Methods of Bacteria Recognition Relying on Simple Hardware Techniques

Alexander Chamski

Sheffield Hallam University

This dissertation is submitted for the degree of Master of Philosophy

January 2011

Contents

Acknowledgements

With gratitude to my aunt Ingrid, to my parents, and to Professor Marcos Rodrigues.

Abstract

Bacterial contamination puts the public at risk and is costly for the food-processing industry. Traditional (biochemical) methods of bacteria recognition require complicated sample preparation for reliable results. Automated technologies exist for the identification of bacterial cells in suspension, but are relatively expensive with only limited success. Therefore, an early warning system that could be applied with little effort and expenditure to give an indication of whether or not more in-depth analytical procedures would be commendable has a high potential on the market. The work presented here demonstrates two methods utilizing flexible and low-cost equipment together with pattern-recognition techniques to form a first-stage bacteria recognition system. Bacterial colonies are excited with laser light and electromagnetic power and their actions are recorded with simple optical sensors. The generated data are the basis for pattern generation algorithms and are evaluated statistically and with Fourier and Principal Component Analysis methods. Focusing on three bacteria species, namely *Escherichia coli*, *Proteus mirabilis*, and *Bacillus subtilis*, the two systems as described here distinguish the species and indicate typical classes to provide the user with a first impression on the sample content.

1 Introduction

1.1 Problem Definition

Principally everybody, from individuals to members of certain business branches, legal authorities, and scientific organizations, is actively or passively interested in - or will at least profit from - the ability to detect and identify microbial organisms in water or food.

The most trivial example is probably that of individuals who want to make sure that what they eat or drink is free from bacteria which may cause illnesses or are, in the worst case, life threatening. In the case of potable water, for example, the water industry is legally responsible to ensure a certain degree of purity (see Section 2.2.2), while breweries need to monitor the bacterial contents within the beer during the production process (see Section 2.2.3). Health laboratories are concerned with the rapid identification of microorganisms in order to save people's lives and to prevent the spreading of epidemics (see Section 2.2.1), and scientists are investigating samples taken from all parts of the world (and even from worlds in outer space!) in order to detect and examine microorganisms (see Section 2.2.4).

The common aim of scientists around the world, in our case, is to create a largely automated identification system that has the potential to speed up the detection and identification process, however, without sacrificing reliability and precision of the analysis. Not only would such a system be expected to result in a faster and more efficient analysis of microbial activity, but since it lends itself to an automated procedure which would be controlled by a computer with only minimal manual interaction, it would help in reducing the cost of an analysis by abolishing lengthy and labour-intensive procedures. Additionally, by minimizing manual interaction, it would help reducing the probability of human errors during the analysis process. In principal, an ideal bacteria identification system would operate fast and efficiently, offer reliable and precise analytic results, and - last but not least - be as cheap and simple to use and maintain as possible.

Spinning radical future prospects, a robust device made of relatively cheap hardware with a simple (e.g. personal computer) interface, could even have the potential to be made available to the public sector, where it could be used to test the water quality of private households. However, today's semi-automated methods are quite far from achieving immediate and accurate prognoses about the bacteria content in food or water samples.

1.2 Scope of the Dissertation

1.2.1 The Cornerstones of a Classification System: Data Collection, Pattern Generation, Pattern Recognition

A computer system to classify microorganisms would be regarded as an 'artificially intelligent' system (see Figure 1.1). At the heart of such a *classification system* lies a series of pattern recognition algorithms to analyze data that have been recorded by the system while monitoring the environment, and within it the microorganisms to be identified. The cleverness or power of these algorithms is hence the degree to which they discover and process patterns from within the available observation data (pattern generation), and to relate the processed patterns to particular bacteria geniuses or species (pattern recognition): Clearly, the higher the level of classification the more intelligent - and probably useful - the system will be regarded as.

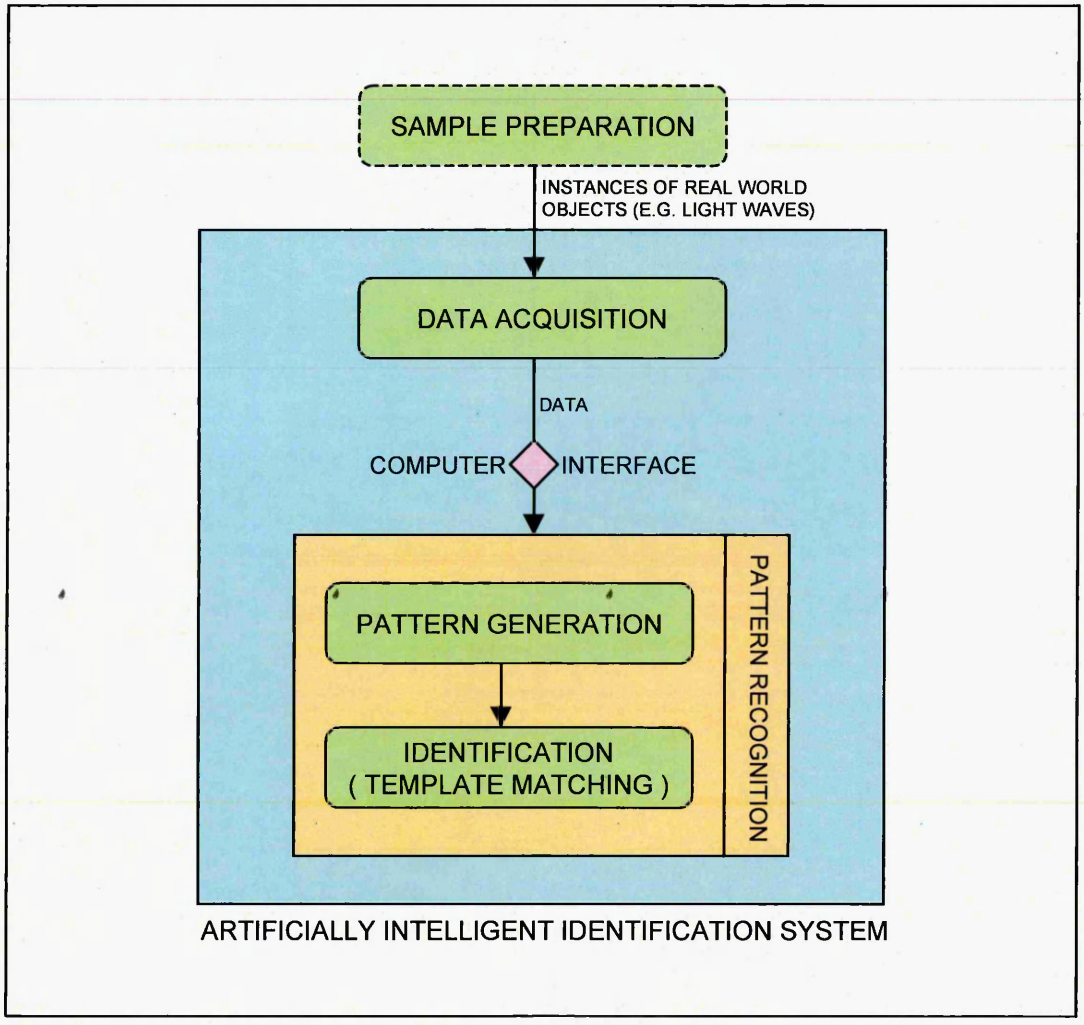

Fig. 1.1 Entities constituting a classification system's high level working principle

However, being able to successfully relate a pattern to the occurrence of particular types of bacteria is only the second intelligent objective of a classification system. Prior to the classification task, those patterns used for the classification have to be generated from the observation data stream. Generating meaningful patterns from the vast amount of recorded data can make or break a successful classification system: The more meaningful a pattern is the more it aids the next step of the system to match it to characteristically templates in order to classify the real-world source of the pattern.

The task to efficiently generate patterns relies to a considerable extent on the technique by which the raw data are physically collected from the observation environment. Various devices with 'artificial eyes' or sensors exist to monitor the environment, the most straight forward probably being the optical camera. In addition, a wide range of sensor techniques exist: Measuring biochemical conditions of the observation environment, energetic exposition, optical or mass spectroscopy, and even 'artificial ears' to listen to the objects to be sensed. The data acquired by the various methods differ in the amount collected and their complexity. By tuning the nature of the data collection to the way the pattern generation algorithm works, a suitable observation technique is chosen. Also, a combination of the data of multiple acquisition techniques may be useful, depending on how the pattern generation algorithm of the system works.

To summarize, an artificially intelligent computer system for the classification of microorganisms consists of three high-level sections: a) the application of a physical technique to record observation data, b) the extraction or generation of patterns from the data, and c) the analysis of the patterns to classify the related microorganisms.

1.2.2 The Sensibility and the Sensibleness of a Classification System

Most if not all identification or classification techniques, even so their technical working principles may be very different, generally consist of three high level stages. By applying a technique, the system may manipulate bacteria, then senses their behaviour or their reaction to the manipulation, and finally analyzes the data recorded by the sensors to make an assumption about the bacteria content (see Figure 1.2). The more information about the bacteria is encoded in the recorded data, the more reliable will the system's assumption be regarding the identity of the monitored bacteria.

It is questionable, however, whether an automated system could be invented which would be capable of undoubtely identifying particular types of bacteria. With thousands of different genuses and species of bacteria existing, many of them inhibit the same or similar properties as others. In order to make a sound assumption about the identity of bacterial cells, their behaviour has to be very precisely observed.

A clear identification of the bacteria under observation, however, might not be required. For many purposes it may be sufficient to classify certain groups of bacteria with similar properties. For example, instead of distinguishing one strain of the Escherichia Coli family from other Escherichia Coli members, as well as from other bacteria and their sub-types, it might well be sufficient for a system to distinguish any type of coli bacteria from all other major bacterial genuses. Thus, the requirements for the sensors and the analysis would be relaxed and the system's reliability could be enhanced. Classifying groups of bacteria rather than identifying cells e.g. on the species level brings up the question to what extent the various forms of bacteria can be sensibly grouped or

clustered; many small groupings or classes with rigid property constraints are as well possible as larger groups incorporating more diverse types of bacteria, therefore abiding looser property constraints. Alas, the way in which bacteria classes are generated could also be a direct result of the system's limitations to sense additional information required for a better ability to distinguish some bacteria from others.

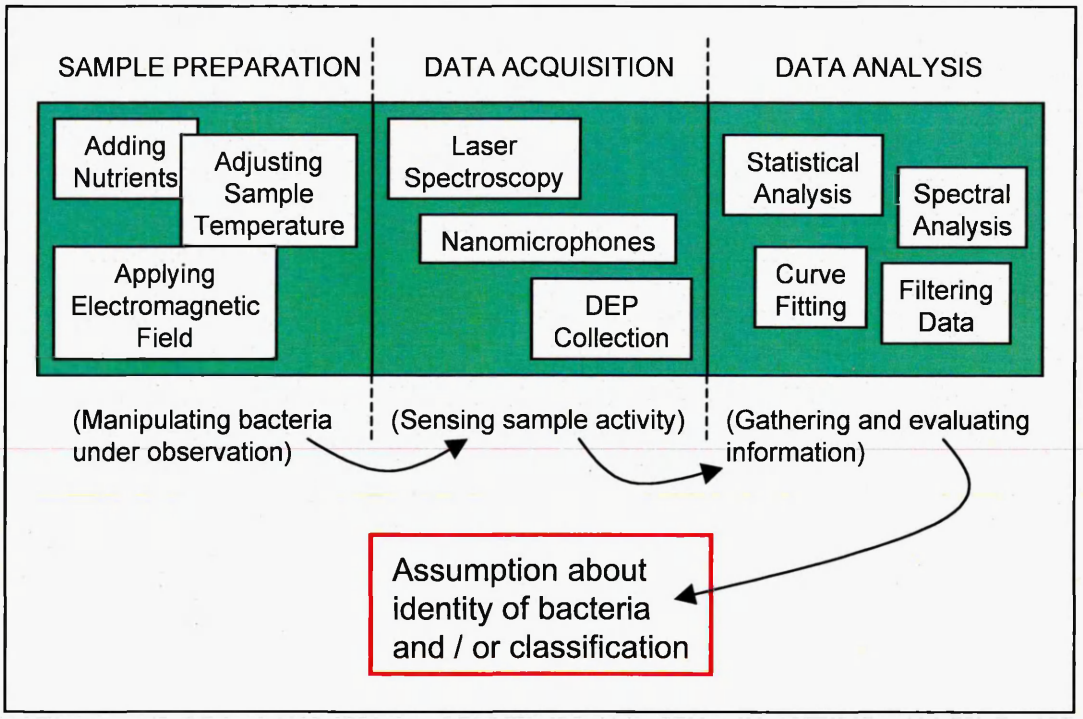

Fig. 1.2 The three stages and their methods constituting a classification system's high level working principle

1.2.3 Practical Experiments vs Model Systems

Greve and Puppels (1993) point out that many biological processes are very complicated and can hardly be mimicked in model systems. Accordingly, the goal of biophysical research - obtaining an understanding of how a biological system functions - requires experiments in which processes like cell division, cell-cell interactions, etc., are studied Following this theory, two practical experimental approaches on how to distinguish between certain types of bacteria comprise the main part of this study: 1) In applying an optical sensory device correlating with a laser scattering technique certain micro-biological species are to be distinguished (see Chapter 3), and 2) using an electro-magnetic method together with image analysis hard- and software bacterial strains from the species 'Escherichia coli' are to be identified (see Chapter 4).

1.3 Aim and Objectives

The aim of this research is trying to solve the question whether or not, or to what extent, a system can be created to utilize (relatively) simple and straight forward technological methods in order to detect and to distinguish particular aquatic microorganisms from others, i.e. to classify microorganisms. Since the methods are simplistic, the required hardware is readily available on the market at low cost.

Prior to this research, such hardware had been assembled to produce two microbiological devices. These are used in conjunction with information technology, in particular with the application of pattern generation and recognition techniques, to produce two methods (or systems) which are separately explored in the course of the research: The data generated by these systems are retrieved and analyzed by computational algorithms in an attempt to transform the data into scientifically useful information in order to accomplish the classification task largely *automatically*.

The objectives are identified as follows:

- Investigation of morphological, physiological and biochemical characteristics of bacteria
- Investigation of light theory, light source emitters and receivers, and assessment of current spectroscopic methods and laser scattering technology
- Investigation of electromagnetic power in view of excitability of bacteria and assessment of current electrophoretic technology
- Functional analysis of existing system hardware with regard to measurable parameters like intensity of scattered light, patterns of scattering, motion detection, growth detection, statistical patterns
- Design of experiments in order to result in the collection of data from a variety of microorganisms
- Design, implementation and testing of two classification systems using a) optical and b) electromagnetic excitation of bacteria
- Investigation of sensible data transformations for classification procedures
- Analysis and interpretation of results with indication of areas for future research

1.4 Methodology

During this work the application of a quantitative research technique leads to exploratory data analysis, while becoming acquainted with microorganisms and their behaviour, designing microbiological experiments, and designing pattern generation and recognition algorithms.

The following sections describe the high level concept, or methodology, for an automated system to classify bacteria, which serves as the framework for the (two) diverse classification approaches outlined in Chapters 3 and 4 during the course of this document.

1.4.1 Generating a Knowledge Database

Prior to the system's operation, a series of initial experiments is carried out, each of which analyzes a sample containing a particular number of bacteria of a particular type. Thus, characteristic data for a range of different bacteria types are obtained. These data, along with their respective system parameters (see Section 1.4.3), also referred to as training set, are stored in a database and serve as reference data for the system.

1.4.2 System Operation

Once the data for the 'a priori' knowledge database have been collected, a test experiment is carried out, similar to the initial experiments. In a test experiment, a sample with an unknown concentration of unknown particles and cells is analyzed. The resulting characteristics which are obtained in the course of the experiment are used as input data to the system and enables the system to start its operation.

The system's task is to determine whether the characteristics of the data gained from the sample of the test experiment indicate the occurrence of bacteria cells. In this case, the system tries to predict the type of bacteria by using a set of algorithms to compare the data of the test experiment with the reference data of the initial (together with the already successfully classified) experiments stored in the knowledge database (see Figure 1.3).

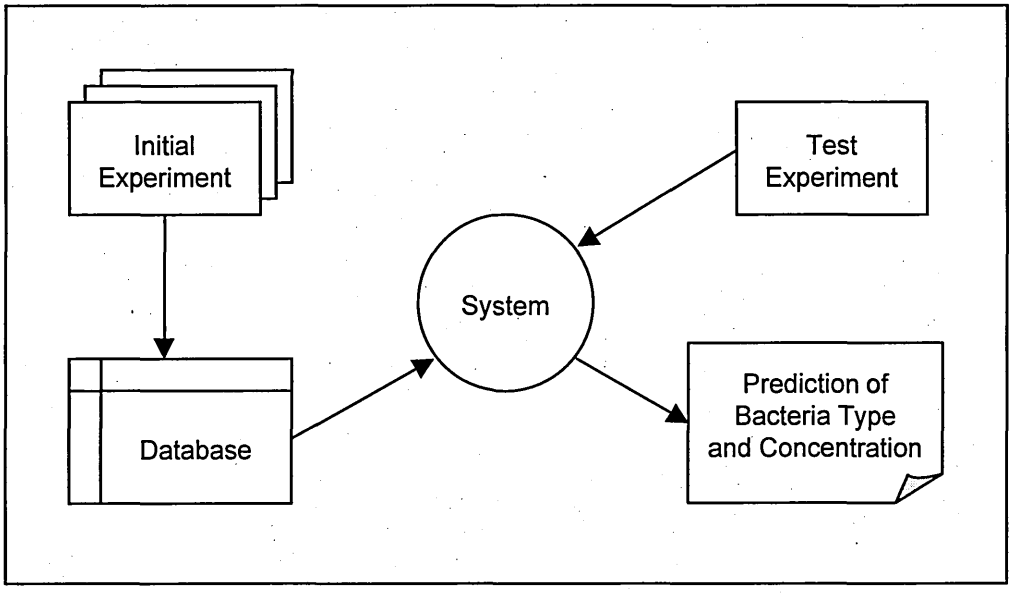

Fig. 1.3 Context Diagram of a Bacteria Classification System

After a successful classification of the bacteria being identified in a test experiment, the data and the classification results, possibly together with key qualifiers to be able to determine how the classification process had been carried out, are being incorporated into the system's database to dynamically update the 'knowledge' of the system, and to adapt it to constantly changing environmental and experimental factors. Thus, the scope of the knowledge database increases with time – the system gains experience.

1.4.3 System Parameters

The system makes use of a number of parameters which are set either by man, i.e. the experimental operator (before each experiment), or by the system itself (in the course of an experiment) to govern the classification process in an optimal way. Static parameters remain constant for each experiment and may include suspension properties, experimental hardware setup, or rules for conducting the experimental process. Dynamic parameters are established during the experimental analysis according to the sample contents and are derived from the collected data, resulting in signals, spectra, or any kind of patterns.

1.4.4 Qualitative and Quantitative Detail

The detail of each experiment and its classification procedure, on which the system bases its judgements about the sample contents, is decisive for a critical analysis. Accordingly, the greater the number of system parameters, and the higher the accuracy with which the values of these parameters are determined, the higher the qualitative detail of the system's operation will be. On a different scale the sheer number of experiments in the database plays a role: the more experiments are carried out, or in other words the more compound the knowledge database, the higher the quantitative detail of the system will be.

1.4.5 Tracing the System's Operation: The Algorithms

The heart of the system is a selection of algorithms which operate on the data in the knowledge database and on the system parameters. It is their very construction which dictates the level of 'cleverness' of the system. The task of an algorithm is to make decisions and draw conclusions about the problem at hand. Every major problem is split into numerous sub-problems which are to be tackled on a step-by-step basis. A sound system demands individual solutions to every small problem that occurs during the step-by-step realization of a high level master plan.

When software specialists start to implement solutions to given problems, they will almost certainly go through a lot of hoops which at first nobody would have thought to encounter: At the start of a project the final objective and major milestones are identified, and a high level roadmap is clarifying the path along which they have to be achieved. But only on the way one will figure out what has to be done in detail in order to get ahead. Individual solutions to the smallest problems pave the way to the overall goal, and the concatenation of these single steps is the advent of an algorithm.

1.5 Structure of the Dissertation

Chapter 1 defines the core problem and issues that this research is concerned with and presents a high level concept of how these issues are to be tackled. After discussing the scope of the dissertation it follows the assessment of the aim and objectives and methodology in a standardized fashion.

Chapter 2 serves as a general introduction to the topic. It wants to make the reader acquainted with the objects of interest, the bacteria, and gives an overview about their characteristics which are relevant within the scope of this work. Further, it lets the novice view different dimensions of the topic, aiming to answer the obvious questions as to who does what, why and how, before one is taken into depth by more detailed and specific investigations in the subsequent chapters.

Chapter 3 comprises the presentation of a low-cost, straight forward laser scattering system, and gives the reader an understanding of the principal features of optical spectroscopy and the way it works. The system consists of a laser beaming through a transparent container (cuvette), where bacteria are floating and metabolizing within an aquatic environment. In an attempt to distinguish three different bacteria species, the reflected light is collected by optical sensors, and the varying intensities due to the motion of the cells and particles are recorded and digitally converted to be analyzed by computer algorithms.

Chapter 4 examines the capabilities of 'dielectrophoresis' (DEP) in order to help the detection and classification of bacteria within liquid samples. This technique presents a formidable opportunity to filter bacteria and thus to raise the concentration of cells at a 'spot' where a monitoring system would look closer to enhance the resolution of the collected data and to reduce the time of an analysis. An algorithm to classify various strains of 'Escherichia coli' bacteria is developed using this method.

Chapter 5 restates the results of this work. The difficulty of the problem, the experiments, and of evaluating the results is reiterated. The achievements and limitations of the methods used are discussed.

Chapter 6 raises issues for improvement and future research. Hardware and method alterations are outlined, and the theoretical issue of joining the two techniques of chapters 3 and 4 is proposed. The question is in how far it would improve the capabilities of a classification system if both DEP and laser scattering methods are combined.

2 Elements of Bacteria Recognition

2.1 General Introduction to the Identification of Bacteria

Being neither animals, nor plants, according to Darwin's 'On the Origin of Species' (Darwin 1859), the world of bacteria is categorized as part of a 'third kingdom', serving as the transition form between the plant and animal kingdoms. The so-called 'Protista', as Haeckel (1894) terms this kingdom of microorganisms, are divided into two divisions: The eucaryotes and the procaryotes. Eucaryotes include protozoa, fungi, and most algae, while all bacteria, together with the blue-green algae, form the group of procaryotes. In the following, characteristics of the latter are discussed by applying morphological, physiological, and biochemical viewpoints.

As a result of the optical nature of the technique described in this document (i.e. due to the focus on light spectroscopy as the means to acquire data), its classification procedures may depend to an extent on the determination of factors such as shape, size, and structure of the examined bacteria. In addition to the morphological characteristics of single isolated cells, diverse spatial arrangements of whole groups of bacteria, i.e. the forming of clusters, chains, or pairs by a number of individual cells, also represent visible characteristics which might aid in the identification of the organisms. Thus, the morphology of bacteria, in theory, serves as a major source to obtain features for the classification process.

In order to generate patterns resulting from light reflected by bacteria colonies, often even more meaningful than morphological characteristics, another indicator to the task of classifying the various forms of bacteria may be their diverse motion. Bacteria are able to move by means of a structure called a flagellum (plural, flagella). The rotation of flagella propels the cell through liquids (Madigan et al. 2000).

In contrast to morphological and physiological properties of bacteria, the manifestation of biochemical activities of microorganisms are directly related to the organisms' metabolism. Biochemical transformations are produced by bacteria due to the enzyme-catalyzed reactions occuring within the organism: During growth, enzymes of the organism are frequently transforming organic (nutrient) molecules in order to produce energy. In other words, the uptake, breaking down and use of organic molecules, and the excretion of by-products, are the principles according to which bacteria cause biochemical transformations (Mitruka 1976). These, at first glance, do not seem to produce any visible patterns. But in fact, a majority of today's proven techniques, e.g. the use of indicator dyes, gas chromatography plus mass spectrometry, or impedance measurements (see Section 2.4), do visualize biochemical states or transformations by monitoring end products or cell components of bacteria.

2.2 Main Players on the Market

2.2.1 Health Organizations

In October 2001, several Anthrax contaminated letters reached their recipients working for US governmental authorities and news agencies. Those were the first of an assassination series with anthrax in the USA in the coming months, and several citizens had to die in its wake. Would a speedier way to identify the bacillus have saved some lives? At least, a rapid and secure way to analyze the bacteria contents of the received items which the affected people came in contact with would have taken the haunt of uncertainty off their minds – hopefully as quickly as possible. However, lacking any promising automated methods, only rudimentary checks can be performed in the clinics, and samples have to be sent away to health laboratories for a close examination. There, the routines applicable to detect anthrax are lengthy and usually last another couple of days (Cheun et al. 2001, AAVLD 2002).

2.2.2 Water Industry

Water quality is an ever-present concern of drinking water utilities and the public health community as contaminated water is known to be a significant source of infection (Edberg and Melson 2001). This has been evidenced by large waterborne outbreaks of bacteria (e.g. Pontius 1993). For example in the 1990s, in just two years, 17 states and territories in the USA reported 34 outbreaks of disease associated with drinking water, and these affected over 17000 people (Moore et al. 1994). In view of such (ever recurring) problems, drinking water utilities would benefit if the water industry had the means to apply appropriate methods to monitor bacteria throughout the public water supply systems.

In many countries, over the past decades, regulations have been continuosly developed as to how drinking water needs to be monitored in order to guarantee a certain level of purity. In the USA, the Safe Drinking Water Act of 1974 was a result of a Community Water Supply Study which generated interest in federal safe drinking water legislation. The 1974 Act specified the process by which national drinking water regulations were to be adopted, and since then it has been subject to many stepwise improvements (Pontius 2003). Nevertheless, the detection of coliforms remains to be a problem which the water industry has to overcome: According to the Massachusetts Water Resource Authority 53 of 56 sampling locations violated governmental regulations in 1996 (Edberg and Melson 2001).

2.2.3 Beer Breweries

Modern large-scale breweries are looking to achieve very high hygienic standards by trying to effectively monitor microbial activities during the production process The brewing process itself is a long production run from water boiling to beer packaging and is at every stage prone to the growth of microorganisms. Fermenting processes of up to several weeks give plenty of time and opportunities for unwanted microorganisms to develop. In the case of microbiological spoilage of the brews an entire batch ranging from 200,000 to 500,000 litres would have to be discarded, posing a serious economical and environmental problem to the brewery. If the image of a beer was to suffer because of quality losses due to microbiological problems in the beer production process, it would

be a disaster to any beer brewing company that needs to be avoided at all cost (Storgards 2001).

2.2.4 Scientific Organizations

A very well known organization in this field is the North Atlantic Space Agency (NASA), which one would most probably relate to interstellar research rather than to the world of bacteria recognition. However, NASA's interest in the detection and recognition of bacteria is multifaceted: For example, the health of astronauts in space is maintained by preventing the breeding of bacteria within recycled air and recycled drinking water supplying the International Space Station and space vehicles. But, NASA's scientists do not only search for ways of how harmful bacteria could be detected and rendered harmless, they are also developing ways in which "good" bacteria can be used to detect "bad" bacteria to assist in keeping astronauts healthier and safer in space (Marconi 2007).

Another one of NASA's projects relating to the detection of microorganisms is the use of 'artificial ears' in an attempt to reconstruct nature's acoustic systems with state-of-the-art technology: The development of acoustic sensor technology based on artificial stereocilia could result in a "nanostethoscope", a device to probe nano/micro-scale biological activity by "listening to the music of life". Thus, the quality of water can be tested by detecting microflows generated by bacterial motion. The instrument can also be used to search for signatures of extant life in the form of microorganisms on other planets (Noca et al. 2000).

2.3 General Problems of Bacteria Recognition

The classification of bacteria is historically governed by rules having been devised by zoologists and botanists over the last centuries. They divided bacteria up into genuses and species in order to suggest something about the general characteristics of particular organisms (Doetsch and Cook 1973). In the absence of technologically more adequate and precise, phylogenetic studies of bacteria in sufficient numbers, this rather vague classification framework largely based on morphological, physiological and biochemical characteristics still exists to date.

2.3.1 Problems of the morphological approach

By analyzing their shape or visual appearing (i.e. their morphology), bacteria are being identified using classes like 'cocci', 'cylindrical', or 'spiral' (etc.). However, their simplicity makes it difficult for an anatomical analysis to result in prominent and stable differences between bacterial species. For example, two genetically different kinds of species might look very similar under the microscope, whereas, on the other hand, two genetically related species might differ to an extent where they seem to belong to seperate classes (Doetsch and Cook 1973). This is why usually a physiological or biochemical analysis is preferred method of identification (see next Section).

2.3.2 Problems of the physiological and biochemical approach

In theory, physiological and especially biochemical diversities of bacteria reveal much about their affinity to a particular (existing) class. However, it plays a role whether these activities are being discovered for microorganisms grown in a laboratory or whether they are found in the natural environment (see Section 2.3.3). Biochemical tests can be applied to successfully identify species as being used in the scientific laboratory environment, but there are simply too many species or even strains of species existing in natural habitats to be coverable. Nevertheless, biochemical tests still form the backbone of today's standard procedures to identify certain bacterial species (see Section 2.4.4).

2.3.3 In vivo and in vitro: microbial ecology and laboratory cultures

In real world scenarios it would be appreciable if remedies against bacterial infections or, equally, remedies due to bacterial activities would be developed with bacteria cultures as found in the natural environment, i.e. 'in vivo' cultures. However, scientific research is more often based on bacteria species cultured in laboratories, i.e. 'in vitro'. In this context, Doetsch and Cook (1973) point out that the life of bacteria in the laboratory is not equivalent to the life of bacteria in nature, which is further supported by Madigan et al. (2000), who state that laboratory cultures of bacteria are rather artificial situations, which lack natural influences and therefore promote less natural metabolic activities. Alas, for the purpose of classifying species, as presented by this research, the microorganisms in question are entirely laboratory, in vitro cultures.

2.4 Current Proven Methods to Detect and Identify Bacteria

Ever since the late 19th century, scientists have been relying on manual methods in order to investigate the bacteria contents in samples for various reasons. The avantgarde in

bacteriology like the famous Robert Koch (1843 – 1910), Martinus Beijerinck (1851 – 1931), or Sergei Winogradsky (1856 – 1953), to name only three, pioneered in developing techniques to isolate and study bacteria, in establishing taxonomy to describe the world of microorganisms.

Since the late 1960s and 1970s, scientists have been trying to use computers in order to aid in the detection and characterisation of bacteria. Since then, by constantly improving these methods and introducing new promising automation techniques, procedures to detect or analyze bacteria have been accelerated or made easier to be carried out.

In order to assess the potential for an improvement of existing detection and identification methods, some of the standard techniques which are commonly applied by today's bacteria analysts are outlined in the following sub-sections. All of these techniques are heavily relying on manual interactions.

2.4.1 Incubation and Isolation

Bacteria can be detected by attempting to incubate a sample of the object suspected to be contaminated in specific culture media and under specific incubation conditions. This means that small numbers of existing bacteria are encouraged to multiply in favourable conditions. By selecting several specific culture media, usually broth (fluid) or agar (solid), for every tested sample respectively, the growth of only a single type of organism is favoured while the growth of others within the same incubation environment is constrained. This leads to isolated bacteria cultures which can hence be easily detected and further analyzed.

2.4.2 Total Cell Count

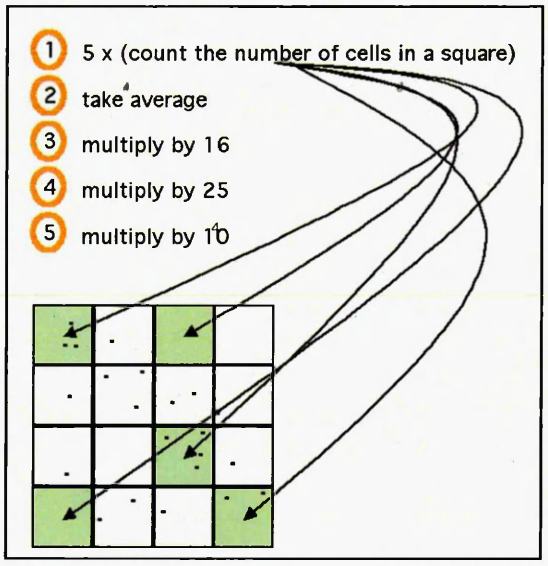

Fig. 2.1 The method by which a counting chamber (haemocytometer) estimates bacterial cells

A method to estimate the amount of bacterial cells within a sample is to use a microscope to manually count cells within discrete small volumes of a haemocytometer, otherwise known as 'counting chamber'. A counting chamber is a glass slide on which surface a grid of numerous precisely measured squares is embossed. The 'chamber' becomes a chamber when a coverslip is slit on top, so that in between the two surfaces and within the marked squares cells can be observed. Then, by using mathematical projections, the total number of cells to be found within the original sample can be estimated (see Figure 2.1).
The drawback of this method is that, in order to be able to properly count the cells, the total number of cells in the sample has to come to a given high concentration per ml, otherwise the view through the microscope lens may reveal too many individual cells to be counted – moreover as the cells are constantly floating in and out of the field

of vision, even if they are non-motile due to brownian motion – or too few to be spotted at all.

2.4.3 Viable Count

Another method to estimate the number of cells within a sample is called 'plate count' or 'colony count'. Its feature is that it only counts alive cells, which is in many cases desirably. Compared with the direct microscopic observation it has the advantage that it allows a wider range of either high or low concentrations of cells per ml in the sample to be counted. Altogether this makes it probably the most widely accepted and standardized method to detect and enumerate bacteria.

In the process, a series of dilutions of the bacteria sample is set up, which is then incubated over time on an equal series of agar plates to reveal the forming of bacteria colonies. A colony can be spotted without the use of a microscope and admits the conclusion that a single bacterium has been its origin. Thus, following the guidelines of the procedure, one is able to estimate the numbers of bacteria in the original sample by counting the colonies which have been forming on the agar plate which is the most suitable one for the counting process.

2.4.4 Monitoring Biochemical Transformations with the Help of Indicator Dyes

Biochemical transformations of bacteria are so numerous and diverse that their examination is today's preferred method to identify a particular class of microorganisms.

In order to test the biochemical activities of bacteria, a carrier medium may be enriched with organic preparations to which the bacteria react. The manifestation of biochemical reactions of bacteria, induce the production of gas and/or acid in the carrier medium. Most common tests observe the organisms' ability to ferment sugars, e.g. glucose, lactose, or mannitol. Other tests measure the organisms' enzyme activities (such as catalase, oxidase, coagulase, DNAse, urease, gelatinase, and many other enzyme activity tests), or measure various biochemical transformations (such as citrate utilization, malonate utilization, hydrogen sulfide production, or nitrate reduction) (Mitruka 1976). These processes can be made visible by adding an indicator dye to the medium, which alters its colour while the acid and/or gas content varies. According to the type of test, sometimes the medium is coloured, while sometimes the bacteria cells themselves are stained, consequenty indicating the presence or absence of a particular class of microorganism. Hundreds of these biochemical tests have been developed for clinical use, but only about 20 are used routinely (Madigan et al. 2000).

Thus, according to the respective class of microorganisms to be identified, different indicators have to be chosen, which may sometimes result in very specific tests rather than in a more universal identification method. The time span for the tests to indicate results ranges from hours to days. Also it should be noted that, due to the nature of most biochemical tests, demanding several sequential action-reaction steps, these and similar tests involve a considerable amount of manual interaction and provide little means for an increased automation.

2.4.5 Gas Chromatography and Mass Spectrometry (GC/MS)

For many decades Gas chromatography (GC) has been a very popular analytical tool for the seperation of microscopical quantities of matter in a mixture of compounds and in its combination with mass spectrometry (MS) it has become a versatile technique for the detection and identification of microorganisms. However, its greatest weakness is its requirement for volatile compounds, which means that bacterial cells have to undergo special preparations prior to the analysis and are killed in the process (McMaster and McMaster 1998).

A system called 'matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry' is an example of how GC/MS can be used for the identification of microorganisms: The MALDI-TOF system uses a laser to dismember a sample held in a vaccum chamber into its basic components by accelerating fragments of the original sample onto a microchannel plate. Here, the mass of the particles is calculated from their flight time to the detector. However, due to the necessity of preparing the organisms by adding chemicals and/or breaking them apart to expose signature molecules for detection, such experiments make real-time and in situ analysis difficult (Legett 2001).

2.4.6 Impedance Measurements

Impedance Measurements have made the detection and enumerating of bacteria cells in samples easier as opposed to the universal manual laboratory methods. In the 1970s, scientists found out that during the growth of microorganisms, due to their metabolismic activity and depending on the frequency applied during the measurement process, the carrier medium's impedance values decrease significantly (Ur and Brown 1975). This was the starting signal for microbiologists to exploit this indicator in view of inventing more efficient methods for the detection and identification of bacteria.

Electrical impedance has been proposed as a convenient alternative to other (mainly optical) methods of bacteria classification (Fistenberg-Eden and Eden 1984, Felice et al. 1988). One advantage is that by measuring the impedance of a medium in which bacteria are reproducing optical clarity is not required. Also, the measurements can be readily automated and little or no sample preparation is required (Cady et al. 1978).

3 Exciting Bacteria with Laser Light

Along with the technological development of light emitting devices, especially with the advent of the laser in 1960, which provided a dramatic increase in photon density in a scattering volume and hence greatly improved the generation of spectral patterns (Nibler and Pubanz 1988), optical spectroscopy has evolved over the past decades in uncountable variations of applications, instrumentation, and measurement techniques. Their difference constitutes the way how to excite or illuminate samples by using diverse technological

equipment to produce and propagate light at various wavelengths, and by establishing several methods as to how the emitted light after it has excited the sample is sensed and recorded, and is henceforth processed. In this chapter, the most common methods (see Section 3.2) and their hardware (see Section 3.1) are presented, before the Optical Biosensor system is to be introduced (see Section 3.3).

3.1 Hardware used with Optical Spectroscopy

3.1.1 Light Emitting Devices

The following paragraphs constitute an overview of those devices which serve as the light source in spectroscopic experiments to excite samples so they reflect, transmit, scatter, or absorb light; these reactions can then be sensed and used for the analysis. The list of devices is not exhaustive, but lets the reader gain an insight into the technology.

3.1.1.1 Photo-optic Lamps

Although laser devices comprise the vast majority of light emitting sources used in optical spectroscopy, some applications make use of lamps, sometimes in combination with laser light, to illuminate samples. Typically photo-optic devices include halogen cycle lamps (with or without reflectors), which operate from 15 W to 250 W, and mercury short-arc lamps, operating from 50 W to 200 W with a wavelength bandwidth from 300 nm to 4.8 μm (Gilway Technical Lamp 2003).

3.1.1.2 Helium-Neon Laser

The 'Volkswagen' amongst the choice of laser technology is the Helium-Neon (HeNe) laser. The most common HeNe laser is a sealed plasma tube with internal mirrors using a high voltage power supply, external mirror HeNe laboratory lasers are also available but are more expensive. Components for HeNe lasers are widely available, they are relatively inexpensive, robust, and have a long live. The output is well collimated without external optics and has excellent coherence length (10 cm to several meters or more) and monochromicity, resulting in an excellent beam quality. Although simple in principle, HeNe lasers are complex to manufacture, consequently they are most widely used for scientific and industrial purposes. A small or medium size HeNe device emits light at around 633 nm and has an output power of 0.5 to 5 mW. Green (543.5 nm), yellow (594.1 nm), and orange (611.9 nm) HeNe lasers are also available but render themselves not nearly as 'efficient' as the common red type. Thus, 'other color' HeNe lasers must be much larger for the same output power and use higher quality mirrors. In the IR range, a HeNe laser commonly radiates at 1152 nm, 1523 nm, or 3391 nm (Goldwasser 2006).

3.1.1.3 Argon/Krypton Ion Laser

The argon/krypton (Ar/Kr) ion laser may be regarded as the superlative amongst the available laser devices, expressed in electrical and handling requirements, output power, beam quality, and its cost. High output power means the devices have to be air- or water-cooled, which may result in vibrational problems and bulky apparatures. The common

wavelengths for krypton ion lasers are 521 and 532 nm (green), 568 nm (yellow), and 647 nm (red). Other wavelengths throughout the visible spectrum and beyond are available but generally have weaker outputs (Goldwasser 2006).

3.1.1.4 Diode Laser

Diode lasers share basic material and manufacturing techniques with semiconductor chips, and this booming industry has made them efficient, small, and inexpensive. However, compared to HeNe, Ar/Kr ion, or other common gas lasers, the output beam of a diode laser may have some undesirable characteristics, as it suffers from two asymmetries: astigmatism and an elliptical beam profile. In an attempt to focus a laser diode, the result will be elongated, and due to slight differences in manufacture, the precise angles will not be exactly the same even for various samples from the same batch of laser diodes (Goldwasser 2006). Another drawback is that for many medical procedures and biotechnology applications the power of diode lasers may be too low, most commonly from 0.1 mW to 5 mW (Hogan 2002a). On the other hand, making them very attractive for many applications, diode lasers are very compact and merely require low voltage input power supply; the variety of available wavelengths includes UV, violet, blue, green, red, and many wavelengths in the IR range.

3.1.1.5 Solid-State Laser

The original laser invented in 1960 was a solid state laser; it produced an intense pulse of coherent red light at 694.3 nm. The solid-state laser uses a solid crystalline material as the lasing medium and is usually optically pumped. Solid-state lasers are different from semiconductor or diode lasers, which are in fact also 'solid state', but are almost always electrically pumped (Goldwasser 2006). Compared to Argon/Krypton Ion or Helium Neon lasers, the advantages of solid-state lasers may be higher reliability, durability, and portability, less maintenance, cooling, and lower electrical power requirements. Which makes them the only solution in many applications.

3.1.2 Light sensing devices

Light which has been reflected or scattered by (or has transmitted through) a sample can be analyzed after it has been sensed by suitable detector devices. Such devices transform the optical energy, the photons, into an electric current, which can then be recorded and displayed as an analog signal or is converted into a digital representation for computerized analysis. Three common types of detectors are being used in bioscience research, which are presented in the following chapters: The photodiode, the photomultiplier tube (PMT), and the charged coupled device (CCD).

3.1.2.1 Photodiode

Based entirely on semiconductors, photodiodes are being produced very cost-effectively. They can efficiently count single photons over a wavelength range from 400 nm to 1100 nm, detecting about 70 percent of incoming photon barrels. A photodiode traps incoming light and converts it into free electrons, which then multiply and lead to an amplified signal (Hogan 2002b). Thus, much more sensitive than a camera, photodiodes can sense very weak signals at specific wavelengths. However, in contrast to a camera, the active detector area of a photodiode is for technological reasons comparatively small, it typically measures only about 200 µm in diameter. In order to capture a more prominent signal, photodiodes can be combined into an array or matrix, where each diode is dedicated to measuring a finite but narrow band of the spectrum, and the combined signal output of all diodes comprises the variation in light intensity over the entire wavelength range (Poole and Kalnenieks 2000).

3.1.2.2 Photomultiplier Tube

Photomultiplier tubes are among the most common non-camera photon-detecting technologies. Their working principle is to let photons collide with a photocathode which results in fast moving electrons being propelled onto specially coated electrodes, so-called 'dynodes', in order to set off a chain reaction which multiplies the photon-induced current up to a million times. Thus photomultiplier tubes can detect very low light levels. They are available in the range from 200 to 600 nm, with maximum sensitivity obtained in the 300 - 500 nm range. Red-sensitive photomultiplier beyond 600 nm tubes are less common (Sablinskas 2003).

3.1.2.3 Charged Coupled Device

A Charge-coupled device (CCD) is a type of photographic memory which is applied in numbers to constitute arrays of photoelectric light sensors. These can be used in digital photography, electron microscopy, medical fluoroscopy, and optical spectroscopy. As the numbers of colours to be detected grows, modern CCD chips can simplify microbiological analysis: So for example, in contrast to sensors utilizing photomultiplier tubes, limited to detecting only a few wavelength bands, CCD sensor systems enhance the captured light spectrum while maintaining the sensitivity of confocal microscopy. As a further bonus, cooling the CCD is not required to achieve high response rates (Hing and Müller 2003).

3.2 Optical Measurement Techniques

As mentioned before, the scope of optical spectroscopy comprises numerous variations of experimental applications, instrumentation, and measurement techniques, which are in many cases adapted to the types of samples and to the capabilities of the exciting and monitoring devices. Not all, but some of the most prominent methods are mentioned in the following sections.

3.2.1 Transmission Spectroscopy

According to Steiner (2003), *transmission spectroscopy* is the most widely used measurement technique. It is simple and can be applied to characterize gases, liquids and solids. By applying this technique, the light that passes through a sample is measured, whilst the portions of light which may be reflected, scattered, or absorbed and re-emitted by the sample are excluded.

