SIGNAL PROCESSING HIERARCHIES FOR PORTABLE, LOW-POWER SAW-BASED CHEMICAL SENSING SYSTEMS

Denise M. Wilson
Department of Electrical Engineering, University of Washington
Box 352500, Seattle, Washington 98195-2500

Thaddeus A. Roppel
Department of Electrical and Computer Engineering, Auburn University
Auburn, Alabama 36849

ABSTRACT

This paper presents a comparison of multi-stage linear and single-stage nonlinear processing techniques for accomplishing chemical discrimination. Data variance, computational overhead, and memory storage requirements are compared between (linear) multistage principal components analysis, non-linear VERI (visually empirical region of influence) and non-linear artificial neural network techniques. For the linear techniques, data variance is reduced by 88%, compared to that of raw data during data preprocessing. Computational overhead is reduced up to 82.5% and 77% for non-linear clustering and artificial neural network techniques respectively. These improvements offer clear promise for significant reduction in power and space consumption for portable chemical sensing systems design.

INTRODUCTION

In the last decade, linear signal processing techniques for chemical sensor arrays have largely been reserved for preliminary analysis of array performance, for array optimization, or for simple sensing applications involving few analytes or few sensors (e.g. [1]-[3]). Non-linear techniques, such as artificial neural networks(e.g. [4]-[7]), non-linear clustering (e.g. [8]), genetic algorithms (e.g. [9]), fuzzy logic (e.g. [10][11]), and similar techniques. Principal component and cluster analysis, in particular, is a linear processing technique that has frequently been used to quickly evaluate, visually and quantitatively, the ability of an array to separate characteristics of individual chemicals. However, as the number of analytes associated with a particular application increases, principal components analysis quickly loses its effectiveness as cluster overlap increases, making discrimination difficult and reducing accuracy. In many cases, non-linear signal processing techniques have proven effective in solving discrimination problems that simply cannot be solved using linear processing techniques. However, most effort in both non-linear and linear signal processing targeted at chemical discrimination has been done without data preprocessing and in a single stage of processing. In this paper, we intend to support the argument that data preprocessing and multi-stage analysis can reduce superfluous information and optimize computation capability sufficiently to allow linear processing techniques to effectively solve discrimination problems.

The inherent limitations of linear signal processing are also what make them attractive to portable sensing systems. Linear signal processing, especially principal components analysis, is often conducive to hardware implementation and requires relatively few floating point operations compared to nonlinear signal processing techniques. Fewer floating point operations translate directly to decreased computational overhead and increased speed of operation. Data compression, minimal memory requirements, and low computational overhead are essential for the design of portable sensing systems that meet user-defined weight, battery-life, and size constraints.

As a proof of concept, this paper presents linear signal processing techniques for implementation in portable, low-power SAW-based chemical sensing systems. This paper provides a demonstration of how three-stage signal processing involving a combination of principal components analysis (PCA) and linear signal preprocessing can provide performance for chemical discrimination comparable to more complex, non-linear analysis techniques, but at significantly reduced numbers of floating point operations.

THEORY

A linearly separable problem, whether two-dimensional or thirty-dimensional, can be solved using linear signal processing techniques. The issue in determining linear separability, however, lies in the ability to visualize the problem, which is often not possible in dimensions greater than three or arrays that contain more than three elements. A number of software programs are available to assist in examining multi-dimensional data, but are limited in their ability to provide all possible perspectives of the data and, in addition, can only provide one (or few) of many perspectives at a given moment in time. For this reason, it is often more straightforward to allow a non-linear signal processing method to determine the separability of the data using an algorithm that has been proven mathematically to converge for a solvable problem. Efforts to discriminate among analytes using arrays of chemical sensors, more often than not, rely on such non-linear signal processing techniques as multi-dimensional, non-linear clustering, artificial neural networks, fuzzy logic, genetic algorithms and others. In this work, we seek to establish linear separability of the problem using multiple (two) stages of signal processing performed in sequence rather than the single stage that characterizes most other efforts. Once linear separability is established, the problem of analyte discrimination, in field operation, is completed more efficiently using linear signal processing techniques. Ultimately, in power and space limited applications for portable chemical sensing, we wish to minimize the number of mathematical operations and stored parameters needed to solve the discrimination problem.

