TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Brau- und Getränketechnologie

Wissenschaftszentrum Weihenstephan
für Ernährung, Landnutzung und Umwelt

Digital Processing Based Solutions for Life Science Engineering Recognition Problems

Walid G. B. Hussein

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. H.-Chr. Langowski

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Th. Becker
2. Univ.-Prof. Dr. G. H. Müller

Die Dissertation wurde am 03.04.2013 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 13.09.2013 angenommen.

Acknowledgement

All praise is due to Allah, who said "We raise in degrees whom we will, but over every possessor of knowledge is one [more] knowing", Quran (12:76). Looking back, I am surprised and at the same time very grateful for all I have received throughout these years. It has certainly shaped me as a person and has led me where I am now.

I am heartily thankful to my supervisor, Prof. Dr.-Ing. Thomas Becker, for all the hope he has put on me, before I thought I could do any research at all. His encouragement and support from the preliminary to the concluding level enabled me to develop an understanding of the subject. I am also thankful for the excellent example he has provided as a successful researcher and professor. Which exceptionally inspire and enrich my growth as a student, a researcher and a scientist want to be. His Involvement has activated my thinking that I will benefit from, for a long time to come.

In my daily work I have been blessed with friendly and cheerful research cooperation and insightful discussions by all my colleagues in the group of Bio-process analysis and technology. These have added immensely to my personal and professional time. In the first place I am particularly Indebted to the good advice, crucial contribution and friendship of our group leader Dr.-Ing. Mohamed Hussein. His ideas have been invaluable on both an academic and a personal level, for which I am extremely grateful. It is also a pleasure to thank all my old colleagues and friends in Universität Hohenheim, Stuttgart, especially Dipl.-Ing. Andreas Houben for facilitating my landing in Germany.

I would like to thank my parents and my sisters for all their love, inseparable support, prayers, caring and gently love. It was under their watchful eye that I gained so much drive and an ability to tackle challenges head on. It would not have been possible to do research without the help and support of the kind people around me. Thanks to my late grandparents, whose memory has only increased reminding me of their endless encouragement to success. I am working on it, still.

Words fail me to express my appreciation to my wife, Sara, for her support, encouragement; quiet patience and unwavering love were undeniably the bedrock upon which the past five years of my life have been built. As always, for which my mere expression of thanks likewise does not suffice. Finally, I would like to thank my children, Omar and Yasmina, for adding a delicious taste to my life, and for giving me quiet times to write this thesis.

Walid Hussein

Table of contents

Abstract

The field of Life Science Engineering (LSE) is rapidly expanding and predicted to grow strongly in the next decades. It covers areas of food and medical research, plant and pests' research, and environmental research. In each research area, engineers try to find equations that model a certain life science problem. Once found, they research different numerical techniques to solve for the unknown variables of these equations. Afterwards, solution improvement is examined by adopting more accurate conventional techniques, or developing novel algorithms. In particular, signal and image processing techniques are widely used to solve those LSE problems require pattern recognition. However, due to the continuous evolution of the life science problems and their natures, these solution techniques can not cover all aspects, and therefore demanding further enhancement and improvement.

The thesis presents numerical algorithms of digital signal and image processing to help in improving the pattern recognition based solution of some LSE problems. These problems are selected randomly from the different areas covered by LSE, including those involved in hidden animal detection, biological tissues recognition, animal taxonomy, and bioprocess monitoring problems.

Hidden weevils are traditionally detected by pheromone traps which are not able to perform the task in an early stage of infestation. Hardly seen animals such like bats are usually detected through their echolocations, with the need of accurate handling and filtration of the recorded sound streams for correct detection. In this thesis, a signal processing system is developed including the extraction of large list of conventional/unconventional bioacoustics features. The filtration process and application of window functions are investigated, and different algorithms for the selection of distinctive features are proposed. The system is applied to accurately detect the existence of red palm weevils through the analysis of recordings made by an insertion sensor into palm trees.

Meanwhile, automatic recognition of objects in biological tissues is achieved with image processing by detecting objects' boundaries in over surface picture, x ray image, ultrasound image, or magnetic resonance image. Nevertheless, the main challenge is to apply a robust edge detector. In this direction, a novel edge detector based on the energy and skewness features of the original image is developed. These features behave as smoothed versions of the image and avoid the application of prior smoothing filters and their corrupting influence on the boundaries. Non-maximum suppression approach with sub-pixel accuracy is established to thin out the detected boundaries to one pixel width. Flux equilibrium check is conducted to fill any discontinuities take place in the constructed edges image. With respect to subjective and objective measures, the developed edge detector presents competitive results in comparison to other commonly used approaches. Several two dimensional features are extract from the edges image to completely define the objects boundaries. And the developed edge detector has been applied efficiently to recognize the intramuscular fat contents in non-living animal slices images.

Animal taxonomy can also be performed by image processing through the analysis of their bioacoustics calls represented in spectrogram images. It is a modern method

which presents fast and accurate classification down to species levels. However, the collections of bioacoustics calls in animal natural environment add more difficulties to the classification process due to the attached field noise. Many enhancement approaches have been considered to suppress this background noise, but their common challenge is the degradation of the produced temporal and/or spectral accuracies. Hence, an improved spectrogram enhancement approach has been developed. The approach limits the dynamic range of the spectrogram to the enclosed high energy patterns. The crest factor image is extracted as a smoothed version of the spectrogram. The developed edge detector is applied to detect the boundaries of sound patterns, at which their surrounding noise are eliminated. The method is compared to other enhancement methods, and applied successfully to classify some birds/bats species through their bioacoustics calls spectrograms.

Alternatively, several in situ sensors and techniques are applied to monitor bioprocesses. However, some challenges are accompanying regarding the biofouling formation on the sensor surface, limited measuring range, base line drift, cost of application, and calibration complications. Particularly, ultrasound sensors are promising tools to perform online, noncontact, and non-invasive monitoring. The main parameters obtained by these sensors are the time of flight of the propagating echoes, and its corresponding speed of sound. Numerical approaches to calculate these parameters are mostly the threshold method and cross correlation method. Whereas with the first method echoes reach the threshold level sometimes after their exact starts, while with the second method the calculation is highly affected by existent noise spikes. In this thesis, a time of flight and speed of sound calculation approach is presented. The ultrasound signal is restricted to its dominant frequency. The involved power spectrum and phase shift distributions are handled to detect the time of flight between echoes corrected by their individual phase shifts. Afterwards the speed of sound is calculated by the information of the signal path length. Validations and sensitivity analyses are conducted to check the consistency and repeatability of the results. The proposed method is applied to estimate the time of flight and to monitor the speed of sound variation during online yeast fermentation process. Furthermore, the signal features are combined with temperature measurements in an artificial neural network to instantaneously predict the mixture density with high accuracy.

The developed approaches enlighten the passage which help in decreasing the challenges of LSE problems, and open the horizon to think in more improvements for the already existent solutions. Therefore, further applications of these approaches as well as their limitations and constrained are discussed.

1.   Introduction

Life Science Engineering (LSE) is a research field for the intersection between engineering and life science, with the focus of finding technical solutions for existing life science problems. It became an increasingly important subject area during the last few years due to its implementation in food technology, pharmacy, biology, metabolic engineering, bioreactors design and operation, pollution abatement technology, environmental impact assessment and life cycle analysis. LSE can be divided into three major areas, as clarified in figure (1.1). The red area is engaged in the research of food and medical technologies containing their production processes, analysis, instrumentation, automation and robotics. The green area is for the plant and pests research, including the methods of cultivation and resistance to grain weevils. While the grey area concerns with the protection of the environment, covering research in topics such as water, air pollution, as well as their analysis techniques.
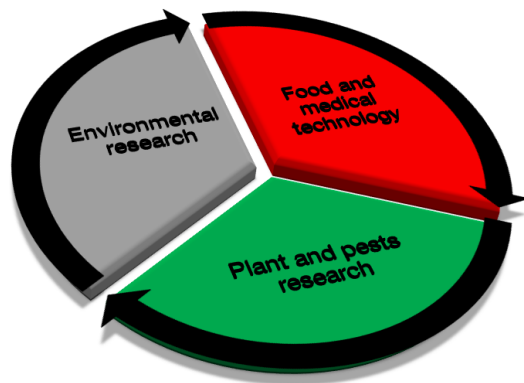

Figure (1.1): The three major fields covered by Life Science Engineering (LSE) research.

1.1   Problem Formulation

The thesis is oriented to help in improving the solution of some LSE problems by signal and image processing techniques. These problems are the difficulties in *detecting hidden or hardly seen animals*, difficulties in *objects recognition in biological tissues*, difficulties in *animal taxonomy in species level*, and difficulties in *instantaneous monitoring of bioprocesses*.

Problem of hidden animal detection
The importance of detecting hidden or hardly seen animals is gained in from the exchanged influence between these objects and environment. One example of these animals is the night-flying bat, which is primary predator of night-flying insects. These include cucumber, potato, and snout beetles; corn-earworm, cotton-bollworm, and grain moths; leafhoppers; and mosquitoes. A mine roosting colony of only 150 bats eat about 33 million cucumber beetles each summer. This saves huge amount of pesticides which could cost the farmers an estimated billion dollars [1]. Despite their critical role in our environment and economy, available studies suggest that millions of bats have already been lost. The loss of bats can increase our reliance on chemical pesticides, which often threaten both environmental and human health.
There is a rapidly growing body of evidence indicating that bat fatalities at wind mills are considerably high enough to endanger their existence, especially for those presented in figure (1.2). the lungs of bats as typical mammalian unlike those of birds,

and it has been hypothesized they are more sensitive to sudden air pressure changes in their immediate vicinity such as wind turbines, and are more liable to rupture them to explain their apparent higher rate of mortality with such devices.

Figure (1.2): Three of the most frequently killed bats at windmills. (a) Hoary bat (*Lasiurus cinereus*), (b) eastern red bat (*Lasiurus borealis*), and (c) silver-haired bat (*Lasionycteris noctivagans*).

About 48 bats are killed per wind blade annually at mountaineer [2]. From the other side, bats destroy these blades, as shown in figure (1.3), which offer source of renewable and pollution-free energy. Clearly, further construction of wind farms on wooded ridge tops, prior to finding solutions to prevent or minimize bat kills, poses potentially devastating cumulative threats to bats and wind turbines [3].
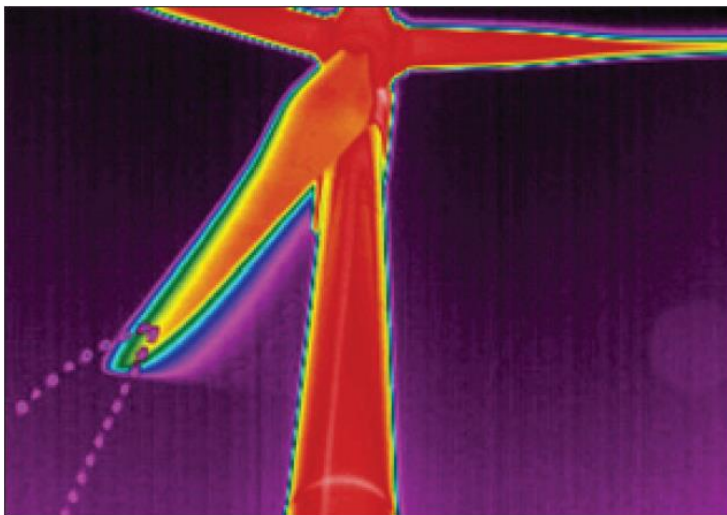

Figure (1.3): Thermal infrared image for a wind turbine, showing the trajectory of a bat that was struck by a moving blade (lower left).[3]

Automatic accurate detection of bats' existence can lead to excellent progress toward identifying causes and possible solutions for the problem. One primitive solution is to stop or reduce the turbine speed when bats are detected in the near zone. However, due to the unique behaviours of bats to fly in late night, it is difficult to visually detect their species. The recognition is always subject to the external characteristics and its accuracy was dependent on the experience and ability of the observer [4]. Governments and mammal rights organizations cannot condone further turbine construction until solutions are found to minimize or prevent bat kills.

Meanwhile, an example for hidden weevils is the red palm weevil (*Rhynchophorus ferrugineus*), shown in figure (1.4), which spends its approximately four month's life cycle into the trunk of trees. It is the most destructive pest of date palms in the world and a serious pest of coconuts.
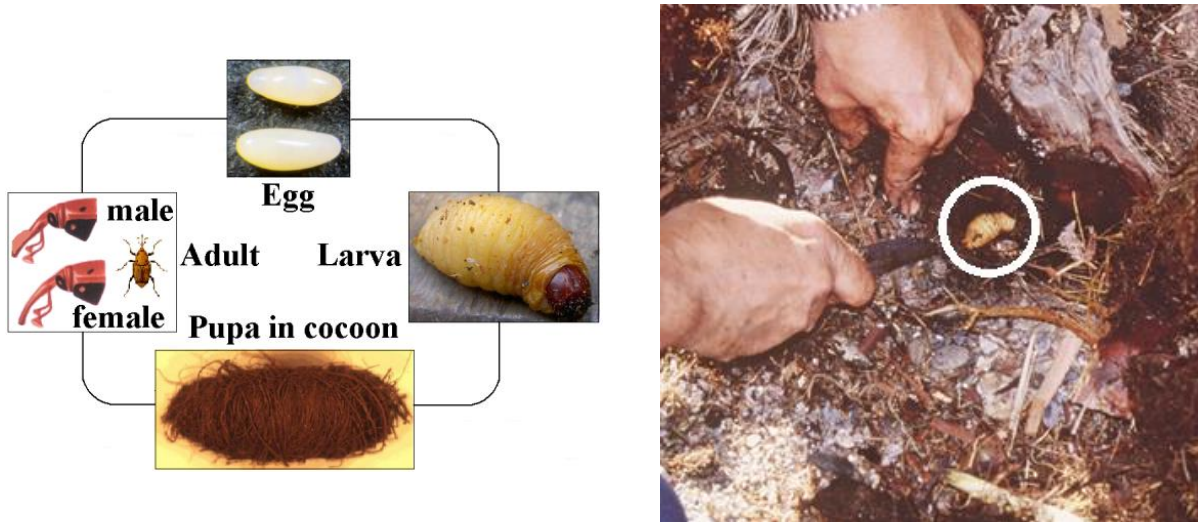

Figure (1.4): The complete life cycle stages of the Red Palm Weevil, and its larva in date palm trunk.

Red palm weevil (RPW) has caused up to 20% loss of these plantations in Asia and Middle East. Since 1980s, it has rapidly expanded its geographical range westwards. It reached Saudi Arabia and the United Arab Emirates in 1985, spreading throughout the Middle East and into Egypt. Between 1994 and 1999, it was detected in Spain, Israel, Jordan and Palestine. Then it is appeared in Italy in 2004, Canary Islands in 2005, Balearic Islands, France, and Greece in 2006, and Turkey in 2007 [5]. It has been reported in several other countries as schematically shown in figure (1.5). The hidden kind of its living enables the spread of infestation without visual appearance on the tree, and therefore no treatment is considered. European Commission is in the process of introducing emergency measures to prevent the further spread of RPW within the community.
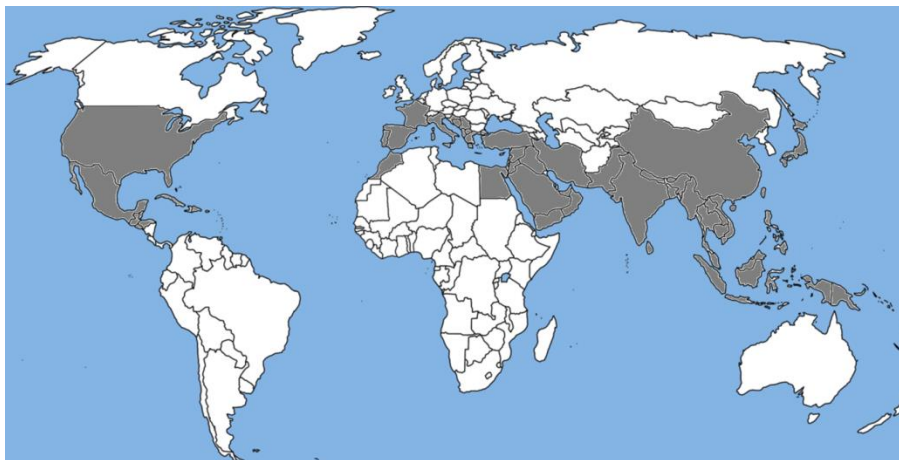

Figure (1.5): worldwide map showing the reported infested regions (in grey) by red palm weevils in 2007.

The females deposit about 300 eggs in separate holes or injuries on the palm trunk. Eggs hatch in 2 to 5 days into legless grubs which bore into the interior of the palms,

moving by peristaltic muscular contractions of the body and feed on the soft succulent tissues, discarding all fibrous material. The larval period varies from 1 to 3 months. The grubs pupate in an elongate oval, cylindrical cocoon made out of fibrous strands. At the end of the pupation period which lasts 14 to 21 days, the adult weevils emerge [6]. Larvae and adults destroy the interior of the palm tree, often without the plant showing signs of deterioration unless damage is severe. Hollowing out of the trunk reduces its mechanical resistance, making the plant susceptible to collapse and a danger to the public, as given in figure (1.6), [7]. In most cases, attack leads to the mortality of trees whatever their sizes.

Figure (1.6): Mortality and fall of the palm tree due to the Red Palm Weevil infestation.

Problem of biological tissue recognition

The second problem is elaborating in recognition of objects in biological tissues, which can be divided into two categories; living and non-living tissues. One example for the problem in living tissues is the diagnosis of Adrenoleukodystrophy [8], which is inherited disorder that leads to progressive brain damage, failure of the adrenal glands and eventually death, as shown in figure (1.7). It is essential to diagnose this disease as early as possible. Once the disease has progressed, there are currently no treatments that can replace the lost myelin. The diagnosis is mainly established by blood test for the detection of serum very long chain free fatty acid levels [9]. The test has the highest accuracy in males, however, it sometimes present negative results in women who carry the disease.
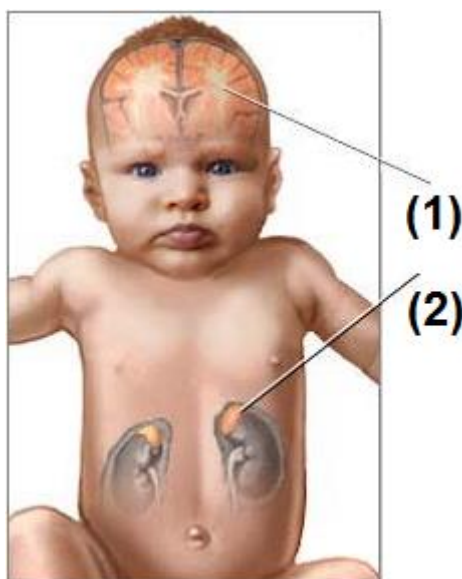

Figure (1.7): The Adrenoleukodystrophy effects which are (1) damaging the white matter of the brain and (2) impairing the adrenal glands.

On the other hand, an example for the problem in non-living tissues is the recognition of fat areas in an animal's meat slice, as shown in figure (1.8). The intramuscular fat content is one of the most important criteria for quality in meat grading systems. It has a large influence on the meat nutrition, functional properties, sensory quality, storage conditions and commercial value [10].
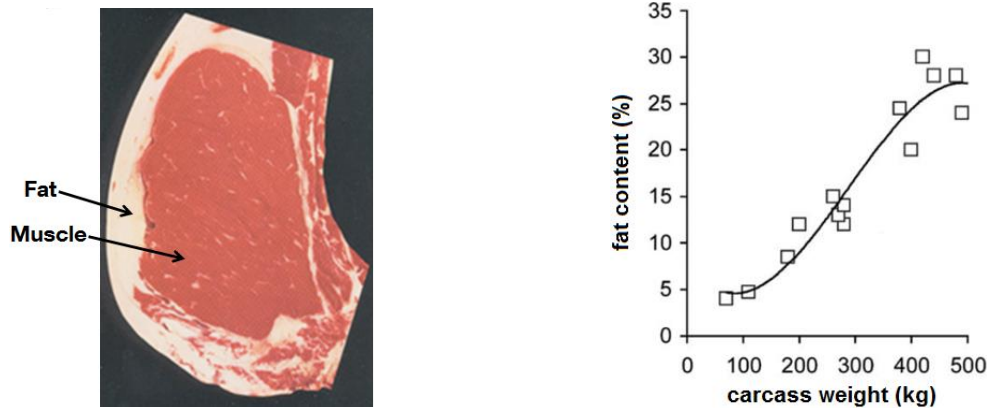

Figure (1.8): An approximate relationship between fat content and carcass weight in Japanese black x Holstein cows, [11].

This problem can be also solved by visual inspection or chemical investigation. However, visual inspection has historically been performed by use of the only "tool" available, the human eye. The process accompanying data acquisition, automatic evaluation and direct control of the ambient conditions is possible only to a very limited extent, because the human assessment furnishes above all qualitative but hardly quantitative data and, on top of this, such inspection is time consuming and cost-intensive. Instead, the chemical investigation is performed through the gravimetric ether extraction method [12], which is known as a standard reference analysis method for measuring the fat contents in simple, accurate, and robust technique [13]. But it has several drawbacks such as its long drying and extraction times, lack of automation, and the amount of solvent used per sample. Furthermore, the method is considered as expensive, destructive, and tedious method, because meat as a raw material is extremely variable and may range from 1 to 65% fat.

Problem of animal taxonomy

The third problem involved in the difficulty of animal classification down to species level (i.e., animal taxonomy). Using morphological, behavioural, genetic and biochemical observations, taxonomists identify species following the taxonomy tree shown in figure (1.9). Unfortunately, taxonomic knowledge is far from complete. In the past 250 years of research, taxonomists have named about 1.78 million species of animals, while the number of species is expected to be between 5 and 30 million, according to the Guide to the Global Taxonomy Initiative on 2007.
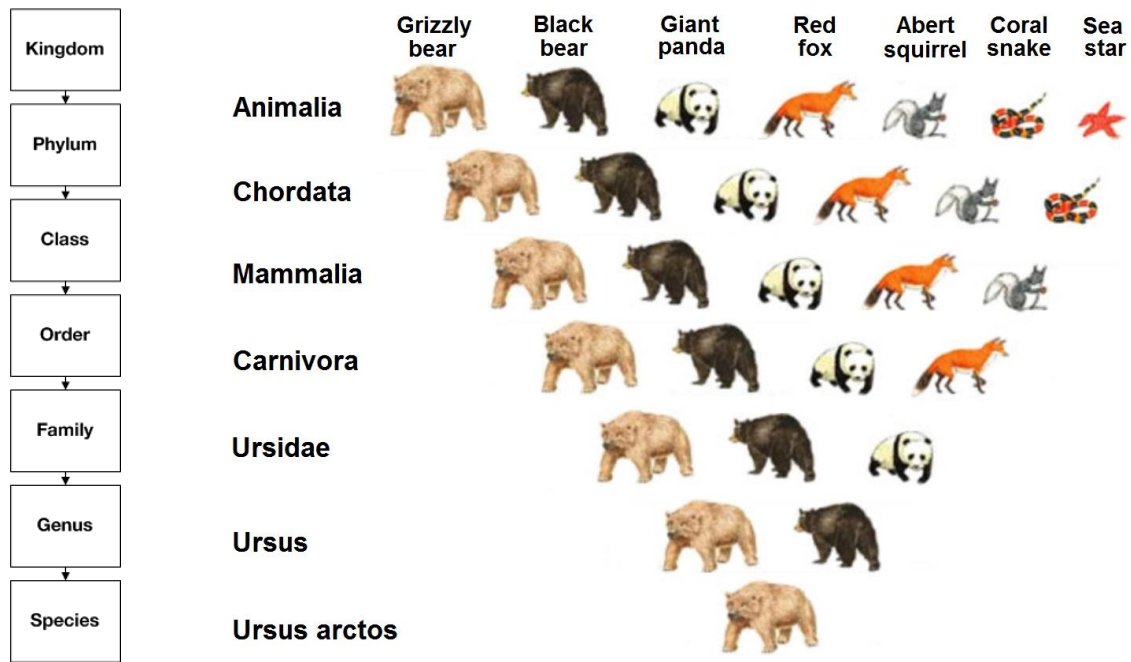
Figure (1.9): Basic taxonomy tree up to species level, with an example for the classification of Grizzly bear (*Ursus arctos*).

The secretariat of convention on biological diversity stated that taxonomists start by sorting specimens to separate sets they believe represent species. The next job is to see whether or not they already have names. This may involve working through identification guides to check similarity of physical structure, evolutionary relationships, homologies and analogies features. Such comparison should cover all of these checks because specimen may be similar to a known species in some features but differ in others. If there is no match the specimens may be considered as new species, not previously given a name. The taxonomist then has to write a description, including ways in which the new species can be distinguished from others, and make up a name for it. From finding the specimens to the name appearing in print can take several years. Due to the continuous evolutionary process of many specimens, they may achieve sensitive changes before they are already being classified. Therefore, a fast and accurate identification approach is missing, which can help the taxonomists classify specimens in reasonable times.

Problem of bioprocess monitoring

Nevertheless, the fourth problem appears through the difficulties accompanying online monitoring of bioprocesses in an accurate, non-invasive, and non-destructive policy. Microorganisms, plant and animal cells supply wide range of pharmaceutical and food products, and maximum economic benefit can only obtained if the involved bioprocesses are well monitored and controlled. The bioprocess, as displayed in figure (1.10), is a process applied on any kind of living cells (e.g., bacteria, enzymes, chloroplasts) to obtain the desired product. And bioprocess monitoring means access to continuous real-time information about all relevant variables of the process. Therefore, a detailed monitoring of bioprocess is necessary to optimize the recovery process with regard to both the quantity (e.g., microorganism growth rate) and quality (e.g., biological activity), [14].
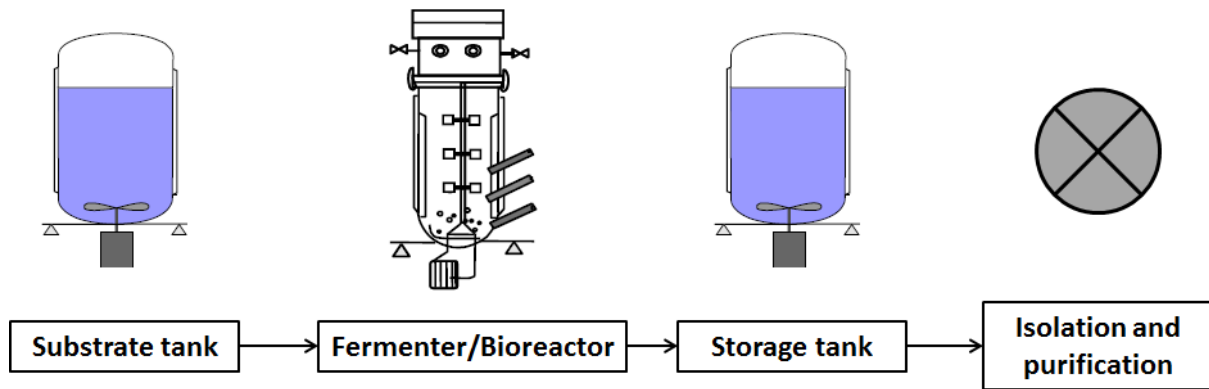
Figure (1.10): Main steps of a bioprocess consist of upstream processing stage through the substrate tank and fermenter, and the downstream processing stage through the storage tank and isolation/purification operation of the final product.

In general, bioprocesses are considered complex for monitoring due to the intricate nature of biological system and its interaction with the surrounding physical and chemical environment [15].

Furthermore, since bioprocess is harsh environment for sensors, they should not contaminate the process and be able to operate over the process period which may take weeks without recalibration. Meanwhile, one major challenge in bioprocess monitoring is the shortage of sensors which measure the key process variables that requires finding robust relationships from the sensed variables to detect the non-sensed variables. Also, offline analysis is function of the ability of automating the sample removal and preparation, which is not an easy task. This sample removal increases the risk of contamination and could perturb the physiology of the organism [16].

Software sensors (i.e., virtual sensors) may be implemented to monitor traditional bioprocess variables, suchlike pH, temperature, dissolved oxygen, pressure, and level. However, their performance depends on the initial conditions applied and understanding of the start-up procedure. As a result, the reliability of these sensors during product synthesis is low, as substantial cell death and product inhibition may occur [17].

Direct tools in bioprocess monitoring are possible through techniques like bio-sensing, using biosensors, as shown in figure (1.11), which detect an analyst that combines a biological component with a physicochemical component. They help to generate database for elucidating origin effect mechanisms of the underlying cells, and facilitate the developing of better process model for state estimation. However, biosensors usage in bioreactors is scarce owing to drift and fouling. Hence they operate under conditions not adapted for noncontact and continuous measurements [18].
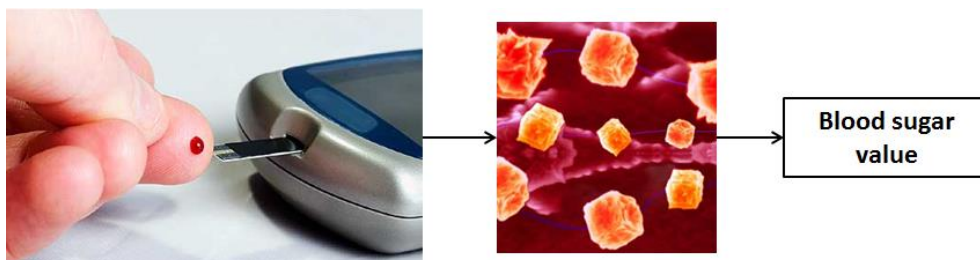

Figure (1.11): The blood glucose meter as example of the common biosensors, which uses the glucose oxidase enzyme to break blood glucose. The produced value is a measure of the glucose concentration.

1.2   Solution Approaches

Hidden animals detection approaches

There are many approaches have been developed to detect animals having hidden kind of living. For grain weevils, these approaches are pheromone based traps, as seen in figure (1.12a). However, this method is unsuitable for quarantine inspections of planting material [19]. Consequently, infested planting material is often transported to a new location before the first detectable symptoms of infestation appear. Moreover, x rays technology enables the detection of early phases of infestation, as shown in figure (1.12b), but its usage is dramatically expensive.

Figure (1.12): (a) Pheromone trap used for detection of adults Red Palm Weevil. (b) X ray image for two mango fruits at which the left fruit is identified as infested (as evidenced by the dark areas in the seed kernel) [20].

Several approaches are also applied to detect bats such as visual observation, infrared cameras [21], thermal imaging [22], and radar [23]. However, the visual observation can only be done in daylight, and with infrared cameras the deriving of a count from the recordings is tedious and time consuming. Thermal imaging and radar based detectors are very costly in equipment and man hours.

All of the hidden animals produce sounds in form of direct call, feeding sound, movement sound, spinning a cocoon sound, etc. Therefore bioacoustics analysis presents simple and quick solution to the detection problem. With bioacoustics, scientists are able to detect and study animals in dense vegetation and over distances in a non-invasive and economic way. Making acoustic recording very useful in aquatic habitats and where visual observations are difficult or impossible [24]. In recent years, the new technologies have improved the way that sounds can be sampled, analysed, stored and accessed. As a consequence, currently the collections of animal sounds are widely used and applied for research. The analysis is based on defining dominant features in these sounds, and search for these features whenever a sound is listened to detect the existence of an animal. This approach is widely used to detect bats through their echolocation calls [25, 26, 27], as expressed in figure (1.13), and can be also used in the case of RPW as will be explained in chapter 3. The challenges to this direction are the correct handling and filtration of sound streams with the available signal processing technology. And the extraction and selection of dominant features which completely and sufficiently define a sound.
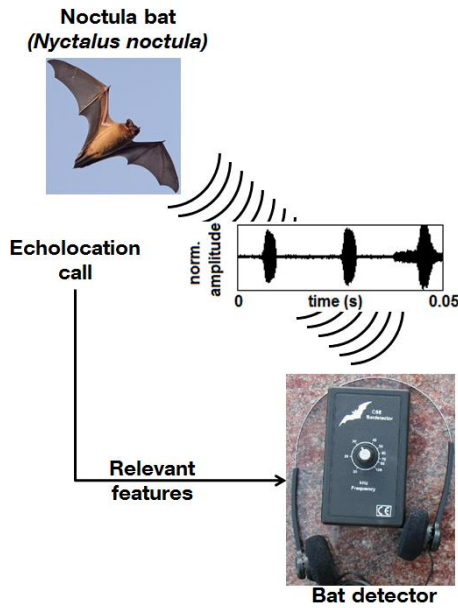
Figure (1.13): detection of Noctula bat (*Nyctalus noctula*) by reviewing the dominant features of its echolocation call.

Biological tissues recognition approaches

In the meantime, for many years computer vision techniques have been applied to recognize objects in biological tissues. These techniques include magnetic resonance imaging (MRI), computed tomography (CT), x ray, ultrasound imaging (USI) for under layer detection, and ordinary camera imaging for over surface detection. For example, MRI analysis on the brain is a must – beside a blood test – to assure the correctness of Adrenoleukodystrophy diagnosis. Lesions on the brain caused by the destruction of the myelin appear on MRI, as shown in figure (1.14), before any neurological or psychological symptoms appear [9].
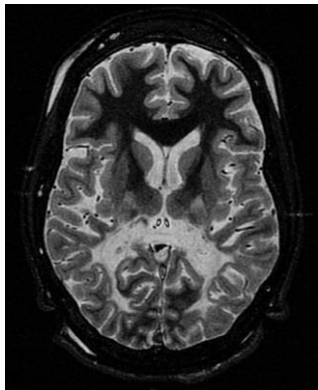
Figure (1.14): MRI showing axial scan at the level of the caudate heads demonstrates the adrenoleukodystrophy with marked loss of posterior white matter.

Some cases for the implementation of other computer vision techniques are presented in figure (1.15), implying how they added a great value and enriched the recognition of different objects in biological tissues. Over surface imaging approaches are mainly used to define objects and accurately recognize their areas and boundaries distributions, such like the fat contents in a meat slice of a non-living animal, as will be explained in chapter 3.

Figure (1.15): Recognition of objects in some biological tissues by computer vision techniques. (a) Ct images for a human right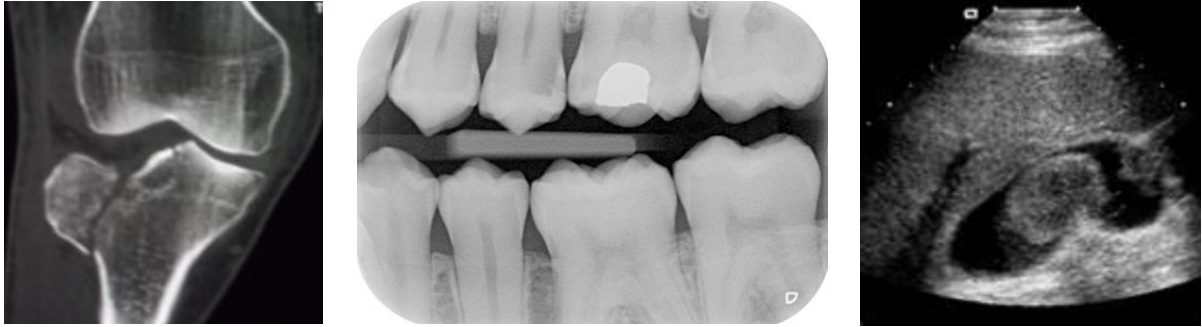 knee showing fracture, (b) x ray image of an amalgam filling (the white spot) with a significant decay in the tooth, and (c) USI shows a polypoid gallbladder cancer.

Automatic investigation of the collected images produced by these techniques is achieved by image processing methods, especially those concern with detecting the objects boundaries.

Edges are mostly detected using either the first derivatives, called gradient, or the second derivatives, called Laplacien. Laplacien is more sensitive to noise since it uses more information due to the nature of the second derivatives [28]. Various gradient based detectors were developed suchlike Prewitt, Roberts, and Sobel detectors, as seen in figure (1.16b, c, and d); respectively, for a finger print image given in figure (1.16a). They compute an estimation of gradient for the pixels, and look for local maxima to localize step edges. Typically, they are simple in computation and capable to detect the edges and their orientations, but due to lack of smoothing step, they are sensitive to noise and inaccurate [29].

| (a) | (b) | (c) | (d) | (e) |

Figure (1.16): (a) Finger print image, (b) Edges image by 3*3 Prewitt edge detector, (c) Edges image by 2*2 Roberts detector, (d) Edges image by 3*3 Sobel detector, and (e) Edges image by Canny detector with filter size of 1.5

Therefore, Canny edge detector was developed, as shown in figure (1.16d), which is probably the most widely detector and considered to be the standard edge detection algorithm in the computer vision applications [30].

The main drawback in Canny detector is its inclusion of adjustable parameter which can affect the computational time and effectiveness of the algorithm [31]. This parameter is the size of the implemented smoothing Gaussian filter which is hardly to take generic value that works well on all images.

Animal taxonomy approaches

For the difficulties involved in animal taxonomy process, the implementation of bioacoustics based classifier – for sound producing animals – has the potential to squeeze up the process in a non-invasive and noncontact approach [32]. Since

bioacoustical signals are species specific, the technique is further increase the classification accuracy. Moreover, several new species have been discovered because of their distinct vocalizations, after being wrongly classified according to their morphologically similarity with other species.

The produced sound is usually analysed through its spectrogram representation, which is a 2D image for the sound power distribution with time and frequency, as given in figure (1.17). Thus the structure of the enclosed sound patterns is measured and correlated to known species database.
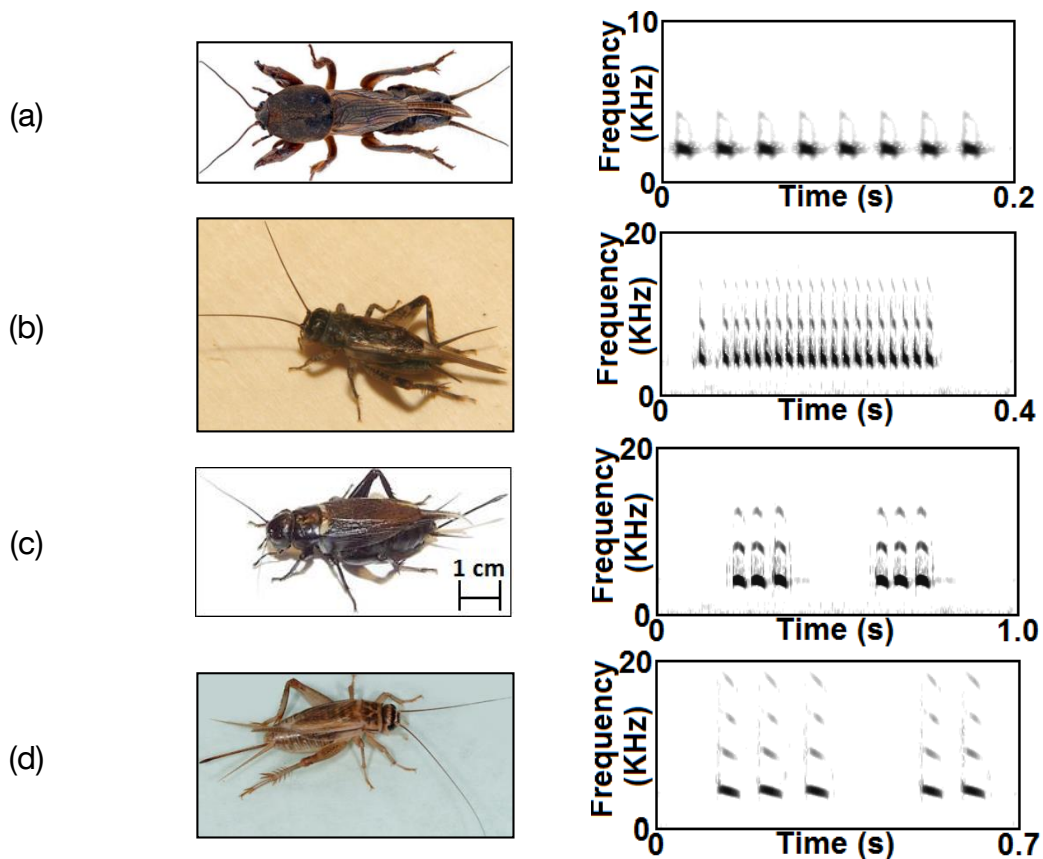

Figure (1.17): Spectrogram representation for the sound calls produced by four species. (a) *Gryllotalpa gryllotalpa*, (b) *Eumodicogryllus bordigalensis*, (c) *Gryllus bimaculatus*, and (d) *Acheta domesticus*.

To ensure the efficiency of the animal sound databases, sounds must be collected in the animal natural environment [33]. This will add more restrictions to the collected sounds in the form of attached field noises, which directly reduce the quality of the waveform, deteriorate the worth of the extracted features and thus lead to inaccuracy in classification of the sound patterns [34]. Many enhancement approaches have been considered to suppress spectrogram background noise, but the common challenge is the degradation in temporal and/or spectral accuracy of the output spectrogram. Further difficulty is added to the problem due to variability of the sound patterns structure, which can vary greatly including vertical straight, sloped straight, sinusoidal type and random patterns.

Widely applied techniques are basic band pass filtering [35], spectral subtraction [36], Wiener filter [37], and wavelet packet decomposition [38]. The band pass filtering is successful when removing low or high frequency noise and does not provide satisfactory results for sound patterns overlapped with their attached noises [39]. The

multi-band spectral subtraction method is simple and efficient, but can produce sounds with musical artefacts that are often more objectionable than the original noise. The Wiener filter technique basically considers the beginning of a signal is noise, and its adaptive type removes noise based on a training data [40]. However, during the operation on data with unknown noise, the noise level can be underestimated and the enhancement can be slightly milder [41]. The wavelet packet decomposition is effective in removing background noise in the spectrogram. But it cannot suppress much of the noise generated during the Fourier transformation, because the former noise is usually random Gaussian distribution while the latter may be modelled by Rayleigh distribution [42]. An example for spectrogram filtration by the above four approaches is given in figure (1.18).
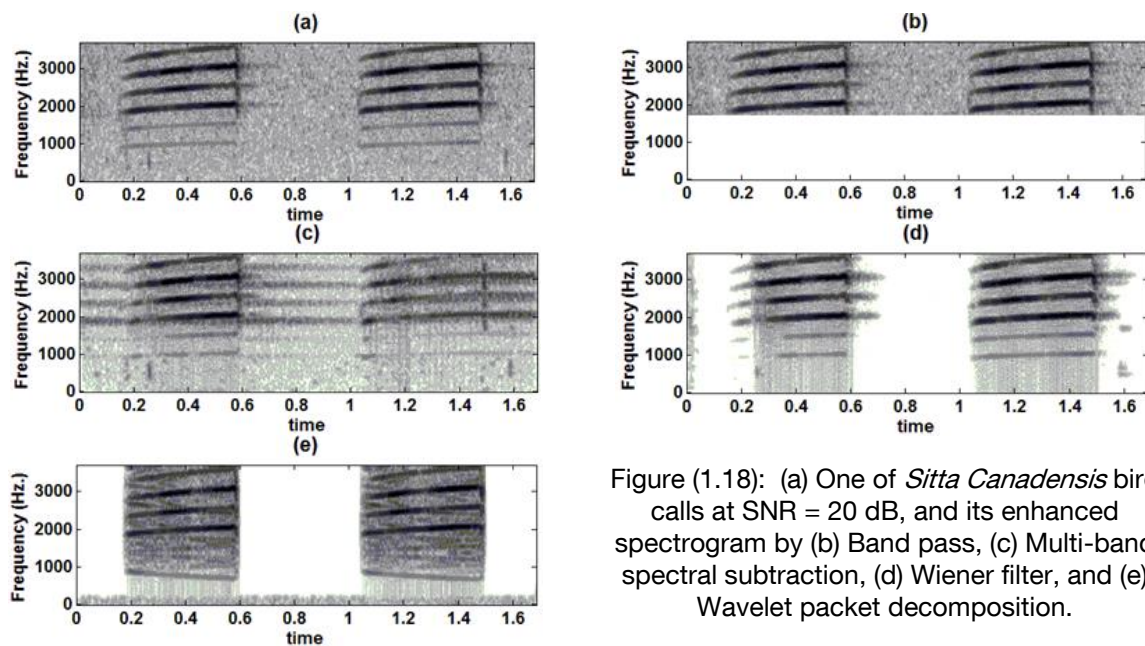

Figure (1.18): (a) One of *Sitta Canadensis* bird calls at SNR = 20 dB, and its enhanced spectrogram by (b) Band pass, (c) Multi-band spectral subtraction, (d) Wiener filter, and (e) Wavelet packet decomposition.

Bioprocess monitoring approaches

For the bioprocess monitoring problems, some improvements have been made to both biosensors and virtual sensors. However, their implementations still very challenging because of either bio-fouling of the surface and irreversible inactivation of the affinity probe due to fermentation additives, or the complex nature of the cell culture environment [43]. Other routinely applied in situ monitoring techniques are classical electrochemical sensors, optical density measurement, fluorescent spectrophotometry, infrared spectroscopy, Raman spectroscopy, and ultrasound measurements.

The classical electrochemical sensors, as described in figure (1.19a), provide the bioprocess data of temperature, pH, dissolved oxygen, dissolved CO_2, etc., that a biotechnologist relies on to determine the process status. Often they are systematic with limited mathematical models which relate measured parameters to some important process variables. Hence there is a strong need for more sensors to measure other important variables [44]. Optical density measurement is the most used method for biomass monitoring [45], but it suffers from limited measuring range and its sensitivity to bubbles and suspended solids. Fluorescent spectrophotometry measure the colour intensity of a pH sensitive fluorescent dye adsorbed on a tip of a light conducting optical fibre [46]. General weakness of this approach is photo bleaching which causes base line drift. Infrared spectroscopy, seen in figure (1.19b), can be used for glucose, ammonia, and CO_2 measurements in bioprocess media. But this method is

expensive and calibrations can be complicated [45]. Meanwhile, the Raman spectroscopy clarified in figure (1.19c) is not very developed because of interference of some cells which fluoresce in the Raman bands. Shifted subtracted Raman spectra method is developed to reduce these interferences, but it is a non-economic solution [47].
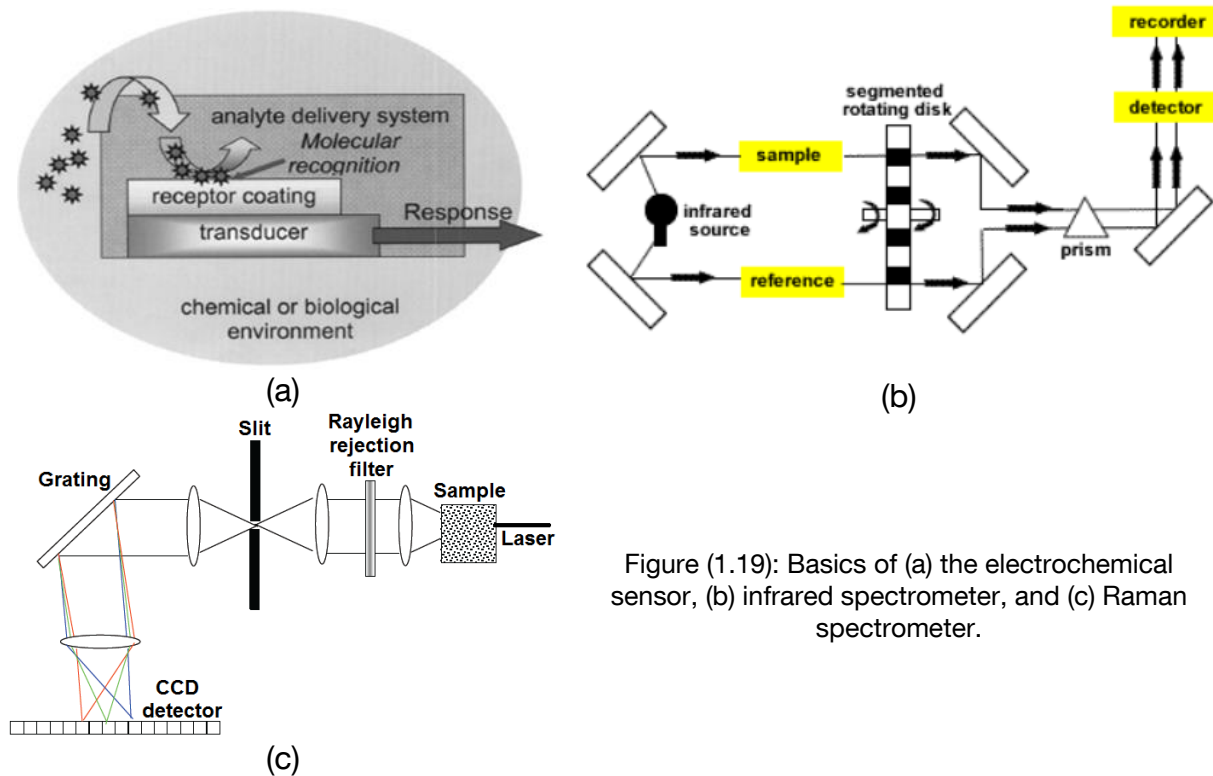

(a)

(b)

(c)

Figure (1.19): Basics of (a) the electrochemical sensor, (b) infrared spectrometer, and (c) Raman spectrometer.

Ultrasound sensor systems are of increasing interest in bioprocess monitoring because they are very sensitive and offer the possibility of online noncontact and non-invasive monitoring. Many non-sensed process key variables can be monitored through the variation of the ultrasound signal features, such like the mixture density which may be estimated from speed of sound and acoustic impedance information. The main ultrasound parameter for process monitoring and control is the speed of sound, as shown in figure (1.20). It is calculated by dividing the signal path length over the time of flight between the propagated ultrasound echoes. Time of flight is estimated by the threshold method or the cross correlation method. The first method detects the indices corresponding to the time instants when the signal amplitude crosses certain threshold, and the time of flight is then the interval between these two instances [48]. The second method searches for the instant at the relative maximum in the correlation function between the first echo and the rest of the signal, and define the time of flight to be the interval between this instant and the start of first echo [49].

Although the results obtained in the previous studies lead to the practical conclusion that the cross correlation method ensures up to 40% increasing in the accuracy of time of flight estimation with respect to the threshold method, it requires a greater computational cost and highly influenced by noise spikes in the signal. Moreover, the overall accuracy depends as well on the phase shift between the selected echoes, which is hardly noticed in time domain and requires a spectral analysis in the frequency domain.
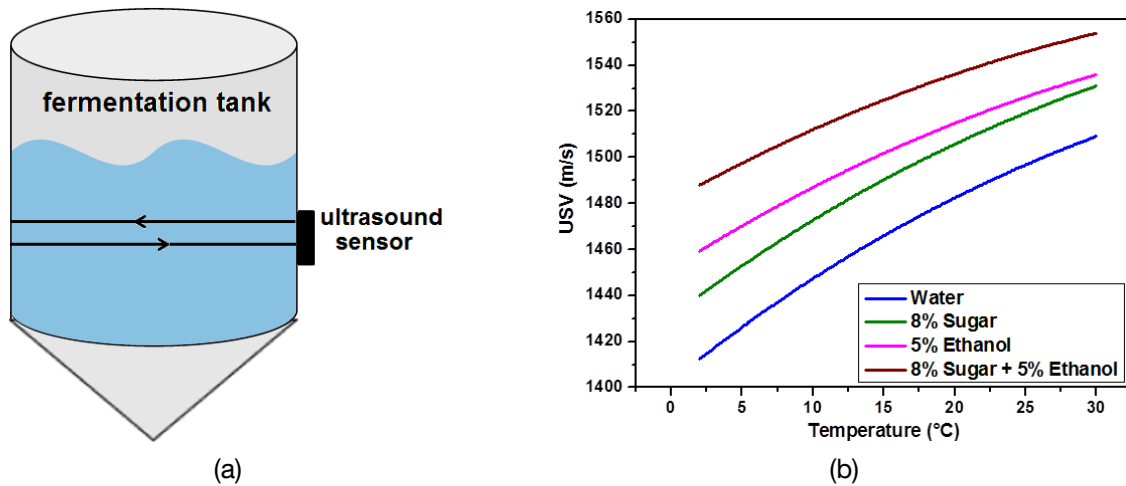
Figure (1.20): (a) Placement of ultrasound sensor to monitor bioprocess and (b) variation of the detected speed of sound for some mediums at different fermentation temperatures values [50].

1.3   Contribution of the thesis

The thesis premises on the advances in signal and image processing which help to improve current solutions approaches of the above discussed LSE problems.

For the detection of hardly seen animals and hidden weevils, a signal processing system is developed including the extraction of large list of conventional/unconventional bioacoustics features. The filtration process before extracting the time domain features, and the application of window functions before extracting the spectral features extraction are investigated. Several algorithms are presented for the selection of relevant and distinctive features. The developed system is employed to successfully detect the existence of hidden RPW through its feeding sound [51]. The system is thereafter engaged in various life science detection problems such as those presented in [52, 53, 54].

For the recognition of objects in biological tissues, a novel edge detector based on the energy and skewness features is developed. Each feature presents a smoothed version of the original image, avoiding the application of smoothing filters, and therefore prevents their accomplished drawbacks. Non-maximum suppression algorithm with sub-pixel accuracy is constructed and applied to suppress thick edges in both images (i.e., energy image and skewness image). Still discontinuities in the output images are filled by a flux equilibrium check. A reasonable combination between edges calculated by the energy image and those calculated by the skewness image is made; to produce an accurate edges image in comparison to those presented by other gradient based edge detectors. Several subjective and objective evaluation measures are formed and applied to define the accuracy of the final edges image. Afterwards, the method has been applied efficiently to recognize the intramuscular fat contents in non-living animal slices images [55].

Meanwhile, to increase the power of implementing bioacoustics spectrograms for animal taxonomy, an improved spectrogram enhancement approach has been developed. The dynamic power range of the spectrogram is restricted to avoid the problem of low level portions of the spectrogram expanding and thereby obscuring the detail of the energetic portions. Afterwards, the crest factor image is calculated as a smoothed version of the original spectrogram image. Sound patterns are detected by the above developed edge detector, while their surrounding noise are eliminated. The

method is compared to other filtration methods, and applied successfully to classify some bird species from their bioacoustics calls spectrograms [56].

Finally, for the instantaneous monitoring of bioprocesses by ultrasound techniques, an accurate time of flight estimation approach is established. The method initially limits the ultrasound signal to its dominant frequency through a high resolution spectral analysis based on short time Fourier transformation. Start times of consequent echoes are detected from the instantaneous power spectrum distribution. Afterwards, phase shift correction is applied to the detected times through the investigation of the instantaneous phase shift distribution. Validations and sensitivity analyses are conducted to check the consistency and repeatability of the results. The proposed method is applied to estimate the time of flight and to monitor the speed of sound variation during online yeast fermentation process. Furthermore, the signal features are combined with temperature measurements in an artificial neural network to predict instantaneously the mixture density with high accuracy [57].

1.4   Structure of the thesis

The thesis starts in (chapter 1) with short overview of the areas covered by LSE, with focus to four associated problems (detection of hidden animals, objects recognition in biological tissues, animal taxonomy in species level, and instantaneous monitoring of bioprocesses). Their available engineering solutions, involved challenges, and the contribution of the current work to overcome these challenges are presented within this chapter.

The remainder of the thesis is divided into three main chapters, starts with detailed explanation for the developed numerical models in both signal and image processing fields (chapter 2). The applications of these algorithms to the previously mentioned LSE problems are presented in (chapter 3), which is mainly formulated by sample of accepted peer-reviewed paper for each application. The first application is the detection of RPW in an early stage of infestation. Secondly, the noncontact recognition of intramuscular fat contents in meat slices. The third application is bioacoustics classification of bird species through their enhanced spectrograms. And the Fourth application is for robust online estimation of speed of sound and mixture density in yeast fermentation bioprocess. Afterwards, a conclusion of the presented work, outlook of the unsolved problems, and proposes for extended applications are given in (chapter 4).

2. Construction of The Mathematical Models

The LSE difficulties discussed in chapter one are pattern recognition problems, and can be mathematically solved by designing adequate pattern recognition system. This system consists of sensor to gather the signals (e.g., bioacoustics sound or ultrasound signal) or camera to gather the images (e.g., image of meat slice contains intramuscular fat). In addition to, a feature extraction mechanism that computes informative numeric from the gathered data, and supervised learning schemes that perform the recognition job based on the extracted features, as schematically shown in figure (2.1).
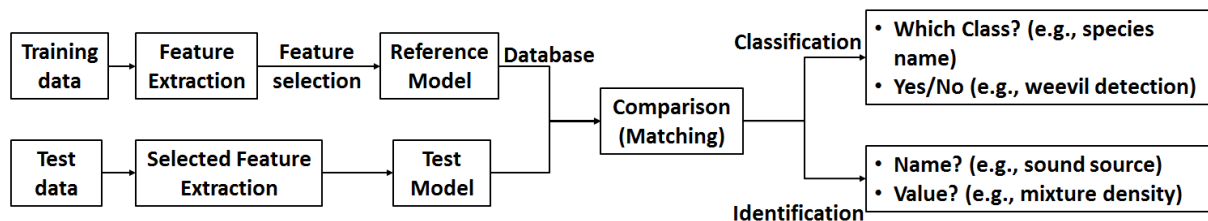

Figure (2.1): Schematic diagram for a pattern recognition system. The features are extracted from the training data, followed by selection of the most dominant features. These features build a refernece model which is compared (or matched) with the one obtained for a test data to be classified (or identified).

A supervised learning algorithm is similar to the "concept learning" in human and animal psychology. It analyses the extracted features from the training data and produces an inferred function, which is called a *classifier* (if the output is discrete) or a *regression function* (if the output is continuous). The inferred function should predict the correct output value for any valid test data.

In the following sections, the mathematical models which have been used to design the recognition system for each LSE problem are constructed. All training and test data are numerically handled by digital signal and/or image processing techniques for data filtration, features extraction, features selection, up to the recognition stage.

2.1. Extracted temporal and spectral features

Some unconventional features to those usually extracted during signal processing analysis, are explained in this section. Typical procedure follows the one given in figure (2.2), starting from the signal wave form (x) at discrete time samples (n), and ending with few features which are carefully selected to represent the important information of this signal (i.e., data mining). Signal samples are regularly spaced by sampling time(T_S), which is determined by the sampling rate (i.e., sampling frequency) of the signal, $F_S = 1/T_S$, as expressed in the sinusoidal signal shown in figure (2.3)
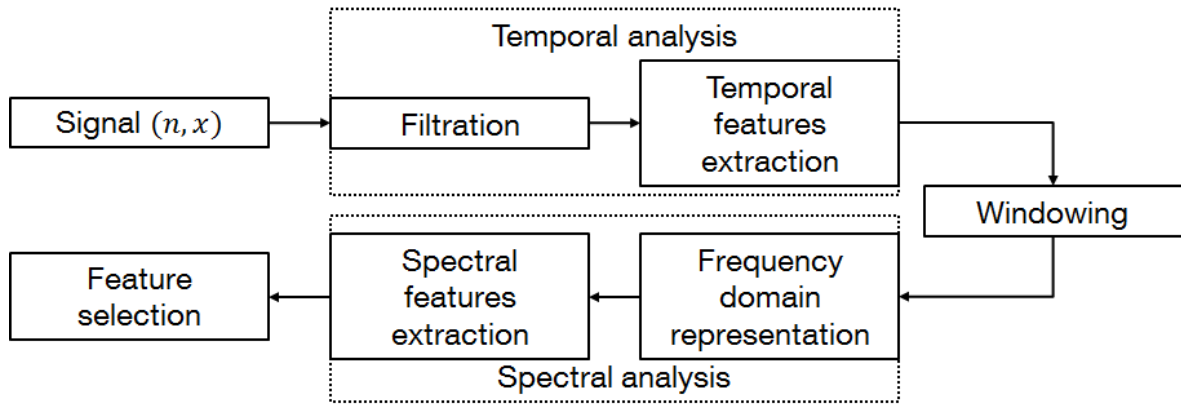
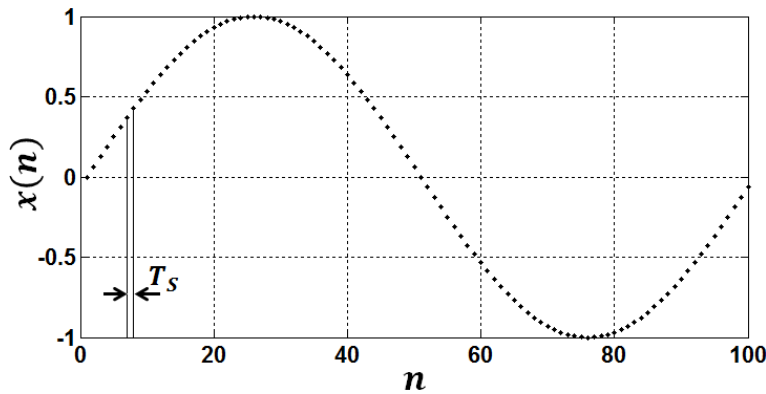Figure (2.2): Flowchart for the typical feature extraction procedure.

Figure (2.3) Discrete wave form of a sinusoidal signal in time domain, where samples are equally separated by an interval equals to the sampling time T_S.