3.2.2 Attenuated Total Reflectance Spectroscopy

In contrast, *attenuated total reflectance spectroscopy* (ATR) is a measurement technique where the reflection of light is made the centre of interest. Here, the beam from the spectrometer is coupled, creating an evanescent wave before it reaches the sensor. However, ATR spectroscopy is only one of several techniques to analyze light reflections, other reflectance measurement methods include *reflection absorption*, *reflection at thin films*, and *diffuse reflection*, the latter being a mixture of light reflections and light that has been scattered diffusely, meaning the light beam has been absorbed by the sample, diffracted within it, and re-emerged to intermingle with the reflected parts.

3.2.3 Rayleigh and Raman Scattering

Rayleigh scattering is the fundamental scattering technique, which has been explored as early as in the late 19th century. Nowadays, it is most widely used in liquids to study more complicated solutions and to determine molecular weights of macromolecules. In This method scatters light when it impinges on a sample, exciting its molecules to a vibrational energy level and quickly relaxing back to the ground energy level, and then it is re-emitted. *Raman scattering* observes the shift of light to a different wavelength; this is called first- or second-order Raman scattering, depending on the amount of energy involved during the light excitations. Raman scattering is always of very low intensity, and therefore its investigation requires high-quality instrumentation (Walker 1975). The Raman effect can be excited in the UV region, the visible region or in the NIR region (Hof 2003).

3.2.4 Infrared Spectroscopy

According to Hof (2003), *Infrared spectroscopy* is the most commonly used spectroscopic method; it is rapid, sensitive, easy to handle and provides many different sampling techniques for gases, liquids and solids. Its IR spectrum provides a unique molecular fingerprint of an individual compound. This has been demonstrated by Cross and Jones as early as 1969 when they used the IR spectrum in comparative studies of two substances to clearly identify these. Another reason why infrared spectroscopy may be famous in science is that in addition to its high sensitivity, its IR excitation force does not perturb photolabile systems, while with some other methods of spectroscopy extreme precautions must be taken not to spoil the sample (Woodruff et al. 1993). A more

specialized way of applying IR spectroscopy is Fourier Transform Infrared Spectroscopy (FTIR), where the fundamental measurement is an interferogram, which is Fourier transformed to give a spectrum (Smith 1996).

3.2.5 Photoacoustic Spectroscopy

The methods seem to become even more special in the case of *photoacoustic spectroscopy*. Here, thermodynamic parameters such as temperature or pressure are measured as acoustic waves in the surroundings of a sample. This kind of listening requires specialized ultrasound detectors using piezoelectric compounds (Anscombe 2003) However, according to Macková (et al. 2003), photoacoustic spectroscopy is not considered suitable for fast measurements, as it is limited by the speed of sound and the slow microphone or piezoelectric response time.

3.2.6 Flow Cytometry

A fast measurement technique is *flow cytometry*, which exploits features of *fluorescence spectroscopy*, where a sample is 'dyed' or 'labelled' with an analyte substance, which radiates luminescence when it is excited at a specific wavelength. Flow cytometers can rapidly examine physical and chemical properties of cells by employing laser-induced light scatter and fluorescence to measure parameters such as DNA content or cell surface antigens (Ginouves 2003).

3.3 The Optical Biosensor System Overview

The Optical Biosensor System (OBS) has originally been invented for the detection of bacteria in liquid samples and the analysis of bacterial growth curves. The device operates by shining a laser beam through a liquid sample containing numerous bacterial cells. Two photo diodes collect the amount of scattered light and output a voltage accordingly. The voltage output is then fed into an analog-digital converter which samples and quantizes the analog voltage data into its digital representation. The data in digital form can then be used to analyze the light which has been scattered onto the photo diodes by means of computational algorithms. The task of the algorithms is then to work on the data so as to find patterns characteristical for the bacteria under investigation.

Tests at the General Hospital PHL Laboratory in Southampton, UK, and at a City Clinic in Moscow, Russia, demonstrated that the application of the device conforms well to commonly 'accepted' medical microbiological practices, and that satisfactory results, unlike with most other approved techniques, were rapidly obtained between 2 to 4 hours (Rusteck 2000).

In the context of this research the OBS system is not utilized for the analysis of cell growth; instead, a two-step analysis is performed. In a first step, and in accordance with the device's original purpose, it is used to detect the presence of bacteria in liquid samples. In a second step, the system provides the platform for a technique to monitor characteristics of bacteria inferred mainly through motility, and thus to draw conclusions about the observed species.

3.3.1 Nephelometry

Resembling a spectroscopic setup, but in contrast to the applications described in Section 3.2, the OBS system is not used to generate patterns in the frequency domain, but in the time-domain: it observes a dynamic sample over a time-span while the sample space is excited with laser light at a fixed frequency. In other words, while commonly in applying optical spectroscopy the dynamic features of the measurements are evoked by changing the wavelength of light to which a (static) sample is exposed, generating a spectral curve, the OBS technique makes use of (static) light at a certain wavelength to generate a time domain curve by measuring dynamic features within the amount of light scattered by the sample at every instant in time (see Figure 3.1). This is a measurement technique called *nephelometry*, where light scattered by a sample is digitized and recorded over time.

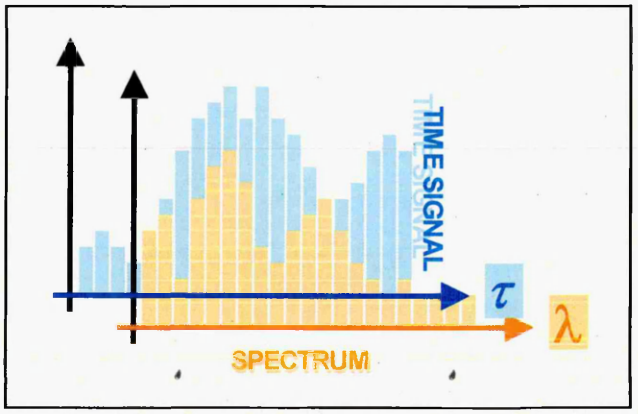

Fig. 3.1 Spectroscopic data may lead to two different optical signals, a spectral curve and a time curve

3.3.2 The Optical Biosensor System's Specifications

As much as the OBS system conforms to the classic principle of nephelometric turbidity, according to Rastopov (1998) there are differences related to the hardware assembly and the device output. The OBS system uses a two element photodiode and an electronic filter before the diode output is digitally amplified.

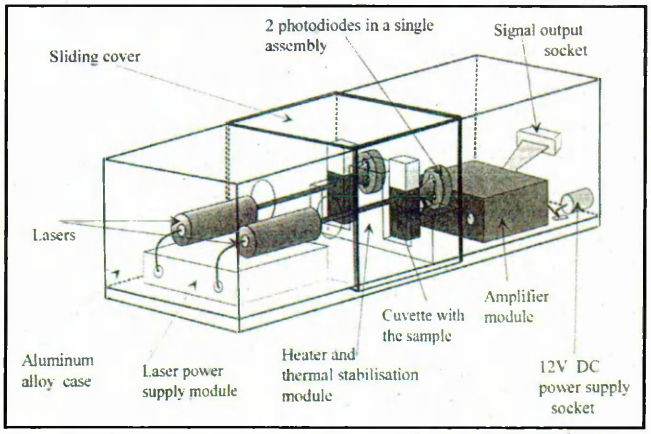

Fig. 3.2 The Optical Biosensor System (Rusteck 2000)

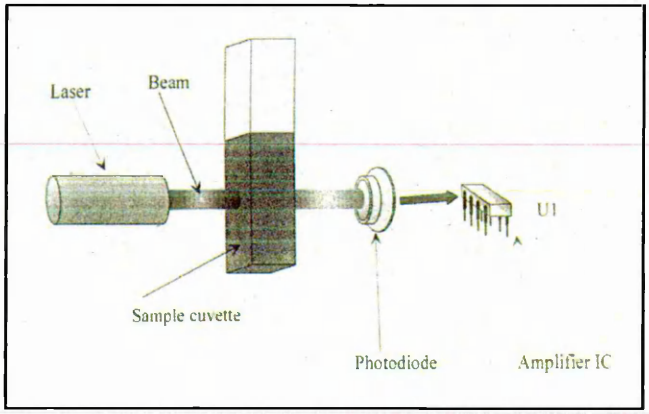

Fig 3.3 The standard nephelometric technique (Rusteck 2000)

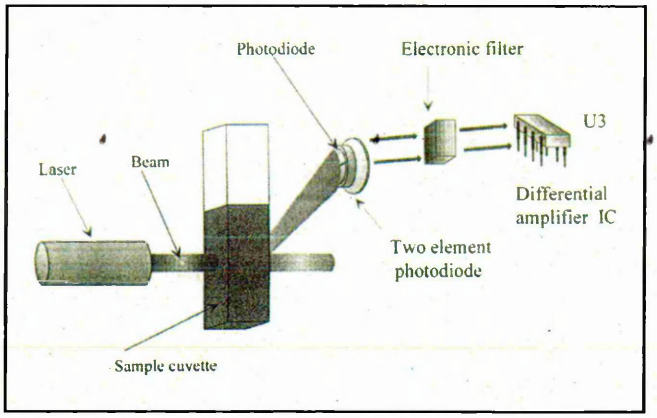

Fig 3.4 The Optical Biosensor System (Rusteck 2000)

3.3.3 The Output Signal

The output signal U of a standard nephelometric turbidity device can be represented as follows:

$$U_{neph}(K, I) = K * I_0 * (I_{bact.scat.} + I_{mol.scat.} + I_{absorb} - I_{cuv.scat.} - I_{win.scat+absorb}) \quad (3.1)$$

where:

K is the sensitivity of the photo diode (light to photo-current conversion coefficient)
I_0 is the light scattering without any particles present in the sample
$I_{bact.scat.}$ is the scattering caused by the bacteria and particles in the sample
$I_{mol.scat.}$ is the molecular scattering of the liquid
I_{absorb} is the light absorption in the liquid and cuvette material
$I_{cuv.scat.}$ is the scattering by the cuvette windows (fingerprints, scratches)
$I_{win.scat+absorb}$ is the scattering by the photo diode window

In contrast, the output signal of the OBS system is described as:

$$U_{OBS}(K, I) = K * I_0 * \delta(I_{bact.scat.}) \quad (3.2)$$

where:

$\delta(I_{bact.scat.})$ is the intensity of the fluctuation of the scattered light.

The OBS system uses coherent light with a relatively short wavelength at 670 nm to increase the intensity of the scattered light. In fact, the above relationship only exists when a coherent light source is used, as the output signal is formed by the interference of light scattered on the sensitive area of the photo diode by each particle in the liquid. Thus, according to Rastopov (1998), the amplitude of the variations of the scattered light intensity is also representative of the number of scattering particles within the sample volume; in other words, the more bacteria cells are present within a sample the higher the amplitude of the recorded output curve. This is an important feature of the device which needs to be taken into account when analyzing the output patterns.

3.3.4 Twin Photo Diodes

By using a pair of photo diodes instead of a single diode the signal to noise ratio of the output signal can be substantially increased. In this setup, the two detectors provide two different signals where correlated signal variations received by both detectors are rejected in the difference signal. On the other hand, the intensity variations received by each detector which are dependent on interference in the scattered light are not correlated, and so the amplitude of the non correlated signal variations will be additive in the difference signal to produce a total amplitude signal to $\sqrt{2}$ times the amplitude of variation from each of the two detectors (Rastopov 1998).

3.3.5 Light Source

The laser implemented in the OBS system is a solid state semiconductor laser, operating at 670 nm. It produces a collimated beam having a cross-sectional area greater than about 1 mm^2, and typically having a diameter of about 3 mm. According to Rastopov (1998), the width of the interior of the cuvette in the direction of the optical axis is normally greater than 3 mm and may be about 1 cm. As a result, the illuminated region may have a volume of about 70 mm^3.

3.3.6 Electronic Filter

The relatively large illuminated volume of the OBS system has the consequence that the output of the photodetector has a substantial dc component. However, because the amplitude of intensity variations in the detector signal are in fact also increased, filtering out the dc component of the detector signal can result in an intensity variation signal which is a sensitive measure of relatively low concentrations of scattering particles in the sample. Typically, the filter will also be arranged to have an upper frequency cut off to eliminate higher frequency noise and other disturbing elements from the signal (Rastopov 1998).

3.4 Determining Features for the Classification of Bacteria

Tailored to the observation technique, this approach is an attempt to establish possible characteristical features, which could be used as parameters for the classification algorithms in order to group bacteria species together. In order to make a decision about the affinity of a sample bacteria species to an existing bacteria class, the classification system relies on the analysis of those patterns, generated by the OBS system, which describe the microbial activities and/or morphology. Some possible characteristics, or features, of bacteria species are outlined in the following sections.

3.4.1 Size

Instances of this feature range from *very small* to *very large*, and *intermediate*, not so explicit values occur quite clearly with more average sizes of microorganisms. In a description of how bacteria vary in size, Mitruka (1976) points out that organisms may develop from elementary bodies (100 nm in diameter) into actively growing forms having diameters of from 200 to 250 nm. Their average volume is approximately $4 * 10^6$ nm^3. In contrast, one of the largest prokaryotic organisms, Thiospirillum jenese, has a volume of about $5 * 10^3$ μm^3, and a length up to 50 μm. If the size of the cell under observation was to give rise to characteristical recording patterns, clearly, the examination of these two contrasting organisms could demonstrate this. On the other hand, it is unlikely that this parameter plays a role when trying to distinguish types like Escherichia coli and Bacillus megaterium (which are quite similar in size).

However, size is not as easily assigned to a cell as it might appear at first thought. Indeed, for spherical cells, one is able to respectably compare the size of bacteria A to that of bacteria B, but many species have odd shapes, probably with extensions like flagella or stalks, which makes the judgement whether a typical cell of species A is larger or smaller than a typical cell of species B all but vague.

Examples of bacteria with diverse dimensions, either extremely small or extremely large, include Escherichia coli, Chromatium weissii, Spirillum volutans, Caryophanon latum Peshkoff, Thiovulum majus, Thiomargarita, Lineola longa, Sarcina ureae, Epulopiscium fishelsoni.

3.4.2 Motion

Instances of this parameter range from *very slow*, even *non-motile*, as opposed to *very fast*, where cells dash across the microscopic lense almost like a flash. Also, *diverse motion patterns* can be observed. There are different techniques by which bacteria achieve movement. Certain aquatic microorganisms can regulate their position in a water column by gas-filled structures called gas vesicles. However, most species are motile by means of flagella. Here is what Madigan (2000) describes as flagellar motion:

> The motions of polarly and lophotrichously flagellated organisms are different from those of peritrichously flagellated organisms. Peritrichously flagellated organisms generally move in a straight line in a slow, stately fashion. Polarly flagellated organisms, on the other hand, move more rapidly, spinning around and dashing from place to place.

Thus, flagellar diversities, including their complete absence, govern to a large extent the motion pattern of individual organisms.

Some examples of species, which may inhibit diverse motion patterns, or are non-motile, include Chromatium minus, Thiovulum majus, Thiococcus, Streptococcus agalactiae, Thiocapsa, Rhodothece, Bacillus subtilis, Proteus mirabilis.

3.4.3 Structure

Examples of structural features of prokaryotes are the composition of the *cell wall*, diverse *flagella*, the existence of *endospores*, and certainly the *internal structure* the cell's body.

Cell walls of gram-negative bacteria are generally thinner in profile, more complex in structure and composition, and mechanically weaker than those of gram-positive bacteria. Different bacteria species may have different flagellar arrangements (peritrichous, polar, lophotrichous), and the type of flagellation is often used to classify bacteria (Madigan et al. 2000). Endospores are special structures within a cell which are only produced by certain species of bacteria. According to Madigan et al. (2000) they are strongly refractive bodies, and thus they may have a measurable impact when excited by laser light. The internal structure of a prokaryote consists of proteins, nucleic acids, polysaccharides, and lipids, and every species consists of a unique composition of these.

While all these characteristics may be negligible to the applied technique of the OBS system when observing single or just a few cells, they could have the potential to generate a signal change significant enough to be recognizable when large quantities of bacteria assemble within the optical target area.

3.5 Utilizing a simple Classification Framework

The expected limitation of the OBS technique to monitor bacteria and their activities is the insufficiency to reveal all or most of their microscopic characteristics. This prohibits grouping species into a sophisticated framework such as the one made up of today's acknowledged phylogenetic classes or species as found in 'Bergey's Manual of Determinative Bacteriology' (Hensyl 1994) or as seen on the first pages of 'Brock - Biology of Microorganisms' (Madigan et al. 2000). Instead, the classification approach presented in this chapter inevitably results in a much simpler framework of classes: Here, the emphasis lies on making use of the features which the observation technique *can* provide to establish a new class structure, rather than trying to discover as many features as possible to be fairly certain to be able to group a species into one of the already existing classes of the (today's) biologically acknowledged framework.

3.6 Motion as the Carrier to transport Clues for distinguishing Cells

In essence, applying laser light to observe bacteria, as presented in this chapter, results in a series of energy fluctuations recorded over time, which are then digitized to be analyzed by a computer algorithm. This pattern of fluctuations recorded by the sensors is the representation of the behaviour of the bacteria under examination, i.e. a representation of their morphological and their physiological characteristics. In order to classify or even identify the thus observed microorganisms, a first step is to distinguish individual cells, or groups of cells, inhibiting similar characteristics from other cells with other characteristics.

In pursuing the question as to what kind of features or clues may be detectable within a recorded signal; a careful selection of bacterial species to be observed has to be undertaken: In an initial approach, it is desirable that the analysis of the fluctuations deals with the occurrence of explicit features encoded within the signal. In other words, features in the signal pattern should be easily recognizable due to an extremely clear manifestation of the related cell characteristics. This would, in theory, enable the analyzing system to distinguish feature A from feature B, hence distinguishing the underlying cells bacteria A (exhibiting feature A) from bacteria B (showing feature B). As mentioned before, distinguishing types of cells from other types is regarded to be the first step in classifying and - even further - identifying bacteria.

In a static system the recorded fluctuations would not change over time, or more precisely would not even result in fluctuations but in a static amount of energy (optical reflection) to be recorded - a stable line without slopes across time. Therefore, it is the motion or the changes of the recorded energy over time, within the observed system revealing the clues for an inference about the characteristics of the cell or group of cells giving rise to a certain pattern.

Attached to the recorded motion, or in other words to the changes of the energy curve recorded over time, are clues like how large the organism would be, how fast it moves, about the general shape of the organism, and about its surface and inner structure (properties and existence or lack of cell walls, flagella, capsules, spores, DNA, RNA, polysaccharides, or lipids). However, individual cells and some of their structural features are entities which are extremely small in relation to the laser beam target area,

and are therefore beyond the resolving capabilities of the applied optical laser-sensor technique. Whereas the target area of the laser beam measures more than 1 millimeter in diameter, the dimensions of bacteria cells and their cytoplasmic features range from only 20 nanometers to at most 1 to 2 micrometers. Hence, may their characteristics be as diverse as possible; they nevertheless fail to result in features detectable in the signal recorded by the optical sensors. As a consequence, motion itself, i.e. the motility of (groups of) microorganisms, is regarded to be by far the most important parameter, and the grade of its detectability makes or breaks the classification system as presented in this chapter.

3.7 Description of the Experiments

3.7.1 Bacteria used in the Experiments

The objective of the experiment is to distinguish individual species. In the course of the experiments, three different bacteria type species are separately observed. In order to choose suitable species, the underlying consideration is to attain a high level of diversity of the species as experimental test objects. The higher their diversity amongst each other, the easier would the task be to distinguish them from each other with the applied method. As has been suggested in the previous section, particular attention has to be paid to the motility of each species.

Accordingly, the species chosen for the experiments are:

a) Proteus mirabilis

b) Bacillus subtilis

c) Escherichia coli

Here is an exact description of each of the three species, taken from "Bergey's Manual of Determinative Bacteriology" (Holt et al. 1994):

> Genus Proteus
>
> Straight rods, 0.4 - 0.8 μm in diameter * 1 - 3 μm in length. Gram negative. Motile by peritrichous flagella. Most strains swarm with periodic cycles of migration producing concentric zones, or spread in a uniform film, over moist surfaces of nutrient media solidified with agar or gelatin. Facultatively anaerobic and chemoorganotrophic, having both a respiratory and a fermentative type of metabolism. Optimal temperature is 37°C. D-Glucose and a few other carbohydrates are catabolized with production of acid and usually gas. Oxidase negative, catalase positive, and methyl red positive; species vary in indole, Voges-Proskauer, and Simmons citrate tests. Lysine decarboxylase and arginine dihydrolase negative; only Proteus mirabilis decarboxylates ornithine. Phenylalanine and tryptophan are oxidatively deaminated, and urea is hydrolized. Decompose tyrosine to produce a clearing on agar media on which the insoluble amino acid is incorporated. Grow on KCN. H_2S is usually produced. Malonate is not utilized. Reduces nitrates. One or more species ferment glycerol, maltose,sucrose, trehalose,

and D-xylose. Occur [...] in polluted waters. Human pathogens, causing urinary tract infections [...].

Genus Bacillus

Cells are rod-shaped and straight, 0.5 - 2.5 μm * 1.2 - 10 μm, and are often arranged in pairs or chains, with rounded or squared ends. Cells stain Gram positive and are motile by peritrichous flagella. Endospores are oval or sometimes round or cylindrical and are very resistant to many adverse conditions. There is not more than one spore per cell, and sporulation is not repressed by exposure to air. Aerobic, or facultatively anaerobic, with wide diversity of physiological ability with respect to heat, pH, and salinity. Chemoorganotrophs, with a fermentative, or respiratory metabolism. Usually catalase positive. Found in a wide range of habitats; a few species are pathogenic to vertebrates or invertebrates.

Genus Escherichia

Straight rods, 1.1 - 1.5 μm * 2.0 - 6.0 μm, occur singly or in pairs. Capsules or microcapsules occur in many strains. Gram negative. Motile by peritrichous flagella [...]. Facultatively anaerobic. Chemoorganotrophic, having both a respiratory and a fermentative type of metabolism. Optimal temperature is 37°C. D-Glucose and other carbohydrates are catabolized with the formation of acid and gas. Oxidase negative, catalase positive, methyl red positive, Voges-Proskauer negative, and usually citrate negative. Negative for H_2S, urea hydrolysis, and lipase. [...] Reduces nitrates. All or most strains ferment L-arabinose, maltose, D-mannitol, D-mannose, L-rhamnose,trehalose, and D-xylose. O-Nitrophenyl-β-D-galactopyranoside positive. [...] Escherichia coli strains that contain enterotoxins and/or other virulence factors, including invasiveness and colonization factors, cause diarrheal disease. Escherichia coli is also a major cause of urinary tract infections and nosocomial infections including septicemia and meningitis. [...] Escherichia coli is often subdivided serologically or by the presence of virulence factors to identify and characterize epidemiologically pathogenic strains. Complete serotyping includes somatic (O), capsular (K), and flagellar (H) antigens. [...].

As much as (motile) diversity is highly important, the factors governing the decision to employ these three species in the experiments are in fact a trade-off between 1) their diverse properties, 2) their scientific relevance, and 3) their availability.

1) As for their motility, the species Proteus mirabilis has the fastest moving cells[1]. In contrast to the above description of Bacillus subtilis, the strain of this species used for the experiments here is, as microscopic observation shows, almost non-motile. Escherichia coli cells are observed to be fast, however, not as fast as Proteus mirabilis. Apart from their differences in motility, by using a microscope with a 40 times magnification, no other differences (e.g. size, shape, cell structure) are apparent. This

[1] In fact the motility of the strain of Proteus mirabilis used in these experiments is very high so that counting alive cells with the help of a haemocytometer (see Section 2.4.2) proves challenging. As a result one has to cut back on precisely estimating numbers.

seems to prove the assumption that motion is the most prominent characteristic to be detected[2].

2) All three species used here constitute examples of widely spread and frequently used laboratory cultures. One reason for this is the important role they play in different scenarios, demanding continuous research in many laboratories throughout the world. Especially desirable is the speedy and automated detection of Escherichia coli: It is regarded as the workhorse and model organism for both research and applications of genetic engineering (Madigan et al. 2000).

3) To the author's knowledge, species do exist which inhibit a greater degree of diversity as opposed to the ones employed here (see Section 3.4); thus, for example, non-motile species, or species many times larger and odd-shaped would further make the differences between the species more expressive, and the task for the classification system would be more relaxed. Alas, the limitations in laboratory support and finances, for the project at hand makes, there acquisition impossible.

3.7.2 Sample Preparation for the Experiments

A sample for the OBS system is prepared from an isolated bacteria colony growing on agar. Thus, a swab of the desired organism is inoculated into a liquid growing solution environment. Escherichia coli, Bacillus subtilis, and Proteus mirabilis all grow utilizing the same standard nutrient broth; this comes as a pre-made powder and has to be dissolved in the appropriate amount of water which is then autoclaved for sterilization prior to inoculating a swab.

Then, all three cultures grow for 16 to 18 hours (i.e. over night) at 37°C in a chamber which is shaking each sample at 150 rpm to reach a desired concentration of approximately $5 * 10^9$ cells per ml. At this level the haemocytometer (see Section 2.3.2) can be used in order to determine the actual concentration of the so-called 'overnight culture'. As some samples will inoculate faster, others slower, the result is a slightly different cell concentration per culture for relatively long growing periods.

However, in order to provide a consistent observation system, the concentration of bacterial cells has to be somehow calibrated, so that all the individual samples are comparable. Moreover, as per the working principle of the OBS system, Rastopov (1998) points out that it is important that the relative magnitude between the DC component of the detector signal and the intensity variation component is not so great that the detector is either saturated by the DC component or the intensity variation component becomes comparable to the amplitude of the noise signal produced by the detector. This means that the concentration of each overnight culture has to be considerably reduced in order to reach a consistent, ideal cell concentration for a sample to be used in the observing sensory device. After the actual number of cells has been determined with the haemocytometer, a factor can be calculated by which each individual sample needs to be diluted. Trials prior to the experiments have shown that an optimal digitized reading is

[2] It is expected that an optical, manual impression of the cells at a 40 times magnification is superior to the capabilities of the analysis of a static single laser beam reflection (on a spot of a few mm^2) in detecting individual cell properties.

attained at a concentration c_{cal} of around $8*10^7$ cells per ml to which each sample has to be reduced by stepwise diluting the overnight culture.

After reaching the final experimental concentration c_{cal}, each bacteria culture is separated from the growing solution by means of centrifuging and is then re-inoculated into the respective medium which is used while the sample will be under observation. Three different observation media have been chosen for the experiments:

a) Sterilized and autoclaved water,
   Is supposed to be a clear observation medium with as little interference from other particles as possible, however, with no means to support the organism's biochemical functionality; as there are no nutrients to be metabolized, organisms activities will cease after prolonged exposition

b) Ringer solution,
   A medium to help the organism to keep up its biochemical functionality to a certain extent and stay alive for longer exposition; a trade-off between genuineness of the liquid medium and bacterial activity

c) Nutrient broth,
   Enhances biochemical functions and activity of the organism, with the drawback of an increased amount of undesired interference, in other words increased cross-reflexions of the exciting light waves within the sample target area

3.7.3 Experimental Procedure

The bacteria sample with the calibrated concentration c_{cal} is now ready to be used in the OBS system (see Figure 3.5). Within the OBS system, a bacteria sample is residing in a polystyrene semimicro cuvette with a 1 ml capacity. Their favorable optical transmission range lies between 400 to 800 nm (VWR International 2006). When the OBS system is activated, its solid state semiconductor laser produces a collimated beam at 670 nm with a cross-sectional area greater than 1 mm^2 and a typical diameter of about 3 mm. The width of the interior of the cuvette in the direction of the optical axis is 5 mm which indicates an illuminated region of about 70 mm^2. The scattered light impinges at an angle of between 5° and 7° from the laser beam axis onto the detector, which is in fact an assembly of two seperate photo diodes, each of which has a sensitive area of about 1 mm^2 in size. Here the collected photons are converted into two streams of electric current, and the thus generated signals are supplied via respective DC blocking capacitors to a band pass filter, which filters out frequencies below 10 Hz and above 300 Hz, before they are further supplied to the inverting and non-inverting inputs of a differential amplifier. This way any correlated varying components of the two signals tend to cancel out in the single output signal from the differential amplifier (Rastopov 1998).

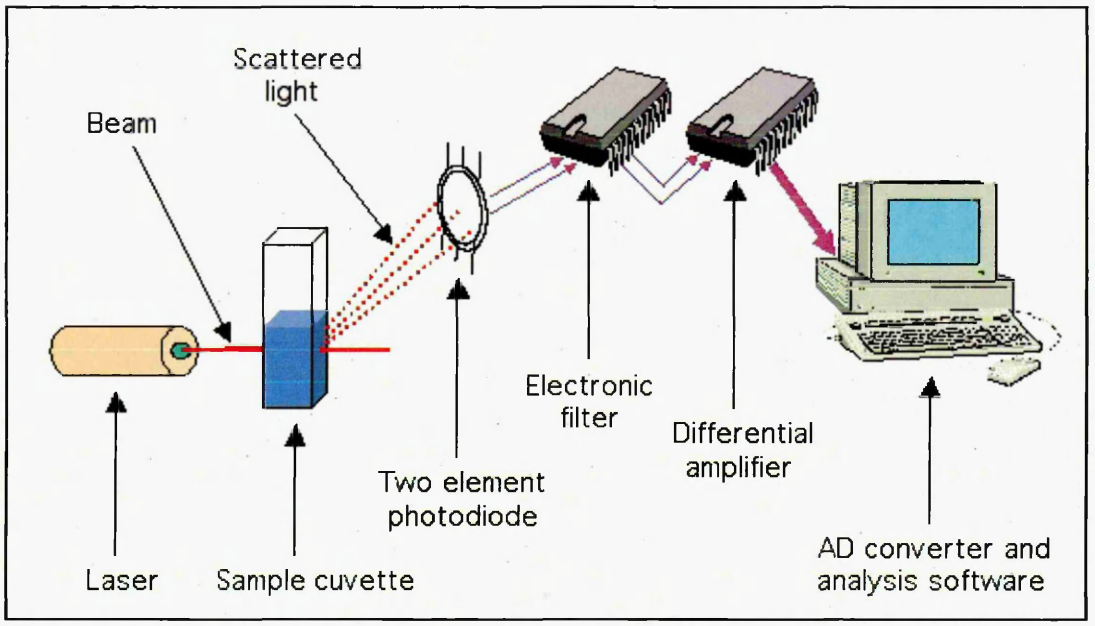

Fig 3.5 Operation of the Optical Biosensor System

At this point the signal leaves the OBS system and is connected to the input of a PC soundcard via a standard RCA jack. The soundcard used in the experiments is an "Audiosystem EWX 24/96" which has a 24-bit converter chip with a signal to noise ratio of 110 dB(A) and an analog input level sensitivity of -10 dBV (Terratec 2000). On the part of the software, the incoming signal is sampled by a hard disk recording program ("Samplitude 2496") at 22 kHz, 16 Bit, 705 kBit/s (stereo format, i.e. sampling 2 experiments at a time), and transformed into the standard PC wavefile format (.WAV). As the energy of the signal is expected to be in the range of 10 Hz to 300 Hz the sampling frequency of 22050 Hz may seem extraordinarily high, but is nevertheless chosen due to the standardized hard- and software handicap. In addition to this, according to Nyquist's theoreme a signal should be sampled at a rate which is larger than twice the highest occuring frequency. This means that an adequate sampling frequency in the present case would be approching 1 kHz, however according to Proakis and Manolakis (1996) only a very high sampling frequency guarantees the correctness of the representation of a true analog signal. Therefore, a much higher sampling frequency is all but beneficial regarding the correctness of the digital representation of the analog signal. The duration of a thus conducted experiment is 18 minutes while it produces 48 megabytes of wave data.

3.7.4 The Experiments at a Glance

Each of the chosen species results in a total of 9 experiments (i.e. 9 * 18 minutes of data recording / observation): 3 observations in sterile water, 3 in ringer solution, and 3 in nutrient broth. Table 3.6 shows the identifier of each experiment and its duration, and the respective media into which the microorganisms are immersed for each experiment[3]. For an exact description of each experiment see Appendix A.

	BROTH	WATER	RINGER
Escherichia coli	EC_B_1 [77 min] EC_B_2 [77 min] EC_B_3 [77 min]	EC_W_1 [18 min] EC_W_2 [18 min] EC_W_3 [18 min]	EC_R_1 [18 min] EC_R_2 [18 min] EC_R_3 [18 min]
Bacillus subtilis	BS_B_1 [77 min] BS_B_2 [77 min] BS_B_3 [77 min]	BS_W_1 [18 min] BS_W_2 [18 min] BS_W_3 [18 min]	BS_R_1 [18 min] BS_R_2 [18 min] BS_R_3 [18 min]
Proteus mirabilis	PM_B_1 [77 min] PM_B_2 [77 min] PM_B_3 [77 min]	PM_W_1 [18 min] PM_W_2 [18 min] PM_W_3 [18 min]	PM_R_1 [18 min] PM_R_2 [18 min] PM_R_3 [18 min]

Tab. 3.6 Table of conducted experiments

Although previously mentioned and listed in the above table, it needs to be said here that the data analysis (as presented in the subsequent paragraphs) excludes all experiments in broth (they have been mentioned for completeness of reasoning only). This is because the examination of 'test signals' reveals that the level of noise in the signal varies considerably relating to the environmental conditions within which sample cells are contained. Accordingly, the OBS system has been step-wise calibrated in absence of any microorganisms by running experiments with no cuvettes present, with empty cuvettes, and with cuvettes filled with sterile water, ringer solution, and broth. Then, every effort is made to successfully distinguish the three bacteria species by, at first, focusing on the sterile water and ringer solution media, both of which cause less noise than nutrient broth in the recorded signal and are thus expected to be less difficult to be analyzed. To anticipate the results of the data analysis, it becomes apparent in the subsequent paragraphs that none of the applied (pattern generation) algorithms performs well enough to enable an adequate classification of the 18 analyzed microorganisms (6 per species, 3 in sterile water, 3 in ringer solution) into their respective classes. As a result, the further (more difficult) analysis of broth-experiments is henceforth abandoned.

3.8 Data Analysis

The following paragraphs comprising the data analysis are, after a general explanatory Section (3.8.1), sectioned into two parts: Firstly, the analysis using statistical methods

[3] Table 3.6 shows that experiments conducted in broth are 77 minutes long as opposed to 18 minutes for the other media. This is due to the assumption that bacterial growth does occur in broth and may play a role for distinguishing the observed species, as different species grow at different rates. As for the other experiments, 18 minutes is too short a time span to reveal changes in the recordings due to bacterial growth.

(Section 3.8.2), and secondly, the analysis using mathematical methods (Section 3.9.3). The subsections within these two main sections are in a chronological order, such that each subsection, apart from the first one, builds on the findings of the previous one.

While describing the algorithms in each section, diagrams and tables of data are displayed. In order to improve visualization and readability, sometimes images may display only parts of diagrams, or not all generated diagrams (e.g. for every experiment conducted) are shown in a section. For an exhaustive display of all diagrams and tables see Appendix B.

3.8.1 Splitting the Data of an Experiment into Sections for Ease of Processing

To handle the rather large amount of generated data, the complete data of a single experiment (48 MByte) is divided into so-called 'stints' of 250000 samples each; this means that each stint is a portion of the entire 18 minutes long experiment representing 11338 milliseconds, resulting in 95 stints altogether. Comparing stints from the beginning, middle, and end sections of an experiment shows that the average amplitude and fluctuations of the signal of every stint are similar to each other, which negates the assumption that bacterial growth or death of cells during the entire experiment could occur to an extent where it becomes relevant for the analyzing algorithms. In other words, whether an algorithm uses stint #1 or stint #45 or stint #90 of an experiment in order to analyze its data with regard to a possible classification evidence is irrelevant, as waveform changes from one stint to another (of the same experiment) are not significant and can be disregarded[4].

3.8.2 Statistical Analysis

Given that the only output of an experiment, which is likewise the only input for the statistical analysis, is a series of sampled data values over a certain period in time, the most straight forward statistical method is usually operating on the values of the samples, or on the values of sections of samples. In the present case, however, it has to be taken into account that all conducted experiments are subject to differences concerning the total number of bacteria cells (per single experiment), which reflects different, i.e. less high or higher, photonic impulses sensed by the system. This in turn results in a less high or higher recorded amplitude across an entire experiment. As a consequence, the concentration c_{cal} varies to some degree for every experiment, which means that, after the digitizing process, each experiment has an average sample value which is different from every other experiment. This is a direct result from the insufficiency to precisely determine the number of cells of bacteria cultures with a haemocytometer (see Section 2.4.2).

Prior to the application of statistical methods, for the above reasons, it should be stressed that differences in photonic impulses and thus differences in the values of the digital samples are not a result of diverse properties of different bacteria cultures, but are solely caused by an experimental constraint. Therefore comparing and operating directly on

[4] In fact, in the course of processing several stints with different algorithms it has become evident that, in order to draw conclusions about a possible classification of the experimental objects (the recorded bacteria), the analysis of only a single stint per experiment is sufficient.

data values of samples would be statistically misleading and is hence disregarded in the following statistical analysis.

3.8.2.1 The Wave Concept

As dealing with direct sample values of an experiment is not advisable (for reasons described in the previous paragraphs), alternative features are sought which can be statistically determined. Consequently, all statistical pattern generation algorithms presented in this section make use of the concept according to which each experiment is comprised of a number of (adjacent) waves. A wave consists of several samples and 'up and downs', or slopes, i.e. increasing or decreasing sample values in the direction of the time signal. Accordingly, the size of a wave is governed by the number of its samples, and it can be further characterized by its number of changes in slope direction (see Figure 3.7).

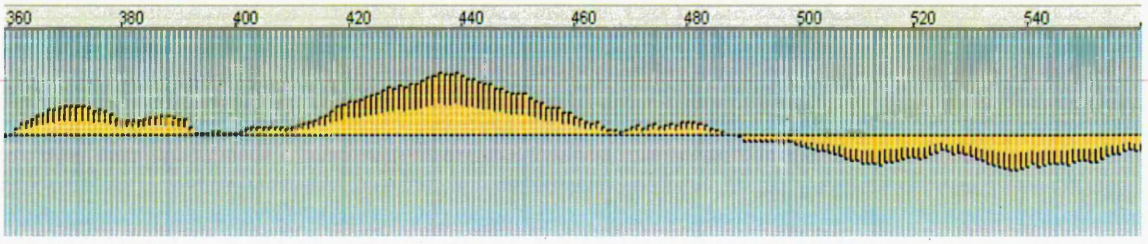

Fig. 3.7 A signal slice of 8 ms, corresponding to 180 samples, showing the formation of waves (scale in samples)

3.8.2.2 Counting the basic statistical Features of an Experiment

The basic statistical features of every time-slice, or stint, of an experiment are identified as:

a) Number of waves detected in a stint

b) Number of samples of every wave detected in a stint

c) Number of slope changes detected in every wave of a stint

The following algorithm is used to count features a) b) and c):

```
% function to generate basic pattern for a stint of an experiment
% features a) b) and c)

wave = read wavefile('stint.wav')

% presets

currentslope=1;
previousslope=1;
w=1;                                % counter: total number of waves
j=1;                                % counter: samples per waves
nofslopechanges(1..250000)=0;       % array: max size of vector
nofsamples(1..250000)=0;            % -"-
thresholdtensionup=0;
thresholdtensiondown=0;


% determine values

for i = 1 to 250000
    if (( wave(i)<0 ) and ( wave(i+1)>0 )) or (( wave(i)>0 ) and ( wave(i+1)<0 ))
       add 1 to w;                  % a new wave starts
       set j to the for-counter i;  % set j to position 1 of the wave
    end;

    % now find out whether a slope change has occured

    threshold=abs(wave(i+1)-wave(i));
    if threshold > 0.015            % always reset tension usually
       reset thresholdtensiondown to 0;
       reset thresholdtensionup to 0;
    end;

    if abs(wave(i)) < abs(wave(i+1))
       if (threshold > 0.015) or (thresholdtensionup > 3)
          set currentslope to 'up';
          release thresholdtensionup back to 0;
       else
          increase thresholdtensionup by 1;
          if thresholdtensiondown > 0 decrease thresholdtensiondown by 1;
       end;
    end;

    if abs(wave(i)) > abs(wave(i+1))
       if (threshold > 0.015) or (thresholdtensiondown > 3)
          set currentslope to 'down';
          release thresholdtensiondown back to 0;
       else
          increase thresholdtensiondown by 1;
          if thresholdtensionup > 0 decrease thresholdtensionup by 1;
       end;
    end;

    if currentslope <> previousslope
       increase nofslopechanges(w) by 1;
       set nofsamples(w) to i-j;    % how large is the wave (so far)?
    end;

    previousslope=currentslope;
end;


% write values in output file

for j = 1 to w                      % for every wave
    output (nofslopechanges(j),nofsamples(j));
end;
```

According to the algorithm, feature a) is represented as the counter *w* which results in *w* entries of both feature c) *(nofslopechanges)* and feature b) *(nofsamples)* being saved, i.e. a 2*n matrix whereby the values for the two features c) and b) are in each row of the matrix, and the value of feature a) is equal to the total number of rows of the matrix.

In order to determine whether the slope of the curve of a wave changes its direction (from upwards to downwards or vice versa), the algorithm makes use of a threshold value: Is the sample value of the current sample considerably lower or higher as the value of the next sample, i.e. it lies above the threshold (positively or negatively), and also depending on the current slope direction (up or down), a slope change occurs. Here, the threshold is set to 0.015 (see Figure 3.8).

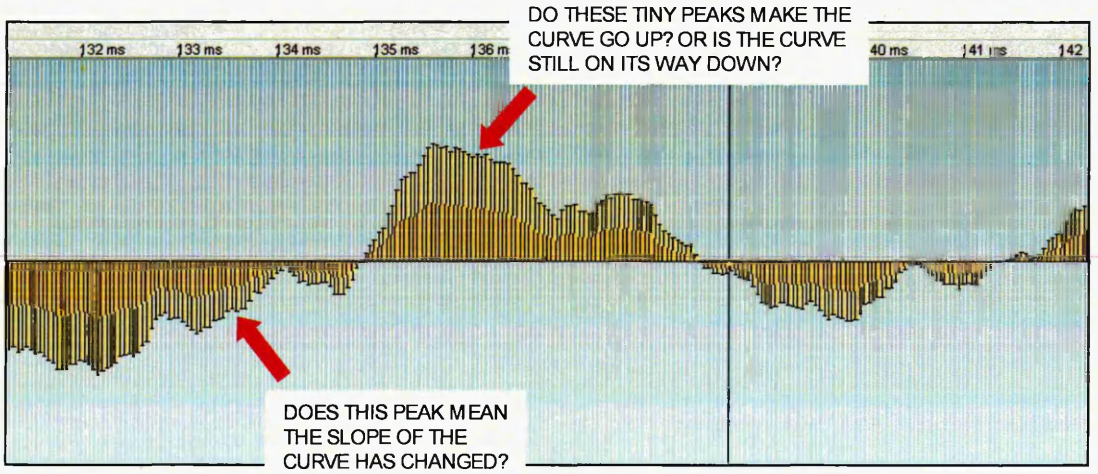

Fig. 3.8 The threshold will make the algorithm skip these 'epitomes'

The concept of a 'threshold tension', as seen in the algorithm, out-rules the threshold in that it ensures that slope changes do occur after a certain number of consecutive ups or downs (here 4 times), even if at any time the changes from one sample to the next are below the threshold. This caters for the impression that a curve indeed changes its direction if only enough, however very small, consecutive changes from one sample to the next (in the same direction) occur.[5]

The values generated by the counting of features a) b) and c) are transformed into a diagram (= pattern) for an improved overview (see Figures 3.9, 3.10, 3.11). Note from the algorithm that numbers for c) have been increased by factor 5 such that the graphical effect in the diagram is enhanced.

A different diagramatical pattern for feature c), resembling a machine code, condenses the graphical data such that more of its underlying statistical data can be seen at a glance, and patterns from different experiments are more easily compared (see Figure 3.49).

In the same graphical way, another (condensed machine code) diagram creates a pattern which can be used for distinguishing between experiments of different species or

[5] It has to be mentioned that the question as to whether a curve is already changing direction or not yet is a matter of individual perception (and probably individual eyesight capabilities). The values for the threshold and the 'tension' as applied in the current context have been chosen to reflect the author's perception

grouping experiments of identical species together: It is the representation of mathematically relating feature c) to feature b), i.e. [number of slope changes of a wave] divided by [number of samples of this wave], for an entire stint (see Figure 3.50).