In this work, we demonstrate that an analyte discrimination problem that has been solved using a single-stage, non-linear signal processing technique can also be solved using two stages of linear

signal processing. The problem is that of a 7-element SAW device array used to discriminate 16 analytes that consist of aliphatic, aromatic, and chlorinated hydrocarbons, alcohols, ketones, and organophosphorous compounds across a range of concentrations from 0.40 - 48.0 P/P_{sat}, where P and P_{sat} are the partial pressure and saturated partial pressure of the analyte of interest respectively. This wide range of analytes and concentrations requires a large computing space. Our first step is to reduce the dynamic range of the data by reducing the influence of concentration. After examination of the data, 59 of 112 possible response characteristics (isotherms) across all sensors and analytes in the array are linear. Because the array characteristics are dominated by linear response curves (53%), we logically choose a linear normalization method to reduce the influence of concentration. During normalization, every point on every response curve in the training set is normalized to the value of coating 1, the SAW device coated with bare quartz; this reference sensor demonstrates linear response characteristics more frequently (69%) than other coatings for all 16 analytes and is used to normalize all response points for the remaining six coatings as follows:

$$\Delta f_{normalized(a, b)} = (\Delta f_{a, b})/(\Delta f_{1, b}) \qquad (1)$$

where Δf refers to the frequency shift experienced by the SAW device in the presence of an analyte, a refers to the coating (numbered 1 through 7) and b refers to the analyte type (1 - 16).

After normalization, a standard linear signal processing technique, principal components analysis, is used to transform the seven dimensional data into two dimensions which contain most of the variance in the original data. The data are then analyzed using standard linear regression techniques (fitting the points to a line) and ambiguous classes are determined (analytes that are not distinguishable in the first stage of processing). A second stage of linear signal processing (principal components analysis) uses the most distinguishable features of the ambiguous classes of the first stage. Linear regression of the first two principal components in this second stage then discriminates or separates the remaining analytes in the test set. Once the principal components are determined using the test set, they can be implemented in field discrimination tasks by storing small, two-dimensional matrices containing the principal components. PCA, in contrast to non-linear signal processing techniques, also involves a significantly reduced number of mathematical operations required to multiple each seven dimensional data point by its 7X2 principal component matrix and then, to evaluate its distance to the 16 lines that characterize the analytes in the training set.

Any point, once transformed, that lies further than two standard deviations from any of the 16 lines in the training set, is labelled an outlier. In this way, unusual ambient conditions, new analytes, or broken sensors can be flagged because each data point is not necessarily forced into a particular classification group. The conversion of this problem to linear signal processing guarantees a significant reduction in the computational demands of the system which can provide lower power, more compact operation or more computational capability to address erroneous or unexpected conditions in the environment or sensor array.

In summary, the 7-element, SAW device array is processed to discriminate 16 analytes using:

- Multi-stage principal components analysis: Stage 1 is pre-processing (normalization of data); Stage 2 is principal components analysis on all 16 analytes simultaneously; Stage 3 is principal components analysis only on members of classes

that could not be separated in Stage 2. Stage 3 is preceded by feature extraction which extracts the most important features of each group to discriminating members of that group.

- VERI: creates a non-linear, multi-dimensional cluster that encompass all training points for a particular analyte. Distance from each testing point to all 16 clusters is then evaluated to determine which (if any) cluster the training point belongs.

- Artificial Neural Network: uses a fully interconnected, feed forward, back propagation trained perceptron-style architecture (7 input nodes, 14 hidden nodes, 16 output nodes) to train on the data set.

In the next section, we discuss the results of linear analysis of a test set for all 16 analytes and compare computational overhead (clock cycles and memory) for this technique and the two other common, non-linear techniques (VERI, Artificial neural networks) used to process the outputs of the SAW array.