Signal filtering

The first step is to explore the essential characteristics of the signal by eliminating attached parasitic frequencies. The three main filters are low-pass, band-pass, and high-pass filters, as described in figure (2.4), are individually implemented in the applications of this thesis when applicable.
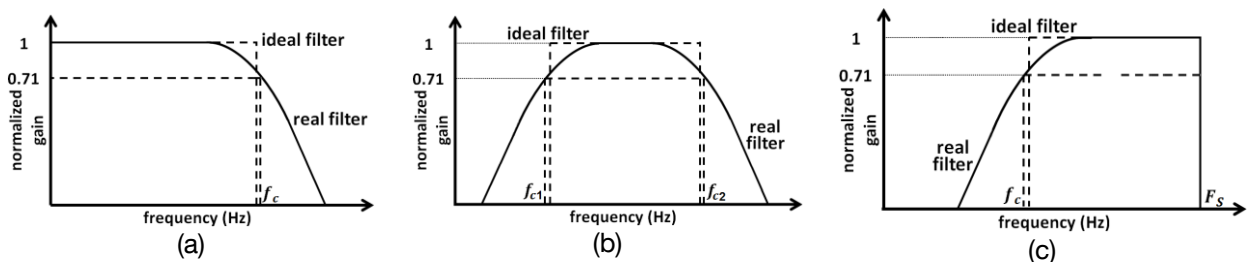
Figure (2.4): Typical performance of (a) low-pass, (b) band-pass, and (c) high-pass filters. The cutoff frequency f_c is determined by the bandwidth which is measured at half-power (associated to gain of -3 dB, or 0.707 relative to peak).

One advantage of the digital signal processing is the possibility to design digital filters that are impractical to be made in analogue processing. Therefore, the ideal filters described in figure (2.4) are directly implemented to filter the digital signals, as schematically shown in figure (2.5).

17

Figure (2.5): An example for applying ideal low-pass filter to a digital signal. (a) sinusoidal signal: $x(t) = \sin(2\pi * 0.13 * t) + \sin(2\pi * 0.93 * t)$. (b) frequency domain representation shows two frequencies at 0.13 and 0.93 Hz with F_s = 2 Hz. (c) ideal low-pass filter permits all frequencies smaller than 0.5 Hz. (d) the filtered signal with one frequency at 0.13 Hz.

Temporal Features Extraction

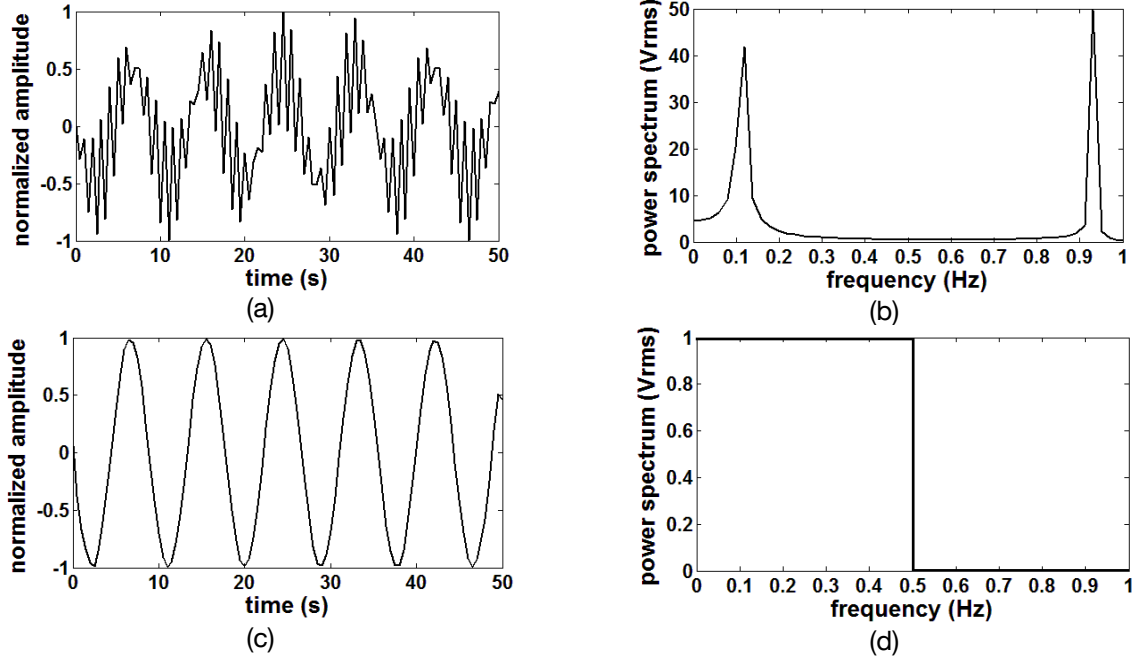The unconventional time domain features which have been extracted and implemented to the applications of this thesis are: temporal roll-off, temporal slope, and temporal spread. Temporal roll-off (n_{rf}) describes the time below which 90% of the energy distribution is concentrated, as given in figure (2.6a) and equation (2.1). Temporal slope (t_{slope}) represents the decreasing or increasing of the signal amplitudes which is computed by linear regression, as described in figure (2.6b) and equation (2.2). And temporal spread (tsp) which denotes the variance of the signal amplitudes around its mean value, as expressed in figure (2.6c) and equation (2.3).

$$n_{rf}: \sum_{n=1}^{n_{rf}} |x(n)|^2 = 0.9 * \sum_{n=1}^{N} |x(n)|^2 \tag{2.1}$$

$$t_{slope} = \frac{N \sum_{n=1}^{N}\big(n * x(n)\big) - \sum_{n=1}^{N} n * \sum_{n=1}^{N} x(n)}{\sum_{n=1}^{N} x(n) * \big(\sum_{n=1}^{N} n^2 - (\sum_{n=1}^{N} n)^2\big)} \tag{2.2}$$

$$tsp = \int_{t(1)}^{t(N)} \big(t(n) - t(C_t)\big)^2 * tpmf(n). dt \tag{2.3}$$

Where $x(n)$ is the signal amplitude at sample n, n_{rf} is the sample number for the temporal roll-off, N is the number of samples, $tpmf(n)$ is the temporal probability mass function $= \frac{|x(n)|}{\sum_{n=1}^{N}|x(n)|}$, and C_t is the temporal centroid, given in equation (2.7).
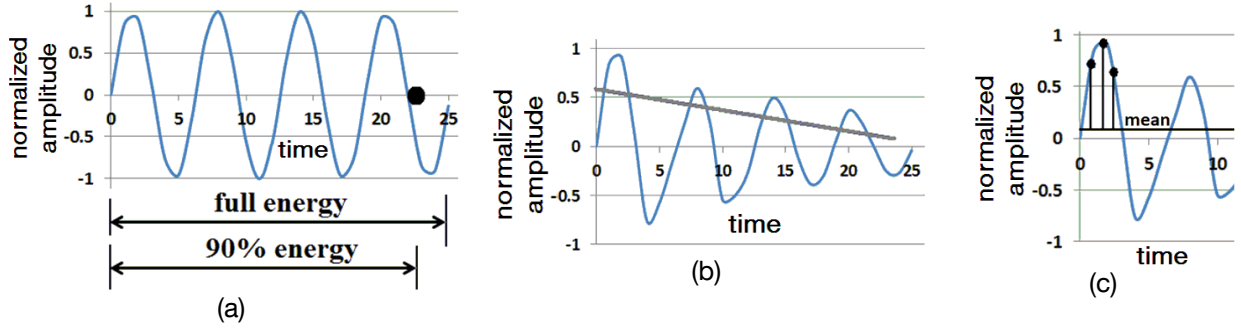
18

Figure (2.6): Schematic diagrams for the three unconventional time domain features which are implemented in this thesis. (a) temporal roll-off, (b) temporal slope, and (c) temporal spread.

Additionally, other conventional and commonly used time domain features are also extracted, following their mathematical forms presented in table (2.1). The graphical and physical descriptions of all time domain features are given in appendix A.

Table (2.1): List of the extracted conventional temporal features.

Name	Equation	
Zero crossing rate	$\frac{1}{2*N}\sum_{n=2}^{N}\left\|sign(x(n)) - sign(x(n-1))\right\|,$ $\begin{cases} sign = 1 & ; x(n) > 0 \\ sign = 0 & ; x(n) = 0 \\ sign = -1 & ; x(n) < 0 \end{cases}$	(2.4)
Short time energy	$\int (x(n))^2 dn$	(2.5)
Energy root mean square	$\sqrt{\dfrac{\sum_{n=1}^{N}(x(n))^2}{N}}$	(2.6)
Temporal centroid	$C_t = \dfrac{\sum_{n=1}^{N} n*(x(n))^2}{\sum_{n=1}^{N}(x(n))^2}$	(2.7)
Temporal entropy	$-\sum_{n=1}^{N}\left(\dfrac{\|x(n)\|}{\sum_{n=1}^{N}\|x(n)\|}\right)^2 ln\left(\dfrac{\|x(n)\|}{\sum_{n=1}^{N}\|x(n)\|}\right)^2$	(2.8)
Temporal flux	$\sum_{n=1}^{N}\left(\left.\dfrac{\|x(n)\|}{max(\|x(n)\|)}\right\|_t - \left.\dfrac{\|x(n)\|}{max(\|x(n)\|)}\right\|_{t-1}\right)^2$	(2.9)
Temporal crest-factor	$\dfrac{max(\|x(n)\|)}{\frac{1}{N}\sum_{n=1}^{N}\|x(n)\|}$	(2.10)
Temporal decrease	$\left(\dfrac{1}{\sum_{n=2}^{N}\|x(n)\|}\right)*\sum_{n=2}^{N}\dfrac{\|x(n)\| - \|x(1)\|}{n-1}$	(2.11)
Temporal skewness	$m_3 = \int_{t(1)}^{t(N)}\left(t(n) - t(C_t)\right)^3 * tpmf(n).dt$ $temporal\ skew = \dfrac{m_3}{tsp^{3/2}}$	(2.12)

| Temporal smoothness | $20 * \displaystyle\sum_{n=2}^{N-1} \left\| log\|x(n)\| - \dfrac{log\|x(n-1)\| + log\|x(n)\| + log\|x(n+1)\|}{3} \right\|$ | (2.13) |

Windowing and frequency domain representation

Wave form of a signal in time domain is not always the best representation in most signal processing related applications. In many cases, the distinguished information is hidden in the frequency spectrum, which shows which frequencies exist in the signal. Hence, there are only two ways that are common for information to be represented in naturally occurring signals. These are information represented in time domain, and information represented in frequency domain [58], and information that cannot be readily seen in time domain can be seen in the frequency domain. The signal can be transformed to frequency domain by several approaches, but the best known and commonly used approach is the Fast Fourier transform (FFT), performed through equation (2.14).

$$X(m) = \sum_{n=1}^{N} x(n)e^{-i2\pi(n-1)(m-1)/N} \qquad , m = 1,2,\dots M \tag{2.14}$$

Where $X(m)$ is the Fourier transform coefficient at m^{th} frequency bin, and M is the total number of frequency bins in power of 2 (e.g., 512, 1024, 2048,…).

The FFT computations assume a signal is periodic [i.e., has integer number of complete cycles], and determines its frequency contents, as displayed in figure (2.7b) for a periodic sine wave given in figure (2.7a). However, since most of real signals are non-periodic, the resulting frequency spectrum by FFT suffers from leakage, as exposed in figure (2.7d) for non-periodic sine wave of figure (2.7c). Leakage results in the signal energy smearing out over a wide frequency range when it should be in a narrow frequency range. The dispersed shape of the FFT makes it more difficult to identify the frequency content of the measured signal and thereafter inaccurate information extraction.
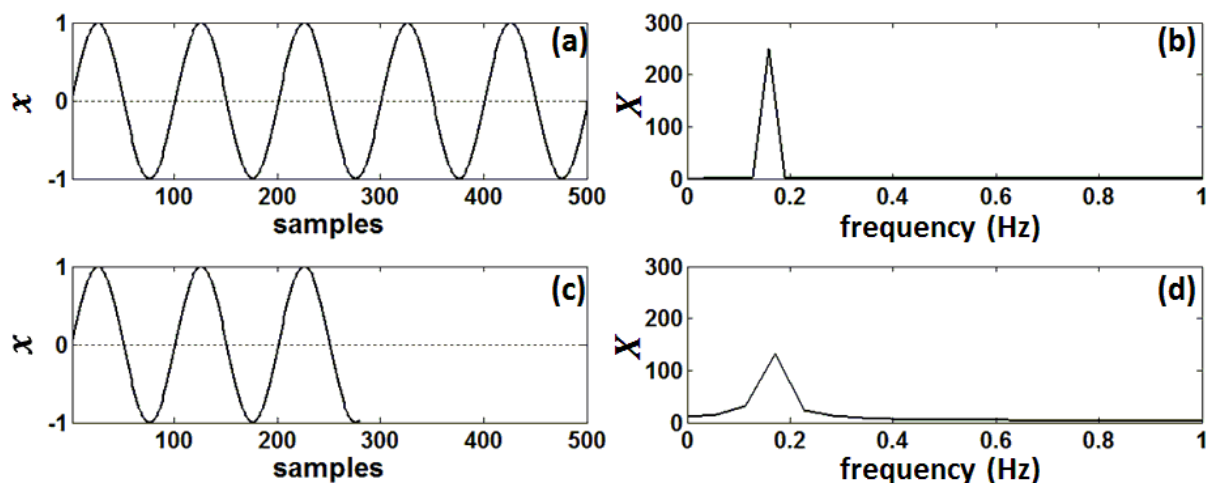

Figure (2.7): Illustration for the spectral leakage occurs when FFT is implemented to compute the frequency contents of non-periodic signal. (a) A $(1/2\pi)$ Hz periodic sine wave, (b) The resulting FFT shows one peak in a narrow band around $(1/2\pi)$ Hz, (c) A $(1/2\pi)$ Hz non-periodic sine wave, (d) The resulting FFT having more energy dispersion.

As a solution, before FFT computations a symmetrically shaped window function $w(n)$ is multiplied by the signal forcing it to be periodic. The most common window functions are Hamming (equation 2.15), Hanning (equation 2.16), Blackman (equation 2.17), and Bartlett (equation 2.18). The time domain representations of these window functions are clarified in figure (2.8a), in comparison to the rectangular window (i.e., no-window) function (equation 2.19).

$$w(n) = 0.54 - 0.46 * cos\left(\frac{2\pi n}{N}\right) \tag{2.15}$$

$$w(n) = 0.5 - 0.5 * cos\left(\frac{2\pi n}{N}\right) \tag{2.16}$$

$$w(n) = 0.42 - 0.5 * cos\left(\frac{2\pi n}{N}\right) + 0.08 * cos\left(\frac{4\pi n}{N}\right) \tag{2.17}$$

$$w(n) = \begin{cases} \dfrac{2n}{N} & , 0 \leq n \leq \dfrac{N}{2} \\ 2 - \dfrac{2n}{N} & , \dfrac{N}{2} < n \leq N \end{cases} \tag{2.18}$$

$$w(n) = 1 \tag{2.19}$$
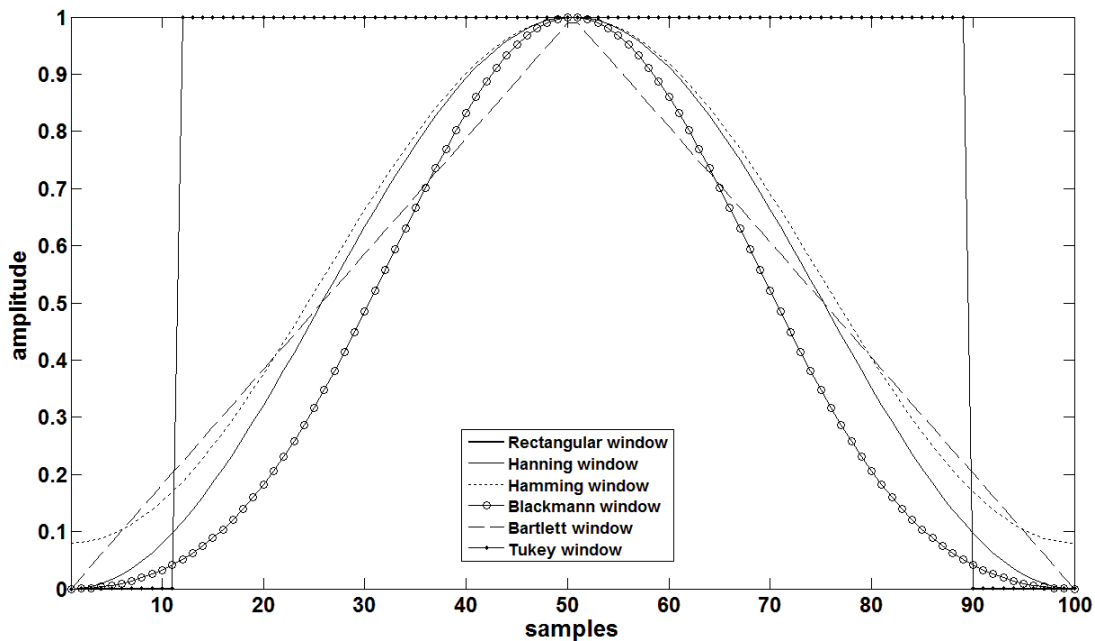
Figure (2.8a): Time domain representation for some of the commonly applied window functions.

Each type of window affects the resultant spectrum in a slightly different way, with its own advantage and disadvantage over the others. The frequency domain representations of these window functions, as displayed in figure (2.8b), determine their performance on improving the spectrum of a signal. The major controlling parameters are the main lobe width, and the drop rate between the main and first side lobes. The tighter the main lobe width, the better the frequency resolution of the resulting FFT, makes it easier to detect the exact frequency of a peak in the spectrum. The higher the drop rate between main and first side lobes, the less spectral leakage occurs around the dominant frequency. An example for applying the above window function to obtain the spectrum of a non-periodic sinusoidal signal is shown in figure (2.9).
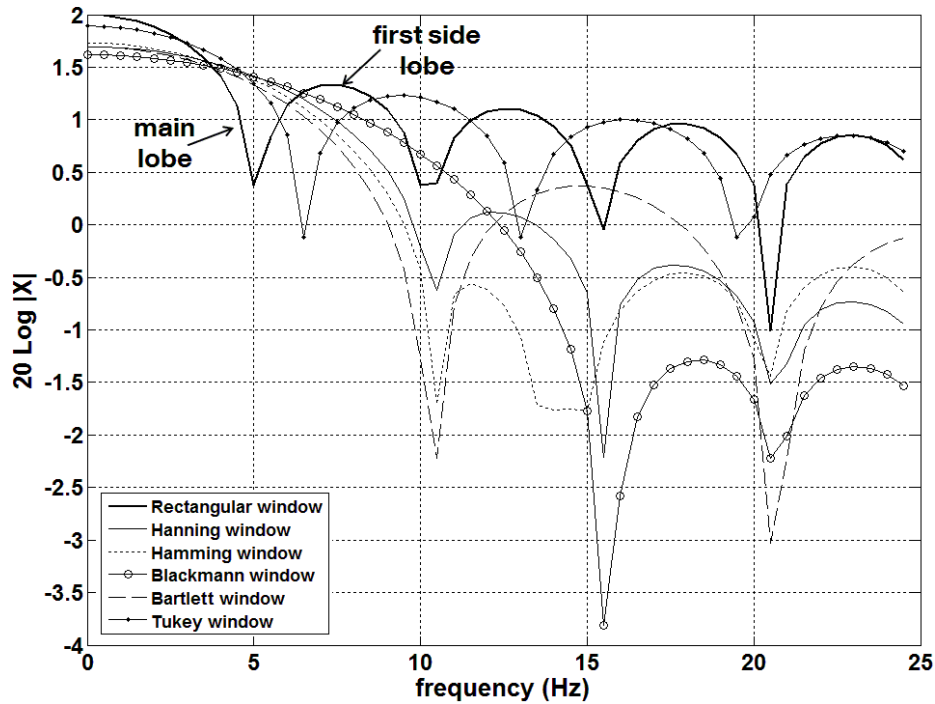
Figure (2.8b): Frequency domain representation for the window functions given in figure (2.7a)
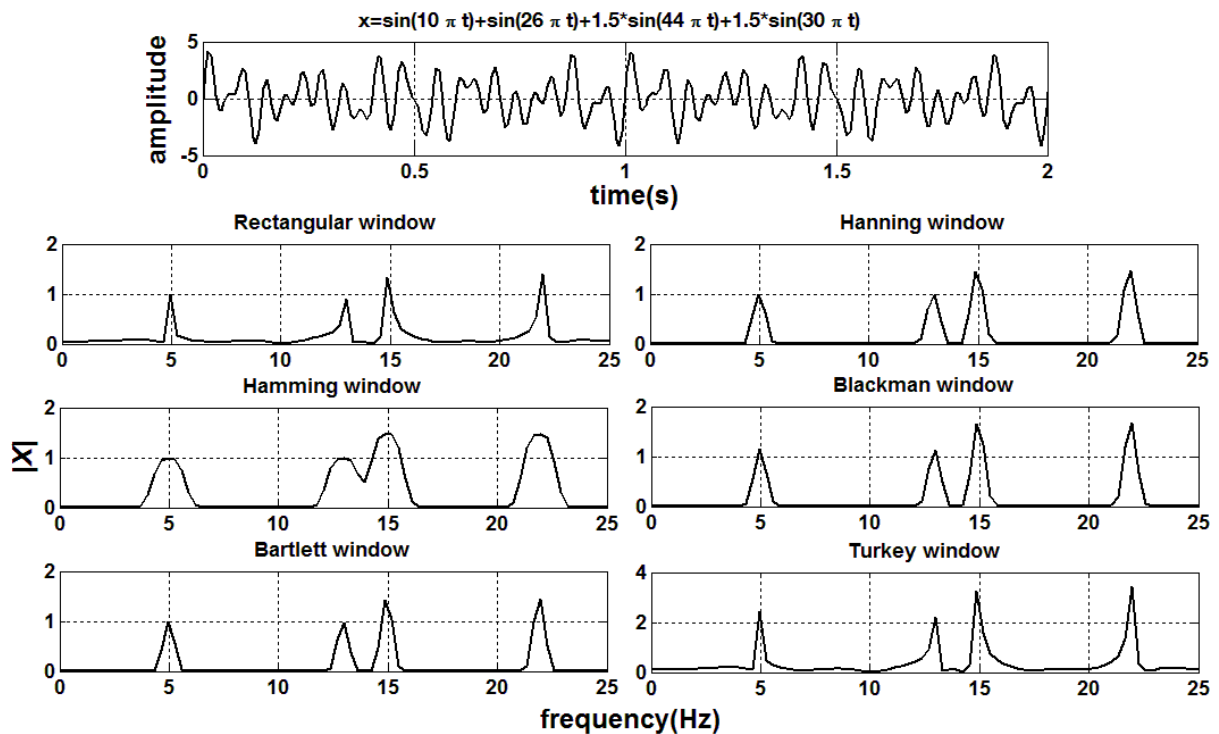
Figure (2.9): Overcoming the inefficiency in calculating the spectrum by applying window functions for non-periodic sinusoidal signal of 4 frequencies.

Spectral features extraction

By selecting the appropriate window function, in correspondence to the required accuracy in the resulting spectrum, equation (2.20) is applied to calculate FFT coefficients which form the frequency domain representation of the signal.

$$X(m) = \sum_{n=1}^{N} w(n).x(n).e^{-i2\pi(n-1)(m-1)/N} \qquad , m = 1,2,\dots M \qquad (2.20)$$

Afterwards, a list of spectral features is extracted from this representation, following an approach similar to that for temporal features extraction. Some of the unconventional spectral features are extracted suchlike spectral crest factor, spectral increase, ascent band, spectral decrease, and descent band. The crest factor measures the compactness of the spectrum and calculated by dividing the maximum spectrum over the average value, as given in (equation 2.21). The spectral increase defines how sharp the spectrum increases from 0.8% to 80% of its maximum value, and the ascent band is for the covered frequency range during this increment. The spectral decrease represents how sharp the spectrum decays from 80% to 0.8% of its maximum value, and the descent band is for the covered frequency range during this decay, as displayed in figure (2.10).

$$\text{spectral crest factor} = \frac{\max(|X(m)|)}{\frac{1}{M}\sum_{m=1}^{M}|X(m)|} \qquad (2.21)$$
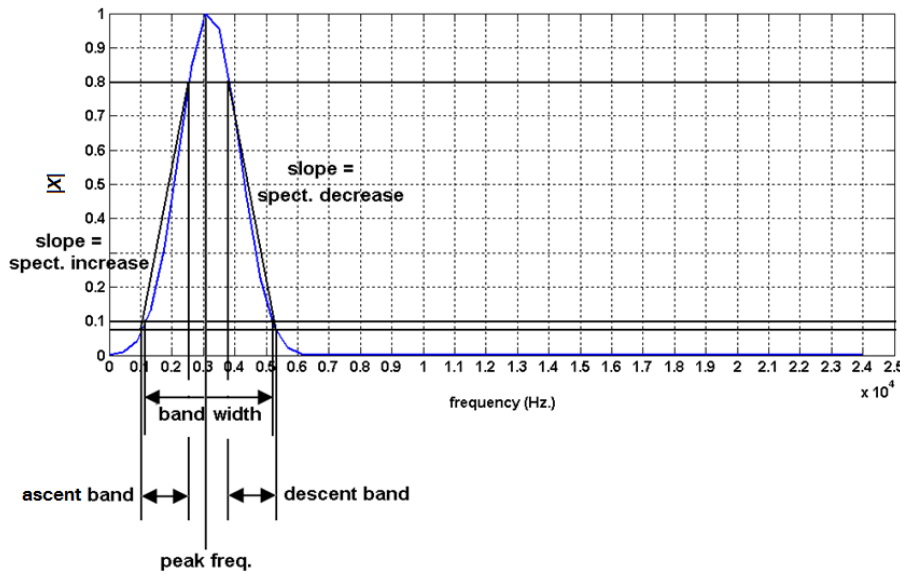
Figure (2.10): Schematic diagram for the spectral increase, spectral decrease, ascent band, and descent band.

In the meantime, a list of conventional spectral features is also extracted following their mathematical forms given in table (2.2). Graphical and physical descriptions for all spectral features are presented in appendix A.

Table (2.2): List of the extracted conventional spectral features. Where C_s is the spectral centroid, pmf is the probability mass function, m_{rf} is the frequency bin at spectral roll-off, and ssp is the spectral spread.

Name	Equation	
Spectral Flux	$\sum_{m=1}^{M-1} \left(\dfrac{|X(m)| - |X(m+1)|}{max(|X(m)|)} \right)^2$	(2.22)

|		
Spectral roll-off	$$\sum_{m=1}^{m_{rf}} |X(m)| = 0.9 * \sum_{m=1}^{M} |X(m)|$$	(2.23)
Spectral Centroid	$$C_s = \frac{\sum_{m=1}^{M} n.\,|X(m)|}{\sum_{m=1}^{M} |X(m)|}$$	(2.24)
Spectral Root Mean Square	$$\sqrt{\frac{\sum_{m=1}^{M} (|X(m)|)^2}{M}}$$	(2.25)
Spectral Slope	$$\frac{M * \sum_{m=1}^{M}(m * |X(m)|) - \sum_{m=1}^{M} m * \sum_{m=1}^{M}|X(m)|}{\sum_{m=1}^{M}|X(m)| * (\sum_{m=1}^{M} m^2 - (\sum_{m=1}^{M} m)^2)}$$	(2.26)
Spectral Energy	$$\sum_{m=1}^{M} |X(m)|^2$$	(2.27)
Spectral Smoothness	$$20 * \sum_{m=2}^{M-1} \left| log|X(m)| - \frac{log|X(m-1)| + log|X(m)| + log|X(m+1)|}{3} \right|$$	(2.28)
Spectral Skewness	$$\frac{\sum_{m=1}^{M}(m - C_s)^3 * pmf(m)}{ssp^{3/2}} \begin{cases} < 0 & \text{more energy on the right} \\ = 0 & \text{symmetric distribution} \\ > 0 & \text{more energy on the left} \end{cases}$$	(2.29)
Spectral Entropy	$$-\sum_{m=1}^{M} pmf(m) * ln\big(pmf(m)\big)$$	(2.30)
Spectral Spread	$$ssp = \sum_{m=1}^{M}(m - C_s)^2 * pmf(m) \,, \qquad pmf(m) = \frac{|X(m)|}{\sum_{m=1}^{M}|X(m)|}$$	(2.31)
Spectral Kurtosis	$$\frac{\sum_{m=1}^{M}(m - C_s)^4 * pmf(m)}{ssp^2} \begin{cases} < 3 & \text{flatter distribution} \\ = 3 & \text{normal distribution} \\ > 3 & \text{peaker distribution} \end{cases}$$	(2.32)

It is worth mentioning that signals do not maintain their characteristics constant over time, which means they are non-stationary and their features vary with time. Therefore, before extracting these features, the signal is divided into short time segments with 10% overlapping, and each segment is considered "quasi-stationary". The feature of each segment is then being extracted, and this process is called short term feature extraction, as shown in figure (2.11). Segment length should be large enough (e.g. > 50 samples) for the feature calculation to have enough data, and short enough for the

assumption of stationary signal still valid. Afterwards, the extracted feature values can be written in a vector form with a size equal to the number of segments. The short term features extraction is followed by a midterm extraction process, at which the feature vector is implemented to calculate one of its statistics (typically its mean value), and this value is used to represent the feature over the whole signal.
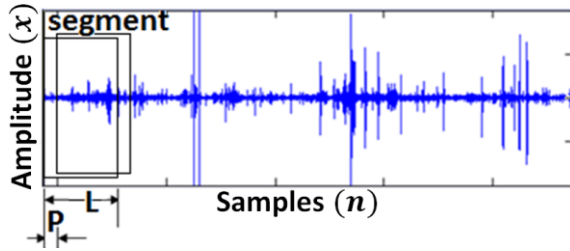
$$M = \left(\frac{N - [L - P]}{L}\right) + 1$$

$$F_1 = mean[f_{1_i}] : i = 1,2,\dots,M$$
$$F_2 = mean[f_{2_i}] : i = 1,2,\dots,M$$
$$\vdots$$
$$F_K = mean[f_{K_i}] : i = 1,2,\dots,M$$

Figure (2.11): Short term feature extraction algorithm for a signal. Where L is the segment length (i.e., number of its included samples), P is the number of overlapped samples between two following segments, N is the number of signal samples, and M is the number of signal segments. f_{1_i} is the value of feature F_1 on segment i , and K is the total number of the extracted features.

Feature selection

Better pattern recognition results, whether in regression or classification, are obtained using a limited number of relevant and non-correlated features. These features are carefully selected from the large extracted features list according to their relevancy and performance in the recognition task. One of the common approaches for feature selection is S*equential Forward Selection* [59], where one feature is added to the selected ones at each step, if it maximizes the recognition performance. Meanwhile, *Sequential Backward Selection* method starts with all features list and removes those reducing the performance [60]. The drawbacks in these algorithms are their high complexity and time consumption. Alternatively, *Fisher Discriminant Ratio* and *Multimodal Overlap Measure* algorithms [61], give rank to each feature by measuring the overlap of its probability density function between different classes. However, no integration for the correlation between the selected features is considered.

In this thesis, two additional simple feature selection approaches are proposed, which are *Jacobian based feature selection* approach (for regression problems) and *steadiness based feature selection* approach (for classification problems). These approaches are implemented in the life science applications, given in chapter 3, if applicable, otherwise; sequential backward selection or multimodal overlap measure is used.

The Jacobian based feature selection follows the algorithm described in figure (2.12). For all training data, where the objective variable (v) is known, features $(f_i)_{i=1:N}$ are extracted and their first-order derivatives ($\left|\frac{\partial f_i}{\partial v}\right|$) are numerically calculated. This derivative expresses the slope of change of a feature with the change of the objective variable. Therefore, low slopes (angles $< 20°$) indicate small change in f_i with respect to changes in v, while high slopes (angles $> 70°$) indicate small change in v with respect to changes in f_i. Intermediate slopes ($20° <$ angles $< 70°$) handle features with significant changes with respect to changes in v, making them the best relevant features for the regression task.
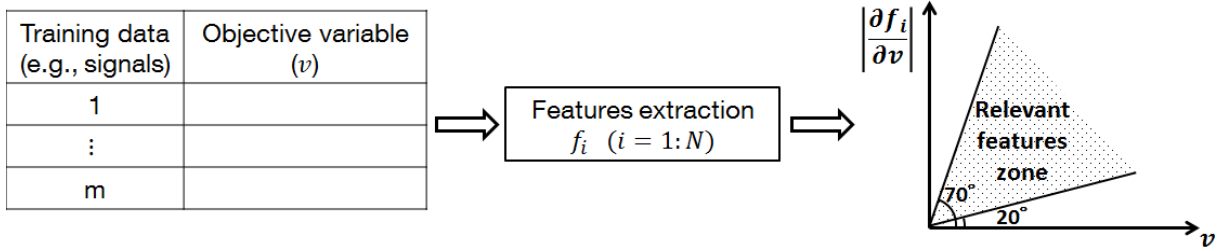
Figure (2.12): The procedure of Jacobian based feature selection approach. Where f_i is the extracted feature for m training data, N is the total number of features, and v is the given objective variable for each training data.

On the other hand, the steadiness based feature selection follows the algorithm given in figure (2.13). The objective variable for classification problems is a predefined class (e.g., $\{C1, C2\}$). If the feature values, assigned to one class, have small standard deviation (σ) and their average is far from that of the other class (with difference d), this feature is relevant. All features are ranked according to their (σ, d) values, followed by a selection of features with higher ranks.
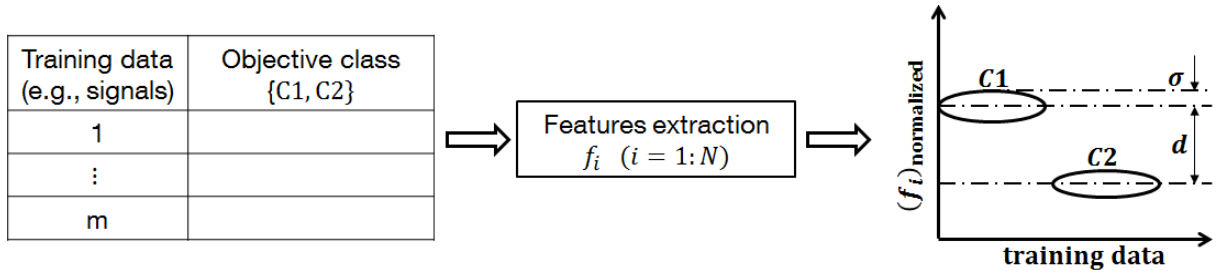
Figure (2.13): The procedure of steadiness based feature selection approach. Where f_i is the extracted feature for m training data, N is the total number of features, $\{C1, C2\}$ are the objective classes, σ is the standard deviation of feature values in one class, and d is the difference between the averages of the two classes values.

Before applying any of the feature selection approaches, the correlation coefficient (r) between each pair of features is calculated by equation (2.33). If $r \geq 0.85$, one of the two features is considered redundant and removed from the features list.

$$ r = \left| \frac{\sum f_i f_j - \frac{\sum f_i \sum f_j}{m}}{\sqrt{\left[\left(\sum f_i^2 - \frac{(\sum f_i)^2}{m} \right) \left(\sum f_j^2 - \frac{(\sum f_j)^2}{m} \right) \right]}} \right| \quad , i \neq j \qquad (2.33) $$

Where f_i and f_j are two different features from the extracted features list, and the summation is performed on the extracted feature values over m training data.

2.2.   Novel edge detection approach

To recognize objects in an image, their boundaries (i.e., edges) should be well detected and clarified. Afterwards, two dimensional features are extracted from the areas enclosed by these boundaries, and used to identify each specified object (e.g., sound pattern in a spectrogram image, amount of fat in meat slice image). A novel edge

detection approach has been developed, which generalizes the process by avoiding the usage of pre-processing "size dependent" filters, and improves the process by working in sub-pixels accuracy, following the procedure which is presented in figure (2.14).

```
┌─────────────────┐   ┌─────────────────┐   ┌─────────────────┐   ┌─────────────────┐
│  Image I(x,y)   │──▶│ Features images │──▶│  Edge detection │──▶│   Hysteresis    │
└─────────────────┘   └─────────────────┘   └─────────────────┘   │  thresholding   │
                                                                   └─────────────────┘
                                                                            │
                                                                            ▼
┌─────────────────┐   ┌─────────────────┐   ┌─────────────────┐   ┌─────────────────┐
│   2D features   │◀──│   Edges image   │◀──│ Flux equilibrium│◀──│  Non-Maximum    │
│   extraction    │   └─────────────────┘   │     check       │   │ suppression with│
└─────────────────┘                         └─────────────────┘   │ sub-pixel accuracy│
                                                                   └─────────────────┘
```
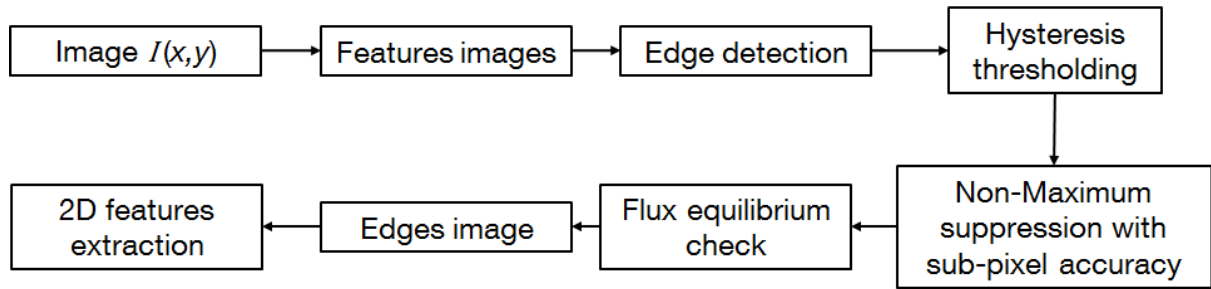
Figure (2.14): Edge detection and 2D features extraction procedure. The intensities of a grey level image $I(x, y)$ are processed to produce its features images. Subsequently, the objects edges are detected in these images. Noisy and weak edges are removed by the Hysteresis thresholding. While remaining edges are suppressed by a Non-Maximum Suppression algorithm and connected through a flux equilibrium check. The dominant 2D features are then extracted from the edges image and implemented to identify the enclosed patterns.

Features images

The objective of developing the features images is to have smoothed version(s) of the original image, avoiding the application of prior smoothing filters and their degrading influences on the contained objects. The algorithm of obtaining features images starts by sliding a $n * n$ mask over columns and then rows of the original image, with a step of one pixel, as described in figure (2.15a). At each mask placement, the intensities of the covered pixels are used to calculate the mask centroid (equation 2.33) and certain feature(s) (e.g., energy (equation (2.34)) or skewness (equation (2.35))), as clarified in figure (2.15b). The feature image $F(x, y)$ is formed by the calculated feature values along all mask placements, as described in figure (2.15c).

$$\text{mask centroid} = \sum_{i=1}^{n^2} i * I_i \setminus \sum_{i=1}^{n*n} I_i \tag{2.33}$$

$$\text{Energy} = \frac{\sum_{i=1}^{n} I_i^2}{n * n} \tag{2.34}$$

$$\text{Skewness} = \frac{\sum_{i=1}^{n^2} I_i * (i - \text{mask centroid})^3}{\sqrt{\left(\sum_{i=1}^{n^2} I_i * (i - \text{mask centroid})^2\right)^3}} \tag{2.35}$$

Where I_i is the grey level intensity value at pixel i, and the back slash means that only quotient of the division is considered. The mask centroid is returned to its corresponding two coordinates (x, y) with respect to the image global axes.
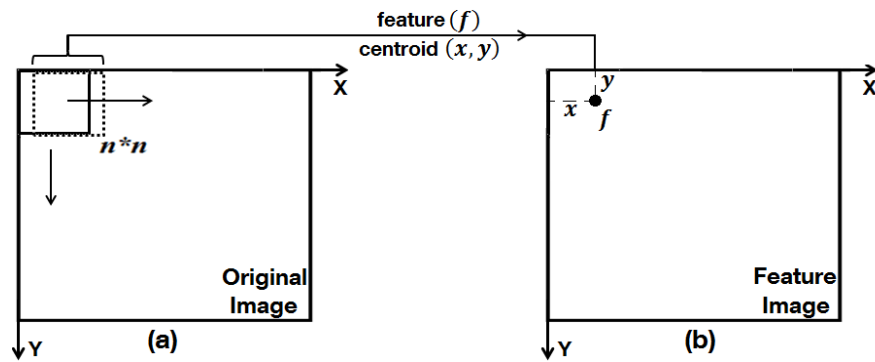
Figure (2.15): Description for the procedure of establishing the feature image. (a) an original image at which a mask of size $(n*n)$ is sliding with a step of one pixel. (b) the feature image formed by feature value (f) and centroid (x, y) of each mask placement. (c) example for an original image (left) and two of its features images, namely energy image and skewness (equation (2.35)) image; respectively.

Edge detection

Edges are detected in areas where the intensity fluctuates sharply, and the more rapid this value changes the stronger the edge is. Therefore, edges can be found either in the gradient image (i.e., first derivative of the feature image) at pixels have local maxima or local minima, or in the second derivative image at pixels of zero crossings, as sketched in Figure (2.16) for noiseless two edges image.
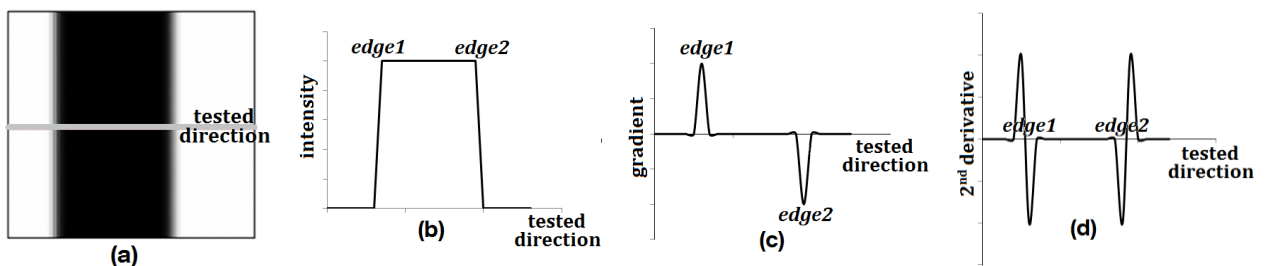
Figure (2.16): (a) A noiseless two edges image. (b) The intensity distribution along the tested direction. (c) The edges are detected in the gradient distribution as the local maxima and local minima points. (d) The edges are detected in the 2nd derivative distribution as the points at which the curve changes its sign

Although edge detection through the 2nd derivative image attains well localized edges, it is highly influenced by image noise and the probability of missing and wrong edges is high. Therefore, in this thesis the gradient based edge detection is considered.

The gradient (G) of a feature image $F(x,y)$ at location (x,y) is a vector of two components, G_x and G_y, which measures how the feature value changes with respect to x and y directions, as given in equations (2.36, and 2.37); respectively.

$$G_x = \frac{\partial F}{\partial x} \tag{2.36}$$

$$G_y = \frac{\partial F}{\partial y} \tag{2.37}$$

Calculating the gradient of feature image is similar to convolving it with two gradient operators, one in x direction and the other in y direction. The commonly used gradient operators are those presented by Roberts, Prewitt and Sobel edge detectors, as described in figure (2.17).
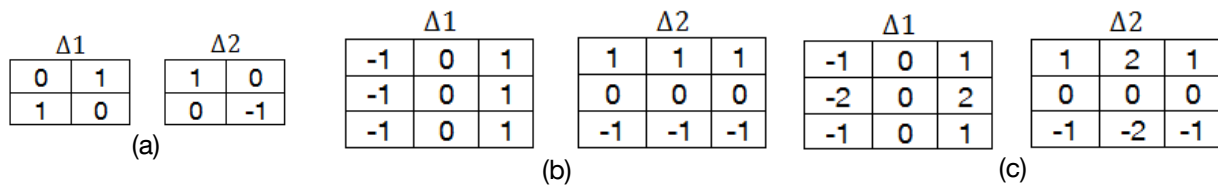

Figure (2.17): Gradient operators of (a) Roberts, (b) Prewitt, and (c) Sobel edge detectors. Δ_1 is the gradient operator in x direction, while Δ_2 is in y direction.

The 3*3 Sobel operators are selected because their basis finite difference scheme has a second degree order of error, $O(\delta x^2)$ and $O(\delta y^2)$, in the calculation of gradient in x and y directions; respectively. Where δ is the distance between two pixels in the indicated direction. Therefore, at any pixel(i,j), the values of G_x and G_y are calculated by convolving the 3*3 surrounded mask with Δ_1 and Δ_2 of the Sobel operators; respectively, as shown in figure (2.18).
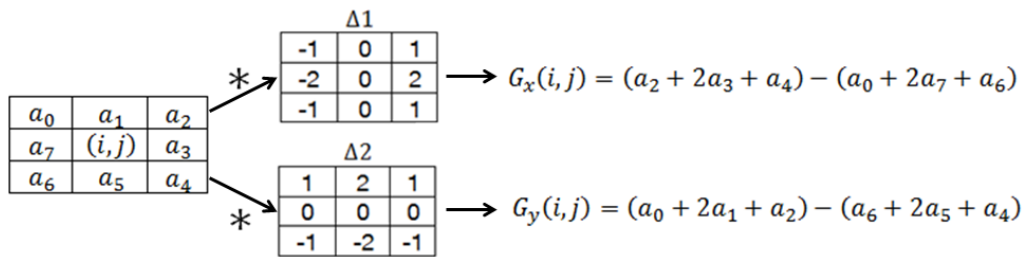
Figure (2.18): The convolution operation between Sobel operators and 3*3 mask to calculate the gradient components at the middle pixel.

The edges image is thereafter formed by obtaining the magnitude (Gm) of the gradient components at each pixel, as explained in equation (2.38). This magnitude represents the edge strength at the associated pixel. The edge direction (θ) is perpendicular to the orientation of the gradient components, as given by equation (2.39) and graphically shown in figure (2.19).

$$G_m = \sqrt{G_x^2 + G_y^2} \tag{2.38}$$

$$\theta = tan^{-1}\left(\frac{G_y}{G_x}\right) + \frac{\pi}{2} \tag{2.39}$$

Figure (2.19): Diagram for the relation between (edge strength and direction) with (gradient magnitude and orientation).

Hysteresis thresholding

At this stage, the edges image contains weak edges result from the weak local peaks in the gradient image. Therefore, Hysteresis thresholding [62] defines two thresholds, T_{high} and T_{low}. These thresholds usually equal 90% and 10% of $max(G_m)$; respectively. Pixel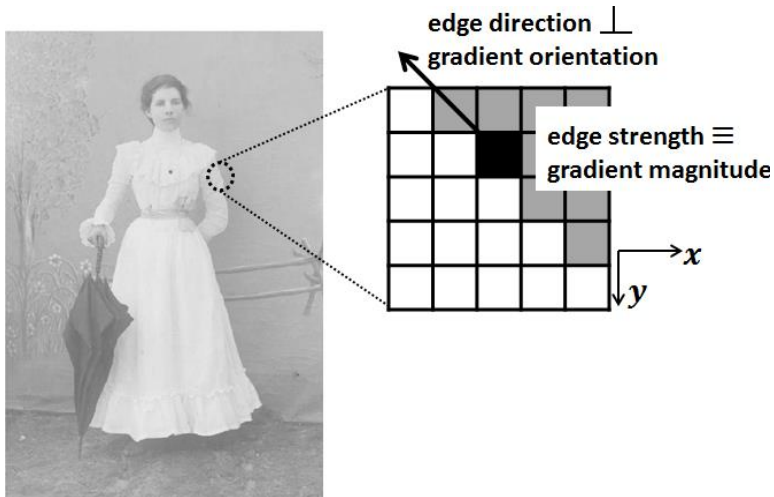s with $G_m > T_{high}$ are called strong edges and preserved, while those with $G_m \leq T_{low}$ are called weak edges and discarded. All other pixels are called candidate pixels. If a pixel is candidate, the chain of its connected pixels is followed in both directions as long as $G_m > T_{low}$. The candidate pixel is preserved as edge only if the starting pixel on its chain is connected to a strong edge.

Non-maximum suppression with sub-pixel accuracy

The non-maximum suppression method is applied to remove unnecessary edges by suppressing the non-maximum magnitude in each cross section of the edge direction in their local neighbourhood. In depth, the edge strength at a pixel is checked with its two neighbours in a direction normal to the edge direction. If this strength is greater than both, its magnitude value set to one, otherwise it is set to zero (i.e., suppressed), as described in figure (2.20).

$$\text{if } a \geq \max(a_1, a_2)$$
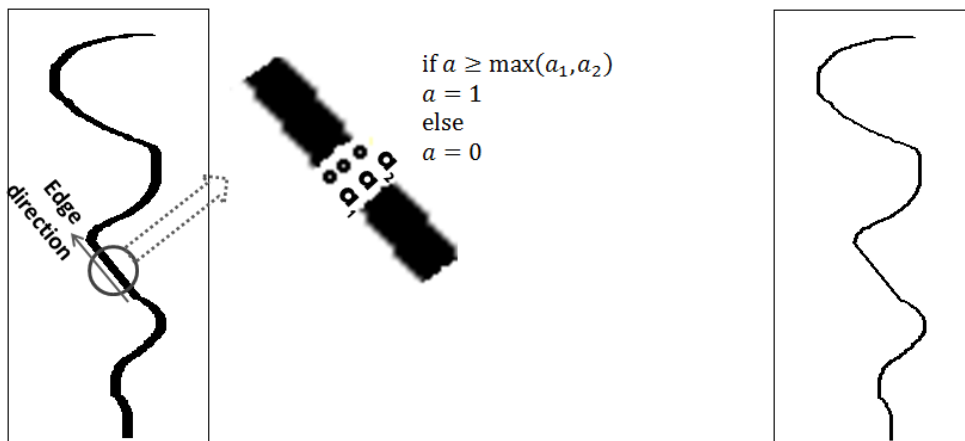$$a = 1$$
$$\text{else}$$
$$a = 0$$

Figure (2.20): Explanation for how the non-maximum suppression algorithm is performed to thin out edges image (left) to "one-pixel" width edges image (right)

As a result, the edges image is transformed to a binary image (i.e., {0,1}), and all thick edges are thinned out to a "one pixel" width, as shown in the edges image of figure

(2.21). Thinning the edges to one pixel width increases the accuracy in edges localization and boundaries recognition.

Figure (2.21): The final binary edges image for the original grey level image in figure (2.15c), calculated on its energy image (as feature image) with mask size = 5*5.

In most cases, edges and their perpendicular axes (i.e., gradients orientations) are not directed to the center of the neighboured pixel, which means they usually have values not equal to {0,∓45, ∓90, ∓135,180}, as schematically described by example given in figure (2.22). Therefore, an interpolation is made to estimate the sub-pixels strengths (a_1 and a_2) before applying the non-maximum suppression algorithm. A quadratic interpolation, given by equation (2.40), using the strengths of the three neighboured pixels gives sufficient estimation of a_1 and/or a_2.

$$G_m = c_1\bar{\theta}^2 + c_2\bar{\theta} + c_3 \tag{2.40}$$

Where $\bar{\theta} = \theta - \frac{\pi}{2}$ is the gradient orientation, and the coefficients c_1, c_2 and c_3 are calculated by strengths and directions of the neighboured pixels.
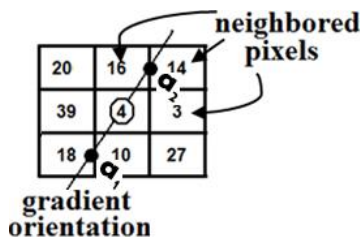
Figure (2.22): Representation for how the gradient orientation does not pass by the centers of the pixels, and passes by an intermediate pixel which can be calculated by quadratic interpolation among the neighbored pixels

Flux equilibrium check

The edges image may still suffer from discontinuities, in the form of zero pixel(s) intersects the edges along their directions. Restoring the missing edges data is very important in many edge detection applications, and is magnified in life science applications where the boundaries are huddled and may be easily overlap or interconnected openings take place by imaging. The flux equilibrium condition states that the derivative of edge strength along its direction should be zero everywhere, as given in equation (2.41). Found discontinuities are filled if they are located between two edges, and in their direction, as explained in figure (2.23).

$$\frac{dE}{d\theta} = 0 \tag{2.41}$$

31

Figure (2.23): A disconnected edge and how it is filled using the flux equilibrium condition.

During this work, several features images are developed during this research and investigated with the aim of having the most accurate final edges image. For example, the crest factor image (equation (2.42)) is efficiently used to obtain the edges of sound patterns in spectrogram ima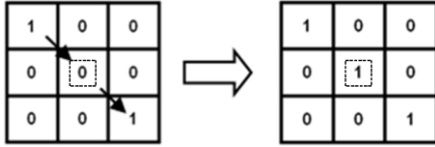ges, while a combination between energy and skewness images are accomplished to obtain accurate edges image for biological tissue images.

$$\text{Crest factor} = \frac{\max(I)}{\frac{1}{n^2}\sum_{i=1}^{n^2} I_i} \tag{2.42}$$

Efficiency of the proposed method

The conventional approach in gradient based edge detectors, such like those of Roberts, Prewitt, and Sobel [63], is to choose a threshold value at which the pixels with gradient magnitudes greater than or equal to this threshold are considered edges; otherwise, they are not edges. Since they do not smooth the image before gradient calculation, they are only suitable for well contrasted noiseless images. Meanwhile, Canny edge detection method considers the edges are step edges corrupted by additive Gaussian noise. It starts by applying Gauss filtration (equation 2.43) to the original image, before finding the gradient image. The edges strengths and directions are then calculated by equations (2.38) and (2.39); respectively. The hysteresis thresholding algorithm is used to name edges and non-edges and the multi pixel wide edges are thinned to single pixel width using the non-maximum suppression algorithm and

$$Gauss(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.43}$$

One drawback in Canny edge detector is the sensitivity of the output edges image to the size (σ) of the implemented Gaussian filter. The results of applying Canny detector on two different natures images with two values of σ are presented in figure (2.24), implying the absence of a general optimum value for the filter size.
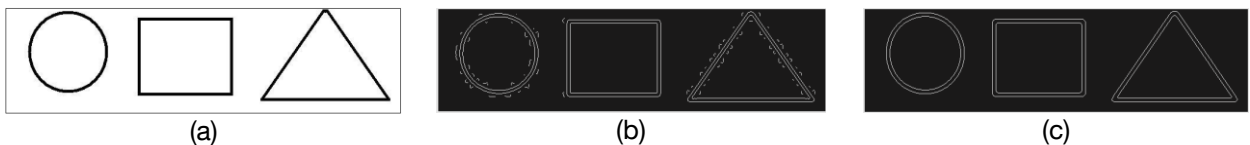

(a)   (b)   (c)

(d)   (e)   (f)

Figure (2.24): Canny detector outputs for two grey scale images indexed by (a) and (d). (b) and (e) are the edges images when σ =1, while (c) and (f) the edges images when σ = 2.

The final edges image obtained by the developed edge detection algorithm is compared with those of the conventional gradient based detectors for a finger print image in figure (2.25). Results indicate the efficiency of the developed algorithm to produce more accurate and detailed edges image.

(a)   (b)   (c)   (d)   (e)   (f)
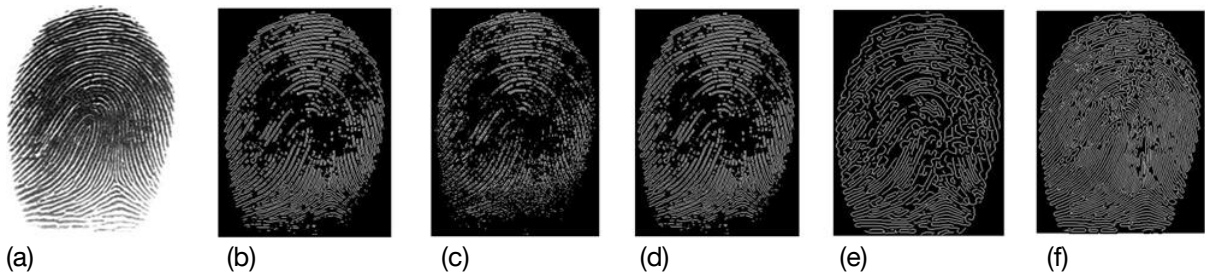
Figure (2.25): Revisit to finger print image of figure (1.16) updated by the results of applying the developed edge detection approach. (a) original finger print image, (b) edges image by 3*3 Prewitt edge detector, (c) edges image by 2*2 Roberts detector, (d) edges image by 3*3 Sobel detector, (e) edges image by Canny detector with filter size of 1.5, and (f) edges image by the developed method calculated on a combination of energy and skewness images (as feature images) [mask size = 5*5, flux check matrix size = 3*3, combination coefficient = 0.75]

2D features extraction

The final edges image identifies the boundaries of the involved objects in the original image. For simple differentiation, image pixels (x, y) enclosed by a boundary (B) are given a general shape function $f(x, y)$ value of 1, while other pixels are assigned with value of 0, as described in figure (2.26).

$$f(x, y) = 0$$

$$f(x, y) = 1$$

$$B$$
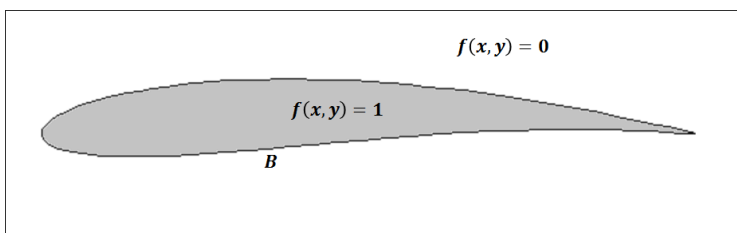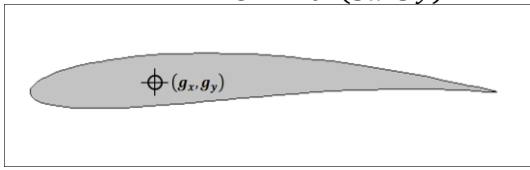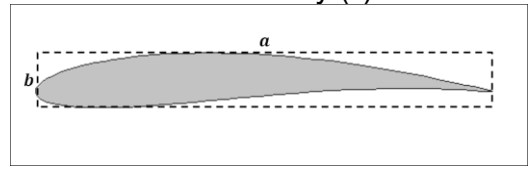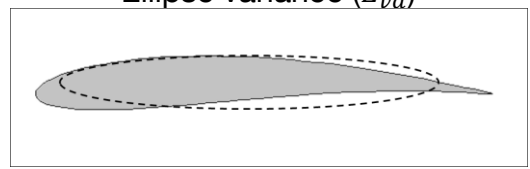
Figure (2.26): An edges image of one object enclosed by a boundary (B). The general shape function $f(x, y)$ is defined to be 1 for $(x, y) \in B$ and 0 otherwise.

Two dimensional features can be extracted from this edges image and implemented in pattern recognition tasks [63, 64, 65]. Examples of these tasks are calculating the area

enclosed by objects to get the fat contents in meat slice image, or to identify the existence of known sound pattern in a spectrogram image. Some of the widely implemented image features are presented in table (2.3).

Table (2.3): some of the commonly implemented 2D features to define objects in edges image. Where N is the number of image pixels enclosed in the object, α is the angle between the axis of least inertia and the x axis. a and b are the length and width of the minimum bounding rectangle contains the object. The ellipse variance introduces an ellipse that has an equal covariance matrix $C_{ellipse}$ as the object.

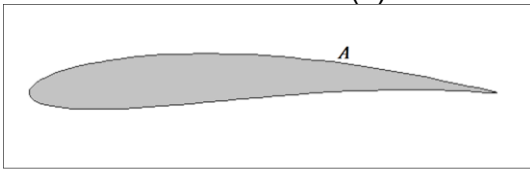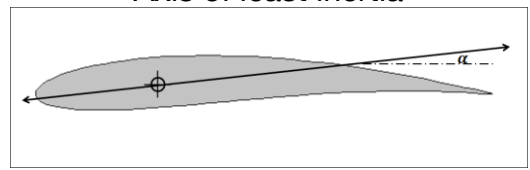| Feature	Equation
Center of gravity (g_x, g_y)	$g_x = \frac{1}{N}\sum_{i=1}^{N} x_i$, $g_y = \frac{1}{N}\sum_{i=1}^{N} y_i$ $(x_i, y_i) \in \{(x_i, y_i)\|f(x_i, y_i) = 1\}$
Contour Area (A)	$A = \frac{1}{2}\left\|\sum_{i=1}^{N-1}(x_i y_{i+1} - x_{i+1}y_i)\right\|$
Axis of least inertia	$p = \sum_{i=1}^{N} x_i^2,\ q = 2\sum_{i=1}^{N} x_i y_i,\ r = \sum_{i=1}^{N} y_i^2$ $\alpha = \frac{1}{2}\tan^{-1}\left(\frac{q}{p-r}\right)$
Eccentricity (e)	$e = \frac{a}{b}$
Ellipse variance (E_{va})	$V_i = \begin{pmatrix} x_i - g_x \\ y_i - g_y \end{pmatrix},\ d_i = \sqrt{V_i^T C_{ellipse}^{-1} V_i},\ \mu = \frac{1}{N}\sum_{i=1}^{N} d_i,$ $\tau = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \mu)^2}$ $E_{va} = \frac{\tau}{\mu}$

Another common 2D feature is the chain code, which describes an object by a sequence of line segments with a given orientation [66]. The movement along the object boundary is encoded by 8 connectivity scheme $\{i\| i = 0,1,2,\dots,7\}$, as shown in the example given in figure (2.27).
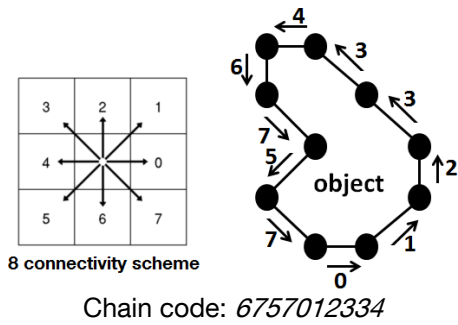
Figure (2.27): Example for chain code representation of an object boundary in 8 connectivity scheme.