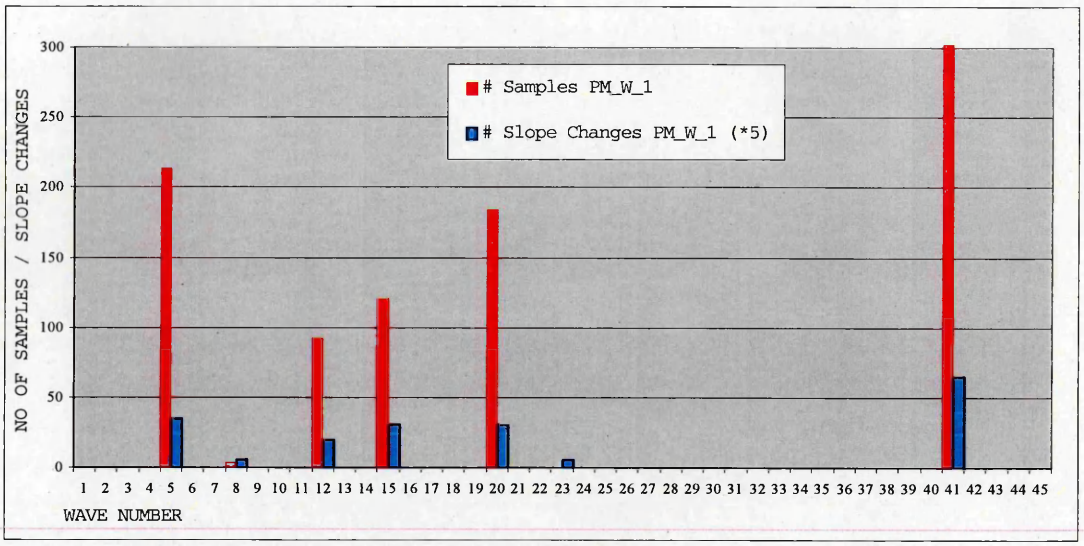

Fig 3.9 Plot of experiment PM_W_1 with Proteus mirabilis in water: Number of samples (feature b) and number of slope changes (feature c) counted per wave (total number of waves = feature a)

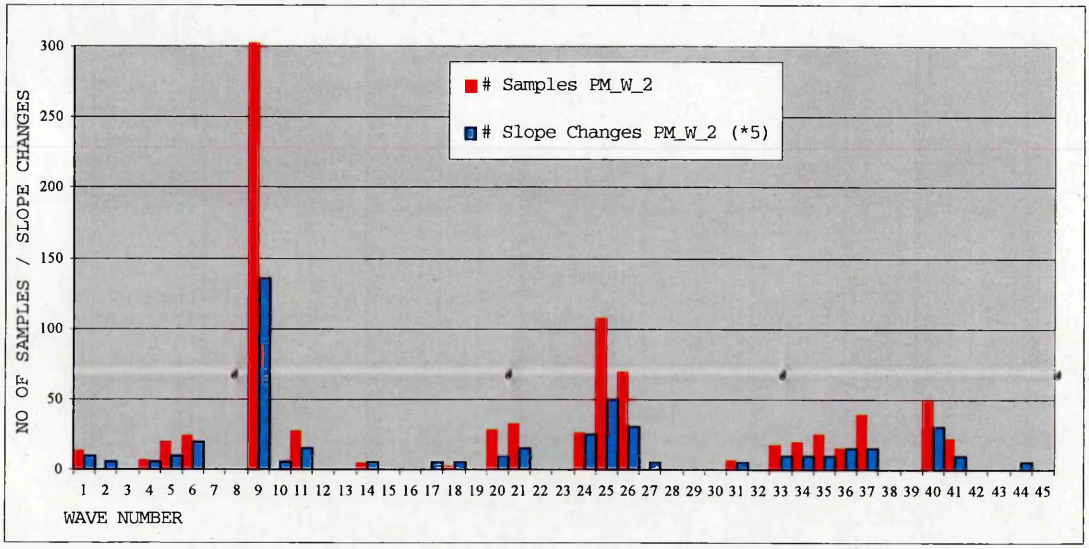

Fig 3.10 Plot of experiment PM_W_2 with Proteus mirabilis in water: Number of samples (feature b) and number of slope changes (feature c) counted per wave (total number of waves = feature a)

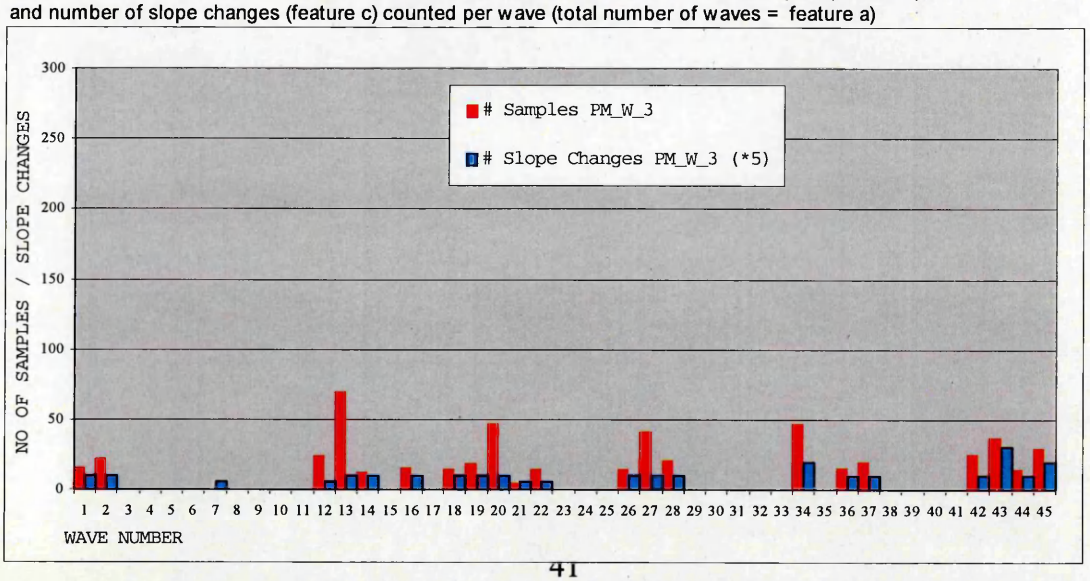

Fig 3.11 Plot of experiment PM_W_3 with Proteus mirabilis in water: Number of samples (feature b) and number of slope changes (feature c) counted per wave (total number of waves = feature a)

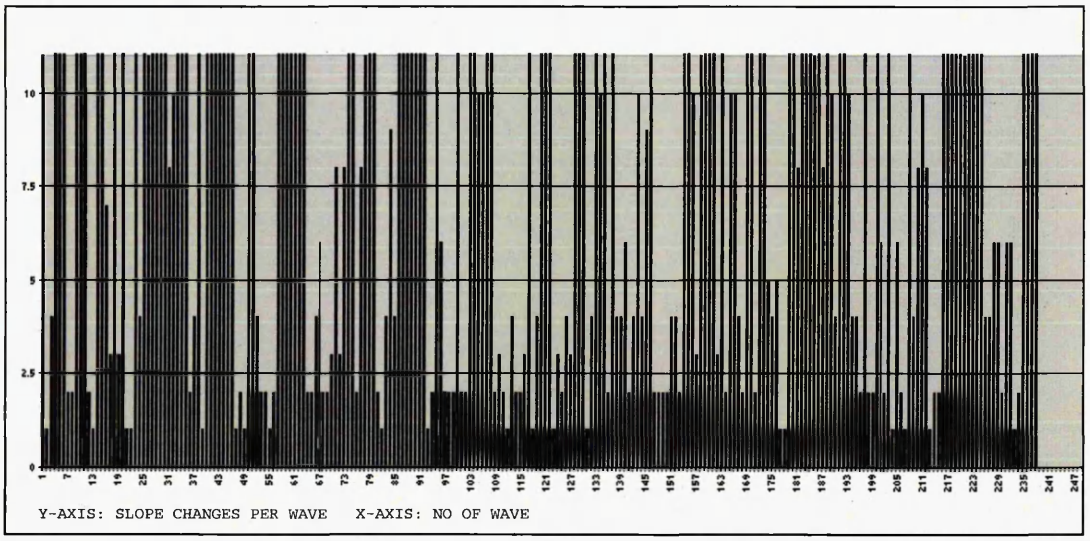

Fig. 3.12 'Machine code' pattern of feature C (slope changes per wave)of Bacillus subtilis in water

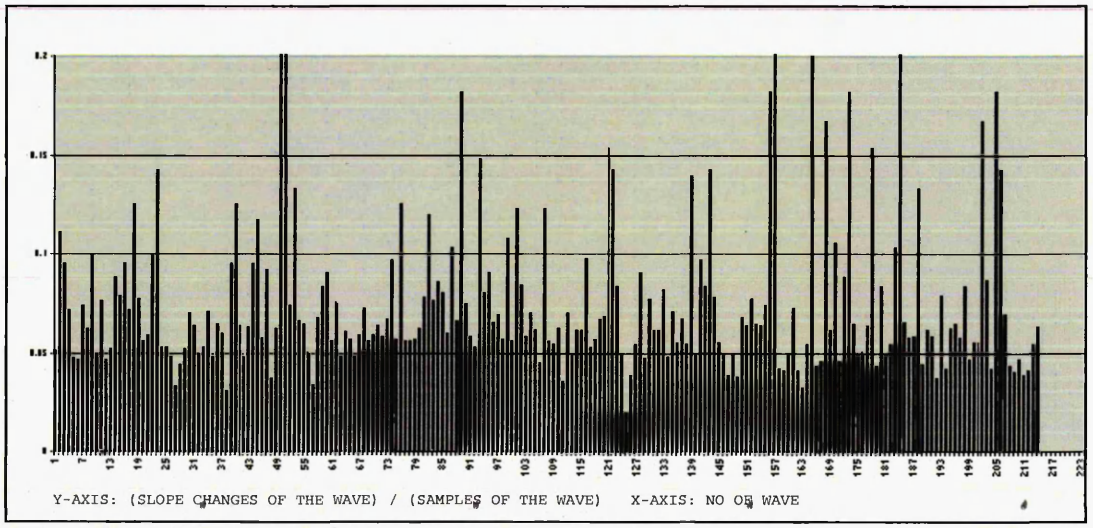

Fig. 3.13 'Machine code' pattern plotting feature C (slope changes of the wave) / feature b (samples of the wave) (Bacillus subtilis in water)

3.8.2.3 Counting Statistical Features to generate a Fingerprint Pattern

This section makes use of the features b) and c), as described in the previous section, but divides these values into categories, which results in more statistical values in order to generate a more complex pattern. This pattern, consisting of a number of subpatterns, may then be interpreted as a so-called 'fingerprint' for bacterial species. Additionally, a new feature is obtained:

d) Numbers of consecutive large waves[6]

This feature originates from the observation that slower moving species result in data streams with less fluctuations, than faster moving species. In terms of wave properties

[6] It has the same impact on the generation of a pattern whether consecutive small waves or consecutive large waves are counted; here the author's decision is to focus on displaying data for large waves.

this means that there can either be a lower number of waves (per stint) with a comparable high number of samples (slow moving species), or a higher number of waves with a comparable low number of samples (fast moving species).

A pattern to be constructed by merely two values, i.e. [number of large waves] and [number of small waves], can be fairly limited in its meaningfulness regarding the distinguishing and grouping of experiments. By further counting the concatenation of large or small waves within an experiment and relating these two parameters, this additional information considerably enhances the resulting pattern.

The following algorithm extends the previous one described in section 3.8.2.2; it executes directly afterwards:

```
% function to generate data for a graphical fingerprint

highnofslopechanges(1..3)=0;
mednofslopechanges(1..3)=0;
lownofslopechanges(1..3)=0;
highnofsamples=0;
mednofsamples=0;
lownofsamples=0;
widthcounter=0;
width(1..15)=0;

for j = 1 to w                                  % for every wave

    if nofsamples(j) >= 290
       increase highnofsamples by 1;
       if nofslopechanges(j)*5 <= 33% of nofsamples(j)
          increase highnofsamples by 1;
       if nofslopechanges(j)*5 > 33% of nofsamples(j)
             and nofslopechanges(j)*5 < 66% of nofsamples(j)
          increase highnofslopechanges(2) by 1;
       end;
       if nofslopechanges(j)*5 >= 66% of nofsamples(j)
          increase highnofslopechanges(3) by 1;
       end;
    end;

    if 100 < nofsamples(j) < 290
       increase mednofsamples by 1;
       if nofslopechanges(j)*5 <= 33% of nofsamples(j)
          increase highnofsamples by 1;
       if nofslopechanges(j)*5 > 33% of nofsamples(j)
             and nofslopechanges(j)*5 < 66% of nofsamples(j)
          increase highnofslopechanges(2) by 1;
       end;
       if nofslopechanges(j)*5 >= 66% of nofsamples(j)
          increase highnofslopechanges(3) by 1;
       end;
    end;

    if 30 < nofsamples(j) <= 100
       increase lownofsamples by 1;
       if nofslopechanges(j)*5 <= 33% of nofsamples(j)
          increase highnofsamples by 1;
       if nofslopechanges(j)*5 > 33% of nofsamples(j)
             and nofslopechanges(j)*5 < 66% of nofsamples(j)
          increase highnofslopechanges(2) by 1;
       end;
       if nofslopechanges(j)*5 >= 66% of nofsamples(j)
          increase highnofslopechanges(3) by 1;
       end;
    end;

    if nofsamples(j) > 30                % relatively large wave
       increase widthcounter by 1;       % count how many consecutive large waves
    else
       if widthcounter > 0
          increase width(widthcounter) by 1;
```

```
            reset widthcounter to 0;
        end;
    end;

end; % for

output highnofsamples,highnofslopechanges(1),highnofslopechanges(2),
highnofslopechanges(3),mednofsamples,mednofslopechanges(1),mednofslopechanges(2),
mednofslopechanges(3),lownofsamples,lownofslopechanges(1),lownofslopechanges(2),
lownofslopechanges(3),width(1),width(2),width(3),width(4),width(5),width(6),width(7),
width(8),width(9),width(10),width(11),width(12),width(13),width(14),width(15));

end;
```

The first part of the algorithm divides the values of features b) and c) into 3 and 3*3 subfeatures respectively. Instead of merely counting a single [number of samples of a wave] and a single [number of slope changes of this wave], both numbers are divided into high values, medium values, and low values (see further below).

The second part of the algorithm counts the number of consecutive large waves[7] and increases a counter value accordingly, i.e. if two adjacent waves have been identified as large and the next wave is small, then the value at the second position in the array *width* is increased; if 7 adjacent (or consecutive) waves have been determined before a small wave, then *width(7)* is increased. The array is of length 15 as no more than 15 consecutive large waves ever appear in the experimental data.

In the case of feature b), its subfeatures are:

b1) [no of samples] (> 290) = [high]

b2) [no of samples] (< 290 AND > 100) = [medium]

b3) [no of samples] (< 100) = [low]

Within each subfeature category of feature b) the subfeatures of feature c) are:

c1) [no of slope changes *5] (>= 2/3) of {[no of samples] of b?)} = [high]

c2) [no of slope changes *5] (< 2/3 AND > 1/3) of {[no of samples] of b?)} = [medium]

c3) [no of slope changes *5] (<= 1/3) of {[no of samples] of b?)} = [low]

[7] The question of how many samples constitute a large wave and how many a small wave is, like setting a sensible threshold value as in the previous section, a matter of opinion and of manually analyzing the data of several experiments. In the author's opinion a value of 30 samples as the boundary of small and large waves is adequate for the task at hand.

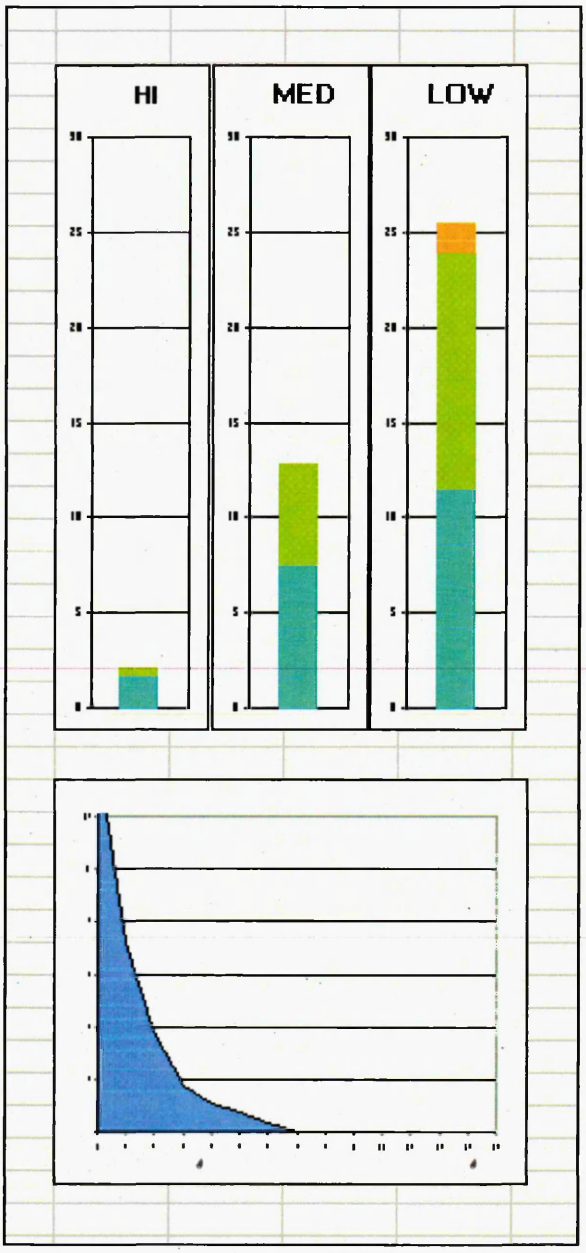

Fig. 3.14 'Fingerprint pattern' resulting from the enhanced statistical data processing (Proteus mirabilis in water)

This means that for each subfeature value b?) subfeature values c1) c2) and c3) exist, altogether resulting in 12 values (per stint). Additionally, feature d) comprises 15 values (per stint). This splitting up of features, as opposed to their usage in the previous section, and the additional values of feature d), result in an increased number of statistical data (see Figure 3.14) and hence in a more detailed pattern (see Figure 3.15) to be generated. Also, it should be mentioned that the values to generate this pattern are not merely derived from discretly processing stints, but are averaged across numerous stints.

3.8.2.4 The Simplification of Waves

PM_W_1

26	26	0	0	120	119	1	0	56	47	9	0	143	20	2	3	0	0	0	0	0	0	0	0	0	0
26	26	0	0	117	113	4	0	70	57	13	0	153	21	3	1	1	0	0	0	0	0	0	0	0	0
30	30	0	0	121	119	2	0	57	44	13	0	156	14	6	1	0	0	0	0	0	0	0	0	0	0
30	30	0	0	109	108	1	0	60	48	12	0	131	21	7	1	0	0	0	0	0	0	0	0	0	0
24	24	0	0	136	135	1	0	57	47	9	1	142	26	6	1	0	0	0	0	0	0	0	0	0	0
24	24	0	0	129	124	5	0	62	54	8	0	176	13	4	0	0	0	0	0	0	0	0	0	0	0
24	24	0	0	126	126	0	0	55	50	5	0	156	19	3	0	0	0	0	0	0	0	0	0	0	0
30	30	0	0	122	119	3	0	67	60	7	0	172	17	2	0	0	1	0	0	0	0	0	0	0	0
35	35	0	0	108	105	3	0	57	52	5	0	144	22	2	0	1	0	0	0	0	0	0	0	0	0
28	28	0	0	118	117	1	0	60	52	8	0	149	27	1	0	0	0	0	0	0	0	0	0	0	0
28	28	0	0	121	119	2.1	0	60	51	8.9	0.1	152	20	3.6	0.7	0.2	0.1	0	0	0	0	0	0	0	0
5.3	5.3	0	0	11	11	1.4	0	7.8	7.1	3	0.3	12	4.5	1.9	0.8	0.4	0.3	0	0	0	0	0	0	0	0

Fig. 3.15 Statistical data divided into sections of subfeature b1) (orange), b2) (green), b3) (blue), and feature d) (yellow). Each coloured line represents data counts from a stint of the experiment, the white lines on the bottom form the average and the square root of the average respectively (Proteus mirabilis in water)

Simplifying curves (or waves in the current context) by removing (optically) unnecessary samples uses a standard mathematical technique, called line generalization. Altering the sampled data for each experiment in such a way may reveal different properties for different bacteria species. In order to establish these properties the data resulting from the application of this method are statistically analyzed. This is why this section is mentioned here, and not in 3.8.3 (mathematical analysis).

In order to be able to make use of a line generalization technique, the samples of the experimental wave data have to be transformed into a so-called 'polyline'. Within a (two-dimensional) polyline each point of the line is identified by two coordinates (X and Y). The transformation of wave data (with n samples) into a polyline can hence be realized by turning the data stream into a 2*n matrix, the X-coordinate being the offset of a sample in the data, and the Y-coordinate being the actual value of the sample.

There are a few line generalization algorithms with different properties; the algorithm applied here is probably the most well known and widely used algorithm developed by Douglas and Peucker in 1973 (Robinson et al. 1995). When it was developed it was regarded as the most effective algorithm for line generalization and hence has found widespread use in various applications. It is effective because its generalization achievements are comparable or even outclass those of other, relatively complex algorithms, while its recursive functionality is simple and can be easily implemented. The Douglas-Peucker algorithm selects extreme points of a line and retains the selected points on the generalized line, whereas all intermediate points are discarded (see Figure 3.16)[8].

[8] However, it has been argued by Visvalingam and Whyatt (1993) that the extreme points of a line might not always be the points required to preserve its character, which is the aim of any generalization process.

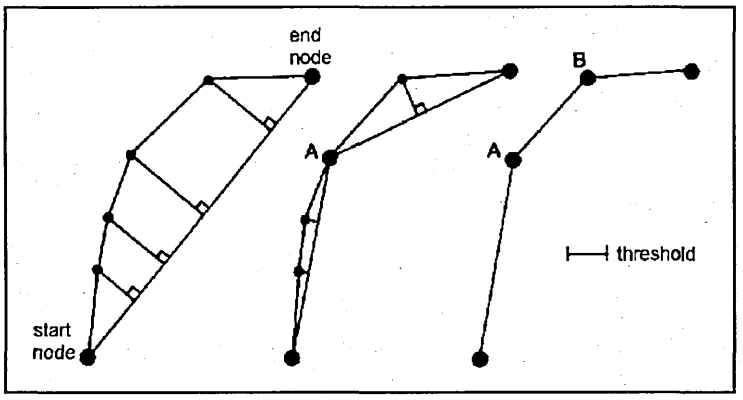

Fig. 3.16 Working principle of the Douglas-Peucker algorithm
(Robinson et al 1995)

The algorithm consists of two parts: The main body loads the wave data of an experiment, presets variables, calls the generalization function for each wave detected in the data, and saves the results into the respective output files. The function *'simplifyDP'* is the actual generalization function, which recursively calls itself until the simplification of the wave (according to a fixed tolerance value *tol*) is finished. The exact description of the algorithm is as follows:

```
% main body: function to simplify waves with douglas-peucker line genralization method

read wavefile from 1 to 250000 bytes into array wave;

firstwave=1;
j=0;
tol=0.01;                                  % tolerance for elimination of points

for i = 1 to 250000

    increase j by 1;
    vt(j,1:2)=[j abs(wave(i))];

    if (( wave(i)< 0 ) and ( wave(i+1) > 0 )) or (( wave(i) > 0 ) and ( wave(i+1) < 0 ))

       if firstwave == 1
          firstwave=0;        % discard first (incomplete) wave-segment
       else
          if j > 50           % only handle waves > 50 samples

             % simplify wave-segment vt with number of points=j

            mk(1)=1;
            mk(j)=1;
            mk=simplifyDP( tol, vt, 1, j, mk );     % recursive!

            m=1;
            n=1;
            evt(1,1..2)=[0 0];
            for z = 1 to j
                if mk(z) == 1                % save new simplified wave-segment, i.e.
                   svt(m,1..2)=vt(z,1..2);   % copy only points which were marked in mk
                   increase m by 1;
                else
                   evt(n,1..2)=vt(z,1..2);   % 0 in mk = evt (matrix with erased points)
                   increase n by 1;
                end;
            end;

      % write (no of points in vt, no of eliminated points, elimination in %)

      p=size(svt);
```

```
      q=size(evt);
      r=size(vt);

      output svt(p(1),1), svt(p(1),1)-p(1), (100/svt(p(1),1))*(svt(p(1),1)-p(1)))
             into file("statistics");

      % write 3 files:
      % evt = eliminated vertex
      % svt = simplified vertex
      % ovt = original vertex

      for wu=1 to size(evt)
          output evt(wu,1), evt(wu,2)) into file("evt");
      end;

      for wu=1 to size(svt)
          output svt(wu,1), svt(wu,2)) into file("svt");
      end;

      for wu=1 to size(vt)
          output vt(wu,1), vt(wu,2)) into file("ovt");
      end;

      end;    % if j > 50
   end;

 end;

end;
```

The output generated by the algorithm, for every wave identified (number of samples > 50) within a stint, is threefold:

ovt) 'original vertex'; represents a wave consisting of the original sampled wave data;

svt) 'simplified vertex'; is a simplified representation of a wave after it has been generalized., i.e. less number of points are being used to display a shape as similar as possible to the original vertex (or wave). The higher the tolerance value the more reduced the resolution will be;

evt) 'eliminated vertex'; this is not a curve or line like the two previous outputs, but records the coordinates of all points which are eliminated from the original wave data to generate the new, simplified 'wave curve'. As these points are the unnecessary remnants of the original wave (after the generalization procedure), it makes no sense to connect them into a curve.

Additionally, properties of the wave generalization, namely the [number of points in the original wave], [number of eliminated points in the simplified wave], and the [elimination in percent] are output into a file; these can be evaluated statistically (see further below).

The effect the algorithm has on wave data can be demonstrated by stepwise increasing the tolerance value *(tol)*, which results in a more and more abstracted wave while trying to retain its original character (see Figures 3.17, 3.18, 3.19, 3.20).

In the author's opinion, and in the context of this research, a tolerance value of 0,1 results in a level of simplification which can be easily recognized with the human eye, still returning a reasonable grade of abstraction of the original features of a wave, or in other words generating a wave which is not highly exaggerated. This applies to all waves being processed by the algorithm in the course of this research, e.g. [wave #1, stint #10,

experiment PM_W_1 (Proteus mirabilis)] (see Figures 3.21, 3.22, 3.23). The alterations may be more clearly seen if ovt), svt) and evt) are displayed in a single frame (see Figure 3.24).

Apart from investigating the sizes of the processed waves, whereby differences from one experiment (or maybe species) to another may be noticed at a first glance, the important statistical value, regardless of the actual wave size, is the [elimination (of points of a wave) in percent] (as opposed to the original wave size) (see Figure 3.25).

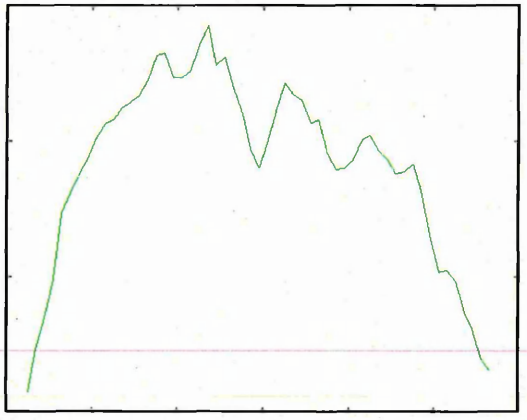

Fig. 3.17 Original wave (Proteus mirabilis)

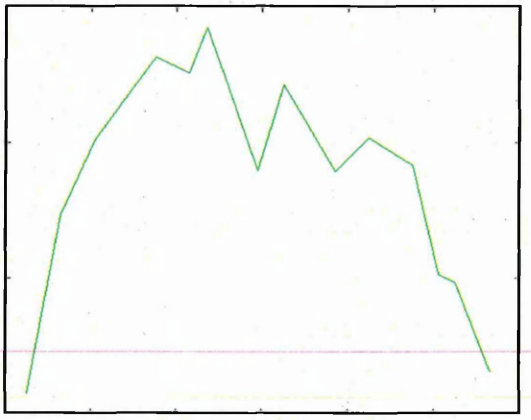

Fig. 3.18 Simplified wave with tolerance 0.008 (Proteus mirabilis)

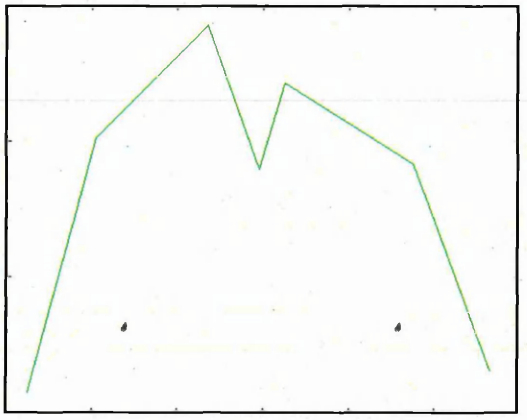

Fig. 3.19 Simplified wave with tolerance 0.02 (Proteus mirabilis)

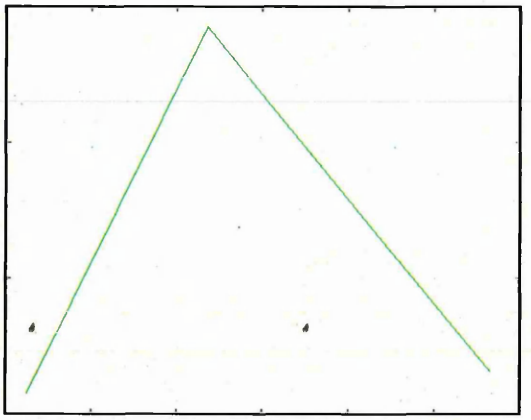

Fig. 3.20 Simplified wave with tolerance 0.08 (Proteus mirabilis)

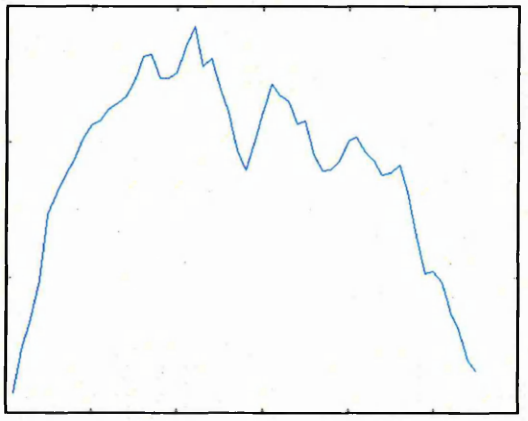

Fig. 3.21 An original wave (Proteus mirabilis)

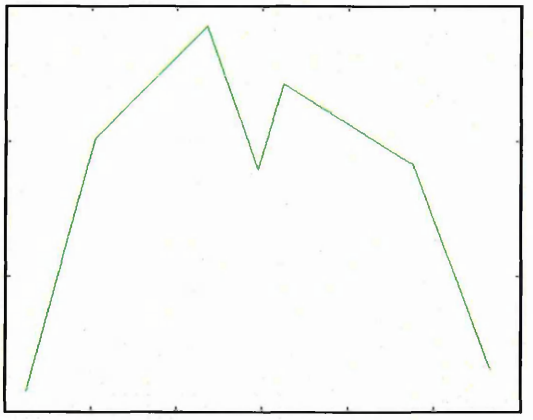

Fig. 3.22 A simplified wave (Proteus mirabilis)

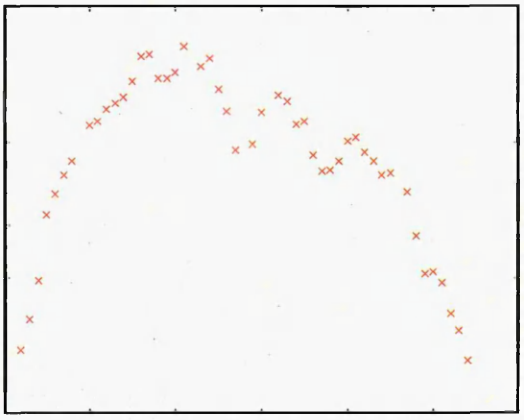

Fig. 3.23 Eliminated points of a wave (Proteus mirabilis)

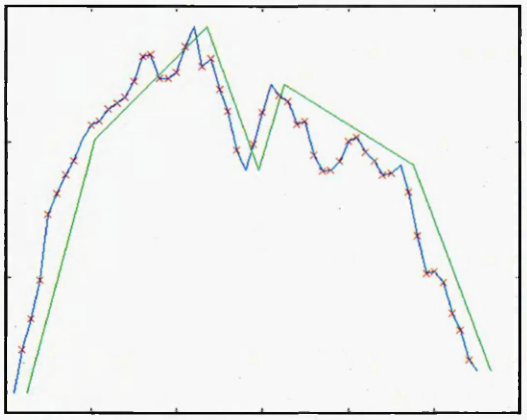

Fig. 3.24 Original wave, simplified wave, and eliminated points of the wave in a single frame (Proteus mirabilis)

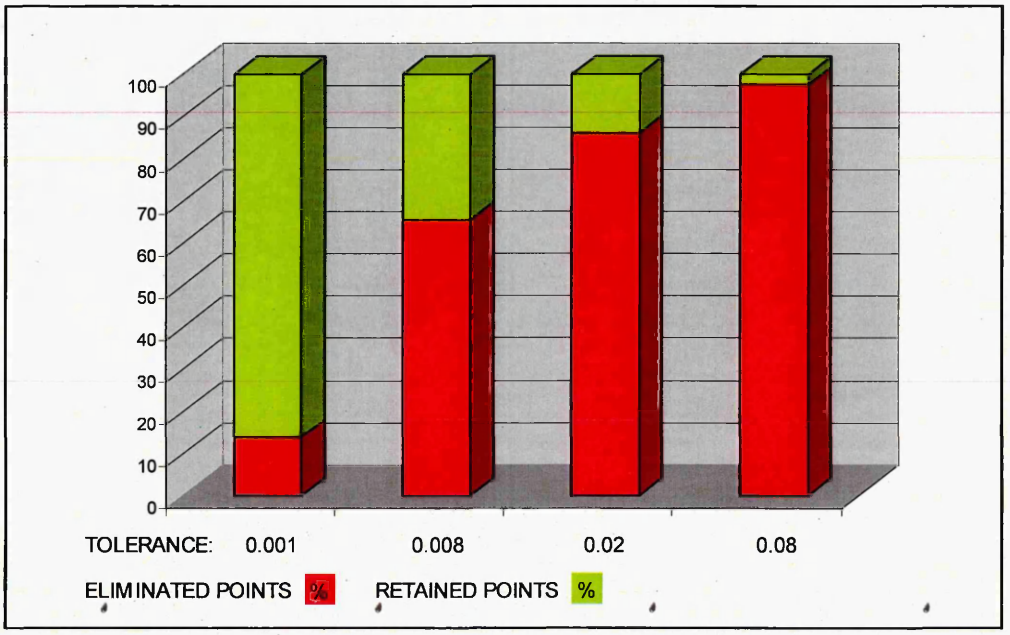

Fig. 3.25 Typical diagram showing the averaged elimination of points of waves for different tolerance settings (Bacillus subtilis)

3.8.3.1 Using the Discrete Fourier Transform to classify Spectral Signals

In the 19th century the mathematician Jean Baptiste Fourier developed a mathematical method capable of describing contrasting processes. With this sophisticated technique, called Fourier analysis, it is possible to split up complex energy waves (or other periodical, recurrent, compound patterns) into their very basic fluctuating components. Its counterpart, the Fourier synthesis, is the mathematical procedure to restore the whole from its numerous single components.

Thus, additionally to the plot of a signal's amplitude (e.g. voltage) over a period of time, there is another form of representing the same signal: The spectral density is a plot of the signal's energy over a range of frequencies.

Any data which have been recorded over a period of time, and hence exist as a signal in the time domain, can be transformed into a frequency domain signal using the Discrete Fourier Transform (DFT). Based on the Fourier Series representation of a signal, which requires the original signal to be a continuous periodic signal, in the case of a random aperiodic signal the DFT assumes the entire set of discrete data values to represent the fundamental period of a time domain signal. Thus any finite aperiodic signal can be represented as an infinite periodic signal by simply repeating the fundamental period (i.e. the signal itself) an infinite number of times (see Figure 3.26).

Unlike the Fourier Series, which serve to (mathematically) reproduce a given signal in the same domain, the DFT transforms a time domain signal into a frequency domain signal and vice versa. The result is the original set of discrete data values being represented by a continuous signal, which is called the energy density spectrum (or spectral density, or power spectrum) of the time domain signal. The energy density spectrum can be used to gather additional information about the original signal, e.g. the amount of energy within a given frequency range (A), the amount of energy at the respective maxima (B), or the frequencies at which the signal has zero energy (C) (see Figure 3.27). Using this information one might be able to improve the accuracy when clustering signals with similar characteristics or distinguishing between signals with adverse characteristics (see Figure 3.28).

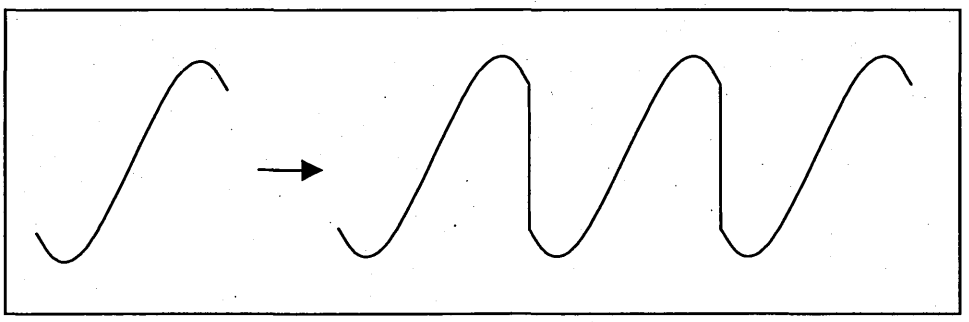

Fig. 3.26 Repetition of the fundamental period to generate an infinite periodic signal to be processed by the Discrete Fourier Transform

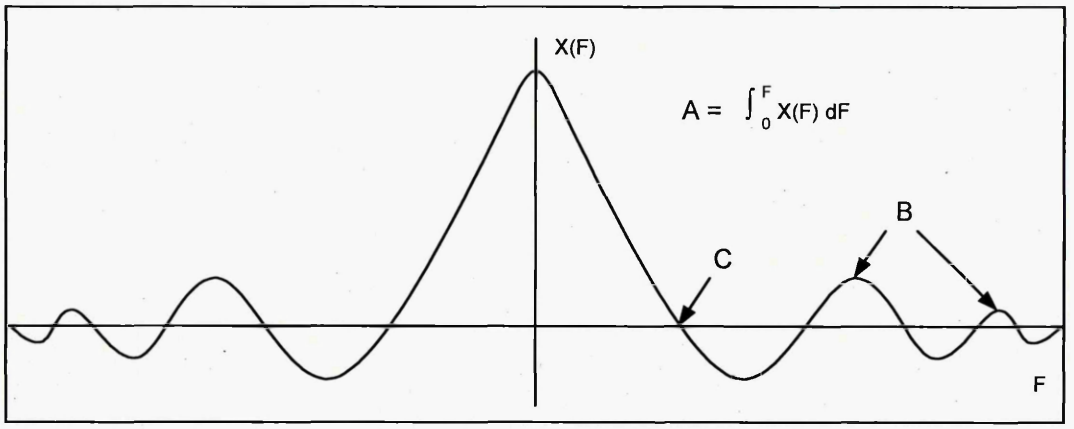

Fig. 3.27 Mathematical clues (A, B, C) gained by transforming a time-domain signal into its frequency domain representation

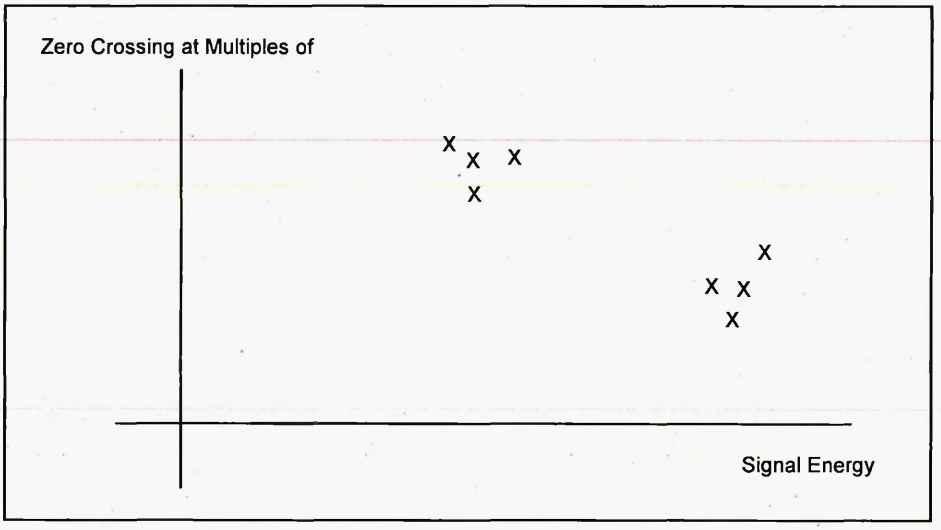

Fig. 3.28 Clustering high-frequent and low-frequent signals with the Fourier Transform

In order to demonstrate its practical capabilities and usefulness for characterizing signals, the following images (see Figures 3.29 to 3.36) visualize the effects of a Fourier Transform on two different signals. The two sine-wave signals (50 Hz and 5000 Hz) are analyzed by altering the scale of the X-axis (frequency) and Y-axis (signal strength) of their respective power spectrum into linear and logarithmic representations. It can be easily seen that, with this simple analysis, each of the two signals has a characteristic 'spectral fingerprint' in the frequency domain.

