vector of open-shell nuts

$$\underline{\omega}_o = [\omega_{o,1} \omega_{o,2} ... \omega_{o,L'}] \qquad (5)$$

where

$$\omega_{o,k} = \frac{2}{L} \sum_{i=1}^{L/2} u_k \cdot (x_i - x_m)\ , \quad k = 1,2,...,L' \quad (6)$$

Similarly, let $\underline{\omega}_1$ be the representative feature vector of closed-shell nuts

$$\underline{\omega}_1 = [\omega_{1,1} \omega_{1,2} ... \omega_{1,L'}] \qquad (7)$$

where $\omega_{1,k}$ is obtained, as in Eq.(6), from the rest of the training data set.

The training phase of the algorithm is completed when $\underline{\omega}_o$ and $\underline{\omega}_1$ are obtained from the data set. Training was attempted with $L/2$ values of five, 10, 15, 20 and 30 nuts from each split category, and 280 nuts from each split type that were not used in the training set were used to test the classification accuracy. However, in the case where 30 nuts from each split type were used in the training set, the remaining 270 nuts were used to test classification accuracy in the recognition phase. In all cases except where 5 nuts from each split type were used in the training set ($L/2=5$), only the eigenvectors corresponding to the 10 largest eigenvalues were used to find the representative vectors of closed and open-shell nuts, i.e., the size of the feature vector was $L'=10$. For the training case where $L/2 =5$, only the five largest eigenvalues were computed, i.e. $L'=5$.

3.3. Recognition Phase

During the recognition phase of the algorithm, features are extracted from the current sound data and compared with the representative feature vectors of each class. Given a vector of sound data, x, its projection, $\underline{\omega}_x$, onto the eigenvectors u_k is computed using Eq. (4), i.e.,

$$\underline{\omega}_x = [\omega_{x,1} \omega_{x,2} ... \omega_{x,L'}] \qquad (8)$$

where

$$\omega_{x,k} = u_k \cdot (x - x_m) \quad k=1,2,..., L' \quad (9)$$

Then, the distance between this feature vector $\underline{\omega}_x$ and $\underline{\omega}_o$ and $\underline{\omega}_1$ is computed. If:

$$\left\| \underline{\omega}_x - \underline{\omega}_o \right\| < \left\| \underline{\omega}_x - \underline{\omega}_1 \right\| \qquad (10)$$

then it is assumed that the current sound vector x belongs to the open-shell nut class. Otherwise, it is assumed that it belongs to the closed-shell nut class.

For the case where both PCA of sound amplitudes and mel-cepstral features are used together, a feature vector representing eigenvectors of the sound data was computed using (8) and (9). This vector was compared to the representative vectors of each class, and distances were computed as in (10). Then, a mel-cepstrum vector representing the same sound data was computed, and compared with the mel-cepstrum vectors of each class. A final decision was based on the sum of the PCA and cost functions derived from mel-cepstral analysis.

Since we have only one mel-cepstral vector describing the impact sound one can also use a Gaussian-mixture model (GMM) based classifier [6].

The computational cost of this scheme during the recognition phase is not high. During the recognition phase, L' inner products are computed. Computationally expensive eigenanalysis is only carried out during the training phase, which can be implemented off-line [4][7].

4. RESULTS AND DISCUSSION

In Table 1, classification results based on PCA of sound amplitudes are presented. In the first column, the number of training sounds for each class are listed. In the second (third) column, the percentage of correctly classified closed- and open-shell nuts in the test set containing 280 sounds are listed. Only two out of 280 closed-shell nuts are misclassified in all cases, corresponding to 99.3 % recognition accuracy for closed-shell nuts. The number of misclassified open-shell nuts decreases as the number of training sounds increases up to the sixth row, in which 20 sound vectors are used in training each representative vector. Beyond this level, improvement in the recognition performance was not observed. In Table 2, classification results based on PCA of the mel-cepstrum coefficients are presented. In the first column of Table 2 the number of training vectors (or sounds) for each class is listed. In the second (third) column, the percentage of correctly

classified closed (open) shell nuts in the test set containing 280 sounds are listed. Open-shell nuts are correctly classified in all cases.