EXPERIMENTAL DETAILS

The array of sensors used in this demonstration is a 7-element array of SAW sensors fabricated and tested at Sandia National Laboratories, each coated with a different material and tested across a wide range of concentrations (0.40 - 48.0 P/P_{sat}) for 16 analytes of interest (see Figure 1). The 7-element array is a reduced form of a 30-element array optimized using a nonlinear signal processing program called VERI (visually empirical region of influence) which relies on human visual perceptions of clusters to separate analytes from one another; although VERI works as well with linear clusters as non-linear, most of the clusters in this analysis are determined to be non-linear. The seven coatings used in this array are a bare quartz coating, a dendrimer coating, two polymer coatings (SE-30, PECH), and three metal coatings (CuO, Au, Pt). Details of the SAW device fabrication and the array optimization and non-linear clustering routine (VERI) can be found in 12 and 8 respectively.

The first stage of the linear signal processing process is preprocessing, which is designed to reduce the influence of parameters that are not relevant to analyte discrimination. In this case, the largest variance in the raw data. is concentration, which is reflected in the principal component analysis of the raw data in Figure 1a. Normalization according to the most linear coating (the bare quartz reference) results in an 88% reduction in data variance (corresponding to variations in the first principal component) as shown in Figure 1b.

The second stage of processing is a straightforward principal components analysis of the data set after normalization (Figure 1b, 1c, 1d). After PCA, all but 5 of the 16 analytes separate into clearly defined clusters, where members of each cluster can be identified within 100% accuracy using a simple linear regression analysis. The remaining 5 analytes separate into two classes, one containing three analytes, cyclohexane(a), hexane(b) and isoctane(c) and the other containing two analytes, propanol(k) and DMMP(p), which are differentiated in the third and last stage of signal processing (Figure 1e, 1f) by first preprocessing the normalized signals used as the input to stage 2. Preprocessing involves highlighting the differences between members of a group. It is the differences between coatings 6 and 3, 7 and 6, and 4 and 1 that dominate the distinguishing features of the response curves in each of the two

classes. After preprocessing to extract only these features, another principal components analysis yields the results in Figures 1e and 1f where the five remaining analytes are clearly distinguishable.

Comparisons in computational overhead for these three types of processing these 16 analytes in this SAW array are made in Table 1. The Motorola MC68HC12 is used to translate the number of add, multiply, and compare operations in each technique to clock cycles which translate directly to processing time and power consumption. The linear processing techniques consume 82.5% and 77% less processing time than the VERI and ANN methods respectively, with no loss in accuracy. The linear techniques also consume 91% and 71% less memory than the VERI and ANN methods respectively. All three pattern recognition techniques achieve 100% discrimination accuracy in this experiment.

Table 1: Computational Overhead Comparison

| Method	Clock Cycles	Memory
PCA (linear)	12,142	344 bytes
VERI	69,390	4 kbytes
ANN	51,740	1.2 kbytes

CONCLUSIONS

We have demonstrated the use of linear signal processing techniques to perform chemical discrimination at the same accuracy as comparable non-linear techniques using a multi-stage hierarchy of processing. Multi-stage processing facilitates the removal of superfluous information to the chemical discrimination problem (e.g. concentration) and also allows the computational effort to be focused on areas of high activity in the chemical signature space. Speed improvements up to 82.5% and memory reductions up to 91% have been computed for the principal components analysis-based techniques presented on SAW-device based arrays in this paper. Comparable improvements can be achieved using other arrays for other analytes. The focus of increased speed, unlike many other sensing applications, is not in improving system response time, but rather in reducing power consumed by the overall chemical sensing system. These algorithms, as presented, can be implemented in a simple microcontroller, such as the Motorola MC68HC12, evaluated here to generate a self-contained system that is truly portable for field applications.

