Application of Phase Only Filter on Electronic Nose data

Karl S. Gudmundsson*, Abdul Ahad S. Awwal**, Ricardo Gutierrez-Osuna#

* Department of Electrical Engineering, Wright State University, Dayton, OH 45435
** Lawrence Livermore National Laboratory, Livermore, CA 94551
# Department of Computer Science, Texas A&M University
E-mail : kgudmund@cs.wright.edu

ABSTRACT:

This article presents an application of Phase Only Filter (POF) to the classification of volatile compound samples with chemical sensor arrays. Sinusoidal temperature modulator excites the chemical sensor array. A system composed of multiple sensors for data acquisition requires the analysis of multiple data signals in order to classify the input data. One such system is the Electronic Nose system (eNose). The eNose data is in fact 1D data from multiple data sources. In this work five samples of three different kinds of coffee are used to build an odor database. An unknown test sample is then classified. In this research data, from such system will be analyzed and classified using the POF and compared to the more conventional K Nearest Neighbors (KNN) classifier. Both classifiers use only the mean and standard deviation as the classification feature space. The difference of the correlation between odor database and the auto correlation of the test sample is the measure of closeness for the POF. For the KNN, the measure of closeness is the difference between the odor database and the test sample.

Keywords: phase only filter, multi-sensor data classification, eNose, pattern recognition

1. INTRODUCTION

A system composed of multiple sensors for data acquisition requires the analysis of multiple data signals in order to classify the input data. One such system is the electronic nose system (eNose). An electronic nose (e-nose) is an instrument that combines gas sensor arrays and pattern analysis techniques for the detection, identification or quantification of volatile compounds[1]. The multivariate response of an array of chemical gas sensors with broad and partially overlapping selectivity can be utilized as an "electronic fingerprint" to characterize a wide range of odors or volatile compound by pattern-recognition means. In this research data from such system will be analyzed and classified using the Phase Only Filter (POF) and the more conventional K Nearest Neighbors (KNN) classifier.

eNose data is in fact 1D data from multiple data sources, thus multi-dimensional data. However, in order to analyze and classify such data using conventional methods such as the K Nearest Neighbors (KNN) is computationally demanding. The test sample is compared to the trained database one by one requiring much computational time.

However, application wise the eNose is just one such possible application area of ultra fast correlation technique. Mass spectrometers, Fourier Transform InfraRed (FTIR) spectrometers, and environmental detectors that can be used in urban environments for early detection of biological aerosols are just few of the instruments that could benefit from ultra fast classification technique.

In this day and age, urban biological defense is reality. Environmental monitoring, medical monitoring and public health surveillance are few of the key components to be considered in such situation. Similarly, applications in bomb identification for airport security might use multiple sensors sensitive to different chemicals, or other quantifiable signal generating material. What all these different application areas and methods have in common is the need for an ultra fast classification technique. Optical POF classifier might be able to fulfill this need. Optically implemented POF will lend itself equally well on large and small feature space.

Proceedings of SPIE Vol. 5201 Photonic Devices and Algorithms for Computing V, Edited by Khan M. Iftekharuddin, Abdul Ahad S. Awwal (SPIE, Bellingham, WA, 2003) · 0277-786X/03/$15.00

2. EXPERIMENT

In the following example, the eNose data is collected using array of four Figaro gas sensors. Each sensor is sensitive to different gases. TGS 2600, and TGS 2602 are sensitive to gaseous air contaminants. TGS 2610 is sensitive to methane, propane, and butane, and the TGS 2620 sensor is sensitive to alcoholic and organic solvent vapor. When using an array of four sensors a wide variety of odors can be analyzed and classified. In this experiment we use the odors of three different coffee types, French Vanilla, Hazelnut, and Jamaican coffee as test specimen. All three can easily be classified using relatively small database.

The system we built was composed of the following main parts: four Figaro sensors, sensor chamber, gas delivery system, control circuit, and a Personal Computer (PC) with general purpose interface board and a Labview package for data collection. The following figure 1 shows one such experiment.