Chain code: *6757012334*

2.3. Enriched spectrogram enhancement approach

From the temporal and spectral features, signal structure and performance is completely represented in time and frequency domains; *separately*. Thus, the need for a combined time-frequency representation stemmed from the inadequacy of either time domain or frequency domain analysis to fully describe the nature of signals. In this section, the two common time-frequency representations; spectrogram and scalogram, are briefly discussed and compared with respect to their time and frequency generated resolutions. Afterwards, the spectrogram is reasonably selected and its dynamic range is limited to only explore the involved high energetic sound patterns. Spectrogram, as a 2D image, is then processed by the developed edge detection approach of section (2.2), to accurately define the boundaries of the enclosed sound patterns and eliminate the attached noise, following the procedure given in figure (2.28).
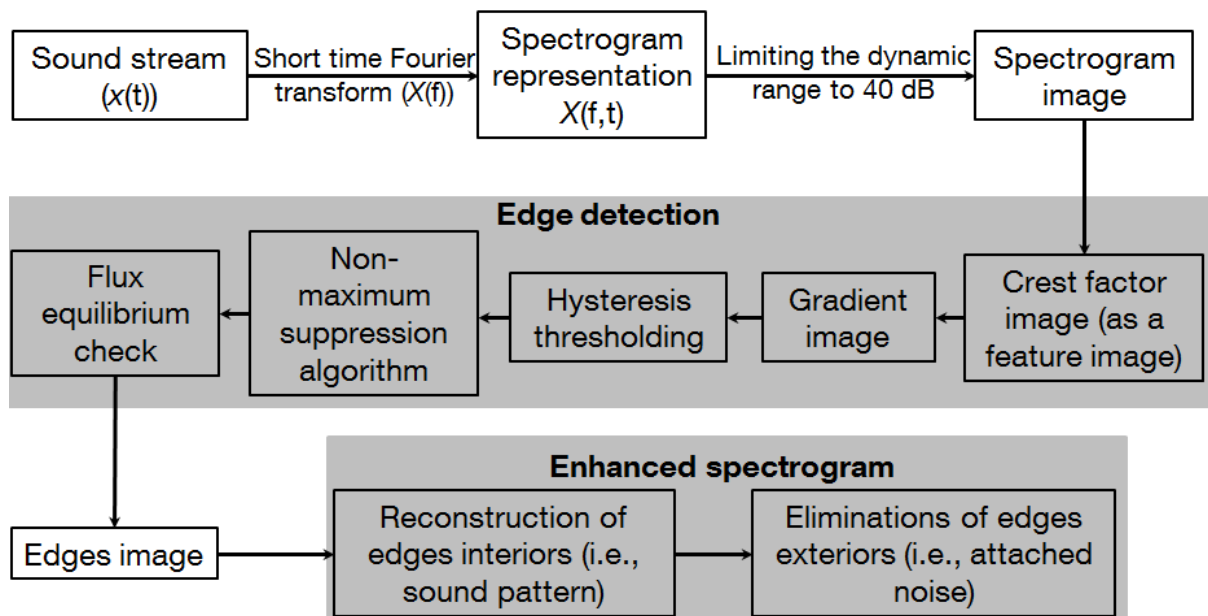
Figure (2.28): Procedure for the developed spectrogram enhancement approach. The spectrogram is limited to 40 dB range which perceives the human speech range. The developed edge detection approach of section (2.2) is applied to the crest factor image, as a feature image. The produced edges image defines sound patterns to be reconstructed and attached noise to be eliminated.

Combined time-frequency representation of sound (2D representation)

A time-frequency distribution of a signal provides information about how the spectral content of the signal evolves with time, thus providing an ideal tool to dissect, analyse

and interpret non-stationary signals. This is commonly performed by applying FFT to short and overlapped segments of the signal in an algorithm called Short Time Fourier Transform (STFT), as described in figure (2.29).
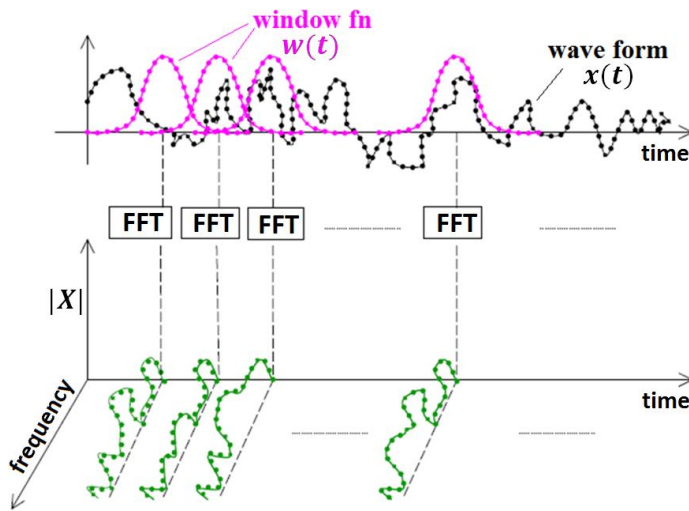
Figure (2.29): Spectrogram time-frequency representation obtained by dividing the signal into overlapped windowed segments and calculates the FFT for each segment by (equation (2.20)). FFT coefficients form one curve in the spectrogram which is located at the center of the processed segment.

The obtained representation is called spectrogram, and controlled by the selected window function and overlapping percentage between segments. The fundamental weakness of spectrogram involved in its subject to the Heisenberg Uncertainty Principle [67], given in equation (2.44). This principle implies that decreasing the segment length in time (increasing time resolution) must result in an increase in the deviation in frequency (decreasing frequency resolution) and vice versa.

$$\Delta f . \Delta t \geq \frac{1}{4\pi} \tag{2.44}$$

The other commonly used method to obtain the time-frequency representation is Continuous Wavelet Transform (CWT). It compares the signal with dilated and time shifted versions of a mother wavelet, as given in equation (2.45) and described in figure (2.30).

$$C(a, b) = \sum_{n=1}^{N} x(n) \frac{1}{\sqrt{|a|}} \emptyset \left(\frac{n - b}{a} \right) \tag{2.45}$$

where $\emptyset \left(\frac{n-b}{a} \right)$ is the selected mother wavelet, a is the scale $\left(= \frac{f_c}{2\pi f} \right)$, b is the space (i.e., shift in time), f_c is the wavelet center frequency, and C is the wavelet coefficient.
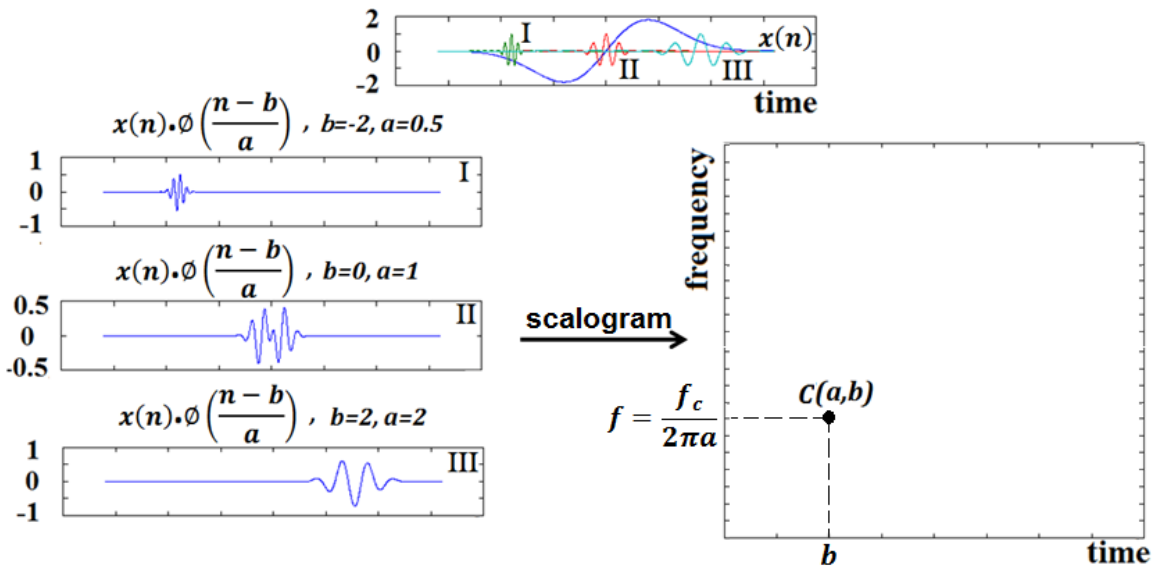
Figure (2.30): Scalogram time-frequency representation obtained by finding the correlation coefficient between the signal and a Morlet mother wavelet $[\emptyset(t) = cos(5t) * e^{-t^2/2}]$ at specific space and scale values. The wavelet is shifted until the whole signal is covered, then it is stretched and the process is repeated.

The produced time-frequency representation is called scalogram, and controlled by the selected mother wavelet, the step in space, and the step in scale [68]. From figure (2.30), at higher frequency (low scale), the resolution in time is good but in frequency is low, because the wavelet is well localized in time but poorly localized in frequency. While at low frequency (high scale), the frequency resolution is good and the time resolution is poor.

A schematic comparison between spectrogram and scalogram with respect to their time and frequency resolutions is presented in figure (2.31). Consequently, spectrogram is selected as a reasonable time-frequency representation due to its constant time and frequency resolutions over the whole frequency range and time length of the signal.
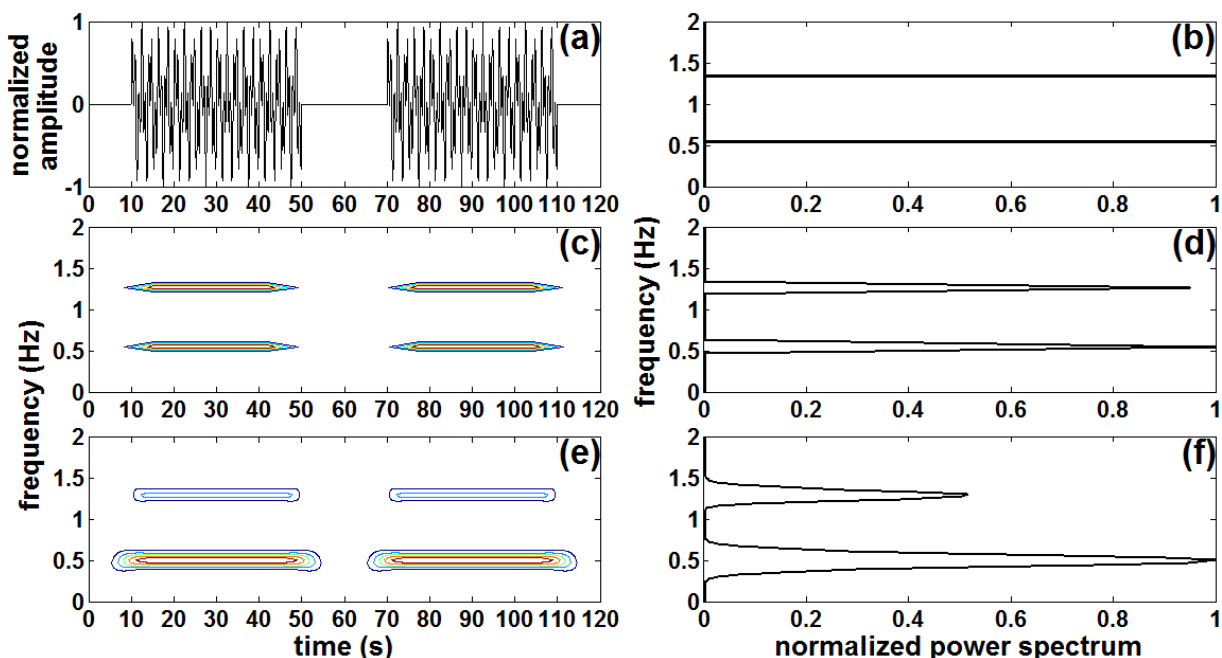
Figure (2.31): Schematic diagram comparing the time and frequency resolutions achieved in

37

spectrogram and scalogram. (a) a signal constituted by two sinusoids of two frequencies 0.5 and 1.3 Hz, sampled at a rate of 4 Hz. (b) the exact (analytical) power spectrum distribution. (c) contour plot of the spectrogram representation, and (d) its corresponding power spectrum distribution [Blackman window, segment length = 12 s, 50% overlapping]. (e) contour plot of the scalogram representation, and (f) its corresponding power spectrum distribution [Morlet wavelet, 300 scales with 0.05 scale spacing, wave number = 10].

In general, sound patterns of the spectrogram appear immersing in attached noise, as described in figure (2.32). These noise are not only the base noise attached to the time domain signal, but also those generated during Fourier transformation. Therefore, filtering the signal in time domain will not ensure clean spectrogram represntation.
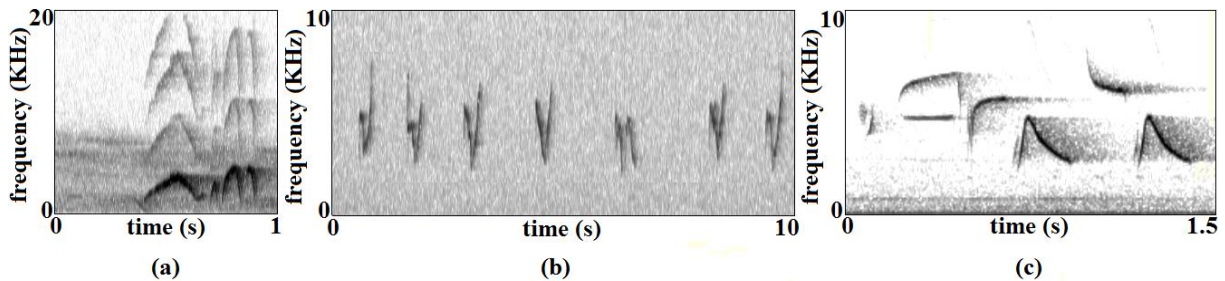
Figure (2.32): Spectrogram of several streams displaying the sound patterns are immersed in attached noise. (a) spectrogram of *Sayornis phoebe* bird, (b) spectrogram of *Phylloscopus inornatus* bird, and (c) spectrogram of *Thyromanes bewickii* bird.

Limiting the dynamic range

Although whisper cannot be heard in loud surroundings, spectrogram will contain all details about whisper as well as loud sound. The power spectrum value (P) is obtained from the relation ($P = 20\,log\,|X|$). Since whisper has low power values, it can be removed by limiting the dynamic range of the spectrogram to cover only high power values. Most life science sound sources are expanded or slowed to the human speech range, which is normally perceived over 40 dB dynamic range [69]. Therefore, all points with power values outside the range [$P_{max} - 40 : P_{max}$] are removed, as addressed in figure (2.33) for the spectrogram of *Rhinolophus hipposideros* bat. The procedure starts by subtracting ($P_{max} - 40$) from all spectrogram values, followed by setting any resultant negative power values to zero. Afterwards, the value of ($P_{max} - 40$) is re-added.

(c)
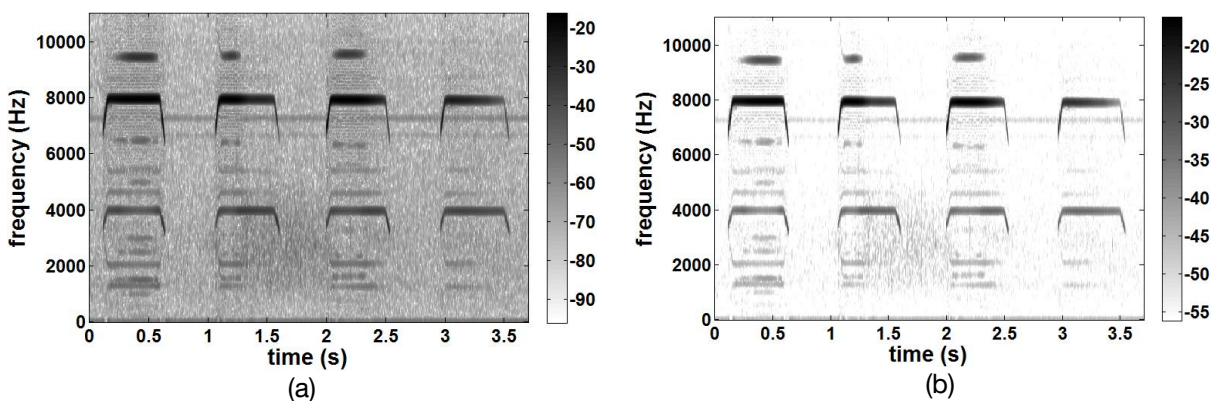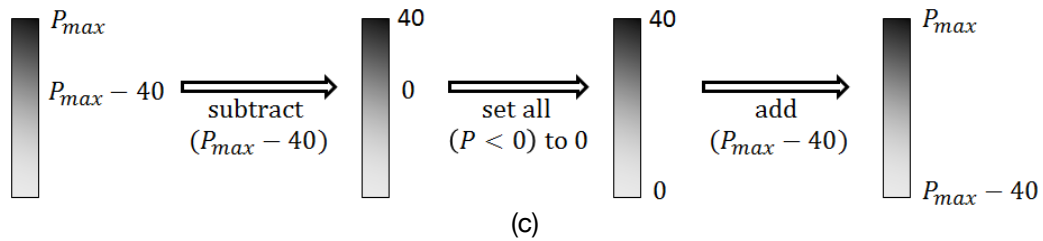
Figure (2.33): (a) The spectrogram of *Rhinolophus hipposideros* sound stream with expansion factor of 13 [segment length = 1% of the total signal length, 90% overlapping, and Bartlett window function]. (b) The resultant spectrogram after its power values are limited to 40 dB dynamic range, following the algorithm given in (c).

Edge detection of the sound patterns

The crest factor image is extracted as a feature image for the limited dynamic range spectrogram following equation (2.42), as shown in figure (2.34).
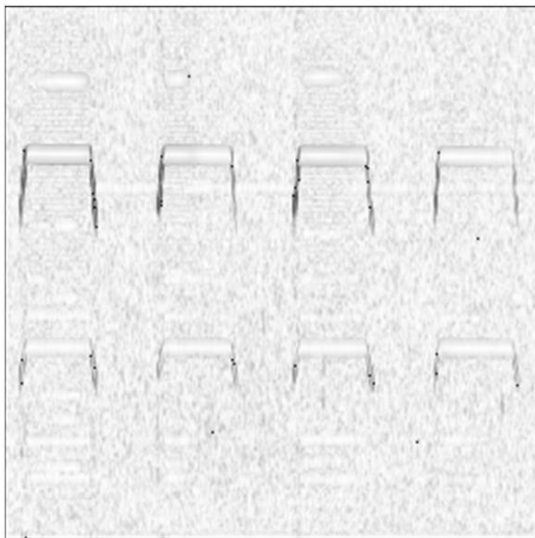
Figure (2.34): The crest factor image calculated by equation (2.42) on 5*5 mask slides over the limited dynamic range spectrogram of figure (2.33b) with step of one pixel.

The developed edge detection approach in section (2.2) is applied to accurately define boundaries of the sound patterns in the crest factor image, producing the edges image displayed in figure (2.35), which separates the patterns from their surrounding noise. However, the edges do not provide information about where exactly are the inner of the patterns and where are their surroundings.
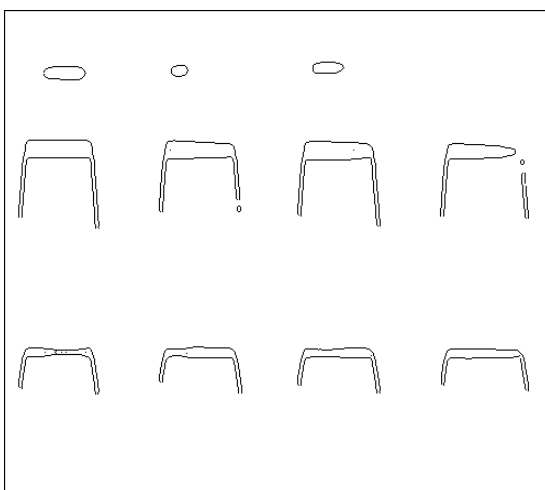
Figure (2.35): The edges image of the limited dynamic range spectrogram for Rhinolophus hipposideros, by the developed edge detection method of section (2.2), applied to the crest factor image (as feature image) [mask size = 5*5, flux check matrix size = 3*3].

39

Reconstructing the spectrogram

To define the patterns and restore their power values and define the surroundings noise and eliminate their power values, a classification algorithm was set and applied for each row and afterwards each column of the edges image. This algorithm states that if pixels among two adjacent edges in one row (column) originally have average power value higher than or equal to the original average power values at these two edges, these pixels belong to a pattern and their power values are restored from the original spectrogram. Otherwise, these pixels belong to the surrounding noise and their power values are eliminated, as presented below and explained in figure (2.36)

$$I = \{i_1, i_2, \ldots, i_n\}, J = \{j_1, j_2, \ldots, j_m\}$$
$$\text{for } k = 1:n-1$$
$$Pixels(i_k:i_{k+1}) = \begin{cases} \text{sound pattern} & ; mean(Power(i_k:i_{k+1})) \geq mean(Power(i_k, i_{k+1})) \\ \text{surrounding noise} & ; mean(Power(i_k:i_{k+1})) < mean(Power(i_k, i_{k+1})) \end{cases}$$
$$\text{end}$$

$$\text{for } k = 1:m-1$$
$$Pixels(j_k:j_{k+1}) = \begin{cases} \text{sound pattern} & ; mean(Power(j_k:j_{k+1})) \geq mean(Power(j_k, j_{k+1})) \\ \text{surrounding noise} & ; mean(Power(j_k:j_{k+1})) < mean(Power(j_k, j_{k+1})) \end{cases}$$
$$\text{end}$$
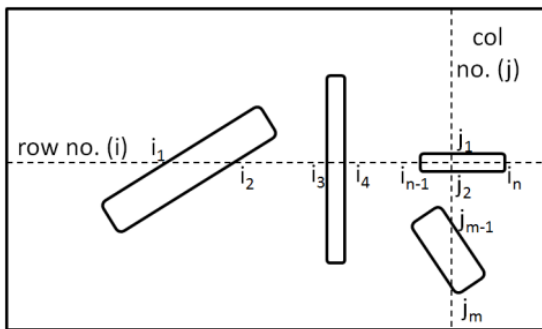
Figure (2.36): Descriptive edges image for the classification algorithm which defines sound patterns from the surrounding noise. *i* is a counter on the image rows, while *j* is a counter on the image columns.

As a result, the enhanced spectrogram is generated, as shown in figure (2.37).
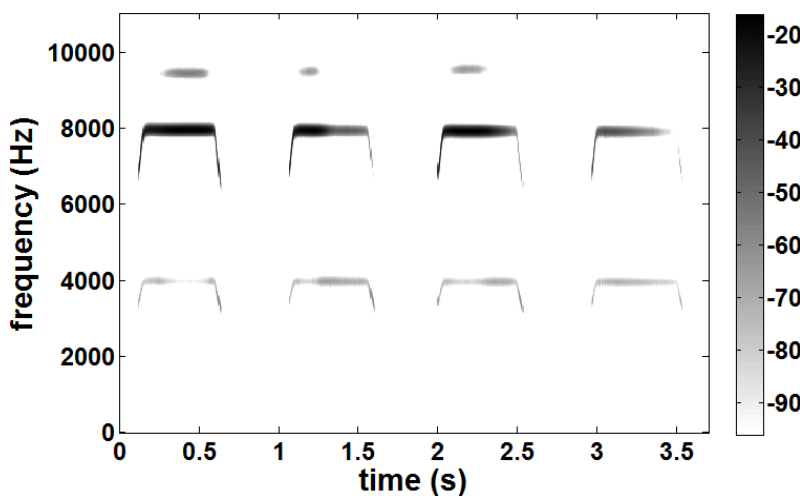
Figure (2.37): The enhanced spectrogram by the proposed method revealing the sound patterns and eliminating their surrounding noise.

Meanwhile, figure (2.38f) is the enhanced spectrogram of *Sitta Canadensis* bird by the proposed method, in comparison to other enhancement approaches which are presented in figure (1.18) and repeated here as figure (2.38b to 2.38e).
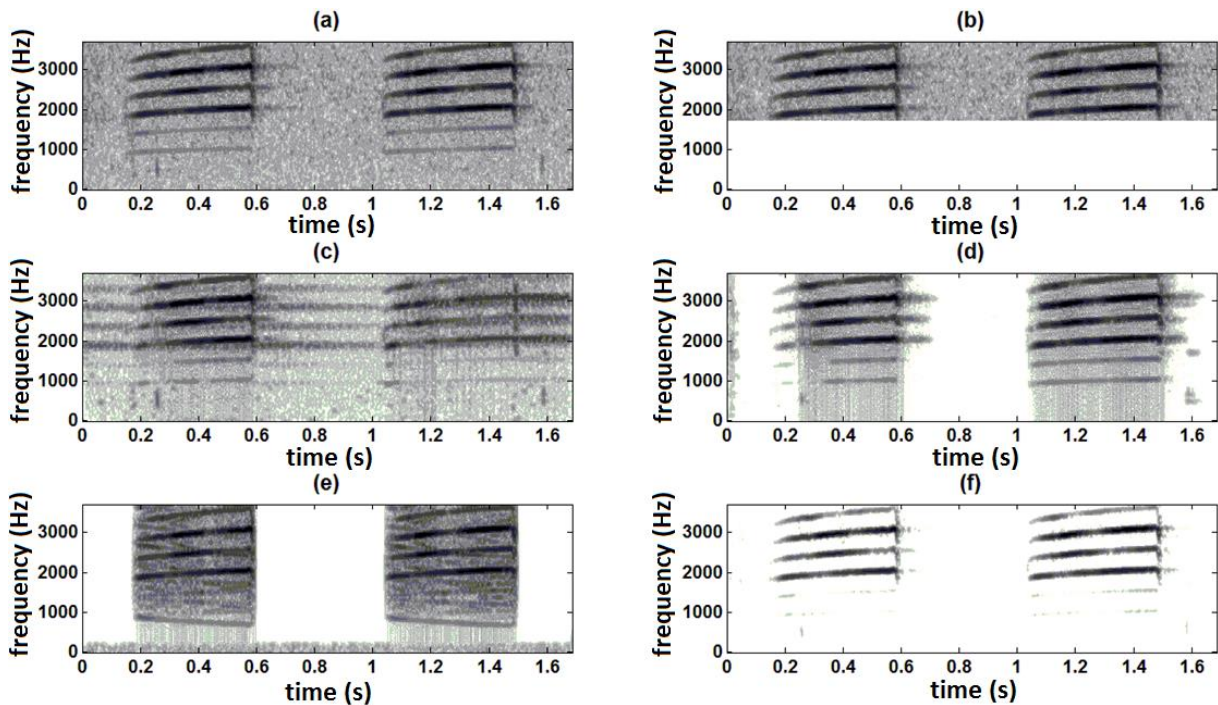
Figure (2.38): Revisit to the spectrogram enhancement of *Sitta Canadensis* bird call of figure (1.18) updated by the results of applying the developed approach. (a) One of *Sitta Canadensis* bird calls at SNR = 20 dB, and its enhanced spectrogram by (b) band pass, (c) multi-band spectral subtraction, (d) Wiener filter, (e) wavelet packet decomposition, and (f) the proposed method.

Since the enhancement process considered the spectrogram as an image and applied an accurate edge detection approach, the resultant spectrogram preserves the sound patterns into their almost original temporal and spectral locations. This is very important issue for any further pattern recognition assignment based on this enhanced spectrogram.

2.4.   Improved speed of sound calculation with phase shift correction

Speed of sound (*USV*) is the main ultrasound signal parameter, which is used to monitor variations in food samples in noncontact and non-destructive testing. When signal is excited from an ultrasound transducer, its main pulse propagates in the sample up to the other end, at which the signal is reflected once more towards the transducer forming its first echo. The process is repeated until the signal is fully damped into the sample, after several echoes are being sensed by the transducer as clarified in figure (2.39).
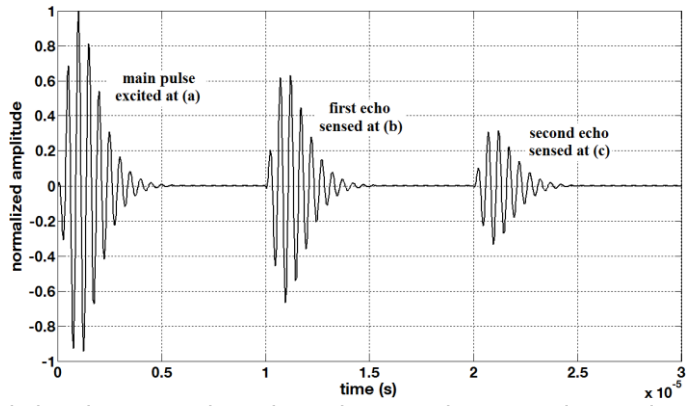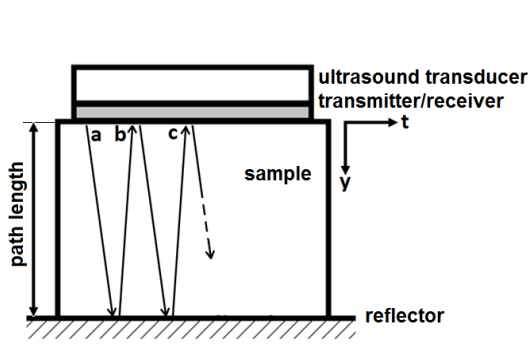
Figure (2.39): An example for ideal ultrasound signal propagation where the transducer sends a main pulse into the sample, and receives its reflected echoes until the signal is fully damped

Time of flight (TOF) is defined as the interval between two following echoes, and its corresponding USV is double of path length, d, taken by the signal, divided by this TOF, as explained in equation (2.46).

$$USV = \frac{2 * d}{TOF}$$

(2.46)

For ideal ultrasound signals, as the one given in figure (2.39), echoes are clearly separated and there is no phase shift among them. Whereas estimating an accurate value for TOF (and USV) in real signal is not an easy task, due to the difficulties in determining an exact start of the echoes, as shown in figure (2.40).
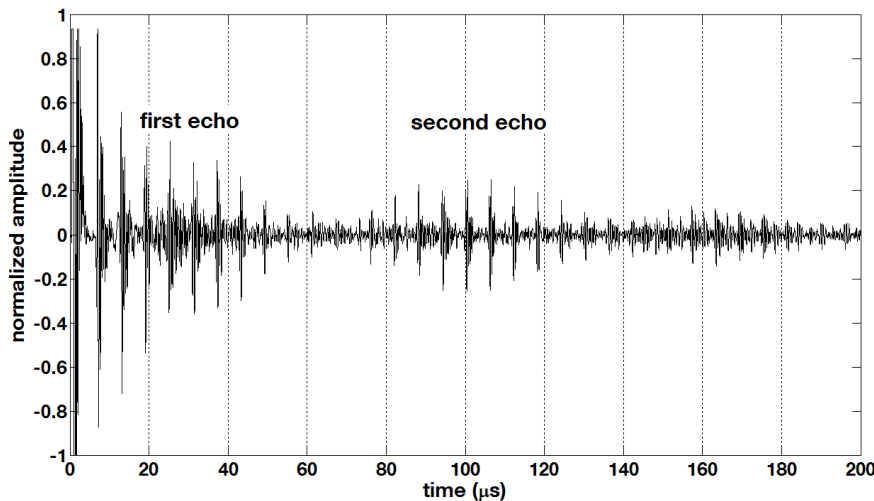
Figure (2.40) ultrasound signal propagation in demineralized water at 10 ˚C. Only received echoes are sensed with the exclusion of the main pulse part.

Signal amplitudes, x_1 and x_2 , at the starting times of first and second echoes, t_1 and t_2, can be mathematically represented by equations (2.47 and 2.48); respectively [70].

$$x_1(t_1) = A_{r1}(t_1)^m e^{-\frac{t_1}{u}} \cos(2\pi f_d t_1 + \varphi_1)$$

(2.47)

$$x_2(t_2) = A_{r2}(t_2)^m e^{-\frac{t_2}{u}} \cos(2\pi f_d t_2 + \varphi_2)$$

(2.48)

Where A_r, f_d, and φ are the echo amplitude, dominant frequency and phase shift; respectively, while m models the initial finite slope of the echo and u determines the final slope. Both m and u are parameters that depend on the type of the ultrasound

transducer. Applying some mathematical operations to equation (2.48), it can be rewritten as:

$$x_2(t_2) = A_{r2}(t_2)^m e^{-\frac{t_2}{u}} \cos(2\pi f_d t_2 + \varphi_2 - \varphi_1 + \varphi_1)$$

$$x_2(t_2) = A_{r2}(t_2)^m e^{-\frac{t_2}{u}} \cos\left(2\pi f_d \left(t_2 + \left[\frac{\varphi_2 - \varphi_1}{2\pi f_d}\right]\right) + \varphi_1\right) \tag{2.49}$$

Thus, TOF between the first and second echoes is calculated by subtracting arguments of equations (2.49) and equation (2.47), as explained in equation (2.50).

$$TOF = \left(t_2 + \left[\frac{\varphi_2 - \varphi_1}{2\pi f_d}\right]\right) - t_1 = t_2 - t_1 + \left[\frac{\varphi_2 - \varphi_1}{2\pi f_d}\right] \tag{2.50}$$

The third term in equation (2.50) is the phase correction term, which is eliminated for ideal signal case, because $\varphi_2 = \varphi_1$. In this thesis, a method is proposed to accurately determine the values for t_1, t_2, φ_1, and φ_2 based on spectral analysis. The spectrogram shown in figure (2.41a) is obtained for the ultrasound signal of figure (2.40), after being divided into overlapped segments with high time resolution and multiplied by suitable window function ensures less spectral leakage (e.g., Hanning window). Afterwards, the spectrogram information at the dominant frequency is separated. This information is basically the FFT coefficients (X's) at this dominant frequency, and can be parted into power spectrum (P) and phase shift (φ) distributions, as shown in figures (2.41b and 2.41c), and explained in equations (2.51 and 2.52); respectively.

$$P = |X| = \sqrt{real(X)^2 + imag(X)^2} \tag{2.51}$$

$$\varphi = \langle X \rangle = \tan^{-1}\left(\frac{imag(X)}{real(X))}\right) \in [-\pi, \pi] \tag{2.52}$$

Where $real(X)$ and $imag(X)$ are the real and imaginary parts of the Fourier coefficient; respectively.
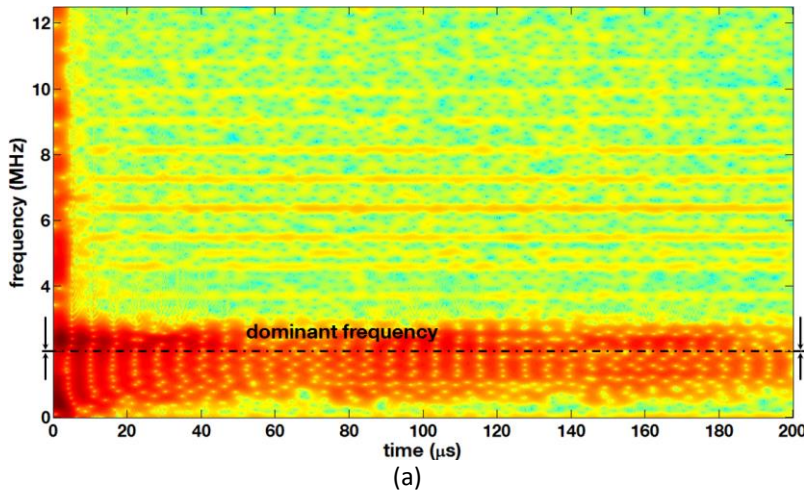
Figure (2.41): (a) Spectrogram for the ultrasound signal of figure (2.40) with 2MHz dominant frequency, calculated by STFT [segment length = 20 μs, Hanning window, and 90% overlapping]. (b) Instantaneous power spectrum distribution at the dominant frequency in voltage-root mean square (Vrms) and (c) phase shift values at the detected times for start of first and second echoes, in radians (rad).

43

(b)

(c)
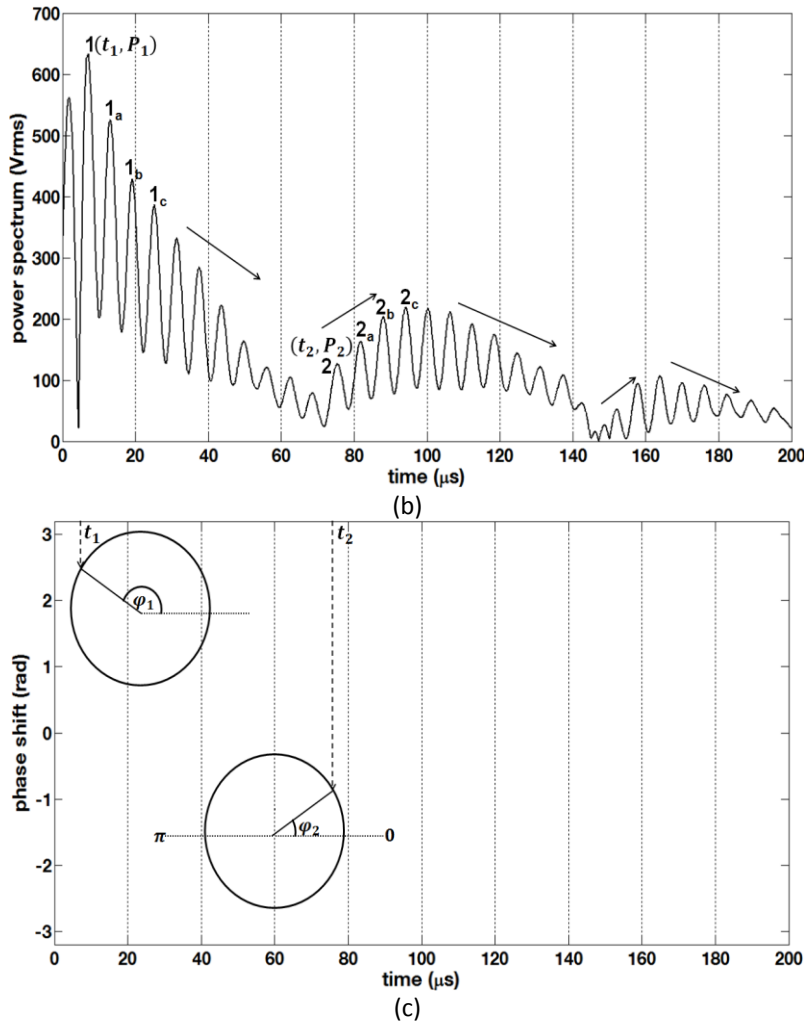
The power spectrum distribution in figure (2.41b) clarifies how the signal is developed in the sample with details about its instantaneous energy variation, arrival times of the consequent echoes, attenuation and damping behaviour. In depth, point 1 is at maximum power spectrum value and indicates the start of first echo (t_1, P_1), while its following peaks at $(1_a, 1_b, ...)$ represent the gradual loss of the intensity (i.e., attenuation) with time. Meanwhile, point 2 is the point at which power spectrum starts to increase again after damping of the first echo, and indicates starting of the second echo (t_2, P_2). Subsequent peaks $(2_a, 2_b, 2_c)$ represent the gradual growth for the arrivals of leading constructive chirps of the second echo to the transducer, while $(2_d, 2_e, ...)$ represent the trailing gradual loss of the second echo intensity. Further echoes are functions of the dramatically damping of the signal in the sample; therefore they are less important for extraction and calculation of signal parameters. Meanwhile, from figure (2.41c), phase shifts of the first and second echoes, φ_1 and φ_2, are obtained at the agreeing times of points 1 and 2, respectively. Finally, TOF can now be calculated by equation (2.50) and its USV is computed with the aid of equation (2.46).

Results of the proposed method are compared to reference speed of sound values presented by [71], for demineralized water. These reference values are calculated through an empirical fifth-degree polynomial equation (equation (2.53)), and valid in temperature (T) range [0 - 95] ˚C with an accuracy of 0.02 m/s.

$$USV = 1.402385 * 10^3 + 5.038813T - 5.799136 * 10^{-2}T^2 + 3.287156$$
$$* 10^{-4}T^3 - 1.398845 * 10^{-6}T^4 + 2.78786 * 10^{-9}T^5$$

(2.53)

Experiments on demineralized water are performed in temperature range [9 – 30.8] ℃, at which one ultrasound signal is investigated by the proposed method every 0.1℃ to obtain the associated USV. The comparison of the proposed values to those found by equation (2.53) shows a maximum error of 0.217% (at T = 20.8℃) and a correlation value R^2 higher than 0.99 as seen in figure (2.42).
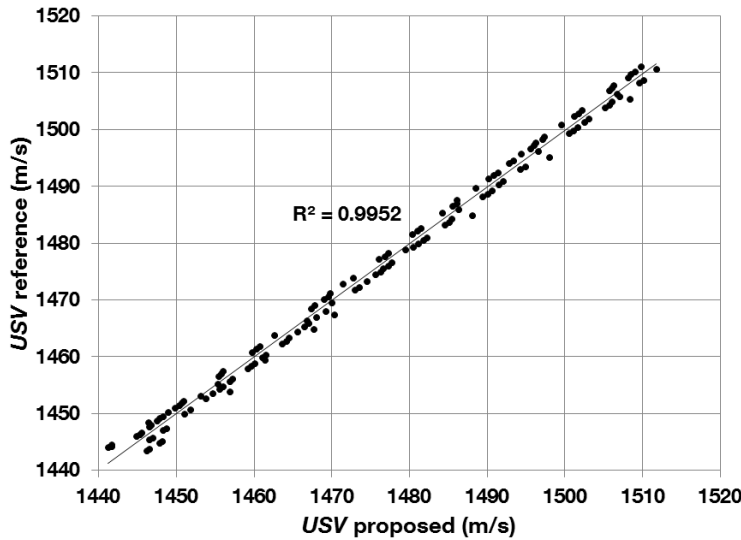
Figure (2.42): Speed of sound in demineralized water at [9 - 30.8] ℃, obtained by the proposed method in comparison to the reference values calculated by equation (2.53).

Although the proposed method requires high resolution spectrogram, its computational cost is relatively low (around 2 seconds for analysing 50 signals on 22.4 GFlops computer). The reason is that only small region of the spectrogram around its dominant frequency is calculated and investigated.

2.5.   Evaluation of the developed tools

Two techniques are commonly employed to evaluate the developed mathematical tools in signal and image processing. They measure certain standards on the output signals and/or images with respect to their exact forms. These techniques are subjective and objective measures, which indicate whether the output by the developed tool is augmented for further pattern recognition tasks [72].

Subjective measures
Subjective measures are borrowed from the field of psychology and human judgment, at which the evaluation is made through the individual experience. In this thesis, sensual inspection and Mean Opinion Score (*MOS*) are considered to perform a subjective evaluation. The sensual inspection is applied to images generated by the developed edge detection and spectrogram enhancement methods, to visually check the existence of significant edges and the removing of attached noise. It is also applied to aurally check existence/absence of certain sound pattern, as a judgment to any detector built by the established temporal and spectral features.
Alternatively, *MOS* presents a numerical subjective evaluation by giving a score to the enhanced spectrogram through its sound signal [73]. In most applications, ten

candidates (5 females, and 5 males) are asked to give a score [1 = bad, 2 = poor, 3 = fair, 4 = good, and 5 = excellent] to a sound. Afterwards, MOS is calculated by averaging the given scores by equation (2.54), while its confidence interval (CI) is computed at 95% confidence level, as described in figure (2.43) and explained in equation (2.55).

$$MOS = \frac{\sum_{i=1}^{l} S(i)}{l} \tag{2.54}$$

Where $S(i)$ is the score given by candidate number i, and l is the total number of candidates.
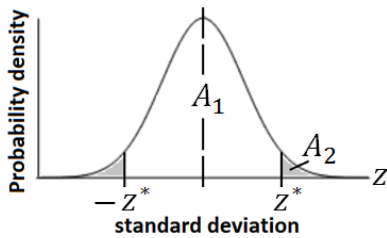
Figure (2.43): Normal distribution curve with 95% confidence level.

$$A_1 = 0.95, A_2 = \frac{1 - 0.95}{2} = 0.025 = P(z > z^*)$$

$$\therefore P(z \le z^*) = 1 - 0.025 = 0.975 \xrightarrow{\text{normal distribution table}} z^* = 1.96$$

$$CI = [MOS - (1.96 * \sigma), MOS + (1.96 * \sigma)] \tag{2.55}$$

Where $A_1, [-z^*, z^*], A_2, P, \sigma$ are the area defines a 95% confidence interval, its enclosing interval, remaining curve area, probability density function, and standard deviation of the given MOS; respectively.

Although subjective measures seem easy to be put into practice, they have some disadvantages. The number of characteristics a human eye/ear can distinguish is limited, and the judgment depends on the image/sound type.

Objective measures

Objective measures are borrowed from digital signal processing field and information theory, providing equations that calculate amount of deviation in the output image/sound in comparison to the exact shape (i.e., what it should be), and they include:

1. Signal to Noise Ratio (SNR) is a relation between the amounts of information in the output form to its differences from the exact form (i.e., noise), as given by equation (2.56).

$$SNR = \sqrt{\frac{\sum [O(x,y)]^2}{\sum [E(x,y) - O(x,y)]^2}} \tag{2.56}$$

Where $O(x,y)$ is the output form of an image/sound by one of the developed methods and $E(x,y)$ is its exact form. The form has two coordinates (x,y) for an image while one coordinate (x) for a sound, and can be pixels intensities of an edges image, power

spectrum of a spectrogram, or signal amplitude of a sound, as schematically described in figure (2.44).

(a)

(b)

For signal filtering evaluation: $O(x, y)$ and $E(x, y) \equiv$ signal amplitude

(c)

(d)

For edge detection evaluation: $O(x, y)$ and $E(x, y) \equiv$ pixel intensity

(e)

(f)

For spectrogram enhancement evaluation: $O(x, y)$ and $E(x, y) \equiv$ power spectrum
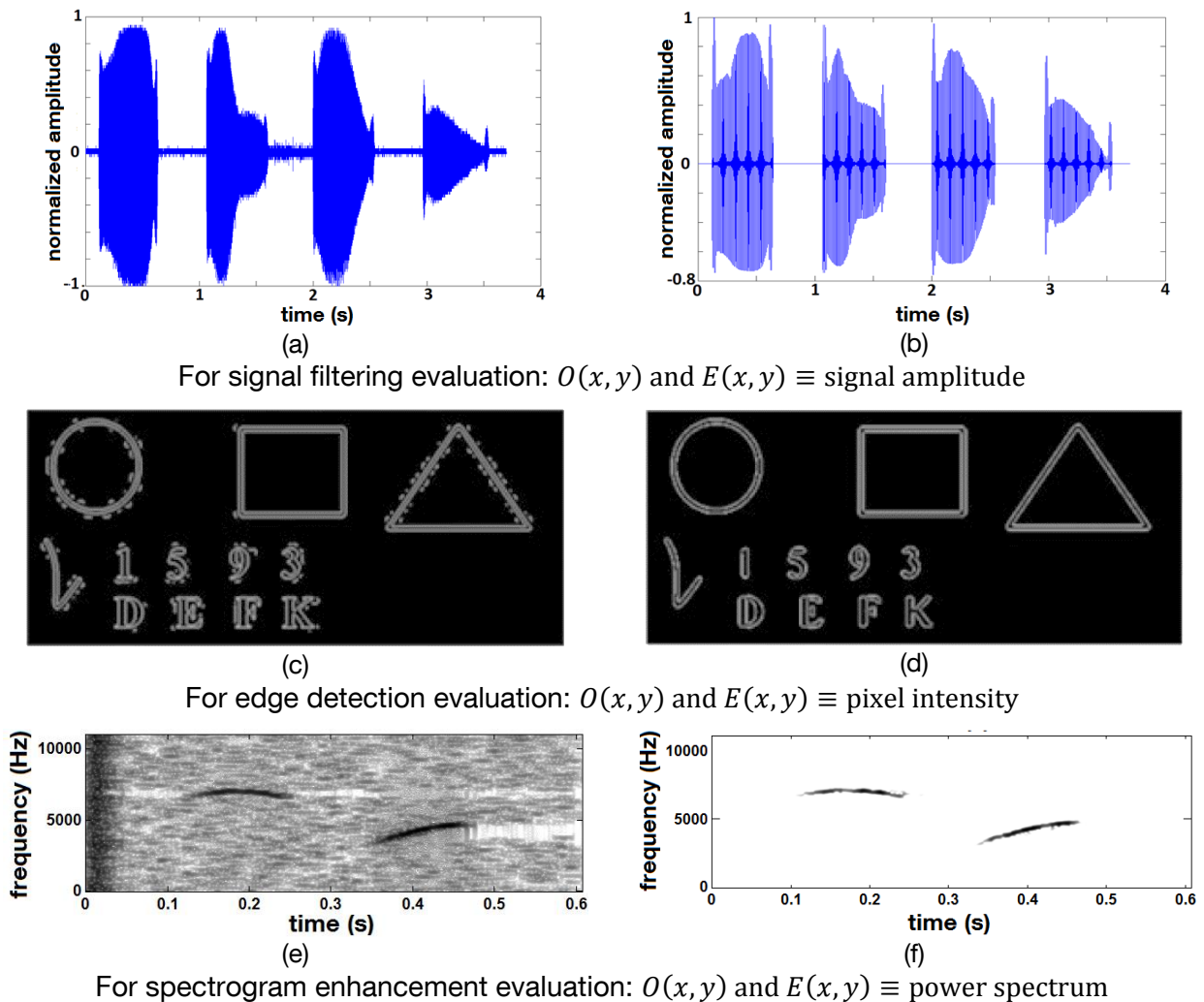
Figure (2.44): The representation of $O(x, y)$ and $E(x, y)$ in some images and sounds produced by the numerical tools (a, c, and e), and their exact forms (b, d, and f).

2. Error Root Mean Square ($ERMS$) is the overall differences between the output and exact forms, as clarified in equation (2.57).

$$ERMS = \sqrt{\frac{1}{M*N} \sum_{r=1}^{M} \sum_{c=1}^{N} [E(r,c) - O(r,c)]^2} \qquad (2.57)$$

Where M is the number of rows in an image [which equals one for a sound], and N is the number of columns in an image or total number of sound samples.

3. Segmental Signal to Noise Ratio ($SSNR$) is the average of SNR values over segments with sound activity in a spectrogram, as given by equation (2.58).

$$SSNR = \frac{10}{M} \sum_{m=0}^{M-1} log_{10} \sum_{i=Nm}^{Nm+N-1} \left(\frac{\sum P_{exact}}{\sum |P_{exact} - P_{output}|} \right) \tag{2.58}$$

Where P_{exact} and P_{output} are the exact and output spectrogram power spectrum; respectively. M is the number of spectrogram segments (usually set to 20), N is the number of samples in one segment.

4. Log Spectral Distance (LSD) is the spectral distance or distortion measure, expressed in dB, between the output and exact spectrograms, as given by equation (2.59).

$$LSD = \sqrt{\frac{2}{FS} \sum_{0}^{FS/2} \left[10log_{10} \left(\frac{P_{exact}}{P_{output}} \right) \right]^2} \tag{2.59}$$

Where FS is the sampling rate of the sound stream.

5. Itakura Saito (IS) is a measure of the perceptual difference between the exact and output spectrograms, as displayed in equation (2.60).

$$IS = \frac{2}{FS} \sum_{0}^{FS/2} \left[\frac{P_{exact}}{P_{output}} - log_{10} \frac{P_{exact}}{P_{output}} - 1 \right] \tag{2.60}$$

6. Average Eccentricity (AE) which simply and roughly shows if the shape of sound patterns (or image objects) is changed from those in the exact form, and given by equation (2.61).

$$AE = \frac{1}{K} \sum_{j=1}^{K} \frac{L_j}{W_j} \tag{2.61}$$

Where L and W are the length and width of the minimum bounding rectangle to the sound pattern/image object; respectively, while K is the number of patterns/objects in the spectrogram/image.

7. Localization Error (LE) is the average difference in locating the edges between exact and output edges image. It is calculated by inspecting the output edges distributions in x and y directions, and comparing these distributions to those of the exact edges, as clarified by equation (2.62) and schematically shown in figure (2.45).

$$LE = \frac{|\sum pixels\ of\ exact\ edges - \sum pixels\ of\ output\ edges|}{N} * 100\% \tag{2.62}$$

Where N is the total number of pixels in the evaluation direction (i.e., row direction or column direction).
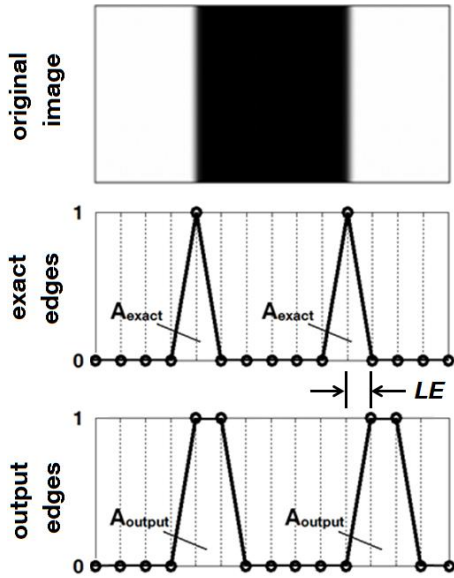
Figure (2.45): Schematic diagram for estimating the edges localization error (LE) between the exact and output edges images.

8. Energy Dissipation (ED) results from inefficiency of reducing the edge width to exactly one pixel during suppression. The edge remains covering more than one pixel, as shown in figure (2.45), results in more energy loss around this edge. This dissipation can be displayed by the intensities distribution around an edge. The area under this distribution is expected to be more dissipated for output edges image than for exact edges image, and can be used to calculate the energy dissipation at the associated edge by equation (2.63). The total energy dissipation is the average value of all rows and columns energy dissipations.

$$ED = log\left(\frac{A_{output}}{A_{exact}}\right) \tag{2.63}$$

Where A_{exact} and A_{output} are the areas under the intensity distribution curves corresponding to exact and output edges images; respectively.

Since objective measures require information for exact image/sound, they are applied in this thesis for validation purposes when the reference exact form is already known, as will be shown in the next chapter.

2.6.   Closed form neural network model

Artificial Neural Network (ANN) as an emulation of a biological network can establish almost any relationship among data by building models between a set of input and output vectors [74]. Usually, a feed-forward multilayer perceptron with an input, one or several hidden, and output layers is designed, as seen in figure (2.46a). Input layer contains a number of neurons equal to the number of the selected relevant features, while the output layer may contain one neuron for regression or two-class classification {0,1} problems, or many neurons for multi-class classification problems.
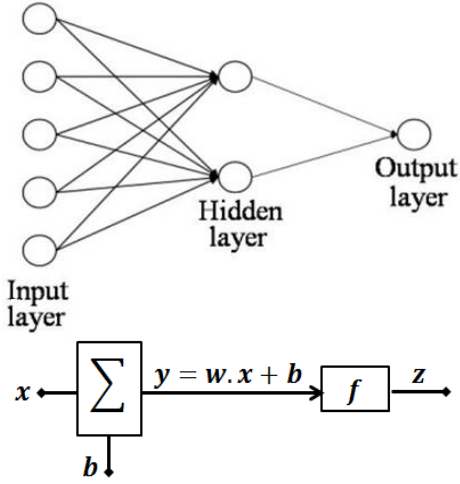
Each neuron weights the input with a weight (w) and sums it up with a bias (b). Then, it presents it using a transfer function (f) to the next layer, as shown in figure (2.46b). The widely implemented technique to train an ANN is the back propagation algorithm [75], which tries to minimize the network mean square error (MSE) in obtaining the desired output, by modifying the weights and bias. On the other hand, iterations on types of the implemented transfer functions may be essential to ensure best network performance. The commonly used transfer functions are Logsig and Tansig, given in equation (2.64) and equation (2.65); respectively.

$$f(y) = \frac{1}{1 + e^{-y}} \tag{2.64}$$

$$f(y) = \frac{2}{1 + e^{-2y}} - 1 \tag{2.65}$$

Although complications and time consumed to obtain an efficient ANN, once the model with best performance is found, no more training is made and it is used in classification/regression of real life samples.

In this section, a closed form is presented for the designed model, which decreases the necessary computations for model implementations to few simple equations solver. The performance of the input layer can be mathematically written in two equations, as explained in equations (2.65 and 2.67).

$$[y_1]_{n*1} = [W_{11}]_{n*n} \cdot [X]_{n*1} + [b_1]_{n*1} \tag{2.66}$$
$$[z_1]_{n*1} = f_1([y_1]_{n*1}) \tag{2.67}$$

Where W_{11}, and b_1 are the weights and biases for the neurons of the input layer. X is a vector contains n input features. y_1 is the output vector, which is transformed through a transfer function f_1 to produce the final output of this layer z_1. Similarly, the hidden layer(s) are mathematically represented as given in equations (2.68 and 2.69).

$$[y_{(i)}]_{m_{(i)}*1} = [W_{(i)(i-1)}]_{m_{(i)}*m_{(i-1)}} \cdot [z_{i-1}]_{m_{(i-1)}*1} + [b_{(i)}]_{m_{(i)}*1} \qquad , i = 2:K \tag{2.68}$$

$$[z_{(i)}]_{m_{(i)}*1} = f_{(i)}\left([y_{(i)}]_{m_{(i)}*1}\right) \tag{2.69}$$

Where K is the total number of hidden layers in the network, $m_{(i)}$ and $m_{(i-1)}$ are the number of neurons in layers number (i) and $(i-1)$; respectively. Finally, the network output (z_l) from the output layer can now be obtained by equations (2.70 and 2.71).

$$[y_l]_{C*1} = [W_{lK}]_{C*m_{(K)}} \cdot [z_K]_{m_{(K)}*1} + [b_l]_{C*1} \tag{2.70}$$
$$[z_l]_{C*1} = f_l([y_l]_{C*1}) \tag{2.71}$$

Where l is the total number of network layers, and C is its number of neurons on the output layer.

Equations (2.66, 2.68, and 2.70) are linear which calculated in a straightforward way, while the calculation speed of equations (2.67, 2.69, and 2.71) depends on the engaged transfer function(s).

3.   Applications To Life Sciences Engineering Problems

The developed numerical approaches are applied to solve four Life Science Engineering (LSE) problems. For each application, brief overview and a printout for the associated published paper are presented. The case studies of these applications are selected due to the availability of their database and reference validation results, as well as they are representatives to wide range of further LSE problems.

3.1   Bioacoustics detection of Red Palm Weevil (RPW)

This work is conducted under the hypothesis that distinctive spectral and temporal features in RPW larval sounds can be combined to construct improved indicator for automated detection of the infested trees, as described in figure (3.1).
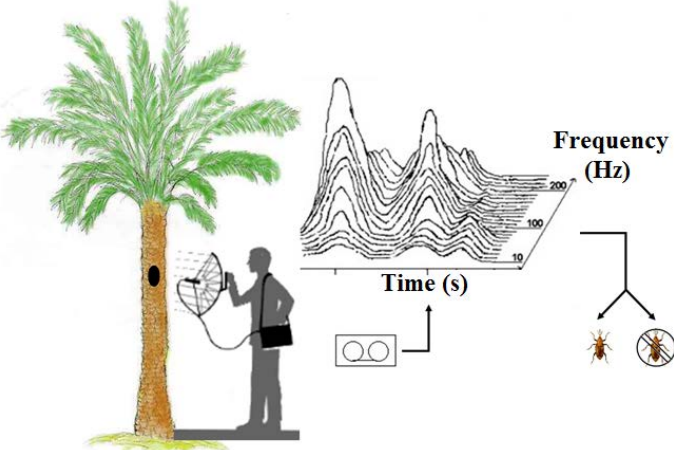
Figure (3.1): Clarification for detection of the hidden red palm weevil by its distinctive sound features.

Because of its high reproduction rate, RPW prefers to live in isolation from other insects; therefore all the recorded sounds are owing to RPW activities which give a good base for the acoustic detection. Alternatively; the greatest difference between RPW and other possible beetle species is that the activities of RPW seem to be very aggressive (i.e., not singular). The aggressive activities of RPW mean its sound streams are not distinctive, and it is difficult to separate the adjacent pulses from each other, as expressed in figure (3.2a). Whereas; the patterned sounds produced during the communication of some insects suchlike *Copris hispanus* [76] are more distinctive and clearer to spot, as shown in figure (3.2b).
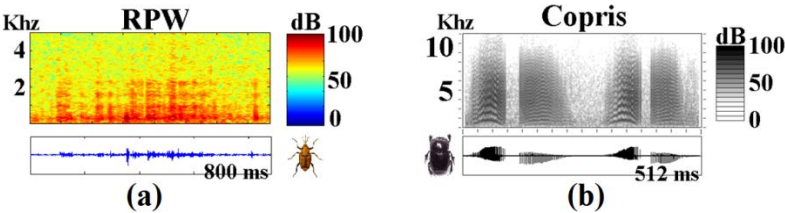
Figure (3.2): Spectrogram and waveforms emitted by (a) RPW feeding sound and (b) *Copris hispanus* call. [FFT length = 1024, Hanning window of length 256, and 85% overlapping].

