Algorithm for Recognition of Optical Spectra in an Environment Containing Interferants

B. Wade, C. Estes, C. Lloyd, and L. Powers
National Center for the Design of Molecular Function
Utah State University, Logan, UT, USA 84322-4155

Abstract - **An algorithm for quickly determining the presence of bacteria based on their intrinsic fluorescence signals has been developed. Applications of this algorithm are discussed.**
Keywords **- Algorithm, microbial detection**

I. Introduction

The constant threat of bacterial infection affects today's society, since these infections cause sickness and often death. These threats include bio-terrorism/bio-warfare, outbreaks of new and re-emerging diseases, and food and water contamination.

Current technologies detect microbes, but are limited in their ability to do so. Most require an outgrowth step, requiring the proper media and time (usually at least one day). In addition, these require sample contact. To overcome these limitations, a prototype instrument has been developed that detects bacteria based on their intrinsic fluorescence properties [see C. Lloyd, H.-Y. Mason, M. Dice, C. Estes, A. Duncan, B. Wade, W. Ellis, Jr. and L. Powers, these proceedings].

II. Detection algorithm development

A. Fluorophores

To detect bacteria, one needs a set of characteristics that determine when bacteria are present. Bacteria contain instrinsic fluorescent compounds, known as fluorophores, that allow for their detection. NADH, or reduced nicotinamide dinucleotide, and reduced nicotinamide dinucleotide phosphate [NADPH] are present in all living cells as they are involved in cell respiration. These compounds have well-documented fluorescence properties. Ca^{++} dipicolinic acid complex also fluoresces and provides excellent detection of spores, since it makes up approximately 5-15% of spore composition [1], and rarely occurs elsewhere in nature. Hemeproteins are found in the respiratory chain of microbial cells. These proteins are found in both live and dead cells and are used as fluorescent markers for both types of cells. However, pure spores do not contain these proteins, although they become visible soon after a spore begins to germinate. Finally, other pyridine nucleotides, in addition to NAD[P]H, are also involved in metabolism and are used in detection.

Each of these fluorophores is characterized by its excitation/emission energies. Simultaneously exciting bacteria with all of the excitation wavelengths from the above fluorophores reduces the chance of interference since the probability of an interference source duplicating the characteristics of more than one fluorophore is extremely small. This probability is further reduced since ratios of the intrinsic fluorescence signals are required to fall within a narrow physiological range.

B. Adjusting data

In our prototype instrument, the sample is excited and the fluorescence is collected using a reflective optic assembly [see C. Estes and L. Powers, these proceedings]. The light is then detected by an array of photomultiplier tubes (PMTs) and then amplified to a level readable by a standard A/D converter [see A. Duncan and L. Powers, these proceedings]. Data collected by the prototype instrument must be corrected in two ways. First, the control voltage for each PMT determines the range of the signals read into the A/D converters. The data from each PMT must be normalized to a specific control voltage. Second, the difference in transmission properties of the interference filters must be taken into account. This is done by mapping the control voltage-compensated values to another fluorescence instrument, a Shimadzu 5301, which has taken this consideration into account.

C. Separating fluorescence

This emitted light contains two main components: fluorescence from the intrinsic fluorophores and Rayleigh scattering. Since only the fluorescence provides detection information, the Rayleigh scattering is removed by finding two additional emission wavelengths at which essentially the entire signal is due to Rayleigh scattering. These two null emission points are used to fit a background curve using the Rayleigh scattering equation

$$I = a / l^4 + c ,$$

where a is due to the intensity of the incident light, the number of scatterers, the polarization, the distance between the scatterers and the observer, and the angle between the incident light and the observer, and c is due to instrumentation. This curve then allows for the subtraction of the Rayleigh scattering from the total emission signal, leaving only the desired fluorescence signals (Figure 1).

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25OCT2001	N/A	-

Title and Subtitle Algorithm for Recognition of Optical Spectra in an Environment Containing Interferants	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) National Center for the Design of Molecular Function Utah State University, Logan, UT, USA 84322-4155	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
3

III. Detection Algorithm Implementation and Results

Ratios of the signals from the fluorescent peaks are used as the basis for detection. Furthermore, the relative amounts of these intrinsic fluorescence signals fall within a narrow physiological range. Three different algorithms were examined to determine their ability to distinguish samples with bacteria from those without. For each algorithm, samples from different species of non-pathogenic bacteria in a minimal media provided the data for samples containing bacteria, while several sources of sterile bacterial media provided the data for samples without bacteria.

Before testing the three algorithms, pairs of ratios were plotted against each other to investigate the separation of the two sample types and the detection capabilities of the algorithms. A typical ratio pair plot as seen in Figure 2 showed that the two sample types had enough separation to guarantee high detection and low false alarm rates.

A. Neyman-Pearson test

The Neyman-Pearson test was examined first. This test gives a probability of detection, p_D, the probability that the test states that bacteria are present when they indeed are, assuming a probability of false alarm, p_{FA}, the probability that the test states that bacteria are present when in fact they are not. This test requires a knowledge of the ratio distributions for both types of samples. A plot of the p_{FA} versus the p_D gives the receiver operating characteristic (ROC) which graphically presents the best choice for p_{FA} and p_D. While the computations needed to produce the ROC for this case required extensive computing, applying the test only requires a few multiplication and comparison steps.

