Eggshell Crack Detection Based on Acoustic Impulse Response and Supervised Pattern Recognition

HAO LIN, JIE-WEN ZHAO, QUAN-SHENG CHEN, JIAN-RONG CAI and PING ZHOU

School of Food and Biological Engineering, Jiangsu University, Zhenjiang, Jiangsu, People’s Republic of China

Abstract


A system based on acoustic resonance was developed for eggshell crack detection. It was achieved by the analysis of the measured frequency response of eggshell excited with a light mechanism. The response signal was processed by recursive least squares adaptive filter, which resulted in the signal-to-noise ratio of the acoustic impulse response reing remarkably enhanced. Five features variables were exacted from the response frequency signals. To develop a robust discrimination model, three pattern recognition algorithms (i.e. K-nearest neighbours, artificial neural network, and support vector machine) were examined comparatively in this work. Some parameters of the model were optimised by cross-validation in the building model. The experimental results showed that the performance of the support vector machine model is the best in comparison to k-nearest neighbours and artificial neural network models. The optimal support vector machine model was obtained with the identification rates of 95.1% in the calibration set, and 97.1% in the prediction set, respectively. Based on the results, it was concluded that the acoustic resonance system combined with the supervised pattern recognition has a significant potential for the cracked eggs detection.

Keywords: eggshell; crack; detection; acoustic resonance; supervised pattern recognition

Although many operations in the egg production have been automated, the crack detection of the eggshell still relies mainly on the manual candling (Jindal & Sritham 2003). Eggs may be contaminated through damage such as cracks, and persons may be exposed to a high health risk when eating damaged eggs (Lin et al. 1995). It mostly results in significant economic losses to the egg industry (Bain 1990).

The acoustic resonance technique is a promising method for the detection and analysis of agro-products (Primo-Martín et al. 2009; Taniwaki et al. 2009; Elbatawi 2008). The characteristic sounds produced by agro-products may provide information on their quality. Therefore, agro-products with different external or internal qualities can be measured by analysis of their characteristic sounds. When cracks are present in an eggshell, its structure is disturbed and its vibration damp is enhanced. The characteristic sound produced by impacting the shell of a cracked egg is different from that of the intact egg. Based on the principle, the eggs with intact or cracked shells can be discriminated. Coucke (1998) presented a non-destructive method for the determination of the eggshell strength based on acoustic resonance analysis. This technique can also be used for the eggshell cracks detection (Cho

Supported by National Key Technology R&D Program of China, Grant No. 2006BAD11A12-06.
et al. 2000; De Ketelaere et al. 2000). De Ketelaere et al. (2003) evaluated several parameters for the eggshell measurement. Kemps et al. (2004) calculated the elasticity modulus based on the resonant frequency of a curved shell segment. Wang et al. (2004) established the relationship between the dominant frequency and the egg physical properties. Jindal et al. (2003) employed ANN model to detect cracked eggs. Furthermore, dominant frequency and the normalisation average of the frequency domain were investigated (Wang & Jiang 2005). The qualities of eggshells coming from different strains of laying hens were compared based on the acoustic impulse analysis (Amer Eissa 2009). The mechanical properties of the rupture force, specific deformation, rupture energy, and firmness were examined (Altuntas & Şekeroğlu 2008). These studies mainly focused on the acoustic system parameters optimisation and on the characteristic frequency investigation of the response signals. However, few reports have appeared on the comparison or selection of the calibration models for the cracked eggs discrimination. Generally, a robust discrimination model is a useful way to improve the identification rates of intact and cracked eggs.