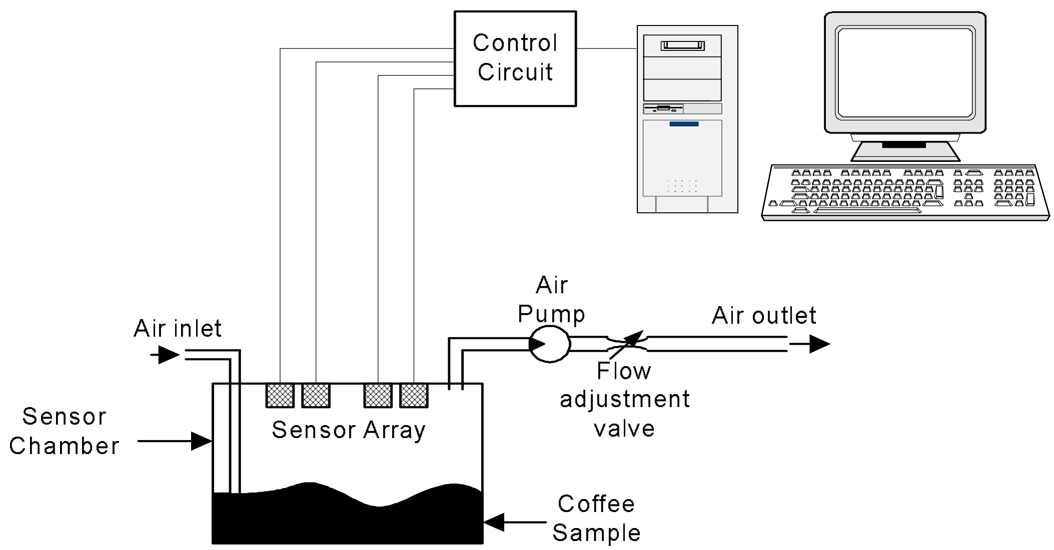

Figure 1. Experimental setup

In this experiment a database of 5 samples from each of the 3 different coffee types was collected. A sinusoidal temperature modulation was used to excite the sensors. For each test sample data was collected for five minutes. Only the last complete sample cycle of each collection was used to build the database. A one complete cycle is twenty data points. The system was purged between each sample collection by running for five minutes without any specimen present. The database was trained by using the five samples from each of the three different coffee types. Finally, three test samples one of each kind of coffee was collected the following day, using the same method. It should be noted that the ambient odor between days changes depending on who is working in the laboratory and what colon he or she uses. This fact is reflected in the experiment.

Preprocessing of the data consisted of removing the DC offset by normalizing the data from each sensor to the magnitude of one. Then, the mean and standard deviation of each signal was computed and used for feature space. However, other features could be used as well as the raw signal. In the case of the POF classifier, the number of features plays a less significant role than in the case of any software classifier. In deed more features result in higher resolution, thus better correlation. For simplicity, and for ease of performance comparison of the KNN classifier and the POF classifier, in this experiment we choose to use only the mean and standard deviation of each sensor. The normalized signal of each-class, each sample is shown in figure 2. As well as one of the test samples used in this experiment.