DETECTION OF THE RED PALM WEEVIL RHYNCHOPHORUS FERRUGINEUS USING ITS BIOACOUSTICS FEATURES

WALID BARAKAT HUSSEIN [a] , MOHAMED AHMED HUSSEIN [a] & THOMAS BECKER [a]

[a] Center of Life Science Engineering, Technical University of Munich, Weihenstephaner steig 20, D-85354, Freising, Germany

PLEASE SCROLL DOWN FOR ARTICLE

DETECTION OF THE RED PALM WEEVIL *RHYNCHOPHORUS FERRUGINEUS* USING ITS BIOACOUSTICS FEATURES

WALID BARAKAT HUSSEIN*, MOHAMED AHMED HUSSEIN
AND THOMAS BECKER

Center of Life Science Engineering, Technical University of Munich,
Weihenstephaner steig 20, D-85354 Freising, Germany

ABSTRACT

The Red Palm Weevil (RPW) is the most destructive pest of the date palm in the world and a serious pest of coconuts. The insect has caused up to 20% loss of these plantations in Asia and the Middle East. The economic damage to palm crops due to RPW could be mitigated significantly by bioacoustics recognition in an earlier phase of infestation and by applying the appropriate treatment. This study is conducted under the hypothesis that distinctive spectral and temporal features in RPW larval sounds can be combined to construct improved indicators for automated detection of infestations. In this paper, a signal processing system is developed with available acoustic technology to detect the existence of RPW in a tree through its feeding sounds. A large set of features are extracted, including unconventional features such as temporal roll-off, temporal slope and temporal spread. Additionally, an analysis is provided of the criteria for the choice of the optimum frame length, as well as the selection of the suitable window function. The results confirm the efficiency of the developed system with the selected representative features, window functions and frame length to detect the existence of the RPW.

Keywords: Red Palm Weevil, bioacoustics detection, signal processing, feature extraction, feature selection.

INTRODUCTION

Since the 1980s, the Red Palm Weevil (RPW) has rapidly expanded its geographical range westwards. It reached Saudi Arabia and the United Arab Emirates in 1985, spreading throughout the Middle East and into Egypt. In 1994 it was detected in Spain, Israel, Jordan and Palestine in 1999, Italy in 2004, Canary Islands in 2005, Balearic Islands, France, and Greece in 2006, and Turkey in 2007 (Malumphy & Moran 2007). The European Commission is in the process of

*Corresponding author. Email: walid.hussein@wzw.tum.de

introducing emergency measures to prevent the further spread of RPW within the European Community.

All stages (egg, larva, pupa and adult), as described in Figure 1, are spent inside palm trunks and the life cycle cannot be completed outside them. The females deposit about 300 eggs in separate holes or injuries on the palm. Eggs hatch in 2 to 5 days into legless grubs which bore into the interior of the palms, and feed on the soft succulent tissues, discarding all fibrous material. The larval period varies from 1 to 3 months. The grubs pupate in an elongate oval, cylindrical cocoon made out of fibrous strands. At the end of the pupation period which lasts 14 to 21 days the adult weevils emerge (Murphy & Briscoe 1999).

Larvae and adults destroy the interior of the palm tree, often without the plant showing signs of deterioration until damage is severe (Blumberg *et al*. 2001). Hollowing out of the trunk reduces its mechanical resistance, making the plant susceptible to collapse and a danger to the public (Howard *et al*. 2001). In most cases, attack leads to the mortality of trees whatever their sizes. The larvae are large, but being hidden makes early detection with traditional methods very difficult. Adults of RPW can be efficiently monitored using pheromone based traps (Oeshlager *et al*. 1993; Soroker *et al*. 2005). However, these methods are unsuitable for quarantine inspections of planting material. Consequently, infested planting material is often transported to a new location before the first detectable symptoms of infestation appear (Mankin *et al*. 2008).
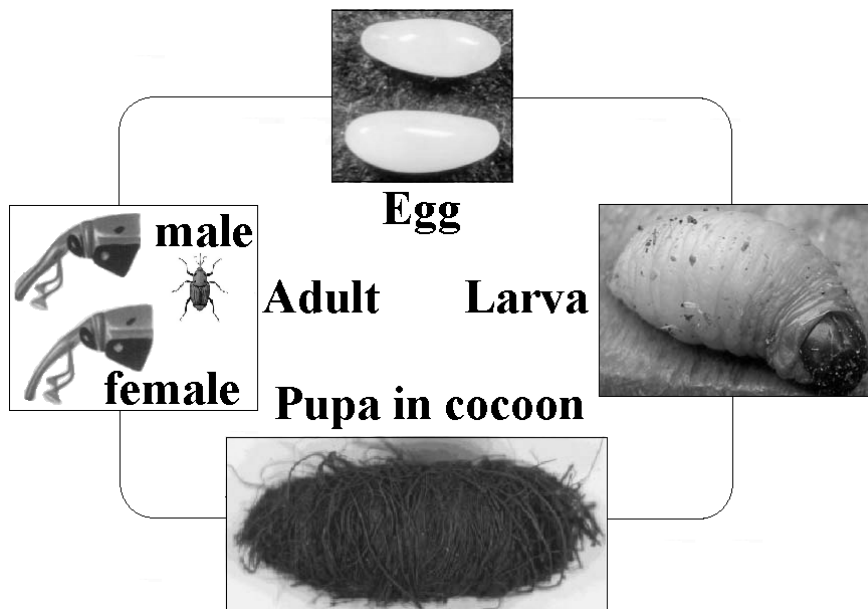
Figure 1. The complete life cycle stages of the Red Palm Weevil.

The economic damage to date palm crops due to RPW could be mitigated significantly by early detection and treatment. Bioacoustics technology and x-rays enable the detection of the early phase of infestation, however the usage of x-rays is expensive. Acoustic technology has potential for reducing the expense and dangers involved in tree inspection. Acoustic recordings from insects in trees often reveal signals with spectral and temporal features that make them distinctive and easily detectable (Mankin *et al.* 2002). Preliminary studies (Mizrach *et al.* 2003; Hetzroni *et al.* 2004; Soroker *et al.* 2004, 2006) have demonstrated that sensitive microphones and dedicated amplifiers enable detection of the movement and feeding sounds of RPW larvae in palm trees.

Automated acoustic analyses of sounds produced by hidden insect infestations have focused primarily on spectral aspects of the insect-produced signals (Mankin *et al.* 2000, 2001, 2007). Temporal patterns were considered by Zhang *et al.* (2003) in classifying larval sounds as snaps, rustles, or repeated pulses, and it has been observed that different stored product insects produce distinctly different sound patterns (Andrieu & Lessard 1990; Mankin *et al.* 1997).

Different sounds of the RPW activities can be recorded, including moving, spinning the cocoon and feeding sounds. It is worthless to take into consideration the moving sound of the RPW to detect its existence in an infected tree, because the moving sounds of many wood insects and beetles are nearly identical. Meanwhile, the waveform and sonogram of the feeding and cocoon spinning sounds are shown in Figure 2.

For the cocoon spinning sound, the strongest energy is below 1 kHz, while it reaches to 2.2 kHz for the feeding sound. Furthermore, one complete acoustic pattern takes approximately 2 seconds, while it is only 0.8 second for the feeding sound. Although the cocoon

FFT	Window	Window Length	Overlap
1024	Hanning	256	85%
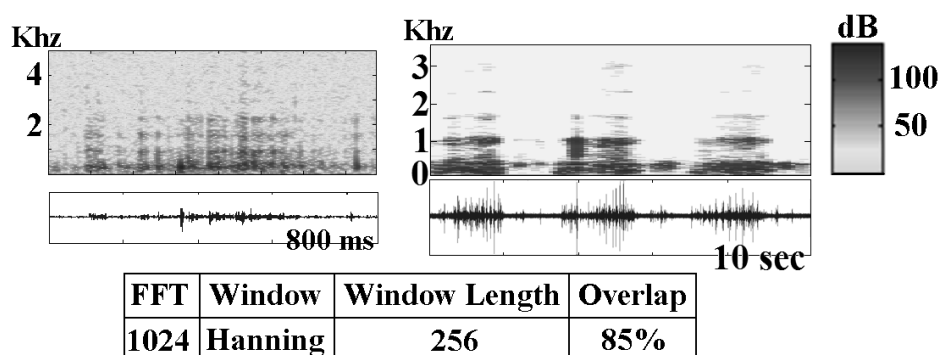
Figure 2. Spectrogram, envelope and detailed waveform of pulse trains for the (a) feeding sound and (b) spinning the cocoon sound of the Red Palm Weevil.

spinning sound may be a characteristic for the RPW, it is a seasonal sound that takes place only in certain periods of the overall life cycle. Therefore, it is less useful for the bioacoustic detection of the RPW. On the other hand; the feeding sound is the most frequently occurred sound in an infected tree. As RPW has a distinctive body shape and internal mouth structure compared to other coexisting beetles, its feeding sound has some unique features that make it more reliable for the RPW detection task.

In this paper, the audio detection system for the RPW feeding sound is presented in detail. The next section describes the recording method and the families of detection features which have been used in the system. This is followed by a section explaining the criteria behind the choice of the window function, before applying Fourier Transform to each time frame of the recorded sound. Finally, we discuss the selection of the representative features and the optimum frame length.

MATERIALS AND METHODS

Recording device

An infested date palm sample trunk (70 cm length, 38-43 cm diameters, 117 kg weight) was transferred from Saudi Arabia to Germany and stored in a quarantine room. The trunk was further divided into small blocks to focus on the weevil development and investigate the activities of the larvae during recording. This resulted in a complete picture for the RPW activities hidden in the trunk. Afterwards, equipment was installed including a mercury steam lamp, heater, humidity device, thermometer, infrared video cameras, as well as the recording device.

Two different recording devices were assessed for this task. The first was a Laar Ultrasound Gate hard disk recording System (frequency range 50 Hz – 250 kHz). The second was a Laar WD 60 detector with amplifying system and insertion sensors of different types (Contact microphone, airborne ultrasound microphone, contact acceleration sensor and a combined contact/airborne probe sensor). The best recordings were obtained via the second device which was attached to a Laar/Avisoft SASLab Pro sound analyzer (50 Hz to 30 kHz). Since there was no contact from the recording person, every recorded sound was caused by the RPW activities in the trunk. For a period of 13 subsequent months (~3 life cycles), covering different environments (i.e., pressure and temperature changes), 251 adult weevils emereged from the trunk (and were later eradicated) and 980 successful feeding sound recordings (each of 0.8 second) were made of the RPW.

Signal processing

The recordings were digitized at 11,025 Hz sampling rate on a computer and saved in .wav file format. The recorded signals were then high-pass filtered with a 200-Hz cutoff frequency to eliminate low-frequency background noise. The recorded signals were then divided into time frames with 90% overlap between frames, as shown in Figure 3. A program was developed using MATLAB, with user-friendly graphical interface, to hold all the signal processing procedures up to the detection step.

Time domain features

The temporal domain features listed in Table 1 were extracted for each time frame in order to have a time distribution of the features along the recorded signal. In Table 1, three unconventional features were included: temporal roll-off that describes the time below which 90% of the energy distribution is concentrated; temporal slope of decrease or increase of the signal amplitudes, computed by linear regression; temporal spread which represents the spread of the signal amplitudes around its mean value.

To extract the spectral domain features, each time frame was scaled by a suitable window function and then transformed in the frequency domain using the Fast Fourier Transform (FFT) method. The frequency axis was divided into 2,048 bins from zero frequency (i.e., DC value) up to 11,025 Hz (i.e., the sampling rate). Due to the symmetry of the resultant spectrum at half the sampling rate,
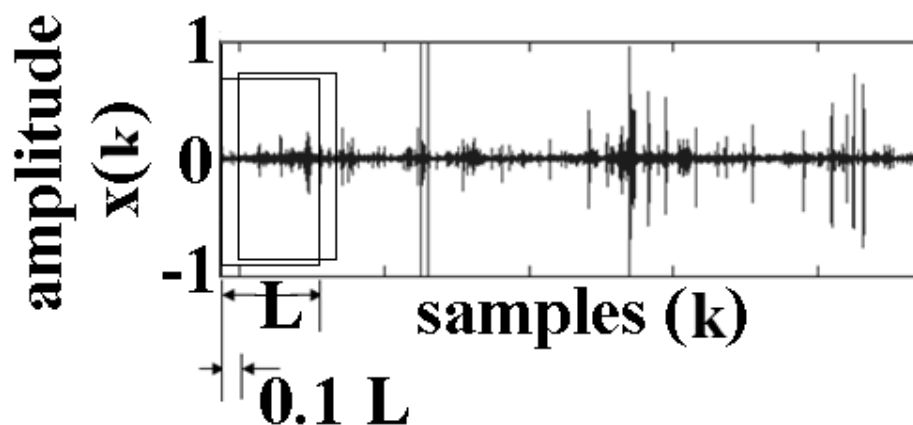
Figure 3. the recorded sound signal sampled in time domain with sampling rate of 11,025 Hz and divided to frames with 90% overlap.

TABLE 1

The time domain features which are used in this study to detect the existence of the Red Palm Weevil. Where $x(k)$ is the amplitude at sample k, k_{rf} is the bin number of the roll-off frequency, and N is the number of samples. One complete RPW feeding sound is 0.8 sec (i.e., 8,821 samples)

Name	Equation
Zero crossing rate	$$\frac{1}{2*N}\sum_{k=2}^{N}\left| sign\left(x(k)\right) - sign\left(x(k-1)\right)\right|, \quad \begin{cases} sign = 1 & ; x(k) > 0 \\ sign = 0 & ; x(k) = 0 \\ sign = -1 & ; x(k) < 0 \end{cases}$$
Short time energy	$$\sum_{k=1}^{N}(x(k))^2$$
Energy root mean square	$$\sqrt{\frac{\Sigma_{k=1}^{N}\left(x(k)\right)^2}{N}}$$
Temporal entropy	$$-\sum_{k=1}^{N}\left(\frac{|x(k)|}{\Sigma_{k=1}^{N}|x(k)|}\right)^2 \ln\left(\frac{|x(k)|}{\Sigma_{k=1}^{N}|x(k)|}\right)^2$$
Temporal roll-off	$$\sum_{k=1}^{k_{rf}}|x(k)|^2 = 0.9*\sum_{k=1}^{N}|x(k)|^2$$
Temporal slope	$$\frac{N\sum_{k=1}^{N}\left(k*x(k)\right) - \sum_{k=1}^{N}k*\sum_{k=1}^{N}x(k)}{\Sigma_{k=1}^{N}x(k)*\left(\Sigma_{k=1}^{N}k^2 - \left(\Sigma_{k=1}^{N}k\right)^2\right)}$$
Temporal flux	$$\sum_{k=1}^{N}\left(\frac{|x(k)|}{max\left(|x(k)|\right)}\bigg|_t - \frac{|x(k)|}{max\left(|x(k)|\right)}\bigg|_{t-1}\right)^2$$
Temporal crest-factor	$$\frac{max\left(|x(k)|\right)}{\frac{1}{N}\Sigma_{k=1}^{N}|x(k)|}$$
Temporal decrease	$$\left(\frac{1}{\Sigma_{k=2}^{N}|x(k)|}\right)*\sum_{k=2}^{N}\frac{|x(k)| - |x(1)|}{k-1}$$
Temporal smoothness	$$20*\sum_{k=2}^{N-1}\left|\log|x(k)| - \frac{\log|x(k-1)| + \log|x(k)| + \log|x(k+1)|}{3}\right|$$
Temporal centroid	$$C_t = \frac{\Sigma_{k=1}^{N}k*\left(x(k)\right)^2}{\Sigma_{k=1}^{N}\left(x(k)\right)^2}$$
Temporal spread	$$tsp = \sum_{k=1}^{N}(k-C_t)^2 * tpmf(k), \quad tpmf(k) = \frac{|x(k)|}{\Sigma_{k=1}^{N}|x(k)|}$$
Temporal skewness	$$m_3 = \sum_{k=1}^{N}(k-C_t)^3 * tpmf(k)$$ $$temporal\ skew = \frac{m_3}{tsp^{\frac{3}{2}}}$$
Temporal flatness	$$tf = \frac{\left(\prod_{k=1}^{N}|x(k)|\right)^{\frac{1}{k}}}{\frac{1}{N}\Sigma_{k=1}^{N}|x(k)|}$$
Temporal tonality	$$\min\left(\frac{10*\log(tf)}{-60}, 1\right)$$

only the spectrum values corresponding to the first 1,024 bins were taken into consideration, resulting in a significant reduction in data processing.

Windowing

The most common window functions are rectangular, Hamming, Hanning, flat top and Blackman windows. Each window function has slightly different wave form from the others, as shown in Figure 4, which results in a different performance in the frequency domain. In Figure 4, the tighter the main lobe width, the better the frequency resolution of the window functions, as in the rectangular window function. Additionally, the higher the reduction rate between the main lobe and the first side lobe, the less the spectral leakage around the central frequency, as in the Blackmann window.

The efficiency of each window function according to their frequency resolution and spectral leakage reduction criteria are shown in Table 2. As a result, the rectangular window was chosen to calculate the spectral features that require high frequency resolution (e.g. spectral roll-off), while the Blackmann window was chosen to calculate the spectral features that require less spectral leakage (e.g. spectral centroid).
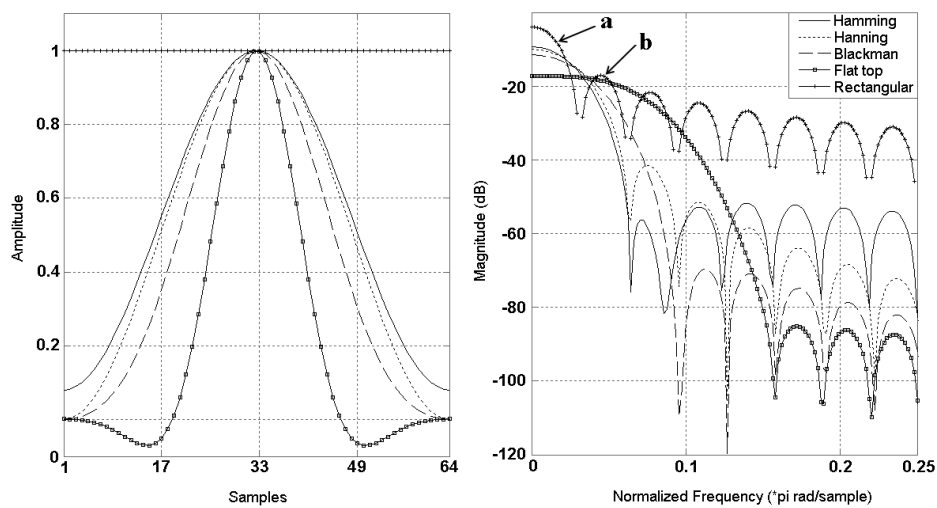
Figure 4. The distribution of the most common window functions in time domain (left), and their spectral domain representations in the frequency domain (right). The letter (a) points to the main lobe of the spectrum for the rectangular window, while (b) points to its first side lobe.

TABLE 2

Comparison of the frequency resolution and spectral leakage analyses
of the different window functions of Figure 5.

Window	Frequency resolution	Spectral leakage reduction
Hamming	Good	Fair
Hanning	Good	Good
Blackman	Poor	Best
Flat top	Poor	Good
Rectangular	Best	Poor

(a)

(b)
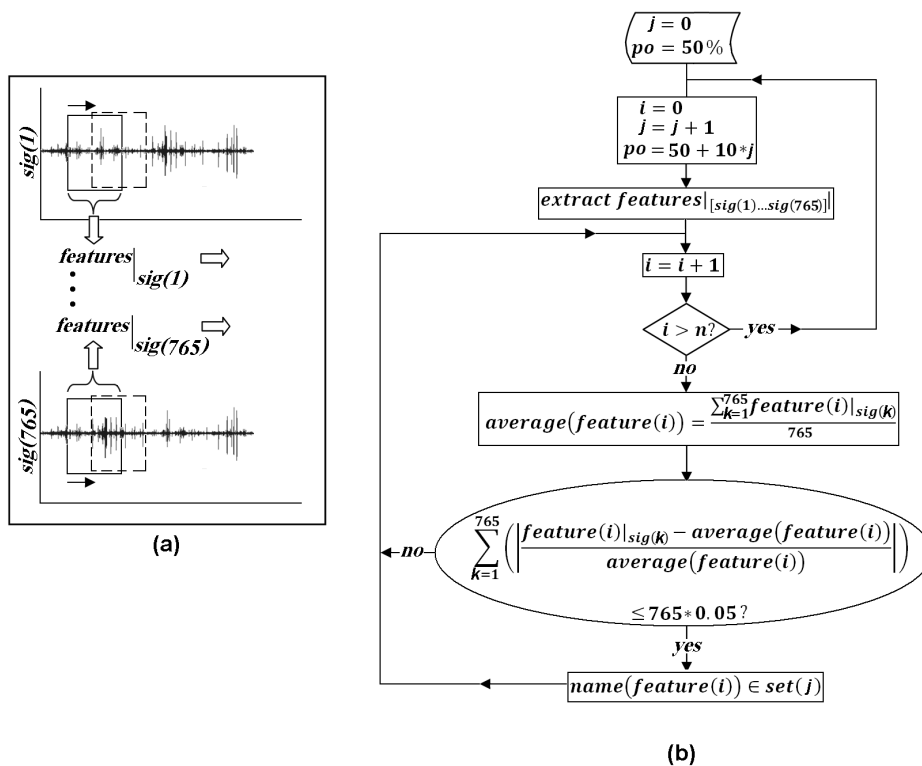
Figure 5: (a) Schematic diagram for the features selection algorithm. (b) Flow chart for the features selection algorithm. (*po*) is the percentage of frames overlap for a frame length of 0.4 sec., (*j*) is the number of the set, and (*n*) is the total number for the extracted features out of the signal.

Frequency domain features

The spectral domain features listed in Table 3 were extracted after applying the FFT to the windowing time frame. Adding these spectral features to the previously calculated temporal features, a group of 31 features were extracted for each time frame.

TABLE 3

The frequency domain features which are used in this study to detect the existence of the Red Palm Weevil. Where $X(n)$ is the spectrum at bin number n and spectral kurtosis determines whether the probability density distribution of the noisy signals is peaked or flat topped with respect to the Gaussian normal distribution

Name	Equation
Spectral roll-off	$\sum_{n=1}^{n_{rf}} |X(n)| = 0.9 * \sum_{n=1}^{1024} |X(n)|$
Spectral flux	$\sum_{n=1}^{1024} \left(\frac{|X(n)|}{\max(|X(n)|)} \Big|_t - \frac{|X(n)|}{\max(|X(n)|)} \Big|_{t-1} \right)^2$
Spectral energy	$\sum_{n=1}^{1024} |X(n)|^2$
Spectral root mean square	$\sqrt{\frac{\sum_1^{1024}(|X(n)|)^2}{1024}}$
Spectral slope	$\frac{1024 * \sum_{n=1}^{1024}(n*|X(n)|) - \sum_{n=1}^{1024} n * \sum_{n=1}^{1024}|X(n)|}{\sum_{n=1}^{1024}|X(n)| * \left(\sum_{n=1}^{1024} n^2 - \left(\sum_{n=1}^{1024} n\right)^2\right)}$
Spectral crest factor	$\frac{\max\left(|X(n)|\right)}{\frac{1}{1024}\sum_{n=1}^{1024}|X(n)|}$
Spectral smoothness	$20 * \sum_{n=2}^{1023} \left| \log|X(n)| - \frac{\log|X(n-1)| + \log|X(n)| + \log|X(n+1)|}{3} \right|$
Spectral decrease	$\left(\frac{1}{\sum_{n=2}^{1024}|X(n)|}\right) * \sum_{n=2}^{1024} \frac{|X(n)| - |X(1)|}{n-1}$
Spectral centroid	$C_t = \frac{\sum_{n=1}^{1024} n.|X(n)|}{\sum_{n=1}^{1024}|X(n)|}$
Spectral band width	$\sqrt{\frac{\sum_{n=1}^{M}(n-C_t)^2.(|X(n)|)^2}{\sum_{n=1}^{M}(|X(n)|)^2}}$, M: the last bin number at which $X = \frac{\max(|X|)}{2}$
Spectral spread	$ssp = \sum_{n=1}^{1024}(n-C_t)^2 * pmf(n)$, $pmf(n) = \frac{|X(n)|}{\sum_{n=1}^{1024}|X(n)|}$
Spectral skewness	$\frac{\sum_{n=1}^{1024}(n-C_t)^3 * pmf(n)}{ssp^{\frac{3}{2}}} \begin{cases} < 0 & \textit{more energy on the right} \\ = 0 & \textbf{symmetric distribution} \\ > 0 & \textit{more energy on the left} \end{cases}$
Spectral kurtosis	$\frac{\sum_{n=1}^{1024}(n-C_t)^4 * pmf(n)}{ssp^2} \begin{cases} < 3 & \textit{flatter distribution} \\ = 3 & \textbf{normal distribution} \\ > 3 & \textit{peaker distribution} \end{cases}$
Spectral entropy	$-\sum_{n=1}^{1024} pmf(n) * \ln(pmf(n))$
Spectral flatness	$sf = \frac{(\prod_{n=1}^{1024}|X(n)|)^{\frac{1}{n}}}{\frac{1}{1024}\sum_{n=1}^{1024}|X(n)|}$
Spectral tonality	$\min\left(\frac{10 * \log(sf)}{-60}, 1\right)$

Features selection

While some audio features may be considered distinctive and representative of the RPW feeding activity, others may be worthless for this particular diagnosis task. The purpose of applying a features' selection algorithm is to select the representative features from a

large set of features. Feature selection algorithms typically fall into two categories: feature ranking (Peng *et al*. 2005) and subset selection (Singhi & Liu 2006). Feature ranking ranks the features by a metric and eliminates all features that do not achieve an adequate score. Subset selection searches the set of possible features for the optimal subset by removing the most irrelevant and redundant features from the data. In this study the feature ranking method was applied. To handle this, the temporal and spectral features were extracted for 765 out of the 980 RPW feeding sound recordings. Subsequently, the average feature value was calculated for the 765 values (based on the 765 recordings). If more than 80% of the recordings have a feature value that lies within 5% deviation from the corresponding average feature value, that feature will be selected as a distinctive feature. The algorithm, as shown in Figure 5, was implemented for 60, 70, 80, and 90% frames overlap and the resultant selected features were stored in set(1), set(2), set(3), and set(4); respectively, as clarified in Table 4.

Optimum frame length

As mentioned above, the recorded signal is divided into a number of time frames. The time frame length depends on the total duration of the desired sound (e.g. the feeding sound), the percentage of overlap, and the required number of time frames. Equation (1) may be used to calculate the frame duration

$$y = \frac{x}{1 + (n-1).(1 - percentage\ of\ overlapping)} \tag{1}$$

where x is the whole sound duration; y is the frame duration; and n is the number of frames. For example, it requires 0.208 second for each frame to divide the feeding sound of the RPW (0.8 sec) into 20 time frames with 85% overlap. Whereas there is an optimum selection for the frame length which is affected by the intended representative features. In depth, if the mean zero crossing rate is considered as identification feature for the feeding sound of the RPW, the frame period should ensure that the value of the mean zero crossing rate is approximately the same for all recorded feeding sounds of the RPW. In Figure 6, the predetermined sets of features (i.e., set(1), set (2), set(3), and set(4)) were extracted for two audio signals (signal (1), and signal (2)), as well as for a reference signal of a typical feeding sound of the RPW, over different choices of frame lengths. Signal (1) contains a feeding sound for a RPW, while signal (2) does not contain a feeding sound. The mean square error between the selected features

TABLE 4

The enclosing set for the selected features, results from
the features selection algorithm of Figure 5 and frame
overlap values of 60, 70, 80, and 90%.

Feature	Enclosing Set
Zero crossing rate	1,3,4
Short time energy	2,4
Energy root mean square	3,4
Temporal entropy	1,2
Temporal roll-off	1,3,4
Temporal slope	2,4
Temporal flux	2,3
Temporal crest-factor	2,3,4
Temporal decrease	1,2
Temporal smoothness	1
Temporal centroid	1,4
Temporal spread	1,4
Temporal skewness	2,3
Temporal flatness	1,3
Temporal tonality	2,3
Spectral roll-off	2,3,4
Spectral flux	1,3
Spectral energy	1,2
Spectral root mean square	1,2,4
Spectral slope	2,4
Spectral crest factor	1,3
Spectral smoothness	1,3
Spectral decrease	2,4
Spectral centroid	1,4
Spectral band width	4
Spectral spread	3,4
Spectral skewness	4
Spectral kurtosis	1,4
Spectral entropy	2,3,4
Spectral flatness	2,3
Spectral tonality	

of each signal and the reference signal features was calculated and
scanned with the corresponding frame length.

In this paper, features set(4), with an overlap of 90%, is chosen
in the detection of the feeding sound of the RPW, and as a result of
Figure 6, an optimum frame length of 0.42 sec is selected. The reason
for the choice of features set (4) is its large content of worldwide
implemented features in the task of sound-based surveillance, and
also it contains the three innovative features of this study (i.e.,
temporal roll-off, temporal slope and temporal spread).

188

$$MSE_{sig(1)} = \left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{feature(i)_{ref.signal} - feature(i)_{sig(1)}}{feature(i)_{ref.signal}}\right]^2\right) * 100\%$$
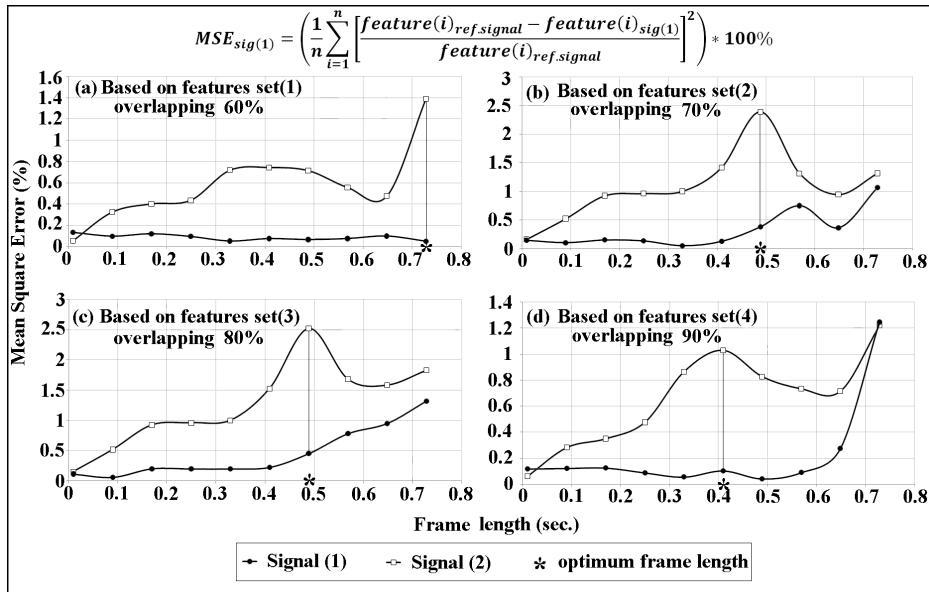
Figure 6. The mean square error between recorded sounds and a reference feeding sound of Red Palm Weevil versus the frame length. Based on features of (a) set(1), (b) set(2), (c) set(3), or (d) set(4). Signal (1) contains a feeding sound while signal (2) does not.

RESULTS AND DISCUSSION

Developing the system

The selected temporal as well as the selected spectral features were extracted for 765 out of the 980 recordings, using features listed in Table 1 and Table 3 respectively. The distributions of these features with time are shown in Figure 7 for a typical RPW feeding sound. In Figure 7, some features experience equivalent standard deviations, which make it difficult to differentiate among them, e.g. temporal roll-off and temporal centroid, spectral slope and spectral centroid; and spectral kurtosis and spectral skewness. Meanwhile the mean value of most features is more representative than the standard deviation. This necessitates careful attention to exclude redundant values in the choice of the related statistic for each feature. Table 5 shows which statistic was assigned to each feature. Therefore, one representative value was considered for each feature over one recording and this value averaged for the 765 recordings. The resultant system contained 18 representative features (i.e., set(4)) for the RPW feeding sound, and these features values were saved as the *system features*.
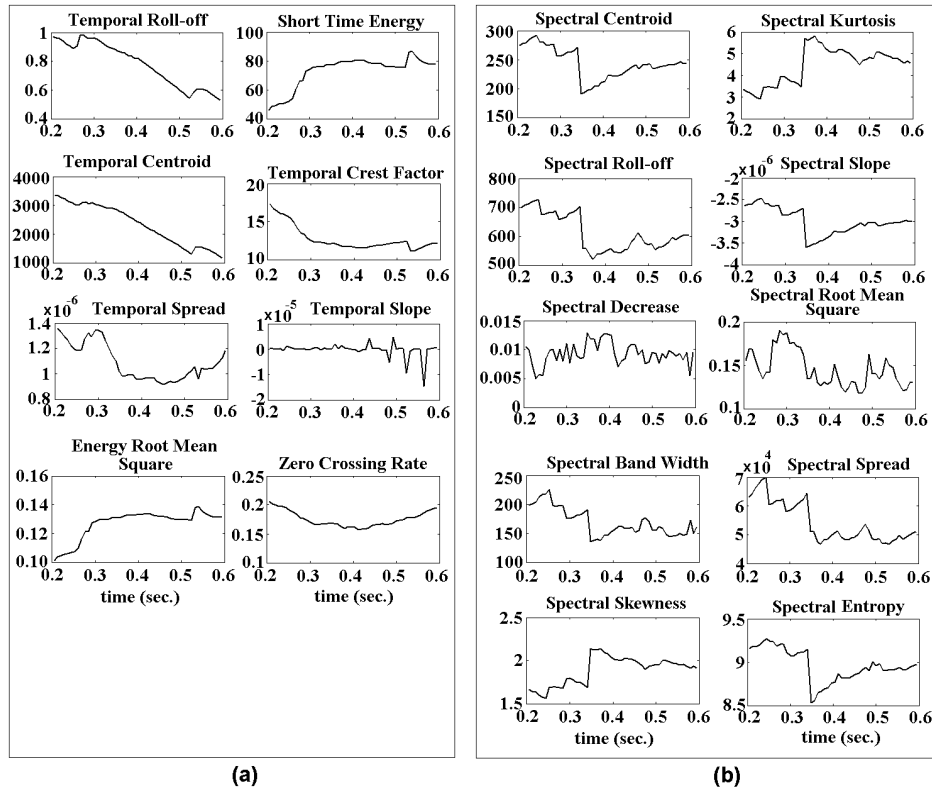
Figure 7. Distribution of the selected (a) temporal features and (b) spectral features with time for a typical Red Palm Weevil feeding sound of total duration 0.8 sec., divided into 10 frames each of 0.42 sec. by means of 90% overlap.

Validation of the system

The remaining 215 out of the 980 feeding sound recordings were used for the validation of the developed system from which the pre-selected 18 features are extracted for each recording, and then compared to the *system features*. 203 records were successfully detected since their extracted features were within 5% from those of the *system features*. On the other hand, 12 records were not detected since more than 10 of their extracted features had more than 12% percent deviation from the corresponding *system features*. This means that the system efficiency is higher than 94% in the detection of the RPW feeding sounds, when the feeding sound was defined to be the one with features that have a maximum of 5% deviation from the corresponding *system features*.

TABLE 5

The related statistic and the window function assigned to each feature.

Feature	Related statistic	Window Function
Zero crossing rate	mean	—
Short time energy	std	—
Energy root mean square	mean	—
Temporal roll-off	mean	—
Temporal slope	mean	—
Temporal crest-factor	mean	—
Temporal centroid	std	—
Temporal spread	mean	—
Spectral roll-off	mean	Rectangular
Spectral root mean square	mean	Blackman
Spectral slope	mean	Blackman
Spectral decrease	mean	Blackman
Spectral centroid	mean	Blackman
Spectral band width	mean	Blackman
Spectral spread	mean	Blackman
Spectral skewness	mean	Blackman
Spectral kurtosis	mean	Blackman
Spectral entropy	mean	Blackman

Evaluation of the system

In order to test the power of the developed system to detect the existence of the RPW in the field, the recording device with its insertion microphone was used. The insertion microphone brings the sensor near the sound source and avoids wind noise effects, resulting in high quality recordings. Human, animal, machinery and footsteps sounds were avoided as much as possible, since they can greatly affect the waveforms and result in a wrong detection.

An audio stream of 5 minutes duration was recorded on one infected palm tree and the sound features were extracted for each 0.8 second batch, with 90% overlap. The extracted features were scanned for each batch as shown in Figure 8, as well as the *system features* (the horizontal solid lines). Similarly, an expert listener (see Acknowledgments) was able to detect the feeding sound regions by band pass filtering the whole stream between 1-3 kHz, followed by analysing the filtered stream using BVL Spectrogram V8 Sound analysis software. The analysis revealed trains of impulses separated by intervals of less than 500 ms that were considered as potential RPW feeding sounds (the grey regions in Figure 8).

From Figure 8, the features distributions in some batches have the same values as those for the reference feeding sound. These batches lie on the same times that were detected by an expert as a RPW feeding sound. Mean temporal roll-off, mean spectral decrease, STD short time energy, and STD temporal root mean square introduce one
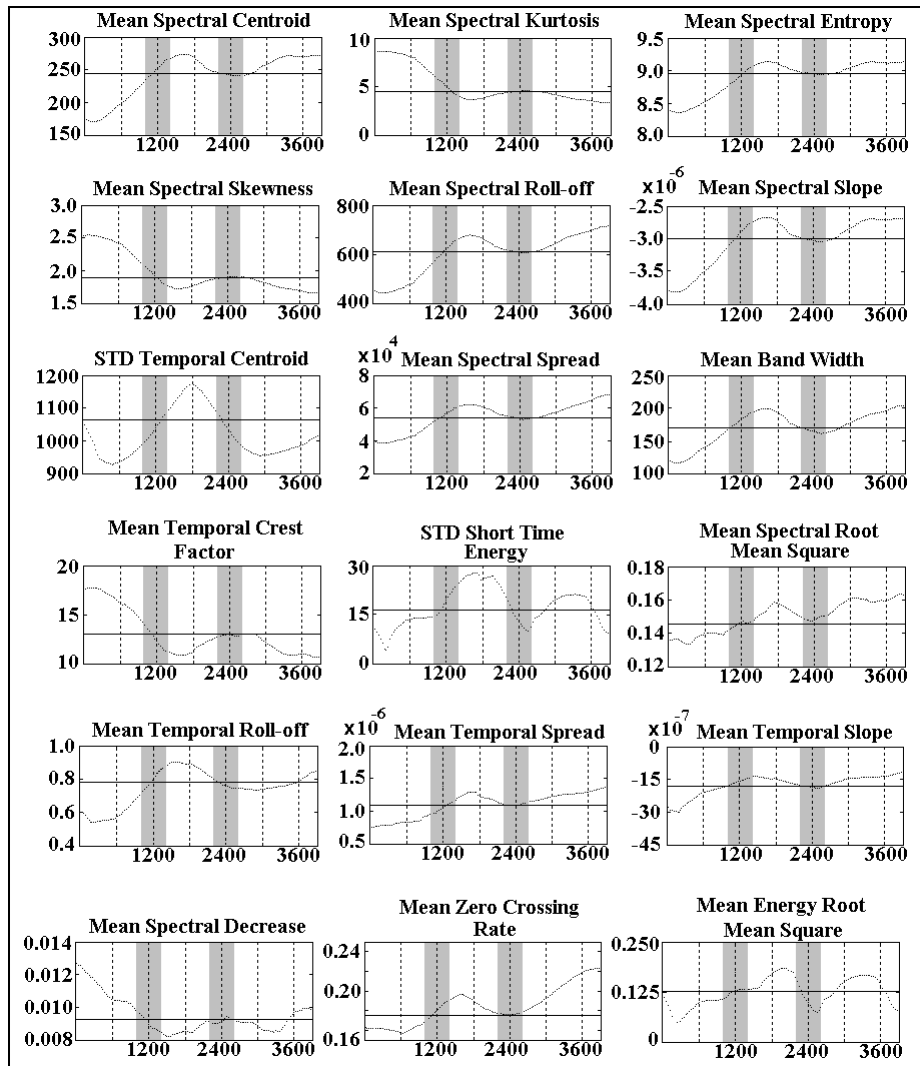
Figure 8. The extracted features from an audio stream of 5 minutes duration plotted against the batch number. The stream is divided into 3750 batches of duration 0.8 sec. with 90% overlap. The solid horizontal lines represent the feature value for a typical Red Palm Weevil feeding sound, while the grey regions are detected by an expert as containing Red Palm Weevil feeding sounds.

more feeding sound near batch number 3600 [ca 4.75 minutes]. But it is not matched by the other features. Although all the above features were used in the developed system, some of them only are sufficient to detect the RPW feeding sound in the 5-minute stream, as mean spectral entropy, mean spectral roll-off and mean zero crossing rate.

Another two 50-second streams were recorded from two uninfected trunks – having average diameters of 36.4 and 45.8 centimetres – revealing sounds that did not match the *system features,* indicating no existence for the RPW in these trunks. The results demonstrate the efficiency of the developed system with the selected representative features, window functions, and frame period to detect the existence of the RPW through its feeding sound.

CONCLUSIONS

The economic damage to palm crops due to RPW could be mitigated significantly by bioacoustic recognition in an earlier phase of infestation and applying the appropriate treatment. While their secretive behaviour makes an early detection with traditional methods (e.g. pheromone traps) not applicable, it can be performed using bioacoustics recognition. Acoustic recordings from insects in trees often reveal sound waves with spectral and temporal features that make them distinctive and easily detectable. In this study, before dividing the sound wave into time frames, the optimum frame length – for the case of the RPW feeding sound – was investigated for several overlap percentages (60, 70, 80, 90%). The developed system applies 90% overlap, with its corresponding optimum frame interval of 0.42 sec, to divide the recorded sounds. For each frame, 15 time domain features were extracted. In addition, the criteria behind the selection of the window function were studied, followed by the extraction of 16 frequency domain features, using their appropriate window functions.

Because some features may be useless for RPW detection, the values of the selected feature should have small deviation for all reference RPW feeding sounds. As a result, ten frequency domain features were selected, beside eight time domain features containing temporal roll-off, temporal slope and temporal spread. These three features are introduced as new features, and are competently implemented in the developed detection system. The developed system was validated to have an efficiency higher than 94% Therefore it was well able to detect the existence of the RPW feeding sound in a five-minute sound stream recorded on an infected trunk and its absence in two 50-second streams recorded on an uninfected trunk.

After the positive determination of the presence of Red Palm Weevil in a palm, it is necessary to remove the palm tree, because to date no real efficient treatment method exists. Also the disinfection of the soil needs to be taken under consideration. The next step of this research is to develop a machine learning program to handle all the available values for the feeding sound features in different environmental conditions. In order to decide whether a new recorded

sound represents a feeding sound, methods such as artificial neural networks or support vector machines may be implemented.

ACKNOWLEDGMENTS

REFERENCES

Andrieu, A. J. & Fleurat-Lessard, F. (1990). Insect noise in stored foodstuffs. Compact disk, *INRA*, Versailles, France.

Blumberg, D., Navon, A., Kehat, E. & Lavski, S. (2001). Date palm pests in Israel early second millennium. *Alon Hanotea*, **55**, 42-48.

Drosopoulos, S. & Claridge, M. (2006). *Insect sounds and communication. Physiology, behaviour, ecology and evolution*. New York:CRC Press LLC.

Hetzroni, A., Mizrach, A., Nakache, Y. & Soroker, V. (2004). Developing Spectral Model to Monitor Activity of Red Palm Weevil. *Alon Hanotea*, **58**, 466-469.

Howard, F. W., Moore, D., Giblin-Davis, R. & Abad, R. (2001). *Insects on Palms*. Oxford: CABI Publishing, Wallingford.

Malumphy, C. & Moran, H. (2007). Plant Pest Notice: Red palm weevil *(Rhynchophorus ferrugineu). Central Science Laboratory*, Sand Hutton, York, U.K.

Mankin, R. W., Shuman, D. & Coffelt, J. A. (1997). Acoustic counting of adult insects with differing rates and intensities of sound production in stored wheat. *Journal of Economic Entomology*, **90**, 1032-1038.

Mankin, R. W., Brandhorst-Hubbard, J., Flanders, K. L., Zhang, M., Crocker, R. L., Lapointe, S. L., McCoy, C. W., Fisher, J. R. & Weaver, D. K. (2000). Eavesdropping on insects hidden in soil and interior structures of plants. *Journal of Economic Entomology*, **93**, 1173-1182.

Mankin, R. W., Lapointe, S. L. & Franqui, R. A. (2001). Acoustic surveying of subterranean insect populations in citrus groves. *Journal of Economic Entomology*, **94**, 853-859.

Mankin, R. W., Osbrink, W. L., Oi, F. M. & Anderson, J. B. (2002). Acoustic Detection of Termite Infestations in Urban Trees. *Journal of Economic Entomology*, **95**, 981-988.

Mankin, R. W., Hubbard, J. L. & Flanders, K. L. (2007). Acoustic indicators for mapping infestation probabilities of soil invertebrates. *Journal of Economic Entomology*, **100**, 790-800.

Mankin, R. W., Mizrach, A., Hetzroni, A., Levsky, S., Nakache, Y. & Soroker, V. (2008). Temporal and spectral features of sounds of wood-boring beetle larvae: identifiable patterns of activity enable improved discrimination from background noise. *Florida Entomologist*, **91**, 241-248.

Mizrach, A., Hetzroni, A., Nacashe, Y. & Soroker, V. (2003). Acoustic detection of Red Palm Weevil activity in the tree trunk. *Proceeding of the International Congress on Information Technology in Agriculture, Food and Environment* (Ed. by Sindir, K. O.), p. 25. Izmir, Turkey.

Murphy, S. T. & Briscoe, B. R. (1999). The red palm weevil as an alien invasive: biology and prospects for biological control as a component of IPM. *Biocontrol News and Information*, **20**, 35-45.

194

Oehlschlager, A. C., Chinchilla, C. M., Jiron, L. F., Morgan, B. & Mexzon, R. G. (1993). Development of an effective pheromone based trapping system for the American palm weevil, *Rhynchophorus palmarum*, in oil palm plantations. *Journal of Economic Entomology*, **86**, 1381-1392.

Palestrini C., Pavan G. & Zunino M. (1991). Acoustic signals in *Copris incertus* Say (*Coleoptera Scarabaeidae Coprinae*). *Ethology, Ecology & Evolution*, **1**, 143-146.

Peng, H., Long, F. & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 1226-1238.

Singhi, S. & Liu, H. (2006). Feature subset selection bias for classification learning. *Proceedings of the 23rd international conference on machine learning.* Pittsburgh, Pennsylvania 2006, pp. 849-856.

Soroker, V., Nakache, Y., Landau, U., Mizrach, A., Hetzroni, A. & Gerling, D. (2004). Utilization of sounding methodology to detect infestation by *Rhynchophorus ferrugineus* on palm offshoots. *Phytoparasitica*, **32**, 6-8.

Soroker, V., Blumberg, D., Haberman, A., Hamburger-Rishard, M., Reneh, S., Talebaev, S., Anshelevich, L. & Harari, A. R. (2005). Current status of red palm weevil infestation in date palm plantations in Israel. *Phytoparasitica*, **33**, 97-106.

Soroker, V., Gindin, G., Glazer, I., Pinhas, J., Levsky, S., Eliahu, M., Biton, S., Haberman, A., Nakache, Y., Mizrach, A. & Hetzroni, A. (2006). The red palm weevil infestation in Israel: occurrence and management. *I Jornada International sobre el Picudo Rojo de la Palmeras*. Agroalimed, Generalitat Valenciana 2006, pp. 59-79.

Zhang, M., Crocker, R. L., Mankin, R. W., Flanders, K. L. & Brandhorst-Hubbard, J. L. (2003). Acoustic identification and measurement of activity patterns of white grubs in soil. *Journal of Economic Entomology*, **96**, 1704-1710.

3.2   Fat contents recognition in meat slices

The developed edge detection approach is applied to recognize the fat boundaries in meat slices images, as an example for non-living biological tissues. Since Sobel operators provide both differencing and smoothing effects, they are employed to obtain gradients of each feature image. Non-maximum suppression and flux equilibrium are applied to the produced edges images to thin out edges and fill existent discontinuities; respectively. Edges images found by the feature images are then combined into one final edges image using a reasonable combination factor. Parametric investigations have been performed for the selection of mask size, flux equilibrium mask size, and combination coefficient of the edges image.

The meat slices images are prepared before applying the edge detector by histogram thresholding analysis and anisotropic diffusion filter. The histogram thresholding analysis eliminates the background area while the anisotropic diffusion filter reduces the specular reflections and speckles without influencing the edges. Once edges are detected, a region growing method is performed on the interiors of the detected boundaries to define the area of the enclosed region(s). A flow chart for the structure of the accompanying paper is presented in figure (3.3).
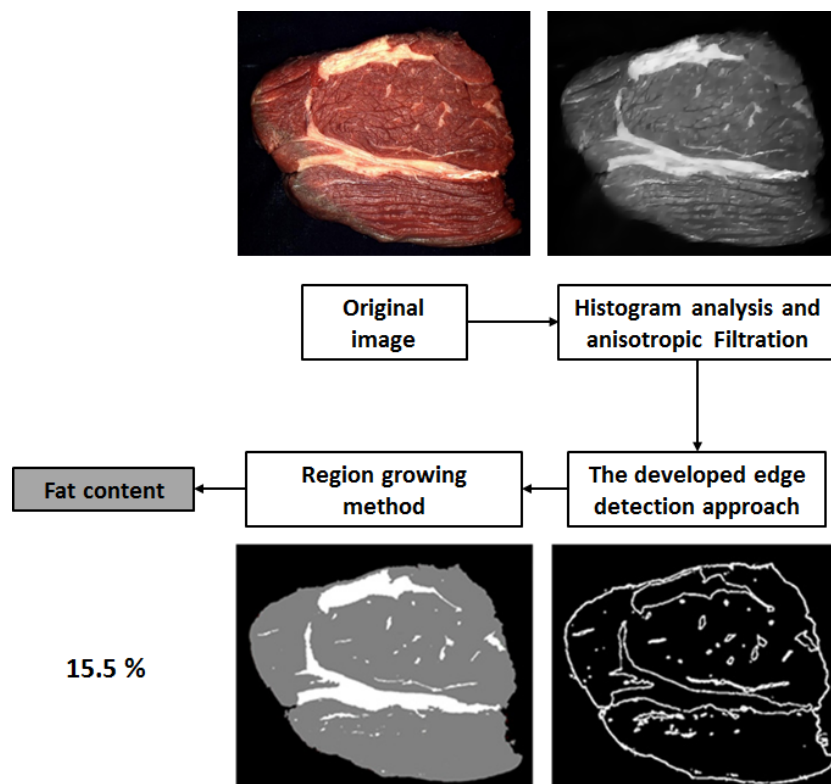
Figure (3.3): Flow chart describes the steps of recognizing the fat contents in meat slices.

A novel edge detection method with application to the fat content prediction in marbled meat

W.B. Hussein*, A.A. Moaty, M.A. Hussein, T. Becker

Group of (Bio-) Process Technology and Process Analysis, Faculty of Life Science Engineering, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany

ARTICLE INFO

ABSTRACT

Intramuscular fat is an important quality criterion, notably juiciness, in meat grading. But traditional visual inspectors are time consuming and destructive. However, edge detection techniques characterize meat surface in consistence, rapid, and non-destructive approach. In this paper, novel edge detection method applied on intramuscular fat is presented based on the energy and skewness as two smoothed versions of the image. Parametric analyses were investigated and the method was tested on several images, producing minimum improvements of 6.451%, 1.667% and 7.826% in signal to noise ratio, mean square error and edges localization, respectively, in comparison to Roberts, Prewitt, Sobel, and Canny detectors.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The fat content, visual texture and chemical composition of meat based products have a large influence on their nutritional value, functional properties, sensory quality, storage conditions and commercial value. Different studies have proved that juiciness and acceptance of the meat is mainly influenced by its fat content, either in calf [1], or lamb meat [2]. Thus, fat estimation knowledge is of great importance for quality control, not only for producers but also for retailers and consumers. Since the visual appearance of food and other biological products is a major factor in the judgment of quality, visual inspection is an important part of quality control in these industries. This inspection has historically been performed by the use of the only "tool" available, the human eye. The process accompanying data acquisition, automatic evaluation and direct control of the ambient conditions is possible only to a very limited extent, because the human assessment furnishes above all qualitative but hardly quantitative data and, on top of this, such inspection is time consuming and cost intensive. This visual inspection offers one of the challenges that is to make the production more automatic and flexible, ideally by the use of autonomous robotic systems. Alternatively, chemical analysis is frequently used method to determine the intramuscular fat content. But this technique is expensive, destructive and tedious, because meat as a raw material is extremely variable and may range from 1% to 65% fat.

As long as the measurements are carried out on the meat surface, image processing might offer a rapid and non-destructive method for analyzing the overall composition for quality control purposes. Furthermore, hand held image processing devices could be developed and used on an industrial basis to determine meat based products composition with the advantage that they are non-contact sensing solutions.

In the past, several works were proposed in order to introduce computer vision on meat quality evaluation. In these works, meat grading was based exclusively on the analysis of meat surface images. Image analysis was described as a highly promising approach for objectively assessing online quality control of industrial meat products [3–5]. In fact, it has been used since the early 1980s [6,7] to determine crude fat content and textural properties. Gerrard et al. [8] have developed an image processing system to evaluate meat quality, in particular to determine beef marbling scores. In the work of Li et al. [9], marbling and image texture features were used to develop tenderness prediction models by using statistical and neural network techniques. Texture features have also been used in [10,11]. They studied a method to detect meat quality using the concept of marbling score and texture analysis. They also proposed the implementation of a meat grading system using neural network techniques and multiple regression analysis.

Although the digital image processing is increasingly used in meat processing field, some bottle necks were shown up, that do not allow an easy adaptation of its techniques. These bottle necks include the difficulty of division of the meat images into the main contents of muscle, ribs, and fat. As well, limitations in detecting accurately the edges of these contents at sufficient resolution,

---

* Corresponding author. Tel.: +49 8161 712624; fax: +49 8161 713883.
E-mail address: whussein@wzw.tum.de (W.B. Hussein).

causing propagation of the errors in the further image acquisition and processing steps.

The representation of the image information in terms of edges is a method of compaction, meaning compacting the two dimensional image patterns into a set of one dimensional curve [12]. Detection of edges is a fundamental step for most computer vision applications such as remote sensing, fingerprint analysis, cytology, optical character recognition and industrial inspection. In addition, edges image reduces the space and computational time required in further steps for analyzing the image.

Edges are mostly detected using either the first derivatives, called gradient, or the second derivatives, called Laplacian. Laplacian is more sensitive to noise since it uses more information because of the nature of the second derivatives [13,14]. Along four decades, various edge detection algorithms were developed. The earliest popular works in this category include the algorithms developed in [15–17]. These classical methods compute an estimation of gradient for the pixels, and look for local maxima to localize step edges. Typically, they are simple in computation and capable to detect the edges and their orientations, but due to lack of smoothing step, they are sensitive to noise and inaccurate [18]. One general problem with the gradient based detection methods is that their output edges may be thicker (i.e., several pixel widths) than a connected unit width line around the boundary of an object. In many applications, further steps such as edge thinning and edge linking are necessary to extract relevant and useful edges. Non-maximum suppression is a simple and widely used edge thinning method. It removes all edge responses that are not maximal in each cross section of the edge direction in their local neighborhood.

The optimal detector should minimize the probability of false positives (i.e., detecting spurious edges caused by noise), as well as that of false negatives (i.e., missing real edges). Additionally, the detected edges must be as close as possible to the true edges. Finally, the detector must return one point only for each true edge point. Out of these criteria, the Canny edge detector was developed, which is probably the most widely used detector and considered to be the standard edge detection algorithm in the computer vision applications [19].

Canny dealt with the edge detection problem as a signal processing optimization problem, so he developed an objective function to be optimized, at which its solution was a rather complex exponential function [20]. The main drawback in the Canny algorithm is its inclusion of a number of adjustable parameters, which can affect the computational time and effectiveness of the algorithm [21]. The first parameter is the size of the smoothing Gaussian filter, which influence highly the detection results. The second parameter is the threshold values used in the implemented hysteresis thresholding method, which can either miss important information or identify irrelevant noise information. On the other hand, it is still difficult to give generic filter size or thresholds that work well on all images.

This paper proposes a novel gray level edge detection method based on the energy and skewness features of an image, as a progress and enhancement to the work given by Hussein and Becker [22]. Each feature presents a smoothed version of the original image, avoiding the application of the smoothing filters, and therefore prevents their accomplished drawbacks. These features are borrowed from the digital signal processing field, as they are implemented for pattern recognition and identification applications [23].

In this work, the energy and skewness, and their gradients are calculated for the original gray level image. Non-maximum suppression algorithm with sub-pixel accuracy is applied to suppress thick edges in both images (i.e., energy image and skewness image). Finally, a combination occurs between the edges calculated by the energy image and those calculated by the skewness image, to produce a final form of the edges image. Many parametric analyses had been investigated to efficiently select the mask size, the mask represented pixel, the flux equilibrium matrix size, and the combination coefficient between the energy and skewness edges images. The proposed method was tested on several images, and the results were competitive in comparison with Roberts, Prewitt, Sobel, and Canny detectors in detection of fine edges with higher signal to noise ratio, less root mean square error, and less total edge localization error. Afterwards, the method has been applied to efficiently detect the intramuscular fat content in a meat slice image. In the following section a detailed algorithm and explanation of the proposed method will be presented. Further sections include parametric studies, validation, and experimental results of the proposed method in comparison with the other conventional gradient based detectors. As well as its application to evaluate the meat quality by intramuscular fat content prediction in the image of marbled meat slices, which is a quite complicated issue due to the difficulty arising from the muscle and fat structural overlapping. Furthermore, intramuscular fat content were chemically analyzed for all the slices and compared with the proposed method outcomes.

2. Material and method

2.1. Gradient based edge detection

One way to detect edges in the image is by computing the local intensity gradient at each pixel of the image. The gradient G of an image intensity $f(x,y)$ at location (x,y) is a vector of two components, G_x and G_y, which measures how the gray level values change in the x and y directions, respectively. Edges are detected in areas where the gray level value fluctuates sharply, and the more rapid this value changes the stronger the edge is. Therefore, edges that are detected at pixels have local maxima or local minima of the gradient, as sketched in Fig. 1 for noiseless two edges image. The edge strength is the magnitude of the gradient G_m at this edge, while the direction of the edge θ is perpendicular to the gradient orientation of this edge, given as

$$G_x = \frac{\partial f}{\partial x}, \quad G_y = \frac{\partial f}{\partial y}$$

$$G_m = \sqrt{G_x^2 + G_y^2} \tag{1}$$

$$\theta = \tan^{-1}\left(\frac{G_y}{G_x}\right) + \frac{\pi}{2} \tag{2}$$

Calculating the gradient of an image is similar to convolving it with two gradient operators, one in x direction and the other in y direction. The commonly used gradient operators are those presented by Roberts, Prewitt and Sobel edge detectors, as shown in Fig. 2. It is important to notice that the 3∗3 operators, such like those of Prewitt and Sobel, have a second degree order of error, $Ó(\delta x^2)$ and $Ó(\delta y^2)$, in the calculation of gradient in x and y directions; respectively. Where δ is the distance between two pixels in the indicated directions.

The conventional approach in these detectors is to choose a threshold value at which the pixels with gradient magnitudes greater than or equal to this threshold are considered edges; otherwise, they are not edges. Furthermore, since they do not smooth the image before gradient calculation, they are only suitable for well contrasted noiseless images.

Meanwhile, Canny edge detection method considers the edges are step edges corrupted by additive Gaussian noise. It starts with image filtration by the derivative of the Gaussian function, which is given in Eq. (3), then finding the magnitude and orientation of the gradient. The edges strengths and directions are calculated by Eqs. (1) and (2), respectively. The multi-pixel wide edges are
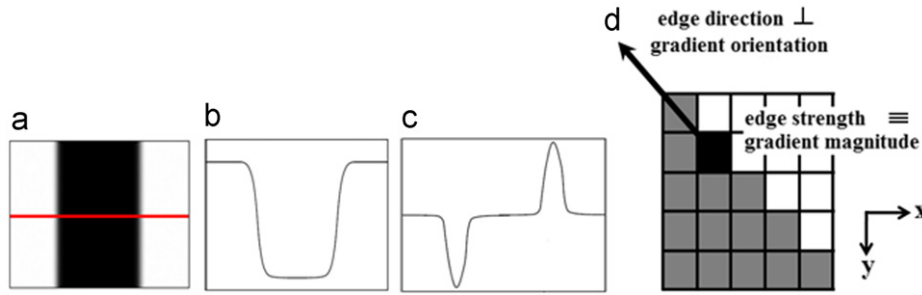
Fig. 1. Schematic diagram for a noiseless two edges image (a), with the intensities and gradient distributions of the pixels row which are covered by the indicated horizontal line, (b) and (c); respectively. The relation between the gradient magnitude and orientation with the edge strength and direction is displayed in (d).

Fig. 2. Basic gradient operators for (a) Roberts, (b) Prewitt, and (c) Sobel edge detectors. Δ_1 is the gradient operator in x direction, while Δ_2 is in y direction.

thinned down to single pixel width using the non-maximum suppression algorithm. Finally, the hysteresis thresholding algorithm is used to name edges and non-edges

$$Gauss(x,y) = \frac{1}{2\pi\sigma^2}\, e^{-(x^2+y^2)/2\sigma^2} \tag{3}$$

The optimum choice of the filter size (σ) is application dependent, and relates to the desired behavior of the detector, as shown in Fig. 3 for the results of application of the Canny detector on two images with two different σ values ($\sigma = 1, 2$).

2.2. The proposed edge detection method

2.2.1. Image segmentation and mask centroid

Initially a small square matrix (i.e., $n*n$ mask) is slid over the original gray level image in x direction and then in y direction, with step of one pixel. Instantaneously, the intensities of the pixels covered by the mask are used to calculate the mask energy and skewness. The calculated values of energy and skewness are dependent on the mask size, which controls the order of error in calculating the gradient image, and may influence the output edges image. Therefore, investigations of different mask sizes were applied, and are presented in Section 3 with estimation for the optimum mask size.

In addition, the simple choice for a mask represented pixel is its center, but a better choice would be its centroid (i.e., weighted center) which set the represented pixel according to the mask intensities distribution. A comparison between the influence of using mask center and using mask centroid on the output edges image is presented in Section 3, and a formula for calculating the mask centroid is

$$mask\ centroid = \sum_{i=1}^{n} i*f_i \Big\backslash \sum_{i=1}^{n} f_i \tag{4}$$

where f_i is the gray level intensity value of pixel i, and the back slash means that only the quotient of the division is considered.