The supervised pattern recognition refers to the techniques with the priori knowledge about the category membership of samples used for the classification (Berrueta et al. 2007; Twelmann et al. 2008). The identification model is developed on a training set of samples within categories. The model performance is evaluated in the prediction set. The application of the supervised pattern recognition within chemistry, biology, pharmacology, and food science is becoming ever and more important (Berrueta et al. 2007; Jin et al. 2007). In order to highlight a good performance in the discrimination between intact and cracked eggs, three supervised pattern recognition algorithms were attempted to develop a robust discrimination model. These algorithms are K-Nearest Neighbours (KNN), Artificial Neural Network (ANN), and Support Vector Machine (SVM). Among them, KNN is a linear method, and both ANN and SVM are non-linear methods. In addition, recursive least squares (RLS) adaptive filter was used to enhance the signal-to-noise ratio. Some excitation resonant frequency characteristics of eggs were used as input vectors of the supervised pattern recognition.

**MATERIAL AND METHODS**

**Samples preparation.** In this work, 130 eggs with intact shells and 130 eggs with shell cracks were collected from a farm within 2 days after laying. All egg sizes from peewee to jumbo were used and cracks were inflicted on some eggs for the experiment. Eggs with irregular cracus were not incorporated in the data analysis. The cracks which were measured with a micrometer were 10–35mm long and less than 10-µm wide.

**Experimental system.** A system based on acoustic resonance was developed for the eggshell crack detection. The system consisted of a product support, a light exciting mechanism, a microphone,
signal amplifiers, a personal computer (PC), and the software to control the experimental setup and to analyse the results. A schematic diagram of the system is presented in Figure 1.

The frequency response is greatly affected by the selection of the egg support. Based on previous experiments, a pair of rolls made of hard rubber was used to support the eggs. The shape of the support was focused normally onto the eggshell surface. During the measurements, the distortion of the egg natural motion caused by this support was minimal. The excitation set included an electromagnetic driver, adjustable voltage (d. c.), and a light stick. The excitation mechanism was a cylinder duralumin stick. Total mass of the stick was 6 g, with the length and diameter of 6 cm and 0.2 cm, respectively. The force of excitation is an important factor affecting the magnitude and width of the pulse. In this work, adjustable voltage (d. c.) is used to control the current of the electromagnetic driver so as to control the force of excitation. Based on a previous test, the voltage of excitation was set at 30 V. In this case, a proper frequency content of the force pulse was maintained and thus the maximum signal was achieved without any instrumentation overload. The impacting point was close to the crack in the cracked eggshell, and was placed randomly in the intact eggshell.

The response signal obtained from a microphone was amplified, filtered, and captured using a 16-bit data acquisition card. The microphone was placed on a shelf and isolated from the egg supporting structure so that no disturbing vibrations were introduced when performing the measurements.

**Data acquisition and analysis.** The software program was written in LabVIEW 8.2 (National Instruments, Texas, USA) that allows a fast acquisition and processing of the response signal. The sampling rate was 22.05 kHz. The time signal was transformed to a frequency signal using a 512-point fast Fourier (FFT) transformation. The linear frequency spectrum obtained was transformed into the power spectrum. A band-pass filter was used to preserve the information of the frequency band between 1000 and 8000 Hz. Due to the signal-to-noise ratio, this favourable in the frequency band.

**Software.** All data processing algorithms were implemented with the statistical software Matlab7.1 (Mathworks, Massachusetts, USA) under Windows XP. SVM Matlab codes were downloaded free of charge from http://www.esat.kuleuven.ac.be/sista/lssvmlab/.

**RESULT AND DISCUSSION**

**Response signals**

Figure 2 shows typical acquisition signals of the intact and cracked eggs in the time domain. Due to the acoustic response an instantaneous impulse occurred thus it was difficult to discriminate between the response signals of cracked and intact eggs. The corresponding frequency domain signals transformed by FFT are shown in Figure 3. Here, the difference between them became remarkable. As a result of the analysis, the dominant resonance frequency could be observed and the average frequency value was found to be higher with the cracked eggs. As to the intact eggs, the peak frequencies were prominent and generally found in the middle position (3500–5000 Hz). In contrast, the peak frequencies of the cracked eggs were dispersed and not prominent.