The Neyman-Pearson test gave excellent results. While assuming a p_{FA} of only 0.01%, the test stated it could detect

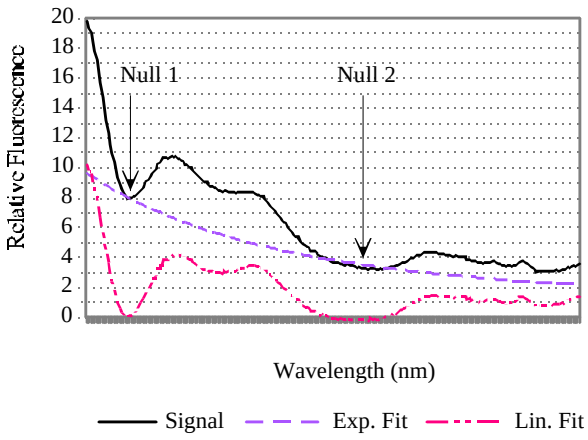

Fig 1. Emission spectra from *Escherishia coli*.

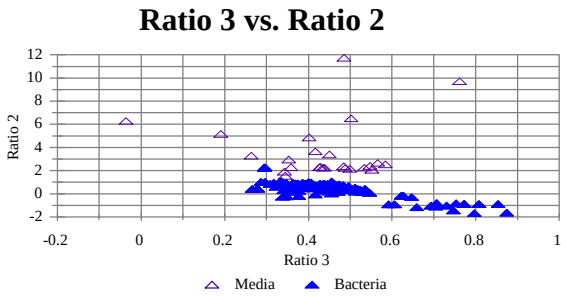

Fig 2. Typical ratio pair plot.

bacteria with a p_D of more than 99% (Figure 3). While this works well for this binary case, extending the capabilities to detect live cells from dead cells would require extensive mathematics and computing power to compute the ROC. Also, the ratios were assumed to be normally distributed, which they were not. These two issues encouraged the use of other algorithms.

B. Fuzzy logic

Fuzzy logic was employed next. Fuzzy logic explicitly describes the degree to which an object belongs to a set by assigning it a fuzzy variable, a number between 0 and 1, inclusive, with numbers closer to 1 determining that an object belongs to the set with higher degree. In this application, fuzzy variables describe the degree to which a set of ratios from an acquired sample belongs to the set of ratios described when no bacteria are present versus those with bacteria present. Numbers above 0.5 identify the sample as containing bacteria while those below 0.5 as not containing bacteria.

Fuzzy logic also worked well. It successfully classified ~91% of the samples with bacteria correctly and 100% of samples without bacteria correctly. While it gives good detection results and takes less computing power to determine its detection capabilities than the Neyman-Pearson test, its

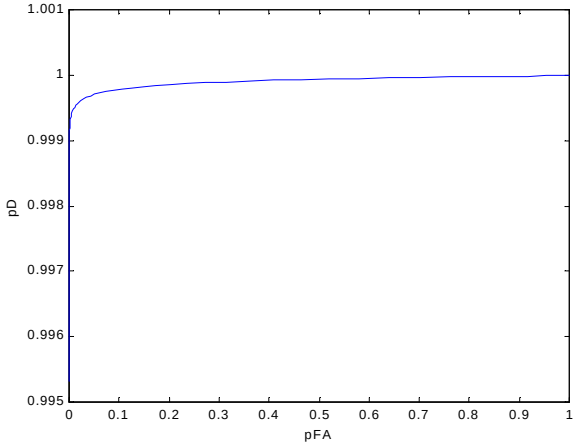

Fig 3. ROC for Neyman-Pearson test.

greatest shortcoming is that the test is not necessarily optimized.

C. Neural network

The last algorithm examined was a neural network known as a multilayer perceptron. This neural network contains an input layer, an output layer, and one or more hidden layers, each with a given number of processing blocks known as neurons. The neurons for each layer are interconnected by weighted connections which are linearly summed at each neuron and passed through a nonlinear function. The ratios are fed into the input layer, processed in the hidden layers, and then output on the output layer with a number between 0 and 1, with numbers above 0.5 identifying the sample as containing bacteria and numbers below 0.5 as not containing bacteria.

To function properly, the neural network must first be trained with data collected from samples that contain bacteria and samples that do not contain bacteria to obtain the weights for each connection. Different configurations of neural networks with 1-2 hidden layers and 1-7 neurons per hidden layer were trained using this data and tested with additional data. The neural network producing the greatest accuracy with the test data was chosen as the one used to implement the detection algorithm.

The neural network provided similar detection capabilities as the Neyman-Pearson test while providing its results in about the same amount of time as the fuzzy logic. Several neural network configurations successfully classified 100% of the bacterial samples and 100% of the non-bacterial samples.

These results verified the expectations from the ratio pair plots.

IV. Conclusion

Detecting bacteria using their intrinsic fluorescence properties provides an excellent detection method that is fast and does not require sample contact. These results suggest that the neural network provides the best solution for the detection algorithm. The algorithm can now be further developed to detect bacteria on real-world samples, such as medical instrumentation and food samples.

References

[1] R. Lundin and L. Sacks, “High-resolution solid-state ^{13}C nuclear magnetic resonance of bacterial spores: Identification of the alpha-carbon signal of dipicolinic acid,” *Appl. Environ. Microbil.*, vol. 54, no. 4, pp. 923-928, Apr. 1988.