2.2.2. Energy and skewness images

The energy feature (En), which is calculated by Eq. (5), performs a two dimensional spatially restricted scaling of the input image, to provide an even contrast output energy image. It is useful for getting some discernable output for images that are either obviously dark or bright

$$En = \frac{\sum_{i=1}^{n} f_i^2}{n*n} \tag{5}$$

While the skewness feature (S), which is calculated by Eq. (6), is a measure for the degree of asymmetry of the mask intensities distribution, in comparison with the normal distribution. The skewness is positive if it is tailed to the right side, as given in Fig. 4a, and negative if the distribution is tailed to the left side, as given in Fig. 4b. The symmetric (normal) distribution has an equal mean and median value, and therefore its skewness is zero. The skewness image may be viewed as a detailed water mark of the original image

$$S = \frac{\sum_{i=1}^{n^2} f_i*(i - maskcentroid)^3}{\sqrt{\left(\sum_{i=1}^{n^2} f_i*(i - maskcentroid)^2\right)^3}} \tag{6}$$

Since Sobel operators provide both differencing and smoothing effects at the same time, they are employed to obtain the gradient images (En_x, En_y), and (S_x, S_y) for energy and skewness images, respectively. Afterwards, with the aid of Eqs. (1) and (2), the edges strengths and directions are calculated for each image separately.

2.2.3. Non-maximum suppression with sub-pixel accuracy

The method of non-maximum suppression is applied to remove unnecessary edges by suppressing the non-maximum magnitude in each cross section of the edge direction in their local neighborhood. For each pixel, we check whether the gradient magnitude is greater at one pixel distance away in either the positive or negative direction of its gradient orientation. If the pixel is greater than both, its magnitude value set to one, otherwise it is set to zero (i.e., suppressed).

In most cases, the gradient orientation is not directed to the center of the neighbored pixel, as schematically described by example given in Fig. 5. Therefore, an interpolation is made to estimate the value of the gradient at this intermediate pixel (i.e., sub-pixel accuracy).

A quadratic interpolation, given by Eq. (7), using the gradient magnitudes of the three neighbored pixels gives sufficiently accurate gradient value at the intermediate pixel. Afterwards, the non-maximum suppression algorithm is applied between the main pixel and the two intermediate pixels on its orientation line

$$G_m = a\bar{\theta}^2 + b\bar{\theta} + c \tag{7}$$

where $\bar{\theta}$ is the gradient orientation of the intermediate pixel, $\bar{\theta} = \theta - (\pi/2)$. The coefficients a, b, and c are calculated by the strengths and directions of the neighbored pixels.

Thinning the edges to one pixel width increases the accuracy in boundaries recognition, and therefore makes the method more
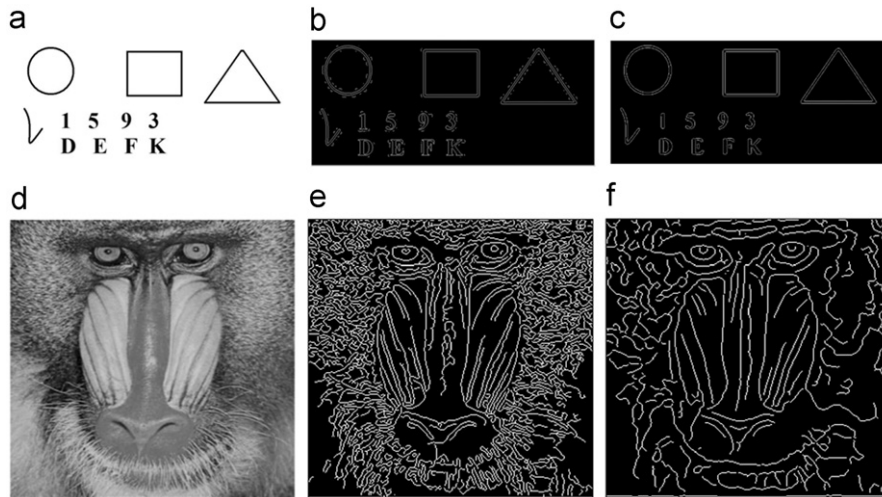
Fig. 3. The Canny detector outputs for two different natures images, (a) and (d). For $\sigma=1$, the detection was noisy for the first image (b), and reasonably accurate for the second image (e). While For $\sigma=2$, the detection became better for the first image (c), and poorer for the second image (f).
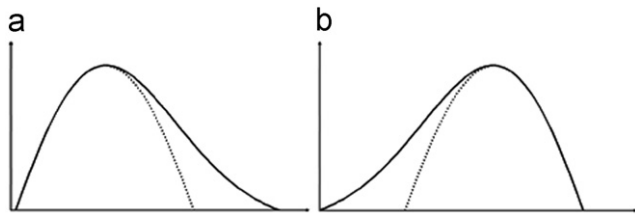
Fig. 4. A schematic diagram for the positive and negative skewness, (a) and (b); respectively, with respect to the normal distribution.

Fig. 6. A schematic diagram for a disconnect edge and how it is transferred to a connected edge using the flux equilibrium condition.

Fig. 5. Schematic diagram for how the gradient orientation does not pass by the centers of the pixels, and passes by an intermediate pixel which can be calculated by quadratic interpolation among the neighbored pixels.

robust and efficient in prediction of the intramuscular fat boundaries in marbled meat slices images.

2.2.4. Flux equilibrium

The edges image may suffer from existence of discontinuities, in the form of zero pixel(s) that intersects the edge direction. Restoring the missing edges data is very important in many edge detection applications, and is magnified in food applications where the boundaries are huddled and may easily overlap; interconnected openings take place by imaging. In particular the problem of intramuscular fat recognition, since many edges points are missing due to the difficulty of the muscle and fat structures interference. The flux equilibrium condition, as given in Eq. (8), checks the edges values along edge direction and fills the zero pixels if they are located between two edges, and in their direction, as shown in Fig. 6

$$\frac{dE}{d\theta} = 0 \tag{8}$$

Eq. (8) states that the derivative of the energy strength along its direction should be zero everywhere, and if this derivative is

not zero at some pixels, the flux equilibrium technique conserve the missing data and force the zero derivative condition to be met. This check is applied on the edges image using a flux check matrix, at which an analysis for different flux check matrix sizes was performed and will be presented in Section 3.

2.2.5. Edges combination

The edges found by the energy image and those found by the skewness image are combined to form the final form of the edges image, using the formula given in Eq. (9). The combination weight coefficient, w, belongs to the interval [0,1], at which a parametric analysis was applied, as will be explained in Section 3, to efficiently choose the value of w. The output of Eq. (9) also belongs to [0,1], and to have binary output, all values greater than or equal to 0.5 were set to 1, otherwise they were set to zero

$$E(x,y) = \frac{w*En_E(x,y) + (1-w)*En_S(x,y)}{\max(w*En_E(x,y) + (1-w)*En_S(x,y))} \tag{9}$$

where E is the final edge strength, w is the combination coefficient, and En_E and En_S are the strengths of edges from energy and skewness images, respectively.

2.3. Evaluation of the edge detector performance

There are two techniques to evaluate the performance of edge detectors, subjective measures and objective measures [24]. Subjective techniques are borrowed from the field of psychology and the human judgment. Although these measures seem easy to be put into practice, they have some disadvantages. The number of characteristics a human eye can distinguish is limited. Furthermore, the judgment depends on the individual experience as well as the image type.

On the other hand, objective measures are borrowed from the digital signal processing field and information theory, and provide equations that can be used to calculate the amount of error in the output edges image in comparison to the exact edges image. In this study, the objective measures are divided into two groups, namely the global measures and the local measures. The global objective measures evaluates the integral differences between the exact and output edges images, and they include the root mean square error (ERMS) and the signal to noise ratio (SNR), which are given by

$$ERMS = \sqrt{\frac{1}{M*N} \sum_{r=1}^{M} \sum_{c=1}^{N} [E(r,c) - O(r,c)]^2} \qquad (10)$$

$$SNR = \sqrt{\frac{\sum_{r=1}^{M} \sum_{c=1}^{N} [E(r,c)]^2}{\sum_{r=1}^{M} \sum_{c=1}^{N} [E(r,c) - O(r,c)]^2}} \qquad (11)$$
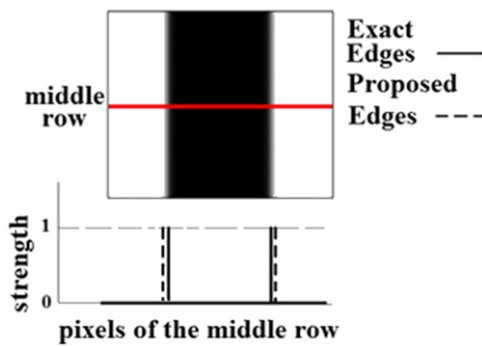
Fig. 7. Schematic diagram for how the differential evaluation technique is used to compare the difference in localization between the exact and output edges.

where $O(r, c)$ is the exact edges image, $E(r, c)$ is the output edges image, M is the number of image rows, and N is the number of image columns.

The local objective measures evaluate the differential differences between the exact and the output edges images, and they include the percentage of the edges localization error and the energy dissipation around the edges. The localization errors of the edges is calculated by inspecting the output edges distributions in x and y directions, and comparing these distributions to those of the exact edges, as given by Eq. (12) and schematically shown in Fig. 7. The total localization error (LE) is the average value of all rows and columns localization errors

$$Localization\ error = \frac{|\sum pixels\ of\ exact\ edges - \sum pixels\ of\ output\ edges|}{N} *100\%$$

$$(12)$$

where N is the total number of pixels in the evaluation direction (i.e., row direction or column direction).

Alternatively, since suppression methods may fail to reduce the edge width to exactly one pixel, the edge still cover more than one pixel, as shown in Fig. 8, results in more energy loss (i.e., energy dissipation) around this edge. This dissipation can be displayed by the strengths distribution around the edge. The area under the first derivative curve is expected to be more dissipated for the output edges image than for the exact edges image, and can be used to calculate the energy dissipation at the edge, as given in Eq. (13). The total energy dissipation (ED) is the average value of all rows and columns energy dissipations and Fig. 8d shows how the local energy dissipation value changes with the ratio between these areas

$$Energy\ dissipation = \log\left(\frac{A_{output}}{A_{exact}}\right) = log(A_{output}) \qquad (13)$$
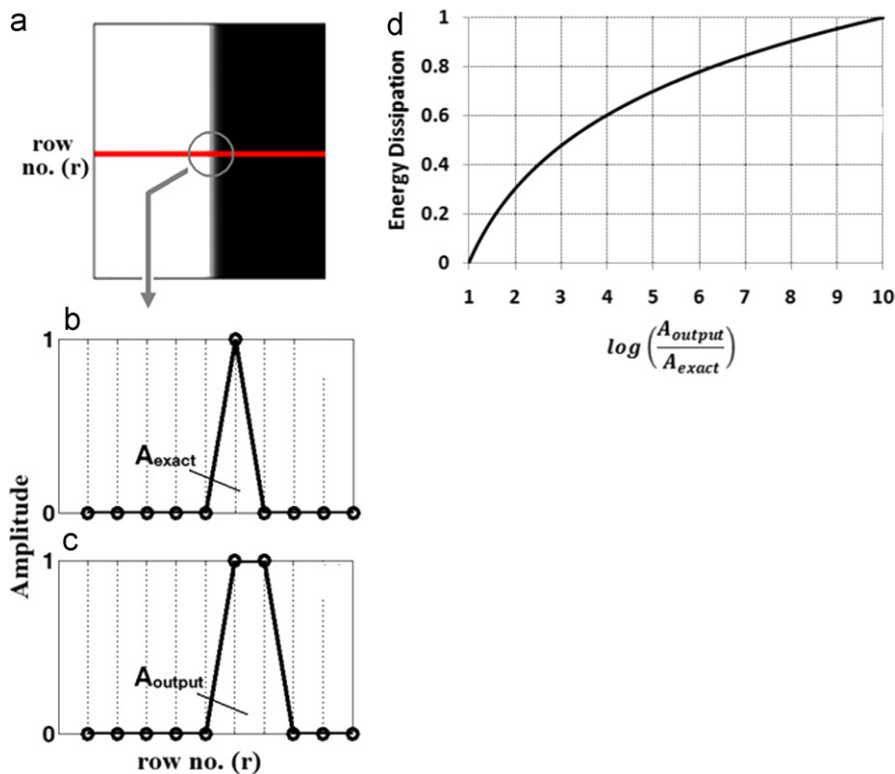
Fig. 8. (a) An one vertical edge image under inspection for the row wise edge strengths distribution. (b) The exact amplitude distribution of pixels in row number (r). (c) The edge detector output for the amplitude distribution of pixels in row number (r). (d) The variation of the local energy dissipation with the ratio between the detected areas in (b) and (c).

where A_{exact} and A_{output} are the areas under the first derivative curves corresponding to exact and output edges images, respectively.

While the local objective measures are not related, they both form a set of two "geometrical" elements (*LE%*, *ED*), since the first measure concerns with location, and the second measure concerns with area. And from definition, they have the same priority in the evaluation of the edge detector performance.

3. Experimental results and discussion

The proposed edge detection method has been tested on images of different natures including primitive features, such as circle, rectangle, alphabets, and numbers, as well as real world images such as human face and hair, fingerprint, and intramuscular fat images. Parametric studies were investigated for the best estimation of the mask size, the mask represented pixel, the flux check matrix size, and the combination coefficient. Afterwards, the subjective and objective evaluation measures were applied to estimate the efficiency of the proposed method in comparison with the other conventional gradient methods. Finally, the proposed method was successfully applied to detect the intramuscular fat boundaries in an image of meat slice, as a life science engineering application.

3.1. The outputs of the proposed method

The proposed edge detector was applied on a gray level test image of human face, hair, and background frames, which is given

in Fig. 9a, with a mask size of 5∗5, a flux check matrix of size 3∗3 and the mask represented pixel is its centroid. The energy image provides a discernable contrast smoothed image, as shown in Fig. 9b, while the skewness image provides a smoothed detailed water mark image, as shown in Fig. 9c. The edges found in the energy and skewness images are shown in Fig. 9d and e, respectively. The output edges image, given in Fig. 9f, is a combined version of the energy and skewness edges, with a combination coefficient $w=0.75$

Fig. 9f clarifies that the output edges image handles the large scale edges such as background frames and shoulders boundaries, as well as the fine scale edges such as hair, eyes, and nose details.

3.2. Sensitivity to the mask represented pixel and mask size

As stated before, either the center or centroid pixel is chosen to represent the mask. But results show how the detection behavior is different for both choices, as shown in Fig. 10b and c for an exact edges image contains some primitive features given in Fig 10a. The objective evaluation measures, showing a significant lower ERMS, higher SNR, lower localization error, and lower energy dissipation for the mask centroid case, as presented in Table 1. Therefore, the mask centroid was selected to represent the mask in the proposed detection method.

An investigation was also made for the influence of the mask size on the detection results, since it does not only affect the resolution of the represented image, but also the accuracy in calculating the gradient image. For two different natures images, which are the primitive features image of Fig. 10a and a
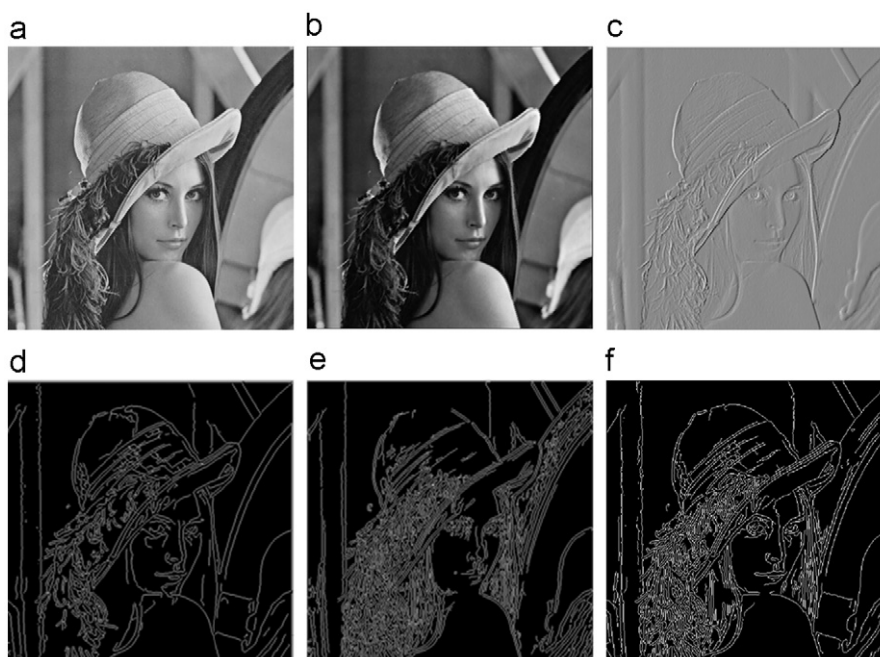
Fig. 9. (a) A gray level test image, (b) its energy image, (c) its skewness image, (d) the energy edges image, (e) the skewness edges image, and (f) the final edges image as a combination between the energy and skewness edges images (mask size=5*5, represented pixel is its centroid, flux check matrix size=3*3, mask, and $w=0.75$).

Fig. 10. (a) Gray level image with primitive features, and its edges image produced by the proposed method (mask size=5*5, flux check matrix size=3*3, $w=0.75$) with a mask represented pixel set at (b) mask centroid, and (c) mask center.

fingerprint image given in Fig. 11a, the proposed method was applied with several mask sizes ($n*n$) and $n=3, 5, 7, 9, 11, 13$, and 15. To implement the objective measures of performance evaluation, the exact edges of the images (i.e., $o(x, y)$) are required, which is quite not possible for many real world images. But for the two test images, they were assumed to be edges images, after being thinned by the non-maximum suppression algorithm. Afterwards, the averaged ERMS and SNR of the output images are plotted in Fig. 11b with respect to the mask size, implying that the best detection was achieved with a 5*5 mask size.

Table 1
Edge detection performance for the primitive features image by the proposed method for two different mask represented pixels, namely the mask center and the mask centroid.

Represented pixel	Global measures		Local measures (LE%, ED)
	ERMS	SNR	
Mask center	0.238	0.542	(23.405, 0.422)
Mask centroid	0.236	0.561	(22.375, 0.401)

3.3. Sensitivity to the flux check matrix size

The level of discontinuity coverage in the edges image is related to the flux equilibrium analysis and therefore the selected flux check matrix size. To observe the influence of this matrix size on the detection results, different sizes were examined on the fingerprint and primitive features images [mask size=5*5, mask represented pixel is its centroid, $w=0.75$], giving the detection and objective measures results displayed in Fig. 12. The corresponding values of $\overline{ERMS}$ and $\overline{SNR}$ reflect how the detection performance downgrades if no flux equilibrium check was applied, and is better and almost the same for different sizes flux check matrices. Therefore, a 3*3 flux check matrix was implemented in the proposed method.

3.4. Sensitivity to the combination coefficient

After obtaining an edges image for each of the energy and skewness images, a combination had to be performed to obtain one output edges image. The suggested formula for this combination, which was given in Eq. (9), depends not only on the edges given by
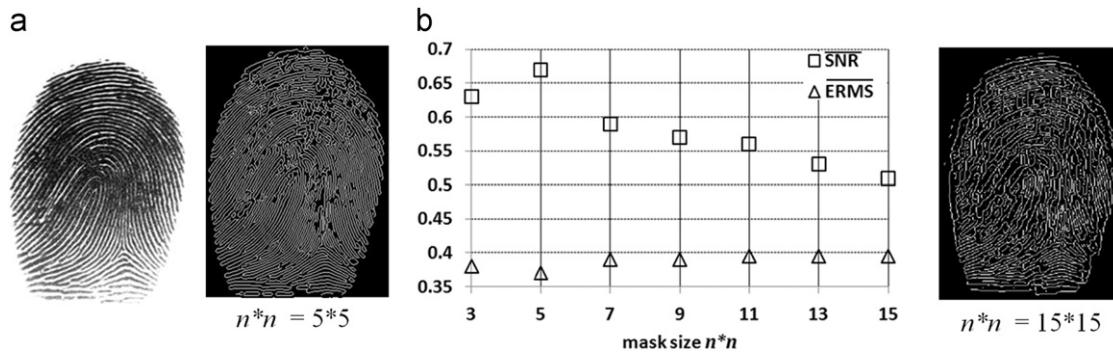
Fig. 11. (a) A gray level fingerprint image. (b) The variation of the $\overline{ERMS}$ and $\overline{SNR}$ with the mask size (mask represented pixel is its centroid, flux check matrix size=3*3, $w=0.75$). Where $\overline{ERMS}$ and $\overline{SNR}$ are the averaged values between those for the primitive features image and those for the fingerprint image.
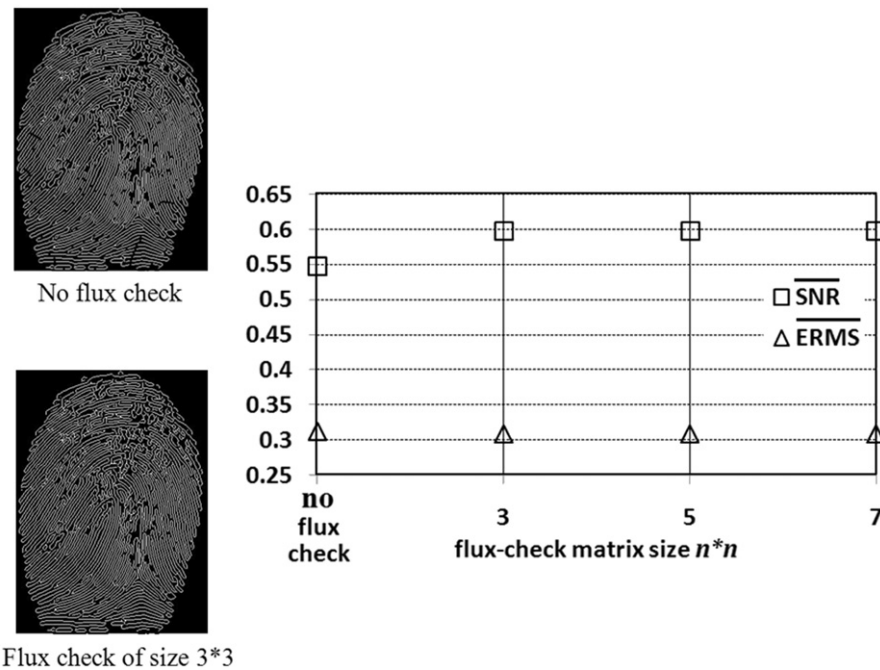
Fig. 12. The variation of the averaged $\overline{ERMS}$ and $\overline{SNR}$ with the size of the flux check matrix (mask size=5*5, mask represented pixel is its centroid, $w=0.75$). Where $\overline{ERMS}$ and $\overline{SNR}$ are the averaged values between those for the primitive features image and those for the fingerprint image.

each particular image, but also on the combination coefficient. Determination of the best value for this combination coefficient is required to produce the most outstanding detection results. Therefore, the proposed method was applied on the primitive features image of Fig. 10a, and the fingerprint image of Fig. 11a, with several combination coefficient values varies from 0 to 1 with a pitch of 0.05. Values for the SNR, ERMS, Localization Error, and Energy Dissipation are averaged for the two test images, producing the results plotted in Fig. 13. When $w=0$, the output edges image is identical to the skewness edges image, and when $w=1$, it is the energy edges image.

From Fig. 13, the variations in the Localization Error and Energy Dissipation are very small with the variation of w. On the other hand, the best values for SNR, ERMS were achieved at $w=0.75$, so it was chosen as the combination coefficient value for the proposed method.

3.5. Comparison to the conventional edge detection methods

3.5.1. Test case 1: fingerprint image

The subjective evaluation measures were used to evaluate the performance of the proposed method in detecting the edges of the fingerprint image of Fig. 11a, in comparison with the results of Roberts, Prewitt, Sobel and Canny detectors, as shown in Fig. 14. By visual inspection of the edges images given in Fig. 14, the proposed method produces more accurate and detailed edges image.

The objective evaluation measures were also calculated and presented in Table 2 for the edges given by canny detector (Fig. 14d) and the proposed detector (Fig. 14e). From the results given in Table 2, the proposed method achieved higher SNR, less ERMS, less LE%, and higher ED, in comparison to the Canny detector, indicating the efficiency of the proposed method in
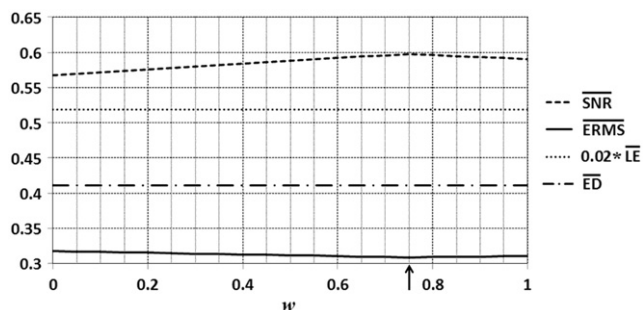
detection and identification of the fingerprints, but with slightly more energy dissipation than the detection of Canny. In general, increment of the ED value means that the suppression technique requires more improvement.

3.5.2. Test case 2: primitive features image

Another experiment was applied to the exact edges image of the primitive features of Fig. 10a, and the ERMS, SNR, LE%, and ED for both Canny detector results, which are displayed in Fig. 15a, and the proposed detector results, which are displayed in Fig. 15b, are presented in Table 3. The results clarify a significant better performance for the proposed method with respect to the first three measures, and a little higher ED value which is thought not to change the results accuracy. This example shows how the proposed method works reasonably in the text recognition.

3.6. Application to fat content prediction in marbled meat

The proposed edge detection method was used to predict the fat content in marbled meat images. Traditionally, the gravimetry ether extraction method [25] was extensively known as a standard reference analysis method for measuring the fat content in meat. The method uses an organic solvent (ether) extract of a dried sample followed by gravimetric measurement of fat. It was known as a simple, accurate and robust technique for estimating the fat content in meat [26]. However, it had several drawbacks such as its long drying and extraction times, lack of automation, and the amount of solvent used per sample, moreover, it is considered as a destructive testing (DT) method. These drawbacks can be avoided by using the image processing and edge detection techniques, and therefore improves the evaluation of the meat quality through prediction of fat content. However, the usages of image processing within the meat products have some common problems such as the specular reflections caused by moisture on the meat surface. Therefore, a pre-processing step is often used to improve the image quality and increase the accuracy of the analysis outputs [27].

Fig. 13. The variation of the objective evaluation measures with the combination coefficient (mask size=5*5, mask represented pixel is its centroid, flux check matrix size=3*3). Where $\overline{SNR}$, $\overline{ERMS}$, $\overline{LE}$ and $\overline{ED}$ are the averaged SNR, ERMS, localization error, and energy dissipation between those for the primitive features image and those for the fingerprint image.

Table 2
Edge detection performance for the fingerprint image by Canny and the proposed methods.

Method	Global measures		Local measures (LE%, ED)
	ERMS	SNR	
Canny	0.390	0.493	(34.661, 0.456)
Proposed	0.381	0.635	(28, 0.458)
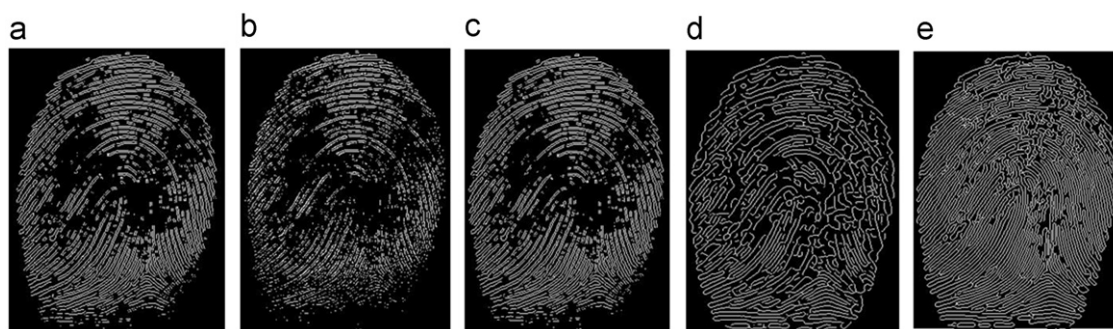
Fig. 14. (a) Edges image by 3*3 Prewitt edge detector, (b) edges image by 2*2 Roberts detector, (c) edges image by 3*3 Sobel detector, (d) edges image by Canny detector with filter size of 1.5, and (e) edges image by the proposed detection method (mask size=5*5, mask represented pixel is its centroid, flux check matrix size=3*3, $w=0.75$).
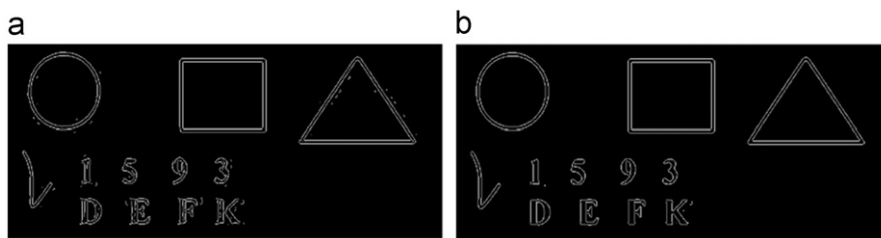
Fig. 15. (a) Edges image by Canny detector with filter size of 1.5, and (b) edges image by the proposed detection method (mask size=5*5, flux check matrix size=3*3, w=0.75).

Table 3
Edge detection performance for the primitive features image by Canny and the proposed method.

Method	Global measures		Local measures (LE%, ED)
	ERMS	SNR	
Canny	0.240	0.527	(24.277, 0.397)
Proposed	0.236	0.561	(22.373, 0.398)
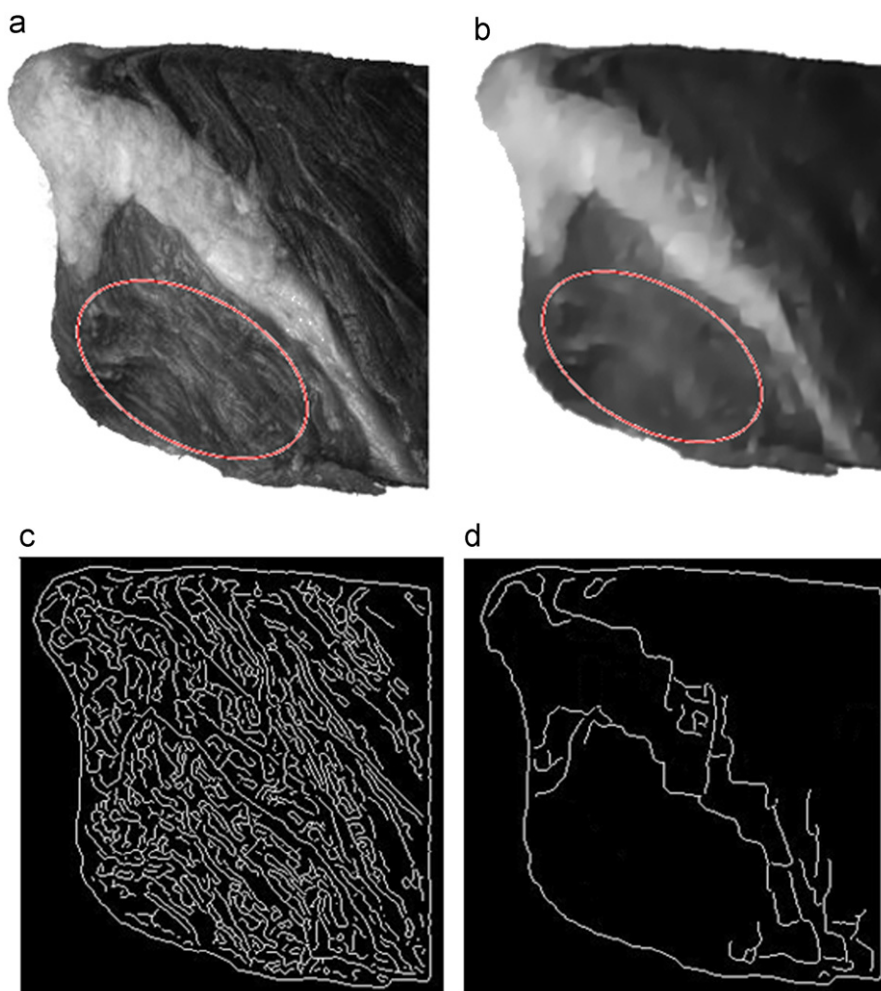
Fig. 16. (a) Part of a meat image showing the reflection of light from the meat surface due to the moisture and the muscle fibrous texture, (b) the meat image after applying the anisotropic diffusion filter, (c) edges image for the original meat image showing the false edges due light speckles, and (d) edges image after removing the speckles.

In this study, the pre-processing step started with the elimination of the background area from the marbled meat slices images using the histogram thresholding analysis techniques [28]. Afterwards, the anisotropic diffusion filter, which is given by Eq. (14), was applied to reduce the specular reflections and speckles without influencing the edges of the images

objects

$$\frac{dI}{dt} = div(D_{(x,y,t)} \Delta I) + \nabla D \nabla I \tag{14}$$

where div is the divergence operator, D is the diffusion coefficient and it is assumed to be function of the gradient of the color intensity $D_{(x,y,t)} = f(\|\nabla I_{(x,y,t)}\|)$, I is the pixels color intensity and ∇ and Δ indicates the gradient and Laplacian operators, respectively, with respect to the space variables.

The anisotropic diffusion filter works to reduce the noises and speckles without removing the significant parts of the image content, typically edges, lines or other details that are important for the interpretation of the image [29]. The application of the anisotropic diffusion filtration is shown in Fig. 16, for one of the meat images. The original image, as displayed in Fig. 16a, has specular reflections and speckles due to the light reflection from the moist areas and also from the meat muscle fibrous surface texture. The filtered image, which is shown in Fig. 16b, is free from these reflections and speckles to be more convenient for recognizing the fat and meat regions and not to analyze the surface texture. The edges image for the original and the filtered images showed the reduction of the false edges due light speckles (reflecting from the flesh texture) after the filtration process, as given in Fig. 16c and d, respectively.

After filtration, the proposed edge detection method was applied in combination with the region growing method to enhance the detection accuracy of the fat content in meat. In general, the region growing method use initial seed points to group the individual pixels of an image into regions if they have similar average gray level [30,31]. The formulation of the method can be given by

$$Region(R) = \bigcup_{i=1}^{n} R_i \; for \; all \; \cap_i R_i \neq \emptyset \tag{15}$$

$$R_i \cap R_j = \emptyset \; for \; all \; i \neq j \tag{16}$$

$$H(R_i) = True \; for \; all \; connected \; regions \; (i) \tag{17}$$

where R_i and R_j are two independent groups of connected regions, $(i, j) = 1, 2, \ldots, n$ and H is homogeneity criterion that control the selection of the enclosed regions.

The selection of a homogeneity condition is depending on the marginal difference in color intensity (Δf) between the starting seed and the neighborhood pixels. The algorithm examines the color intensity of the neighborhood pixels and decides whether to stop the region growth or continue. The edge detection algorithm was used to derive an initial set of seed points. The regions are grown from these initial seed points relative to the region pixel and edge pixel. Image edges act as hard barriers during region growing.
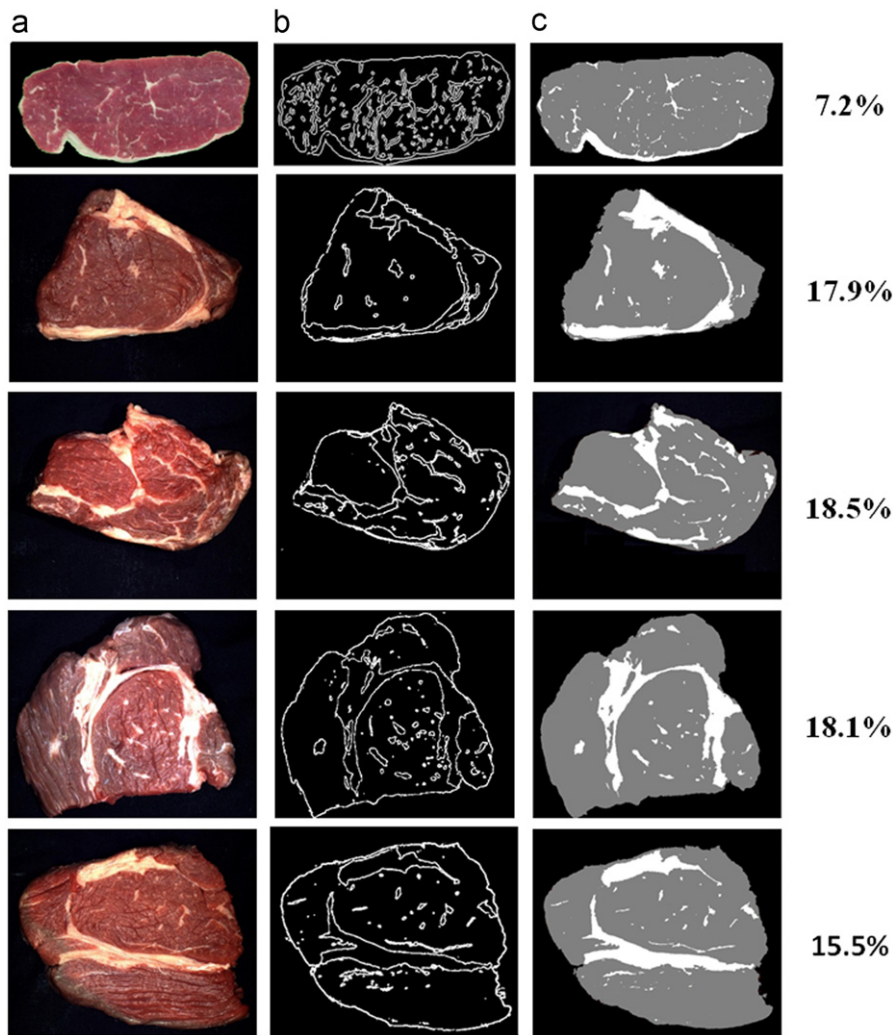
Fig. 17. (a) The original marbled meat images, (b) the edges images using the proposed method (mask size=5*5, flux check matrix size=3*3, w=0.75), (c) the detected fat regions after applying the region growing method, and (d) the percentage of fat content in the images.

For marbled meat samples of different fat content, the proposed edge detection method was applied on their digital images, as shown in Fig. 17a for five selected samples, producing the edges of the fat regions, as shown in Fig. 17b. The participation of the skewness feature in the edge detector made the algorithm sensitive to the small color variations. Therefore, it helped the model to overcome un-resolute distinguishing between the fat and the connective tissues [32]. The edges image was then post processed by the region growing method to recognize the sizes of the detected fat regions, as shown in Fig. 17c. Afterwards, the total percentage η of fat in the meat area was calculated in Fig. 17d, using the formula of Eq. (17). Integration of the proposed method for surface fat content recognition with an ultrasound technique for under layer fat content recognition, will present a complete recognition system in general three dimensional non-uniform marbled meat

$$\eta = \frac{\int_{\Omega} f_{Fat}(x,y)\, dA}{\int_{\Omega} dA} \times 100 \tag{18}$$

where Ω is the total meat area and $f_{Fat}(x, y)$ is the fat presence parameter, which equals one if the pixel is on fat region, and zero otherwise.

4. Conclusion

This work is contributed for automating fat content measurement by means of image processing, since fat content is one of the most important parameter in determining meat quality. Existing traditional evaluation techniques are time consuming, expensive and destructive. Image processing techniques and particularly edge detection techniques have shown the ability to identify the boundaries of the intramuscular fat and therefore the amount of fat content in the marbled meat slices images. However, most of the gradient based edge detectors suffer from sensitivity to noise because the original images are not pre-smoothed. Furthermore, smoothing the original image with Gaussian filter, does not promise optimum detection level, because it is dependent on the filter size. In this study, an edge detection method was developed based on combination and enhancement of two novel image features, which are image energy and image skewness, as smoothed versions of the original image. Subsequently, a flux equilibrium check technique on the edge direction was defined and applied to avoid discontinuity in the resultant edge image. The results implying that the proposed method is more enhanced with significant improvements in comparison to other conventional gradient based edge detection methods, with respect to both subjective and objective evaluation measures.

Further work will be focused on the application of higher order gradient operators, in accordance testing the availability to implement the lattice method to improve the propagation of the mask on the entire image. Ultrasonic detection techniques, which are as well reliable and non-destructive, may also be combined with the proposed method to detect the under layer fat content, and have the overall 3-D fat content for the samples.

References

[1] M.P. Penfield, C.A. Costello, M.A. McNeil, M.J. Rienmannn, Effects of fat level and cooking methods on physical and sensory characteristics of reestructure beef streaks, Journal of Food Quality 11 (1989) 349.

[2] B. Touraine, P. Vigneror, C. Touraille, M. Prud'hom, Influence des onditions d'elevage sur les characteristiques des carcasses et de la viande d'agneaux Merino d'Arles, Bulletin Technique de l'Elevage Ovin 4 (1984) 29.

[3] C. Du, D. Sun, Shape extraction and classification of pizza base using computer vision, Journal of Food Engineering 64 (2004) 489–496.

[4] S. Ötles, A. Önal, Computer-aided engineering software in the food industry, Journal of Food Engineering 65 (2004) 311–315.

[5] A.M. Wyle, D.J. Vote, D.L. Roeber, R.C. Cannell, K.E. Belk, J.A. Scanga, Effectiveness of the SmartMV prototype BeefCam System to sort beef carcasses into expected palatability groups, Journal of Animal Science 81 (2003) 441–448.

[6] H.R. Cross, D.A. Gilliland, P.R. Durland, S. Seideman, Beef carcass evaluation by use of a video image analysis system, Journal of Animal Science 57 (4) (1983) 910–917.

[7] R.L. Wassembergn, D.M. Allen, K.E. Kemp, Video image analysis prediction of total kilograms and percent primal lean and fat yield of beef carcasses, Journal of Animal Science 62 (6) (1986) 1609–1616.

[8] D.E. Gerrard, X. Gao, J. Tan, Beef marbling and color score determination by image processing, Journal of Food Science 61 (1) (1996) 145–148.

[9] J. Li, J. Tan, F.A. Martz, H. Heymann, Image texture features as indicator of beef tenderness, Meat Science 53 (1999) 17–22.

[10] K. Shiranita, T. Miyajima, R. Takiyama, Determination of meat quality by texture analysis, Pattern Recognition Letters 19 (1998) 1319–1324.

[11] K. Shiranita, K. Hayashi, A. Otsubo, T. Miyajima, R. Takiyama, Grading meat quality by image processing, Pattern Recognition 33 (2000) 97–104.

[12] T. Lindeberg, Edge detection, in: M. Hazewinkel (Ed.), Encyclopedia of Mathematics, Kluwer/Springer, 2001.

[13] B. Singh, A.P. Singh, Edge detection in gray level images based on the Shannon entropy, Journal of Computer Science 4 (3) (2008) 186–191.

[14] M. Juneja, P.S. Sandhu, Performance evaluation of edge detection techniques for images in spatial domain, International Journal of Computer Theory and Engineering 1 (5) (2009) 614–621.

[15] I. Sobel, Camera Models and Perception, Ph.D. Thesis, Stanford University, Stanford, CA, 1970.

[16] J. Prewitt, Object enhancement and extraction, in: B. Lipkin, A. Rosenfeld (Eds.), Picture Processing and Psychopictorics, Academic Press, New York, , 1970.

[17] G.S. Robinson, Edge detection by compass gradient masks, CGIP Journal 6 (5) (1977) 492–501.

[18] M. Sharifi, M. Fathy, M.T. Mahmoudi, A classified and comparative study of edge detection algorithms, in: Proceeding of the International Conference on Information Technology: Coding and Computing, 2002, pp. 117–120.

[19] E. Nadernejad, S. Sharifzadeh, H. Hassanpour, Edge detection techniques: evaluations and comparisons, Applied Mathematical Sciences 2 (31) (2008) 1507–1520.

[20] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1986) 679–714.

[21] L. Ding, A. Goshtasby, On the canny edge detector, Pattern Recognition 34 (3) (2001) 721–725.

[22] M.A. Hussein, T. Becker, Numerical modelling of shear and normal stress of micro-porous ceramics for stimulated in-vitro cultivation of bone cells, Microfluidics and Nanofluidics 8 (5) (2010) 665–675.

[23] W.B. Hussein, M.A. Hussein, T. Becker, Detection of the red palm weevil *Rhynchophorus ferrugineus* using its bioacoustics features, Bioacoustics 19 (3) (2010) 177–194.

[24] M. Roushdy, Comparative study of edge detection algorithms applying on the gray scale noisy image using morphological filter, GVIP Journal 6 (4) (2006) 17–23.

[25] AOAC, Official Methods of Analysis, 14th ed., Washington, DC, 1991.

[26] F.J. Eller, J.W. King, Supercritical fluid extraction of fat from ground beef: effects of water on gravimetric and GC-FAME fat determinations, Journal of Agricultural and Food Chemistry 49 (2001) 4609–4614.

[27] M.R. Chandraratne, Vision based methodology for evaluation of meat quality characteristics, Ph.D. Thesis, Lincoln University, Lincoln, 2004.

[28] L.G. Shapiro, G.C. Stockman, Computer Vision, Prentice-Hall, New Jersey, 2002.

[29] G. Sapiro, Geometric Partial Differential Equations and Image Analysis, Cambridge University Press, Cambridge, 2001.

[30] T. Pavlidis, Y.T. Liow, Integrating region growing and edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 3.

[31] L. Gao, J. Jiang, S.Y. Yang, Constrained region-growing and edge enhancement towards automated semantic video object segmentation, Lecture Notes in Computer Science 4179 (2006) 323–331.

[32] L. Ballerini, A. Högberg, K. Lundström, G. Borgefors, Colour image analysis technique for measuring of fat in meat: an application for the meat industry, in: Hunt(Eds.), Machine Vision Applications in Industrial Inspection IX, San Jose, CA, Proceedings of the SPIE 4301 (2001) 113–124.

Walid B. Hussein born 1978, received his B.Sc. and M.Sc. with Honour Grade in Aerospace Engineering from Cairo University, Egypt in 2000 and 2005, respectively. During this period, he worked as research assistant and teacher in the Egyptian Civil Aviation Academy & German University in Cairo. His research interests are in the area of signal and image processing, chemometrics and production optimization, with particular consideration of bioacoustics and pattern recognition. Between 2007 and 2009, he worked as researcher and Ph.D. student in Department of Life Science, University of Hohenheim, Germany. Afterwards, he moved as a Ph.D. student to the Center of Life Science Engineering, Technische Universität München, Germany.

Ahmed A. Moaty born 1980, Received his B.Sc. and M.Sc. in Aerospace Engineering from Cairo University, Egypt in 2001 and 2006, respectively. During that period he worked in the National Authority for Remote Sensing and Space Sciences, where he worked in satellite weather image analysis and weather modeling. Between 2007 and 2009, he worked as researcher and Ph.D. student in Department of Life Science, University of Hohenheim, Germany His research interests are in the area of image processing and modeling, with special focus of using Boltzmann methods in image modeling. Afterwards, he moved to Center of Life Science Engineering, Technische Universität München, where he continued his Ph.D. work.

Mohamed A. Hussein born 1977, received his B.Sc. and M.Sc. with Honour Grade from Aerospace Engineering, Cairo University, Egypt in 2000 and 2003, respectively. From 2000 to 2006 he worked as research assistant and teacher in the Aerospace Department, Cairo University & German University in Cairo. From 2006 to 2008 he worked as Scientific employee and Ph.D. Student in the Universität Hohenheim-Stuttgart, primarily focused on Numerical Modeling, Finite methods, Numerics in Signal and Image Processing. Since 2008 he became the Group Leader of Process Analysis in Universität Hohenheim. Moved in 2009 to Technische Universität München, where he submitted his Ph.D. thesis entitled "On the Theoretical and Numerical Development of lattice Boltzmann Models for Biotechnology and its Applications" and held the group leader post of Bio-Process Analysis Technology & Process Analysis (PAT).

Thomas Becker born 1965, studied Technology and Biotechnology of Food at the Technische Universität München. He received his Ph.D. in 1995: "Development of a computerized enzyme integrated flow injection system and its use in the biotechnological process and quality control." From 1996 to 2004 he worked as Deputy Head of Department and assistant lecturer at the Institute of Fluid Mechanics and Process Automation of Technische Universität München. His habilitation took place in 2002: "Management of Bioprocesses by means of modeling and cognitive tools." In the period from 2004 to 2009 he worked as university professor at the University of Hohenheim (Stuttgart)—until 2005 as head of the Department of Process Analysis and Grain Technology, followed by as head of Department of Process Analysis and Cereal Technology. In 2009, he became the professor (Ordinarius) of Beverage Technology Department at Technische Universität München.

86

3.3   Species taxonomy by bioacoustics classification

The developed spectrogram enhancement approach is applied to clean spectrograms of bioacoustics calls generated by birds and bat species. Since bat echolocations are ultrasound calls, they are expanded in time domain with a factor of 10 or 13 to be in the human hearing range.

Due to its edge detection based method, the enhanced spectrograms by the developed approach return almost the same temporal and spectral structure of the sound patterns. Most of other traditional approaches interact with the spectrogram during enhancement, either in time domain (through wave form of the call) or in frequency domain (through power spectrum and frequency contents of the call). As a result, the structure and shape of the produced sound patterns may be altered.

Having an enhanced spectrogram with undestroyed sound patterns eases the identification and classification processes.

Few two dimension features, which define sound patterns, are extracted for some bird and bat species. These features are employed to classify spectrograms of candidate calls into their associated species, as described in figure (3.4).
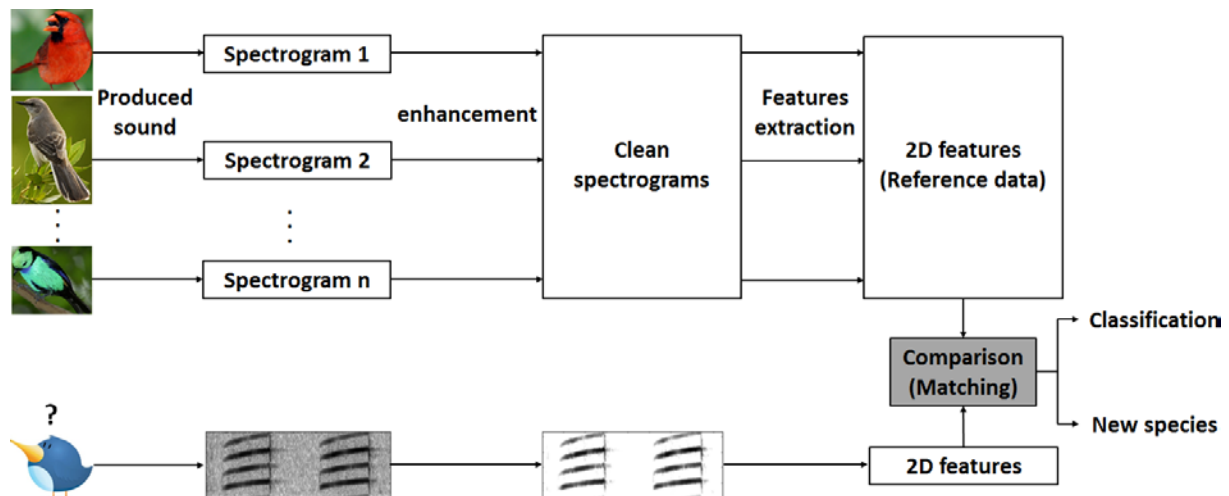
Figure (3.4): Description for the classification of sound producing species by the 2D features of their enhanced spectrograms.

Additionally, the representation of a sound pattern by its edges boundary (i.e., contour) greatly reduces the amount of computation for feature extraction. The reason is that contour pixels are generally a small subset of the total number of pixels in a sound pattern.

Spectrogram Enhancement By Edge Detection Approach Applied To Bioacoustics Calls Classification

W.B. Hussein[*], M.A. Hussein and T. Becker

Group of (Bio-) Process Technology and Process Analysis, Faculty of Life Science Engineering, Technische Universität München, Germany
whussein@wzw.tum.de

ABSTRACT

Accurate recognition of sound patterns in spectrograms is important step for further recognition applications. However, background noise forms fundamental problem regardless the species under study. In this paper, crest factor feature was extracted from the limited dynamic range spectrogram. The developed crest factor image behaved as smoothed version of the spectrogram, at which edges of the involved sound patterns were detected without the need of prior smoothing filters and their scaling constraints. Attached noise – surrounds the detected edges – was removed, to form the enhanced spectrogram. The method was compared to other enhancement approaches such like spectral Subtraction and wavelet packet decomposition. Comparison was performed on different structure patterns of bats and birds. Results indicate how the method is promising for efficiently enhancing the spectrogram while preserving its temporal and spectral accuracy. The method correctly classified three bioacoustics species with an accuracy of 94.59%, using few 2D features of their enhanced spectrograms
.

Keywords

Spectrogram Enhancement, Edge Detection, Bioacoustics Classification, Pattern Recognition.

1. Introduction

1.1. Problem formulation

Bioacoustics calls have been efficiently employed for a long time for species detection, classification, and recognition. These calls handle the sound patterns which are almost unique and oriented for the investigated bio-source. Several temporal and spectral sound features are extracted from these patterns in time and frequency domains; respectively. These features are used to train and develop a learning system, using methods such as Artificial Neural Network, which afterwards is able to successfully recognize the test bioacoustics calls to their corresponding groups or species, according to their features values. The approach have been widely used in many life science problems, such as the bioacoustics detection of hidden grain weevils for early treatment [1], and detection of bat ultrasound echolocation calls in the windmills region to avoid their expected collision with the blades[2].

In general, collection of time domain and frequency domain features are used to develop more accurate detection system, revealing the importance of having a reliable spectrogram (time-frequency intensity 2D image) of the specified sound.Spectrogram is an important representation of sound data looks like the human hearing which is based on a kind of real-time spectrogram

1

encoded by the cochlea of the inner ear [3], to classify and recognize patterns of sound samples. However, spectrogram is usually attached by different forms of noise; including those formed during sound recording[4],and those produced during the transformation to frequency domain result in spectral leakage, and up to 10% error(s) infrequency and/or power spectrum computations [5]. These noises directly degrade the quality of the waveform, deteriorate the worth of the extracted features and thus lead to inaccuracy in recognition of the sound patterns [6].Further difficulty is added to the problem due to variability of patterns structure, which can vary greatly including vertical straight, sloped straight, sinusoidal type and relatively random patterns. However, filtering noisy signals through the spectrogram is considered more effective than separate filtrations in time and/or frequency domain, since sound patterns do not cover the whole spectrogram image, and therefore easier to filter off the noise.

A spectrogram enhancement approach which is independent on the noise type, level, and structure is required. Once established, the pattern recognition algorithms can operate efficiently and smoothly on the clear "only patterns" spectrogram. Therefore, the problem of spectrogram enhancement and accurate detection of the sound patterns has attracted researchers' interest from a variety of backgrounds ranging among signal and image processing, and statistical models [7,8].

1.2. Related work

Common and recent techniques for spectrogram enhancement include basic band pass filtering[9], spectral subtraction[10], Wiener filter[11], and wavelet packet decomposition[12, 13].Simple methods, such as the band pass filtering, originally employed the use of time-domain filtering of the corrupted signal, however, this is only successful when removing low or high frequency noise and does not provide satisfactory results for many species which have frequency range overlaps with their attached noises [14].

Although the base spectral subtraction method is very simple and efficient, it assumes the noise to be additive and uncorrelated with the signal [15]. Moreover, the enhancement by spectral subtraction tends to produce sounds with musical artifacts that are often more objectionable than the original noise [16]. Later, the multi-band spectral subtraction method was proposed, at which the corrupted sound is initially divided into several frequency bands, and then the spectral subtraction method is applied to each band[17]. This method outperforms the standard power spectral subtraction method resulting in superior spectrogram quality and largely reduced musical noise. Meanwhile, the Wiener filter technique basically considers the beginning of a signal is noise, and its adaptive type removes noise based on a training data [18]. However, during the operation on data with unknown noise, the noise level can be underestimated and the enhancement can be slightly milder [19]. The methods based on wavelet packet decomposition are effective in removing background noise in the spectrogram. But they cannot suppress much of the noise generated during the Fourier transformation, because the former noise is usually random Gaussian distribution while the latter may be modeled by Rayleigh distribution [7,20].

Image analysis techniques applied to this area treat the spectrogram as an image, provides a wide range of methods which could be beneficial to this problem. One of these developed methods is the noise suppression using spectrogram morphological filtering [21,22], applying two subsequent operations of erosion and dilation. The erosion was responsible to remove noise from the noisy spectrogram while dilation used to restore any erroneously removed sound patterns. However, it improves the enhancement accuracy by only 10% when combined by nonlinear spectral subtraction with a suitable selection of the threshold. The author in [23] proposed an edge detection method which initially smoothes the spectrogram using a Gaussian filter, followed by thresholding each point by comparison to the background measurement. This allows for time invariant noise conditions and computing independently for each frequency bin, which successfully detected (90%) of whale calls. If the smoothing kernel is quite large, the detrimental effect is reduction in the detection accuracy, especially at low SNR. Meanwhile, the authors in [24] passed the spectrogram through 2D bilateral filter to reduce noise and preserve its patterns

2

edges. The filtered spectrogram is sent to two parallel processing paths, at which the first path extracts significant patterns from background noise. The second path performs the edge detection and restructures the rough patterns that can be used as a mask. The processed image from the first path passes through a mask generated from the second path. However, pattern recognition still severely depends on image processing skills and spectrogram resolutions which lead to concealment of very short patterns. Apartial differential equation technique was used in [25] for edge enhancement and noise reduction based on a regularization of the mean curvature motion equation. However, the comparisons indicate that the method gives almost similar results as the wavelet based methods.

In this paper, an improved spectrogram enhancement method has been developed based on the last advances in the edge detection techniques. The dynamic power range of the spectrogram is limited to avoid the problem of low level portions of the spectrogram expanding and thereby obscuring the detail of the energetic portions. Afterwards, the crest factor image is calculated as a smoothed version of the original spectrogram image, hence escaping the application of smoothing Gaussian filters and their drawbacks [26].Based on the edge detection algorithm presented in [28], the sound patterns in the crest factor image are detected. Afterwards, the original power values of the patterns edges and their interior are reconstructed, while the power values of the patterns surrounding are eliminated, as they represent the attached noises whether attached to the sound or generated during the frequency domain transformation. The proposed method was applied to several bioacoustics calls of different SNR values, and compared to the results given by band pass, multi-band spectral subtraction, Wienerfilter, and wavelet packet decomposition methods, with respect to subjective and objective measures. Finally, possible implementations of the proposed method in obtaining the enhanced frequency and power contours, reconstruction of the enhanced waveform, and simplified pattern recognition operation are presented.

2. Material and method

2.1.Signal processing

Audio sound streams are sampled in time domain with suitable sampling frequencies, selected to be higher than the double of maximum frequency in the sound stream, satisfying the Nyquist sampling theorem [5] and avoid antialiasing in the signal reconstruction. Figure (1) shows an example for the call of Sitta canadensis bird which was sampled at 11025 Hz. The signal is divided into segments with length of 1% of the total signal length and 90% overlapping percentage. Each segment is then multiplied by Bartlett window function and transformed to frequency domain through Fast Fourier Transform (FFT). The frequency domain representation of the signal (i.e., spectrogram) is the power spectrum distribution with frequency, at each time instant, as plotted in figure (2).The implementation of the Bartlett window function is to have better frequency resolution while keeping acceptable spectral leakage and amplitude accuracy [27].

The resultant spectrogram contains important sound patterns of the signal immersed in attached noise. These noises are not only due to the base noise attached to the sound, but also generated during FFT, therefore, cleaning the signal in time domain, will not ensure clean spectrogram.

3
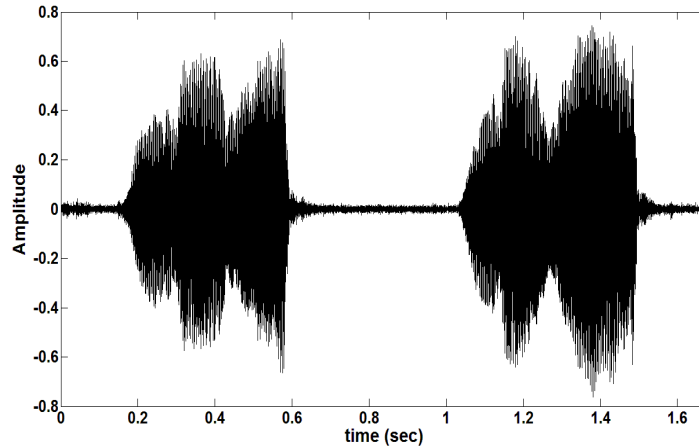
Figure 1. A sound stream for Sitta canadensis bird in time domain revealing its contents of two long pulses and one long inter-interval.
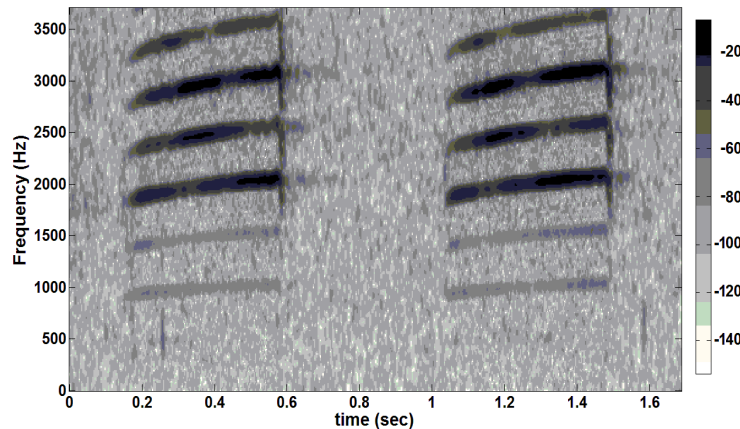
Figure 2. The spectrogram for the time domain signal of figure (1), after being divided into segments of length equal to 1% of the total signal length, multiplied by Bartlett window function and transformed to frequency domain using FFT.