**Adaptive RLS filtering**

Owing to the cracked eggshell detection being based on the acoustic response measurement, it is apt to interference by the surrounding noise. This fact is accentuated by the much damped behaviour of agro-products. Therefore, the response signal should be processed to remove the noise for further more analysis.

Adaptive interference cancelling is a standard approach to remove the environmental noise (Adall & Ardalan 1999; Madsen 2000). The recursive least squares (RLS) is a popular algorithm in the field of the adaptive signal processing. In the adaptive RLS filtering, the coefficients are adjusted from sample-to-sample to minimise the mean square error (MSE) between the measured noisy scalar signal and its modelled value from the filter (Chase et al. 2005). The scalar, real output signal, $y_k$, is measured at a discrete time $k$, in response to a set of scalar input signals, $X_i(i), i = 1, 2, \ldots n$, where $n$ is an arbitrary number of filter taps. In this research, $n$ was set to the number of degrees of freedom to ensure the conformity of the resulting filter matrices. The input and output signals are related by a simple regression model:
\[ y_k = \sum_{i=0}^{n-1} w(i) \times x_k(i) + e_k \]  
\[ \text{where:} \] 
\[ e_k \] – measurement error
\[ w(i) \] – proportion that is contained in the primary scalar signal \( y_k \)

The implementation of the RLS algorithm is optimised by exploiting the inversion matrix lemma and provides fast convergence and small error rates (Djigan 2006).

In this work, FIR filter based on RLS algorithm was used to process the acoustic response signal. The parameters of RLS were optimised as follows:

- the length of FIR filter was set to 32, the forgetting factor was set to 1, and the vector of the initial filter coefficients was set to 0.
- Figure 4 shows the frequency signals before and after the adaptive RLS filtering.

### Variable selection

Based on the difference between the frequency domain response signals of intact and cracked eggs, five characteristics of the response frequency signals were exacted as input vectors of the dis-

<table>
<thead>
<tr>
<th>Variables</th>
<th>Resonance frequency characteristics</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>area of amplitude value</td>
<td>( X1 = \sum_{i=0}^{512} Pi )</td>
</tr>
<tr>
<td>X2</td>
<td>standard deviation of amplitude value</td>
<td>( X2 = \sqrt{\frac{(Pi - \bar{P})^2}{n}} )</td>
</tr>
<tr>
<td>X3</td>
<td>frequency band of max amplitude value</td>
<td>( X3 = \text{Index}_{\text{max}}(Pi) )</td>
</tr>
<tr>
<td>X4</td>
<td>mean of top three frequency amplitude values</td>
<td>( X4 = \text{Max}_{1:3}(Pi)/3 )</td>
</tr>
<tr>
<td>X5</td>
<td>ratio of amplitude values of middle frequency bands to low frequency band</td>
<td>( X5 = \frac{\sum_{i=1}^{200} Pi}{\sum_{i=2}^{400} Pi}/200 )</td>
</tr>
</tbody>
</table>
cification models. The frequency characteristics of the response are given in Table 1.

**Qualitative discrimination between intact and cracked eggs**

In this work, 260 fresh eggs were investigated after having been divided into two subsets. One of the subsets was called the calibration set used to build the model, and the other one was called the prediction set used to test the robustness of the model. The calibration set contained 170 samples, while both the intact and cracked eggs comprised 85 samples each. The remaining 90 samples constituted the prediction set, consisting of 45 intact and 45 cracked eggs.

**K-nearest neighbours**

KNN (k-nearest neighbours) search is an important and classic means in the computer science. It is a linear method. An unknown sample of the prediction set is classified according to the majority of its K-nearest neighbours in the training set.
set (Liu & Fu 2008; Gjertsen 2007). Parameter K has a great influence on the identification rate of KNN model, the optimal values for these parameters being selected in the calibration process. KNN-parameters are optimised with minimum prediction error estimated by cross-validation in the calibration set (Koukal et al. 2007; Chen 2008). That value of K which gives the lowest error rate is selected. Therefore, in this work 10 K values (K=1, 2, ..., 10) were tested by cross-validation. The identification results of the cross validation as influenced by the K values are shown in Figure 5. As seen in Figure 5, the optimal KNN was achieved. When K was equal to 5, the identification rate by cross-validation was 86.1%. Here, the identification rate was 88.9% in the prediction set.