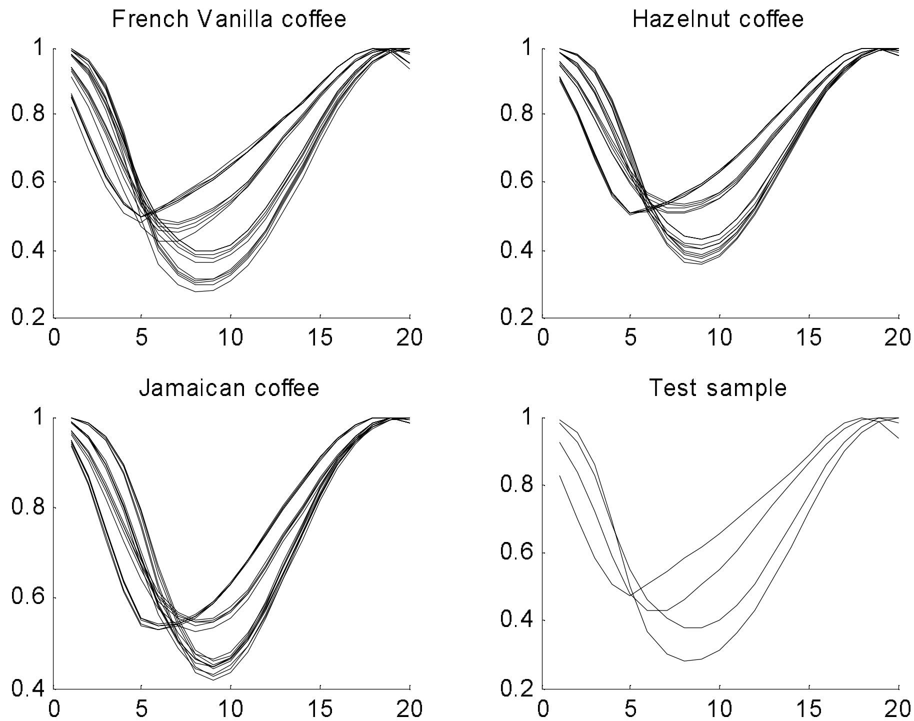

Figure 2. Three classes of patterns and one test pattern

4. PROPOSED POF CLASSIFIER

Oppenheim and Lim laid the groundwork for phase only correlation of images when they showed that much of the information of an image could be retrieved from the phase component of one image and the average amplitude of another image[2]. Same should hold true for one dimensional signals.

The mathematical foundation of the POF filter can easily be derived as follows[3]. If we let the Fourier transform of the signal function $f(x)$ be denoted by:

$$F(U_x) = \left|F(U_x)\right| \exp(j\Phi U_x) \qquad (1)$$

Then a Complex Match Filter (CMF) corresponding to this function $f(x)$ at the filter plane is expected to produce the autocorrelation of the input at the output plane. Therefore, the CMF is given by the complex conjugate of the input Fourier spectrum:

$$H_{CMF}(U_x) = F^*(U_x) = \left|F(U_x)\right| \exp(-j\Phi U_x) \qquad (2)$$

The inverse Fourier transform of the product of $F(U_x)$ and $H_{CMF}(U_x)$ results in the convolution of $f(x)$ and $f(-x)$, which is the equivalent of the autocorrelation of $f(x)$. Moreover, when $|F(U_x)|$ is set to unity, H_{CMF} becomes a phase only filter:

$$H_{POF}(U_x) = \exp(-j\Phi U_x) \qquad (3)$$

Finally, the phase only correlation of input signal and the reference sample is simply:

$$F^{-1}\{F(U_x)\, H_{POF}(U_x)\} \tag{4}$$

Since the convolution operator in space domain is equivalent to the product operator in frequency domain, one can think of the POF as an edge enhancer by way of division by $|\, F(U_x) \,|$ and integrator ,by integrating the product of the input signal and the reference sample.

For each class a different POF needs to be developed from the database samples. A 4f correlator set up may allow for parallel matching of all the input into different classes in parallel. One possible implementation is shown in the following diagram.