ACKNOWLEDGEMENTS

References

1 C. Di Natale, A. Macagnano, F. Davide, A.D'Amico, A. Legin, Y. Vlasov, A. Rudnitskaya, and B. Selezenev, "Multicomponent analysis on polluted waters by means of an electronic tongue," Sensors and Actuators B, vol. 44 (1997) pp. 423-428.

2 J.W. Gardner, "Detection of Vapours and Odours from a Multisensor Array Using Pattern Recognition: Part 1: Principal Component and Cluster Analysis," Sensors and Actuators B, vol. 4 (1991) pp. 109-115.

3 J.W. Gardner, T.C. Pearce, and S. Friel, "A Multisensor System for Beer Flavour Monitoring using an Array of Conducting Polymers and Predictive Classifiers," Sensors and Actuators B, vol. 18-19, (1994) pp. 240-243.

4 M. Hoummady, D. Hauden, B. Hivert, J. Henrious, P. Mielle, and P. Etievant, "SnO2 Sensor Array for Flavour Recognition with Neural Networks," Eurosensors VIII: Toulouse, France, 1994.

5 P.S. Barker, J.R. Chen, N.E. Agbor, A.P. Monkman, P. Mars, and M.C. Petty, "Vapour Recognition using Organic films and Artificial Neural Networks," Sensors and Actuators B, vol. 17 (1994) pp. 143-147.

6 G. Huyberechts, P. Szecowka, J. Roggen, and B.W. Licznerski, "Simultaneous quantification of carbon monoxide and methane in humid air using a sensor array and an artificial neural network," Sensors and Actuators B, vol. 45 (1997) pp. 123-130.

7 Eduard Llobet, Jesus Brezmes, Xavier Vilanova, Jesus E. Sueiras, and Xavier Correig, "Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array," Sensors and Actuators B, vol. 41 (1997) pp. 13-21.

8 G.C. Osbourn and R.F. Martinez, "Empirically Defined Regions of Influence for Clustering Analyses," Pattern Recognition vol. 28 no. 11, (Nov. 1995) pp. 1793-1806.

9 John S. Wagner, Michael W. Trahan, Willie E. Nelson, Philip J. Hargis, Jr., and Gary Tisone, "Chemical Recognition Software," SPIE Conf Optical Sensors for Environmental and Chemical Process Monitoring: McLean, Virginia (1995) pp. 228-238.

10 Wang Ping and Xie Jun, "Novel Recognition Method for Electronic Nose using Artificial Neural Network and Fuzzy Recognition," Sensors and Actuators B, vol. B37, no. 3 (Dec. 1996) pp. 169-174.

11 Dimitrios Vlachos and John Avaritsiotis, "Fuzzy Neural Networks for Gas Sensing," Eurosensors IX: Stockholm, Sweden (June 1995) pp. 703-706.