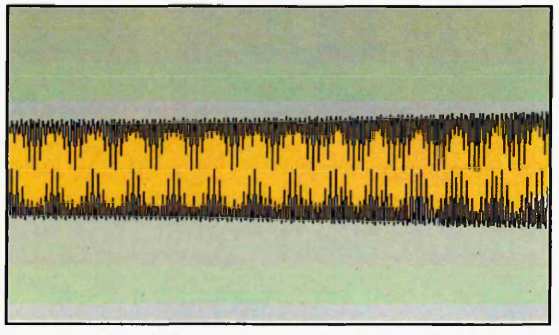

Fig. 3.29 Sine wave at 5000 Hz

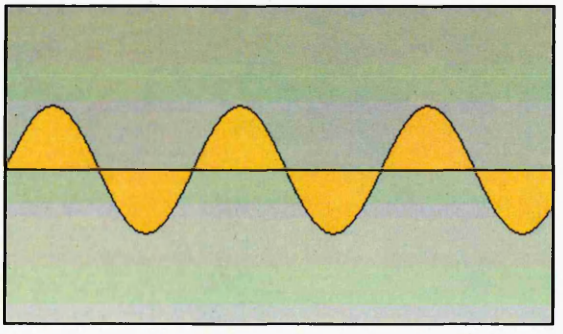

Fig. 3.30 Sine wave at 50 Hz

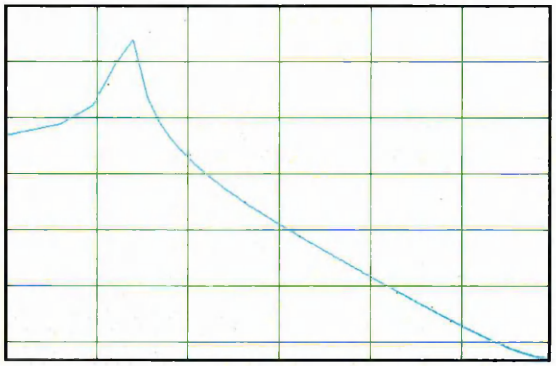

Fig. 3.31 X-axis: Frequency (log) Y-axis: Energy (log)
50 Hz sinewave

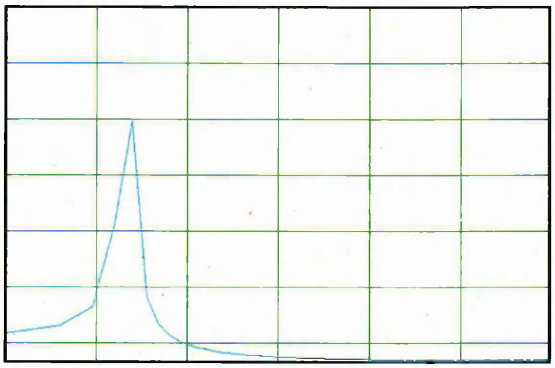

Fig. 3.32 X-axis: Frequency (log) Y-axis: Energy (lin)
50 Hz sinewave

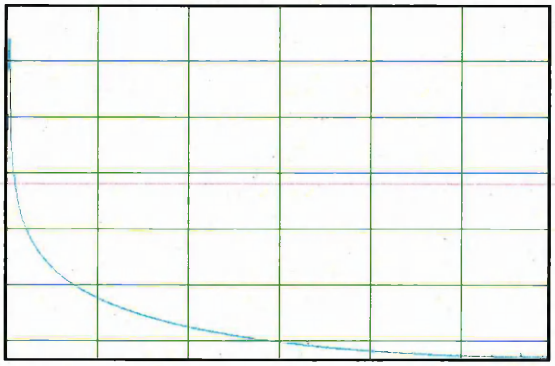

Fig. 3.33 X-axis: Frequency (lin) Y-axis: Energy (log)
50 Hz sinewave

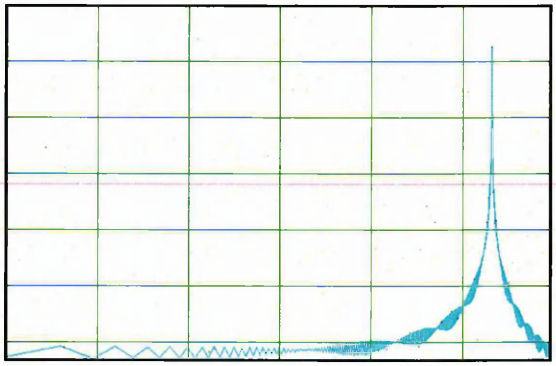

Fig. 3.34 X-axis: Frequency (log) Y-axis: Energy (log)
5000 Hz sinewave

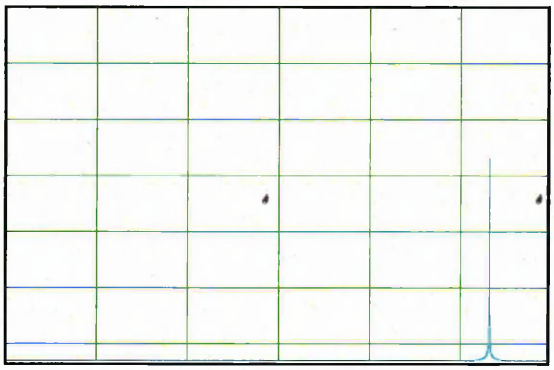

Fig. 3.35 X-axis: Frequency (log) Y-axis: Energy (lin)
5000 Hz sinewave

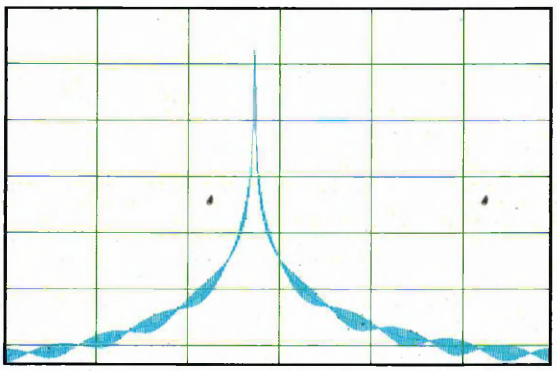

Fig. 3.36 X-axis: Frequency (lin) Y-axis: Energy (log)
5000 Hz sinewave

With regard to the characterization of aquatic microorganisms, Nibler and Pubanz (1988) suggest that both frequency- and time-domain methods are important for the study of dynamic processes in liquids. But in contrast to mathematically constructed signals with a constant pattern, real bacteria signals are a mix of lots of different (energies at lots of different) frequencies and are thus far more difficult to be told apart, as the energy spans over a frequency range rather than being isolated at a single frequency. Also, the energy distribution in a single experiment (i.e. the 'fingerprint' of a single bacterial strain under observation) varies, sometimes remarkably, for every instant in time. This means that for a single experiment, depending on when during the course of the experiment the data are analyzed, and how large the 'window', i.e. the time-slice, for the Fourier analysis is, different spectral representations are obtained.

The window size is a parameter of a DFT affects the precision with which the data to be transformed are analyzed. Apart from executing much faster, especially on large amounts of data, sometimes a low resolution is preferable to a high resolution as it averages out fluctuations of the energy curve; this improves the readability of a signal.

The following two diagrams (see Figures 3.37 and 3.38) show the results of applying a DFT on 10 stints of an experiment respectively using two different window sizes': The first diagram is generated with a window size of 4096, and the second diagram is the result of using a window of the size 65536.

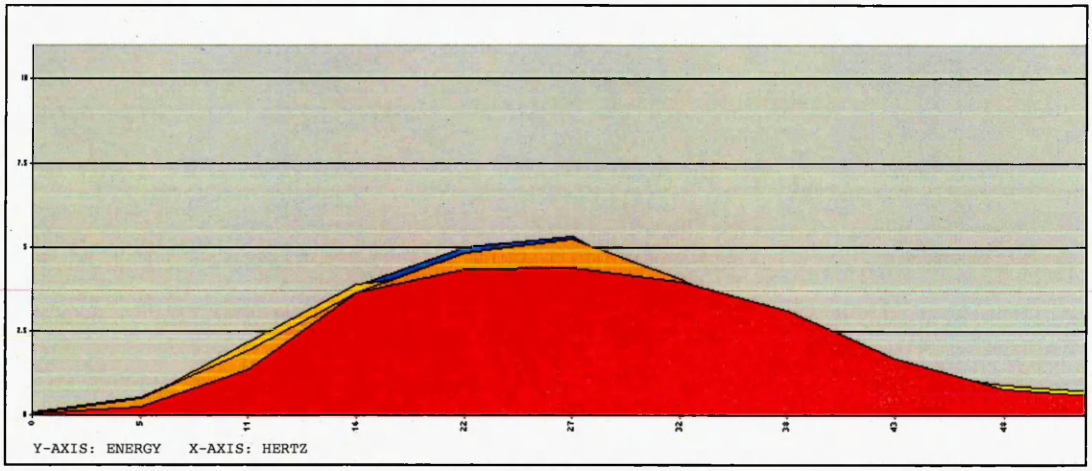

Fig. 3.37 Power spectra of various stints from experiment BC_W_2 with window size 4096 (Bacillus subtilis in water)

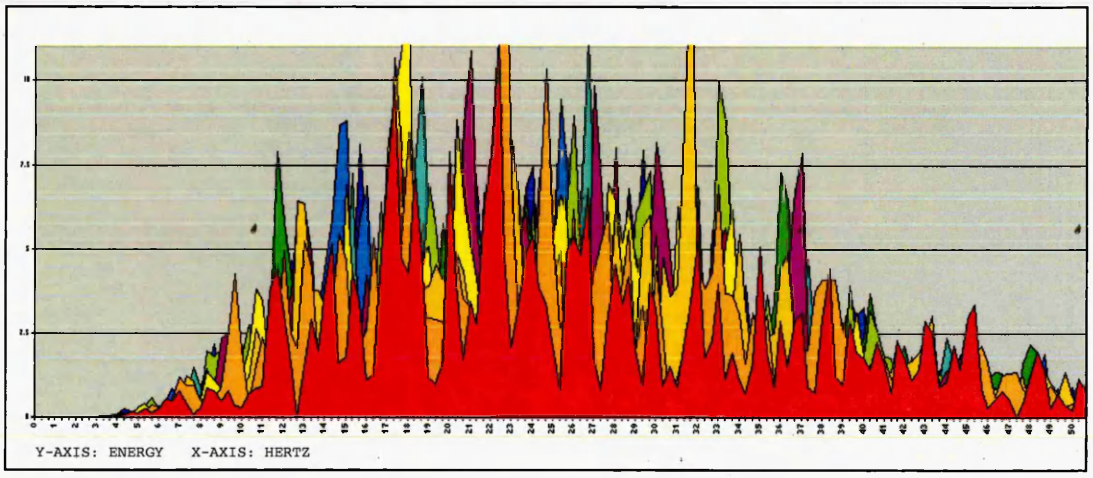

Fig. 3.38 Power spectra of various stints from experiment BC_W_2 with window size 65536 (Bacillus subtilis in water)

At a first glance both diagrams seem to be quite different, but mathematically the amount of energy and its distribution under both curves is identical. The difference in appearance of both energy signals is due to applying two different window parameters: Whereas the first diagram, using a window size of 4096, appears to be quite flat and smooth, the second diagram with a window of the size 65536 appears to have high and spiky energy peaks with a high fluctuation. The reason for this is that the second transform, due to a 16 times higher window parameter, results in a 16 times higher resolution, which means that for every data point calculated in the 4096-window-transform there are 16 data points in the 65536-window- transform. In other words, the steps in Hertz from one data point to

the next are 5.383 Hz for the low resolution transform and 0.336 Hz for the high resolution transform. This means that in the low resolution diagram (with the window size set to 4096) a single data point shows the average energy of 16 seperately calculated data points (or energy fluctuations) as displayed in the diagram with a high resolution (using a window size of 65536).

By nature of the underlying mathematical process, the highest frequency which can be calculated with a Fourier Transform is half the frequency of the original signal in the time domain, i.e. 11025 Hz in the current context, as the experimental data are digitized with 22050 Hz. However, as the signal is filtered above 300 Hz, the display of the energy signal can be sensibly omitted for higher frequencies. In fact, as the Fourier analysis of all experiments shows, there is no (or hardly any) energy above 100 Hz. Consequently, for diagrams with a window size of 4096 only the first 40 data points are displayed; diagrams with a window size of 65536 display the first 600 data points; this means that the display of both types of diagrams is cut off at about 200 Hz.

3.8.3.2 Principal Component Analysis on Power Spectra

Principal component analysis (PCA) is an eigenvector-based multivariate analysis and serves, amongst other disciplines, as a standard tool to generate patterns. The exact working principle is well outlined, e.g. by Wikipedia (2007c).

PCA has already been successfully applied in recognizing particular strains of Escherichia coli by Bayraktar et al. (2006) where bacteria grow on agar, with a system setup utilizing a diode laser as excitation source and a CCD camera as sensor. In conjunction with the OBS system, the experimental data have to be pre-processed prior to be analyzed by PCA. Accordingly, a Fourier Transform is applied to 4 arbitrary stints from each experiment respectively, and thus generates (characteristical) power spectra for every organism. The power spectra are then organized into training and test sets, which includes, the removal of outliers, filtering, the definition of events characterizing a particular organism, and a frequency analysis of events[10]. Once these steps are completed, the thus processed data are then used for PCA-based classification.

The first step of data pre-processing is to remove points at the start and the end of the wave in order to minimize undesirable start-up effects which may be caused by outliers:

```
read WAV file
remove first and last 250 samples to remove start up noise
```

Then, the wave is filtered using the *filtfilt* function, which zero-phase digital filters both in the forward and reverse directions. By default, the *filtfilt* function minimizes start-up

[10] The algorithms for processing and analyzing data are implemented using MATLAB, a high-level scientific programming language. Information about the in-built functions used in the course of this description can be found at http://www.mathworks.com [17-07.2008]

transients by adjusting initial conditions to match the DC component of the signal and by pre-pending several filter lengths of a flipped, reflected copy of the signal:

```
apply filtfilt with step 250
```

The results of such operations are depicted in figure 3.38.

The next step in the data pre-processing is to define an event that characterizes an organism. Since a stint has 250000 data points in total (equal to 11.3 seconds of an experiment) this lends itself to a large number of possibilities. The definition of an event offered here is governed by the time it takes for an organism to cross the line of laser. The collection of such data readings then would characterize that organism.

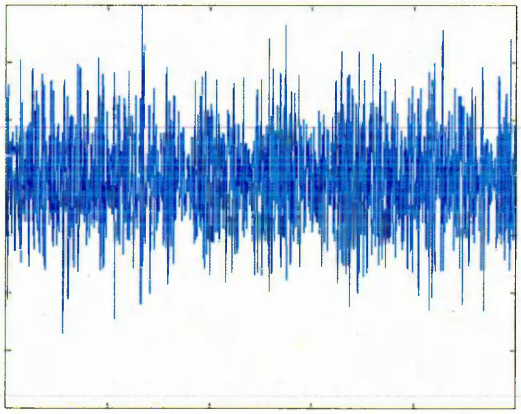

Fig. 3.38 A filtered stint (Escherichia coli)

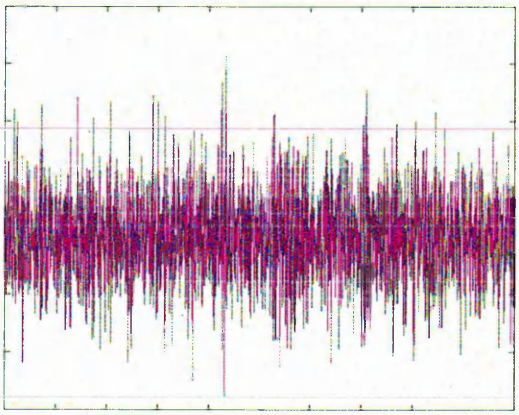

Fig. 3.39 An event comprised of 1000 samples (Escherichia coli)

Considering that unaided bacteria would move at low frequencies close to 10 Hz, and if we are interested in capturing motion characteristics, then the data collection must be at a sampling frequency that is greater than twice this frequency (according to Nyquist's sampling theorem). This has been substantially exceeded with a sampling frequency of 22050 Hz. Each original stint of 250000 samples generates 250 sub-stints, or events, each comprising 1000 samples (see Figure 3.39). These are further organized into a data matrix:

```
for X = 1 to L in steps of S = 250
    build a matrix of X / S columns, sampling each point displaced by step S of
    the original signal
end
```

Once the data are organized into such a matrix, a Fourier Transform is applied to each row of the data using the *pwelch* function[11]:

[11] Power Spectral Density using Welch's method

```
for each row of data matrix
    calculate pwelch with window size of 128
end
```

Note that the output of *pwelch* is consistent across all rows. This procedure is repeated for 4 stints out of every experiment, such that the resulting set of curves can be used to characterize a particular species from any given experiment (see Figures 3.40 to 3.43).

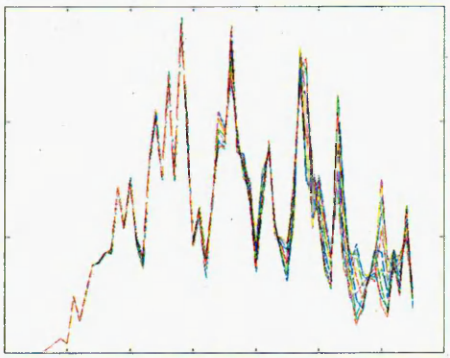

Fig. 3.40 'Pwelch' (1 out of 4)

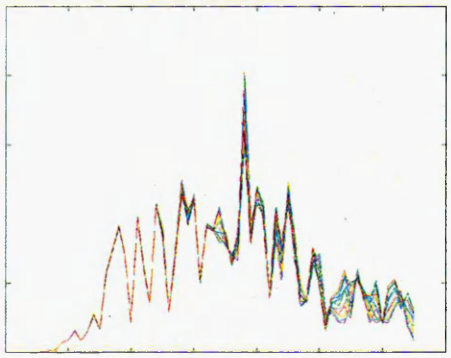

Fig. 3.41 'Pwelch' (2 out of 4)

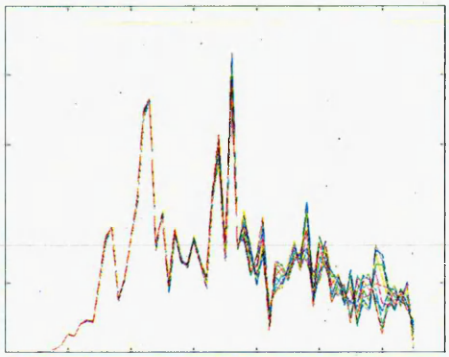

Fig. 3.42 'Pwelch' (3 out of 4)

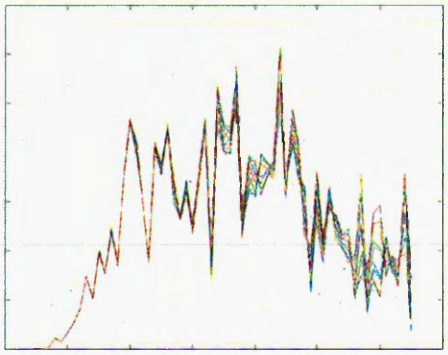

Fig. 3.43 'Pwelch' (4 out of 4)

After the experimental data are pre-processed for all organisms, these are concatenated into a matrix of data D, where each column represents the data of a species. D is the knowledge database, or training set, for the classification system, to which a 'test vector' t is related to in order to measure its affinity to an event. There are 250 'power spectral density columns' per sub-stint, otherwise 1000 columns per organism. The average column vector V in matrix D is estimated, and then each column of D is subtraced from the average, resulting in the scatter matrix A:

$$A = D - V \tag{3.3}$$

This PCA method applies Eigen Analysis to the covariance matrix of the data C, which measures the tendency of two vectors to vary, together for each combination of two vectors in the data set. The covariance matrix C can be obtained by estimating

$$C = A * A^T \tag{3.4}$$

Once the eigenvectors of the matrix C are evaluated, the associated eigenvalues are used to rank the eigenvectors by their usefulness in characterizing the variation amongst the event vectors, by chosing the eigenvector with the highest eigenvalue as the most useful

one, and ranking the others accordingly. The eigenvectors and corresponding eigenvalues of C are estimated using the covariance method. The subset M of eigenvectors, corresponding to the largest eigenvalues, can be used to accurately describe all input data as linear combinations of all event vectors in the training set. The coefficients in these linear combinations are called the weights W, and the 5 largest eigenvectors are used to build the weight matrix.

In order to obtain the actual weights W_k for any vector k in D, D it is multiplied by the matrix of weights W, which then represents the actual basis for the classification: It needs to be determined which vector in the training set is most similar to a test vector t. Thus, the weight vectors for all known vectors in D, as well as for the test vector t, are calculated. Then the best match is the vector whose weight vector has the smallest distance from the weight vector to the unknown test vector, i.e. the V_k which minimizes the distance measure:

$$V_k \rightarrow \| W - W_k \|_{min} \tag{3.5}$$

The method described above chooses the vector in the training set with the shortest distance from the test vector. By looking at the identity of the thus chosen vector, the assumption is made that the test vector has the same characteristics, i.e. t is an event from the same group. After ranking, the 4 closest vectors can be selected by measuring the shortest distances or by setting a minimum threshold to decide whether the test vector is classified into a particular group or not.

Two decision rules are used based on the 4 ranked (shortest) distances. Instead of considering only the shortest distance as a parameter for deciding class membership, two alternative rules are used:

1 out of 4: if one of the ranked distances points to the identity of the organism; this is equivalent to using a threshold for class membership in which distances up to the 4th ranked distance are considered within the threshold;

2 out of 4: this is a stricter rule, as two of the ranked distances must point to the identity of the organism before it is considered to be a correct classification.

Experimental results using these rules are reported in the next section.

3.9 Results and Discussion

The analysis of the experimental data largely comprised the generation of patterns or diagrams. The diagrams were then manually (optically and logically) analyzed in order to prove the existence of characteristic differences. The latter were noticeable, however, only to a minor degree and on an irregular basis, so that no algorithms could be produced to exploit these for an automated classification. This affected all analyzing methods apart from the Principal Component Analysis (3.8.3.2), where a classification algorithm was applied with reasonable success.

3.9.1 Inconspicuousness and Inconsistency of the basic Features

When examining the diagrams depicting the basic statistical features of an experiment (see Section 3.8.2.2), one can see in figures B.1 to B.6 that, on average, Proteus mirabilis exhibits a lower number of samples per wave as both Bacillus subtilis and Escherichia coli, which logically results in a higher number of waves per experiment when compared to the other two organisms. Bacillus subtilis has the largest and most prominent spikes (number of samples per wave), with lasting interludes of almost non-existing waves. Escherichia coli may be interpreted as incorporating a mixture of the features of the other two microorganisms. Overall on these diagrams it shows that feature c) (number of slope changes per wave) does not play a characterizing role, although it can be argued that this feature has a slightly higher markedness on Bacillus subtilis diagrams as compared to the other two.

The display of feature c) on the condensed (and simplified) 'machine code' diagrams (see Figures B.7 to B.42) shows that it actually contributes to the detection of diversity between different experiments: Some patterns appear to be quite different than others, but unfortunately this does not only occur between different species but is also true for data of identical species (e.g. Figure B.19 and Figure B.11 compared to Figure B.17 and Figure B.27). The same is true for those figures where feature c) and b) are set in relation to each other; there are remarkable pattern differences, however, found throughout the whole range of experimental data regardless of species. So, e.g. for Bacillus subtilis, in Figure B.16 and Figure B.8 those diagrams exhibit a dense, long lasting column pattern, whereas in Figure B.30 and in Figure B.14 the characteristical pattern is short with sparse columns.

3.9.2 'Fingerprint' as the best Graphical Representation of Statistical Data

Probably the best representation of statistically analyzing the experimental data is the 'fingerprint pattern' (see Figures B.78 to B.95), which categorizes features b) and c) and adds another feature d) (numbers of consecutive large waves within a stint) to result in a diagram which contains more statistical information at a glance.

Compared to Proteus mirabilis, the combined representation of features b) and c) for Bacillus subtilis is more balanced between *high*, *med*, and *low*, with the highest column being *med*; the same pattern for Proteus mirabilis clearly shows an upwards trend (from left to right), sometimes with the highest column (*low*) weighing twice as much as the average, balanced value of the pattern for Bacillus subtilis. But here again the representation fails to be consistent, such that patterns from experiments PM_W_1 and PM_R_2 (Proteus) are almost identical to the patterns from experiments BS_W_3 and BS_R_3 (Bacillus). Nevertheless, when viewed at a glance, one is able to distinguish these two species. Like previously expressed, Escherichia coli seems to be a mixture of the other two species: These patterns exhibit the upwards trend of Proteus mirabilis, but softened to give a more balanced impression, i.e. the differences between *high*, *med*, and *low* are not so great as for Proteus mirabilis.

The representation of feature d) cannot be interpreted so explicitly. For Proteus mirabilis, this feature seems to result in a longer 'wearing off' period of its curve, although that is not always the case. By and large, it can be compared with the feature's representations of Bacillus subtilis, where the curves slope seems to wear off slightly faster. For

Escherichia coli, this feature is not uniform, with two instances (EC_R_2 and EC_R_3) declining very rapidly compared to the other three.

3.9.3 Line Generalization shows no statistical Differences between Waves

Although line generalization has a remarkable and consistent graphical effect on the wave curves of every experiment, its usefulness for a statistical analysis proves to be very limited. The evaluation of the resulting elimination of points from a wave is the only statistical asset achieved by applying this method. The assumption that different species could give rise to wave patterns which may have different characteristics that could be exploited or visualized by a line simplification process seems invain, at least for all experiments carried out in this research project: The percentage of eliminated points as opposed to the total number of points of a wave from all waves of an experiment is, on average, equal for every experiment, hence also for every species.

3.9.4 Power Spectrum reveals a Classification Clue

As the total number of cells in a bacteria sample varies for every experiment, sometimes quite drastically, the energy level of the experimental recordings varies with the same intensity. The direct result, when applying a Fourier Transform to experimental data, are power spectra with different amounts of total energy. These cannot be compared by evaluating the energy level at certain frequencies, but by evaluating how the energy is overall distributed across the frequency range (regardless of the signal strength).

Consequently, the diagrams depicting the power spectra for each experiment (see Figures B.43 to B.77) are inherently different. However, focusing on the course of the energy curve reveals consistent changes for Bacillus subtilis and Proteus mirabilis respectively. For Proteus mirabilis the relative decline of the curve is much slower than the decline for Bacillus subtilis; in other words, whereas for Bacillus subtilis the amount of energy accumulates at lower frequencies and stagnates quickly towards the higher frequencies, for Proteus mirabilis the higher frequencies have more energy content as the decline of the energy (the curve) drags on for much longer. This is logically undermined due to the fact that Proteus mirabilis moves much faster than Bacillus subtilis. An extremely explicit example of this behaviour is the energy curve from experiment PM_W_3 (Proteus), which stands out against every other power spectrum of the two organisms.

3.9.5 Benefits and Constraints of the Principal Component Analysis

In order to classify the three monitored species the data of the experiments are separated into training and test sets. The method used here is a 'leave-one-out' cross-validation and the results obtained are summarized in tables 3.44 to 3.47. The analysis of variance, provided by the PCA algorithm, yields a classification by ranking distances: The shorter the distance, the more similar two curves are. The implemented code ranks the 4 closest samples to any given (unknown) test sample.

However, by knowing the true identity of the test sample, it is possible to reason about the proposed class suggested by the algorithm. For instance, a 'best of 4' algorithm can be applied to decide whether two species are identical. In the results reported below, if 2 or

more proposed classes point to the real class of the species, then this is counted as a correct classification, otherwise a mis-classification is recorded ('2 out of 4').

Table 3.44 shows the classification rates for each of the three species.

	Proteus mirabilis	Bacillus subtilis	Escherichia coli
total number of vectors	6000	6000	6000
correctly classified (%)	96	100	83
misclassified (%)	4	0	17
overall recognition rate (%)	93		

Tab. 3.44 Classification results, decision rule 2 out of 4

Table 3.45 shows the number of correctly classified species and the number of species which are classified into a wrong class, cross-referenced with the purpose of identifying the rate of false classifications for every species. By looking at a row and checking against the respective column for that organism, one may get an impression as to how well the respective species is recognized. For example, for Bacillus subtilis it can be seen that, per 100 samples tested, 92 are classified correctly, only 4 are classified incorrectly into the Proteus mirabilis class, are classified as 'unknown', which means do not belong to any class.

	Unknown	Proteus mirabilis	Bacillus subtilis	Escherichia coli
Proteus mirabilis	32	67	1	0
Bacillus subtilis	4	4	92	0
Escherichia coli	29	0	0	71

Tab. 3.45 Correct (bold) and false classifications cross-reference, decision rule 2 out of 4

Usually the priority for a classification system would be to keep the rate of 'unknown' classifications as low as possible. However, if the system is merely used to give an indication about a suspected microorganism and subsequent, more precise tests can be carried out (e.g. more labour intensive manual techniques) to approve or decline the indicated result, the system's parameters would not have to be changed and its operation would still be regarded as useful.

Furthermore, the analysis described here only counts correct and incorrect classifications without taking into account the distance measure (between the test vector and a vector in the knowledge database. If the distance is zero then there is a perfect match). The number of samples classified as 'unknown' can be reduced if the classification algorithm makes use of a threshold assigned to the distance measure. Consequently, if the distance for a species is within a certain threshold then it is considered as a correct classification. Sometimes, the application of a classification system demands that the identification of a certain species is critical, such as for pathogenic bacteria. In this case, by using the threshold, the system could be arranged such that even small distance measures flag the presence of a species and result in a possible correct classification. As mentioned before, the system may in such a case only be used to give an indication about the suspected

occurrence of a microorganism, which necessitates subsequent, more precise techniques to approve the indicated result.

As an example of such reasoning, if Escherichia coli were regarded as a critical species, it could be flagged by changing the decision rule from '2 out of 4' to '1 out of 4'. This would mean that from the 4 options given by the algorithm, whereas previously 2 needed to point to the same group of bacteria to rate this as a correct classification, now only 1 needs to point to the class of Escherichia coli bacteria to correctly classify a sample into this class, and the rate of correct classifications would hence increase.

Using this reasoning, the data of table 3.44 are re-calculated for all organisms yielding the recognition rates in table 3.46. Obviously, the down side of such reasoning is concerned with false classifications as per trend analysis discussed earlier. With a decision rule of '1 out of 4', false classifications almost certainly increase, and this may be to such a level that the system may become unworkable. Therefore, in order to increase correct recognition rates and eliminate false classifications as much as possible it is more advisable to work on the analysis of distance thresholds and incorporate this into a decision rule. However, this is an issue left to the requirements of a particular environment in which such a classification system would be used.

	Proteus mirabilis	Bacillus subtilis	Escherichia coli
total number of vectors	6000	6000	6000
correctly classified (%)	100	100	100
misclassified (%)	0	0	0
overall recognition rate (%)	100		

Tab. 3.46 Classification results, decision rule 1 out of 4

Further to the validation analysis provided above, the ability of the algorithm to recognize a species explicitly from *different* stints (i.e. the sample is not taken from the stint that it is compared to during the analysis) is tested. For the (strict) decision rule '1 out of 4', the results are summarized in table 3.47. In this situation, all correct classifications originating from the stint which the sample is taken from are not counted. Consequently, the overall results worsen.

	Proteus mirabilis	Bacillus subtilis	Escherichia coli
total number of vectors	6000	6000	6000
correctly classified (%)	33	67	83
misclassified (%)	67	33	17
average recognition rate (%)	61		

Tab. 3.47 Classification results from different stints, decision rule 1 out of 4

It is apparent from the analysis that it is possible to fine-tune a system to recognize specific organisms, such as Escherichia coli, with a reasonable accuracy level. The problem of a large percentage of mis-classifications still remains. The main reason for this seems to be that the scattering characteristics of some species do not remain constant for different experiments. This may be due to manually preparing samples for the experiments: For every experiment, several steps (inoculation from a swab, incubation

time, incubation medium, statistically counting numbers, dilution to final experimental sample) inevitably result in slight differences between each experimental sample, of which the most prominent one is the total number of cells. This has the consequence that the (overall) amplitude of the recorded signal varies for every experiment, thus providing the system with slightly inconsistent data recordings for the classification process.

4 Exciting Bacteria with Electromagnetic Power

4.1 Introduction to Dielectrophoresis

Conventional techniques to analyze aquatic microorganisms usually require a high level of manual labour input to prepare bacteria cells before these can be subsequently detected or identified (Brown et al. 1998). This may lead to a long time lag before results of the analysis are available. In contrast, the application of dielectrophoresis (DEP) is a more recent method, aiming towards the ambition of 'real-time' analytical microbiology, i.e. providing analytical results of a bacteria sample (extremely) rapidly.

The principle of DEP is to apply a nonuniform electric field to a suspension of microbial cells in order to induce a polarization of the cells, which then move towards the region of highest field strength (see Figures 4.1 and 4.2). By changing the field frequency the amount of particles collected at the electrodes will vary. Consequently, characteristic collection spectra for different bacteria cells can be obtained (see Figure 4.3).

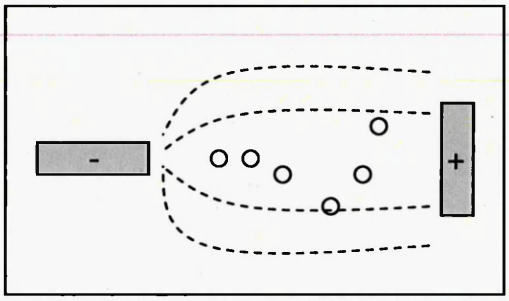

Fig. 4.1 The DEP force agitates cells to move towards the region of highest field strength

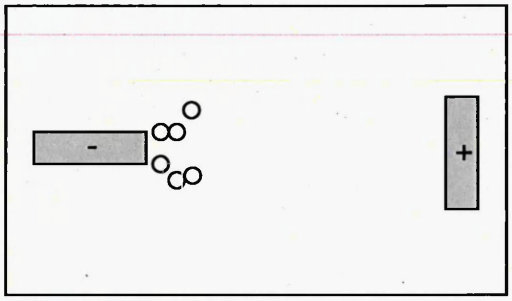

Fig. 4.2 After some time, all cells are captured at an electrode

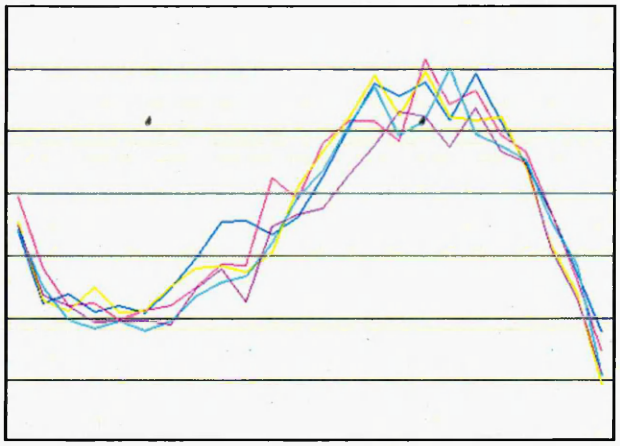

Fig. 4.3 Characteristical DEP collection spectrum of Escherichia coli 15669

DEP is based on a technique called electrophoresis, which has been first mentioned by Reuss (1809) and has been extensively exploited for investigating the surface properties

of cells by Coulter (1920). In electrophoresis, a direct current (D.C.) electric field of uniform intensity is applied to particles suspended in an aqueous medium, which may lead to three effects: 1) the motion of particles, 2) the distortion of the electrical double layer surrounding the particles, 3) electrical charges are induced within the particles' structure (IBMM and School of Electronic Engineering 2002).

While charged particles undergo electrophoretic motion in *uniform* electric fields, neutral particles remain stationary, as they have equal numbers of opposite charges. However, in a *nonuniform* field, as achieved with dieletrophoresis, an imbalance in force on the particles enable them to move toward the region of greatest field intensity such as an electrode (Brown et al. 1998).

The total electric force acting on a particle of net charge in a non-uniform field E is:

$$F = QE + \delta qE(r_+) - \delta qE(r_-) = QE + (m.\nabla).E \tag{4.1}$$

Where; ∇ is the Del vector operator. If the particle is uncharged (i.e. $Q = 0$) or for frequencies above around 1 kHz, where electrophoretic effects are negligible, the term on the right-hand side of equation 4.1 dominates, so that the time-averaged force is given by:

$$F(\omega) = \mathrm{Re}\{m(\omega)\}\,\nabla E^2 / 2E \tag{4.2}$$

Where; Re denotes the real component of the dipole moment. According to Pohl (1958, 1978), this is called *dielectrophoresis* (IBMM and School of Electronic Engineering 2002).

Brown et al. (1998) describes that the dielectrophoretic force is different for charged and neutral particles respectively. Whereas a charged particle reacts by changing its direction of motion each time the field is reversed, a neutral particle, governed by its dielectric properties and those of the suspending medium, shows continuous movement toward regions of increasing field intensity, as demonstrated in equation 4.3:

$$F = 2\pi a^3 \varepsilon_m \,\mathrm{Re}\left[\frac{\varepsilon^*_p - \varepsilon^*_m}{\varepsilon^*_p + 2\varepsilon^*_m}\right] \nabla |E|^2 \tag{4.3}$$

Where F is the dielectrophoretic force, a is the particle radius, E is the electric field (r.m.s.), [...] and ε^* is complex permittivity ($\varepsilon^* = \varepsilon' - \iota\sigma/\omega$) for particle and medium, and where σ is the conductivity and ω is the angular frequency (Brown et al. 1998).

4.2 Flow-Through Grid Electrode DEP System Overview

The heart of a DEP system is an arrangement of electrodes to which the cells to be monitored are exposed: This is the so-called 'dielectrophoretic chamber'; it consists of a pair of electrodes which are suspended in a test suspension. In conventional systems, the sample cells are passed over the electrodes where they are collected and then eluted. To maximize the dielectrophoretic collection, more sophisticated systems using more complex flow-through chambers have been constructed. In a so-called 'traveling-wave dielectrophoretic device', microorganisms are propelled through a stationary supporting fluid whilst high-frequency electrical signals are applied to alternate pairs of electrodes (Pimbley et al. 1999). The result is a concentration of cells at the electrodes, usually in a 'pearl chain' formation (Betts and Brown 1999). The motion of the cells and the concentration effect is governed by a function or signal generator which is used to apply the electric voltages. The cells are monitored with a microscope to which computer hard- and software is linked through a digital camera, in order to enable an image analysis program to investigate bacteria formation patterns and to count the number of cells collected at the electrodes.

The system described here is used for the identification of DEP collection spectra, each of which is being generated by applying electric fields at a range of frequencies to a test sample. The test sample contains a particular strain of a bacteria species, which needs to be identified by the system (see Figure 4.4). A commercial image analysis software package (Domino, Perceptive Instruments) counts the number of bacteria cells of the test sample. These numbers are processed to result in a pattern, i.e. a DEP frequency spectrum (see Figure 4.3), for each test sample. Then, the system's algorithms analyze the spectrum in order to identify the bacteria species found in the test sample.

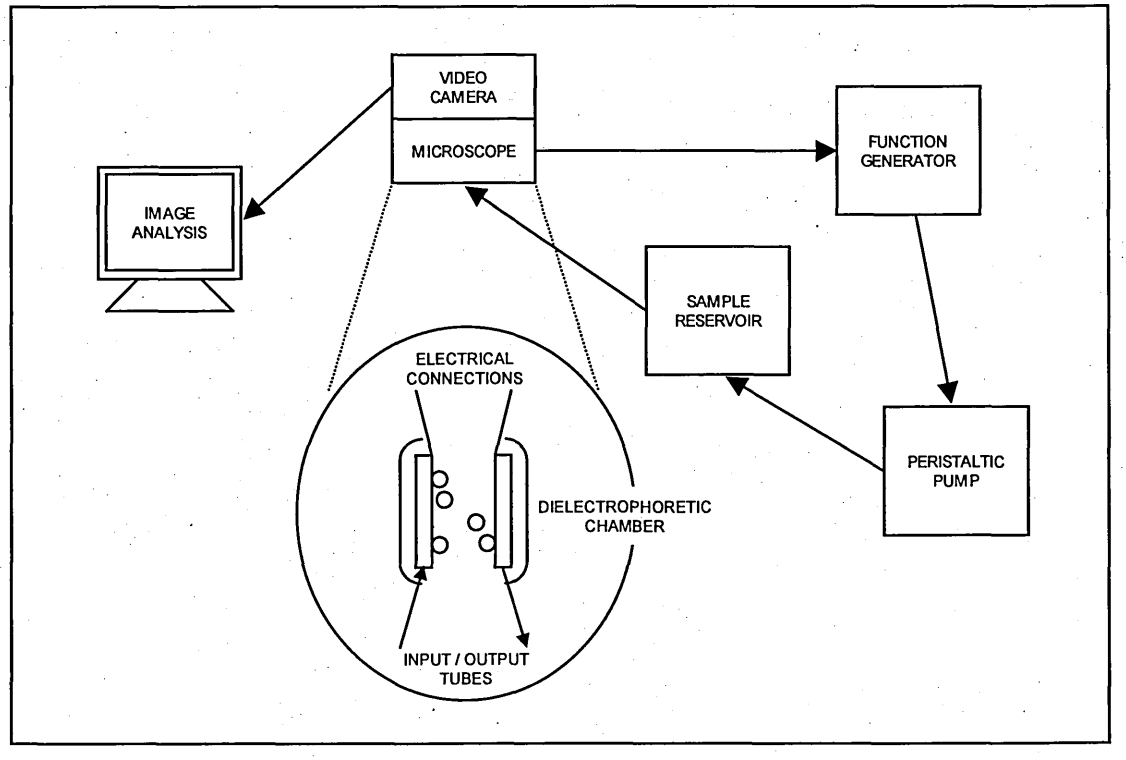

Fig. 4.4 A typical flow-through DEP system (Pimbley et al 1999)

In order to identify the cells present in the test sample, the system accesses a database storing DEP spectra which have already been identified and hence have been categorized into a number of so-called 'bacteria classes'. When the system is activated for the first time, the database is setup so as to contain a number of bacteria classes, each of which contains a few preselected example spectra. In the course of the system's usage more bacteria classes may be added to the database, enabling the system to distinguish between a wider variety of microorganisms.[12]

With the help of the information stored in the database, the analysis aims to identify the test sample, i.e. it seeks to determine the affinity of the test sample to one of the existing bacteria classes. Upon identifying a test sample, its DEP spectrum is added to the relevant bacteria class in the database. As the class access conditions of a bacteria class represent the characteristics of the spectra stored within that class, these have to be updated whenever a new DEP spectrum is added. Therefore, the system does not simply add to the database but also dynamically alters some of its values (see Figure 4.5). This is described in detail in section 4.4.

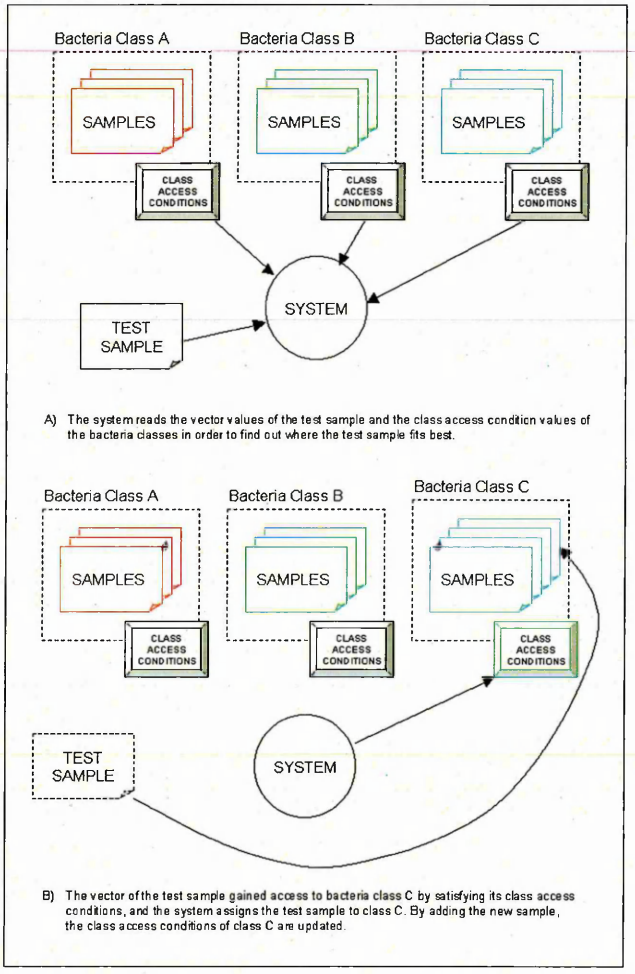

Fig. 4.5 High level working principle of the system

[12] Note that a bacteria class is always created manually and has to be setup so as to contain a few examples of DEP spectra valid for the bacteria class.

4.3 Description of the Experiments

All experiments are carried out using 9 different 'in vitro' strains of the species Escherichia coli, which have been chosen due to their scientific relevance and availability (see Section 3.7). The following Table 4.6 lists all bacterial strains used for the experiments and the amount of experiments carried out for each strain respectively:

Bacterial Strains	No of Experiments
Escherichia coli 32339	5
Escherichia coli 15669	5
Escherichia coli 09019	5
Escherichia coli 10960	5
Escherichia coli p1	5
Escherichia coli BH1	5
Escherichia coli BH2	5
Escherichia coli 12900	5
Escherichia coli 8114	20

Tab. 4.6 Available bacterial strains for the experiments

Brown et al. (1998) explain in detail how the samples are prepared for the experiments and the procedure to attain the dielectrophoretic spectra for every strain. Altogether 60 bacteria samples are processed. Each test sample results in a dielectrophoretic spectrum showing the amount of seperated cells at electric field frequencies between 10^3 and $4*10^7$ Hertz (see Figures in appendix C). At the start of the system's operation, its knowledge database is made up of training sets, or bacteria classes, of 4 sample members each, altogether comprising 9 bacteria classes (an exception is Escherichia coli 8114 where, due to a greater amount of available DEP spectra from the experiments, the training set comprises 10 samples). The system then tries to assign one of the remaining 18 test samples to one of the existing 9 bacteria classes (see Table 4.7). The entries in the knowledge database are several times rotated with the test samples so that numerous constellations exist to increase the amount of possible categorization attempts for the system.

Bacteria Class	Amount of DEP spectra in the knowledge database	Amount of DEP spectra as test samples
Escherichia coli 32339	4	1
Escherichia coli 15669	4	1
Escherichia coli 09019	4	1
Escherichia coli 10960	4	1
Escherichia coli p1	4	1
Escherichia coli BH1	4	1
Escherichia coli BH2	4	1
Escherichia coli 12900	4	1
Escherichia coli 8114	10	10

Tab. 4.7 All bacteria classes and the number of DEP spectra available to the experiments

4.4 Data Analysis

The DEP collection spectra (see Figures in appendix C) have to be processed by suitable algorithms in order to categorize them into different bacteria classes. The first objective is to transform the data gained by the DEP experiments into information which emphasizes their class characteristics. Then, the transformed data are processed in order to evaluate the affinity of the underlying DEP spectrum with each of the known bacteria classes.

4.4.1 Extracting Information

A regression technique is applied in order to abstract the data collected by a DEP experiment so as to yield the desired information, i.e. suitable characteristics which suggest the categorization of the data into a particular class. In other words, collection spectra from bacteria belonging to a common class will result in the same or at least similar characteristics during the abstraction process. Thus, the extracted information can serve as a bacteria class label for the underlying DEP spectrum (see Figure 4.8).