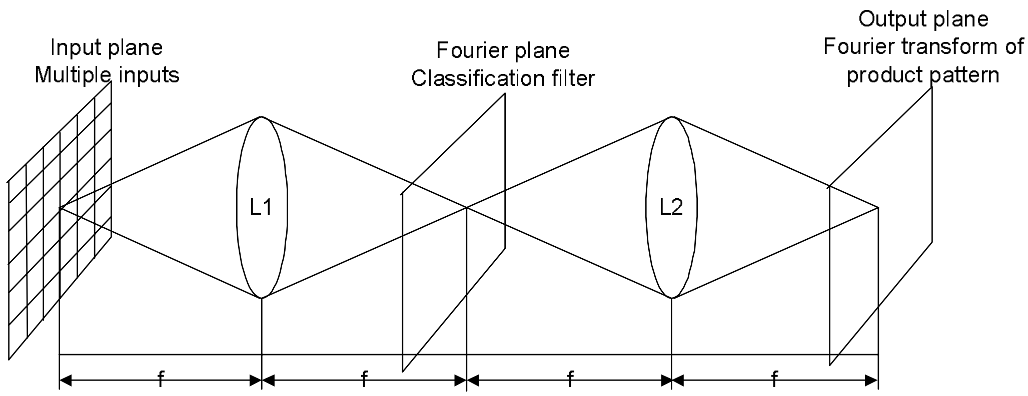

Figure 3. 4f correlator

A simple lens L1 performs Fourier transform. The unknown input sample is placed in the input plane, one focal length from the lens, and its Fourier transform forms at the focal distance from the lens in the Fourier plane. The classification filters will are placed at this plane. After another lens the Fourier transform of the product pattern will appear at the output plane. This output will be generated as a result of the parallel matching producing the match with different classes at spatially varying location. The location of highest intensity in the output plane represents the best match between the test sample and the trained database.

5. POF SIMULATION

The real system was simulated on a PC computer using the Matlab software. The database consists of fifteen samples. Four sensors represent each sample, and two features represent each sensor. Thus the database consists of 120 items. A cross-correlation between the sample and the database was performed to generate the output at the output plane. Each sample can be represented by the following set:

$$sample = \{Sensor1(mean, std), Sensor2(mean, std)Sensor3(mean, std)Sensor4(mean, std)\} \tag{5}$$

and the database by the set of samples:

$$database = \{sample1, \ldots, sample15\} \tag{6}$$

Therefore, the correlation database simply becomes:

$$corrDB = \{test\_sample \otimes database\} \tag{7}$$

The absolute difference between the test sample and each element in the database provides circular boundary or distance measure between the sample and the database. This fact was used for comparisons purposes of the POF and KNN classifier. In the case of the KNN the difference is the direct difference, or distance between the test samples mean and std and the mean and std of each element in the database. In the case of the POF the auto-correlation of the test samples mean and std was computed and subtracted from the database of cross-correlations between the samples mean and std and the mean and std of each sample in the database. The scatter plots in figures 4 and 5, show the sample location in the feature space with respect to the test sample represented by the letter "X".