2.2. Limiting the dynamic range

The attached noise to the spectrogram may be assumed to have almost same power value, which can be removed from the whole spectrogram. However, this will eliminate as well the non-noisy patterns which have this power value.Therefore, this thresholding scheme should be carefully applied through the physical fact of the limited dynamic range. Althoughthe whisper cannot be heard in loud surroundings, the spectrogram will contain all details about whisper and loud sound powers. Thus, the spectrogram powers have to be limited, to avoid much of the attached whisper (i.e., noise).The range is limited to 40 dB below the maximum value for all tested sounds, because most bioacoustics signals are expanded/slowed to the human speech range, which is normally perceived over this range [29]. Therefore, any point with power value outside this range, including those of noise as well as very weak patterns, are eliminated from the spectrogram, as shown in figure (3).

4
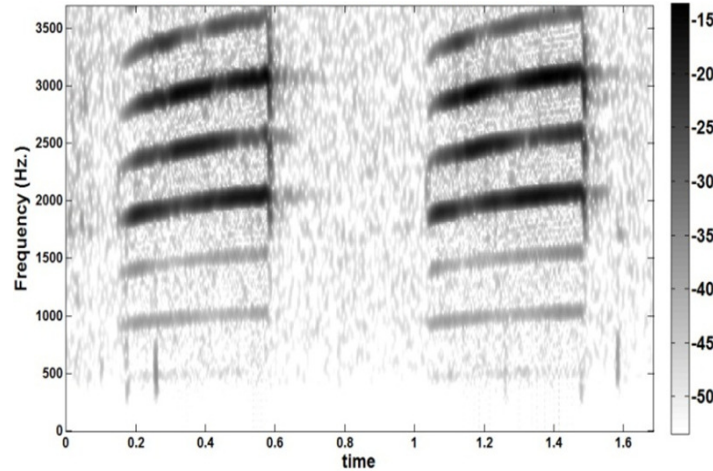
Figure 3. The limited spectrogram after the power values were limited to a dynamic range of 40 dB, clarifying how most of the noise and very weak sound patterns were eliminated.

2.3. Detection of the pattern edges

The algorithm starts by sliding a 5*5 matrix (mask) over the limited dynamic range spectrogram image in *x* direction and then in *y* direction, with step of one pixel. The represented pixel of the mask is its centroid which is calculated by equation (1). The intensities of the pixels (i.e., power values) enclosed by the mask are used to calculate the crest factor feature, given in equation (2), which is a ratio of the maximum value to the root mean square value, indicating how much impacting is occurring inside the mask, as schematically explained in figure (4).

$$mask\ centroid = \sum_{i=1}^{n} i * f_i \setminus \sum_{i=1}^{n} f_i \quad (1)$$

$$C = \frac{f_{peak}}{f_{rms}} \quad (2)$$

Where f_i is the gray level intensity value of pixel i, and the back slash means that only the quotient of the division is considered. f_{peak} and f_{rms} are the peak and root mean square of the pixels intensities; respectively, and C is the crest factor of the mask.
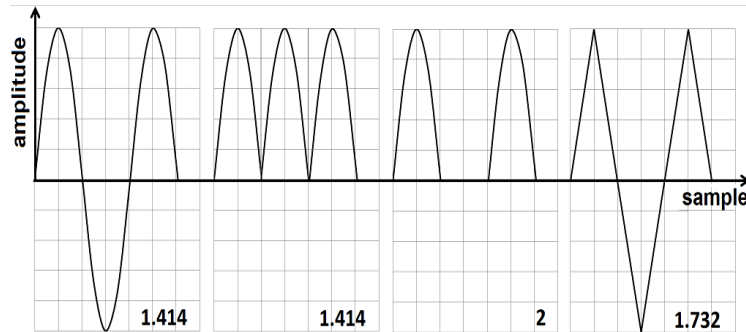
Figure 4. The crest factor for some basic curves, showing how much impacting occurs.

As a result, the crest factor image is obtained, by gathering the local crest factors calculated during the sliding of the mask. Although this edge detection algorithm follows the one presented in [28], it is applied to the crest factor image instead of combination of energy and skewness images, because this combination presents both strong edges (output of the energy feature) and weak edges (output of the skewness feature). Hence, the noises are also detected as patterns, displayed in figure (5a) with a signal to noise ratio (SNR) - given in equation (3) - of 1.95. Alternatively, the crest factor feature of an image presents the edges that have impact to their

5

surrounding (i.e., non-weak edges), as displayed in figure (5b) with SNR of2.82. Furthermore, the limited dynamic range makes the crest factor more meaningful since it is a measure of relative spatial intensity change.

$$SNR = 10log_{10}\frac{\sum P_{clean}}{\sum|P_{clean} - P_{output}|} \qquad (3)$$

Where P$_{clean}$ and P$_{output}$ are the clean and output power values; respectively.
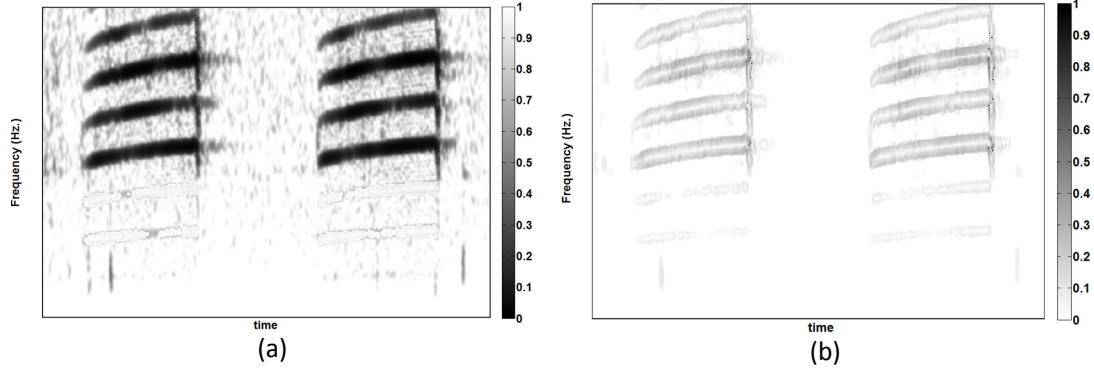
Figure5. Two features from the limited dynamic range spectrogram of figure (3), (a) combination of energy and skewness images, following the algorithm presented in [28], and (b) the crest factor image.

Afterwards, the Sobel operators given in equation (4) are employed on the crest factor image, through equation (5),to get its gradient images ($C_x$andC_y). And with the aid of equations (6) and (7), the edges strength (E) and the edges directions (θ) are calculated; respectively.

$$\Delta_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \Delta_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \qquad (4)$$

$$C_x = \Delta_x * C, C_y = \Delta_y * C \qquad (5)$$

$$E = \sqrt{C_x^2 + C_y^2} \qquad (6)$$

$$\theta = tan^{-1}\left(\frac{C_y}{C_x}\right) + \frac{\pi}{2} \qquad (7)$$

Where$\Delta_x$and$\Delta_y$are the derivative operators in x and y directions; respectively. C_x and C_y are the intensities of the gradient images in x and y direction; respectively. E is the edge strength and θ is the edge direction with the x-axis.

Finally, the edges image is formed by the values of edges strength (E), and executed by the non-maximum suppression algorithm and flux equilibrium check [28], to suppress thick edges to one pixel width and fill the missing pixels in the edge direction. Consequently, the final edges image, given in figure (6) is produced, which separates the patterns from surrounding noise. However, the edges do not provide information about where exactly are the inner of the patterns and where are their surroundings.

6
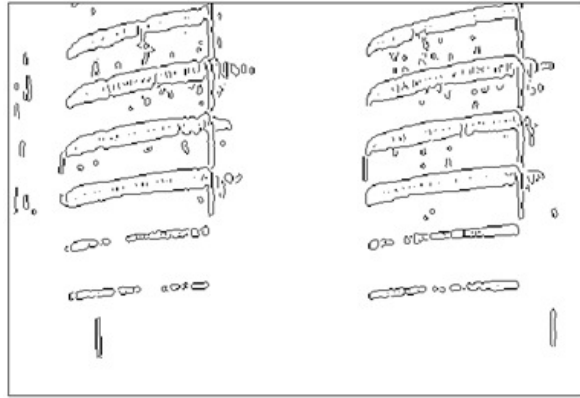
Figure 6. The edges image of the limited dynamic range spectrogram using 5*5 mask and a flux check matrix of size 3*3

2.4. Reconstructing the spectrogram

A classification condition is applied to each row and afterwards each column of the edges image. This condition compares the average power spectrum of all pixels among two subsequent edges in one row (column), with respect to the average power spectrum of these two edges, as clarified in the algorithm given in figure (7).

$I = \{i_1, i_2, \ldots, i_n\}, J = \{j_1, j_2, \ldots, j_m\}$
for $k = 1:n-1$

$$Pixels(i_k:i_{k+1}) = \begin{cases} \text{sound pattern} & ; mean(Power(i_k:i_{k+1})) \geq mean(Power(i_k, i_{k+1})) \\ \text{surrounding noise} & ; mean(Power(i_k:i_{k+1})) < mean(Power(i_k, i_{k+1})) \end{cases}$$

end

for $k = 1:m-1$

$$Pixels(j_k:j_{k+1}) = \begin{cases} \text{sound pattern} & ; mean(Power(j_k:j_{k+1})) \geq mean(Power(j_k, j_{k+1})) \\ \text{surrounding noise} & ; mean(Power(j_k:j_{k+1})) < mean(Power(j_k, j_{k+1})) \end{cases}$$

end
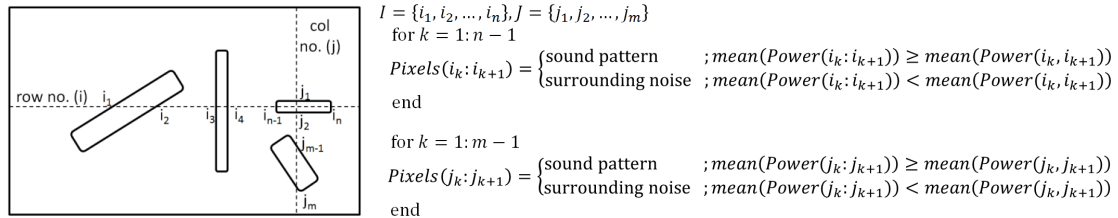
Figure 7. Schematic diagram and algorithm for the classification condition which classifies sound patterns from their surrounding noise.

Thus, the patterns are defined and their power values are restored, and the surrounding noises are also defined and their power values are eliminated, results in the enhanced spectrogram shown in figure (8).
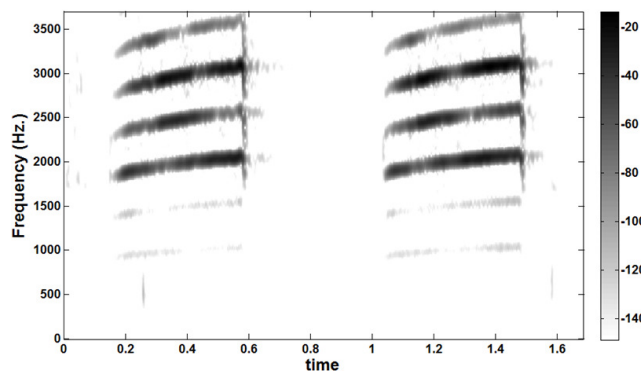
Figure 8. The enhanced spectrogram by the proposed method revealing sound patterns and eliminating their surrounding noise.

3. Experimental results and discussion

The results obtained by the proposed method [PM] were investigated and compared to those obtained by conventional and modern spectrogram enhancement methods. The designing parameters of these methods were carefully selected to give best enhanced spectrogram for the first application (Rhinolophus blasii bat), and were fixed over the following applications to have leading results and trustful comparisons. The first method is the widely used band pass filter [BP] with a band width enclosing the sound patterns, and measured at half-power points (i.e., gain -3 dB relative to peak). The second method is multi-band spectral subtraction [MBSS]using 4 linearly-spaced frequency bands, over subtraction factor of 4 and power factor of 1.5 [7]. The third method employs the Wiener filter [WF]with a spectral distance threshold of 3 and the initial 0.03 seconds considered as noise[11]. The fourth method is wavelet packet decomposition [WPD] with soft thresholding and 5 level decompositions using symlet 8 wavelet [14, 16]. The number of tested sound samples are 42 (each with 23-25 sec for Rhinolophus blasii bat), 37 (each with 23-25 sec for Barbastella barbastellus bat), 48 (each with 0.55-0.9 sec for Vanellus vanellus bird), and 45 (each with 0.55-0.9 sec for Parus major bird), with a frame length of0.025 sec multiplied by Bartlett window function, and 90% overlapping percentage.

The analysis was applied by both subjective and objective measures of enhancement accuracy. The subjective measure is borrowed from the field of psychology and the human judgment of evaluation. One of the commonly used subjective measures is the Mean Opinion Score (MOS),which gives a numerical estimation of the perceived quality of the media received [30]. After enhancing the spectrogram, its time domain signal was reconstructed and played back to 10 listeners. These listeners (5 females, and 5 males) were asked to give a score [1 = bad, 2 = poor, 3 = fair, 4 = good, and 5 = excellent] to estimate the enhanced spectrogram quality. Afterwards, the MOS was calculated by averaging the given scores, and its confidence interval (CI) was computed for 95% confidence level, as described in figure (9).
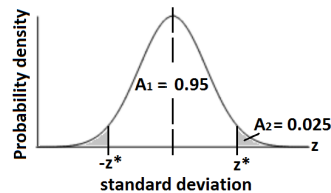
Figure 9. Normal distribution curve with 95% confidence level. $A_2 = P (z > z^*) = (1 - 0.95) / 2 = 0.025$, $P (z <= z^*) = 1 - 0.025 = 0.975$, results in $z^* = 1.96$ (from normal distribution table). The confidence interval is$(MOS - [1.96 * \sigma], MOS + [1.96 * \sigma])$. Where$[-z^*, z^*]$ encloses CI on the standard deviation axis (z) and σ is the standard deviation of the opinion score.

On the other hand, objective measures are borrowed from digital signal processing and information theory, providing equations that can be used to measure the enhancement accuracy of the enhanced spectrogram in comparison to the clean one. Four widely used and easy to implement objective measures were employed, having high correlation with diagnostic acceptability [11,30,31]. These measures are the overall Signal to Noise Ratio (SNR), Segmental Signal to Noise Ratio (SSNR), Log Spectral Distance (LSD), and Itakura Saito (IS), given by equations (3, 8, 9, and 10; respectively).SSNR is defined as the average of SNR values over segments with sound activity, LSD is the spectral distance or distortion measure, expressed in dB, between the enhanced and clean spectrograms, while IS a measure of the perceptual difference between these two spectrograms. Furthermore, the average eccentricity (AE) was calculated by equation (11), to simply check if the shape of sound patterns in the enhanced spectrogram was changed from those in the original spectrogram, results in indication of whether the enhanced spectrogram is augmented for further pattern recognition task. Eccentricity is the aspect ratio of length to width of the minimum rectangle bounding the sound pattern.

8

$$SSNR = \frac{10}{M} \sum_{m=0}^{M-1} log_{10} \sum_{i=Nm}^{Nm+N-1} \left(\frac{\sum P_{clean}}{\sum |P_{clean} - P_{output}|} \right) \quad (8)$$

$$LSD = \sqrt{\frac{2}{FS} \sum_{0}^{FS/2} \left[10log_{10}\left(\frac{P_{clean}}{P_{output}} \right) \right]^2} \quad (9)$$

$$IS = \frac{2}{FS} \sum_{0}^{FS/2} \left[\frac{P_{clean}}{P_{output}} - log_{10}\frac{P_{clean}}{P_{output}} - 1 \right] \quad (10)$$

$$AE = \frac{1}{K} \sum_{j=1}^{K} \frac{L_j}{W_j} \quad (11)$$

Where FS is the sampling rate of the signal, P_{clean} and P_{output} are the clean and enhanced power spectrum respectively. M is the number of spectrogram segments (set to 20), N is the number of samples on a segment. L and W are the length and width; respectively, for the minimum rectangle bounding the sound pattern, and K is the number of sound patterns in the spectrogram.

The applications were selected to cover different spectrogram shapes, including those with narrow band, wide band, constant frequency, frequency modulated, short pulses, and long pulses patterns. The original bioacoustics calls in each application were corrupted by several white Gaussian noises, as descriptively shown in the left part of figure (10) for a saw tooth wave, result in several time domain SNRs (30, 20, 10, 5, and 1 dB), and in correspondence several spectrogram SNRs as demonstrated in the right part of figure (10), with average values of (3.66, 2.89, 2, 1.73, 1.52 dB; respectively).For each application and for each SNR, the five enhancement methods (BP, MBSS, WF, WPD, and PM) were applied, and the results were subjectively and objectively compared. It is worthy to mention that the silent regions were being removed, because they can considerably influence the output objective measures.

9
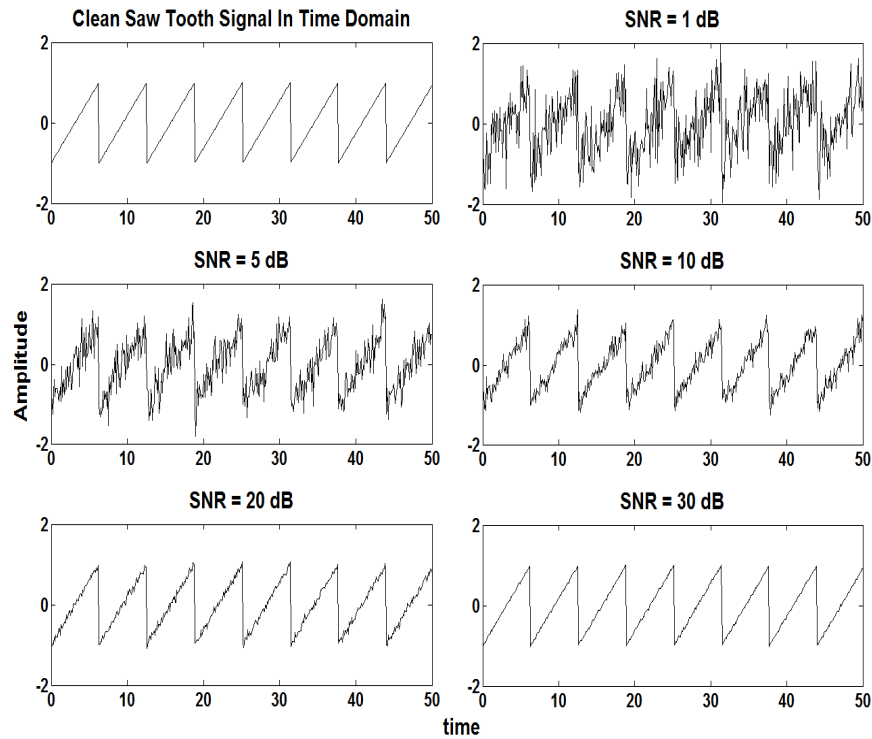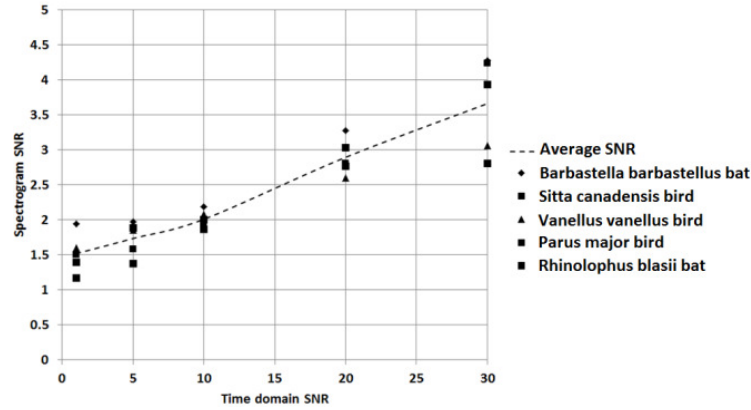
Figure 10. (top) A description for how the addition of different white Gaussian noise changes the structure and SNR of the time domain signal. (bottom) The corresponding changes in the spectrogram SNR averaged for the five applications.

3.1. Bioacoustics call of Rhinolophus blasii bat

The echolocation call of the Rhinolophus blasii bat which was investigated in this section includes medium duration strong sound pulses separated by short intervals. These pulses cover short Frequency Modulated (FM) band around 5 KHz, roughly estimated as Constant Frequency (CF) band, which were slowed down by a time expansion factor of 10 to be in the audible range, as shown in figure (11a).BP approach was able to remove most of the added noise, by rejecting the spectrogram values outside the small band surrounds the frequency of 5 KHz, as graphically shown in figure (11b) and numerically in the second column(s) of Table 1. Instead, the MBSS and WF approaches were not successful to remove reasonable amount of noise, especially for input SNRs less than 20 dB. The spectrograms generated with MBSS approach tended to temporally spread out the sound pulses, while those generated by WF approach tended to temporally cut from the duration of the pulses, as displayed in figure (11c, 11d) and the third and fourth column(s) of Table 1; respectively.

Figure 11. (a) One of Rhinolophus Blasii bat echolocation calls at SNR = 1 dB (with expansion factor of 10), and its enhanced spectrogram by (b) BP, (c) MBSS, (d) WF, (e) WPD, and (f) PM.

10

For higher values of input SNRs, the WF presented better enhancement. The WPD approach removed reasonable amount of noise and presented good estimation for the time domain resolution of the sound pulses. However, the five level decompositions produced repeated patterns along the frequency axis, as shown in figure (11e) and the fifth column(s) of Table 1. The PM dealt with the noisy spectrogram as an image, and was able to preserve the sound pulses while removing most of the attached noise, as expressed in figure (11f) and the sixth column(s) of Table 1.

Table 1. The subjective and objective measures for the spectrogram of Rhinolophus Blasii call enhanced by BP, MBSS, WF, WPD, and PM

(A) MOS

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1 (1,1)	2.8 (2.5,3)	1 (1,1)	1.1 (0.9,1.3)	1.7 (1.4,2)	1.6 (1.3,1.9)
5	1.1 (0.9,1.3)	2.9 (2.7,3.1)	1 (1,1)	1.2 (0.9,1.5)	1.8 (1.5,2)	2 (1.6, 2.4)
10	1.3 (1,1.6)	2.8 (2.5,3)	1.1 (0.9,1.3)	1.3 (1,1.6)	1.8 (1.4,2.2)	2.6 (2.3,2.9)
20	1.3 (0.9,1.7)	3.1 (2.9,3.3)	1.3 (0.9,1.7)	1.8 (1.3,2.3)	1.9 (1.5,2.2)	4.9 (4.7,5.1)
30	1.6 (1.3,1.9)	3.3 (3,3.6)	1.5 (1.1,1.9)	2.5 (2.2,2.8)	2 (1.7,2.3)	5 (5,5)

(B) SNR

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	2.238	15.399	2.1390	2.565	7.734	6.903
5	2.579	15.711	2.505	3.227	8.695	8.936
10	3.097	16.207	3.017	3.983	8.912	13.594
20	4.344	17.322	4.200	7.786	9.319	31.798
30	5.982	18.700	5.696	13.142	10.113	31.930

(C) SSNR

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1.755	3.278	1.744	1.793	2.472	2.292
5	1.794	3.312	1.786	1.869	2.667	2.526
10	1.854	3.369	1.845	1.956	2.765	3.063
20	1.998	3.495	1.981	2.395	2.859	6.470
30	2.187	3.651	2.155	3.057	3.004	5.187

(D) LSD

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1.260	0.537	1.834	2.125	1.011	0.767
5	1.002	0.352	1.826	2.020	0.655	0.620
10	1.050	0.501	1.708	1.743	0.520	0.560
20	0.697	0.312	1.500	1.267	0.218	0.311
30	0.528	0.279	1.256	1.044	0.144	0.279

(E) IS

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1.281	0.186	3.455	5.296	0.833	0.392
5	0.731	0.076	3.410	4.559	0.334	0.242
10	0.815	0.155	2.828	2.997	0.216	0.191
20	0.313	0.056	1.994	1.294	0.049	0.054
30	0.169	0.043	1.266	0.803	0.023	0.043

(F) AE

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	3.51	3.51	4.21	2.75	5.16	3.52
5	3.51	3.51	4.21	2.79	5.15	3.52
10	3.51	3.51	3.97	3.06	4.99	3.52
20	3.51	3.51	3.84	3.11	4.64	3.52
30	3.51	3.51	3.75	3.14	4.32	3.52

11

3.2. Bioacoustics call of Barbastellabarbastellus bat

As an alternative, the echolocation call of the Barbastella barbastellus bat includes very short duration sound pulses separated by non-fixed intervals. The pulses cover long Frequency Modulated (FM) band over the region [25-100] KHz, which were also slowed down by a time expansion factor of 10 to be in the audible range, as demonstrated in figure (12a). Since the frequency band covered by the sound pulses is large, BP approach was not able to remove most of the noise into this band, as graphically shown in figure (12b) and numerically in the second column(s) of Table 2. Meanwhile, the pulses were temporally spread out by MBSS approach; however, it removed reasonable amount of noise, as displayed in figure (12c) and the third column(s) of Table 2. The WF approach removed higher amount of noise except those exist in the beginning of the signal. There are also spectral cut from the pulses along the frequency axis and the temporal resolution is degraded, as given in figure (12d) and the fourth column(s) of Table 2. Although the spectral resolution of the spectrograms generated by the WPD approach still corrupt and low frequency noise was not removed, there were no possibility for the repetition of the pulses along the frequency axis because the covered FM band is considerably high, as clarified in figure (12e) and the fifth column(s) of Table 2. The PM has the same performance as in application 1, by keeping the temporal and spectral properties of the pulses while removed most of the attached noise, as shown in figure (12f) and the sixth column(s) of Table 2.
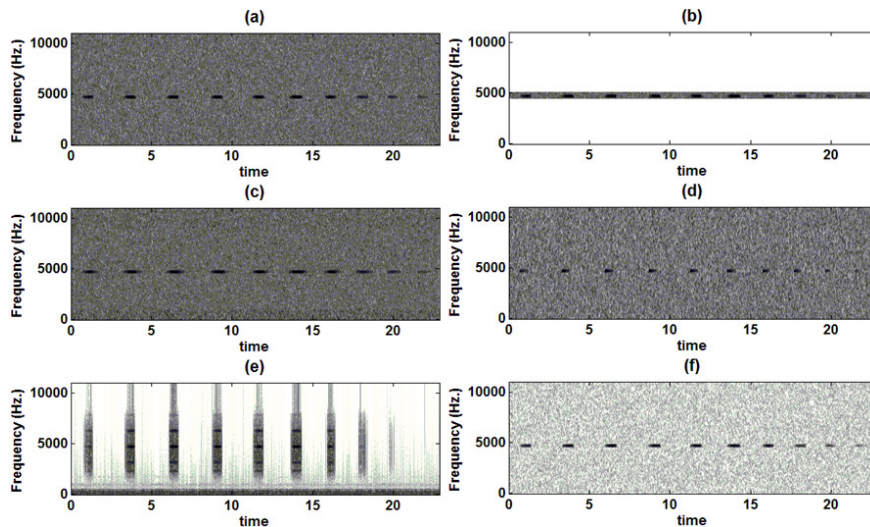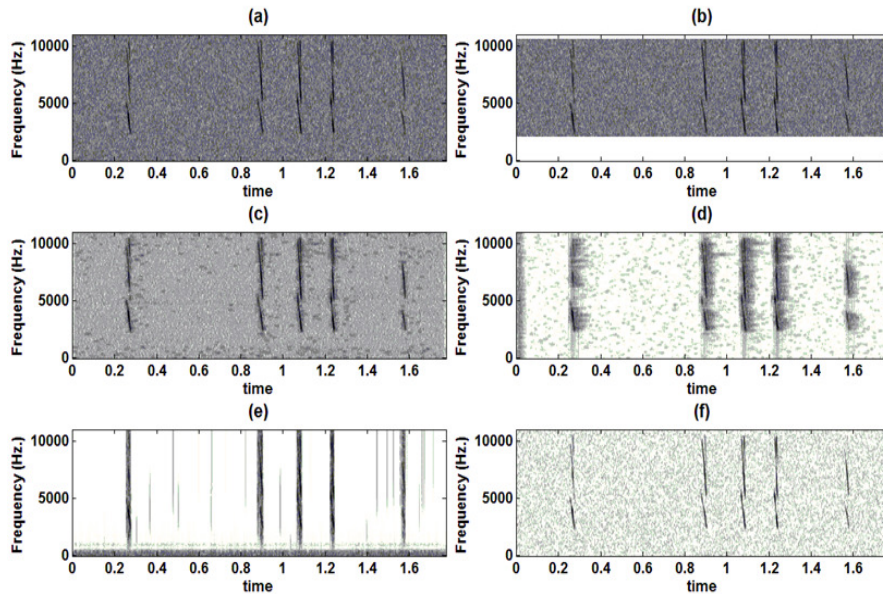

Figure 12. (a) One of Barbastella Barbastellus bat echolocation calls at SNR = 5 dB (with expansion factor of 10), and its enhanced spectrogram by (b) BP, (c) MBSS, (d) WF, (e) WPD, and (f) PM.

Table 2. The subjective and objective measures for the spectrogram of Barbastella barbastellus call enhanced by BP, MBSS, WF, WPD, and PM

(A) MOS

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1 (1,1)	1.2 (0.9,1.5)	1.3 (1,1.6)	2.1 (1.9,2.3)	2.7 (2.4,3)	2 (1.7,2.3)
5	1.1 (0.9,1.3)	1.3 (1,1.6)	1.4 (1.1,1.7)	2.4 (2.1,2.7)	2.8 (2.5,3.1)	2.4 (2.1,2.7)
10	1.3 (0.9,1.7)	1.4 (1,1.8)	1.5 (1.2,1.8)	2.6 (2.3,2.9)	2.9 (2.7,3.1)	3.4 (3.1,3.7)
20	1.4 (1,1.7)	1.6 (1.3,1.9)	1.9 (1.7,2.1)	2.8 (2.4,3.2)	3.1 (2.9,3.3)	5 (5,5)
30	1.7 (1.4,2)	1.9 (1.7,2.1)	2.3 (2,2.6)	2.9 (2.7,3.1)	3.2 (2.9,3.5)	5 (5,5)

(B) SNR

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	2.303	3.379	3.757	7.918	10.939	7.140
5	2.662	3.760	4.307	9.272	11.595	9.500
10	3.181	4.273	5.023	10.456	12.149	14.289
20	4.393	5.454	6.937	11.691	13.089	22.826
30	5.745	6.732	9.236	11.917	13.182	22.872

(C) SSNR

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1.763	1.886	1.931	2.510	2.843	2.320
5	1.804	1.930	1.995	2.782	2.961	2.592
10	1.864	1.989	2.078	3.182	3.077	3.158
20	2.004	2.126	2.312	5.017	3.376	6.814
30	2.163	2.278	2.659	4.233	3.628	3.386

(D) LSD

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	0.942	1.015	0.600	0.194	0.278	0.610
5	0.943	0.888	0.466	0.140	0.191	0.485
10	0.859	0.715	0.417	0.209	0.171	0.360
20	0.571	0.535	0.244	0.201	0.124	0.121
30	0.407	0.367	0.135	0.392	0.173	0.195

(E) IS

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	0.630	0.777	0.228	0.036	0.059	0.234
5	0.631	0.564	0.133	0.021	0.032	0.141
10	0.506	0.341	0.107	0.025	0.020	0.074
20	0.203	0.179	0.042	0.024	0.015	0.008
30	0.100	0.082	0.022	0.071	0.018	0.021

(F) AE

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	57	57	56.6	6	16.27	57
5	57	57	56.8	9.31	19	57
10	57	57	56.8	14.54	34.52	57
20	57	57	56.85	21.13	47.33	57
30	57	57	56.85	30.46	54.18	57

3.3. Bioacoustics call of Vanellus vanellus bird

As an example for a multi harmonic sound stream in the human hearing range, the bioacoustics call of Vanellus vanellus bird was investigated in the region bounded by 6 KHz. The sound stream contains three FM long pulses with dominant frequencies around (1, 2.2, and 4) KHz; respectively, followed by three downstream CF short pulses of fundamental frequencies around (2, 3, and 4) KHz; respectively, as given in figure (13a). Enhancement by the BP approach did not produce clear spectrogram, since the pulses cover much of the frequency axis, given high

13

constraint to the rejected band by this approach, as shown in figure (13b) and the second column(s) of Table 3. The enhanced spectrogram generated by the MBSS approach has reasonable temporal resolution and degraded spectral resolution of the sound pulses, especially for the downstream pulses which corrupted by high spectral distortion, as displayed in figure (13c) and the third column(s) of Table 3.On the other hand, the temporal and spectral resolutions of the spectrogram generated by WF approach are acceptable, although there is little spectral leakage for the downstream pulses and initial sound noise (< 0.03 sec) were not removed, as plotted in figure (13d) and the fourth column(s) of Table 3.Whereas the low frequency noise was not enhanced by the WPD approach, many temporal bands of noises were removed. The decomposition of the sound stream presented spectral mirrors of the weak harmonic patterns and almost eliminated the downstream pulses, as demonstrated in figure (13e) and the fifth column(s) of Table 3. Meanwhile, the enhanced spectrogram by PMre established high temporal and spectral resolutions of the sound pulses, as indicated by the obtained LSD, with high distinction from the attached noise, as designated by the obtained SNR and shown in figure (13f) and the sixth column(s) of Table 3.
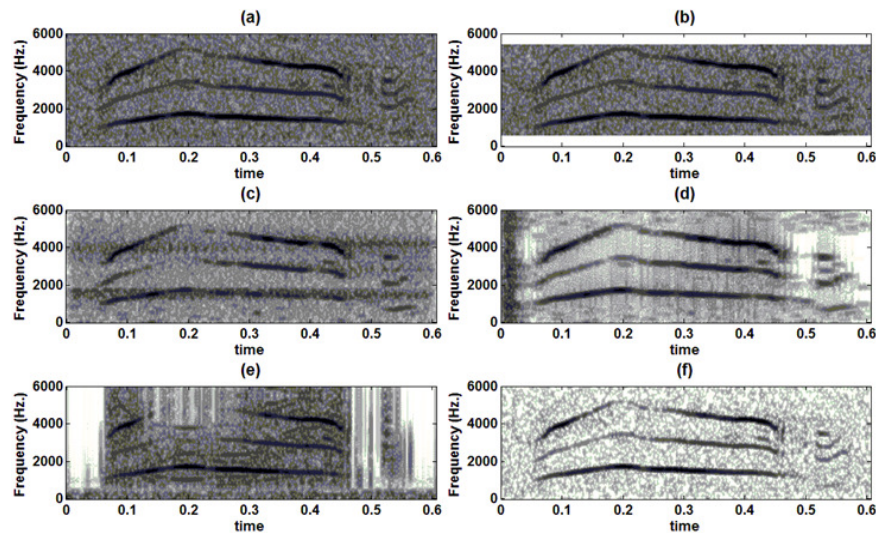
Figure 13. (a) One of Vanellus vanellus bird calls at SNR = 10 dB, and its enhanced spectrogram by (b) BP, (c) MBSS, (d) WF, (e) WPD, and (f) PM.

Table 3. The subjective and objective measures for the spectrogram of Vanellus vanellus bird call enhanced by BP, MBSS, WF, WPD, and PM

(A) MOS

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1 (1,1)	1.5 (1.2,1.8)	1.4 (1.1,1.7)	2 (1.7,2.3)	1.7 (1.4,2)	2.4 (2.1,2.7)
5	1.4 (1.1,1.7)	1.6 (1.3,1.9)	1.5 (1.2,1.8)	2.2 (1.9,2.5)	1.8 (1.4,2.2)	2.9 (2.7,3.1)
10	1.3 (1,1.6)	1.8 (1.4,2.2)	1.7 (1.3, 2.1)	2.4 (2.1,2.7)	1.9 (1.7,2.1)	3.7 (3.4,4)
20	1.5 (1.2,1.8)	1.9 (1.5,2.3)	1.8 (1.4,2.2)	2.6 (2.3,2.9)	1.9 (1.7,2.1)	4.7 (4.4,5)
30	1.6 (1.3,1.9)	2 (1.6,2.4)	1.8 (1.4,2.2)	2.7 (2.4,3)	1.9 (1.7,2.1)	5 (5,5)

(B) SNR

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1.651	2.635	2.489	3.591	2.895	4.371
5	1.941	2.843	2.689	4.046	3.180	5.414
10	2.207	3.153	2.912	4.452	3.317	6.807
20	2.662	3.532	3.108	4.771	3.422	9.078
30	2.814	3.605	3.163	4.884	3.339	9.480

14

(C) SSNR

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1.688	1.801	1.784	1.925	1.896	2.001
5	1.721	1.825	1.807	1.995	1.989	2.121
10	1.752	1.861	1.833	2.059	2.036	2.283
20	1.804	1.904	1.856	2.119	2.120	2.564
30	1.822	1.913	1.863	2.162	1.991	2.621

(D) LSD

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	1.172	1.083	0.982	1.220	1.009	0.790
5	1.042	0.938	0.876	1.141	0.907	0.505
10	0.988	0.888	0.537	1.028	0.783	0.537
20	0.814	0.783	0.753	1.012	0.587	0.267
30	0.718	0.714	0.515	0.968	0.579	0.267

(E) IS

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	0.630	0.922	0.703	1.221	0.856	0.423
5	0.803	0.649	0.538	1.031	0.657	0.156
10	0.707	0.568	0.182	0.804	0.469	0.177
20	0.451	0.425	0.385	0.771	0.248	0.041
30	0.341	0.345	0.169	0.696	0.233	0.041

(F) AE

input SNR	noisy	BP	MBSS	WF	WPD	PM
1	8.47	8.47	14.63	10.32	3.13	8.51
5	8.47	8.47	12.51	10.3	3.82	8.51
10	8.47	8.47	9.78	9.46	4.17	8.51
20	8.47	8.47	9.12	8.72	7.12	8.51
30	8.47	8.47	8.65	8.19	7.44	8.51

From the pattern recognition point of view, different AE values to those of the original spectrogram, reflect changing in the shape of some or all of the sound patterns, which result in non-accuracy in further pattern recognition results. Unlike, similar AE values do not ensure the shape of the sound pattern is similar to its original shape, but it may changes in a way that its aspect ratio is constant.

4. Extended applications

The spectrograms obtained by the proposed method (PM) displayed how it is powerful and consistent to enhance different structure bioacoustics calls. Therefore, these enhanced spectrograms can be implemented in various post processing tasks. In this section, the three most important implementations of the generated spectrogram by PM will be explored.

4.1. Power and frequency contours

The three variables of the enhanced spectrogram (i.e., time, frequency, and power spectrum) may be plotted in different orders to obtain its power and/or frequency contours, as visible for the bioacoustics call of Rhinolophus hipposideros bat (with expansion factor of 13) in figure (14), after its spectrogram was enhanced by PM. The power contours provides an image of the instantaneous power contents of the sound patterns and can be used for specific sound power detection after calculating the areas enclosed by these contours. While the frequency contours provide an image of the instantaneous frequency contents of the recorded bioacoustics call and can be used for designing more reliable frequency filters.

15
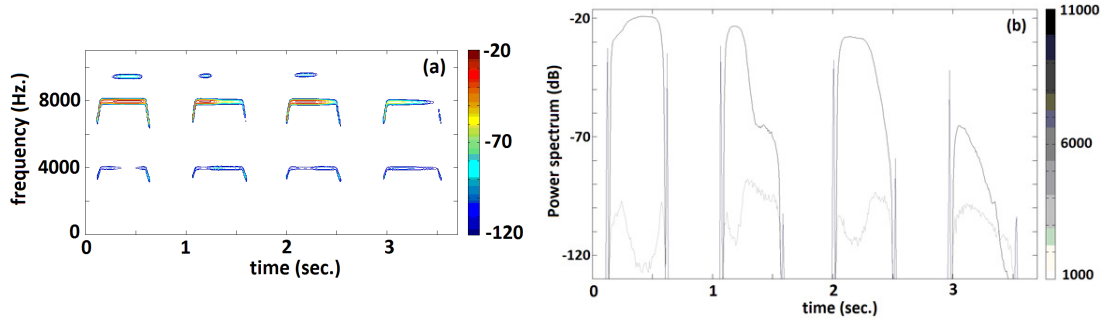
Figure 14. The enhanced (a) power and (b) frequency contours for Rhinolophus hipposideros bat call

4.2. The enhanced wave form of the bioacoustics call

By transforming the enhanced spectrogram variables back to the time domain by Inverse Fast Fourier Transform (IFFT), the enhanced wave form of the call is obtained, as given in figure (15).The phase information obtained through the former Short Time Fourier Transform (STFT) is used to reconstruct the enhanced wave form, following the flowchart of figure (15). The output waveform can be used for reliable extraction of the bioacoustics temporal features suchlike zero crossing rate, short time energy, temporal roll-off, and temporal spread of the sound patterns [1].
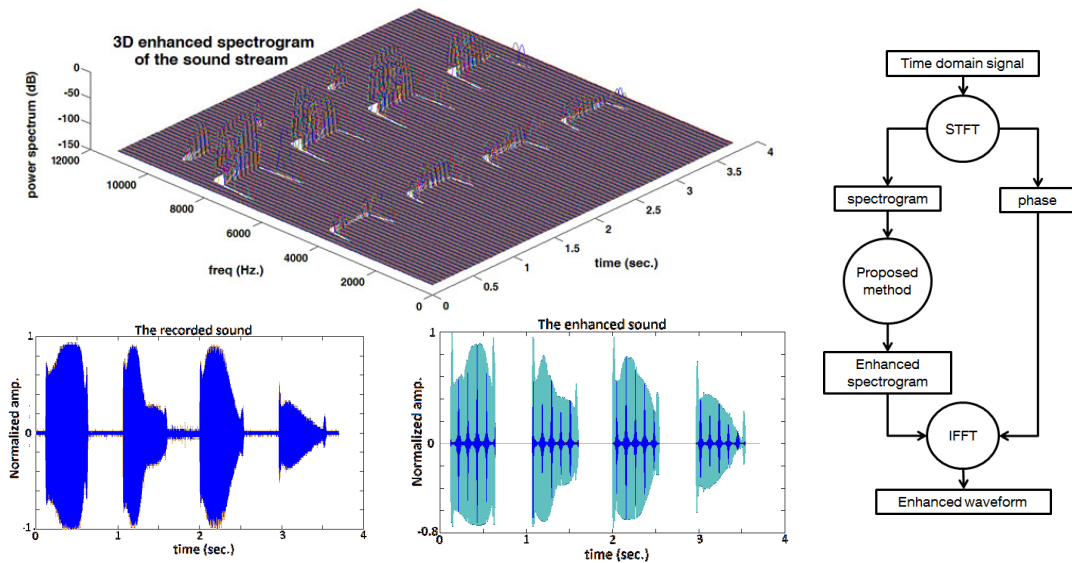
Figure 15. (upper) The 3D enhanced spectrogram for the bioacoustics call of Rhinolophus hipposideros bat(with expansion factor of 13)and flow chart to reconstruct the waveform of its sound stream. (bottom) the original and the reconstructed wave form of the sound stream; respectively.

4.3. Bioacoustics calls classification

From the enhanced spectrogram, simple and reduced number of 1D featuresand/or 2D features can be extracted for complete pattern recognition of the bioacoustics sound. The 1D features are the signal features suchlike the covered frequency band(s), peak frequency, pulse duration, interval between sound pulses, etc. while the 2D features are the image features suchlike eccentricity and centroid. As a test case, a classifier of Vanellus vanellus, Parus major, and Sitta

16

Canadensis birds has been constructed using the eccentricity (a/b) and the vertical coordinate of the centroid (c) features, as shown in figure (16(ii)), extracted from the enhanced spectrogram of each bird sound, as displayed in figure (16(i)). The classifier was trained by 26, 30, and 37 sound patterns of the three birds; respectively, to define the rough dividing contours, given in figure (16(iii)). Afterwards, the classifier was tested by another 7, 8, and 10 sound patterns of the three birds; respectively, beside 7 patterns of Barbastella barbastellus bat and 5 patterns of Rhinolophus Blasii bat, giving 94.59% classification accuracy (two sound patterns of Sitta Canadensis wrongly detected as Vanellus vanellus), as plotted in figure (16(iii)). It may be realized that even simple classifier can separate out the sound patterns into the correct bioacoustics source, providing that distinctive features were selected and sufficient training patterns were used.

(i)

(ii)

(iii)
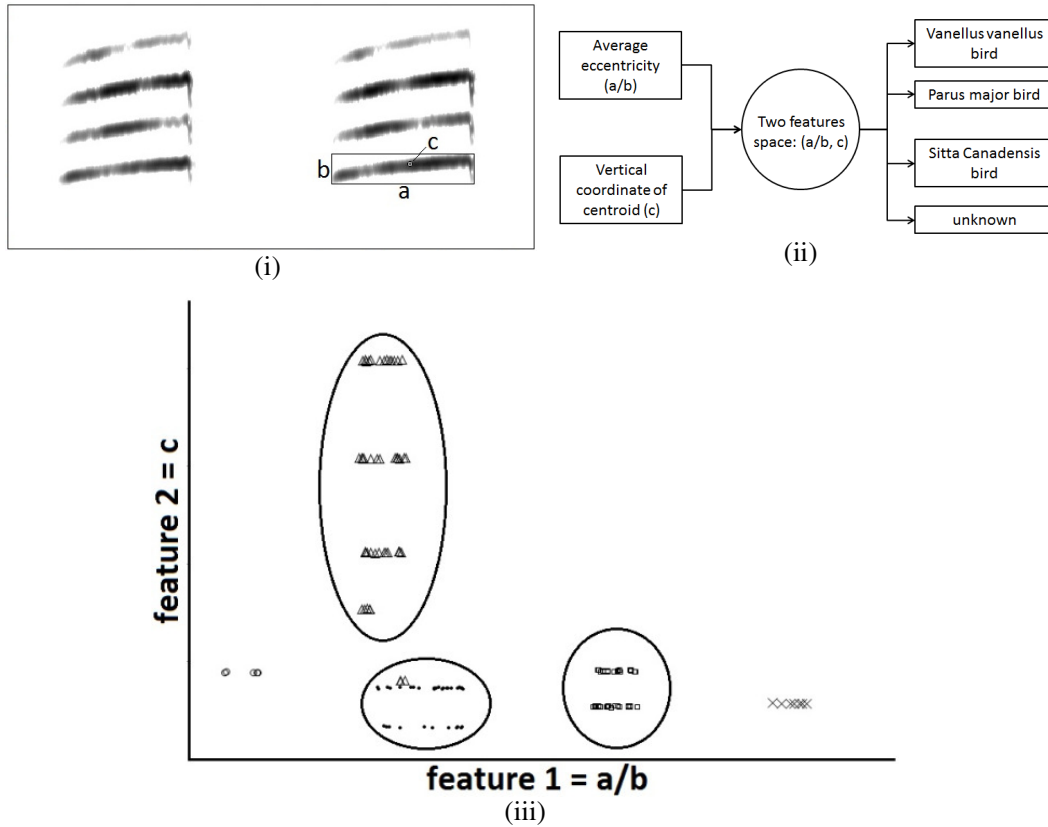
Figure 16. (i) Enhanced spectrogram for one of Sitta Canadensis bird contains the strongestsound patterns, indicating how the eccentricity (a/b) and the vertical coordinate of the centroid (c) are extracted for one of its sound patterns. The units of a and b are in pixeld, and c in pixel number. (ii) Simple classifier structure with input of the two features, which form the classification space, and the outputs are four classes for Vanellus vanellus, Parus major, Sitta Canadensis , and unknown sounds. (iii) The classification space with three dividing contours encounters the features of the three birds, respectively, at which the surrounded region is for unknown sound features. The classification results also included with two sound patterns of Sitta Canadensis wrongly detected as Vanellus vanellus.

5. Conclusion

Spectrogram reading provides a direct method for hands-on learning of the characteristics of bioacoustics calls, therefore, a variety of enhancement techniques have been considered over the past years to remove the attached noises. In this paper, a spectrogram enhancement method was developed based on high accurate edge detection of the enclosed sound patterns and removing the

17

surrounding noise. The crest factor was presented as a smoothed version of the spectrogram image, avoiding the threshold problem of usual smoothing filters, suchlike Gaussian filter in Canny edge detector. The proposed method was applied to enhance the limited dynamic range spectrogram of different structure bioacoustics calls, in comparison to the four commonly used enhancement approaches, which are band pass filter (BP), multi-band spectral subtraction (MBSS), Wiener filter (WF), and wavelet packet decomposition (WPD) approaches. The comparison was established on one subjective measure [mean opinion score] and four objective measures [signal to noise ratio, segmental signal to noise ratio, log spectral distance, and Itakura Saito]of the spectrograms obtained by the five methods at different SNR. The results showed that the shorter the frequency band of FM pulse, the better the enhancement with BP and WPD. The larger the upstream interval before the first pulse, the better the enhancement with WF. The longer the CF pulse at high SNR, the better the enhancement with MBSS because it tends to spread the patterns over time. Meanwhile, the proposed method produced highly efficient enhanced spectrograms for all of the investigated calls.

The temporal and spectral resolutions of the spectrograms produced by the BP approach are of high accuracy, since it does not operate a post processing to the full range of the noisy spectrogram, but only rejects the band which estimated to be noise. This was not the case with (MBSS, WF, and WPD) which post process the spectrograms for enhancement, results in changing the temporal and/or spectral resolutions. In the meantime, the edge detection algorithm of the proposed method was able to preserve the sound pulses into their almost original temporal and spectral locations while processing the noisy spectrogram. This is very important issue for any further pattern recognition assignment based on the enhanced spectrogram.

As a future aspect to this research work, an investigation will be made to avoid the loss of weak patterns done through limiting the dynamic range of the spectrogram. Moreover, improving the original spectrogram generation by adapting the applied STFT settings, this in correspondence improves the enhanced spectrogram.

ACKNOWLEDGMENTS

REFERENCES

[1] Hussein, W. B., Hussein, M.A.& Becker, T., (2009) "Application of audio signal processing in the detection of the Red Palm Weevil", *Proceeding of European Signal Processing Conference EUSIPCO2009*, Glasgow, Scottland, 24-28.09.2009, pp 1597-1601.

[2] Obrist, M. K., Boesch, R. & Flückiger, P. F.,(2004) "Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergetic pattern recognition approach". *Mammalia*, Vol.68, No. 4, pp 307-322.

[3] O'shaughnessy, D., (1987) "*Speech Communication: Human and Machine*", Addison-Wesley Publishing Co., Reading, MA, 1987, pp 56 - 106.

[4] Mallawaarachchi, A., Ong, S. H., Chitre, M., &Taylor, E, (2008) "Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles". *J. Acoust. Soc. Am.*, Vol. 124, No. 2, pp 1-8.

[5] Oppenheim, Alan V., Schafer, Ronald W.& Buck, John A. (1999). "*Discrete-time signal processing*". Upper Saddle River, N.J., Prentice Hall, pp 468–471.

[6] Li, H., Zhang, Y. &Xu, D., (2010) "Noise and Speckle Reduction in Doppler Blood Flow Spectrograms Using an Adaptive Pulse-Coupled Neural Network".*EURASIP Journal on Advances in Signal Processing*, pp 1-11.

18

[7] Ghanbari, Y., Karami, M. R.&Amelifard,B., (2004) "Improved multi-band spectral subtraction method for speech enhancement", *Proceedings of the 6th ISTED International conference SIGNAL AND IMAGE PROCESSING*, pp225-230.

[8] Lampert, T. A. & O'keefe, S.E.M, (2010) "A survey of spectrogram track detection algorithms". *Applied acoustics*, Vol. 71, pp 87- 100.

[9] Mellinger, D. K., (2002). *Ishmael 1.0 User's Guide*, Pacific Marine Enviromental Laboratory, Seattle, USA.

[10] Liu, W. A., Miller, R.C., Merzenich, K.D., &Schreiner, C. E., (2003)"Acoustic variability and distinguishability among mouse ultrasound vocalizations," *J. Acoust. Soc. Am.*, Vol. 114, pp 3412–3422.

[11] Ding, H., Soon, I.Y., Koh, S.N.&Yeo, C.K., (2009) "A spectral filtering method based on hybrid Wiener filters for speech enhancement". *Speech Communication*, Vol. 51, pp 259–267.

[12] Gur, B. M., & Niezrecki, C., (2007). "Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform," *J. Acoust. Soc. Am.*, Vol. 122, pp 188–199.

[13] Ghanbari, Y.&Karami, M.R., (2006)"A new approach for speech enhancement based on adaptive thresholding of wavelet packets' *Speech communication*, Vol. 48, pp 927-940.

[14] Chavan, M.S., Chavan, M.N., &Gaikwad, M.S., (2010) "Studies on Implementation of Wavelet for Denoising Speech Signal". *International Journal of Computer Applications*, Vol. 3, No.2,pp. 1-7

[15] Pandey, P.C., Pratapwar, S.S., &Lehana, P.K., (2004) "Enhancement of Electrolaryngeal Speech by Reducing Leakage Noise Using Spectral Subtraction with Quantile Based Dynamic Estimation of Noise ".*Proceeding of the 18th international congress on acoustics ICA2004*, pp. 3029-3032

[16] Sumithra, A., &Thanushkodi, B., (2009) "Performance Evaluation of Different Thresholding Methods in Time AdaptiveWavelet Based Speech Enhancement". *IACSIT International Journal of Engineering and Technology*,Vol.1,No.5,pp 42-51.

[17] Kamath, S.& Loizou, P., (2002) "A Multi-band spectral subtraction method for Enhancing speech corrupted by colored noise", *proceedings of ICASSP-2002*, pp 37-43.

[18] Hermansky, H., Wan, E. A.& Avendano, Carlos, (1994) "Noise Suppression in Cellular Communications", *in Proceedings IEEE IVTTA'94*, pp 85-88.

[19] Kim, H.G., Obermayer, K., Bode, M.&Ruwisch, D., (2000) "real time noise cancelling based on spectral minimum detection and diffusive gain factors", *in proceeding of 8th Aust. Int. Conf. Speech Sci. and Tech.*, pp 250-255.

[20] Tuthill, T.A., Sperry, R.H. & Parker, K.J., (1988) "Deviations from Rayleigh statistics in ultrasonic speckle," *Ultrasonic Imaging*, Vol. 10, No. 2, pp 81–89.

[21] Evans, N.D., Mason, J.S. & Roach, M.J., (2002) "Noise Compensation using Spectrogram Morphological Filtering". *In Proceeding of 4th IASTED International Conference Signal Image Processes*, pp 157-161.

[22] Steinberg, R. &O'Shaughnessy,D., (2008) "Segmentation of a Speech Spectrogram using Mathematical Morphology", *in proceeding Acoustics, Speech and Signal Processing, ICASSP 2008*, pp 1637 – 1640.

[23] Gillespie D., (2004) "Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram". *Can Acoust 2004*, Vol. 32, pp39–47.

[24] Lin, B.S., Wu, H.D., Chong, F.C.&Chen, S.J., (2006) "Wheeze Recognition Based On 2d Bilateral Filtering Of Spectrogram". *Biomed Eng Appl Basis Comm*, Vol. 18, pp 128-137.

[25] Dugnol, B., Fernández, C., Galiano, G. &Velasco, J., (2007) "Wolves chorus noise reduction by spectrogram image processing", *Signal Processing for Image Enhancement and Multimedia Processing*, Springer Verlag.

[26] Ohtake, Y., Belyaev, A.G.&Seidel, H.P., (2002) "Mesh smoothing by adaptive and anisotropic Gaussian filter". *Proceeding of Vision, Modelling, and visualization 2002*, pp 203-210.

19

[27] Hussein, W. B., Hussein, M.A.,&Becker, T., (2010) "Detection of the Red Palm Weevil Using Its Bioacoustics Features." *Journal of Bioacoustics*, Vol. 19, No. 3, pp 177-194.

[28] Hussein, W. B., Moaty, A. A, Hussein, M. A.& Becker, T., (2011) "A novel edge detection method with application to the fat content prediction in marbled meat". *Pattern recognition*, Vol. 44, No. 12, pp 2959 – 2970.

[29] Eargle, J., (2005). "*Handbook of recording engineering*", 4th edition, Springer, pp 1-27.

[30] Deller, J.R., Hansen, J.H.L.& Proakis, J.G., (2000) "*Discrete-Time Processing of Speech Signals*", second ed., IEEE Press, New York.

[31] Quackenbush, S.R., Barnwell, T.B.& Clements, M.A, (1988) "*Objective Measures of Speech Quality*". Englewood Cliffs, NJ, Prentice-Hall.

20

3.4 Online monitoring of yeast fermentation bioprocess

The medium carries the fermentation process is very dynamic with mixture being formed in different phases of fermentation. The specific product resulting from fermentation is determined by the catalyst of the process and the substrate in which the fermentation occurs. The catalysts, such as bacteria and yeast, vary considerably in their metabolic characteristics. For example, bread and beer are products of yeast fermentation in grain and antibiotics are the products of bacterial fermentation.

During yeast fermentation process, yeast cells use oxygen (O_2) and break down sugar ($C_6H_{12}O_6$) to carbon dioxide (CO_2), water (H_2O), and energy (e), as given in the reaction of equation (3.1). In the absence of oxygen, yeast switches to another pathway and produces ethanol (C_2H_5OH) and carbon dioxide, as shown in the reaction of equation (3.2).

$$C_6H_{12}O_6 + 6\,O_2 \xrightarrow{yeast} 6\,CO_2 + 6\,H_2O + e \tag{3.1}$$

$$C_6H_{12}O_6 \xrightarrow{yeast} 2\,C_2H_5OH + 2\,CO_2 \tag{3.2}$$
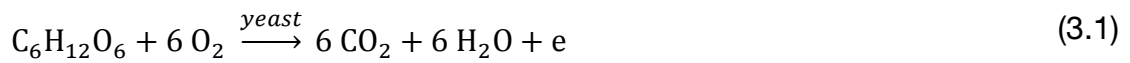
Reference density is measured by an oscillating U-tube technique, based on a mass-spring model and an electronic measurement of the frequency of oscillation [77]. A sample is filled into a U-shaped container with oscillation capacity, at which its center frequency is influenced by the sample's mass, results in electronic excitation of undamped oscillation. Since the volume involved in this tube is limited, the model uses the measured sample's mass to calculate the sample density as function of this oscillation, as explained in equation (3.3).

$$\rho_{ref} = A\tau^2 + B \tag{3.3}$$

Where A and B are the oscillating U-tube constants, f is the oscillation period and ρ_{ref} is the measured reference density of the sample.

In this work, ultrasound signals are excited into a fermentation mixture along the whole process time, where the speed of sound and other signal features are continuously extracted by the developed mathematical tools. These features are combined with the mixture temperature to design an artificial neural network which predicts the mixture density. The network is trained by these features as inputs and the offline measured reference density as an output. The network is employed online in fermentation processes to monitor mixture density in noncontact, non-invasive, and non-destructive approach. The monitored density determines the fermentation level, suchlike how much alcohol is in the beer and the status of gas retention in dough.

As fermentation proceeds, the vaporization of CO_2 may interfere with accurate density measurement. Furthermore, measures such as mixture pressure have to be controlled to prevent or reduce bubble formation. This is done usually by installing a valve in the piping from the fermentation tank outlet.

Walid B. Hussein

Mohamed A. Hussein

Thomas Becker

Group of (Bio-)Process
Technology and Process Analysis
Faculty of Life Science
Engineering, Technische
Universität München, Freising,
Germany

Supporting information
available online

Research Article

Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation processes

Ultrasound techniques are well suited to provide real-time characterization of bioprocesses in non-invasive, non-contact, and non-destructive low-power consumption measurements. In this paper, a spectral analysis method was proposed to estimate time of flight (TOF) between the propagated echoes, and its corresponding speed of sound (USV). Instantaneous power spectrum distribution was used for accurate detection of echo start times, and phase shift distribution for correcting the involved phase shifts. The method was validated by reference USV for pure water at 9–30.8°C, presenting a maximum error of 0.22%, which is less than that produced by the crosscorrelation method. Sensitivity analyses indicated a precision of 6.4×10^{-3}% over 50 repeated experiments, and 0.11% over two different configurations. The method was competently implemented online in a yeast fermentation process, and the calculated USV was combined with temperature and nine signal features in an artificial neural network. The network was designed by back propagation algorithm to estimate the instantaneous density of the fermentation mixture, producing a maximum error of 0.95%.

1 Introduction

Ultrasound sensors are powerful tools that effectively address the problem of performing non-contact and non-invasive determination of process variables, from the viewpoint of performance and cost [1]. Ultrasound has advantages over other traditional analytical techniques because measurements are rapid, non-destructive, precise, fully automated, and might be performed either in laboratory or online [2]. The ultrasound sensor produces a pulse or burst signal to measure the required ultrasound characteristics with low power consumption.