The method based on PCA features of sound amplitudes classifies closed-shell nuts more accurately compared to open-shell nuts. On the other hand, the method based on mel-cepstral features classifies open-shell nuts more accurately compared to closed-shell nuts, as shown in Table 2. The most accurate recognition results were obtained when PCA of sound amplitudes were combined with mel-cepstrum based features, as summarized in Table 3. The best results are obtained when the classification system was trained with 20 closed-shell nuts and 20 open-shell nuts (last row of Table 3). The number of misclassified open-shell nuts dropped to four, which corresponds to 98.6 % recognition accuracy in open-shell nuts. Recognition accuracy of the closed-shell nuts remained the same, 99.3 %, after linear combination. This approach is similar to use of a compound feature vector for representing speech data.

Using the same data set discussed in this study, the discriminant analysis routine described in [1] classifies open-shell nuts with 98.0 % accuracy and closed-shell nuts with 97 % accuracy.

The new method has two advantages over the previous algorithm by Pearson [1]. First, this method is slightly more accurate, which could reduce incorrectly classified pistachio nuts, resulting in a higher quality product and reduced loss of revenue from incorrectly classified open-shell nuts. Second, this method is easily trainable and may work for other types of pistachio or tree nut defects. This will be the basis of future study.

5. REFERENCES

[1] T.C. Pearson,"Detection of pistachio nuts with closed-shells using impact acoustics,"*Applied Eng. in Agriculture* 17-2, 2001.
[2] E. Erzin, A. E. Cetin, and Y. Yardimci, "Sub-band analysis for robust speech recognition in the presence of car noise," in *Proc. of the IEEE ICASSP, 1995*.
[3] F Jabloun, E. Cetin and E. Erzin, "Teager energy based feature parameters for speech recognition in car noise," *IEEE Signal Proc. Lett*ers, pp 259-261, 1999.
[4] R. Kuhn et.al, "Rapid-speaker adaptation in eigen-voice space," *IEEE Trans. Speech Audio Proc.*, 2000.
[5] Quatieri, T. F., *Discrete-time speech signal processing: principles and practice,* Prentice Hall 2001.
[6] D. Reynolds, et.al, "Robust text-independent speaker identification using gaussian-mixture speaker models," *IEEE Trans. on Speech and Audio Proc*, pp.72-83, 1995.
[7] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Neuroscience* 3(1): 71-86, 1991.
[8] F. Younce, et.al , "A dynamic sensor for cherry firmness, *Trans. of the ASAE* 38(5): 1467-1476,1995.
[9] M. Stone et al. Peach firmness prediction by multiple location impulse testing. *Trans of the ASAE* , 1998.
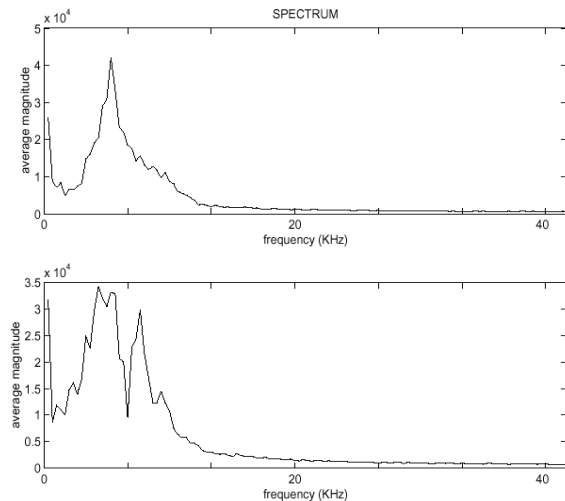
Fig. 1 The top (bottom) plot is obtained by averaging the spectra of impact sounds of 20 open-shell (closed-shell) nuts.

Table 1 Classification results for PCA of sound amplitudes. The second (third) column presents the percent of correctly classified closed (open) shell nuts in a test set containing 280 impact sounds.

No. of training sounds	Closed	Open
5	99.3	87.9
10	99.3	92.1
15	99.3	91.4
20	99.3	92.5
30	99.3	92.5

Table 2 Classification results for PCA of mel-cepstral coefficients. The second (third) column presents the percent of correctly classified closed (open) shell nuts in a test set containing 280 sounds.

No. of training sounds	Closed	Open
5	76.7	100
10	82.9	100
15	91.8	100
20	93.2	100

Table 3 Results for both PCA of sound amplitudes and mel-cepstral coefficients. The second (third) column presents the percent of correctly classified closed (open) shell nuts in a test set containing 280 sounds.

No. of training sounds	closed	Open
5	99.6	96.8
10	99.3	98.2
15	99.3	98.2
20	99.3	98.6