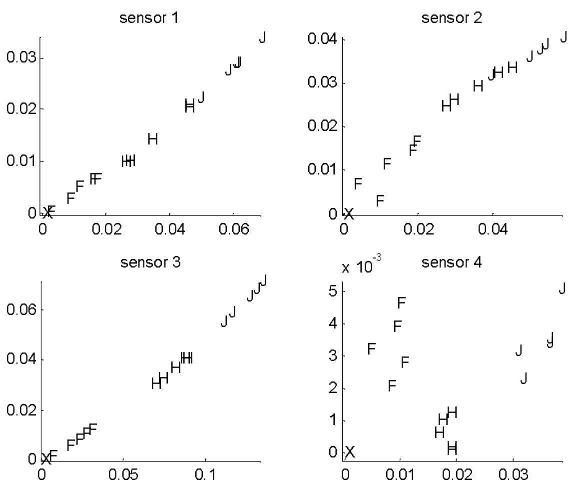

Figure 4. KNN scatter plot, French Vanilla test sample

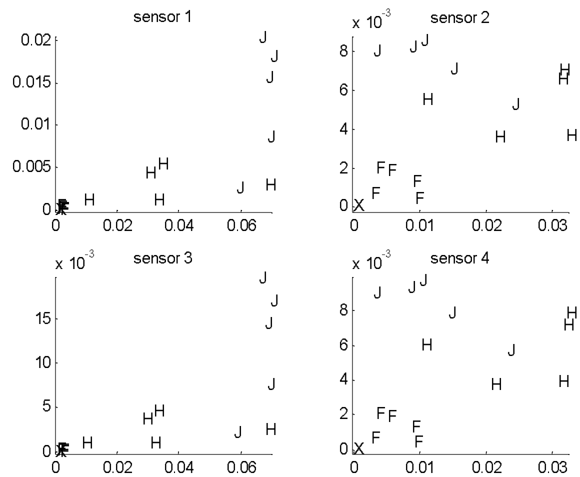

Figure 5. POF scatter plot, French Vanilla test sample

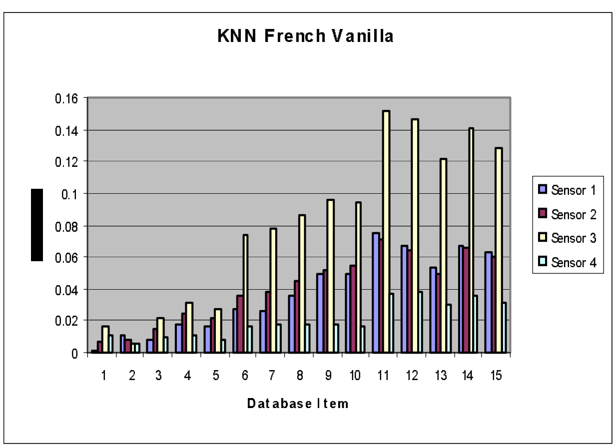

Figure 6. KNN Classifier, French Vanilla test sample

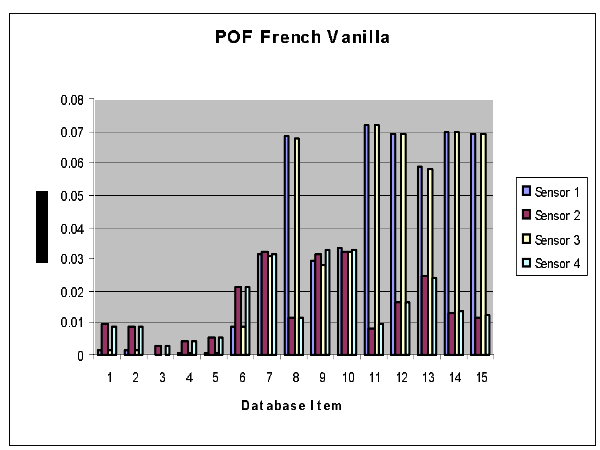

Figure 7. POF Classifier, French Vanilla test sample

The bar diagrams in figures 6 and 7 show the distance from the test sample to each database sample for each sensor. The distance is simply the computed Euclidean distance, sqrt(mean^2+std^2). In the case of the French Vanilla coffee the POF has much better discrimination. However, the results for the Hazelnut Coffee are not as favorable.

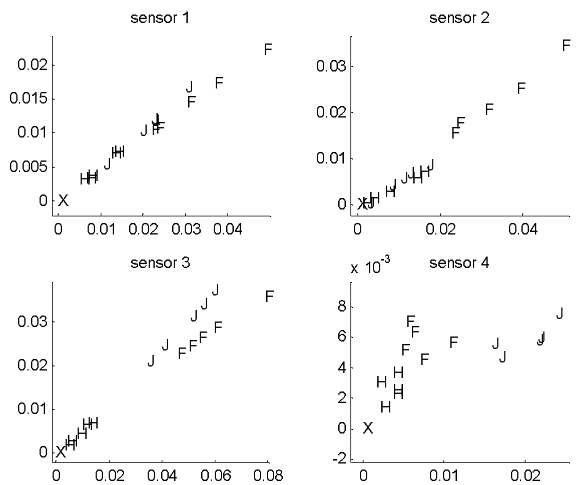

Figure 8. KNN scatter plot, Hazelnut test sample

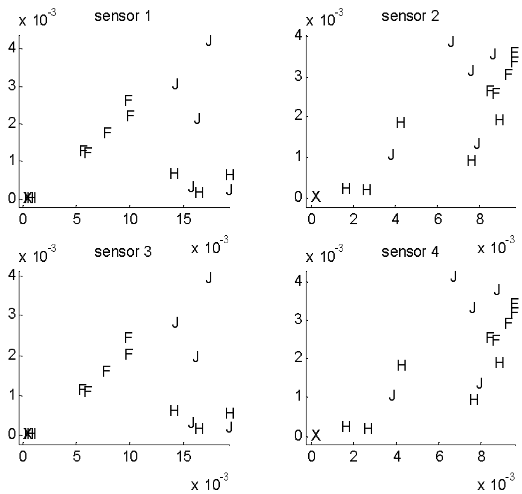

Figure 9. POF scatter plot, Hazelnut test sample

Similarly the scatter plots in figures 8 and 9, show the sample location in the feature space with respect to the test sample represented by the letter “X” for the case of Hazelnut test sample.