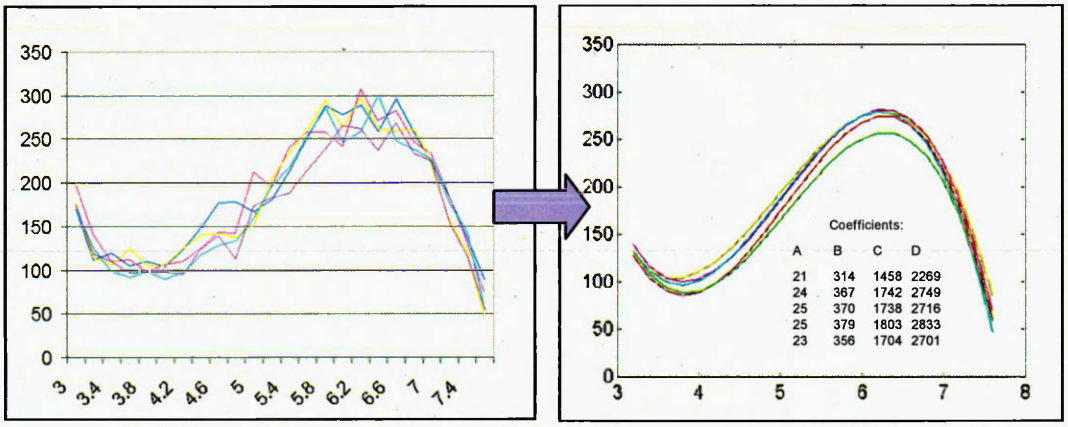

Fig. 4.8 Abstraxtion of DEP spectra into bacteria class labels by means of regression (Escherichichia coli 15669)

4.4.2 Classifying Data using a Two-Step Pattern Matching Concept

A two-step algorithm is used to calculate the affinity a collection spectrum has with respect to a certain bacteria class. The method of the algorithm is to match a set of vectors to the bacteria class access conditions of several bacteria classes, and it is determined which set of vectors work best for a number of bacteria classes. According to this concept every bacteria class is guarded by two seperate access conditions. Each step of the algorithm generates a unique vector using the bacteria class label of a test sample, and the resulting pair of vectors is tried to match the two access conditions of each of the known bacteria classes.

The first vector (α) is represented by the combination of polynomial coefficients which result from the regression of a collection spectrum. The second vector (β) is made up of an array of function values representing the polynomial curve resulting from the same regression. Vector α is used on class access condition A and vector β operates class access condition B.

4.4.2.1 Generation of Vector α using Four Polynomial Coefficients

The least-squares curve fit to a DEP collection spectrum, using a 4th order polynomial, which has been found to resemble the data best, yields four coefficients. These coefficients may be used to indicate whether the association of them with certain spectral patterns or bacteria classes is likely or not. Together, the coefficients form a four-dimensional vector such that each coefficient value represents an element of the vector. A certain combination of elements will gain access to one or more bacteria classes. Also, different combinations of elements might gain access to the same bacteria class (see Figure 4.12).

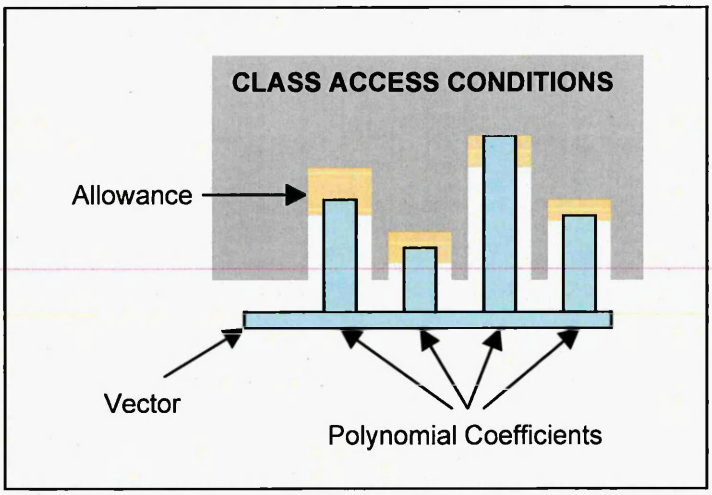

Fig. 4.9 The first test sample vector generated by the four polynomial coefficients of a bacteria class label

Access to one of the bacteria classes is granted if the characteristics of vector α (i.e. its elements or polynomial coefficients) match the characteristics of the respective bacteria class access conditions A to a certain degree. An access condition A is made up of four values each of which represents the average of the n-th coefficient (n=4) of all polynomials (i.e. abstracted DEP spectra) assembled within a bacteria class. Accordingly, a vector matches best if its four polynomial coefficients are identical to the four average polynomial coefficients of the class access condition A. However, the average values of A should only serve as an indicator, as to what would be a suitable vector, allowing several different combinations of polynomial values to successfully match the access condition A. Therefore, an allowance exists for every element of A, which allows vector element which are slightly different from the average elements (best-fit) of the respective elements of class access condition A to be successfully used.

The allowance of a bacteria class access condition A is a concept which adheres to the maximum likelihood rule for classification, as described by ESRI (2010). It creates a frame around the average polynomial values, within which the polynomial values of vector α of a test sample are valid. According to this concept, the individual values of the polynomial coefficients of the members of a bacteria class, which altogether construct the four average polynomial values, one for each of the four polynomial coefficients of every sample within that class, deviate from the average values more or less drastically. The greater the overall deviation, the greater is the allowance for a deviation of the values of

the polynomial coefficients of a potential test sample from the average polynomial values of the bacteria class. The mathematical definition of the allowance is

$$apc_n = \left(\sum_{m=1}^{z} pc_{n,m} \right) / z \tag{4.4}$$

$$ap_n = \left[\sum_{m=1}^{z} \left| \left(100 - \left(100 / apc_n * \left| \left(pc_{n.m} \right) \right| \right) \right) \right| \right] / z \tag{4.5}$$

$$P = \left(\sum_{n=1}^{k} ap_n \right) / k \tag{4.6}$$

$$\text{allowance} = c * P \tag{4.7}$$

where apc_n is the n-th average polynomial coefficient of all members assembled within a bacteria class, z is the total number of samples within the class, and $pc_{n,m}$ is the n-th polynomial coefficient of the m-th sample in the class. Therefore, ap_n is defined to be the average deviation in percent of the n-th polynomial coefficients of all members assembled within a bacteria class from the n-th average polynomial coefficient of the members within that class; k is the total number of coefficients for the type of polynomial function used throughout the system (i.e. 4 in the case of a 4th order function). Therefore, P is defined to be the mean average deviation of all four coefficients in percent; c is a constant which serves as a threshold to define the range of the allowance in units of P. The experiments with 9 different bacteria classes show that in this case $c = 3$ is adequate.

The concept of allowance serves two purposes. First of all, it is necessary in order to enable several test samples with different polynomial coefficients, yet inhibiting similar characteristics, to join a common bacteria class. In addition to that, the property included in the allowance, which reflects the overall amount of deviation from the average n-th coefficient value of all test samples, provides an effective criterium to choose the correct bacteria class for a particular group of coefficients. Thus, in cases where the coefficient values of a test sample are found to lie more or less exactly between the values of two bacteria classes, it is more likely that the test sample belongs to the class with the higher average deviation, as more of its members, as opposed to the class with a lower average deviation, will have polynomial coefficients with values ranging closer to the values of the test sample.

This is also true even if the difference between a test sample's polynomial coefficient values (α) and the A-access condition values of bacteria class 1 is less than between the test sample and bacteria class 2, given that the allowance due to the high average deviation of the polynomial coefficients of class 2 is so much larger than the allowance of class 1, where the average deviation of the polynomial coefficients of its members is relatively small (see Figure 4.10).

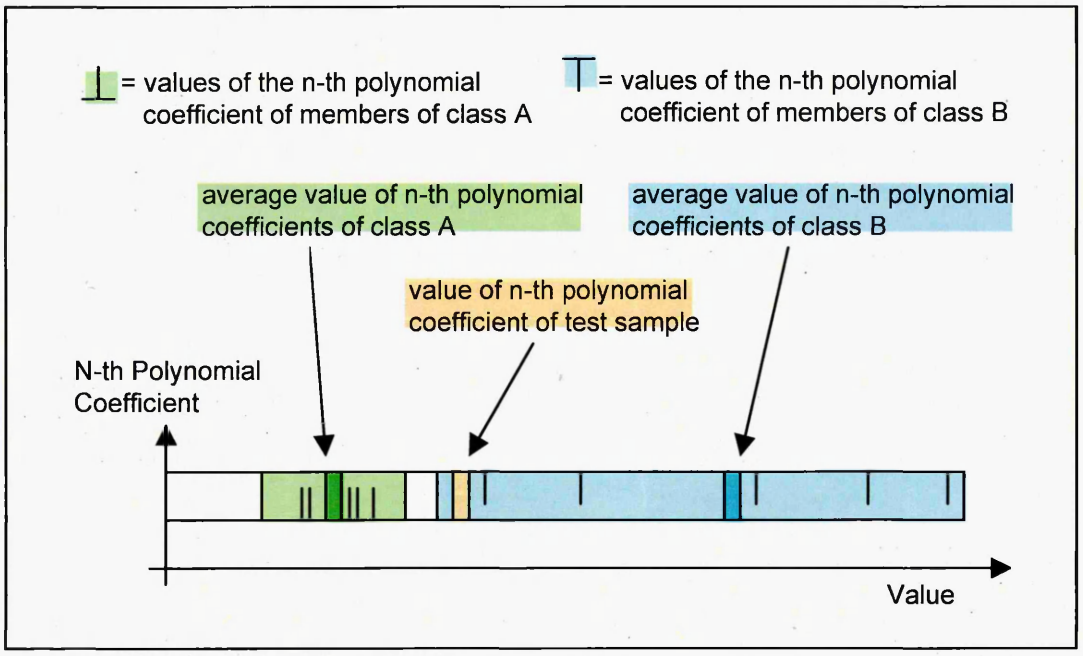

Fig. 4.10 The high deviation of the values of the polynomial coefficients of the members of bacteria class B attracts a test sample despite the values of its polynomial coefficients are situated much nearer to the average values of bacteria class A, however, class A's allowance is smaller as the values of its members do not deviate that much (ESRI 2010)

A formal description of the algorithm is as follows.

```
FUNCTION CALCULATE-ALLOWANCE:

k = ( number of polynomial coefficients of a regression )
m = ( number of members within a class )

FOR x = 1 TO k

    FOR y = 1 TO m
        INCREASE apc[x] BY pc[x,y]
    END-FOR
    apc[x] = apc[x] / m

    ( ∴ apc[k] = average of all kth coefficients within a class )

    FOR y = 1 TO m
        ap[x] = INCREASE ap[x] BY
                  abs[ 100 - ( 100 / apc[x] * abs( pc[x,y] ) ) ]
    END-FOR
    ap[x] = ap[x] / m

    ( ∴ ap[k] = average deviation in percent of all kth
                  coefficients from apc )

    INCREASE P BY ap[x]

END-FOR

P = P / k

allowance = c * P

FUNCTION GRANT-ACCESS:
```

```
kpc = ( polynomial coefficient of key α)

FOR x = 1 TO k
    kp[x] = abs[ 100 - ( 100 / apc[x] * abs( kpc[x] ) ) ]
END-FOR

( ∴ kp[k] = average deviation in percent of the kth coefficient
     from kpc )

FOR x = 1 TO k
    INCREASE test BY kp[x]
END-FOR

test = test / k

IF test < allowance THEN grant access
```

The polynomial coefficients represent the first vector (α) in the two-stage concept. It may successfully match a number of class access conditions A whereas it might fail for several others. This leads to a preselection of possible class candidates which are further analyzed by the second step of the algorithm.

4.4.2.2 Generation of Vector β using a Candidate Pattern

The second vector in the two-stage concept can be described as the fingerprint or individual pattern of the test sample. The elements of vector β are equal to the test sample's (or candidate's) individual pattern c(ω) which is formed out of a range of function values[13] of the polynomial representing the DEP spectrum of the test sample. On trying to gain access to a bacteria class, the candidate pattern c(ω) is compared to the group of corresponding patterns of the bacteria class labels which are contained in that class. This group of polynomial functions represents the second class access condition B of a bacteria class, and its curves are to be imitated in order to perform a successful match. Doing so, the comparison yields the degree of resemblance of the candidate pattern with the patterns of a particular bacteria class. The higher the resemblance, the higher is the probability that the vector will be successful in matching the second of the bacteria class access conditions, and that the test sample will thus become a member of that class (see Figure 4.11).

[13] In the available experiments, DEP measurements are only taken at certain frequencies between 10^3 and $4*10^7$ Hertz. Consequently, the function c(ω) resulting in the candidate pattern is only defined for log(ω) = { 3.2, 3.4, 3.6 ... 7.6 }.

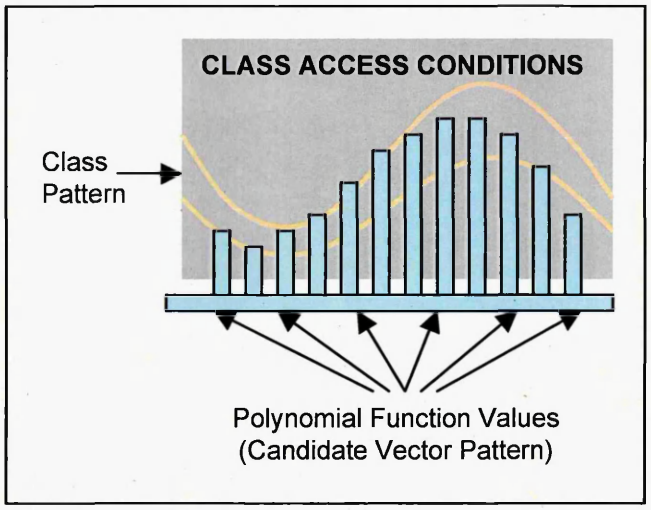

Fig. 4.11 The second vector generated by the polynomial function values of a bacteria class label

As for the identification process, all patterns of a single bacteria class are layered on top of each other so that two functions min(ω) and max(ω) are generated representing the lower border line (lowest collection of micro-organisms) and the upper border line (highest collection of micro-organisms) respectively. When a candidate pattern is compared to the function layer, its degree of resemblance can be expressed by the sum *D* of two values, namely *lo* and *hi*. The *lo* value is equal to the accumulated difference of the values of the candidate function c(ω) and the function min(ω) in cases where c(ω) < min(ω), i.e. in areas where the curve of the candidate pattern falls below the lower border line of the layered pattern of the bacteria class. The *hi* value equals the accumulated difference of the values of the candidate function c(ω) and max(ω) in cases where c(ω) > max(ω), i.e. in areas where the curve of the candidate pattern overshoots the pattern layer of the bacteria class. The sum *D* of *lo* and *hi* thus reflects the degree of failure to stay within the limits min(ω) and max(ω) of the function layer, such that minimizing *D* results in a high resemblance of the candidate pattern with the function layer of a bacteria class, and a large sum *D* stands for a low resemblance (→ figure 4.16).

However, for bacteria classes with a large value *P*, i.e. those having a high average deviation of the coefficients of the polynomials within a class from the average coefficients of that class, the distance ΔF (max(ω) - min(ω)) between the two border lines of the pattern layer is large. In fact, *P* is proportional to ΔF. Consequently, the probability that any polynomial function lies mainly between min(ω) and max(ω) is high, and therefore a high degree of resemblance is likely. In contrast, when a bacteria class is comprised of a compact layer of functions with min(ω) and max(ω) forming a narrow corridor, i.e. for classes with a low mean average deviation *P* and hence a small distance ΔF, the probability that a test sample's polynomial is plotted mainly between min(ω) and max(ω), resulting in a high degree of resemblance, is rather low. Thus, given a high mean average deviation, not only the second access condition B is easier to open with any kind of pattern β, but also, due to the high mean average deviation resulting in a large allowance, the first class access condition A had already been loose in its constraints for granting access to key α. In addition to this, classes which inhibit a high mean average deviation *P*, allowing a high resemblance to be achieved with many different kinds of candidate patterns, will attract more test samples than other classes with more restricted access constraints, and will thus further raise *P*, which leads to an undesirable chain reaction.

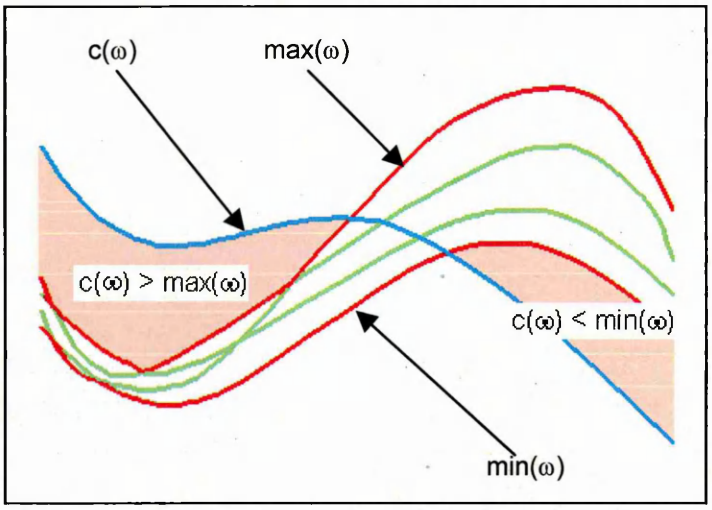

Fig. 4.12 A function layer, comprising 4 polynomial functions and the functions min(w) and max(w), is used to identify the degree of resemblance of candidate function c(w) with the layered functions

Thus, without further manipulating the sum D, a bacteria class, which already contains several diverse patterns and therefore a large distance ΔF, might in the long run act like a magnet, in that it also attracts test samples whose first vector succeeded in matching a few other A-access conditions with possibly more rigid constraints, and whose characteristics would make other classifications likely. This is taken into account by multiplying D with P, the result being the final reciprocal degree of resemblance.

A formal description of the algorithm is as follows.

```
FOR all member functions of a bacteria class
 (sampled at log(ω) = 3.2 to 7.6 in steps of 0.2)
    min(ω) = FIND lowest function value for every ω
    max(ω) = FIND highest function value for every ω
END-FOR

FOR log(ω) = 3.2 TO 7.6 IN STEPS OF 0.2
    IF c(ω) < min(ω) THEN INCREASE lo BY ( min(ω) - c(ω) )
    IF c(ω) > max(ω) THEN INCREASE hi BY ( c(ω) - max(ω) )
END-FOR

D = lo + hi

resemblance = 1 / ( D * P )
```

In order to uniquely determine to which of the known bacteria classes the test sample belongs, its second vector β can only successfully match a single class access condition B, belonging to the bacteria class whose member patterns, when being compared with the candidate pattern, have the highest degree of resemblance. However, in cases where a unique solution is not required, one could define a minimal degree of resemblance which has to be attained in order to match a bacteria class, which would make a number of classifications possible.

4.5 Results and Discussion

In a series of tests the system's algorithms showed a successful categorization of the entire set of test samples into their respective bacteria classes. However, experimental data had been relatively scarce, and further test data have to be processed in order to prove the system's ability to function with more numerous bacteria classes and large amounts of samples to be classified.

18 test samples were analyzed in a single system run, and several system runs were performed, rotating the test sample of every respective bacterial strain, such that every sample out of every bacteria class was tested.

Each test sample was represented by a dielectrophoretic spectrum showing the amount of seperated bacteria at electric field frequencies between 10^3 and $4*10^7$ Hertz. The spectra were then successfully identified by generating their bacteria class labels and applying the system's algorithms to compare them with members from the 9 bacteria classes available to the system.

4.5.1 Constraints of the Algorithms

4.5.1.1 Number of Cells in a Sample needs to remain constant

Making use of bacteria class labels to represent a particular bacterial strain involves the constraint that the number of cells must, to a certain degree, remain constant for every sample. In order to generate bacteria classes, and to classify a bacterial sample into one of a number of these classes, the algorithms are dependent on dealing with discrete numbers, and these cannot be normalized while, at the same time, preserving the functionality of the algorithms. In other words, the classification task fails if the amount of cells in a sample varies between samples.

4.5.1.2 Increasing the Number of processed Samples weakens the System

If the system is to be enhanced to distinguish between a greater number of bacterial strains, and / or if other microorganisms in addition to Escherichia coli are to be classified by the system, this would lead to an increase in categorization classes to an extent where the 'border lines' of bacteria classes may overlap for some classes, thus joining classes of different bacteria together.

In addition to this, a prolonged system operation may result in stretching the border lines: When processing a very high number of samples, despite operating with a limited number of bacteria classes, classes may become distended with every additional sample, as the properties of all samples within a bacteria class govern the extension of its border lines.

However, for a practical system operation, it is not necessarily required to inspect a wide range of existing bacteria. According to Madigan et al. (2000), it is often sufficient to be able to identify only a limited number of microorganisms (see Section 2.4.4).

5 Conclusion

5.1 Achievements

The aim of this research, i.e. to construct a bacteria classification system around relatively simple hardware and to determine its capabilities has been achieved. The results of this undertaking have been, up to a certain degree, successful. The investigations of bacterial properties as being revealed by the application of the experimental hardware available to the system have led to the conclusion that, under the technological constraints, morphological and physiological characteristics are adequate for an optical analysis. Further, in a second approach, electromagnetic characteristics of bacterial cells have been explored for an analysis using dielectrophoresis.

A number of (sequenced) algorithms have been invented to successfully exploit these characteristics in form of electronic data gained by the system's hardware. For the analysis, the data have been statistically, graphically, and mathematically transformed, and for some of these patterns appropriate classification procedures have been established. When manually and optically analyzed, most of the patterns generated by the OBS system (see Chapter 3) hint at correctly distinguishing species. Although for an automated analysis the differences of some of these patterns are not obvious enough for a successful classification. On the contrary, the patterns generated by the DEP system (see Chapter 4), when being manually observed, seem to offer only little means for distinguishing the various samples of Escherichia coli, but when being processed by the analyzing algorithms, they altogether result in correct classifications. Similarly, for the data gained with the optical scattering method by the OBS system, the sequence of PCA algorithms delivers overall good classification results.

On the other hand, one may argue that the achievements of the system described here are to a degree insufficient: Much of the laser scattering method has led to the generation of patterns only, and automated classification procedures could have been ill-founded due to the differences of patterns being either not explicit enough or not consistently achieved whilst repeating experiments. The PCA analysis' grade of success would be greatly diminished if test samples would not be extracted directly from the knowledge database, but would be derived from experiments completely unknown to the classification system. In the case of DEP, the pool of test samples has been quite small, so that experimental data have been scarce with merely a single species used for the analysis, and it has become apparent that considerable differences between samples already exist at the strain type level. With an almost infinite number of strains of microorganisms existing, this means that such a very high number of samples would expand the system to an extent where the imaginary 'classification space' would become exhausted, so that only a limited number of classes could co-exist, or otherwise classes would overlap.

To improve the system's performance, the algorithms presented in this work should not be used in seperate attempts to classify the observed species, as every single one of them proves to be relatively limited in the revelation or exploitation of clues for a successful classification. In a different approach, the entirety of algorithms used in combination may have the potential to weighing up assumptions about a species to be determined, which ultimately may result in an acceptably low error rate for classification advises of a thus automated system. A suggestion of how this may be achieved by combining the two seperate methods as described in this work is outlined in Section 6.4 as a recommendation for future research. However, prior to the attempt to organize methods

and technology into a successful classification system, the study of bacteria themselves, more than the study of most other living organisms, poses a problem to any classification task.

Given the difficulty of the posed problem and in the absence of any other known comparable work, the evaluation of the results is inherently not straight forward. In any case, the work presents a promising first step in the direction of low-cost bacteria recognition methods and provides the biochemist with a valuable first entry into the structural aspects of the demonstrated techniques. This is further undermined by Keiderling and Pancoska (1993) stating that although information obtained from optical analysis often merely concerns the average, bio molecular structure of the analyzed objects, it provides guidance for later more detailed structural investigations.

5.2 Limitations

The nature of the task prohibited the use of advanced and costly equipment. It is indeed a question as to which extent the information content in the data, as provided by a device that merely consists of a single diode laser to excite the observed species and a number of photo detectors to sense the resulting light reflexions, is sufficient to spin a successful classification algorithm around it. It may be that one has to resort to more advanced data acquisition techniques to distinguish between patterns derived from different bacteria species, and to separate the wheat from the chaff, so to speak, i.e. the bacteria patterns from the background noise.

5.2.1 Low Resolution

In contrast to the initial underlying considerations regarding a desirable high level of diversity of the bacteria as experimental test objects, from the extensive description in "Bergey's Manual of Determinative Bacteriology" (Holt et al. 1994, see Section 3.6) it becomes apparent that, although entirely unique in three different genus classes, the three species chosen for the experiments (Escherichia coli, Proteus mirabilis, Bacillus subtilis) - on a less accurate level - inhibit quite similar morphological, physiological, and biochemical properties. In fact, as being realized during the data analysis (see Section 3.9), the technical level of the OBS system represents low cost equipment as opposed to more sophisticated state-of-the-art technology, and this results in a rather low level of laser accuracy. In fact, as demonstrated by Keiderling and Pancoska (1993), most optical spectroscopic approaches to biological structure are of inherently low resolution. The resulting experimental data lead to the consequence that the attempts being made to distinguish the species may be to a considerable extent unsatisfactory, e.g. for professional purposes. In order to cater for an overall low resolution of the experimental data, the manifestation of diverse characteristics would have to be improved. Therefore, the use of more diverse organisms is encouraged, however it has already been pointed out (see Section 3.6) that such bacterial cultures are quite often difficult to be obtained as they have to be purchased from international providers, which makes their acquisition expensive. On a different note, some species are difficult to be handled as they might be very fragile (and thus not easily inocubated) or even dangerous when they are classed as pathogenic. Experimenting with the latter requires a licensed laboratory.

5.2.2 Extensive Noise Generation

The time signal recorded by the system, prior to the analysis, represents the level of light energy, or amount of photons, as detected by the optical sensors at a given moment in time. Thus, per time unit, only a single intensity value is sampled.

While it is desirable to retrieve information (encoded in the light scattering) exclusively from within the sample target area, one has to take into account that before the light waves emitted from the target area can reach the sensors they are subject to many distortions by other particles or cells on the way to the detectors. Moreover, many particles and cells outside the immediate target area are hit by the incident laser beam, too, resulting in ever more light reflections being scattered towards the sensors. Again, these may be distorted by floating cells and particles on their way and further change the pattern of light waves hitting the photodetectors. This puzzle of interacting light reflexions is termed 'noise'.

As mentioned before, the recorded signal is merely comprised of the light reflexions as detected at the point of the sensors at every instant in time. However, it does not contain information about how, for every instant in time, these reflexions are being generated by interacting elements of the sample environment. Consequently, per time unit, a single energy value recorded by the sensors represents the condition of much of the entire interacting sample environment (the target area signal plus the noise), and not just the state of the immediate laser target area, the result of which is a noisy signal. However, with respect to the analyzing algorithms, and to raise the signal to noise ratio, the desire is to focus on the actions taking place exclusively within the target area excited by the laser beam.

5.2.3 Focus merely on Motion of Particles or Cells

Applying a laser scattering device in order to distinguish between certain bacteria species means to focus primarily on the motion of the cells, while other physiological or biochemical characteristics (e.g. metabolating actions) and morphological characteristics (i.e. surface structure, shape and size) cannot be taken into account. The degree as to which this method is able to sense these characteristics is fairly low, i.e. their markedness is not well represented in the recorded signal. Therefore, the performance of this method to generate patterns suitable for the classification task is rather poor. The dominant markings in the signal as created by the OBS system are the energy fluctuations which exclusively relate to the motion of cells and particles in a sample.

6 Suggested Improvements and Future Work

6.1 Sensoric Improvements through Charged Coupled Device

A rather rudimentary setup of a diode laser combined with a set of photodetectors, as demonstrated by the OBS system, benefits from its simplicity and robustness: It is easily assembled, installed, operated, and transportable. Likewise, it can be produced and distributed extremely cost efficiently. On the other hand, relying on this method for the acquisition of data required for a subsequent analysis process seems to be the achilles heel of the entire classification system. Accordingly, the most extensive use of algorithms to analyze the data must fail if, in the first place, there is no (or at best gravely distorted) information related to bacterial activity encoded within the data stream. Thus, the great advantage of the device being simple and cheap loses its attraction if it does not serve well enough to achieve the scientific aim for which it is employed. It seems that the really ambitious task of classifying bacteria from an unknown sample demands the application of a more sensitive technique to retrieve suitable patterns for the analysis.

Compared to a Charged Coupled Device (CCD) chip, for example, to record patterns generated by bacteria, a simple photodetector has a number of drawbacks (see Figure 6.1). As mentioned before, the laser/photodetector setup samples a single value per time unit. This value solely represents the state of the observation environment at every instant in time. This results in a one-dimensional time signal. Due to the relatively simple construction of the signal it can be recorded at a very high frequency (i.e. 1000 to 10000 Hz), enabling the analyzing elements to investigate the behaviour of the sample environment down to microseconds. The refresh rate is where a photodetector should normally outclass a CCD chip. At high sampling frequencies, actions or movements of cells could, in theory, be analyzed in great detail. However, in the present experimental setup, the signal relates to a vast number of light reflexions interacting with each other and thus carries too many patterns which are mixed together before they reach the photodiodes, rendering the signal pattern largely useless as it is undetermined.

Today's high speed cameras are making use of highly sensitive CCD chips and are able to produce snapshots with an exposure time of a few nano seconds (see Section 3.1.2.3). This means that there is no difference between the refresh rate frequency at which a photodetector or a CCD chip operates. However, current data processing and memory techniques struggle to cope with such a high data rate which makes the idea of retrieving information from hundreds or thousands of pictures per second elusive. But it is not only that one has to increase the data flood to gain more information, the data have to be put into perspective. Thus, the real advantage of a CCD chip sensor over a comparable photodetector setup is that its output excels in an increased number of noticable patterns. While photodetectors solely 'digitize' light intensity values, according to the number of photons reaching the sensor, and thus describing the state of the sample environment per time unit, the CCD sensor provides these light intensity values more detailed together with light spectrum and spatial distribution information. In other words, at every instant in time, rather than sampling a single intensity value at the excitation light frequency to cover all the actions taking place within the sample environment, the CCD sensor samples spectral information and the photonic intensity at thousands of specific locations per millimeter. Due to the high information content, these data can be more adequately used to generate patterns to aid in the recognition of bacterial activities. The topological information gained by a CCD image provides the analysis with morphological features such as cell structure, size and shape of bacteria.

In order to justify the use of the herewith presented technique one has to again mention its primary advantage of being able to be applied at very low cost. However, technologically less spectaculous, and with far less information content in the data stream, this method of using photodetectors as the only sensors can most probably be outclassed by the application of CCD chips to sense photonic impulses.

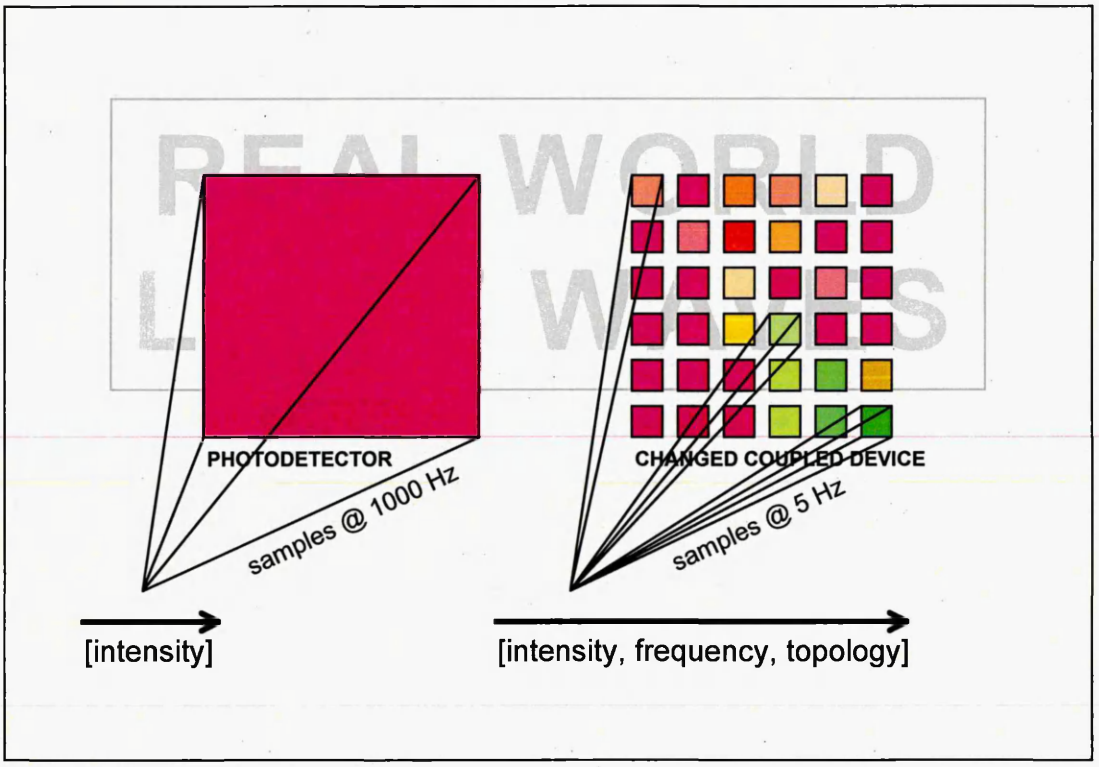

Fig. 6.1 A pair of quite different 'eyes', the CCD-eye resulting in multi-dimensional values being sampled, offering a greater choice of patterns to be generated - compared to the 'all-in-one' samples of the photodetector, however at a much higher refresh rate.

6.2 Altering Wavelengths to improve Laser Scattering Results

Many molecules present within bacteria can be used to generate patterns when they are targeted by laser light at various wavelengths; amongst them are nucleic acids, outer membrane proteins, peptidoglycans, quinones and cytochromes. By collecting a dedicated portion of the light spectrum reflected by the examined bacteria, their respective molecules may cause patterns, which are generated by the scattered intensities at different wavelengths. Since the composition of the molecules within the cell is diverse for every species, the spectrum might result in representative patterns. However, it has to be guaranteed that the analysis is tuned to be sensitive enough to recognize the respective features. This might only be achieved by focusing the light (excitation) source on non-motile, individual cells.

Greve et al. (1993) point out that the resonance Raman scattering at UV light levels is so intense that many molecules present in bacteria (such as nucleic acids, outer membrane proteins, peptidoglycans, quinones and cytochromes) are uniquely represented in the resulting spectrum. Particularly excitations at 222 nm are rated to be useful for

fingerprinting because the spectra are detailed and show many variations between bacterial cells.

In contrast to exciting cells in the UV range, the International Organization for Standardization (1990) recommends a standardized procedure, namely ISO 7027 Section 3, which prescribes that particle light scattering should be measured in the near infrared (λ = 860 nm) spectrum. This is said to minimize interferences compared to measuring between 400 and 600 nm (Hongve and Akesson 1998).

6.3 Genetic Analysis leads to a better Classification Framework

Changes of genuses and species characteristics have occurred throughout the past and will do so in the future. Optical techniques monitor the morphological and physiological characteristics of microorganisms, but these vary considerably within each of today's acknowledged bacteria classes. As a consequence, genuses and species should not be assembled into classes because of their similarities in shape or, more detailed, because of common structural characteristics, but due to their nucleoid similarity. In other words, classes established with optical identification techniques do not adhere to the (advanced) customary class structure of ribosomal RNA-based phylogenies like the one proposed by Madigan et al. (2000). Using this framework, the structure of the nucleid acids found in bacteria cells are examined.

Indeed, it makes further sense to use DNA (deoxyribonucleic acid) or RNA (ribonucleic acid) in order to classify bacteria since they uniquely describe individual organisms and therefore quite naturally represent a genotypic fingerprint of any species. Thus, traditional stain-typing methods (see Section 2.4.4) are nowadays replaced by the analysis of DNA patterns. When bacterial DNA is eletrophoresed, it can be transformed into a molecular fingerprint which is characteristic to a specific clone's genome (Edberg and Melson 2001). Ultimately, the genetic analysis seems to be the only precise means for an identification of microorganisms.

6.4 Combining Laser Spectroscopy and Dielectrophoresis

6.4.1 General Considerations

When applying optical detection methods like the one described in Chapter 3, morphological and behavioural characteristics of the observed microorganisms serve to generate features for the classification process. However, taking into account the target area of the laser as well as the sensitivity of the sensors, it is obvious that single cells, due to their comparatively tiny size, will not give much rise to changing patterns in the recordings. Even clusters of several bacteria attached together, naturally occurring with many species over time when single cells keep splitting, may not provide the required evidence to be detected by the laser/sensor setup. Therefore, the technique of dielectrophoresis could be applied to achieve bacteria formations large enough to indicate patterns which might be detectable. Apart from the potential to collect cells, another highlight of dielectrophoresis is to have an influence on the motion of the cells.

It is assumed that the system which links the previous methods together is a more powerful answer to the question of how bacteria could be classified or distinguished than both methods used on their own. Conditions and properties, which let a single technique system decide upon a classification of an observed bacteria type, can be logically combined to achieve a more profound assumption about the observations (see Figure 6.2).

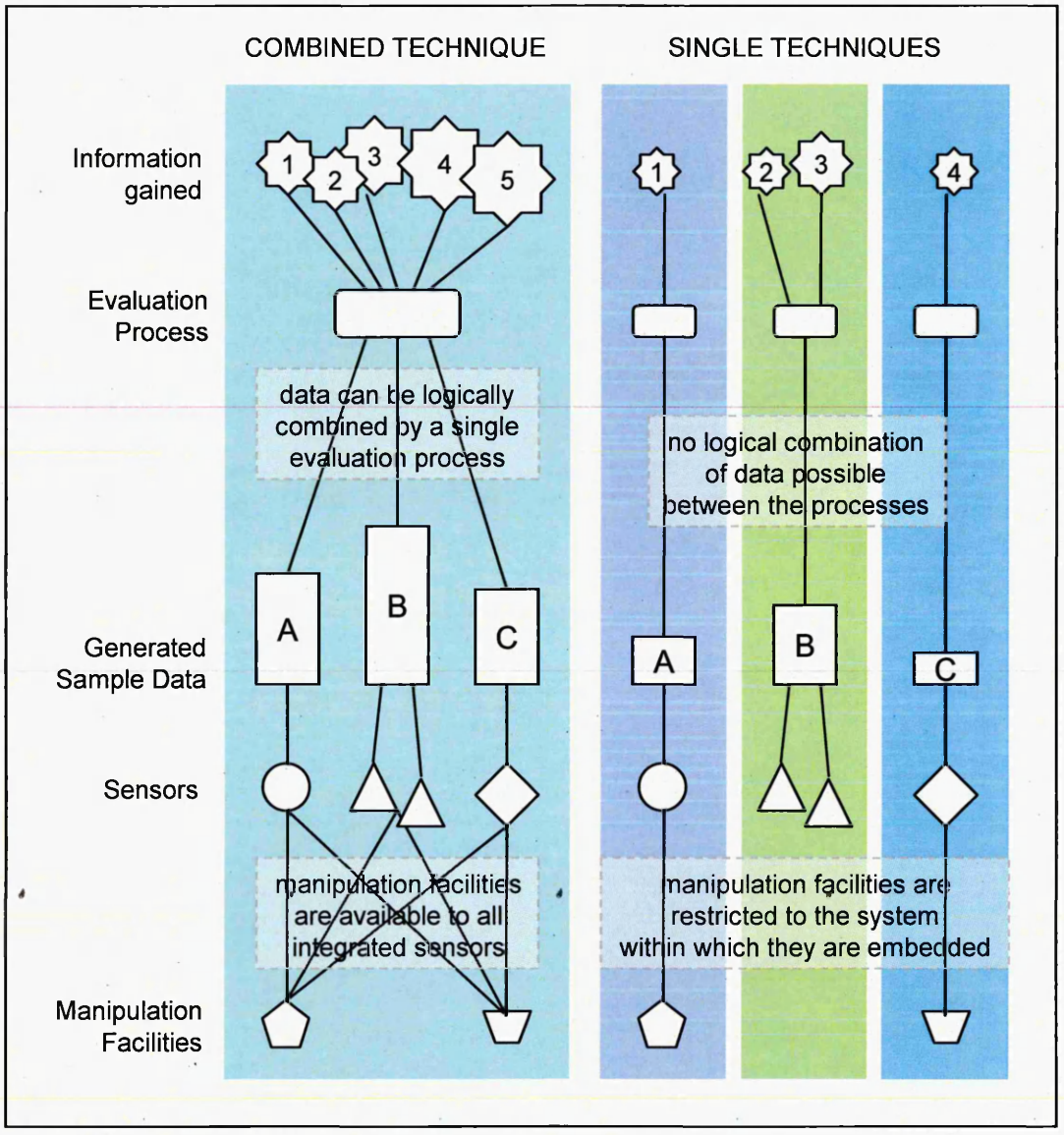

Fig. 6.2 Advantages of the combined technique system in gaining information over the sum of information of each respective single technique system

In addition to the advantage to observe the cells' activity with multiple diverse sensors, dielectrophoresis assembles the bacteria under observation together to result in a much higher density of cells, which let the sensors focus on areas of interest and at the same time results in much less interference within the thus recorded samples. Another very important advantage of the combined technique is the system's ability to excite the cells by means of the DEP force and to subsequently monitor their reactions with a variety of

different sensors. To monitor this play of exposing the cells to a force and letting them react might reveal vital indications about their identity. In the combined technique system, all integrated sensors, as they are recording the behaviour of the bacteria within the sample, benefit from the powerful manipulating capabilities of the dielectrophoresis technique.

6.4.2 Advanced Working Principle of a new Laser Scattering Device

The here described method to monitor the bacteria contents of water samples makes further use of the DEP technique to achieve a multitude of benefits for the analyzing algorithms, which try to classify the bacteria present within the samples. Using any technique, bacteria are classified by generating useful information to be extracted from the data recordings of the sensors. The benefits of the here described method include being able to collect more useful information, while reducing the overall amount of recorded data, which may enable classification algorithms to operate more successfully, and to generally speed up the whole classification process.

In addition to the laser/sensors arrangement, an electrode chamber is placed within the water sample. According to the desired recording methods, the lasers are focused on certain points of interest on and around the electrodes, where, once the DEP action has been induced, they are expected to hit numerous bacterial cells at a very high rate (see Figure 6.3).

In contrast to the mixture of bacteria and other floating particles found in water, giving rise to distorted optical reflections and thus distorting the frequencies and amplitudes of the sensors' recordings, making use of the DEP force to filter the bacteria from the rest of particles may result in much less noisy recordings. Furthermore, large numbers of cells are being located at a single (relatively small) area, where they can be detected without greatly increasing their numbers to intensify the bacteria related information in the recordings of the laser/sensors setup.

6.4.3 Advantages of the Proposed Device

In the following, the advantages of the proposed method are set out in detail.

6.4.3.1 Enhanced Information encoded in the Recordings

The data should inhibit a much greater information density regarding any recorded characteristics of the observed bacteria. Whereas, using the Optical Biosensor System on its own (as described in Chapter 3), the bacteria colonies are spread out among the vast area of the sample cuvette as opposed to the hot spot of the laser, the newly proposed method acts to assemble the majority of cells at a comparably small area. Within minutes, the dielectrophoretic force helps to drag all cells within the active radius of the force onto the electrodes. Being grouped together at known locations, the lasers can be focused on these points of interest and are thus able to produce a much higher rate of (characteristical) optical reflexions.

The larger the entire area of the sample environment, surrounding the immediate sample target area, the more open the system is to interacting elements and light reflexions and thus to noise. While the complete area of the sample environment of the OBS system (not

the immediate laser target area) is in the range of 70 mm^3, this area is considerably reduced in the case of the new system to the order of 1 mm^3; this area could even be further minimized by the application of a more precise laser. Compared with the cubic appearance of the cuvette, as being the sample environment of the OBS system, exhibiting great depth, length and breadth parameters, the sample environment of the advanced system can be described as a very thinly sliced (depth), narrow channel (length and breadth). Here, the light waves do not travel as nearly as far as in the other sample environment (or do so outside the sample environment without distorting elements) before they reach the optical sensor, thus greatly reducing the amount of noise.