12 Antonio J. Ricco, Richard M. Crooks, and Gordon C. Osbourn, "Surface Acoustic Wave Chemical Sensor Arrays: New Chemically Sensitive Interfaces Combined with Novel Cluster Analysis to Detect Volatile Organic Compounds and Mixtures," Acc. Chem. Res., vol. 31, (1998) pp. 289-296.
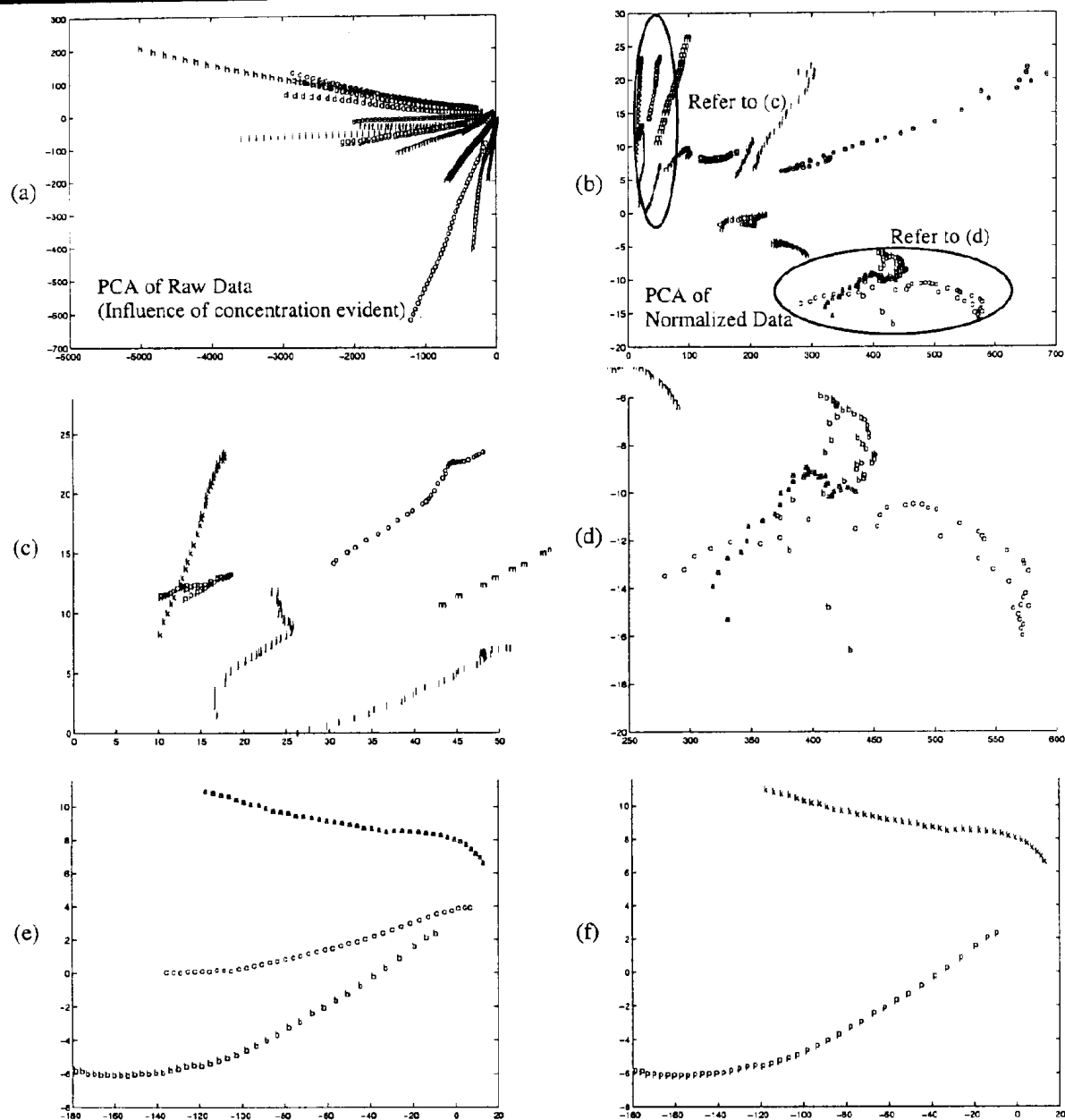
Code to Chemicals:

a: cyclo-hexane	e: benzene	i: trichloroethylene (TCE)	m: acetone
b: hexane	f: toluene	j: methanol	n: methyl isobutyl ketone (MIBK)
c: isooctane	g: chlorobenzene	k: propanol	o: diisopropylmethylphosphonate (DIMP)
d: kerosene	h: carbon-tetrachloride	l: pinacolyl alcohol	p: dimethylmethylphosphonate (DMMP)

Figure 1. Single and Multiple-stage Principal Component Analysis of 16 Chemicals in 7-element SAW Array
In (a) single stage principal components analysis, the analytes tend to group according to the magnitudes of their response curves, directly relating to concentration as well as discrimination. After (b), (c), (d) normalization, data clusters primarily according to fundamental differences in the response characteristics associated with the last six of the seven coatings. During the third stage of processing (e) and (f), members of closely related classes are separated and recognized. After clustering based on distance to the nearest line, all points cluster to their correctly identified analytes.