Therefore, ultrasound sensor systems open in-line applications in many processes for many substances such as product characterization in the food, chemical, pharmaceutical and petrol industries, control of sewage treatment, or polymerization processes, and monitoring of chemical etching [3]. Moreover, typical industrial applications of ultrasound for concentration measurement and process monitoring include those of

chemical and pharmaceutical industry (polymerization, paints, and waste water treatment), food industry (beverage, dairy, and starch production), and biotechnology (fermentation process, enzyme concentration) [4]. One of the widespread applications is the utilization of ultrasound for concentration measurement during yeast fermentation process. The possibility of using low-intensity ultrasound to characterize such food processes was first realized over 60 years ago [5]; however, it is only recently that the full potential of the technique has been realized. There are a number of reasons for the current interest in ultrasound monitoring during fermentation process. From one side, the food industry is becoming increasingly aware of the importance of developing new analytical techniques to study complex food materials, and to monitor properties of foods during processing where strict protocols, issued by the FDA, be maintained to ensure food purity; ultrasound techniques are ideally suited to both of these applications. And from the other side, ultrasound instrumentation can be fully automated, make rapid and precise measurements, and can easily be adapted for online applications [6]. Unlikely, the ultrasound technology has a few limitations such as its sensitivity to air bubbles during the process course, and dependency of the acoustic properties on the specified sample concentration.

Correspondence: Walid B. Hussein (whussein@wzw.tum.de), Group of (Bio-)Process Technology and Process Analysis, Faculty of Life Sciences Engineering, Wissenschaftszentrum Weihenstephan, Weihenstephaner Steig 20, 85354 Freising, Germany.

The two common approaches of ultrasound measurements are the continuous wave approach and the pulse-echo approach. In the continuous wave approach, two separate transmitting and receiving elements are used, which requires more complex hardware [7]. Meanwhile, the pulse-echo approach requires only one transducer that operates alternatively between transmitting and receiving modes. The latter approach offers a simple and low cost solution, even if it yields poorer results owing to the uncertainty in the time domain measurement [8]. The main ultrasound parameter for use in process monitoring and control is the speed of sound (USV). Once it is determined in a sample, the data can be used in a number of ways, such as identification of liquids, concentrations of solutions, behavior of mixtures of liquids, and two-phase liquid systems [3]. USV is calculated by dividing the signal path length over the estimated time of flight between the propagated ultrasound echoes, and this estimation is considered the critical point of the whole measurement. Therefore, various methods have been developed to improve the time of flight (TOF) estimation accuracy. At which some of them based on the proper design of the transmitting and/or receiving system (configuration approach) [9], on the generation of signals with good time localization (sensor approach) [10], or on the usage of sophisticated digital techniques (processing approach) [11–13]. The first two strategies either offer simple and low cost resolutions with very poor results, or produce accurate estimations by requiring complicated and expensive hardware. As a result, it is worth to examine the possibility of improving TOF estimation using digital signal processing techniques rather than hardware adaptation.

Two digital techniques are widely used to estimate TOF, the threshold method, and the cross correlation method. The first method detects the indices corresponding to the time instants when the signal amplitude crosses certain threshold, and TOF is then the interval between these two instances [14]. However, the received echoes reach the threshold level sometime after their exact start. The second method searches for the instant at the relative maximum in the correlation function between the first echo and the rest of the signal, and define TOF to be the interval between this instant and the start of first echo [15].

Although the results obtained in the previous studies lead to the practical conclusion that the cross correlation method ensures up to 40% increase in the accuracy of TOF estimation with respect to the threshold method, it requires a greater computational cost and is highly influenced by noise spikes in the signal. Furthermore, the overall accuracy depends on the measurement of phase shift between the selected echoes, which is hardly noticed in the time domain and requires a spectral analysis in the frequency domain.

The spectral representation of the ultrasound signal can be obtained by one of many common transformations algorithms including Hilbert transformation, short time Fourier transform, or wavelet transform. The former algorithm is very sensitive to noise and works powerfully only if the echoes are with mutual interference [16]. Meanwhile, the other two algorithms are less computational cost, efficient, and produce better noise rejection.

In this paper, a TOF estimation method was developed consisting of two parts; the first is the enhancing of the ultrasound signal to its exactly dominant frequency, through a high-resolution spectral analysis based on short time Fourier transform. Afterwards, start times of the first and second echoes were detected on the generated instantaneous power spectrum distribution. The second part is to apply a phase-shift correction to the detected times by investigating the instantaneous phase-shift distribution. The proposed method was validated by the standard reference data given for USV of demineralized water at elevated temperatures 9–30.8°C. Sensitivity analyses were also conducted to check the consistency and repeatability of the method results. Finally, it was efficiently applied to calculate USV during online yeast fermentation process, and combined with other signal features in a feed-forward artificial neural network to estimate the mixture density.

2   Materials and methods

2.1   Experimental setups

2.1.1   Hypothetical ideal pulse-echo setup
One of the commonly used arrangements in technical ultrasound applications is the one shown in Fig. 1A, at which ultrasound sensor is fixed on one side of the test section, to transmit suitable energy pulse of frequency, usually between 2 and 5 MHz. This pulse penetrates in the test section up to the reflector side, which reflects it back toward the sensor. Once the sensor produces the pulse, it works as a receiver to handle all the reflected and consequent echoes after their fully propagation into the contained medium of the test section.

2.1.2   Real pulse-echo setup
A hygienic and insulated pulse-echo setup that is widely used in food and beverage non-destructive tests was employed for this research, including a stainless steel support armature to carry the inline sensors (ultrasound and temperature sensors) and a heater. The transmission-reception parts contain buffer and reflector lids, each has 17-mm thickness, as displayed in Fig. 1B. Two different nominal diameters configurations were tested for the purpose of sensitivity analysis; the first is DN50 (50 mm) and the second is DN80 (80 mm). The implemented temperature sensor is PT100 (standard platinum resistance sensor) with accuracy of 0.1°C at 0°C. The ultrasound transducer is lead zirconate titanate ceramic with a resin-tungsten backing, generates a rectangular 250 ns (i.e. nano-sec) impulse of 20 V amplitude and 2 MHz dominant frequency. The system was placed vertically to avoid existence of internal air pockets.

2.1.3   Yeast fermentation setup
The bioreactor shown in Fig. 1C was employed to perform and monitor the aerobic yeast fermentation process under brewing relevant conditions. The sensors are measuring turbidity, density, and ultrasound signals and were mounted inline in the circulation pipe before the aeration jet. Moreover, the ultrasonic sensor set up was mounted using a VARINLINE® coupling to insure no dead-space measurements in accordance with high hygienic considerations. All sensor readings are recorded via a Beckhoff EtherCAT® control system to a personal
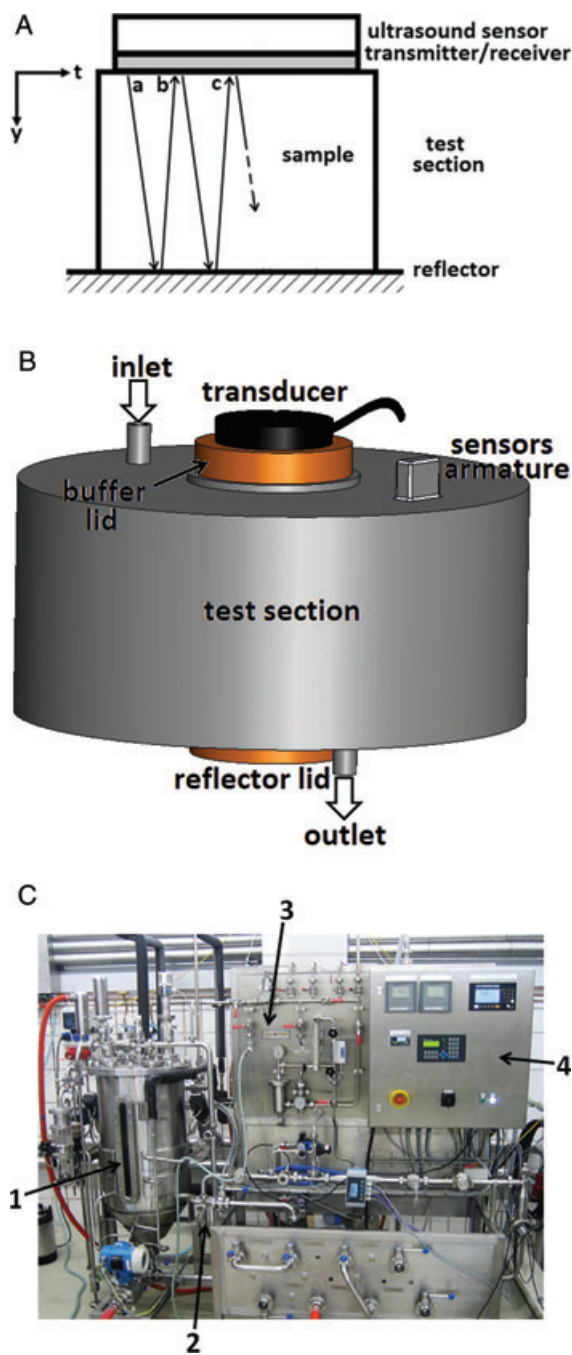
Figure 1. (A) Basic pulse-echo configuration for measuring TOF between consequent echoes. (t) is the time axis carries information about the arrival of echoes while (y) is the distance axis carries information about the path length covered by the signal. Points (A), (B), and (C) represent transmission time of main pulse and reception times for first and second echoes; respectively.
(B) Layout of the pulse echo ultrasound set up consists of transducer, buffer lid, sample test section of DN50 and DN80, reflector lid, as well as attached armature with temperature sensor and heater.
(C) The implemented yeast fermentation setup and its four main items of (1) cylindroconical vessel at which the fermentation process takes place, (2) circulation pipes, (3) gas flow panel, and (4) switch cabinet.

computer using the software package Virtual Expert®. Fermentation process takes place in the stainless steel cylindroconical vessel with a total volume of 129 L and a working volume up to 60 L of wort (∼50% head space for foam formation). The internal pressure was fixed at 0.9 bars to prevent foam and flow homogenization was achieved by a centrifugal pump. An aeration jet "Turbo Air"® (Co. Esau & Hueber) was introduced into the circulation to supply yeast with oxygen. The used substrate (wort) contained approximately 10 g/100 g sugar (main components are maltose, glucose, fructose, and maltotriose) supplied with hops.

2.2 The proposed method

The propagation of ultrasound signals in ideal pulse-echo system (Fig. 1A), follows the one given in Fig. 2A, at which the succeeded echoes can be simply detected and well defined, results in straightforward recognition of TOF and its corresponding USV. In the frequency domain, one frequency value is contained in the signal, which is the dominant frequency f_d, as illustrated by the power spectrum distribution in Fig. 2B. Additionally, although the main pulse and the received echoes are out of phase, which means they have a phase difference between $[0, \pi]$ radians, there is no phase shift among the succeeding echoes, as shown in the phase-shift distribution given in Fig. 2C.

The amplitude (x) of the transmitted ultrasound signal in a medium can be mathematically modelled as a function of its propagation time (t), as reported in [1] and given by equation (1).

$$x(t) = A_s t^m e^{-t/u} \cos(2\pi f_{dt} + \varphi) \tag{1}$$

Where A_s, f_d, and φ are the pulse amplitude, dominant frequency and phase shift, respectively, while m models the initial finite slope of the pulse and u determines the final slope. Both m and u are parameters that depend on the type of ultrasound transducer.

Similarly, the amplitudes at the consequent echoes can be found for first echo, equation 2, and second echo, equation 3.

$$x_1(t_1) = A_{r1}(t_1)^m e^{-t_1/u} \cos(2\pi f_d t_1 + \varphi_1) \tag{2}$$

$$x_2(t_2) = A_{r2}(t_2)^m e^{-t_2/u} \cos(2\pi f_d t_2 + \varphi_1) \tag{3}$$

Where A_{r1} and A_{r2} are the amplitudes for first and second echoes; respectively, t_1 and t_2 are the start times for first and second echoes; respectively. φ_1 is the phase shift of both echoes.

Accordingly, TOF is defined as the interval between these two echoes, and can be obtained directly by subtracting the two detected times of first and second echoes, as given in equation (4).

$$\text{TOF} = t_2 - t_1 \tag{4}$$

Afterwards, USV is calculated in equation (5), by information about the path length of the test section (d).

$$\text{USV} = \frac{2 \times d}{\text{TOF}} \tag{5}$$
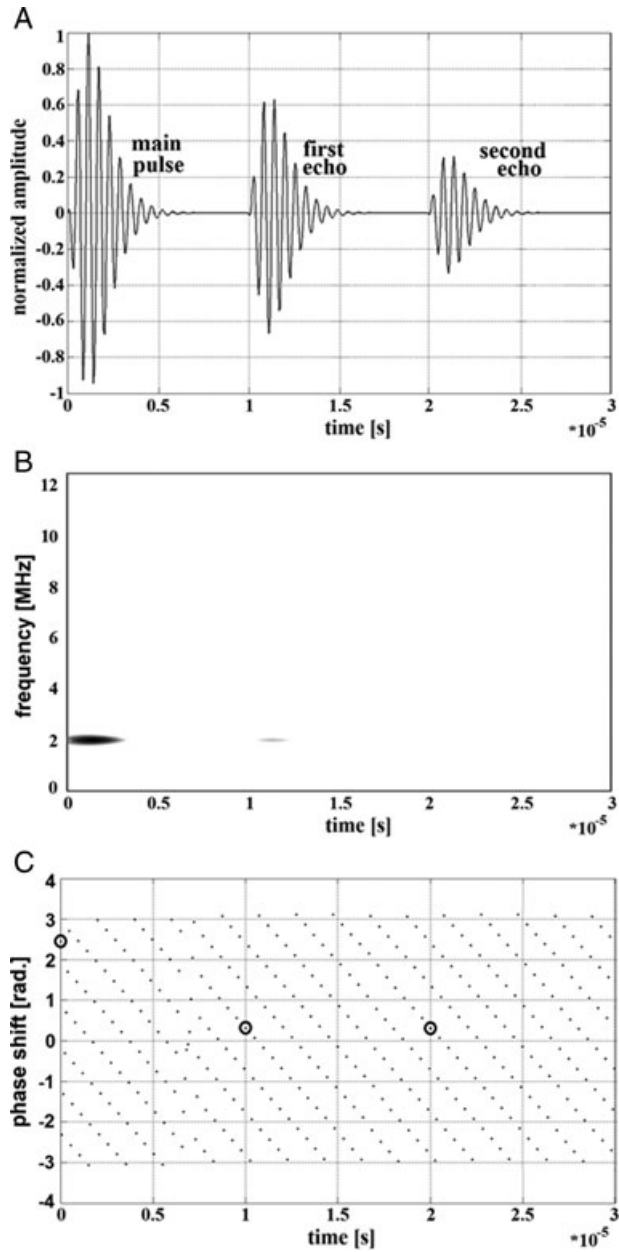
Figure 2. (A) An ideal ultrasound signal follows equation (1) with $m = 2$ and $u = 5 \times 10^{-7}$. (B) Spectrogram representation of this signal exploring its dominant frequency at 2 MHz. (C) Phase-shift distribution with time at the dominant frequency showing equal phase values for the first and second echoes.

Since real applications are not provided by ideal ultrasound signals, recognition of the consequent echoes is not straightforward and requires signal preprocessing. Actual transducers produce pulses with $m \geq 0$, which adds more difficulty to the task of finding the start of an echo. Also, a real signal contains many frequency bands beside its dominant frequency due to sensor non-perfect vibration and/or spectral leakage, at which some of them are mainly noise and removed by applying appropriate filters, while others are related to system vibration, thermodynamic parameters of the medium, or interpropaga-

tion diffraction of the echoes. Moreover, the signal is not fully damped among the echoes intervals, and there is an interfering signal produced by the attenuation and distortion between the propagating sample and the reflected waves. These reality considerations formed two major deviations between ideal and real ultrasound signals, as the real signal suffers from existence of multifrequency bands instead of single-dominant frequency, and existence of phase shifts among the consequent echoes.

The propagation of ultrasound signal in a real pulse-echo setup (Fig. 1B), is presented in Fig. 3A. The main pulse of the ultrasound signal, displayed by (arrow a), is excited at the transducer and propagates downwards into the buffer lid. By approaching the sample, the signal is particularly reflected toward the sensor (arrow b), and the remaining part (arrow c) that hopefully has enough energy (depends on the buffer material) transmitted into the sample. When the transmitted part approaches the reflector lid, much of the signal is reflected (arrow d) and the remaining small amount (depends on the reflector material) is transmitted into the reflector lid (arrow e).

All reflections (i.e. echoes) are sensed by the transducer and mirrored back into the system to repeat the whole process (arrows (a, b, c, d, and e), arrows (a, b, c, d, and e), etc.). With time, most of the signal energy is dissipated into the system elements (buffer, sample, and reflector) and the echoes become weaker until they are fully damped.

Limiting the ultrasound signal to exactly its dominant frequency (f_d) is a challenge due to the necessity of having high-resolution spectrogram (obtained by short time Fourier transform (STFT)), or scalogram (obtained by wavelet transform (WT)). However, the time resolution in WT is less at lower frequencies and high at higher frequencies, while it is fixed and reasonable with STFT (see Supporting information, Fig. S1). The spectral analysis of ultrasound signals can also be performed by Hilbert transformation, but it is very sensitive to noise and only used when the signal has few zero crossings.

As a result, STFT was employed in this work, because most ultrasound signals have low-dominant frequency, around 10% of the sampling frequency, and high time resolution is essential to accurately determine TOF between arrived echoes. The signal is divided into short overlapped segments; each with length equals 10% of the whole signal length, and 90% overlapping between adjacent segments. Since Fourier transformation presents spectral leakage for non-periodic signals, each segment is multiplied by a Hanning window function, equation (6), forcing it to be periodic [17]. The windowed segment is transformed to frequency domain by Fast Fourier Transform (FFT), equation (7), forming one vertical column in the generated spectrogram (see Supporting information, Fig. S2).

$$w(n) = 0.5 \left(1 - \cos \left(\frac{2\pi n}{N - 1} \right) \right) \tag{6}$$

$$X(m)_{\text{segment}} = \sum_{n=1}^{N} x(n) \times w(n - m) \times e^{-i(2\pi f_d)n},$$

$$m = 1, 2, 3, \ldots, M \tag{7}$$

Where *x, X, w, N,* and *M* are signal amplitude, resulting Fourier transform coefficient, window function, segment length, and
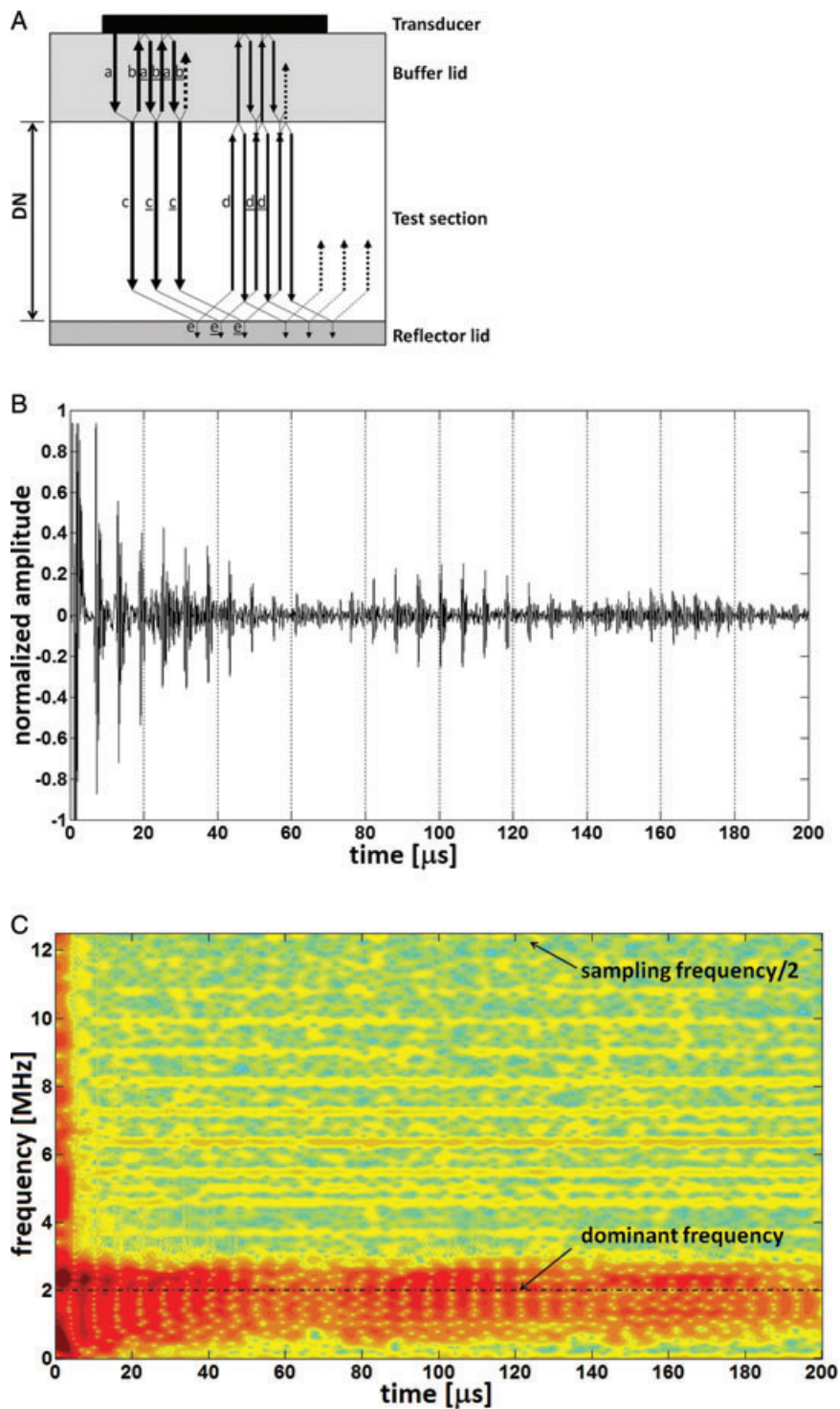
Figure 3. (A) Detailed schematic description for the propagation of the ultrasound signal into the real pulse-echo set up of Fig. 1B. (B) An ultrasound signal obtained for demineralized water sample at 10°C, and its (C) Spectrogram representation calculated by STFT [segment length = [20] μs, Hanning window, and 90% overlapping].

number of frequency pins in powers of 2 (i.e. 512, 1024, 2048, . . .), respectively.

The instantaneous power spectrum (P) at center of the working segment is given by equation (8), as the absolute value of the calculated Fourier coefficients. Accordingly, the instantaneous phase shift (φ) is the argument formulated by the real and imaginary parts of these Fourier coefficients, as given in equation (9).

$$P(m)_{\text{segment}} = \left| X(m)_{\text{segment}} \right|$$
$$= \sqrt{real(X(m)_{\text{segment}})^2 + imag(X(m)_{\text{segment}})^2} \quad (8)$$

$$\varphi\left(m\right)_{\text{segment}} = X\left(m\right)_{\text{segment}}$$

$$= \tan^{-1}\left(\frac{imag(X\left(m\right)_{\text{segment}})}{real(X\left(m\right)_{\text{segment}}))}\right) \in [-\pi, \pi] \quad (9)$$

Where $real(X\left(m\right)_{\text{segment}})$ and $imag(X\left(m\right)_{\text{segment}})$ are the real and imaginary parts of the Fourier coefficient, respectively.

An example for real ultrasound signal in pure water at $10°C$ is shown in Fig. 3B, and its spectrogram is seen in Fig. 3C. The instantaneous power spectrum and phase-shift distributions at $[f_d = 2\,\text{MHz}]$ are separated from this spectrogram and displayed in Figs. 4A and B, respectively.

The instantaneous power spectrum distribution (Fig. 4A) contains the most important information about how the signal is developed in the medium with details about immediate energy variation, arrival times of the consequent echoes, attenuation and damping behavior. In depth, point 1 is at maximum power spectrum value and indicates the start of first echo (t_1, P_1), while its following peaks at (1_a, $1b, \dots$) represent the gradual loss of the intensity (i.e. attenuation) with time. Meanwhile, point 2 is the point at which power spectrum starts to increase again after damping of the first echo, and indicates the start of second echo (t_2, P_2). Subsequent peaks (2_a, 2_b, 2_c) represent the gradual growth for the arrivals of leading constructive chirps of the second echo to the sensor, while (2_d, 2_e,) represent the trailing gradual loss of the second echo intensity. Further echoes are functions of the dramatically damping of the ultrasound signal in the system; therefore, they are less important for extraction and calculation of signal parameters. Meanwhile, from Fig. 4B, phase shifts of the first and second echoes (i.e. point 1 and point 2) were identified as φ_1 and φ_2, respectively.

The signal at point 1 can be mathematically modelled by equation (2), while at point 2, the phase shift is different from that at point 1, implying a modification to equation (3), which is rewritten here as equation (10).

$$x_1\left(t_1\right) = A_{r1}\left(t_1\right)^m e^{-t_1/u} \cos\left(2\pi f_d t_1 + \varphi_1\right) \quad (2)$$

And

$$x_2\left(t_2\right) = A_{r2}\left(t_2\right)^m e^{-\frac{t_2}{u}} \cos\left(2\pi f_d t_2 + \varphi_2\right)$$

$$x_2\left(t_2\right) = A_{r2}\left(t_2\right)^m e^{-\frac{t_2}{u}} \cos\left(2\pi f_d t_2 + \varphi_2 - \varphi_1 + \varphi_1\right)$$

$$x_2\left(t_2\right) = A_{r2}\left(t_2\right)^m e^{-\frac{t_2}{u}} \cos\left(2\pi f_d\left(t_2 + \left[\frac{\varphi_2 - \varphi_1}{2\pi f_d}\right]\right) + \varphi_1\right)$$

$$(10)$$

Thus, TOF is calculated by subtracting arguments of equations (10) and (2), as explained in equation (11).

$$\text{TOF} = \left(t_2 + \left[\frac{\varphi_2 - \varphi_1}{2\pi f_d}\right]\right) - t_1 = t_2 - t_1 + \left[\frac{\varphi_2 - \varphi_1}{2\pi f_d}\right]$$

$$(11)$$

The third term in equation (11) is the phase correction term, which is eliminated for ideal signal propagation. By estimating TOF, equation (5) is considered to calculate USV.

2.3 Density prediction by artificial neural network

During fermentation process, yeast cells use oxygen (O_2) and break down sugar ($C_6H_{12}O_6$) to carbon dioxide (CO_2) and water (H_2O). The density of the fermentation mixture is correlated to the sugar concentration (i.e. extract concentration), the higher the sugar concentration the higher the mixture density. In the absence of oxygen, yeast cells break down sugar to produce ethanol (C_2H_5OH) and (CO_2), wherein the density decreases with the rise of ethanol production. Therefore, mixture density is an important parameter that determines the fermentation level and its low values are used as an indication of less sugar concentration (i.e. apparent extract).

Hence, the calculated USV is combined with the measured Temperature T and relevant signal features to design an artificial neural network (ANN) for non-contact estimation of the mixture density (ρ_{est}) at any fermentation instance.

The procedure starts by extracting a list of signal features through the forms given in [18]. These features are extracted from the main chirp of the first echo, because it contains most of the reflected signal information. Relevant features are selected to be those with adequate variation with respect to their corresponding reference density (ρ_{ref}), satisfying $[\tan(20°) \leq |\overline{\frac{\partial \text{feature}}{\partial \rho_{ref}}}| \leq \tan(70°)]$, where $|\overline{\frac{\partial \text{feature}}{\partial \rho_{ref}}}|$ is average of the absolute feature-reference density variation (see Supporting information, Fig. S3).

Afterwards, a regression three-layer's ANN is designed (see Supporting information, Fig. S4). The input layer has one neuron for USV, one neuron for T, beside neurons for the selected relevant features. The hidden layer has a number of neurons equal 50–60% of those in the input layer, and the output layer has one neuron to calculate ρ_{est}. Each neuron weights the input, sums it up with a bias, and delivers it using a transfer function to the next layer. LOGSIG transfer function (equation (12)) is accomplished for the input layer neurons, and TANSIG (equation (13)) to both hidden and output layers neurons.

$$a = \frac{1}{1 + e^{-n}} \quad (12)$$

$$a = \frac{2}{1 + e^{-2n}} - 1 \quad (13)$$

Where n and a are the input to the transfer function and the neuron output, respectively.

This feed forward ANN loads N training data to the input layer {USV, T, relevant features}$_i$, $i = 1, \dots, N$, and their associated reference density $\{\rho_{ref}\}_i$ is compared to the resultant $\{\rho_{est}\}_i$ at the output layer. The network followed back propagation algorithm, which tries to minimize the mean square error (MSE) in ρ_{est}, according to equation (14), by modifying the weights and biases of the neurons [19].

$$MSE = \frac{1}{\text{No. of signals}} \sum_{i=1}^{\text{No. of signals}} \left(\rho_{ref} - \rho_{est}\right)_i^2 \quad (14)$$
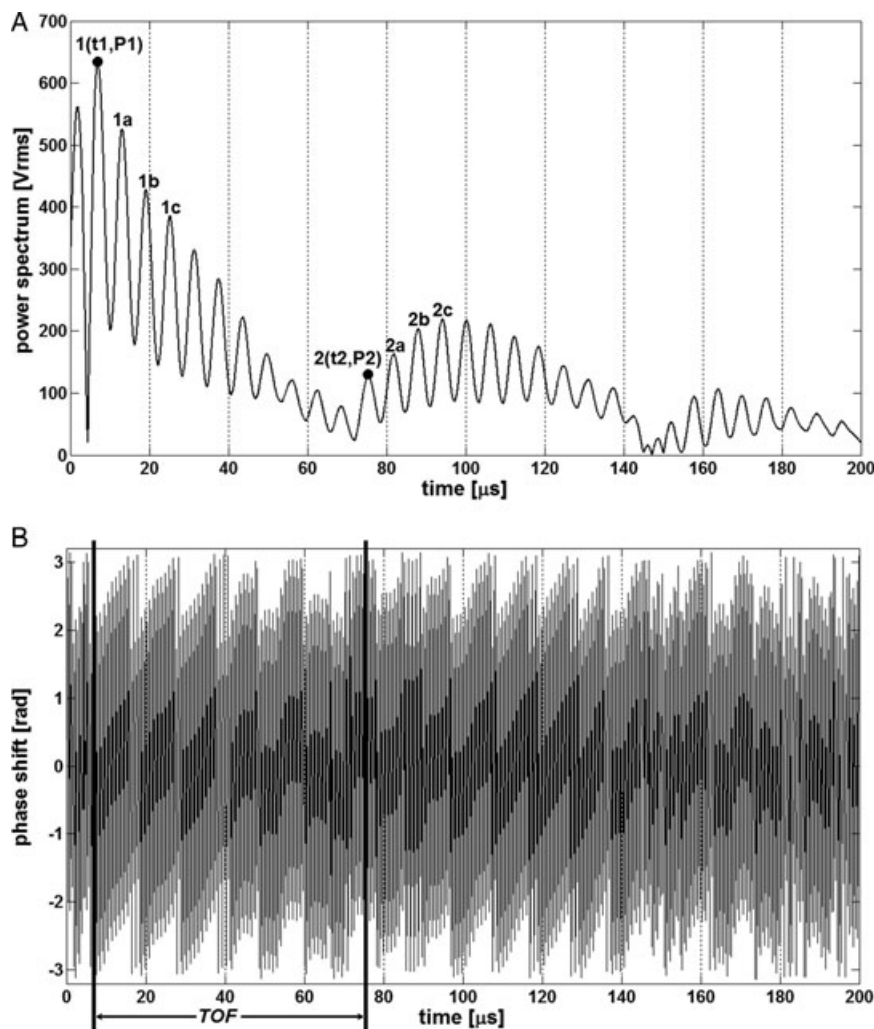
Figure 4. (A) Instantaneous power spectrum distribution in voltage-root mean square (Vrms) and (B) Instantaneous phase-shift distribution in radians (rad), over the signal at its dominant frequency exploring the start and damping behavior of each echo and differences in the phase shifts between first and second echoes.

3   Results

The validation, sensitivity evaluation, and application of the proposed method were performed in this section (see Supporting information, Fig. S5).

3.1   Validation of the proposed method

Demineralized vented water was selected to validate USV values obtained by the proposed method, with the reference results given in [20]. These reference data were fitted in a polynomial equation, displayed in equation (15), of six digits coefficients to ensure a precision of 0.02 m/s at ambient pressure and temperature (T) range 0–95°C.

$$USV = 1.402385 \times 10^3 + 5.038813T - 5.799136 \times 10^{-2} T^2$$
$$+ 3.287156 \times 10^{-4} T^3 - 1.398845 \times 10^{-6} T^4$$
$$+ 2.78786 \times 10^{-9} T^5 (m/s) \tag{15}$$

Experiments were performed using the pulse-echo setup of Fig. 1B with DN50, along a temperature range 9–30.8°C. Re-ceived signals, as shown in Fig. 5A, were analyzed to detect the arrivals of first and second echoes, point 1 and 2; respectively. Afterwards, equation (11) was used to estimate TOF values, and equation (5) to calculate the corresponding USV.

By substituting in equation (11) with ($t_1 = 7.1\,\mu$s, $\varphi_1 = -2.06$ rad) and ($t_2 = 73.72\,\mu$s, $\varphi_2 = -0.11$ rad)

$$TOF = 73.72 - 7.1 + \left[\frac{-0.11 - (-2.06)}{2\pi * 2} \right] \cong 66.77\mu s$$

and

$$USV = \frac{2 \times (50 \times 10^{-3})}{(66.77 \times 10^{-6})} = 1497.60 \text{m/s}$$

While the reference USV at T = 25.6°C, from equation (15) is 1498.301 m/s (error < 0.05%).

Over the temperature range 9–30.8°C, one signal was attained every 0.1°C, and the values of TOF and USV were calculated and compared to the reference values, as revealed from Fig. 5B. The results reflect the efficiency of the proposed method to calculate USV with a maximum error of 0.22% (at temperature = 20.8°C). Moreover, the speed of sound/temperature trend almost coincides with that of reference data.
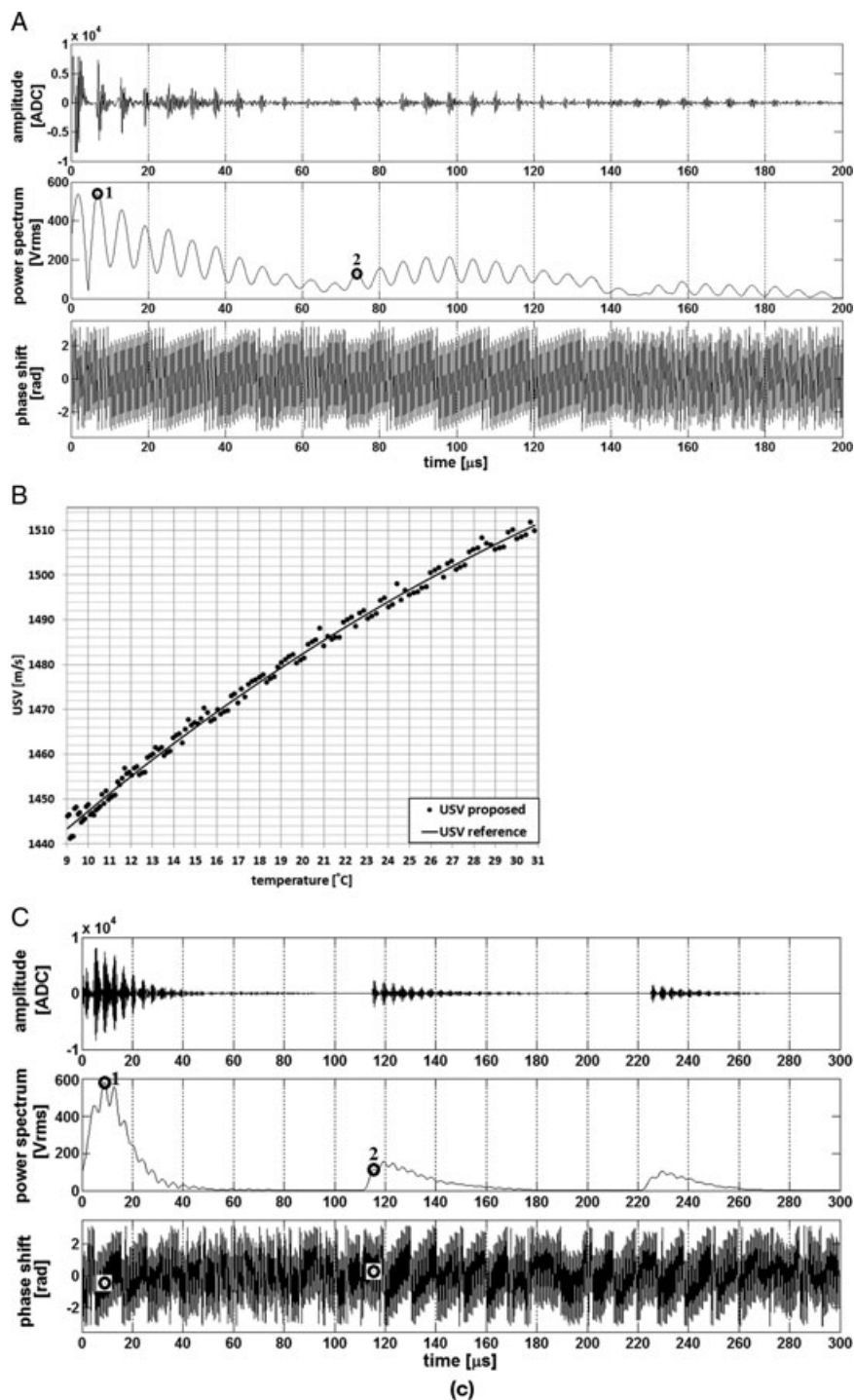
Figure 5. (A) Ultrasound signal obtained in demineralized water using the pulse-echo set up of Fig. 2 with DN50 at 25.6°C (up), and its instantaneous power spectrum (middle) and phase-shift (bottom) distributions at the dominant frequency.
(B) Speed of sound variation with temperature in demineralized water at 9–30.8°C, obtained by the proposed method in comparison to the reference data given by equation 15 [20].
(C) Ultrasound signal obtained for demineralized water using configuration II with DN80 at 25.6°C (top), and the instantaneous power spectrum (middle) and phase shift (bottom) distributions at its dominant frequency.

3.2 Sensitivity of the proposed method

To check the reliability of the results obtained by the proposed method, two sensitivity analyses had been applied. The first analysis is related to the repeatability of the results if the main parameters influencing the USV were fixed. Therefore, 50 experiments were accomplished for demineralized water at 10°C, using the pulse-echo set up of Fig. 1B with DN50, and the re-

sulting power spectrum distributions at dominant frequency are separated (see Supporting information, Fig. S6a).

As a result, no change in the detection of first echo throughout the 50 signals was observed and small perturbations influenced the detection of second echo (see Supporting information, Fig. S6b). The average and standard deviation in the calculated USV are 1447.28 m/s and 0.09 m/s, respectively, implying the consistency of the proposed method to provide reliable

Table 1. Average speed of sound obtained by the crosscorrelation and the proposed methods for demineralized water at temperatures of 10, 15, and 20°C, in comparison to the reference values. USV is the average speed of sound over 50 values at the specified temperature.

Temperature (°C)	USV by cross correlation (m/s)	USV by proposed method (m/s)	USV reference (m/s)
10	1447.152	1447.279	1447.291
15	1465.922	1465.921	1465.962
20	1482.384	1482.383	1482.382

results when the dependent variable (i.e. temperature) was fixed, with a precision of ($[0.09/1447.28] \times 100\% = 6.4 \times 10^{-3}\%$). Furthermore, low computational power is required to detect TOF and calculate USV for the 50 signals (around 2 s with 22.4 GFlops).

The second sensitivity analysis is related to the dependency of the results on the set up dimensions. Hence, the two nominal diameters of the pulse-echo set up (i.e. DN50 and DN80) were implemented to calculate USV in pure water at T = 25.6°C. The received signals by the two setups are presented in Figs. 5A and C, respectively, and analyzed by the proposed method.

For signal with DN50 setup:

$$\text{TOF}_1 = 66.77\mu s, \text{USV}_1 = 1497.60 m/s$$

For signal with DN80 setup:

$$\text{TOF}_2 = 115.6 - 8.99 + \left[\frac{0.27 - (-0.58)}{2\pi*2} \right] = 106.67\mu s,$$

$$\text{USV}_2 = \frac{2 \times (80 \times 10^{-3})}{(106.67 \times 10^{-6})} = 1499.89 m/s \quad (16)$$

Thus, the sensitivity of the proposed method to the configuration changes is:

$$\text{sensitivity} = \frac{\text{standard deviation (USV}_1, \text{USV}_2)}{\text{average (USV}_1, \text{USV}_2)}$$
$$\times 100\% \cong 0.11\% \quad (17)$$

Alternatively, 50 experiments were executed on demineralized water at 10, 15, and 20°C, and the calculated average USV are compared to those produced by the cross correlation method and published in [21], as listed in Table 1. The results explore how both methods (cross correlation, and proposed) produce high accurate USV around T = 20°C (error ca. $0.1 \times 10^{-3}\%$). This accuracy is less at lower temperature, for the cross correlation method (error ca. $9.6 \times 10^{-3}\%$), and for the proposed method (error ca. $0.82 \times 10^{-3}\%$).

3.3 Monitoring of yeast fermentation process

Fermentation process is one of the most important processes during malt production, due to its continuous biomixture change along the period of process, which ends in many progressions up to days. Yeast fermentation is considered to be four mixtures process consisting of yeast, extract, water, and carbon dioxide. As fermentation proceeds, density and refractive index fall rapidly and USV gradually rises [3]. Therefore, the ultrasound measurement becomes nearly independent of the degree of fermentation making it suitable for continuous inline monitoring of the process original gravity [22].

The proposed method was applied to monitor USV during a yeast fermentation process in the fermenter of Fig. 1C. The cylindroconical fermenter was initially filled with 60 liter of wort and 10 million cells of yeast. The temperature was controlled via a fuzzy control system to be in the range of 10–20°C during the process time which was between 24 and 96 h. The yeast suspension was under permanent circulation and was aerated by an aeration jet, until the yeast count reached 80–100 million cells. Moreover, the fermenter—that generated an ultrasound signal every 3 s—was equipped with a built in USV calculator based on the cross correlation method. Both results obtained by the proposed and cross correlation methods are displayed in Fig. 6A for 48 h of the fermentation process.

In comparison to cross correlation method, the proposed method presented a distribution free from extremely out of range values (i.e. outliers) and stable along the process time. And the maximum difference between both distributions in the non-outliers regions is 0.67%, demonstrating the effectiveness of the proposed method to be implemented in such inline processes.

Additionally, the mixture reference density (ρ_{ref}) was measured using an oscillating U-tube technique (Company Centec®) via a bypass connected to the circulation pipe. The signal features were extracted and those relevant to ρ_{ref} were selected (see Supporting information, Fig. S7). In this procedure, nine relevant features were defined and presented in Table 2.

These features were combined with the calculated USV and the measured *T* in the designed ANN for non-contact estimation of the mixture density (ρ_{est}). The input layer of the network has 11 neurons for the normalized values of (USV, *T*, and the selected nine features). The hidden layer was designed with six neurons, and the output layer with one neuron for the density value.

Altogether, 3685 signals (i.e. fermentation instances) have been used, at which (40% [1474 signals]) to train, (25% [921 signals]) to validate, and (35% [1290 signals]) to test the network. After 500 iterations, the training results, as shown in Fig. 6B, had reasonable MSE of 4×10^{-5} kg/m³. And the designed network was examined by the test signals, at which the estimated density values are shown in Fig. 6C, producing a maximum error in ρ_{est} of 0.95%. These results indicate the efficiency of the designed network with the selected signal features to estimate the mixture density values during the fermentation process, in non-invasive and non-contact approach. In practice, different error influences have to be taken into consideration as the pressure dependency, carbon dioxide influence, and the extract dependency on the present gravity.
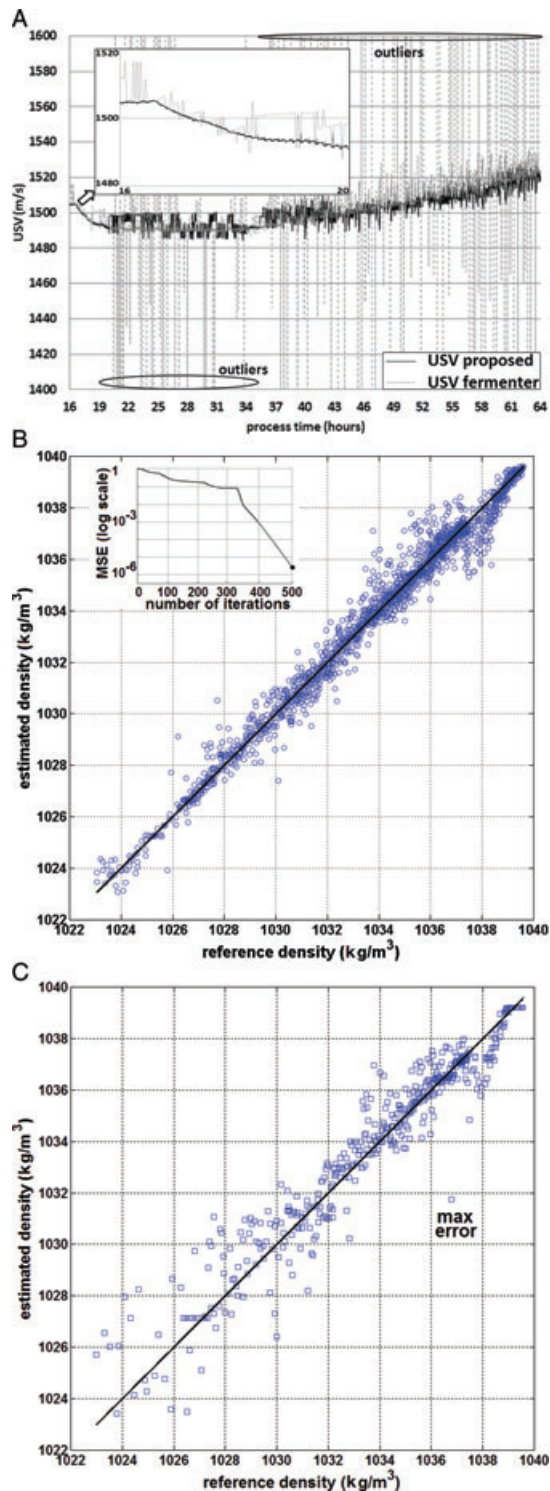
Figure 6. (A) USV distribution between the 16th and 64th h of fermentation, obtained by the proposed method (in solid black) and the built-in crosscorrelation-based calculator of the fermenter (in dashed gray), which suffers from many outliers along the process time. (B) ANN training results for 1474 signals after 500 iterations with MSE of 4×10^{-5} kg/m^3 between the estimated and reference densities of the fermentation mixture. (C) ANN testing results for 1290 wherein maximum error between the estimated and reference densities is 0.95%.

4   Concluding remarks

Ultrasound signal has the ability to interrogate fluids and dense mixtures in a non-destructive way, which makes it an ideal method for characterizing and process monitoring during food production. The two widely used methods to estimate TOF of an ultrasound signal—which is the most important parameter for USV determination and further analysis—are threshold and cross correlation methods. However, the first method showed considerable deviations that are mainly caused by high variations of the echo amplitudes, while the second method held error due to inaccurate determination of the start of the second echo, since other chirps may have higher correlation coefficient with the first echo.

In this paper, a USV estimation method based on the spectral analysis of pulse-echo ultrasound signal at its dominant frequency was presented, as well as a correction for the phase shift between the first and following echoes. To assure high time resolution, short time Fourier transform was applied to perform spectral transformation of the time domain signals.

Despite of the high-resolution Fourier transform required, the computational time is still within that taken by the cross correlation method, because only the small region of the spectrogram around the dominant frequency is considered, avoiding further calculations at other frequencies.

The method was validated by experiments in demineralized water at different temperatures in comparison to the cross correlation method. Two sensitive analyses were applied for the proposed method, proving its tendency to produce identical results if the experiment is repeated with precision of 6.4×10^{-3}%, and a sensitivity of 0.11% to changes in the pulse-echo setup dimensions. The proposed method was applied to automatically determine online USV in yeast fermentation process with a maximum error of 0.67%. Afterwards, USV was combined with temperature and nine signal features in an artificial neural network to estimate the instantaneous density of the fermentation mixture, producing a maximum error of 0.95%.

Although the first and second echoes were handled to calculate TOF for all experiments, subsequent echoes may also be used. However, the first and second echoes are qualitatively the best selection since they are least corrupted by any destructive issues during signal propagation.

In general, results of the proposed method, in comparison to those given by cross correlation method, showed how it can be powerfully applied for ultrasound measurement in food products. Moreover, the results can be improved if higher resolution spectral analysis is applied, with higher overlapping percentage, less frame length, and optimum selection of the window function.

In bioprocess, changes in the precision of temperature measurements should be well considered. The implemented PT100 temperature sensor has high accuracy (0.1°C at 0°C), low drift, and with resistance rate of 0.385 ohm/°C. Using sensors with higher rates allow obtaining higher resolution. For calibration, a voltage meter was applied to adjust the resistance value for an exactly well-known reference temperature. However, the precision of the temperature measurement is reduced slightly at higher temperatures, since it becomes difficult to prevent contamination of the platinum wire. The precision is also degraded due to

Table 2. Relevant features with respect to ρ_{ref}. Where *pmf*, *x*, *N*, and *M* are probability mass function, signal amplitude, segment length, and number of frequency pins; respectively. *X* is the resulting Fourier coefficient (equation (7)).

Feature	Equation
Spectral centroid	$\dfrac{\sum_{m=1}^{M} m.|X(m)|}{\sum_{m=1}^{M}|X(m)|}$
Spectral spread	$\sum_{m=1}^{M}(m - \text{spectral centroid})^2 \times pmf(m)\,,\; pmf(m) = \dfrac{|X(m)|}{\sum_{m=1}^{M}|X(m)|}$
Spectral kurtosis	$\dfrac{\sum_{m=1}^{M}(m - \text{spectral centroid})^4 \times pmf(m)}{(\text{spectral spread})^2}$
Temporal and spectral energy	$\sum_{n=1}^{N}(x(n))^2 \;\&\; \sum_{m=1}^{M}|X(m)|^2$
Temporal and spectral crest factor	$\dfrac{\max(|x(n)|)}{\frac{1}{N}\sum_{n=1}^{N}|x(n)|} \;\&\; \dfrac{\max(|X(m)|)}{\frac{1}{M}\sum_{m=1}^{M}|X(m)|}$
Temporal and spectral entropy	$-\sum_{n=1}^{N}\left(\dfrac{|x(n)|}{\sum_{n=1}^{N}|x(n)|}\right)^2 \ln\left(\dfrac{|x(n)|}{\sum_{n=1}^{N}|x(n)|}\right)^2 \;\&\; -\sum_{m=1}^{M} pmf(m)\times\ln\big(pmf(m)\big)$

the self-heating of the sensor as a result of the power applied to it. Further errors rise during temperature transients because the sensor may not respond to changes fast enough.

Practical application

The proposed method for estimating speed of sound propagation of ultrasound signals can be applied in non-destructive tests to check mediums quality and variations. In particular, monitoring of fermentation processes can be performed successfully in non-contact, non-invasive approaches. Furthermore, the combination of the estimated speed of sound with other signal features are intended to approximate online the important and non-sensed process variables such as mixture density, which thereafter determines the fermentation level and indicates the amount of the remaining sugar.

All measurements were performed in the laboratory of Chair of Brewing and Beverage Technology, Technical University of Munich, Freising, Germany.

The authors have declared no conflict of interest.

Nomenclature

A_s	[Voltage]	Pulse amplitude
f_d	[Hz]	Dominant frequency
m		Initial finite slope of the pulse
u		Final finite slope of the pulse
A_{r1}	[Voltage]	Amplitude on the first echo
A_{r2}	[Voltage]	Amplitude on the second echo
TOF	[s]	Time of flight
USV	[m/s]	Speed of sound
d	[m]	Path length through the test section
t_1	[s]	Detected time for the start of first echo
t_2	[s]	Detected time for the start of second echo

ADC		Analogue-to-digital converted signal amplitude
DN	[m]	Nominal diameter
FFT		Fast Fourier transform
STFT		Short time Fourier transform
WT		Wavelet transform
f_s	[Hz]	Sampling frequency
x	[Voltage]	Signal amplitude
X		Fourier transform coefficient
w		Hanning window function
N		Number of segment samples
M		Number of frequency pins
P	[Vrms]	Power spectrum
P_1	[Vrms]	Power spectrum at the start of first echo
P_2	[Vrms]	Power spectrum at the start of second echo
T	[°C]	Temperature
ANN		Artificial neural network
MSE		Mean square error
a		Neuron activation function
pmf		Probability mass function

Greek letters

φ	[rad]	Phase shift
φ_1	[rad]	Phase shift at the start of first echo
φ_2	[rad]	Phase shift at the start of second echo
ρ_{ref}	[kg/m³]	Reference density of the fermentation mixture
ρ_{est}	[kg/m³]	Estimated density of the fermentation mixture

5   References

[1] Andria, G., Attivissimo, F., Giaquinto, N., Digital signal processing techniques for accurate ultrasound sensor measurement. *Measurement* 2001, *30*, 105–114.

[2] Dolatowski, Z. J., Stadnik, J., Stasiak, D., Applications of ultrasound in food technology. *Acta Sci. Pol. Technol. Aliment.* 2007, *6*(3), 89–99.

[3] Hauptmann, P., Lucklum, R., Puettmer, A., Henning, B., Ultrasonic sensors for process monitoring and chemical analysis: state of the art and trends. *Sensors and Actuators A* 1998, *67*, 32–48.

[4] Henning, B., Rautenberg, J., Process monitoring using ultrasonic sensor systems. *Ultrasonics* 2006, *44*, 1395–1399.

[5] Bamberger, J., Bond, L., Greenwood, M., Ultrasonic measurements for on-line real-time food process monitoring. *Proceeding of 6th Conference on Food Engineering AIChE,* Texas, USA, 1999, pp. 1–4.

[6] Mason, T. J., Sonochemistry and sonoprocessing: the link, the trends and (probably) the future. *Ultrason. Sonochem.* 2003, *10*, 175–179.

[7] Cheeke, J. D., *Fundamentals and Applications of Ultrasound Waves.* First edition, CRC press LLC, USA 2002, 39–122.

[8] Guening, F., Varlan, M., Eugene, C., Dupuis, P., Accurate distance measurement by an autonomous ultrasonic system combining time-of-flight and phase-shift methods. *IEEE Trans. Instr. Meas.* 1997, *46*(6), 1236–1240.

[9] Schoeck, T., Becker, T., Sensor array for the combined analysis of water sugar ethanol mixtures in yeast fermentations by ultrasound. *Food Control* 2010, *21*(4), 362–369.

[10] Hosoda, M., Takagi, K., Ogawa, H., Nomura, H. et al., Rapid and precise measurement system for ultrasonic velocity by pulse correlation method designed for chemical analysis. *Japanese J. Appl. Phys.* 2005, *44*(5A), 3268–3271.

[11] Anderson, W. L., Jensen, C. E., Instrumentation for time-resolved measurements of ultrasound velocity deviation. *IEEE Trans. Instr. Meas.* 1989, *38*(4), 913–916.

[12] Webster, D., A pulsed ultrasonic distance measurement system based upon phase digitizing. *IEEE Trans. Instr. Meas.* 1994, *43*(4), 578–582.

[13] Carullo, A., Ferraris, F., Graziani, S., Grimaldi, U. et al., Ultrasonic distance sensor improvement using a two-levelneural network. *IEEE Trans. Instr. Meas.* 1996, *45*(2), 677–682.

[14] Parrila, M., Anaya, J. J., Fritsch, C., Digital signal processing techniques for high accuracy ultrasonic range measurements. *IEEE Trans. Instr. Meas.* 1991, *40*(4), 759–763.

[15] Marioli, D., Narduzzi, C., Offelli, C., Petri, D. et al., Digital time-of-flight measurement of ultrasonic sensor. *IEEE Trans. Instr. Meas.* 1992, *41*(1), 198–201.

[16] Duncan, M. G., Real-time analytic signal processor for ultrasonic non-destructive testing. *IEEE Trans. Instr. Meas.* 1990, *39*(6), 1024–1029.

[17] Hussein, W. B., Hussein, M. A., Becker, T., Detection Of The red palm weevil rhynchophorus ferrugineus using its bioacoustics features. *Bioacoustics* 2010, *19*, 177–194.

[18] Peeters, G., *A large set of audio features for sound description (similarity and classification).* CUIDADO project report, AES 115th Convention, New York, USA 2004, pp. 3–25.

[19] Chauvin, Y., Rumelhart, D. E., *Backpropagation: Theory, architecture, and applications (1st edn.),* Psychology Press, New Jersey, USA 1995, 1–35.

[20] Marczak, W., Water as a standard in the measurements of speed of sound in liquids. *J. Acoust. Soc. Am.* 1997, *102*(5). 2776–2779.

[21] Hoche, S., Hussein, W., Hussein, M. A., Becker, T., Time of flight prediction for fermentation process monitoring. *Eng. Life Sci.* 2011, *11*(3), 1–12.

[22] Skrgatic, D. M., Mitchinson, J. C., Graham, J. A., Measurement of specific gravity during fermentation. United States Patent 4959228, 1990.

4. Conclusion and Outlook

4.1 Conclusion

In this thesis, numerical models belong to signal and image processing techniques are developed to help in improving the solution of associated Life Science Engineering (LSE) problems. The appropriateness of LSE problem to be possible application for the developed models depends on either it is originated by a sound producer (e.g., weevils detection), or related to process which can be examined by ultrasound measurements (e.g., food process monitoring). In addition to, problem which can be identified in pictures (e.g., swelling recognition in biological tissues) or simply distinct in spectrogram images (e.g., sound pattern based animal classification).

The scheme followed in this thesis can be concluded in the flowchart displayed in figure (4.1), starting by processing the signal/image of LSE problem. The major objective is to improve signal/image analysis through the developed numerical models, and extract more relevant and accurate features. Afterwards, suitable pattern recognition techniques (such as ANN) are applied in detecting, identifying, or classifying the involved patterns and solve the corresponding problem.
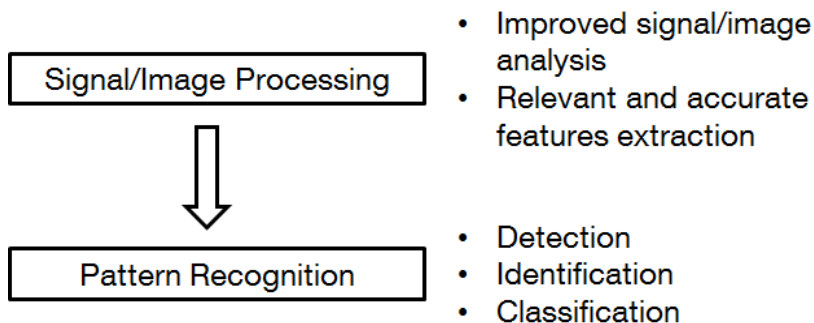

Figure (4.1): Concluding diagram for the scheme which is followed in the thesis to solve LSE problems.

Altogether, four numerical models have been developed and applied to four LSE problems, as summarized in table (4.1).

Table (4.1): The developed numerical models and their solved LSE problems.

Developed model	Solved LSE problem
Relevant features extraction for sound pattern recognition	Detecting the existence of red palm weevil (*Rhynchophorus ferrugineus*)
Improved edge detection approach with sub-pixel accuracy	Fat contents recognition in meat slices
Enriched spectrogram enhancement approach protecting pattern structure	Classification of birds/bats bioacoustics calls
Robust speed of sound estimation approach for ultrasound signals	Monitoring of yeast fermentation bioprocess and its mixture density

The first model concerns with the extraction of large list of temporal and spectral features including some unconventional features, to completely define the sound print of a sound source. Application of window functions to smooth the ends of non-periodic signals and reduce the spectral leakage in calculating Fourier transform coefficients is

discussed. The performance of these window functions is divided into two categories; the first is for those functions increase the frequency resolution, while the second is for those reduce the spectral leakage. A third category may be defined for those functions increase the amplitude resolution; however this is not significant categorization because in most applications the calculated coefficients are normalized before extracting the spectral features.

Beside the common algorithms for dominant features selection, two simple selection approaches are introduced. These approaches are Jacobian based feature selection and Steadiness based feature selection, for regression and classification pattern recognition problems; respectively.

The developed model is applied to detect the existence of Red Palm Weevil (RPW) – which is the most destructive pest of date palm in the world – in an earlier phase of infestation. Its hidden kind of living makes an early detection with traditional methods (e.g. pheromone traps) not applicable, while it can be performed using bioacoustics recognition. Several sensor constructions are built and the best recordings are obtained by an insertion sensor, which brings the sensor near the source of the sound. The recordings are digitized at 11025 Hz, high-pass filtered with cutoff frequency of 200 Hz. The recordings are divided into segments wherein optimum segment length is investigated along several overlapping percentages, and found to be 0.42 sec when 90% overlapping is engaged. Rectangular window is used to calculate spectral features that require high frequency resolution (e.g. spectral roll-off), while Blackman window is applied to calculate those require less spectral leakage (e.g. spectral centroid). For each frame; 8 temporal and 10 spectral relevant features are selected, including some unconventional features, building up a RPW detection system. The developed system detects correctly the location of RPW feeding sounds' patterns in a five minutes recording.

After the positive determination of RPW presence in a palm, it is necessary to remove the palm tree, because up today no real efficient treatment method exists. But with the removing of an infested palms and disinfection of the soil, a strong blow against RPW population is made.

The second model is for an improved edge detection method which extracts features images (energy and skewness) from the original grey scale image. These images behave as smoothed versions of the original image and avoid the using of prior smoothing filters (such as Gaussian filter) and their scaling constraints. The gradient images of these features images are obtained by 3*3 Sobel derivative operators, and edges are detected as the peaks in the gradient images. Afterwards, Hysteresis thresholding is applied to remove noisy weak edges, and non-maximum suppression method is applied to reduce the thickness of the detected edges to one pixel width. Since edges are not always directed to the center of pixels, a quadratic interpolation is made among adjacent pixels to estimate the sub-pixel edge strength. Furthermore, a flux equilibrium check is performed on the obtained edges image to detect any discontinuities along the edge direction and fill them.