**Artificial neural network**

Considering that KNN is a linear classification method that need not provide a complete solution of the classification problem, the non-linear approach such as artificial neural networks (ANN) was also used to compare with KNN. It is a powerful data-modelling tool that is able to capture and represent complex relationships between the inputs and outputs (Yetilmézsoy & Demirel 2008; O’Farrell et al. 2005). The back propagation artificial neural network (BP-ANN) is a classical feed-forward multi-layer network consisting of neurons arranged in layers (an input layer, one or more hidden layers, and an output layer) (Caglar et al. 2008; Jančić et al. 2008). The eigenvectors obtained from the response signals were processed by the neural network; the network output expresses the resemblance between the object and the training pattern. Along with every pass of the training pattern and adjustment of the weight factors, the difference between the desired and calculated network outputs, defined as the network output error, will gradually become smaller until it meets the desired value. One cycle through all the training patterns is defined as an epoch. Before the optimal accordance of the network output errors is achieved for all training patterns, many epochs are required for the back propagation.

In this work, BP-ANN as one of the calibration methods for comparison was applied. Three layers (i.e., the input-hidden- and output layer) of BP-NN were arranged. Five characteristics of the response frequency signals were exacted as input vectors of ANN models. The number of neurons in the

Table 2. Comparison of the identification results coming from three models

<table>
<thead>
<tr>
<th>Discrimination models</th>
<th>Identification rates (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cross-validation calibration set</td>
<td>independent prediction set</td>
</tr>
<tr>
<td>KNN</td>
<td>86.1</td>
<td>88.9</td>
</tr>
<tr>
<td>BP-NN</td>
<td>93.2</td>
<td>92.1</td>
</tr>
<tr>
<td>SVM</td>
<td>95.1</td>
<td>97.1</td>
</tr>
</tbody>
</table>
hidden layer was set to 5. The output of BP-NN was the free amino acid content of interest. Finally, optimum network architecture was obtained with the topological architecture 5–5–1. The learning rate factor and momentum factor were set to 0.1; the initial weights were 0.3; the scale function used was the ‘tan h’ function. The permitted error was set at 0.002 and the maximal training time was 50 000 times. The identification rates of intact and cracked eggs were 95.9% and 90.9% in the cross-validation calibration set, respectively. When the performance of BP-ANN model was evaluated by means of the samples, the identification rates of intact and cracked eggs were 90.9% and 93.2% in the prediction set, respectively. The performance of BP-ANN model was remarkably better than that of KNN in both the cross-validation calibration set and the prediction set.

**Support vector machines**

Another non-linear approach, support vector machines (SVM), was also applied in this work. The support vector machine (SVM) is typically used to describe the classification problems with the support vector methods (Fernández et al. 2006). The basic concept of SVM is to map non-linearly the original data x into a higher dimensional feature space and create a hyperplane between two sets of data for classification. The transformation into the higher-dimensional space is implemented by a kernel function (Thissen et al. 2004). In general, there are three classical kernel functions: Polynomial kernel function, radial basis function (RBF) kernel function, and sigmoid kernel function. The selection of the kernel function has a great influence on the performance of SVM. Comparing with other feasible kernel functions, RBF could handle the linear and nonlinear relationships between the spectra and target attributes. Besides, RBF is able to reduce the computational complexity of the training procedure and give a good performance under general smoothness assumptions (Chen et al. 2007). Thus, RBF was recommended as the kernel function of SVM in this work.