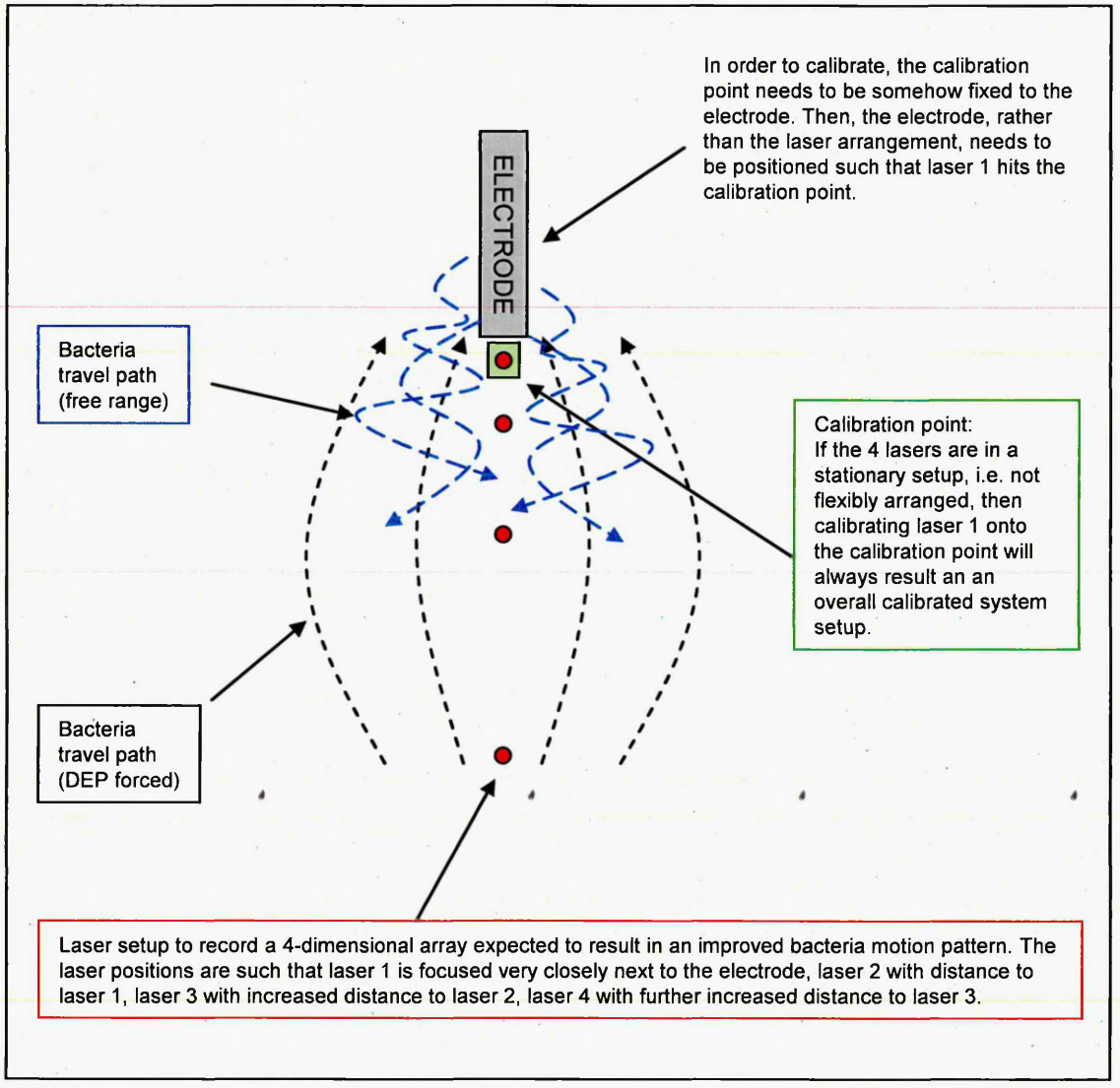

Fig. 6.3 Advanced working principle of a recognition system combining laser scattering with dielectrophoresis

6.4.3.2 Speeding up the Bacteria Detection Process and reducing the Amount of Cells in the Sample

In any case, compared to the working principle of the OBS system, employing the dielectrophoresis technique allows it to speed up the bacteria detection process. Whereas the OBS system begins to detect the bacteria at concentrations of around 10^7, which requires an incubation time of 24 hours, the proposed device, due to the 'assembling capabilities' of the DEP technique, would need a much lower incubation time to facilitate detectable cell concentrations. At the same time this means that the newly proposed

method allows to considerably reducing the bacteria concentrations, probably in the range of 10^2, an amount far too small to be observable by the OBS system. Hence, fewer numbers of cells can be detected in less time.

The 'dragging force' of dielectrophoresis helps to localize great quantities of bacteria around the electrodes, a quantity which needs to be large enough to be detected. This quantity of thus captured cells is governed by the overall amount of bacteria within a sample and the action radius of the dielectrophoretic force, coupled with its activation time. These figures and the resulting possibilities for the detection process would have to be further analyzed and discussed in future research.

6.4.3.3 Altering the Sample Environment to obtain additional useful Information

As already mentioned, the new device would allow the precise observation of cells while they are moving, rotating, and vibrating. In order to increase the amount of information gained by their motility, the behaviour of bacteria could be influenced by the DEP force to result in diverse motion patterns. Accordingly, three scenarios are imaginable (see Figure 6.4):

1) The bacteria are being monitored locally at the electrodes after they have been captured through the DEP force. Thus, while the electromagnetic force is still active, the reflected laser light could be recorded, which may result in a characteristical 'immobile state' pattern (see Figure 6.4 [1]).

2) After the DEP force has captured a number of cells at the electrodes, it will be switched off and the bacteria will start to range freely. The sensor would then record the light reflexions from the laser focused around the electrodes (see Figure 6.4 [2]).

3) In a third approach the bacteria could be sensed while they are agitated by the electromagnetic force. The dielectric force exerted by the electrodes has an impact on the velocity of the cells, further governed by the field distribution and the dielectric properties of the cells themselves. (Pimbley et al. 1999). Thus different DEP frequencies might result in different motion patterns as the bacteria are forced towards the electrodes, because the dielectric characteristics of bacteria influence its dielectrophoretic behaviour. In other words varying bacteria types might react differently to certain DEP frequencies, causing the sensors to record characteristical patterns for each bacteria type under observation (see Figure 6.4 [3]).

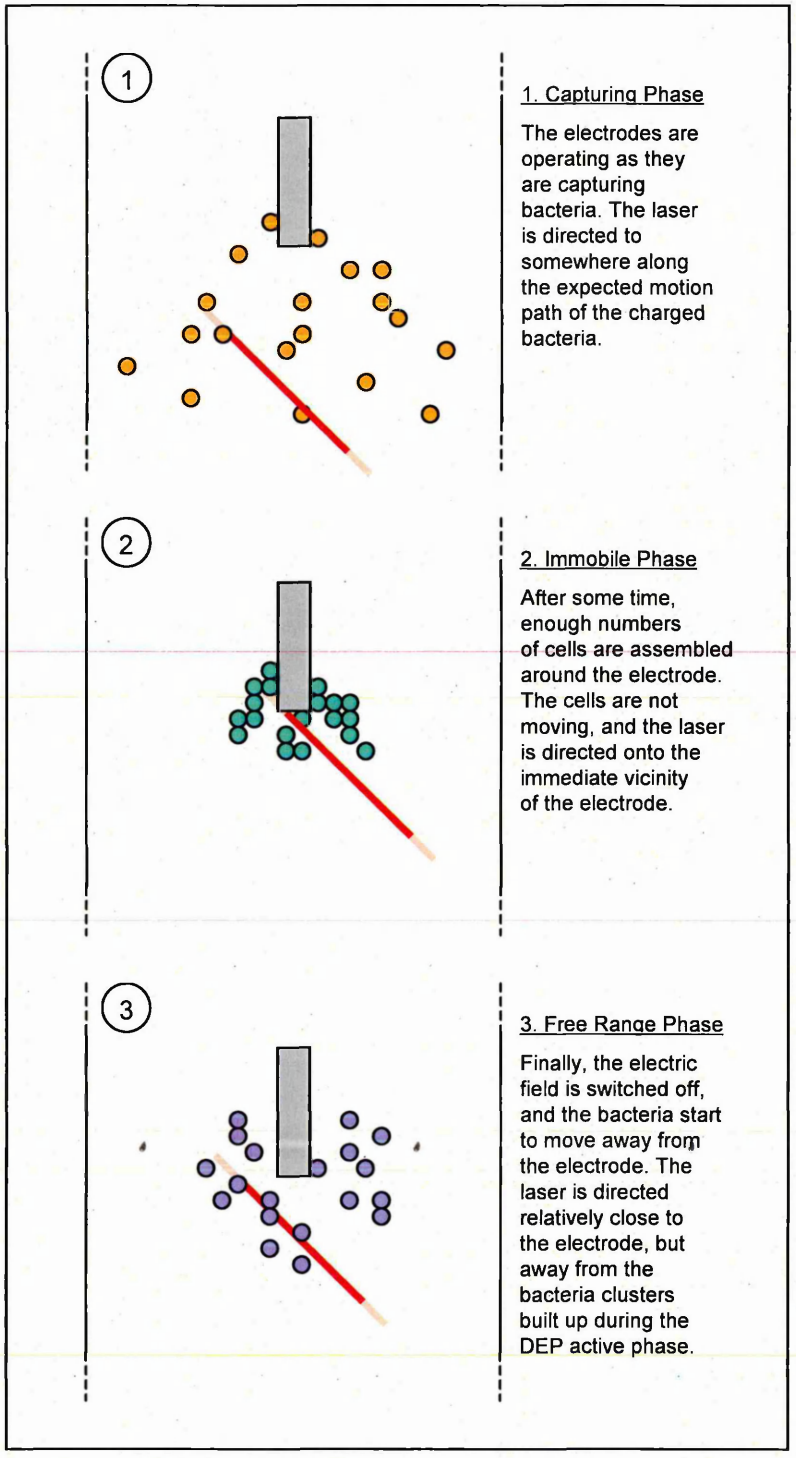

Fig. 6.4 Three possible setups to monitor a multitude of bacteria charactersitics. Setup 1) locates the laser along the bacteria path, setup 2) along the edges of the electrodes, and setup 3) in a circle formation around the electrodes.

6.4.3.4 Benefits at a Glance

a) Immobile states as well as DEP driven states would be recorded, while the self-motility of bacteria would be retained, too;

b) A (much) greater information density within the recorded data could be achieved;

c) A (much) greater speed in detecting cells and obtaining results would be achieved;

d) The manual preparation requirements for the bacteria samples could be reduced;

In summary it can be said that by reasonably enhancing the sensoric precision and working principle of the OBS system, while combining it with the exciting capabilities of dielectrophoresis, the analytical performance of the classification system could be inherently improved. Not only would there be a greater information density within the recorded data, improving the reliability of the classification process, but the detection of cells would be accelerated, too, and results could be obtained in less time. The DEP force would enable the system to record immobile and DEP driven states of the monitored organisms, while their self-motility would be retained as the force is switched off. Finally, the manual preparation requirements for the bacteria samples would be simplified by being able to reduce the cell concentration, which, in turn, would have the benefit of an increased overall consistency.

References

AAVLD (American Association of Veterinary Laboratory Diagnosticians) (2002) *Yersinia Pestis: Isolation and Identification,* www.aavld.org/MainMenu2/NAHLN/YersiniapestisSOP.doc, [26-10-2006]

Anscombe N (2003) "Optoacoustics moves closer to the clinic", *Biophotonics International 04/2003,* ISSN 1081-8693, Laurin Publishing, MA

Bayraktar B, Banada P P, Hirleman E D, Bhunia A K, Robinson J P, Rajwa B (2006) "Feature extraction from light-scatter patterns of listeria colonies for identification and classification", *J. Biomed. Opt., Vol. 11*, p034006, 2006

BD Bioscience Clontech (2002) "Far-red fluorescent protein", *Biophotonics International 09/2002,* p72, ISSN 1081-8693, Laurin Publishing, MA

Betts W B and Brown A P (1999) *Dielectrophoretic analysis of microbes in water*, J Appl Microbiol. Symposium Supplement 1999, 85, pp201S-213S

Brown A P, Betts W B, Harrison A B, O'Neill J G (1998) *Evaluation of a dielectrophoretic bacterial counting technique*, Biosensors & Bioelectronics 14/1999, pp341-351

Cady P, Dufour S W, Shaw J, Kraeger S J (1978) *Electrical Impedance Measurements: Rapid Method for Detecting and Monitoring Microorganisms*, J Clinical Microbiol. 1978 Mar; Vol. 7 No. 3, American Society for Microbiology

Cheun HI, Makino SI, Watarai M, Shirahata T, Uchida I, Takeshi K (2001) *A simple and sensitive detection system for Bacillus anthracis in meat and tissue*, J Appl Microbiol. 2001 Sep;91(3):421-6

Coulter C B (1920) J. Gen. Physiol. 3, pp309-406

Cross A D and Jones R A (1969) *An introduction to practical infra-red spectroscopy*, Third Edition, Butterworth & Co, UK

Darwin C R (1859) *On the Origin of Species,* http://embryology.med.unsw.edu.au/pdf/Origin_of_Species.pdf, [08-10-2004]

Doetsch R N and Cook T M (1973) *Introduction to Bacteria and Their Ecobiology,* Medical and Technical Publishing, Lancaster, UK

Edberg S C and Melson J E (2001) "Evaluation of Indicator Bacteria Using Molecular Fingerprinting", *American Water Works Association,* ISBN 1583211004

ESRI (2010) *ESRI Developer Network,* http://edndoc.esri.com/ascobjects/9.2/net/shared/geoprocessing/dpatial_analyst_tools/how_maximumum_likelihood_classification_works.htm

Felice C J, Clavin O E, Gallo B, Armayor M R, Spinelli J C, Valentinuzzi M E (1988) "Impedancimetric bacterial detection: Theoretical and experimental aspects", *Med. Progr. Technol.*, 14 pp25-33

Fistenberg-Eden R and Eden G (1984) "Impedance Microbiology" Bactomatic, *Div. Med. Technol. Corp.*, p170, Princeton, New Jersey, USA

Gilway Technical Lamp (2003) "Photo-optic lamps", *Biophotonics International 11/2003,* p74, ISSN 1081-8693, Laurin Publishing, MA

Ginouves P (2003) "The Next Generation of Analytical Instruments", *Biophotonics International 01-02/2003,* ISSN 1081-8693, Laurin Publishing, MA

Goldwasser S M and Gieszczykiewicz F M (2006) *Sci.Electronics.Repair FAQ, Version 1.04,* www.repairfaq.org/sam/laserssl.htm, [22-01-2007]

Greve J and Puppels G J (1993) "Raman Microspectroscopy of Single Whole Cells", in: Clark R J H and Hester R E, *Biomolecular Spectroscopy Part A, Advances in Spectroscopy Vol. 20*, pp231-265, John Wiley and Sons, UK

Haeckel E H (1894) *Systematische Phylogenie der Protisten und Pflanzen,* I. G. Reimer, Berlin

Hing P and Müller H W (2003) "CCD Cameras simplify biological analysis", *Biophotonics International 09/2003,* p52, ISSN 1081-8693, Laurin Publishing, MA

Hirleman E D, Bae E W, Huff K, Banada P, Bhunia A (2007) "Light scattering endows bacterial colonies with unique fingerprints", *SPIE - The Int. Soc. for Opt. Eng., 10.1117/2.1200701.0557*

Hogan H (2002a) "This little light of mine: Diode Lasers in Medicine and Biotechnology", *Biophotonics International 06/2002,* ISSN 1081-8693, Laurin Publishing, MA

Hogan H (2002b) "Advances boost photon detection", *Biophotonics International 07-08/2002,* ISSN 1081-8693, Laurin Publishing, MA

Holt J G, Krieg N R, Sneath P H A, Staley J T, Williams S T (1994) *Bergey's Manual of Determinative Bacteriology,* ISBN 0-683-00603-7, Williams and Wilkins, Maryland 21202, USA

Hongve D and Akesson G (1998) "Comparison of Nephelometric Turbidity Measurements using Wavelengths 400 - 600 and 860 nm", *Wat. Res. Vol. 32, No. 10*, pp3143-3145

Hof M (2003) "Section II Methods 1: Optical Spectroscopy (3 Basics of Optical Spectroscopy)", in: Gauglitz G and Vo-Dinh T (eds.), *Handbook of Spectroscopy Vol. 1,* ISBN 3-527-29782-0, Wiley-VCH, Germany

IBMM and School of Electronic Engineering (Bangor University) (2002) *Basic Science*, http://www.ibmmicrotech.co.uk/microeng/dielectrophoresis/science.php, [15-09-2008]

International Organization for Standardization (1990) *International Standard ISO 7027; Water Quality: Determination of Turbidity*, 2nd edition, Geneva

Keiderling T A and Pancoska P (1993) "Structural studies of biological macromolecules using vibrational circular dichroism", in: Clark R J H and Hester R E, *Biomolecular Spectroscopy Part B, Advances in Spectroscopy Vol. 15*, John Wiley and Sons, UK

Legget K (2001) "Mass spectroscopy identifies harmful bacterial spores", *Biophotonics International 12/2001,* p55, ISSN 1081-8693, Laurin Publishing, MA

Macková A, Morton S A, Walker C G H, Volka K (2003) "Section VI Methods 5: Surface Analysis Techniques", in: Gauglitz G and Vo-Dinh T (eds.), *Handbook of Spectroscopy Vol. 1,* ISBN 3-527-29782-0, Wiley-VCH, Germany

Madigan, Martinko, Parker (2000) *Brock - Biology of Microorganisms, Ninth Edition*, Prentice-Hall, Upper Saddle River, New Jersey

Marconi E M (2003) *Helpful Bacteria to the Rescue*, http://www.nasa.gov/missions/science/f_bacteria.html, [15-05-2007]

McMaster M and McMaster C (1998) *GC/MS: a practical user's guide*, ISBN 0-471-24826-6, Wiley-VCH, USA

Milner K R, Brown A P, Allsopp D W E, Betts W B (1998) "Dielectrophoretic classification of bacteria using differential impedance measurements", *Electronic Letters 1998, 34 (1)*, pp66-68, ISSN 0013-5194

Mitruka B M (1976) *Methods of Detection and Identification of Bacteria,* ISBN 0-8493-5116-2, CRC Press

Moore A C, Herwaldt B L, Craun G F, Calderon R L, Highsmith A K, Juranek D D (1994) "Waterborne disease in the United States", *1991 and 1992 Journal of the American Water Works Association*, 86:87-99

Nibler J W and Pubanz G A (1988) "Coherent Raman Spectroscopy of Gases", in: Clark R J H and Hester R E, *Advances in Non-Linear Spectroscopy, Advances in Spectroscopy Vol. 15*, John Wiley and Sons, UK

Nikon Instruments (2003) "CCD camera", *Biophotonics International 11/2003,* p79, ISSN 1081-8693, Laurin Publishing, MA

Noca F, Hoenk M, Hunt B, Choy D, Kowalczyk B, Xu J, Koumoutsakos P, Werder T, Walther J (2000) *Nanoscale Ears based on Artificial Stereocilia*, ASA/NOISE-CON 2000 Meeting, Newport Beach, CA, http://www.acoustics.org/press/140th/noca.htm, [23-09-2004]

Oregon State Public Health Laboratory (2003) *Bacillus anthracis Level A Laboratory Guidelines Flow Chart*, www.dhs.state.or.us/publichealth/phl/bt/anthrax/flowchart.pdf, [26-10-2006]

Pimbley D W, Patel P D, Robertson C J (1999) "Dielectrophoresis", in: Edwards C (ed.), *Environmental Monitoring of Bacteria, Methods in Biotechnology Vol. 12*, Humana Press, NJ

Pohl H A (1958) "Some Effects of Nonuniform Fields on Dielectrics", *J. Appl. Phys. 29*, pp1182-1188

Pohl H A (1978) *Dielectrophoresis*, Cambridge University Press, Cambridge

Pontius F W (1993) "Protecting the public against Cryptosporidium", *Journal of the American Water Works Association*, 84(3), pp36-50

Pontius F W (2003) *Drinking Water Regulation and Health*, ISBN 0471447412, Wiley-IEEE

Poole R K and Kalnenieks U (2000) "Introduction to light absorbtion: visible and ultraviolet spectra", in: Gore M G (ed.), *Spectrophotometry and Spectrofluorimetry,* Oxford University Press

Power Technology (2003) "Violet laser diode", *Biophotonics International 07/2003,* p68, ISSN 1081-8693, Laurin Publishing, MA

Proakis J G and Manolakis D G (1996) *Digital signal processing : principles, algorithms and applications*, Maxwell Macmillan, New York

Rastopov S (1998) *Method of detecting live microorganisms*, US Patent No 5846759

Reuss F F (1809) *Memoires de la Societe Imperiales de Naturalistes de Moskou 2*, pp327-336

Robinson A H, Morrison J L, Mührcke P C, Kimerling A J, Guptill S C (1995) *Elements of Cartography*, p466, 6th Edition, John Wiley & Sons, USA

Rusteck (2000) *Optical detection of particles in a liquid medium (Rastopov et al.)*, P.O. Box 301, Queen's House, Don Road, St. Helier, Jersey JE4 8UQ, UK

Sablinskas V (2003) "Section II Methods 1: Optical Spectroscopy (4 Instrumentation)", in: Gauglitz G and Vo-Dinh T (eds.), *Handbook of Spectroscopy Vol. 1,* ISBN 3-527-29782-0, Wiley-VCH, Germany

Smith B C (1996) *Fundamentals of Fourier Transform Infrared Spectroscopy,* ISBN 0-8493-2461-0, CRC Press, Florida

Steiner G (2003) "Section II Methods 1: Optical Spectroscopy (5 Measurement Techniques)", in: Gauglitz G and Vo-Dinh T (eds.), *Handbook of Spectroscopy Vol. 1,* ISBN 3-527-29782-0, Wiley-VCH, Germany

Storgards E (2001) *Process hygiene control in beer production and dispensing,* ISBN 951-38-5559-7, VTT Publications 410

Terratec (2000) *AudioSystem EWX 24/96 (Manual)*, TerraTec Electronic GmbH, Nettetal, Germany

Ur A and Brown D F J (1975) "Monitoring of Bacterial Activity by Impedance Measurements", in: Heden C G and Illeni T (eds.), *New approaches to the identification of microorganisms,* John Wiley & Sons, USA

Visvalingam M and Whyatt J D (1993) "Line Generalisation by repeated Elimination of Points", *Cartographic. Journal, Vol. 30, No. 1*, pp46-51

VWR International (2006) http://www.vwr.com [18-01-2006]

Walker C T (1975) "Light Scattering", in: Jacobs S F, Sargent M, Scott J F, Scully M O, *Laser Applications to Optics and Spectroscopy*, ISBN 0-201-05682-8, Addison-Wesley, MA

Wikipedia (2007a) *Charge-coupled device*, http://en.wikipedia.org/wiki/DNA [28-06-2007]

Wikipedia (2007b) *Charge-coupled device*, http://en.wikipedia.org/wiki/RNA [28-06-2007]

Wikipedia (2007c) *Principal component analysis*, http://en.wikipedia.org/wiki/Principal_component_analysis [28-06-2007]

Woodruff W H, Dyer R B, Einarsdottir O (1993) "Spectroscopy, Dynamics, and Functions of Cytochrome Oxidase", in: Clark R J H and Hester R E, *Biomolecular Spectroscopy Part B, Advances in Spectroscopy Vol. 15*, John Wiley and Sons, UK

Appendix A - Experimental Details

e3_2
---

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	20h
experimental medium:	water
dilution factor:	2
experimental duration:	18min

e3_3
---

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	20h
experimental medium:	ringer
dilution factor:	2
experimental duration:	18min

e4_2
---

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	20h
experimental medium:	nutrient broth
dilution factor:	3
experimental duration:	77min

e6_1
---

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	70h
experimental medium:	water
dilution factor:	2
experimental duration:	18min

e6_2

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	70h
experimental medium:	water
dilution factor:	2
experimental duration:	18min

e6_3

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	70h
experimental medium:	ringer
dilution factor:	2
experimental duration:	18min

e6_4

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	70h
experimental medium:	ringer
dilution factor:	2
experimental duration:	18min

e7_1

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	17h
experimental medium:	ringer
dilution factor:	2
experimental duration:	18min

e7_4

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	17h
experimental medium:	ringer
dilution factor:	2
experimental duration:	18min

e8_1

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	17h
experimental medium:	water
dilution factor:	2
experimental duration:	18min

e8_4

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	17h
experimental medium:	water
dilution factor:	2
experimental duration:	18min

e9_2

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	17h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

e9_3

species:	escherichia coli
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	17h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

e10_1

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	water
dilution factor:	1
experimental duration:	18min

e10_2

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	water
dilution factor:	1
experimental duration:	18min

e10_3

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	water
dilution factor:	1
experimental duration:	18min

e10_4

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	water
dilution factor:	1
experimental duration:	18min

e11_1

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	ringer
dilution factor:	1
experimental duration:	18min

e11_2

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	ringer
dilution factor:	1
experimental duration:	18min

e11_3

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	ringer
dilution factor:	1
experimental duration:	18min

e11_4

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	ringer
dilution factor:	1
experimental duration:	18min

e13_2

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

e13_3

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	22h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

e14_1

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	24h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

e14_2

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	24h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

e14_3

species:	bacillus subtilis
growing medium:	nutrient broth
growing temperature:	30'C
growing time:	24h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

e14_4

species:	proteus mirabilis
growing medium:	nutrient broth
growing temperature:	30°C
growing time:	24h
experimental medium:	nutrient broth
dilution factor:	2
experimental duration:	77min