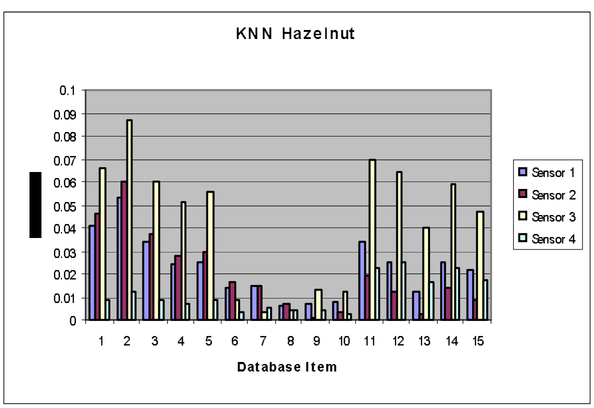

Figure 10. KNN Classifier, Hazelnut test sample

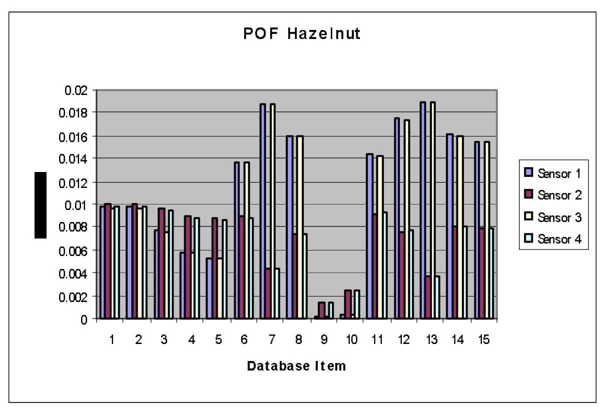

Figure 11. POF Classifier, Hazelnut test sample

Here the bar diagrams in figures 10 and 11 show the distance from the test sample to each database sample for each sensor. Unexpectedly, the results for the Hazelnut Coffee are clearly not as favorable in the case of the POF classifier as for the KNN classifier. Samples 6, 7, and 8 correlate very poorly for sensors 1 and 3. This might be due to the POF’s great sensitivity to relatively small changes, and the fact we are only comparing two features.

Again, the scatter plots in figures 12 and 13 shows the sample location in the feature space with respect to the test sample represented by the letter “X” for the case of the Jamaican test sample.

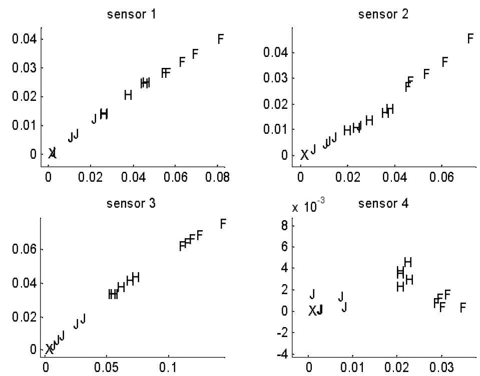

Figure 12. KNN scatter plot, Jamaican test sample

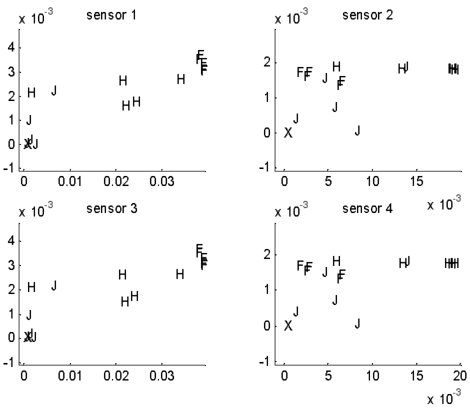

Figure 13. POF scatter plot, Jamaican test sample

In figures 14 and 15 the bar diagram shows the distance from the test sample to each database sample for each sensor. The samples 12, and 13 correlate poorly for sensors 2 and 4 in this case.