Investigations are conducted to find the optimum values for controlling parameters of the model. The first parameter is the size of the mask which slips over the original image to extract the feature images. The second parameter is the size of the flux equilibrium matrix which slips over the edges to check and fill discontinuities. And the third parameter is the combination coefficient between the output edges images (one from energy image and the other from skewness image) to produce the final edges

image. The optimum values for these three parameters are found to be 5*5, 3*3, and 0.75; respectively. The model is competitively compared to other edge detectors through several subjective and objective evaluation measures. For the reference images which are implemented in this comparison, the signal to noise ratio is increased by at least 6.4%. The mean square error is reduced by at least 1.6%, while the localization error and energy dissipation are improved by at least 7.8% and 0.25%; respectively.

The developed model is applied to detect the boundaries of fat content in meat slices images. These images are initially pre-processed by a histogram thresholding analysis which eliminates the background noise. And by an anisotropic diffusion filter which reduces the specular reflections and speckles due to the light reflection from the moist meat areas and muscle fibrous texture. Once boundaries are detected, the region growing method is employed to calculate the area enclosed by them. The traditional and standard reference method to measure fat content in meat is the organic solvent based method. Its drawbacks are the long drying and extraction time, lack of automation, high consumption of solvent, and it is a destructive testing method.

The third model presents an enriched spectrogram enhancement approach which does not affect the structure of the sound patterns. Spectrogram is selected as suitable time-frequency representation because of its reasonable constant time and frequency resolutions. But scalogram, which is an alternative to the spectrogram and obtained by wavelet transform, accomplishes low frequency resolution at higher frequencies and low time resolution at lower frequencies.

The spectrogram dynamic range is limited to 40 dB level to focus on the contained high energetic patterns. Afterwards, a crest factor image is extracted as a feature image for the spectrogram, and processed by the developed edge detection model to accurately define the boundaries of the sound patterns. The usage of crest factor image instead of energy/skewness combination is because this combination presents both strong edges (from energy feature) and weak edges (from skewness feature). As a result the attached noise may also be detected as sound patterns.

Interiors of the detected boundaries are reconstructed with their original sound power, while the exteriors (i.e., attached noise) are eliminated. As a result, the enhanced spectrogram is produced with the sound patterns in their almost original temporal and spectral locations, giving an efficient base for further pattern recognition tasks to be performed.

The model is compared to other leading edge enhancement approaches for reference spectrograms of birds and bats, over a range of added white Gaussian noise. Results show the efficiency of the model to enhance spectrograms with higher signal to noise ratio and segmental signal to noise ratio, as well as lower log spectral distortion and Itakura Saito.

The enhanced spectrogram is implemented to obtain power and frequency contours of the enclosed sound patterns. The power contour may be used for specific sound power detection, and the frequency contour may be used in designing more reliable filter. Furthermore, the enhancement spectrogram is accessed to reconstruct the enhanced wave form of the bioacoustics calls, which can be adopted for accurate feature extraction process.

The developed model is applied to classify bioacoustics calls of bird and bat species. The sound pattern is distinctive for each sound producing animal, and this appears completely in the spectrogram representation because it contains both temporal and

spectral information. Whereas the sound pattern may be mismatched if separately considered in time or frequency domains. From the enhanced spectrogram image, several two dimensional features are extracted to define the sound patterns of the candidate species. The classifier is designed by two relevant features, namely the eccentricity and centroid of the sound pattern. After training the classifier with enough sound patterns of each species, it correctly classified 94.59% of the test sounds.

The fourth model is for an improved calculation of the speed of sound (*USV*) during ultrasound measurements, which is an important parameter to monitor simultaneously variations in the tested samples, in noncontact and non-destructive way. In practice, echoes of the propagated ultrasound signals are not clearly parted and involve phase shifts, result in difficult determination of their time of flight and its corresponding *USV*. Therefore, the signal is divided into short segments (length = 10% signal length) with overlapping > 90%, and multiplied by suitable window function reducing the spectral leakage, to obtain high resolution spectrogram. The Fourier coefficients at the dominant frequency are separated and their power spectrum distribution and phase shift distribution are established. Accordingly the starts of first and second echoes are accurately detected and their phase shifts are obtained. Time of flight is the interval between these starts' times corrected with their phase shifts, and *USV* is calculated by information of the signal path length. The model is validated by reference data of speed of sound in demineralized water at temperatures [9 – 30.8] ℃, producing a maximum error of 0.217%. Sensitivity analyses are performed on the results implying their repeatability with a deviation of $6.4*10^{-3}$% over 50 repeated measurements, and their consistency with a deviation of 0.107% when the dimensions of the implemented ultrasound setup is changed. Results are compared to those produced by cross correlation method on a continuous process monitoring, generating more stable *USV* distribution free from noticeable outliers with less computational time.
The technique offers a deeper insight into bioprocesses and thus their monitoring, optimization, and control. As a result, the model is applied online in a yeast fermentation process accompanied by continuous ultrasound measurements. The calculated *USV* is combined with signal relevant features and the measured temperature values in an artificial neural network to estimate instantaneously the mixture density with a maximum error of 0.95%. This density is important to determine the fermentation level, amount of produced alcohol, as well as an indication for concentration of the remaining sugar.

4.2 Outlook

An important advantage of the developed tools is their provision of non-destructive, noncontact, and low cost solution to the associated LSE problems. Therefore, in some applications, more precautions have to be considered. For example, in recording the sounds produced by hidden weevils such like red palm weevils, holes are made in the palm tree, to bring the sensor to the closest place from the sound source. These holes should be small enough to avoid tree destruction, and the sensor has to be contentiously anti-infectious treated to prevent spread of infestations among trees.
Moreover, higher sensor sensitivity to the produced sounds is double-edged sword. Since it can detect lower energy weevil sounds and in the same time detect more of the weak noise. Meanwhile, a next step to the generated bioacoustics detection

system is to develop a learning scheme to handle all the available features for the feeding sound in different environmental conditions and field sounds, such like machines, animals, and other existing insects. Adding some features for the description of noise-like sounds should enhance the system to become a useful tool in similarity search for the sound calls of insects.

On the other hand, although edges image limits the data included in an original image to the objects boundaries information, edge detectors may require long computational time since pixel-by-pixel investigation is necessary. This time is further extended if sub-pixel accuracy is targeted as in the developed edge detector, because edges directions will cover all the range between [0-360] ° instead of covering only the basic directions {0, ±45, ±90, ±135, 180}.

The developed edge detection approach provides no information about the frame pixels of the original image, due to the implementation of central finite difference in calculating the gradient image. These pixels are located on first and last rows, and first and last columns. This issue can be sidestepped by applying forward or backward finite difference at the frame pixels. More simple solution is to avoid having useful information from taking place in frame pixels when capturing the original image. Furthermore, the 3*3 Sobel operator is applied to obtain the gradient image in a quickly but less accurate manner. Higher order finite difference operators suchlike 5*5 increases the accuracy order from $O(h^2)$ to $O(h^4)$, as described in figure (4.2). Unfortunately, this procedure increases the overall computational time.
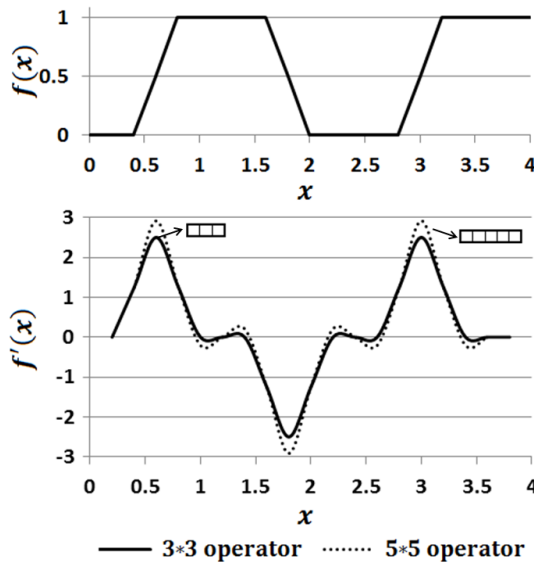

Figure (4.2): Comparison between 3*3 and 5*5 derivative operators in obtaining the gradient of general function $f(x)$.

The flux equilibrium check helped to fill the discontinuities in edges, as long as these discontinuities are seen within the selected 3*3 flux check matrix. Therefore, the selection of such matrix size is application oriented which is dependent on the investigated image and object size.

In determining fat contents in meat slices by the developed approach, the slice is assumed to be 2D or has uniform distribution of fat along the third dimension. For other cases, an under-layer analysis has to be added suchlike ultrasound imaging.

Alternatively, for the developed spectrogram enhancement approach, different parameters have to be studied with respect to the achieved enhancement level. One of these parameters is the selected dynamic range which limits the spectrogram to its

inclusion of high energy sound patterns. Other features images may also be investigated beside the crest factor image to improve the enhancement process. Furthermore, parameters involved in the frequency domain transformation influence the obtained spectrogram resolution, such as overlapping percentage, window function, and window function size as described in figure (4.3).

Clearly, in classification of bioacoustics calls, much care should be paid in selecting the relevant features. The clusters will be far apart for features that have good discriminatory ability, whereas the clusters may overlap for other features. The samples selected as training data are crucial for the recognition success and assumed to have all possible pattern representatives; otherwise, species may be wrongly classified.
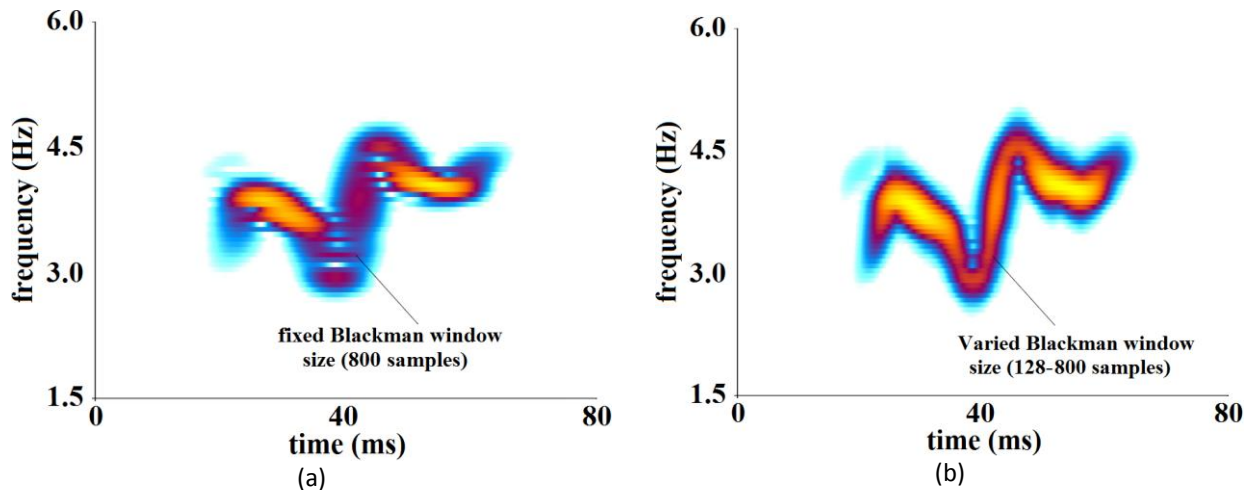

Figure (4.3): Spectrogram of *Prunella collaris* bioacoustics call obtained by: (a) fixed window size, and (b) varied window size. [78]

In regard to the developed speed of sound estimation approach, reverse engineering may help to design more suitable window function which eliminates spectral leakage, and returns accurate frequency and power spectrum. Wavelet transform and Hilbert transform can also be candidates to Fourier transform in seeking more precise spectral analysis. For correct detection of time of flight, complete damping of each echo is essential, which is not only related to the sample under study but also to the implemented transducer system.

Due to the complex nature of microbial growth, monitoring and control of bioprocesses represent an ever-increasing engineering challenge. For further optimisation of bioprocesses and ensuring high product quality, the lack of accurate real time monitoring will be the bottleneck. Therefore, in the years to come, increasing focus has to be given to online monitoring techniques. The use of online measurements to monitor bioprocesses for identifying cell status has become more predominant. All sensors take part in monitoring the bioprocess should be evaluated periodically for any changes to their sensing accuracy. These changes may be a result of their contamination, life time, and/or weak incompetent calibration. Ultrasonic measurements are not applicable when the sonic attenuation of the sample is very high. This can occur in substances with high content of bubbles or voids. Also no effects are measured in composite samples when the speeds of sound of the components are identical.

Generally, more samples collection - whether signals or images – improves the training of the corresponding pattern recognition system and ensures consistent and reliable

performance when applied to new test samples. The typical problems include how to represent complex samples, and how to exclude spurious (unstable) samples to avoid over fitting problems.

So where should it goes from here? The ultimate goal of this thesis is to aid in paving the way for extended research applications out of the tools and applications presented here. Some of these applications and further thoughts are presented in table (4.2).

Table (4.2): Some of the possible future directions and LSE applications where the developed models can prove to be beneficial.

Developed model	Possible applications
Relevant features extraction for sound pattern recognition	• Detection of others hidden or night migrating animals (e.g., rice weevil, flour weevil, and geese). • Scared systems for birds on airplanes (based on generating sound features of birds enemies). • Source identification (based on database for sound features of possible sources). • Bioacoustics monitoring of animal population. • Acoustic communication studies. • Detection of alarm calls in species suchlike red ruffed lemur (for earthquake, tornado, or volcano warnings). • Study environmental influence (e.g., temperature) on sound sensitivity. • Acoustic localization of fishes.
Improved edge detection approach with sub-pixel accuracy	• Identification of general objects in images (based on their edges distribution and 2D features). • Cancer and pictured seen diseases diagnosis. • Movement tracking (based on the changing of center of gravity position of the detected object). • Biometric analysis. • Mobile robot vision system. • Measurement of critical objects dimension and inspection for missing parts. • Analysis of microscopy images.
Enriched spectrogram enhancement approach protecting pattern structure	• Robust identification of sound producing species. • Learning sound characteristics. • Filtration of audio signals for further processing or checking a filter performance. • Phonetics and speech recognition. • Studying frequency modulation in animal calls. • Reverse engineering of sound patterns. • Creation of modern music systems. • Development of radio frequency and microwave systems.
Robust speed of sound estimation approach for ultrasound signals	• Monitoring the development of other food processes such as dough mixing. • Monitoring of medical products processes such as monoclonal antibody in hybridoma cultures.

	• Assisting in the adaptive control of bioprocess progression.
	• Non-destructive inspection of food quality.
	• Shelf life examination of products.
	• Under layer compositions identification.

Major challenges of the future such as food, environment, and health care at a sustainable cost are related to life sciences. The convergence of life sciences with engineering is an emerging opportunity that will help generate many exciting future achievements. Therefore, the implementation of cheap, robust, non-destructive, and noncontact signal and image processing techniques should not be overlooked for investment and development. It is a very large and exciting field of research and there continue to be many developments.

Appendix A: List of temporal and spectral features

Time domain features

These features are extracted from the signal waveform distribution, and used to express the variation of samples amplitudes with time.

1. Zero crossing rate: is the number of points at which the signal passes the time axis, through the whole pattern, as shown in figure (A1.1)

$$\frac{1}{2*N}\sum_{n=2}^{N}\left|\text{sign}(x(n)) - \text{sign}(x(n-1))\right|, \qquad \begin{cases} \text{sign} = 1 & ; x(n) > 0 \\ \text{sign} = 0 & ; x(n) = 0 \\ \text{sign} = -1 & ; x(n) < 0 \end{cases}$$

Where $x(n)$ is the amplitude at sample n, and N is the number of samples.
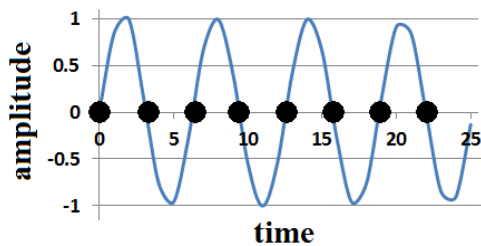

Figure (A1.1): Schematic diagram for calculating zero crossing rate

2. Short Time Energy: Is the energy content of the pattern and calculated by summing up all the areas under the curves of the signal, as shown in figure (A1.2)
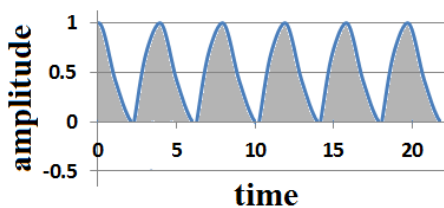
$$\sum_{n=1}^{N}(x(n))^2$$

Figure (A1.2): Schematic diagram for calculating short time energy

3. Energy Root Mean Square: represents also the energy content, but averaged over the pattern length

$$\sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N}}$$

4. Temporal Centroid: is the time at which half of the pattern energy was covered. It is not necessary to be exactly in the middle of the pattern length, as shown in figure (A1.3).

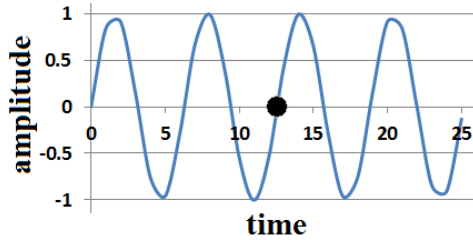$$C_t = \frac{\sum_{n=1}^{N} n*(x(n))^2}{\sum_{n=1}^{N}(x(n))^2}$$

Figure (A1.3): Schematic diagram for temporal centroid

5. Temporal Entropy: It measures the disorder of a signal, with low value for highly ordered signals (i.e., has less noise and fewer bits are sufficient to describe its information).

$$-\sum_{n=1}^{N}\left(\frac{|x(n)|}{\sum_{n=1}^{N}|x(n)|}\right)^{2}\ln\left(\frac{|x(n)|}{\sum_{n=1}^{N}|x(n)|}\right)^{2}$$

6. Temporal Crest-Factor: It is the ratio of the maximum magnitude on the pattern to the average of the amplitudes, and this represents the singularity of the pattern, as shown in figure (A1.4).

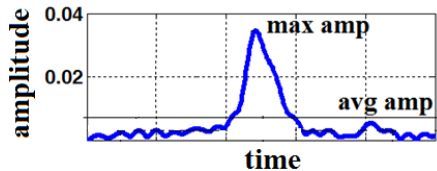$$\frac{\max(|x(n)|)}{\frac{1}{N}\sum_{n=1}^{N}|x(n)|}$$

Figure (A1.4): Schematic diagram for temporal crest factor

7. Temporal Slope: It represents the amount of decreasing or increasing of the pattern amplitude, and is computed by linear regression, as shown in figure (A1.5).

$$\frac{N\sum_{n=1}^{N}\left(n*x(n)\right)-\sum_{n=1}^{N}n*\sum_{n=1}^{N}x(n)}{\sum_{n=1}^{N}x(n)*\left(\sum_{n=1}^{N}n^{2}-\left(\sum_{n=1}^{N}n\right)^{2}\right)}$$

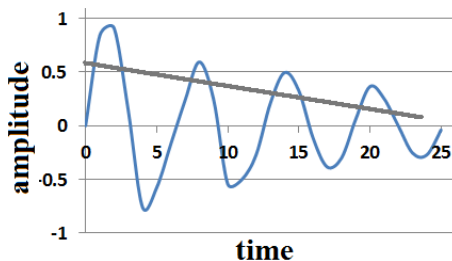Figure (A1.5): Schematic diagram for temporal slope

8. Temporal Flux: represents how sharp the amplitudes change on the same pattern, and measured by summing up all the squaring values for the difference between the normalized magnitudes of adjacent amplitudes, as shown in figure (A1.6).

$$\sum_{n=1}^{N}\left(\left|\frac{|x(n)|}{\max(|x|)}\right|-\left|\frac{|x(n+1)|}{\max(|x|)}\right|\right)^{2}$$
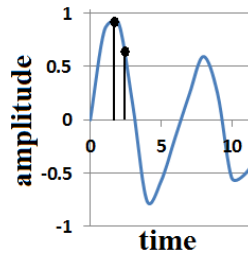
130

Figure (A1.6): Schematic diagram for temporal flux

9. Temporal Decrease: represents the decay part of the pattern, and measured by the slope of the line connecting the last point with amplitude equals 0.8 of the maximum amplitude, and the last point with amplitude equals 0.08 of the maximum amplitude, as shown in figure (A1.7).

10. Descent time: is the period covered by the temporal decrease line, as shown in figure (A1.7).

11. Temporal increase: represents the attack part of the pattern, and measured by the slope of the line connecting the first point with amplitude equals 0.8 of the maximum amplitude, and the first point with amplitude equals 0.08 of the maximum amplitude, as shown in figure (A1.7).

12. Attack time: is the period covered by the temporal increase line, as shown in figure (A1.7)

13. Duration: represents the interval between first and last points with amplitudes equal to 0.1 of the maximum amplitude, as shown in figure (A1.7).
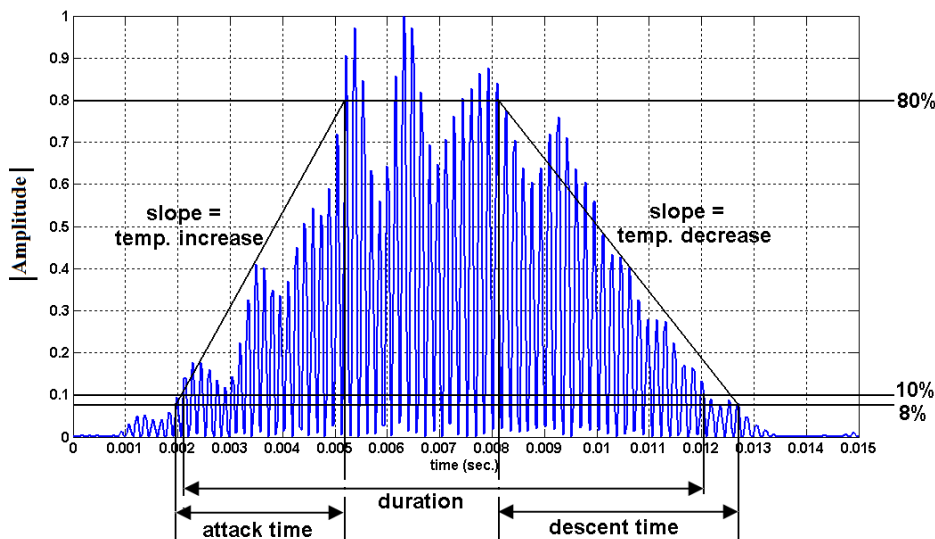

Figure (A1.7): Schematic diagram for temporal decrease, descent time, temporal increase, attack time, and duration

14. Temporal Roll-off: the sample in the time domain below which 90% of the energy distribution is concentrated, as shown in figure (A1.8)

$$\sum_{n=1}^{n_{rf}} |x(n)|^2 = 0.9 * \sum_{n=1}^{N} |x(n)|^2$$

Where n_{rf} is the sample number of the temporal roll-off
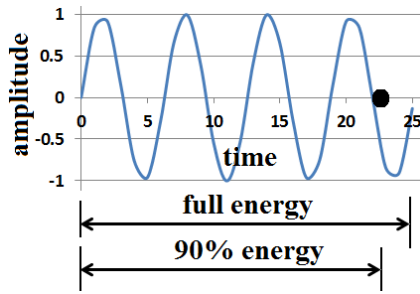
Figure (A1.8): Schematic diagram for temporal roll-off

15. Temporal Flatness: It is a measure for the tendency of the pattern to hold constant (flat) amplitudes, it equals "1" for constant amplitudes pattern and smaller values for other pattern types. It is calculated by the ratio of the geometric mean to the arithmetic mean of the pattern amplitudes, as shown in figure (A1.9).

$$tf = \frac{\left(\prod_{n=1}^{N} |x(n)|\right)^{1/n}}{\frac{1}{N} \sum_{n=1}^{N} |x(n)|}$$

Figure (A1.9): Schematic diagram for temporal flatness

16. Temporal Tonality: It is related to the flatness of the pattern, and indicates whether tone or noise exists in the pattern. It is value is close to "1" for tonal patterns and close to "0" for noisy patterns, as shown in figure (A1.10).

$$\min\left(\frac{10 * \log(\text{temporal flatness})}{-60}, 1\right)$$

Figure (A1.10): Schematic diagram for temporal tonality

132

17. Temporal Spread: Represents the variance of the pattern amplitudes around its mean value.

$$tsp = \sum_{n=1}^{N}(n - \text{centroid})^2 * tpmf(n)\,, \qquad tpmf(n) = \frac{|x(n)|}{\sum_{n=1}^{N}|x(n)|}$$

Where $tpmf$ is the probability mass function

18. Temporal Skewness: gives a measure for the degree of asymmetry of the pattern amplitudes with respect to the normal distribution, as shown in figure (A1.11).

$$m_3 = \sum_{n=1}^{N}(n - \text{centroid})^3 * tpmf(n)\,, \qquad temporal\ skew = \frac{m_3}{tsp^{3/2}}$$

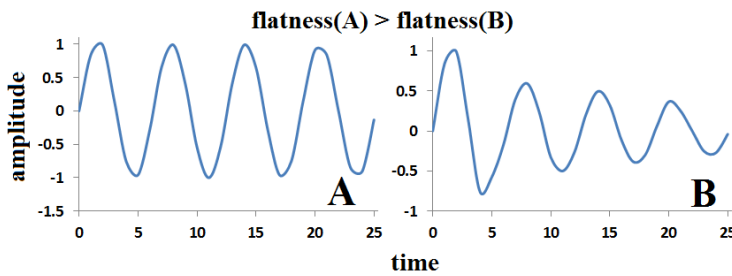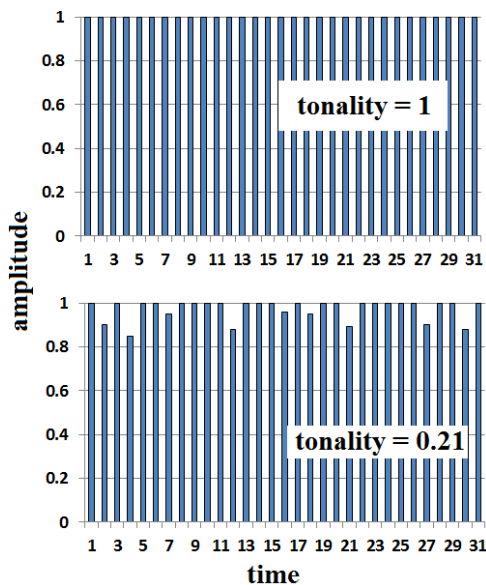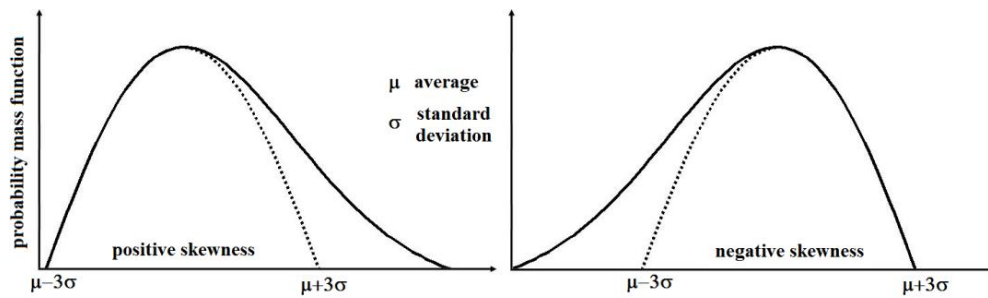Figure (A1.11): Schematic diagram for temporal skewness

19. Temporal Smoothness: It measures the smoothness variation of the pattern amplitude with respect to its two neighbours.

$$20 * \sum_{n=2}^{N-1}\left|\log|x(n)| - \frac{\log|x(n-1)| + \log|x(n)| + \log|x(n+1)|}{3}\right|$$

Frequency domain features

In the frequency domain, many spectral features may be extracted and implemented in combination to those in time domain, to fully characterize the signal. Many of these spectral features are calculated with the same mathematical formulas given for temporal features, but using the power spectrum and frequency instead of the signal amplitude and time; respectively.

20. Spectral Roll-Off:

$$\sum_{m=1}^{m_{rf}}|X(m)| = 0.9 * \sum_{m=1}^{1024}|X(m)|$$

Where m_{rf} is the frequency bin representing the spectral roll-off, $X(m)$ is the spectrum at bin number m, and 1024 is the total number of frequency bins.

21. Spectral Flux:

$$\sum_{m=1}^{1024} \left(\left| \frac{|X(m)|}{\max(|X(m)|)} \right| - \left| \frac{|X(m+1)|}{\max(|X(m)|)} \right| \right)^2$$

22. Spectral Centroid:

$$C_t = \frac{\sum_{m=1}^{1024} n.\,|X(m)|}{\sum_{m=1}^{1024}|X(m)|}$$

23. Spectral Root Mean Square:

$$\sqrt{\frac{\sum_{m=1}^{1024}(|X(m)|)^2}{1024}}$$

24. Spectral Slope:

$$\frac{1024 * \sum_{m=1}^{1024}(m * |X(m)|) - \sum_{m=1}^{1024} m * \sum_{m=1}^{1024}|X(m)|}{\sum_{m=1}^{1024}|X(m)| * (\sum_{m=1}^{1024} m^2 - (\sum_{m=1}^{1024} m)^2)}$$

25. Spectral Flatness:

$$\frac{(\prod_{m=1}^{1024}|X(m)|)^{1/m}}{\frac{1}{1024}\sum_{m=1}^{1024}|X(m)|}$$

26. Spectral Crest Factor:

$$\frac{\max(|X(m)|)}{\frac{1}{1024}\sum_{m=1}^{1024}|X(m)|}$$

27. Spectral Energy:

$$\sum_{m=1}^{1024}|X(m)|^2$$

28. Spectral Tonality:

$$\min\left(\frac{10 * \log(\text{spectral flatness})}{-60}, 1\right)$$

29. Spectral Smoothness:

$$20 * \sum_{m=2}^{1023} \left| \log|X(m)| - \frac{\log|X(m-1)| + \log|X(m)| + \log|X(m+1)|}{3} \right|$$

30. Spectral Spread:

$$ssp = \sum_{m=1}^{1024}(m - C_t)^2 * pmf(m) \, , \qquad pmf(m) = \frac{|X(m)|}{\sum_{m=1}^{1024}|X(m)|}$$

Where pmf is the probability mass function

31. Spectral Skewness:

$$\frac{\sum_{m=1}^{1024}(m - C_t)^3 * pmf(m)}{ssp^{3/2}} \quad \begin{cases} < 0 & \text{more energy on the right} \\ = 0 & \text{symmetric distribution} \\ > 0 & \text{more energy on the left} \end{cases}$$

32. Spectral Entropy:

$$-\sum_{m=1}^{1024} pmf(m) * \ln\big(pmf(m)\big)$$

33. Spectral Kurtosis: a measure of the flatness and singular variation of the spectrum around the mean, as shown in figure (A1.12)

$$\frac{\sum_{m=1}^{1024}(m - C_t)^4 * pmf(m)}{ssp^2} \quad \begin{cases} < 3 & \text{flatter distribution} \\ = 3 & \text{normal distribution} \\ > 3 & \text{peaker distribution} \end{cases}$$
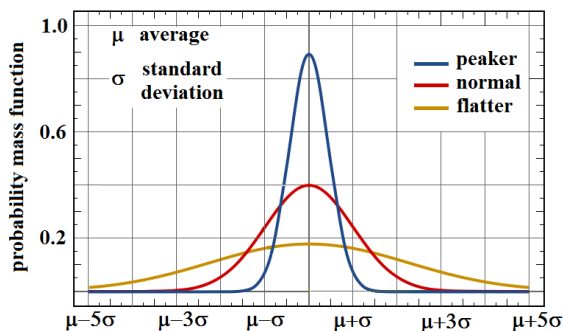
Figure (A1.12): Schematic diagram for spectral kurtosis

34. Spectral Band Width: the frequency band between first and last points with power spectrum equal to 0.1 of the maximum amplitude, as shown in figure (A1.13).

35. Spectral Decrease: the decay part of the spectrum, and measured by the slope of the line connecting the last point with power equals 0.8 of the maximum value, and the last point with power equals 0.08 of the maximum value, as shown in figure (A1.13).

36. Descent band: is the frequency band covered by the spectral decrease line, as shown in figure (A1.13).

37. Spectral increase: the attack part of the spectrum, and measured by the slope of the line connecting the first point with power equals 0.8 of the maximum value, and the first point with power equals 0.08 of the maximum value, as shown in figure (A1.13).

135

38. Ascent band: is the frequency band covered by the spectral increase line, as shown in figure (A1.13).

Figure (A1.13): Schematic diagram for band width, spectral decrease, descent band, spectral increase, and attack band

References

1. Miller, D.A., Arnett, E.B., and Lacki, M.J.: "Habitat management for forest-roosting bats of North America: a critical review of habitat studies". Wildlife Society Bulletin 31, 2003, 30-44.
2. Johnson, G.D.: "A review of bat mortality at wind-energy developments in the United States". Bat Res News 46, 2005, 45–49.
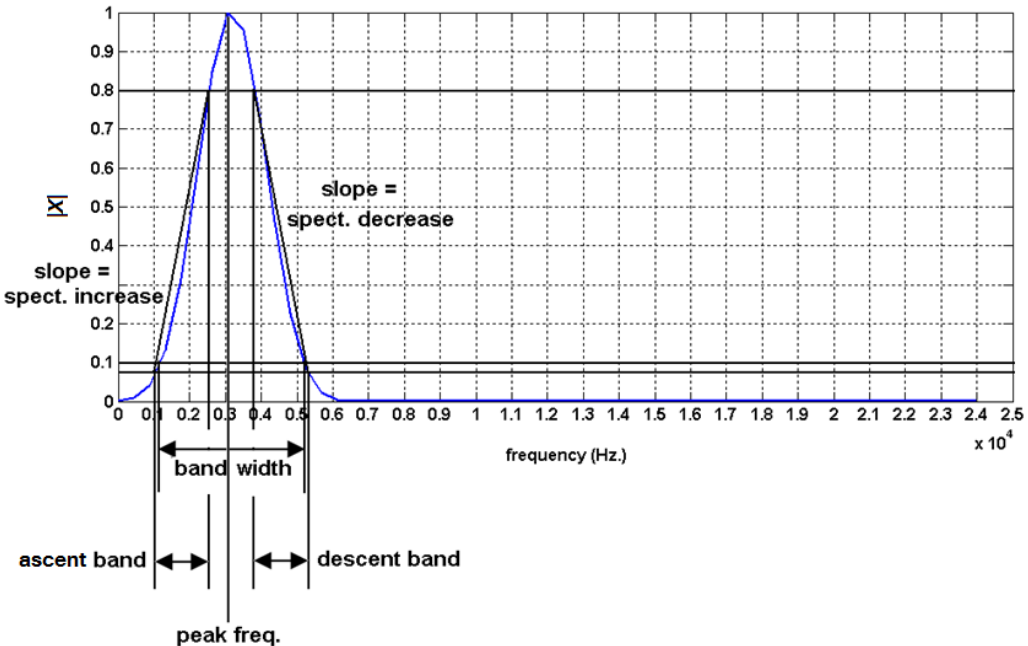3. Kunz, T.H., Arnett, E.B., Erickson, W.B., Hoar, A.R., Johnson, G.D., Larkin, R.P., Strickland, M.L., Thresher, R.W., and Tuttle, M.D.: "Ecological impacts of wind energy development on bats: questions, research needs, and hypotheses". Front Ecol Environ 5(6), 2007, 315-324.
4. Parsons, S. and Jones, G.: "Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks". Journal Exp. Biol. 203, 2000, 2641-2656.
5. Malumphy, C., and Moran, H.: "Plant Pest Notice: Red palm weevil (Rhynchophorus ferrugineu)". Central Science Laboratory, Sand Hutton, York, U.K., 2007.
6. Murphy, S.T., and Briscoe, B.R.: "The red palm weevil as an alien invasive: biology and prospects for biological control as a component of IPM". Biocontrol News and Information, 20, 1999, 35-45.
7. Howard, F.W., Moore, D., Giblin-Davis, R., and Abad, R.: "Insects on Palms". Oxford: CABI Publishing, Wallingford, 2001.
8. James, W.D., and Berger, T.G.: "Andrews' Diseases of the Skin". Clinical Dermatology. Saunders Elsevier. ISBN 0721629210, 2006.
9. Moser, H.W., Raymond, G.V., and Dubey, P.: "Adrenoleukodystrophy: new approaches to a neurodegenerative disease". JAMA 294 (24), 2005, 3131–4.
10. Penfield, M.P., Costello, C.A., McNeil, M.A., and Rienmann, M.J.: "Effects of fat level and cooking methods on physical and sensory characteristics of restructure beef streaks". Journal of Food Quality 11, 1989, 349-356.
11. Pethick, D.W., Harper, G.S., Hocquette, J.F. and Wang, Y.H.: "Marbling biology-what do we know about getting fat into muscle?". Australian Beef-the Leader Conference, Australia, 2006, 103-110.
12. AOAC, "Official Methods of Analysis". 14th ed., Washington, DC, 1991. http://www.eoma.aoac.org/.
13. Eller, F.J., and King, J.W.: "Supercritical fluid extraction of fat from ground beef: effects of water on gravimetric and GC-FAME fat determinations". Journal of Agricultural and Food Chemistry 49, 2001, 4609–4614.
14. Tahkoniemi, H., Helmja, K., Menert, A., and Kaljurand, M.: "Fermentation reactor coupled with capillary electrophoresis for on-line bioprocess monitoring". Journal of Pharmaceutical and Biomedical Analysis 41, 2006, 1585-1591.
15. Scheper, T., Hitzmann, B., Stärk, E., Ulber, R., Faurie, R., Sosnitza, P., and Reardon, K.F.: "Bioanalytics: detailed insight into bioprocesses". Analytica Chimica Acta 400, 1999, 121–134.
16. Lee, H.L., Boccazzi, P., Gorret, N., Ram, R., and Sinskey, A.: "In situ bioprocess monitoring of Escherichia coli bioreactions using raman spectroscopy". Vibrational Spectroscopy 35, 2004, 131-137.
17. Alford, J. S.: "Bioprocess control: Advances and challenges". Computers & Chemical Engineering 30, Issues 10-12, 2006, 1464-1475.
18. Kumar, M. A.: "Biosensors and Automation for Bioprocess Monitoring and Control". PhD thesis, Department of Biotechnology, Lund University, Sweden, 2011.
19. Soroker, V., Blumberg, D., Haberman, A., Hamburger-Rishard, M., Reneh, S., Talebaev, S., Anshelevich, L., and Harari, A.R.: "Current status of red palm weevil infestation in date palm plantations in Israel". Phytoparasitica 33, 2005, 97-106.

20.   Thomas, P., Kannan, A., Degwekar, V.H., and Ramamurthy, M.S.: "Non-destructive detection of seed weevil-infested mango fruits by X-ray imaging". Postharvest Biology and Technology 5 (1-2), 1995, 161-165.

21.   Elliott, W.R., Kaufmann, J.E., Samoray, S.T., and Gardner, S.E.:"The MDC Method: Counting Bats with Infrared Video". proceeding of National Cave and Karst Management Symposium, 2005, 147-153.

22.   Hristov, N.I., Betke, M., and Kunz, T.H.: "Applications of thermal infrared imaging for research in aeroecology".Integr. Comp. Biol. 48(1), 2008, 50-59.

23.   Gauthreaux, S.J., and Livingston, J.W.: "Monitoring bird migration with a fixed-beam radar and a thermal imaging camera". J Field Ornith 77, 2006, 319-328.

24.   Eekhout, X., Solís, G., and Márquez, R.: "Acoustic playback: Contributions to the study of animal communication in the last 13 years (1992-2004)". Advances in Bioacoustics 2, Dissertationes Classis IV: Historia Naturalis, Slovenian Academy of Sciences and Arts (Ljubljana), XLVII-3, 2006, 165–175.

25.   Holderied, M.W., Corine, C., Fenton, M.B., Parsons, S., Robson, S., and Jones. G.: "Echolocation call design and intensity in the aerial hawking bat Eptesicus bottae (Vespertilionidae) studied using stereo videogrammetry". J Exp Biol 208, 2005, 1321-1327.

26.   Surlykke, A., and Kalko, E.: "Echolocating Bats Cry Out Loud to Detect Their Prey". PLoS ONE 3(4), 2008, doi:10.1371/journal.pone.0002036

27.   Schnitzler, H.U., and Kalko, E.: "Echolocation by insect-eating bats". BioScience 51, 2001, 557–569.

28.   Juneja, M., and Sandhu, P.S.: "Performance evaluation of edge detection techniques for images in spatial domain". International Journal of Computer Theory and Engineering 1 (5), 2009, 614–621.

29.   Sharifi, M., Fathy, M., and Mahmoudi, M.T.: "A classified and comparative study of edge detection algorithms" Proceeding of the International Conference on Information Technology: Coding and Computing, 2002, 117–120.

30.   Nadernejad, E., Sharifzadeh, S., and Hassanpour, H.: "Edge detection techniques: evaluations and comparisons". Applied Mathematical Sciences 2 (31), 2008, 1507–1520.

31.   Ding, L., and Goshtasby, A.: "On the canny edge detector". Pattern Recognition 34 (3), 2001, 721–725.

32.   Orci, K.M.: "On the bioacoustics and morphology of some species-groups of Orthoptera". PhD thesis, Faculty of Science, University of Debrecen, Hungary, 2002.

33.   Obrist, M.K., Pavan, G., Sueur, J., Riede, K., Llusia, D., and Márquez, R.: "Bioacoustic approaches in biodiversity inventories". Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories, Abc Taxa. 8, 2010, 68-99.

34.   Li, H., Zhang, Y., and Xu, D.: "Noise and Speckle Reduction in Doppler Blood Flow Spectrograms Using an Adaptive Pulse-Coupled Neural Network". EURASIP Journal on Advances in Signal Processing, 2010, 1-11.

35.   Mellinger, D. K.: "Ishmael 1.0 User's Guide". Pacific Marine Enviromental Laboratory, Seattle, USA, 2002.

36.   Liu, R.C., Miller, K.D., Merzenich, M.N., and Schreiner, C.E.: "Acoustic variability and distinguishability among mouse ultrasound vocalizations," J. Acoust. Soc. Am. 114, 2003, 3412–3422.

37.   Ding, H., Soon, I.Y., Koh, S.N., and Yeo., C.K.: "A spectral filtering method based on hybrid Wiener filters for speech enhancement". Speech Communication 51, 2009, 259–267

38.   Ghanbari, Y., and Karami, M.R.: "A new approach for speech enhancement based on adaptive thresholding of wavelet packets". Speech communication 48, 2006, 927-940.

39.   Chavan, M.S., Chavan, M.N., and Gaikwad, M.S.: "Studies on Implementation of Wavelet for Denoising Speech Signal". International Journal of Computer Applications 3(2), 2010, 1-7

40. Hermansky, H., Wan, E.A., and Avendano, C.: "Noise Suppression in Cellular Communications". Proceedings IEEE IVTTA'94, Kyoto, Japan, 1994, 85-88.

41. Kim, H.G., Obermayer, K., Bode, M., and Ruwisch, D.: "real time noise cancelling based on spectral minimum detection and diffusive gain factors". Proceeding of 8th Aust. Int. Conf. Speech Sci. and Tech., 2000, 250-255.

42. Ghanbari, Y., Karami, M.R., and Amelifard, B.: "Improved multi-band spectral subtraction method for speech enhancement". Proceedings of the 6th ISTED International conference Signal And Image Processing, Hawaii, USA, Agust 23-25, 2004, 225-230.

43. Moore, J.D., Pardo, M.A., Popplewell, J.F., Spencer, S.J., Ray, M., Swann, J., Shard, A.G., Jones, W., Hills, A., and Bracewell D.G.: "Chemical and biological characterisation of a sensor surface for bioprocess monitoring". Biosensors and Bioelectronics 26, 2011, 2940-2947.

44. Bohnke, C., Duroy, H., and Fourquet, J.L.: "PH sensors with lithium lanthanum titanate sensitive material: applications in food industry". Sens. Actuators B Chem. 89, 2003, 240–247.

45. Vojinovic, V., Cabral, J.M.S., and Fonseca. L.P.: "Review: Real-time bioprocess monitoring, Part I: In situ sensors". Sensors and Actuators B 114, 2006, 1083–1091.

46. Kostov, Y., Harms, P., Randers-Eichhorn, L., and Rao, G.: "Low-cost microbioreactor for high-throughput bioprocessing". Biotechnol. Bioeng. 72, 2001, 346–352.

47. Lee, H.L.T., Boccazzi, P., Gorret, N., Ram, R.J., and Sinskey, A.J.: "In situ bioprocess monitoring of Escherichia coli bioreactions using Raman spectroscopy" Vib. Spectrosc. 35, 2004, 131–137.

48. Parrila, M., Anaya, J.J., and Fritsch, C.: "Digital signal processing techniques for high accuracy ultrasonic range measurements" IEEE Trans. Instr. Meas. 40(4), 1991, 759-763.

49. Marioli, D., Narduzzi, C., Offelli, C., Petri, D., and Sardini, A.E.: "Digital time-of-flight measurement of ultrasonic sensor". Proceeding of Instrumentation and Measurement Technology Conference IMTC-91, 1991, 198-201.

50. Schöck, T., and Becker, T.: "Sensor array for the combined analysis of water–sugar–ethanol mixtures in yeast fermentations by ultrasound". Food Control 21(4), 2010, 362–369.

51. Hussein, W.B., Hussein, M.A., and Becker, T.: "Detection of the Red Palm Weevil Using Its Bioacoustics Features". Journal of Bioacoustics 19(3), 2010, 177-194.

52. Schirmer, M., Hussein, W.B., Hussein, M.A., Jekle, M., and Becker, T.: "Impact of air humidity in industrial heating processes on selected quality attributes of bread rolls". Journal of Food Engineering 105, 2011, 647-655.

53. Wallhäußer, E., Hussein, W.B., Hussein, M.A., and Becker, T.: "On the usage of acoustic properties combined with an artificial neural network - a new approach of determining presence of dairy fouling type A". Journal of Food Engineering 103, 2011, 449-456.

54. Elfawakhry, H., Hussein, M.A., Becker, T.: "Possibility of Using Acoustic Techniques for Dough Processing Evaluation". 11th International Congress on Engineering and Food, ICEF11, Athens, Greece, 2011-05-22.

55. Hussein, W.B., Moaty, A.A., Hussein, M.A., and Becker T.: "A novel edge detection method with application to the fat content prediction in marbled meat". Pattern Recognition 44 (12), 2011, 2959-2970.

56. Hussein, W.B., Hussein, M.A., and Becker, T.: "spectrogram enhancement by edge detection approach applied to bioacoustics calls classifications". Signal and Image Processing: An International Journal 3(2), 2012, 1-20.

57. Hussein, W.B., Hussein, M.A., and Becker, T.: "Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation process". Eng. Life Sci. 12(6), 2012, 603-614.

58. Smith, S.W.: "Digital signal processing: a practical guide for engineers and scientists". Newnes Elsevier science, USA, 2003.

59. Cedeno, A.M., Dominguez, J.Q., Januchs, M., and Andina, D.: "Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network". Proceeding of 36th annual conference on IEEE Industrial Electronics Society, 2010, 2845-2850.

60. Jain, A.K., and Zongker, D.: "Feature-Selection: Evaluation, Application, and Small Sample Performance". IEEE Trans. on Pattern Analysis and Machine Intelligence 19(2), 1997, 153-158.

61. Lin, T.H., Li, H.T., and Tsai, K.C.: "Implementing the Fisher's Discriminant Ratio in a k-Means Clustering Algorithm for Feature Selection and Data Set Trimming". J. Chem. Inf. Comput. Sci. 44(1), 2004, 76–87.

62. Nixon, M.S., and Aguado, A.S.: "Feature Extraction & Image Processing". 2nd Edition, Academic Press Elsevier, Oxford, UK, 2008.

63. Gonzalez, R., and Woods, R.: "Digital image processing". Second Edition, Pearson Education North Asia Limited and Publishing House of Electronics Industry, 2002.

64. Zhang, D., and Lu, G.: "Review of shape representation and description techniques". Pattern Recognition 37, 2004, 1-19.

65. Han, S., and Yang, S.: "An invariant feature representation for shape retrieval". Proc. Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies, 2005, 1052-1054.

66. Liu, Y.K., and Zalik, B.: "An efficient chain code with Huffman coding". Pattern Recognition 38 (4), 2005, 553-557.

67. Folland, G., and Sitaram, A.: "The Uncertainty Principle: a mathematical survey". Journal of Fourier Analysis and Applications 3 (3), 1997, 207–238.

68. Boashash, B.: "Time-Frequency Signal Analysis and Processing – A Comprehensive Reference". First edition, Elsevier Science, Oxford, UK, 2003.

69. Eargle, J.: "Handbook of recording engineering", 4th edition, Springer, New York, USA. 2005, 1-27.

70. Andria, G., Attivissimo, F., and Giaquinto., N.: "Digital signal processing techniques for accurate ultrasound sensor measurement". Measurement 30, 2001, 105-114.

71. Marczak, W.: "Water as a standard in the measurements of speed of sound in liquids". Journal Acoustic Society of America 102(5), 1997, 2776-2779.

72. Mingqiang, Y., Kidiyo, K., and Joseph, R.: "A survey of shape feature extraction techniques". Pattern Recognition, chapter 3, IN-TECH, Vienna, Austria, 2008.

73. Deller, J.R., Hansen, J.H.L., and Proakis, J.G.: "Discrete-Time Processing of Speech Signals". Second ed. IEEE Press, New York, 2000.

74. Basheer, I.A., and Hajmeer. M.: "Artificial neural networks: fundamentals, computing, design, and application". Journal of Microbiological Methods 43, 2000, 3–31.

75. Russell, S., and Norvig, P.: "Artificial Intelligence: A Modern Approach". Second edition, Prentice Hall, New Jersey, USA, 2002.

76. Drosopoulos, S. and Claridge, M.: "Insect sounds and communication: Physiology, behaviour, ecology and evolution". CRC Press LLC, New York, USA, 2006.

77. Stabinger, H.: "Density Measurement using modern oscillating transducers". South Yorkshire Trading Standards Unit, Sheffield, 1994.

78. Huebner, S. V.: "Computer aided design of audio signal classifier systems". Journal of acoustical society of America 123(5), 2008, 5711-5715.

Personal Information

Date of Birth   : 24/11/1978
Nationality   : Egyptian
Marital State   : Married since 14/10/2007, 1 son
Home Address   : Kepserstr. 48, 85356 Freising,
                        Deutschland
Tel.   : +49-1701631627
E-mail Address   : whussein@wzw.tum.de

Education

- B.Sc. in aerospace engineering, July 2000, Faculty of Engineering, Cairo University.
    - ➢ Grade: very good with honor degree
    - ➢ Project grade: Excellent

- Pre- master courses in aeronautical engineering, August 2001, Faculty of Engineering, Cairo University.

- M.Sc. in Aerospace Engineering: "On calculation of Aerodynamic Coefficients and derivatives for A complete Aircraft configuration", Faculty of Engineering, Cairo University, August 2005.

- Eligibility of Master Degree from Technische Universität München, (Note: 1.3), August 2012.

Current Job

Wissenschaftlich Mitarbeiter, Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising. (since June 2009)

Previous Jobs

- Wissenschaftlich Mitarbeiter, Institut für Lebensmittelwissenschaft und Biotechnologie, Universität Hohenheim, Stuttgart. (Mai 2007 – Mai 2009).
- Teacher Assistant, German University in Cairo, Cairo. (August 2004 – April 2007)
- Researcher; Institute of Aviation Engineering and Technology, Egyptian Civil Aviation Academy, Cairo. (September 2000 – July 2004).

Professional Affiliations

- Active Member, Egyptian Engineering Syndicate
- Active Member, VDI, Association of German Engineers
- Active Member, DECHEMA, Society for Chemical Engineering and Biotechnology
- Active Member, EURASIP, European Association for Signal Processing
- Active Member, DAGM, German Association for Pattern Recognition
- Active Member, DGPF, German Society for Photogrammetry and remote sensing

Awards

- Efficient Award, Institute of Aviation Engineering and Technology, 2002
- Recipient Outstanding Lecturer Award, German University in Cairo, 2006
- Nominated in Who's Who in Science and Engineering, 11th edition, 2010
- Nominated in Who's Who in The World, 29th edition, 2012

Computer Skills

- Visual Basic® / C++®   : Programming and graphical user interface generation.
- Matlab®   : Programming and implementation of Signal/Image Processing Toolboxes, Filter design, ANN, and Pattern Recognition Toolbox.
- AutoCAD®   : Design and rendering 3D shapes.

Languages

- Arabic   : Mother Tongue
- English   : Excellent
- German   : Level A2 (currently enrolled for level B1)

Hobbies

- Volley ball, chess and swimming.
- Listening to music.
- Reading in history and arts.

Publications list
Selected reviewed papers

[1]   Hussein, W. B.; Hecker, F.; Mitzscherling, M.; Becker, T.: "Modelling and Simulation of bakery production lines for process analysis and optimization". Proceeding of the FOODSIM (2008), Dublin, Ireland, 26-29.06.2008, pp. 45-49

[2]   Hussein, W. B.; Hecker, F.; Mitzscherling, M.; Becker, T.: "Computer Modelling and Simulation of Bakeries' Production Planning". International Journal of Food Engineering, Vol. 5, issue 2, 2008. DOI: 10.2202/1556-3758.1565

[3]   Hecker, F., W. B. Hussein, M. Mitzscherling, T. Becker: "Analysis and optimization of bakery production lines using arena". International Journal of Simulation Modelling 9(4), 2010, pp. 208-216

[4]   F. Hecker, W. B. Hussein, M. A. Hussein, T. Becker: "Numerical Modelling and optimization of bakery production line with particle swarm optimization". Proceeding of the 10th International Conference on Modeling and Applied Simulation, 12 –14.09.2011, Rome, Italy.

[5]   Hussein, W. B.; Hussein, M.A.; Becker, T.: "Application of audio signal processing in the detection of the Red Palm Weevil". Proceeding of European Signal Processing Conference EUSIPCO2009, Glasgow, Scottland, 24-28.09.2009, pp. 1597-1601

[6]   Hussein, W. B., M. A. Hussein, T. Becker: "Detection of the Red Palm Weevil Using Its Bioacoustics Features." Journal of Bioacoustics 19(3), 2010, pp. 177-194

[7]   Hussein, W. B., M. A. Hussein, T. Becker: " Automatic Recognition of the European Bats by Novel Bioacoustics Features and Artificial Neural Network" (in preparation)

[8]   Hussein, W. B.; Hussein, M.A.; Becker, T.: "Automatic bioacoustics detection of Barbastelle bat (Barbastella barbastellus) by artificial neural network". (in preparation)

[9]   Hussein, W. B., Moaty, A. A, Hussein, M. A., Becker, T.: "A Novel Edge Detection Method Based On Image Energy And Skewness With Application To Intramuscular Fat Recognition". Proceeding of IADIS Multi-Conference on Computer Science and Information Systems 2010, 27-29.07.2010, pp. 93-100. *(with best paper award)*

[10] Hussein, W.B., Sayed, A. A., Hussein, M.A., Becker, T.: "A Novel Edge Detection Method with Application to the Fat Content Prediction in Marbled Meat", Pattern Recognition 44 (12), pp 2959 – 2970, 2011

[11] A. A. Moaty, W. B. Hussein, C.F. H. Longin, C.I. Kling, M. A. Hussein, T. Becker: "Development of Numerical Color Calibration Models and its Applications in Quality Evaluation of Food Products". Journal of Cereal Science (under review)

[12] W. B. Hussein, M. A. Hussein, T. Becker: "Spectrogram enhancement by edge detection approach applied to bioacoustics calls classification", International Journal of signal and image processing 3(2), 2012, pp. 1-20. DOI: 10.5121/sipij.2012.3201

[13] Hoche, S., W. B. Hussein, M. A. Hussein, T. Becker: "Time of Flight Prediction for Fermentation Process in-line Application". Engineering in Life Sciences 11, 2011, DOI: 10.1002/elsc.201000177

[14] Wallhäußer, E., W. B. Hussein, M. A. Hussein, T. Becker: "On the usage of acoustic properties combined with an artificial neural network - a new approach of determining presence of dairy fouling type A", Journal of Food Engineering 103(4), 2011, pp. 449-456

[15] Schirmer, M., W. B. Hussein, M. Jekle, M. A. Hussein, T. Becker: "Effect of humidity on the quality of bread rolls in industrial heating processes", Journal of Food Engineering, 2011

[16] W. B. Hussein, M. A. Hussein, T. Becker: "Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation process". Engineering in Life Sciences 12(6), pp. 603-614, 2012. _(selected for"In this issue" main emphasis)_

[17] H. Elfawakhry, W. B. Hussein, M. A. Hussein, T. Becker: "Automatic determination of density in dough samples using ANN and signal processing features". (in preparation)

[18] Hussein, W. B., H. Elfawakhry, M. A. Hussein, T. Becker: "A Novel Approach to Predict the Ultrasonic Signal Parameters of visco-elastic materials in Time Domain". (in preparation)

[19] D. Krause, W. B. Hussein, M. A. Hussein, T. Becker: "Ultrasonic analysis and prediction of sugar growth in bioprocesses using signal processing and multivariate calibration" (in preparation)

Selected conference papers

[20] Hecker, F.; Hussein, W. B.; Mitzscherling, M.; Becker, T.: "Produktionssimulation zur Optimierung von Produktions- und Logistikprozessen in Backbetrieben". GDL-Kongress Lebensmitteltechnologie, Hamburg, Germany, 11.10.2007

[21] Hecker, F.; Hussein, W. B.; Mitzscherling, M.; Becker, T.: "Simulation der Produktion in einer Bäckerei unter Berücksichtigung des Energiebedarfs". 58 Tagung für Bäckerei-Technologie, Detmold, Germany, 06.11.2007

[22] Mitzscherling, M., F. Hecker, W. B. Hussein, T. Becker: "Simulation der Produktion in einer Bäckerei unter Berücksichtigung des Energiebedarfs". 58th Tagung für Bäckerei-Technologie. Detmold, Germany, 08.11.2007

[23] Hussein, W.; Hecker, F.; Mitzscherling, M., Becker, T.: "Production Line Simulation and Optimization of a Small Bakery". Cereal technology congress, Hamburg, 11-13.10.2007

[24] Hecker, F.; Hussein, W.; Mitzscherling, M.; Becker, T.: "Die Virtuelle Bäckerei". Vereinigung der Backbranche, Leonberg, 11.03.2008

[25] Hecker, F.; Hussein, W. B.; Mitzscherling, M.; Becker, T.: "Die Virtuelle Bäckerei". VDB-Qualitätsgespräch, Großostheim, Germany, 25.04.2009

[26] Hussein, W. B., F. Hecker, M. A. Hussein, T. Becker: "Baking Production Line Optimization Using the Particle Swarm Optimization Technique". ProcessNet-Jahrestagung, 21-23.09.2010, Aachen, Germany

[27] Schirmer, M.; Hussein, W. A.; Jekle, M.; Hussein, M. A.; Becker, T.: "Betrachtung von Strukturbildungsreaktionen bei thermischer Behandlung von Getreideprodukten". 27th Jahrestagung der Biotechnologen, Mannheim, 13.09.2009

[28] Hoche, S., W. B. Hussein, M. A. Hussein, T. Becker: "Time of flight prediction for fermentation process in-line application". 9th Dresdner Sensor-Symposium, Dresden, Germany, 2009

[29] Wallhäußer, E., T. Úbeda, W. B. Hussein, M. A. Hussein, T. Becker: "Acoustic impedance analysis for determining fouling presence and cleaning success of dairy fouling". Fouling and Cleaning in Food Processing Cambridge, UK, 2010

[30] Schirmer, M., Hussein, W. B., Jekle, M., Hussein, M. A., Becker, T.: "Interaction of humidity process and roll crust quality in industrial convection oven". 15th World Congress of Food Science and Technology, Cape Town, South Africa, 22-26.10.2010

[31] Wallhäußer, E., W. B. Hussein, M. A. Hussein, T. Becker: "Usage of ultrasound in detecting dairy fouling on stainless steel surfaces". International Workshop of Online sensors for fouling monitoring (Dechema), Frankfurt, Germany, 2010

[32] W. B. Hussein, M. A. Hussein, T. Becker: "From 1D to 2D acoustic signal pattern recognition techniques for determination and monitoring of bio-mixtures processes". Proceeding of the 8th European Congress of Chemical Engineering, 25-29.09.2011, Berlin, Germany.

Selected non-reviewed papers

[33] Hussein, W. B.; Hecker, F.; Mitzscherling, M.; Becker, T.: "Modelling and Simulation of Bakery Production Lines For Process Analysis and Optimization". Getreidetechnologie 4 (2007), pp. 12-17

[34] Hecker, F.; Hussein, W.; Mitzscherling, M.; Becker T.: "Die virtuelle Bäckerei". Jahrbuch der Bäckereitechnik und -technologie 2008, (2008), pp. 148-153

[35] Hecker, F.; Hussein, W.; Mitzscherling, M.; Becker, T.: "Simulierte Prozesse". Backtechnik-Europe 2 (2008), pp. 6-10

[36] Hecker, F.; Hussein, W. B.; Mitzscherling, M.; Becker, T.: "Die virtuelle Bäckerei". Brot und Backwaren 2 (2008), pp. 15-17

[37] Hussein, W. B.; Hecker, F.; Mitzscherling, M.; Becker, T.: "On Simulation and Optimization of Bakery Processes". Getreidetechnologie Heft 2 (2008), pp. 105-110