In order to obtain a good performance, the regularisation parameter C and parameter \( \sigma \) of the kernel function in SVM model have to be optimised. Parameter C determines the trade-off between the minimising of the training error and the minimising of the model complexity. Parameter \( \sigma \) implicitly defines the non-linear mapping from the input space to some high-dimensional

![Figure 6. Contour plot of the optimisation parameters \( C \) and \( \sigma^2 \) of the model using RBF kernel](image-url)
feature space. As shown in Figure 6, the search procedure was carried out in two steps. First, a comparatively large step length in a 10 × 10 grid represented as “·” was applied. Subsequently, a much smaller step length was used to obtain the optimal combination of these parameters; the search grid “×” is also shown in Figure 6. The optimal search area is determined based on the last step. The logarithmic form was employed in the search plane owing to the large magnitude in the investigated ranges of these parameters. Optimal $C$ and $\sigma^2$ for the calibration models were found as the values of 88.13 and 3.05, respectively.

When RBF kernel and optimised parameters of $C$ and $\sigma^2$ were used, the optimal SVM model could be obtained. It was achieved with the identification rates of intact and cracked eggs of 94.7% and 95.5% in the cross-validation calibration set, respectively. When the performance of SVM model was evaluated by means of the samples, the identification rates of intact and cracked eggs were 96.2% and 97.2% in the prediction set, respectively. Compared with KNN and ANN models, SVM model showed superior ability in the discrimination between intact and cracked eggs in both the calibration and prediction sets.

Discussion of discrimination results

In order to get a good performance in the discrimination between intact and cracked eggs, KNN, ANN, and SVM models were examined comparatively. Table 2 shows the identification results coming from the three models. SVM model and ANN model provided a remarkably better performance than KNN model, thus indicating that the nonlinear information was helpful to improve the prediction performance. These results can be explained by some relevant statistical learning theories. Generally, the non-linear method is apter than the linear method at the level of self-learning and self-adjust. Therefore, the non-linear model has often a higher performance in the calibration model.

The performance of SVM model was close to that of ANN model in the calibration set, but better in the prediction set. As to the investigation between ANN and SVM models, the traditional ANN models are based on the empirical risk minimisation (ERM) principle (Fang et al. 2008). ERM minimises the error of the training data, suffering from the problem of producing models that can over-fit the data in generalisation. It means that the ‘best’ training model may always result in a poor performance in the prediction set. The foundation of SVM embodies the structural risk minimisation (SRM) principle, which covers the problem caused by ERM principle. The SRM principle theoretically minimises the expected risk based on the simultaneous minimisation of both the empirical risk and the confidence interval. SRM can maintain a trade-off between the accuracy of the training data and the capacity of the learning machine so as to improve the generalisation of the model (Wu et al. 2008). This equips SVM with a greater potential to generalise the input–output relationship learnt during its training phase for making good predictions for new input data. The difference in RM leads to a better generalisation performance for SVM than ANN. Therefore, SVM embodies excellent generalisation in its theory, which results in better results than those obtained with ANN model in the prediction set.

CONCLUSIONS

The eggshell crack detection based on the acoustic impulse resonance was aimed at in this work. The signal-to-noise ratio of the acoustic impulse response was remarkably enhanced by adaptive filters. Three supervised pattern recognition methods (KNN, ANN, and SVM) were examined comparatively to develop a discrimination model in this work. All three methods provided acceptable results. Compared with KNN and ANN, SVM model showed superior ability in the discrimination between intact and cracked eggs in both the calibration and prediction sets. The results indicated that a robust discrimination model can be a useful way to improve the identification rates of cracked eggs. It can be concluded that the use of the acoustic resonance technique combined with an appropriate supervised pattern recognition is a promising method to detect cracked eggs.

References


Received for publication April 15, 2009
Accepted after corrections July 16, 2009

Corresponding author:
Jiewen Zhao, Ph.D., School of Food and Biological Engineering, Jiangsu University, Xuefu Road 301, Zhenjiang City, Jiangsu Province, 212013, P.R. China
tel.: + 86 511 887 973 08; fax: + 86 511 887 802 01 e-mail: zjw-205@163.com