Appendix B - Diagrams related to Chapter 3

Fig. B.1
e10_2, e10_3, e6_4

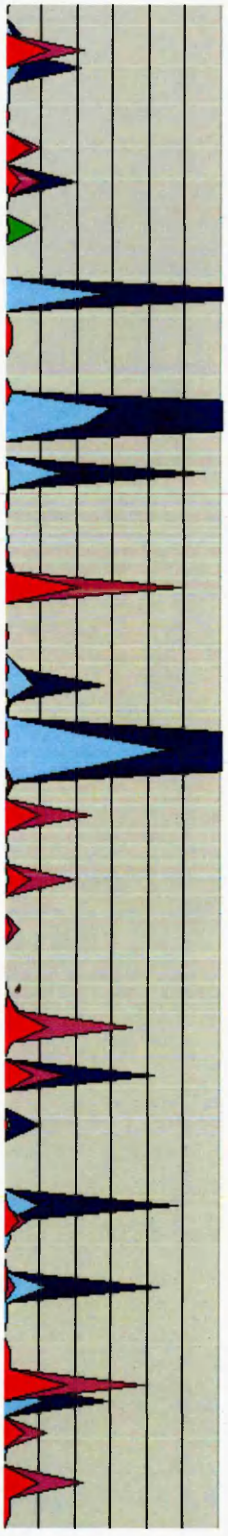

Fig. B.2
e10_1, e10_4, e6_2

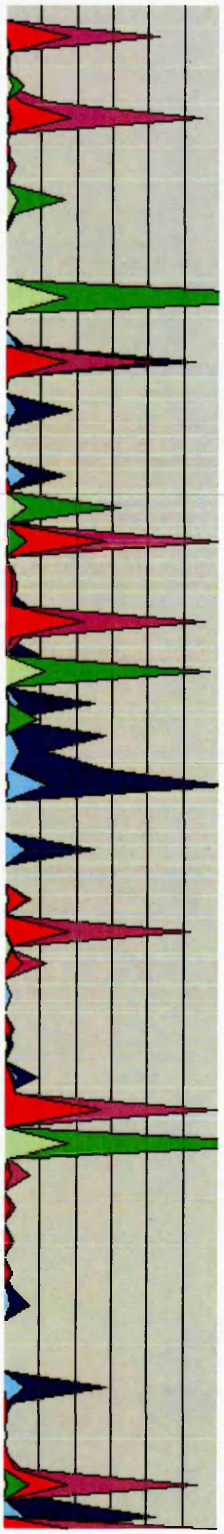

Fig. B.3
e3_2, e8_1, e8_4

Fig. B.4
e11_2, e11_3, e6_1

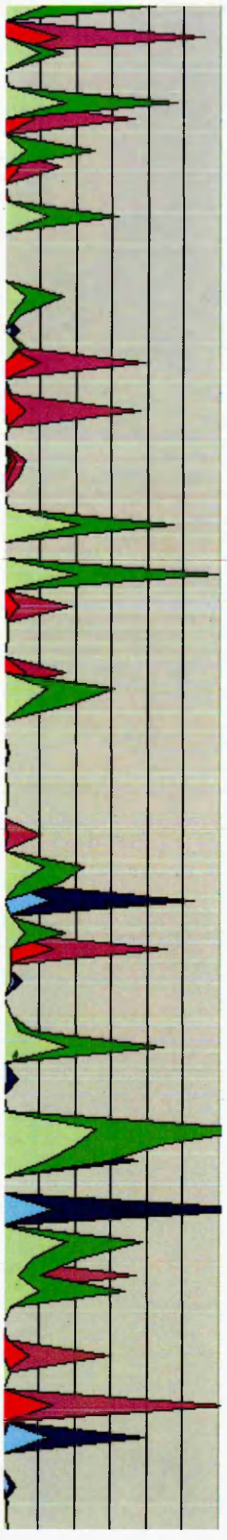

Fig. B.5
e11_1, e11_4, e6_3

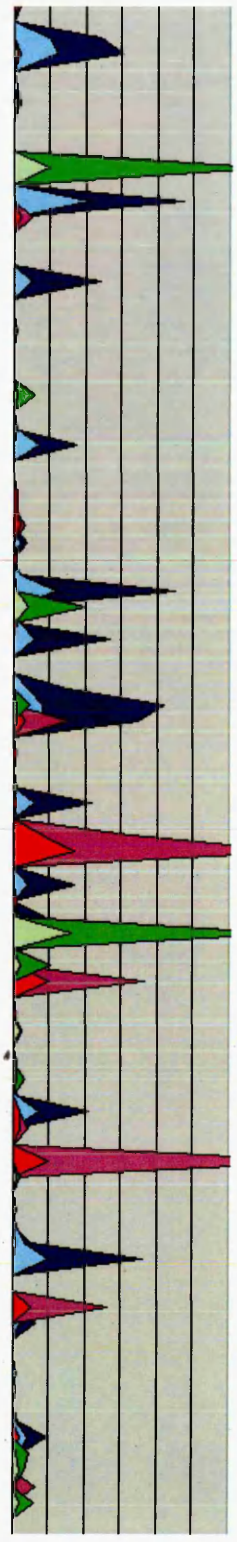

Fig. B.6
e13_3, e7_1, e7_4

Fig. B.7
slope changes e10_1

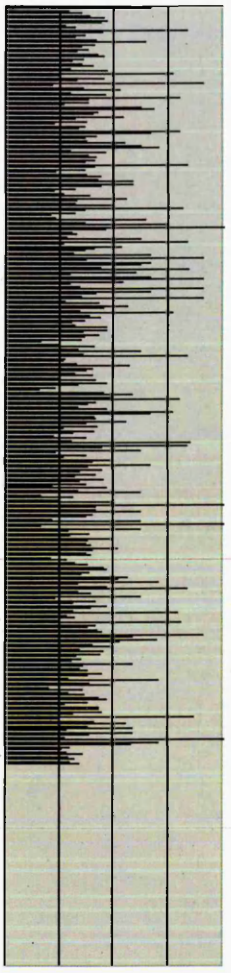

Fig. B.8
slope changes / samples e10_1

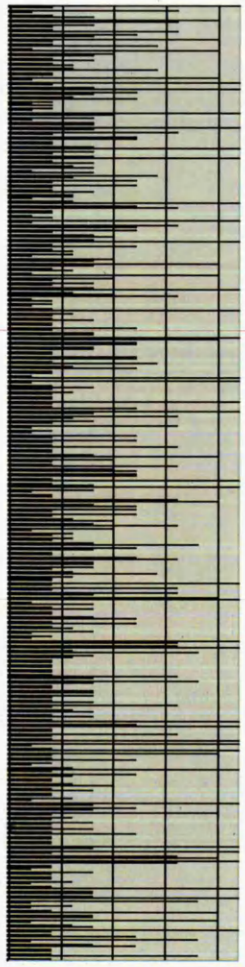

Fig. B.9
slope changes e10_2

Fig. B.10
slope changes /
samples e10_2

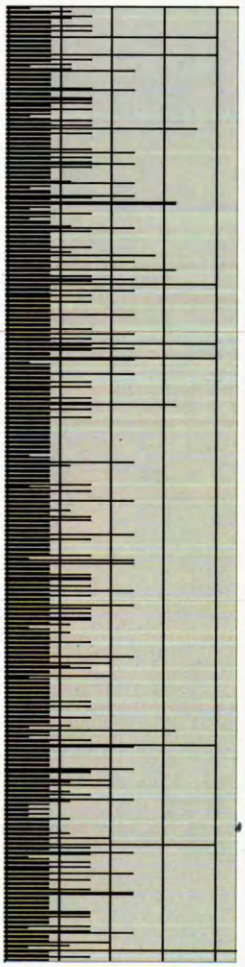

Fig. B.11
slope changes e10_3

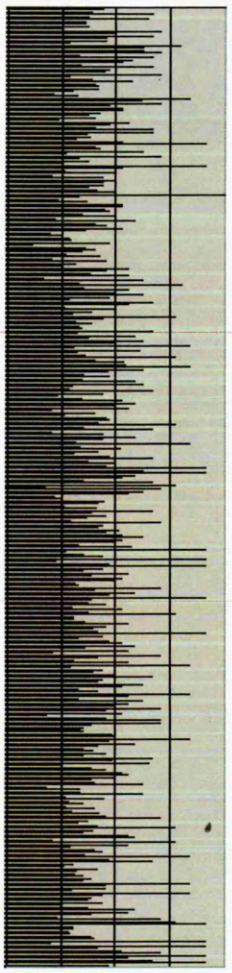

Fig. B.12
slope changes /
samples e10_3

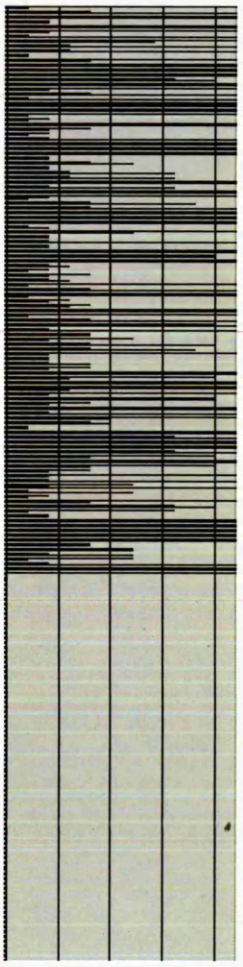

Fig. B.13
slope changes e10_4

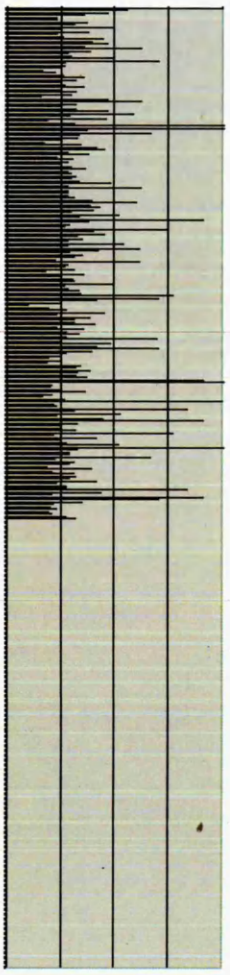

Fig. B.14
slope changes /
samples e10_4

Fig. B.15
slope changes e11_1

Fig. B.16
slope changes /
samples e11_1

Fig. B.17
slope changes e11_2

Fig. B.18
slope changes /
samples e11_2

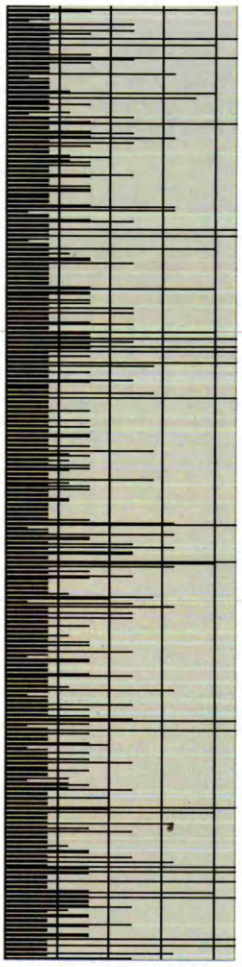

Fig. B.19
slope changes e11_3

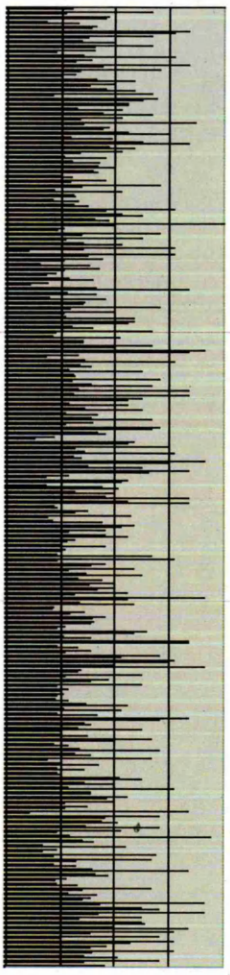

Fig. B.20
slope changes /
samples e11_3

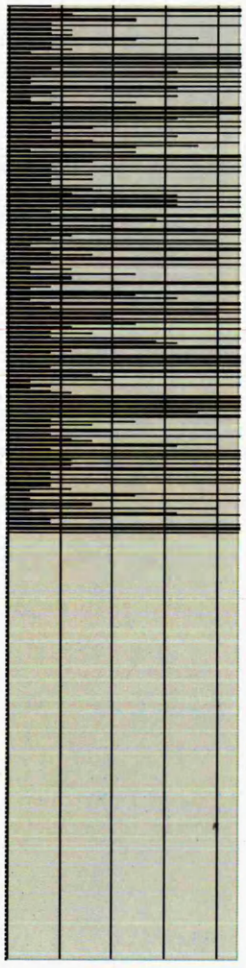

Fig. B.21
slope changes e11_4

Fig. B.22
slope changes /
samples e11_4

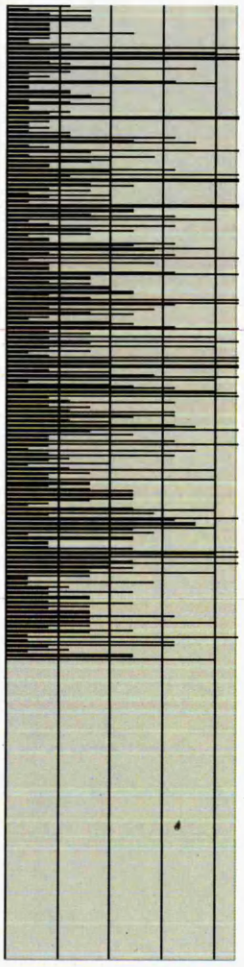

Fig. B.23
slope changes e3_2

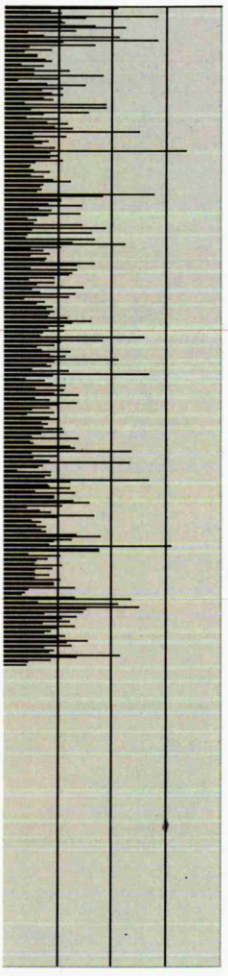

Fig. B.24
slope changes /
samples e3_2

Fig. B.25
slope changes e3_3

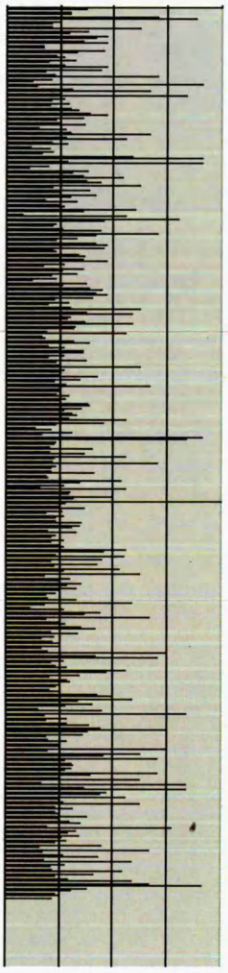

Fig. B.26
slope changes /
samples e3_3

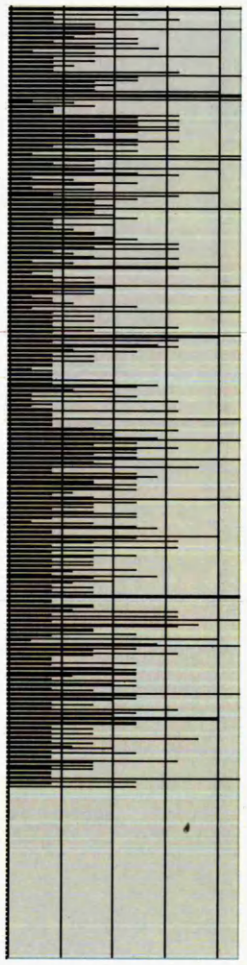

Fig. B.27
slope changes e6_1

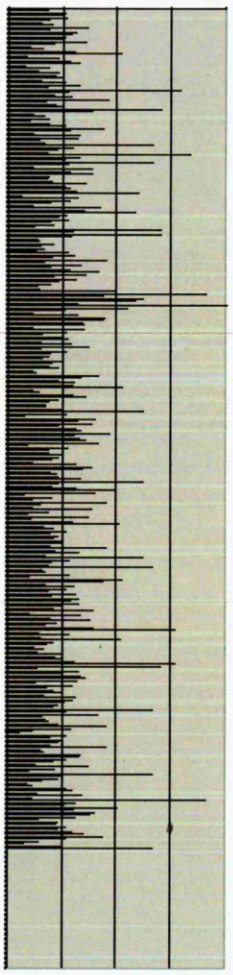

Fig. B.28
slope changes /
samples e6_1

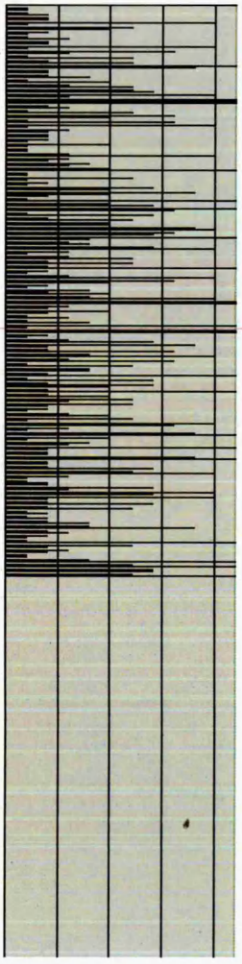

Fig. B.29
slope changes e6_2

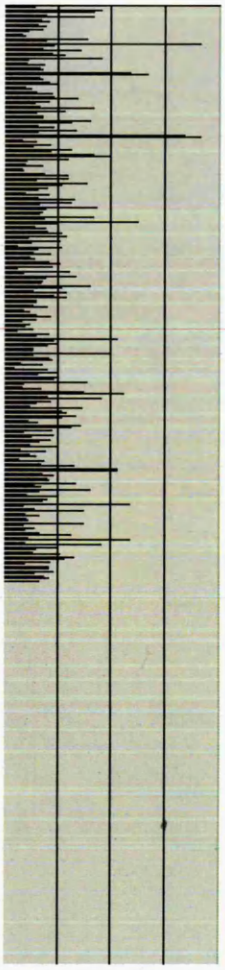

Fig. B.30
slope changes /
samples e6_2

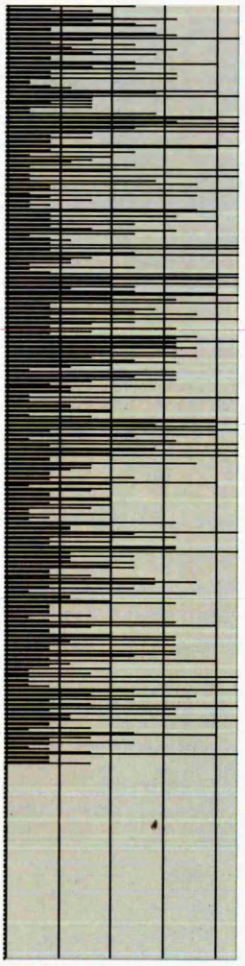

Fig. B.31
slope changes e6_3

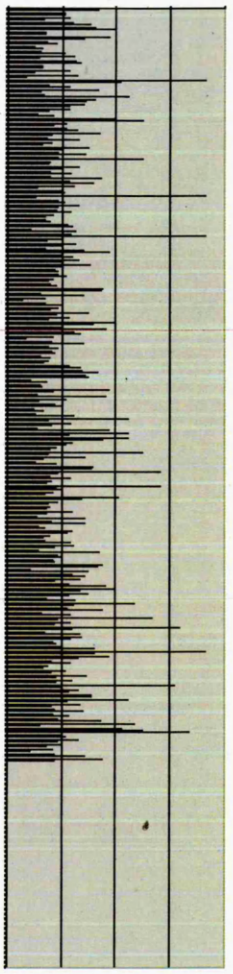

Fig. B.32
slope changes /
samples e6_3

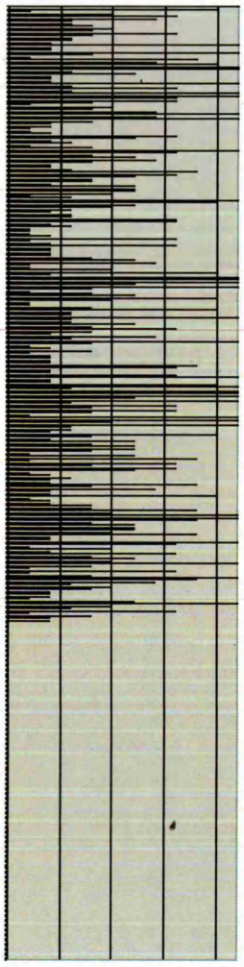

Fig. B.33
slope changes e6_4

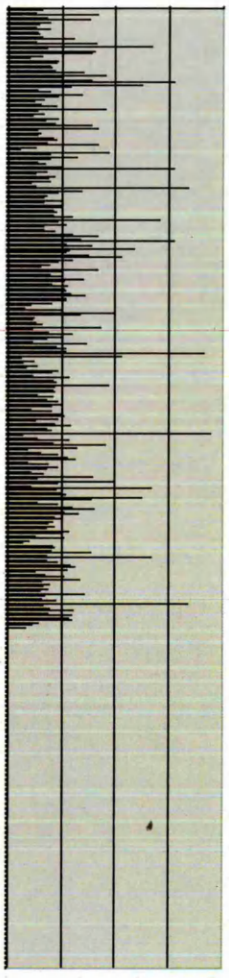

Fig. B.34
slope changes /
samples e6_4

Fig. B.35
slope changes e7_1

Fig. B.36
slope changes /
samples e7_1

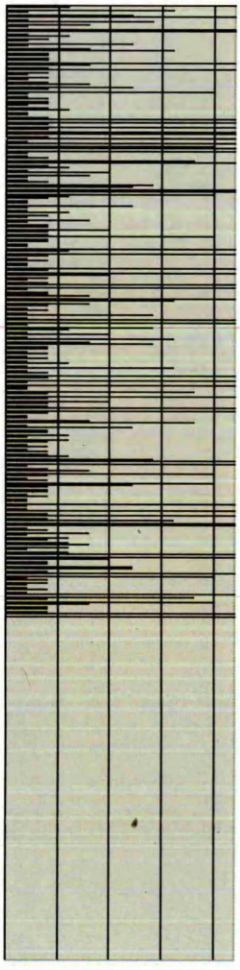

Fig. B.37
slope changes e7_4

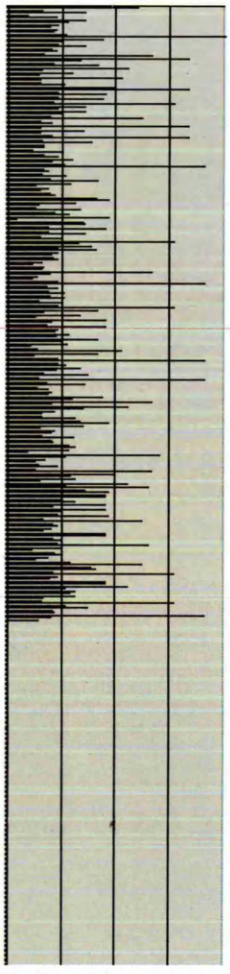

Fig. B.38
slope changes /
samples e7_4

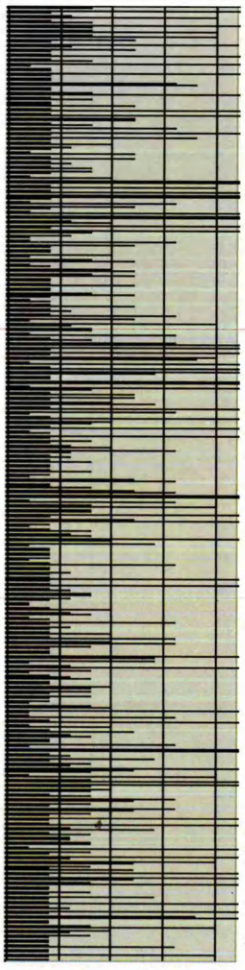

Fig. B.39
slope changes e8_1

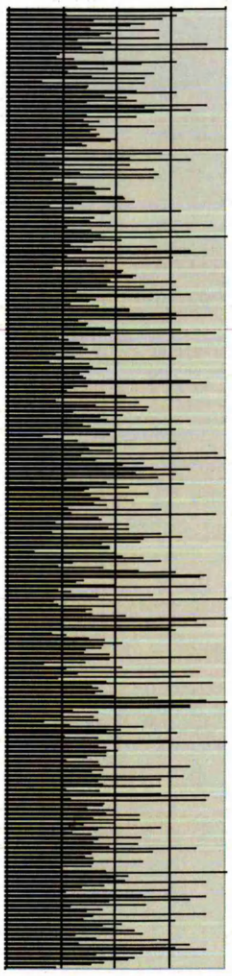

Fig. B.40
slope changes /
samples e8_1

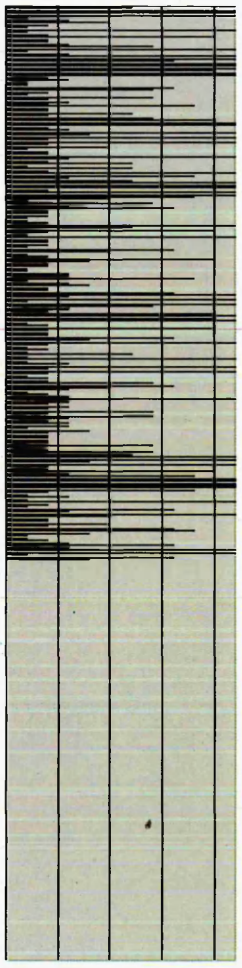

Fig. B.41
slope changes e8_4

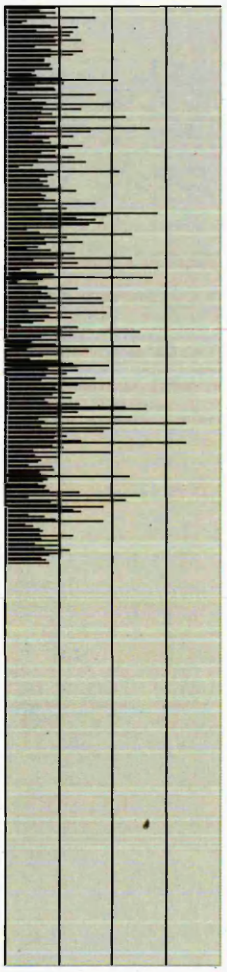

Fig. B.42
slope changes / samples e8_4

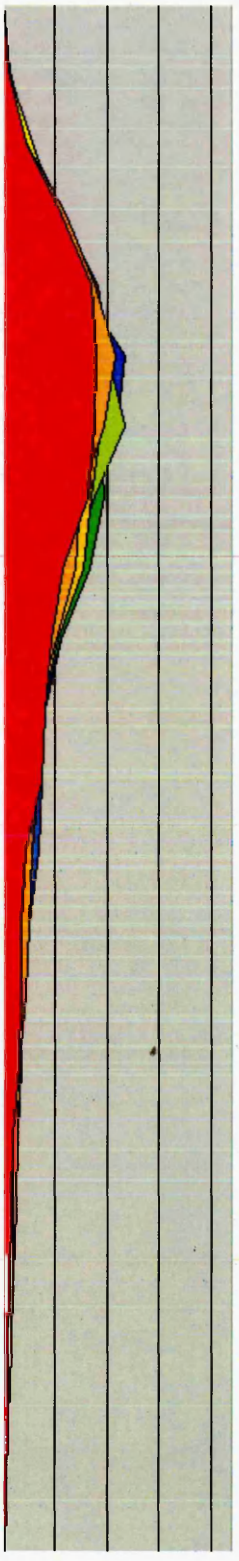

Fig. B.43
psd 4096 e10_1

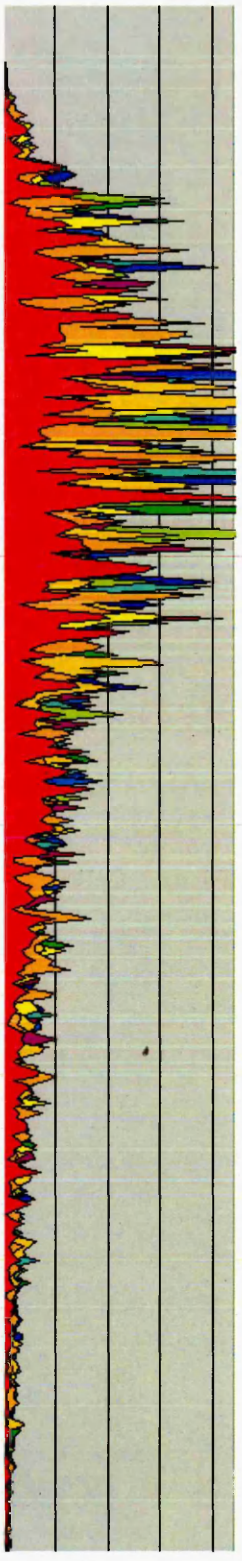

Fig. B.44
psd 65536 e10_1

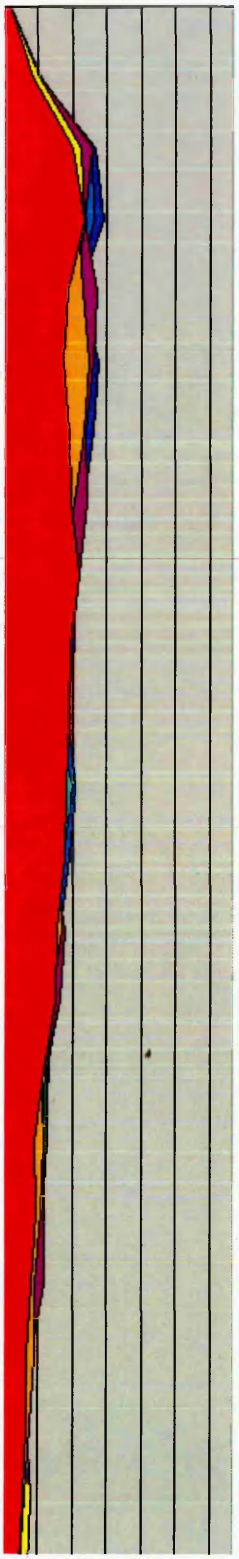

Fig. B.45
psd 4096 e10_2

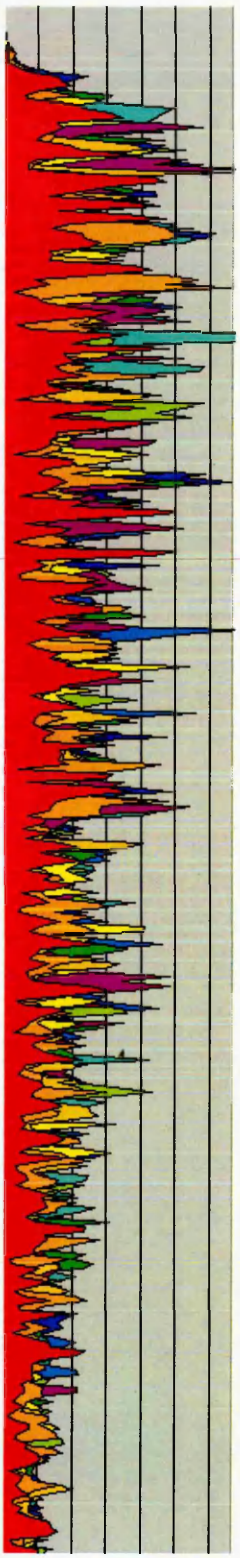

Fig. B.46
psd 65536 e10_2

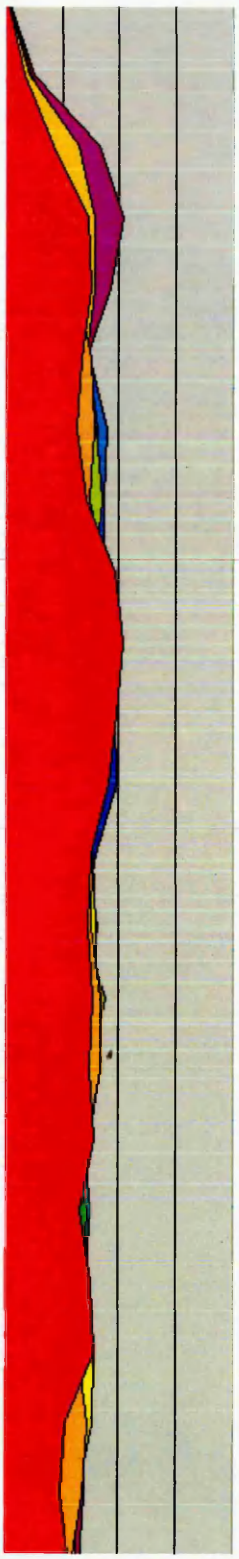

Fig. B.47
psd 4096 e10_3

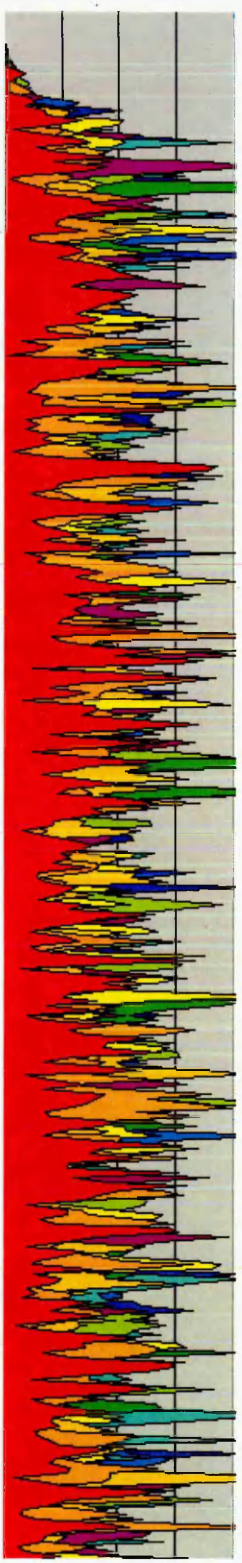

Fig. B.48
psd 65536 e10_3

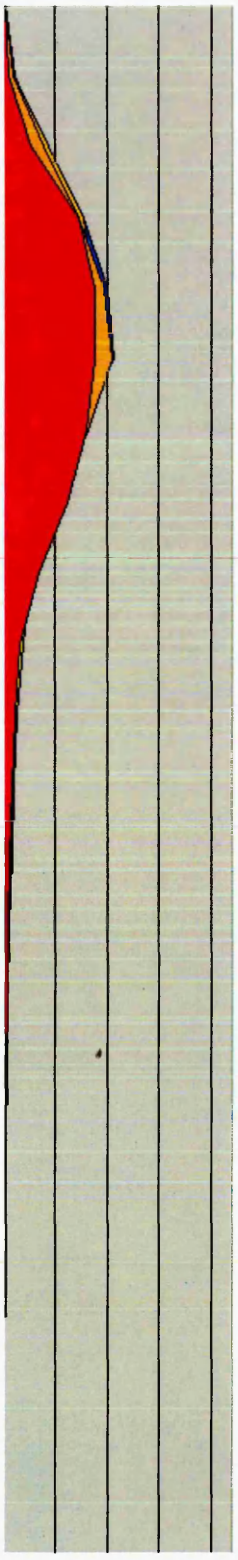

Fig. B.49
psd 4096 e10_4

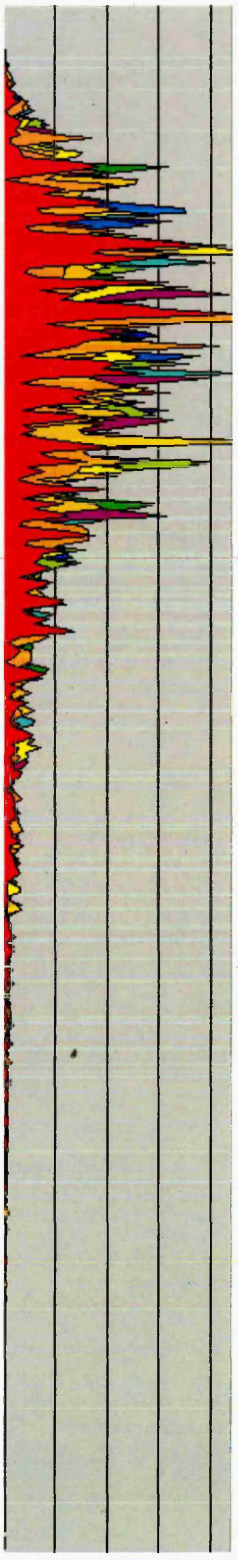

Fig. B.50
psd 65536 e10_4

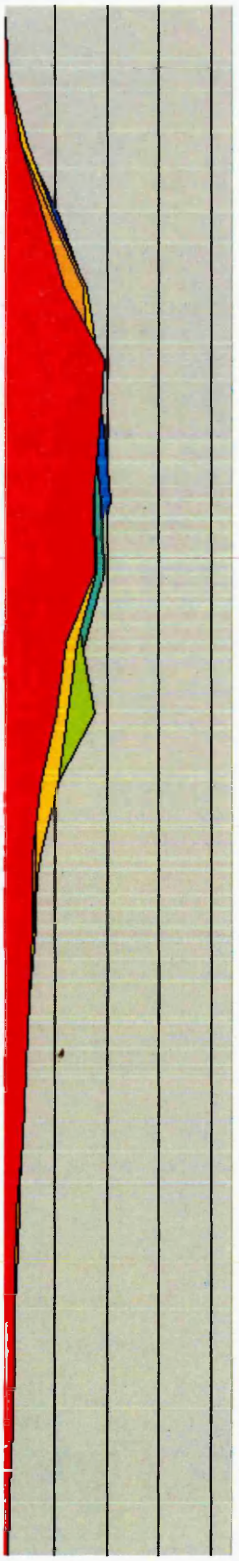

Fig. B.51
psd 4096 e11_1

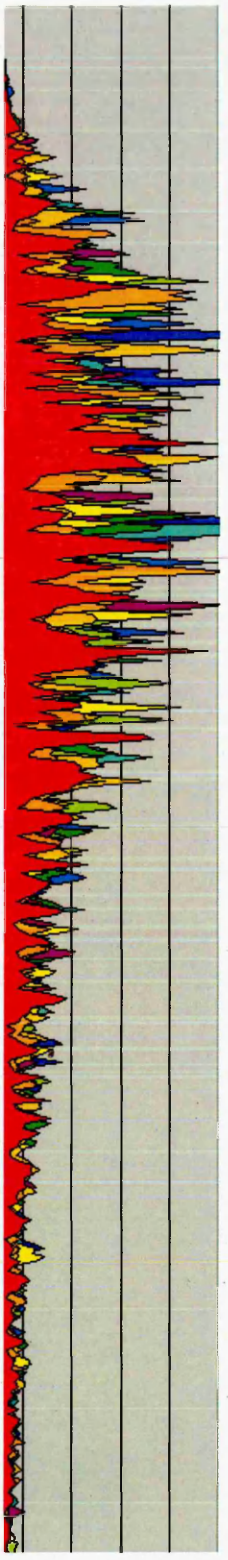

Fig. B.52
psd 65536 e11_1

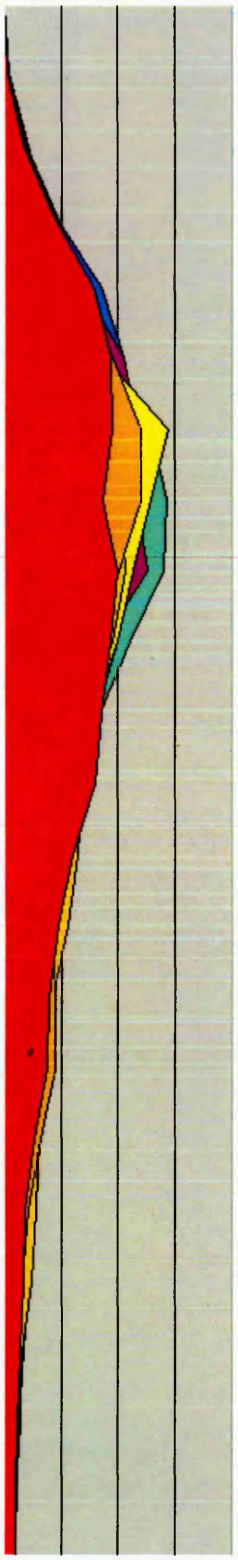

Fig. B.52
psd 4096 e11_2

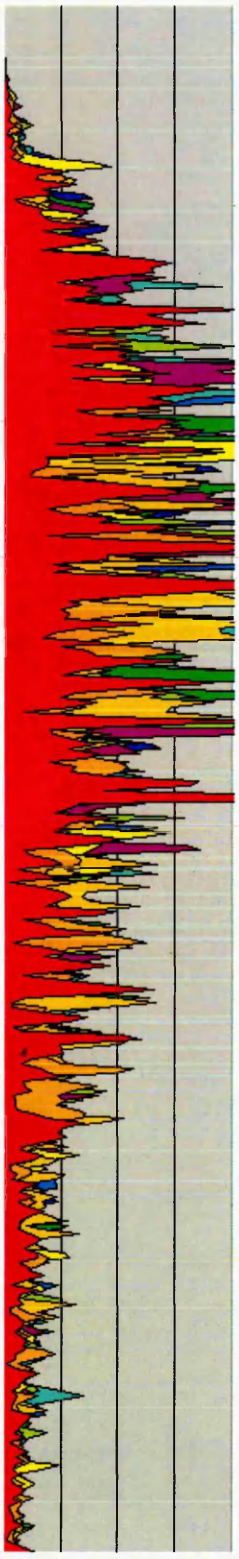

Fig. B.53
psd 65536 e11_2

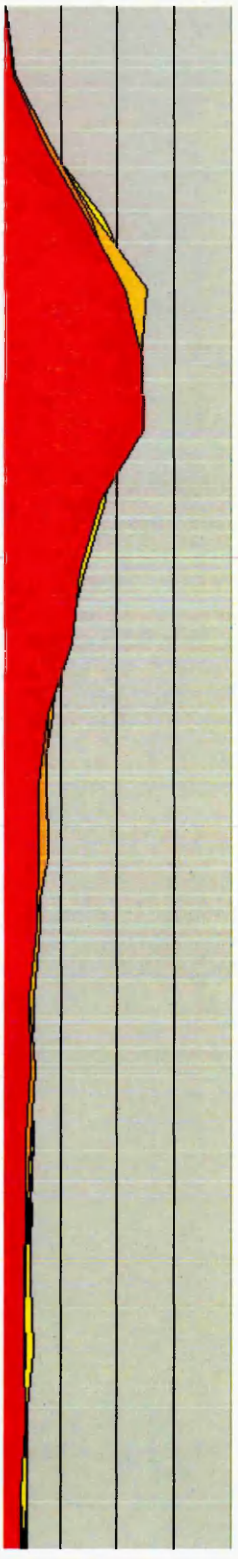

Fig. B.54
psd 4096 e11_3

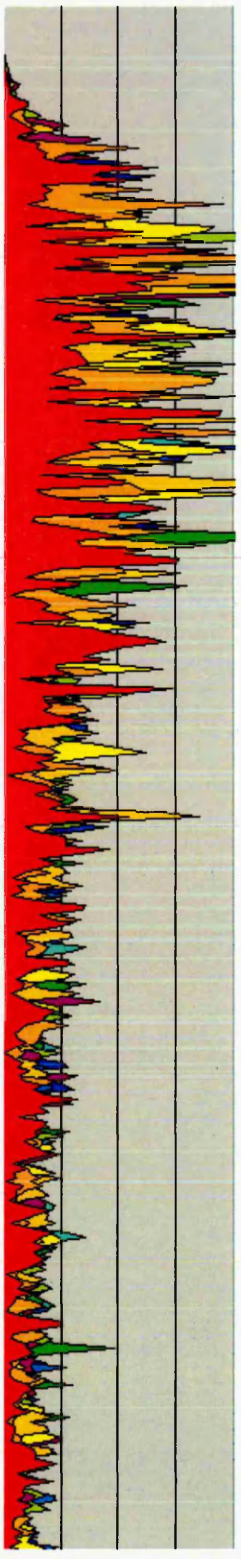

Fig. B.55
psd 65536 e11_3

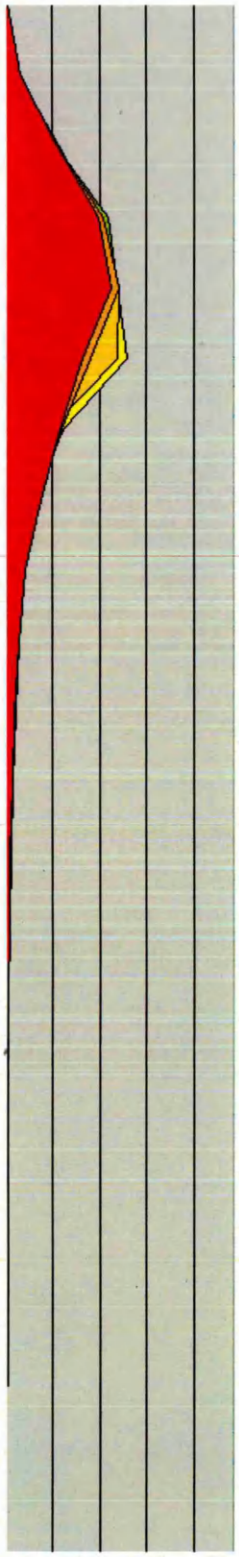

Fig. B.56
psd 4096 e11_4

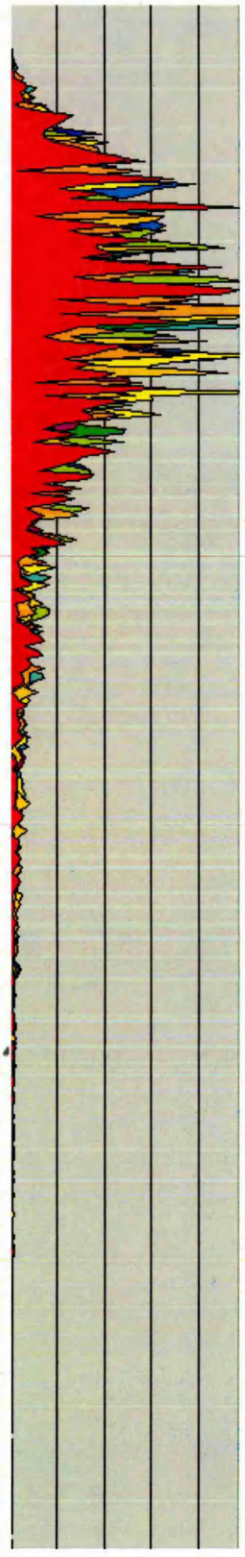

Fig. B.57
psd 65536 e11_4

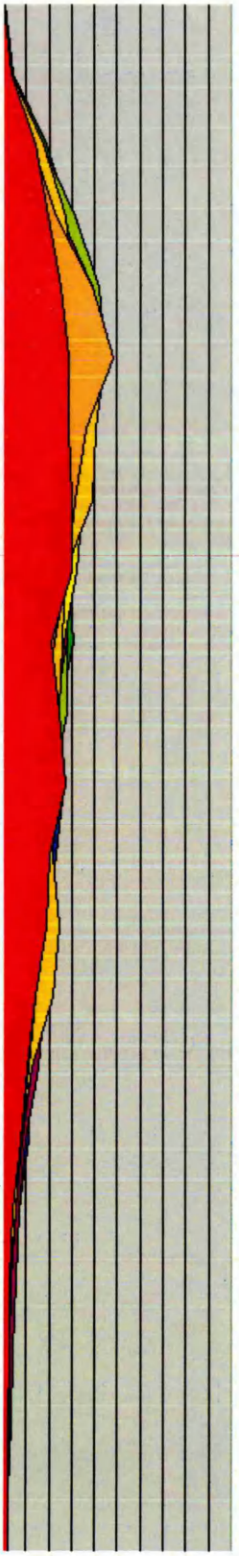

Fig. B.58
psd 4096 e3_2

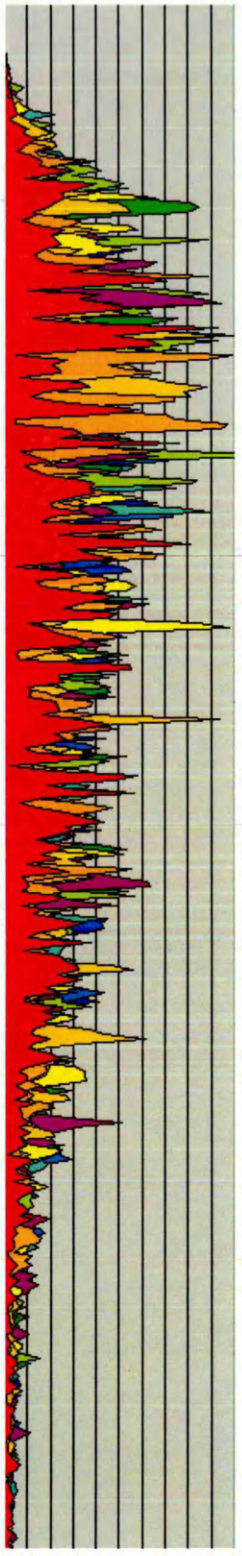

Fig. B.59
psd 65536 e3_2

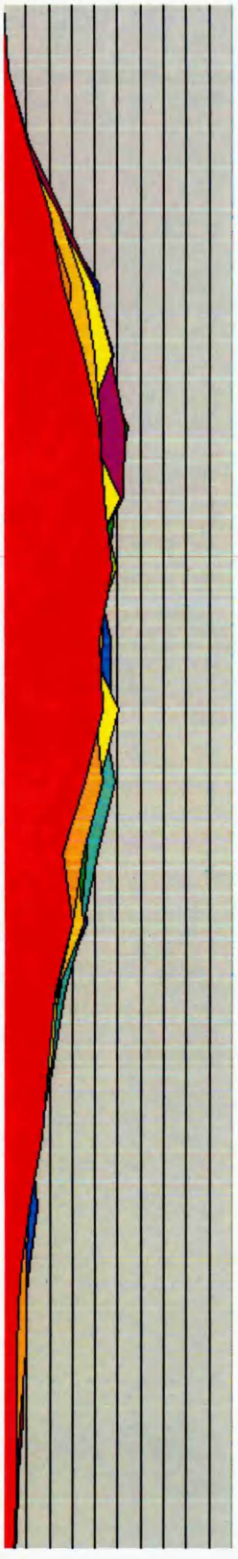

Fig. B.60
psd 4096 e3_3

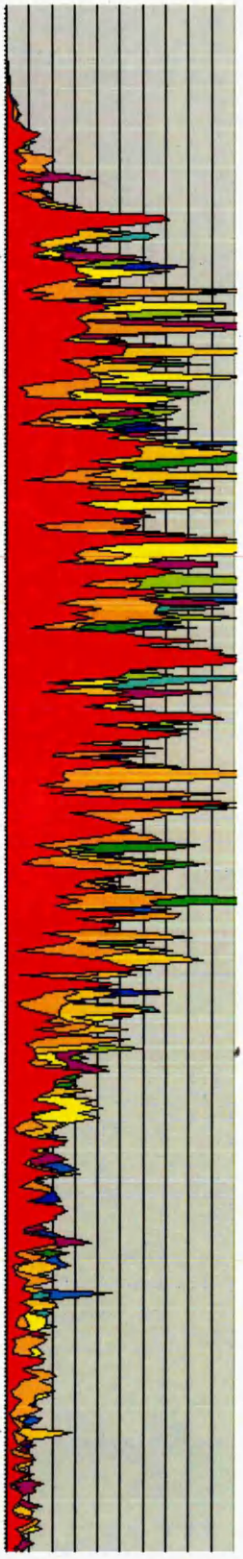

Fig. B.61
psd 65536 e3_3

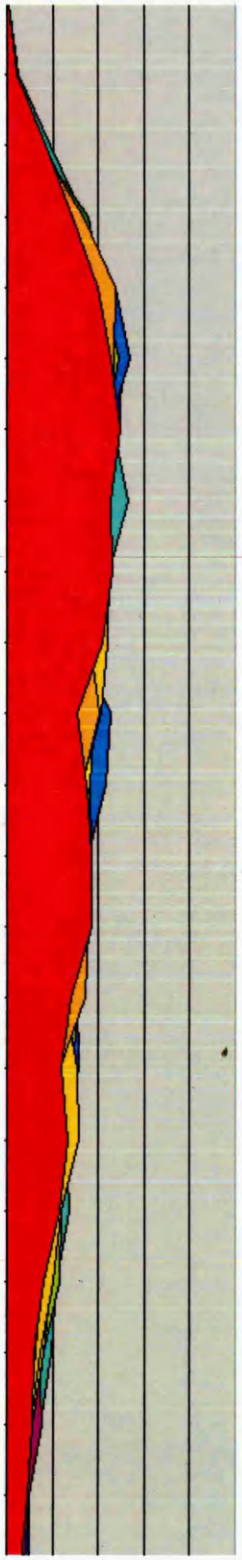

Fig. B.62
psd 4096 e6_1

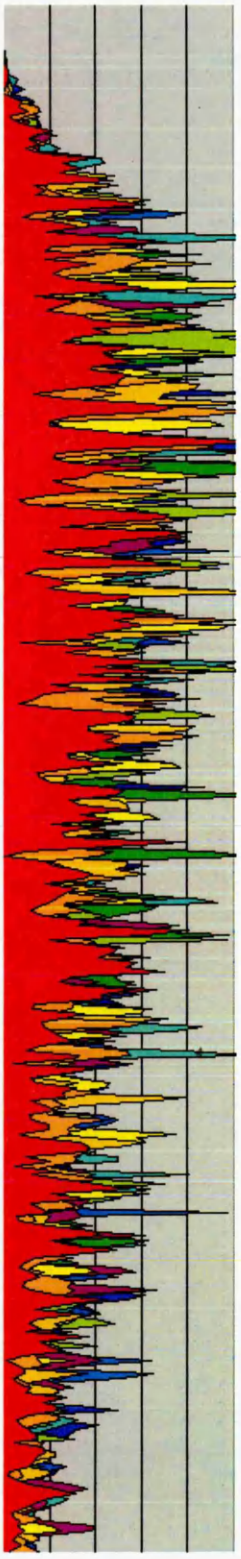

Fig. B.63
psd 65536 e6_1

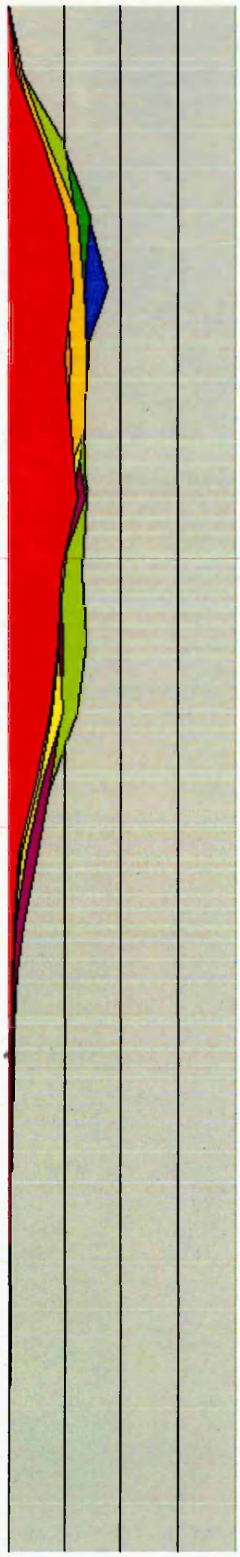

Fig. B.64
psd 4096 e6_2

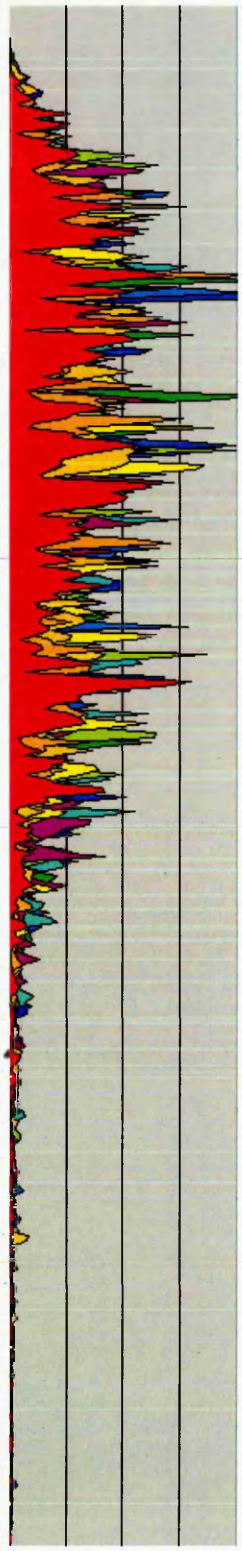

Fig. B.65
psd 65536 e6_2

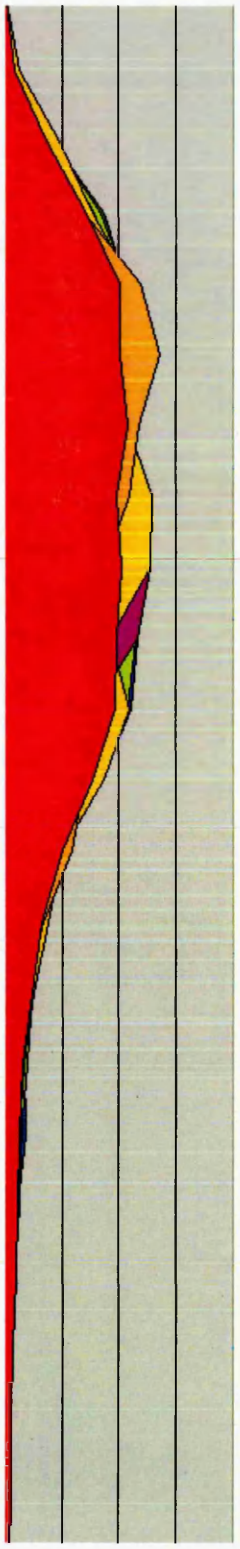

Fig. B.66
psd 4096 e6_3

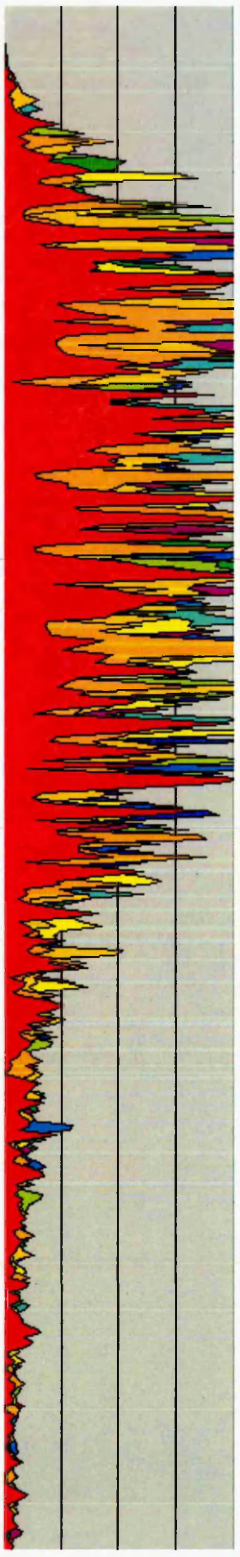

Fig. B.67
psd 65536 e6_3

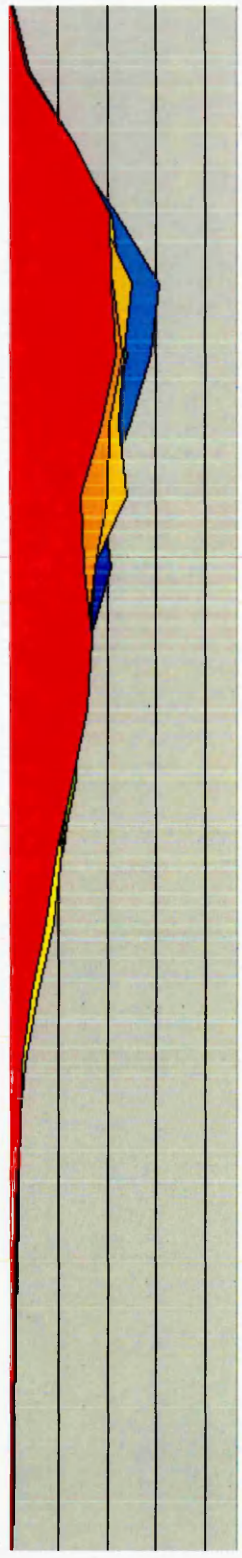

Fig. B.68
psd 4096 e6_4

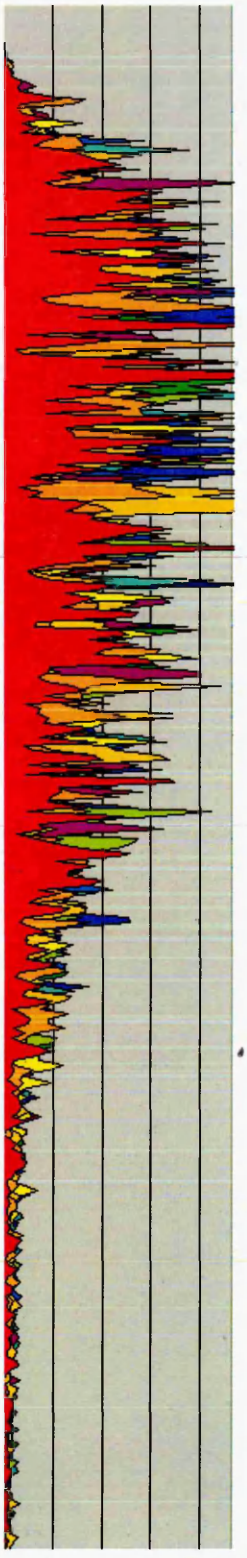

Fig. B.69
psd 65536 e6_4

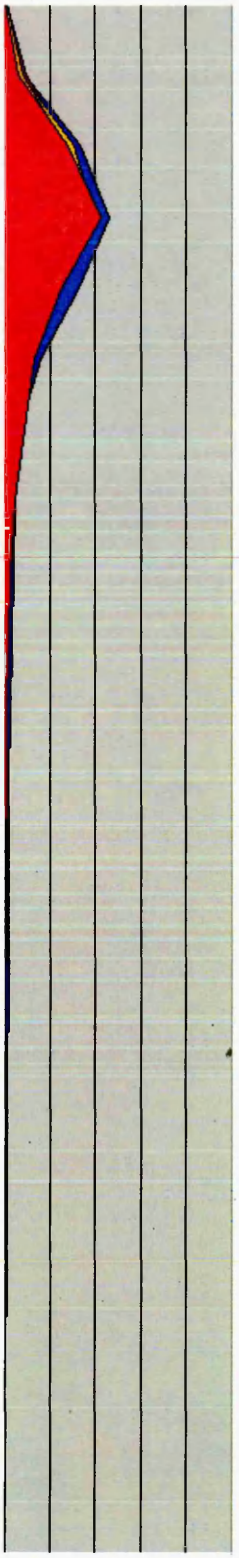

Fig. B.70
psd 4096 e7_1

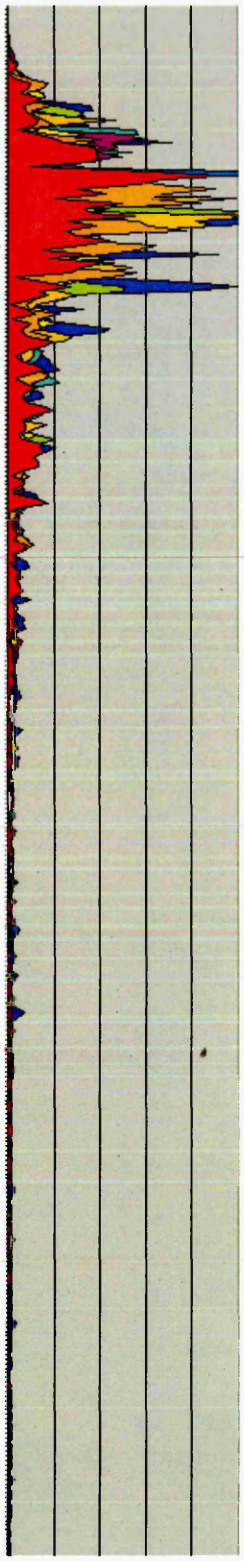

Fig. B.71
psd 65536 e7_1

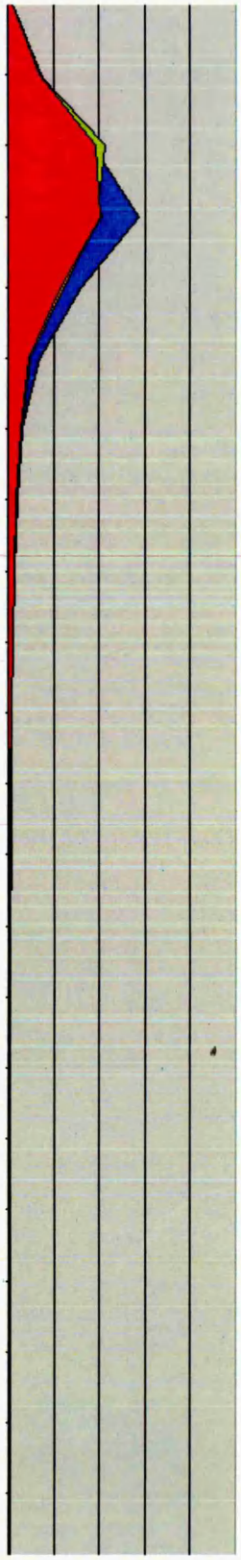

Fig. B.72
psd 4096 e7_4

Fig. B.73
psd 65536 e7_4

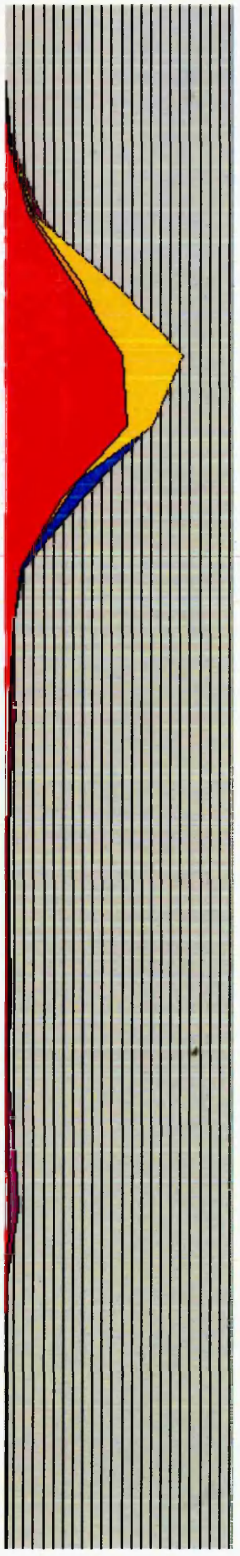

Fig. B.74
psd 4096 e8_1

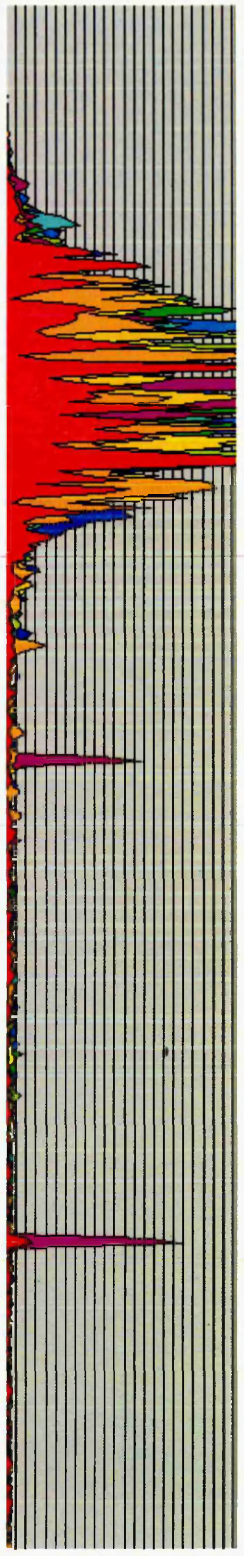

Fig. B.75
psd 65536 e8_1

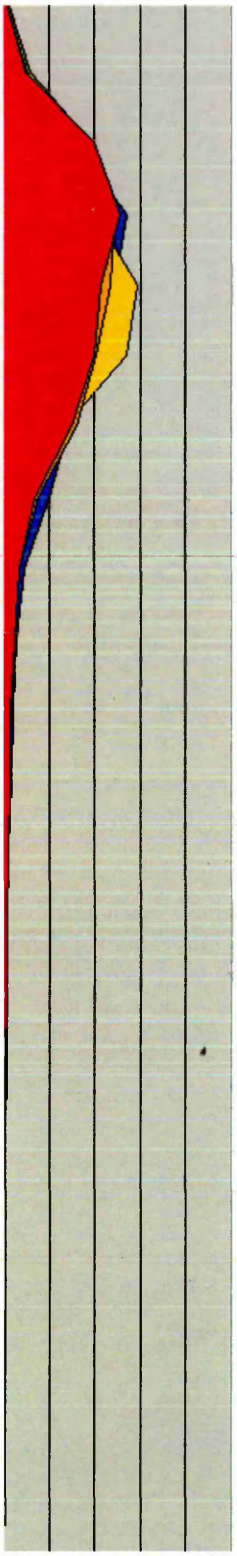

Fig. B.76
psd 4096 e8_4

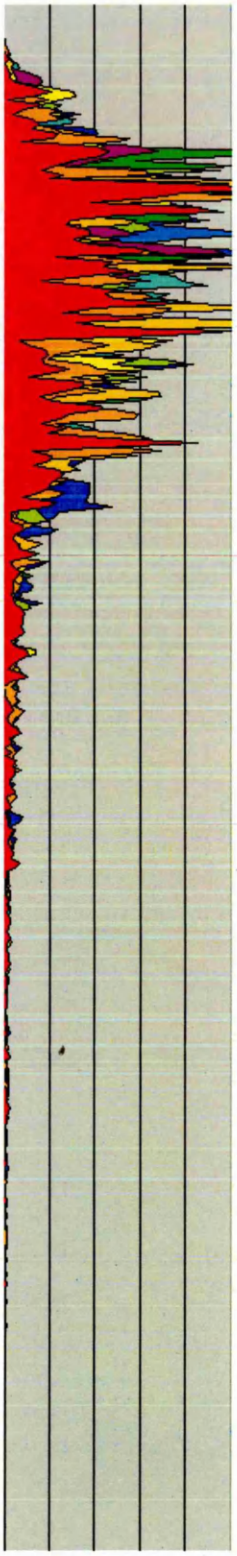

Fig. B.77
psd 65536 e8_4

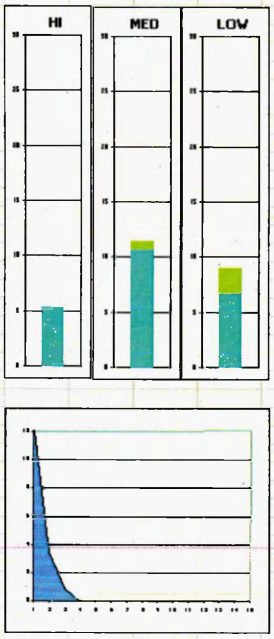

Fig. B.78 'fingerprint' e6_2

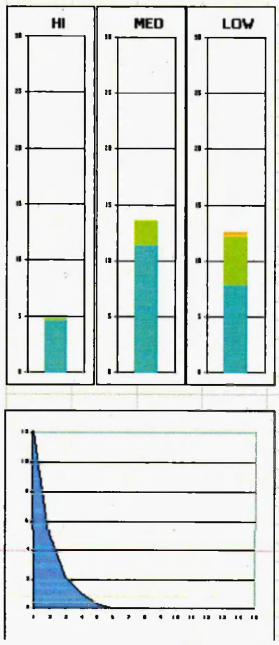

Fig. B.79 'fingerprint' e6_3

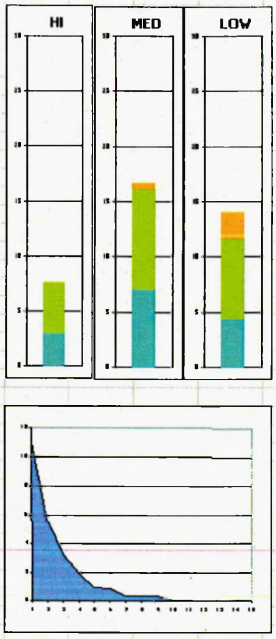

Fig. B.80 'fingerprint' e10_1

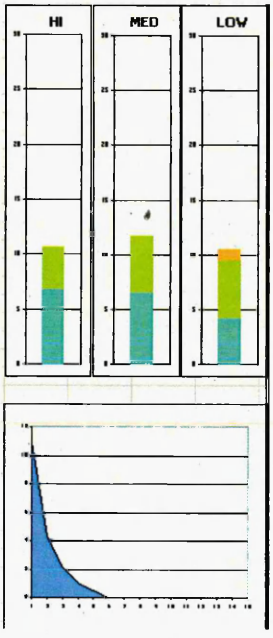

Fig. B.81 'fingerprint' e10_4

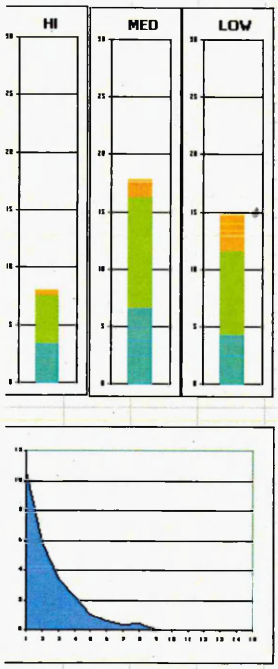

Fig. B.82 'fingerprint' e11_1

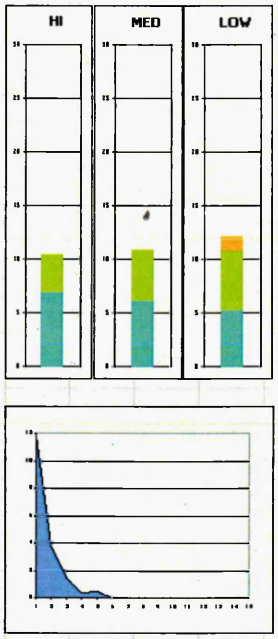

Fig. B.83 'fingerprint' e11_4

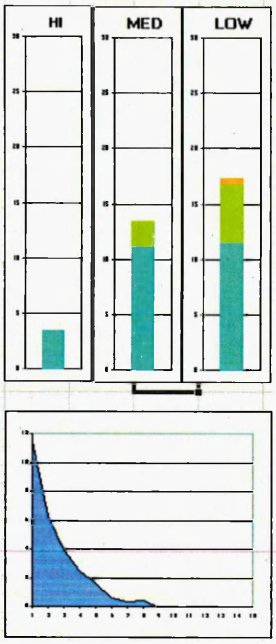

Fig. B.84 'fingerprint' e6_1

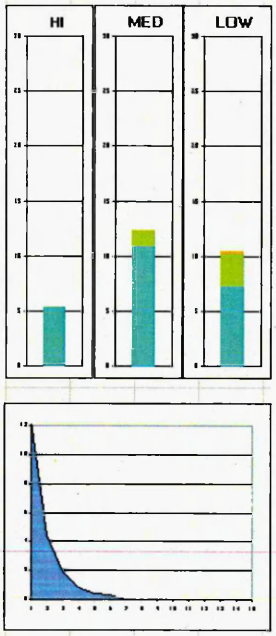

Fig. B.85 'fingerprint' e6_4

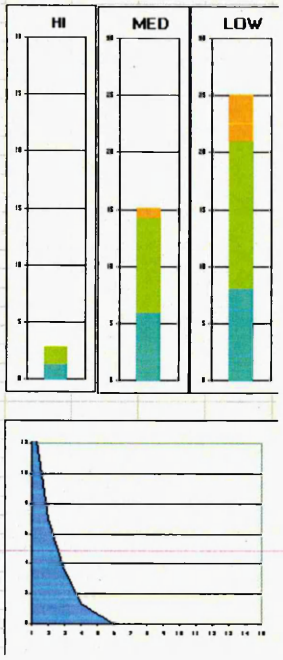

Fig. B.86 'fingerprint' e10_2

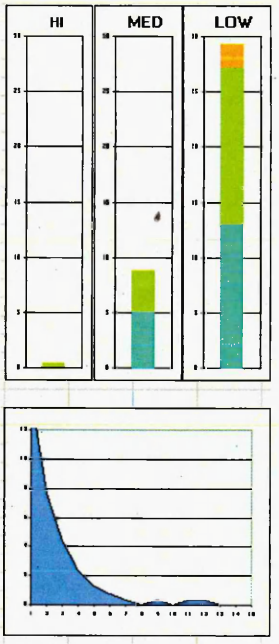

Fig. B.87 'fingerprint' e10_3

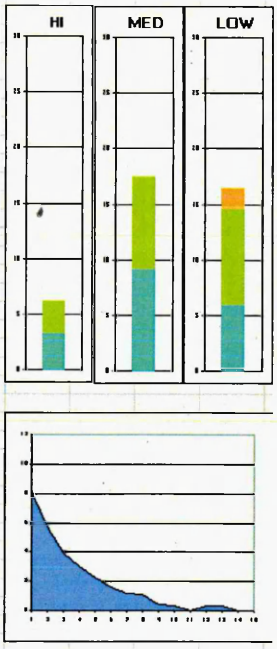

Fig. B.88 'fingerprint' e11_2

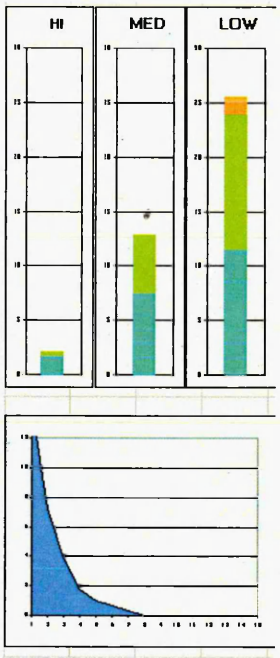

Fig. B.89 'fingerprint' e11_3

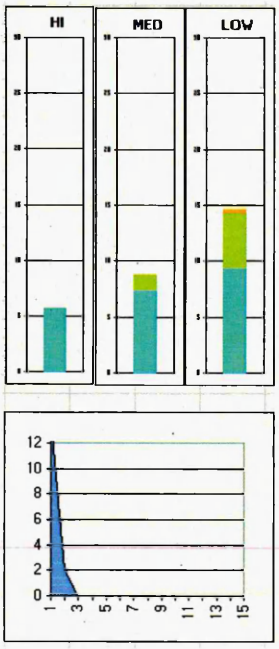

Fig. B.90 'fingerprint' e7_1

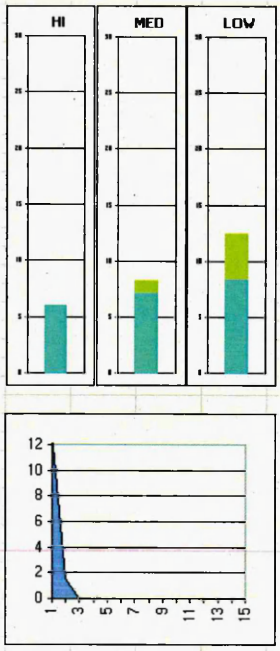

Fig. B.91 'fingerprint' e7_4

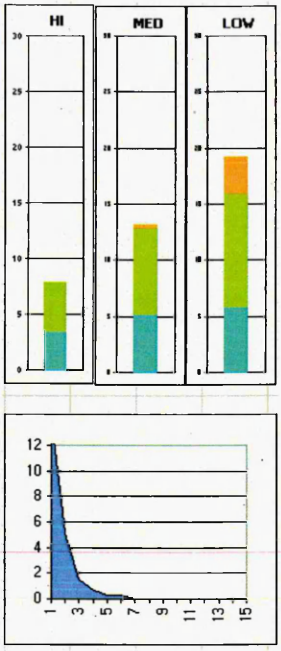

Fig. B.92 'fingerprint' e8_1

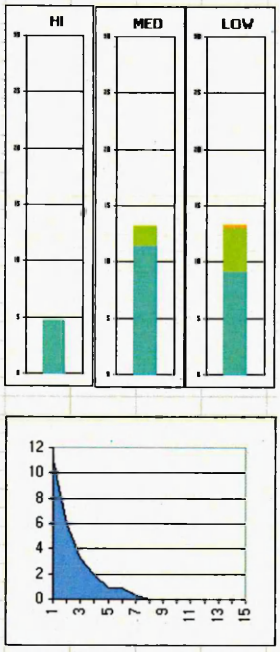

Fig. B.93 'fingerprint' e3_2

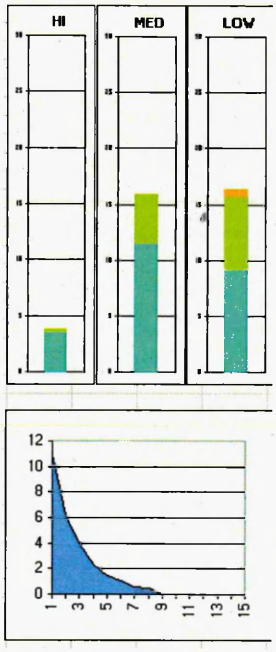

Fig. B.94 'fingerprint' e3_3

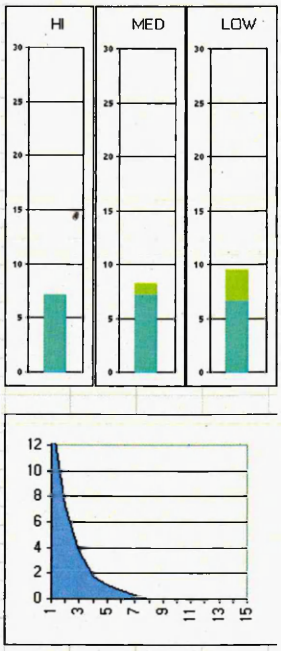

Fig. B.95 'fingerprint' e8_4

Appendix C - Diagrams related to Chapter 4

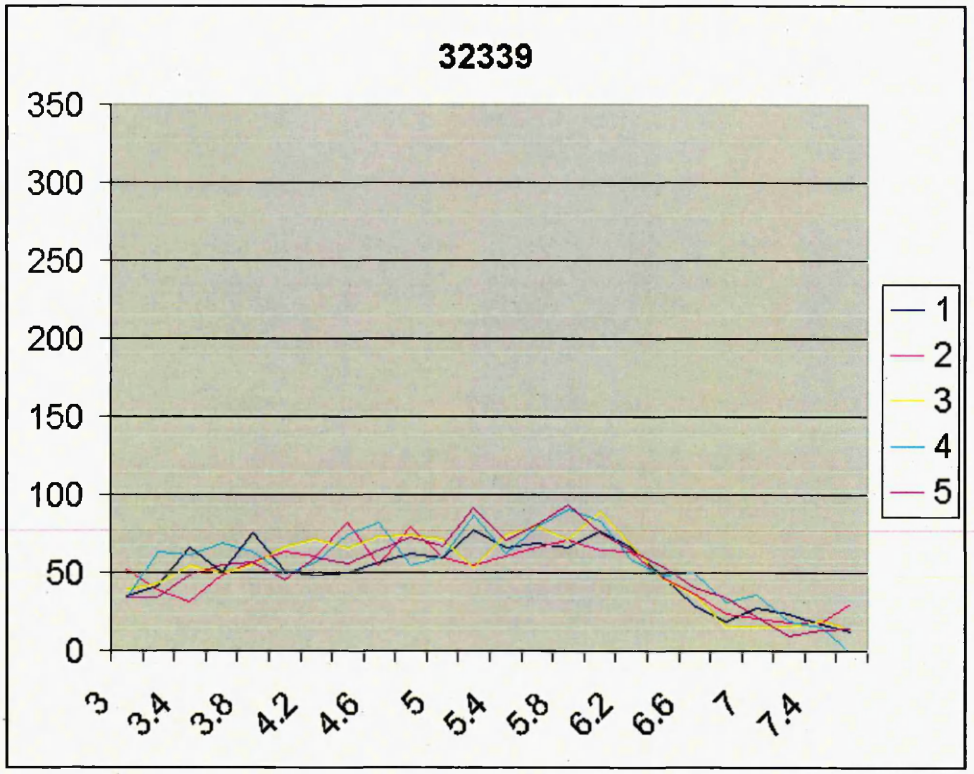

Fig. C.1 DEP collection spectum for Escherichichia coli strain 32339

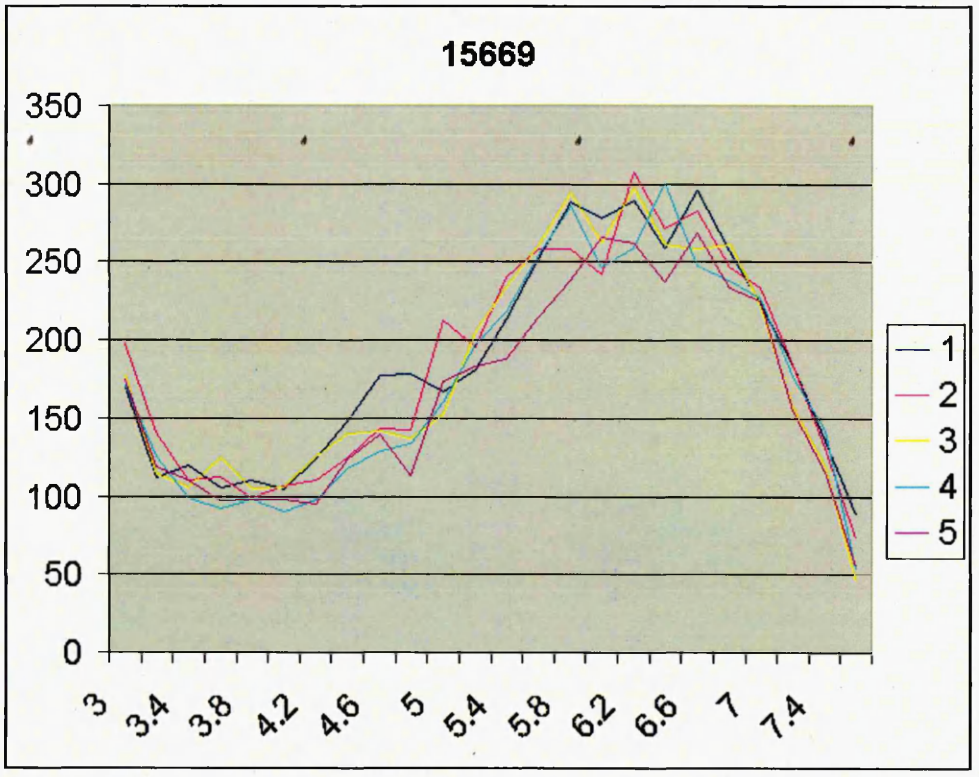

Fig. C.2 DEP collection spectum for Escherichichia coli strain 15669

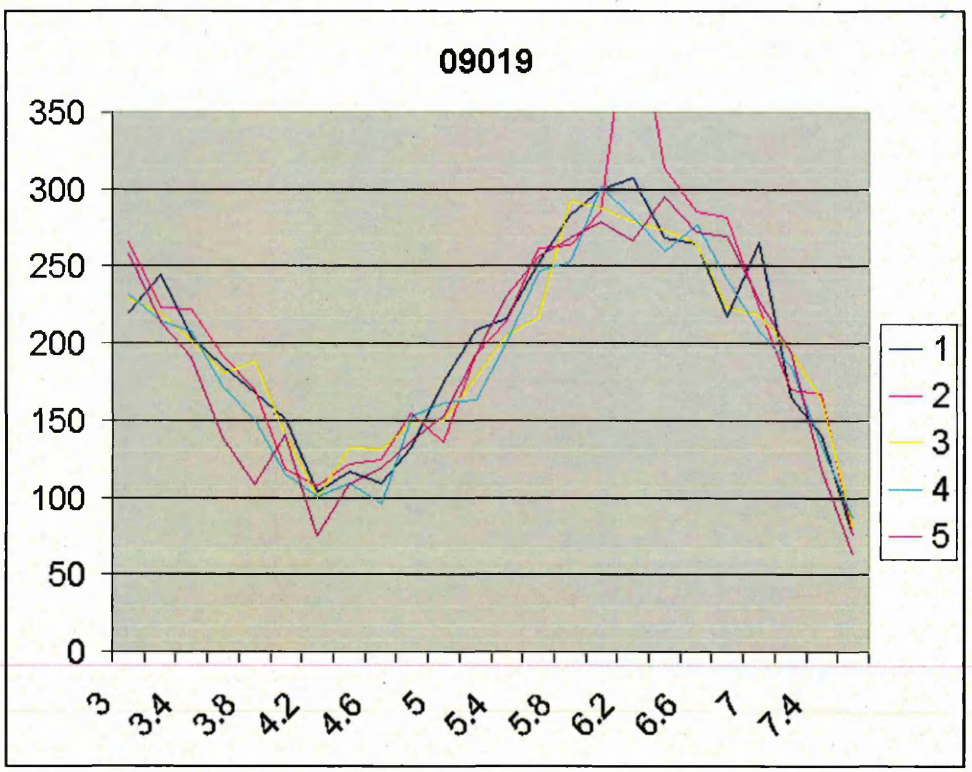

Fig. C.3 DEP collection spectum for Escherichichia coli strain 09019

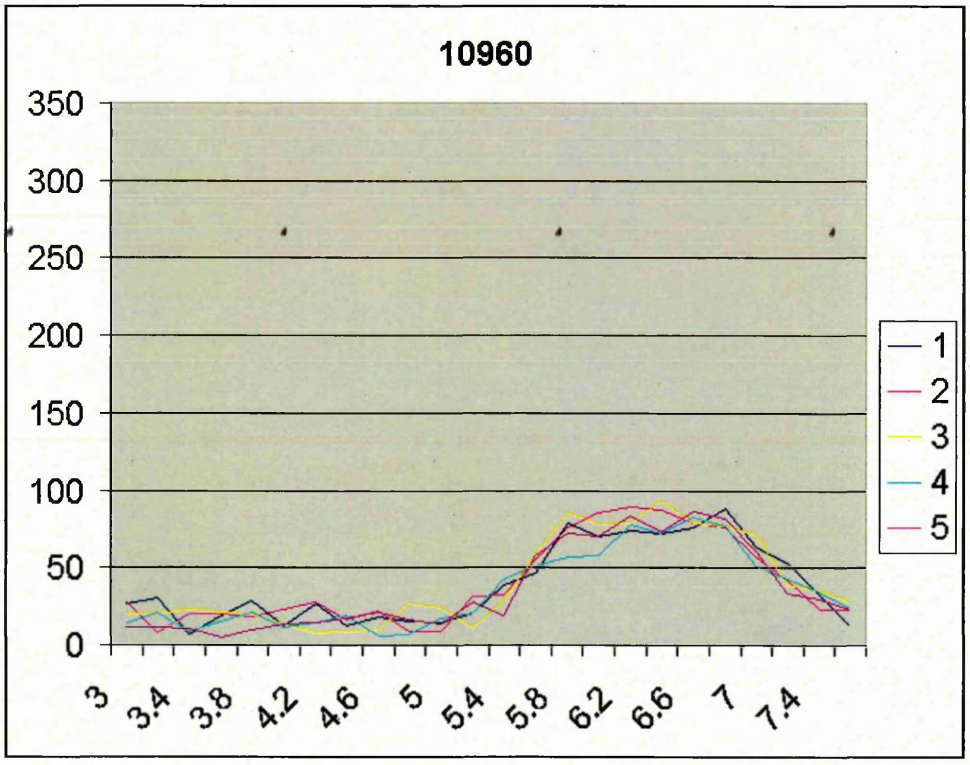

Fig. C.4 DEP collection spectum for Escherichichia coli strain 10960

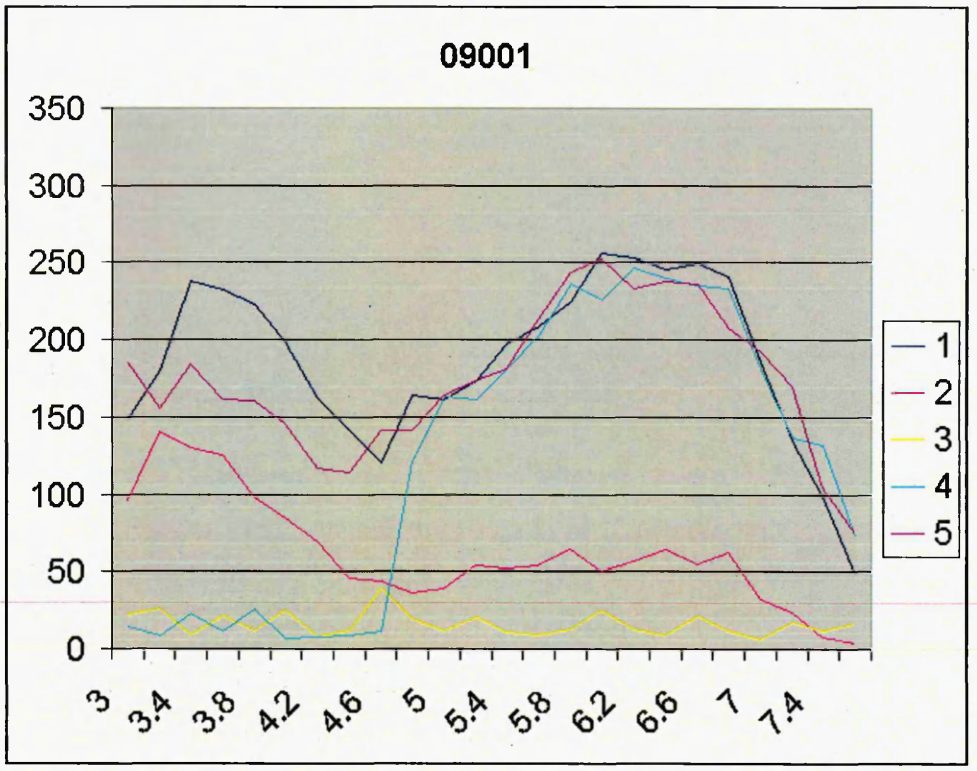

Fig. C.5 DEP collection spectum for Escherichichia coli strain 09001

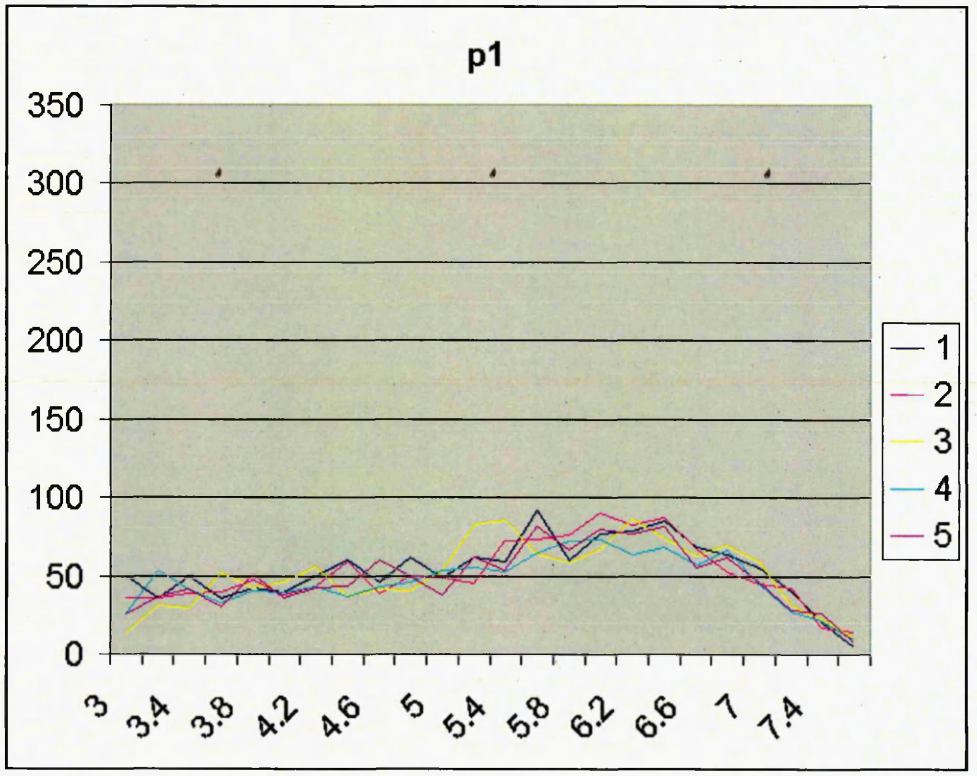

Fig. C.6 DEP collection spectum for Escherichichia coli strain p1

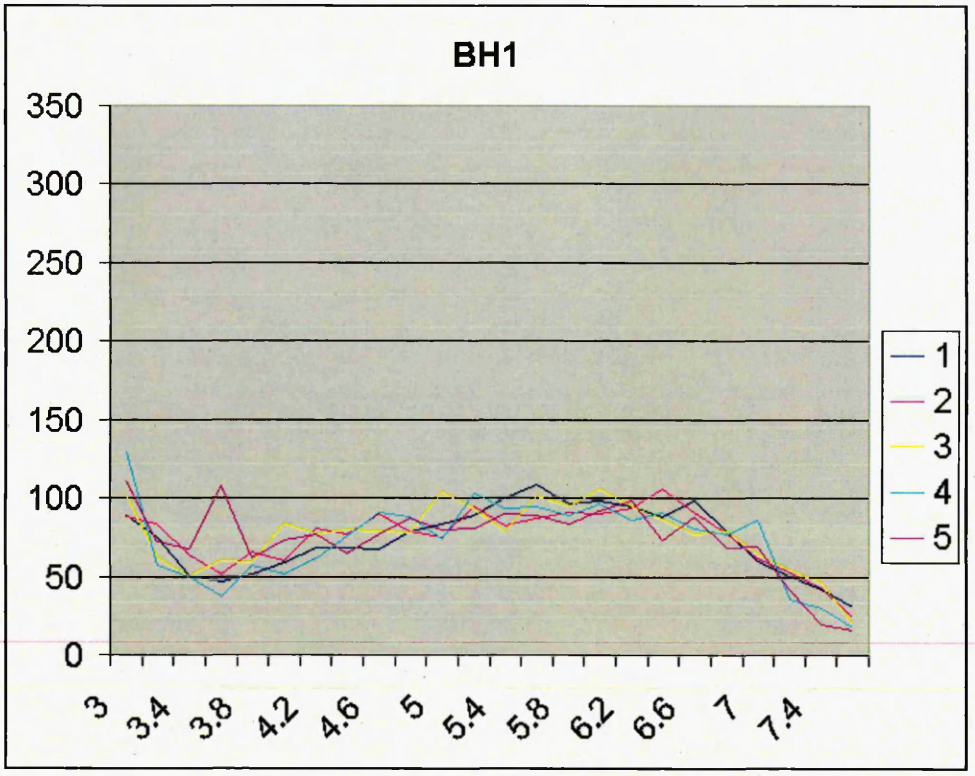

Fig. C.7 DEP collection spectum for Escherichichia coli strain BH1

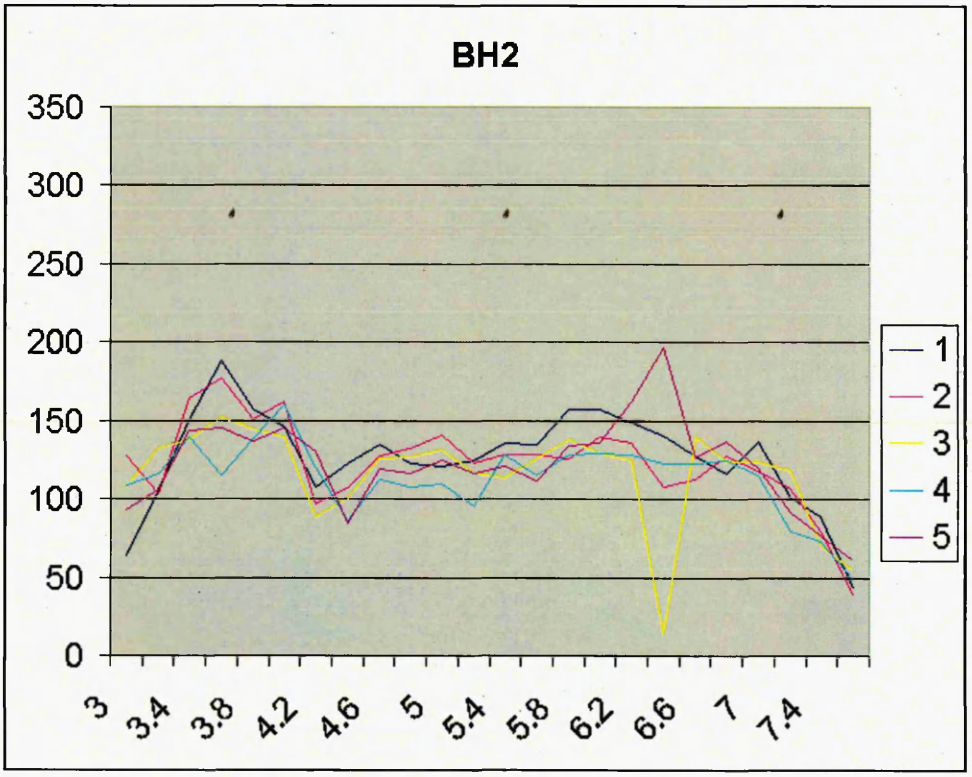

Fig. C.8 DEP collection spectum for Escherichichia coli strain BH2

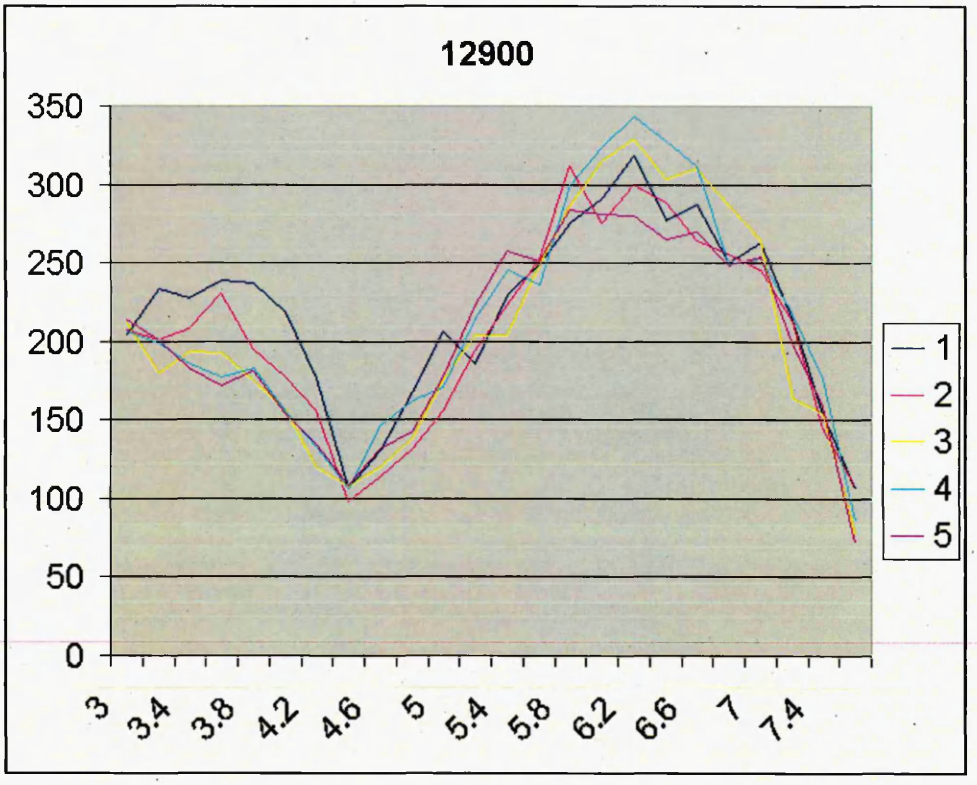

Fig. C.9 DEP collection spectum for Escherichichia coli strain 12900

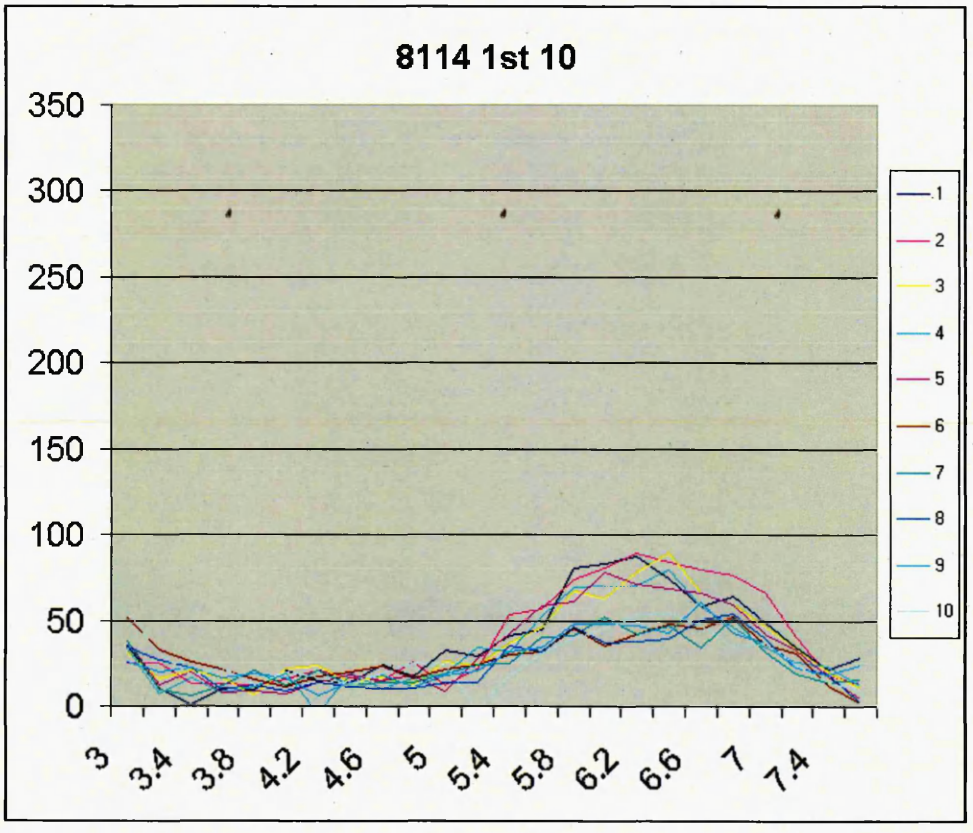

Fig. C.10 DEP collection spectum for Escherichichia coli strain 8114 (1st batch of 20)

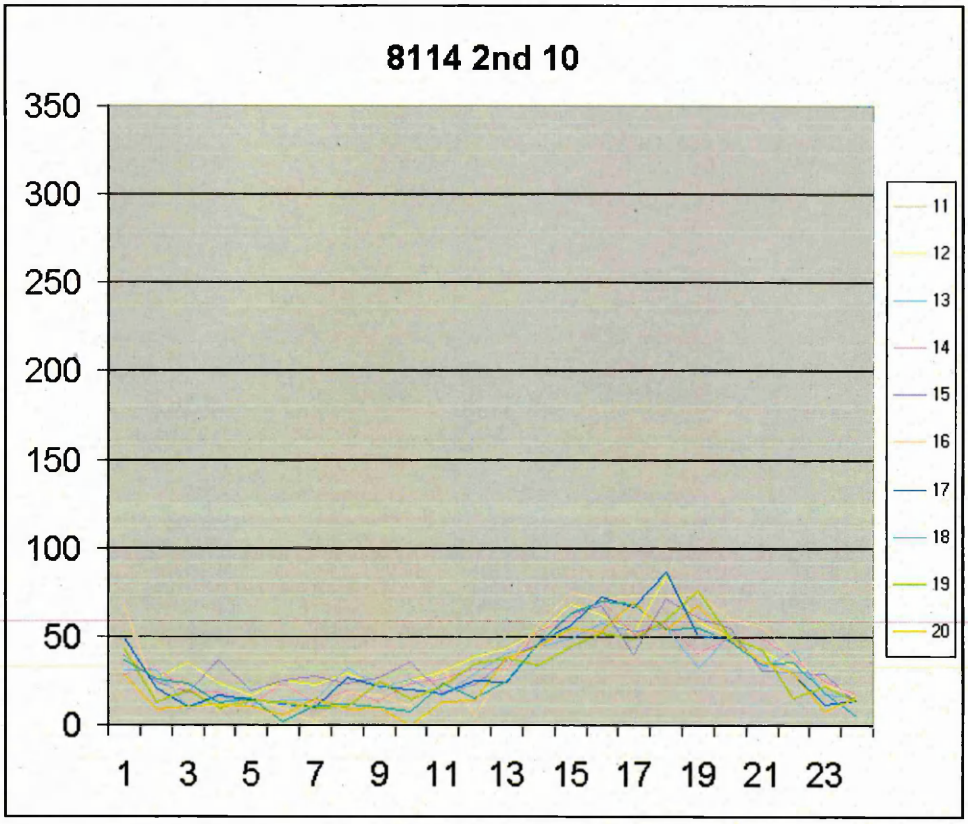

Fig. C.11 DEP collection spectum for Escherichichia coli strain 8114 (2nd batch of 20)