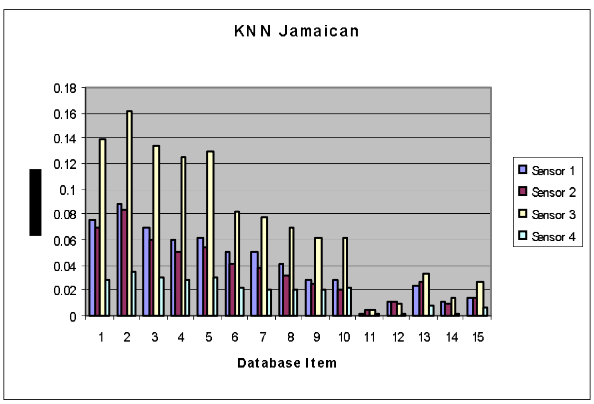

Figure 14.KNN classifier, Jamaican test sample

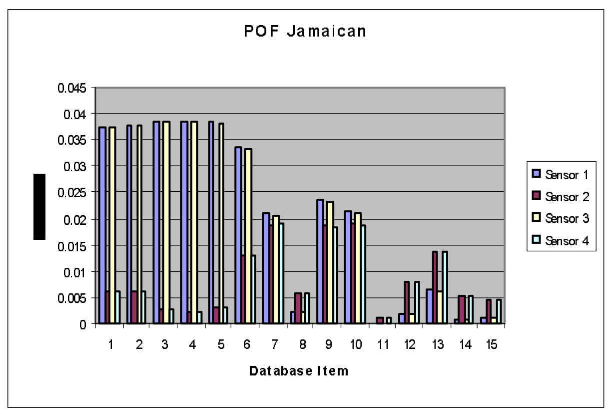

Figure 15.POF classifier, Jamaican test sample

The following table was constructed to further investigate the performance of the POF and KNN classifier. The first column describes the coffee type and method used for the classification. The second column is the mean of the correlation database in the case of the POF classifier, and simply the mean of the database in the case of the KNN classifier. In the third column we have the mean of each coffee type after subtracting the test sample. The fourth column is the ratio of the third column over the second column. Finally, the discrimination ratio is found by dividing the results in column four by the ratio of the match in that category. To illustrate this lets consider the case of the KNN, and the French Vanilla test sample. The match is in the first row were the mean of the difference of the French Vanilla samples and the test sample is 0.013798371. The ratio over the database is 0.013798371/0.045053963 = 0.306263205. The Discrimination ratio with respect to the Hazelnut type is 1.0299604/0.306263205 = 3.363 an so on.

Method and Type	Mean Database	Mean Type	Ratio Type/Database	Discrimination ratio
KNN French Vanilla	0.045053963	0.013798371	0.306263205	
Hazelnut		0.046403798	1.0299604	3.362990999
Jamaican		0.074959721	1.663776395	5.432505018
POF French Vanilla	0.024863286	0.003426112	0.137798031	
Hazelnut		0.029867322	1.201262057	8.717556041
Jamaican		0.041296423	1.660939912	12.0534372
KNN Hazelnut	0.024231869	0.007687871	0.317262824	
French Vanilla		0.03704558	1.528795807	4.818704526
Jamaican		0.027962156	1.153941369	3.637178016
POF Hazelnut	0.009238991	0.007334546	0.793868795	
French Vanilla		0.008502842	0.920321517	1.159286677
Jamaican		0.011879584	1.285809689	1.619675313
KNN Jamaican	0.042638052	0.011829375	0.277437051	
French Vanilla		0.075615707	1.773432511	6.392197824
Hazelnut		0.040469073	0.949130438	3.421065917
POF Jamaican	0.014329485	0.004287313	0.299195183	
French Vanilla		0.021072239	1.470551086	4.915022596
Hazelnut		0.017628902	1.230253731	4.111876796

Table 1. Classification results

6. CONCLUSION

The phase only filter has been successfully used to classify one dimensional multi sensor data. In some cases the POF classifier outperforms the KNN classifier. In this initial research only two features were used for classification to simplify the task of comparing the two methods. Both the KNN and the POF could be used on a raw signal comparing in this case all the twenty points of the signal. The POF would probably benefit from larger feature space, since each point in the correlation would weight less. However, this would burden the KNN since the computational complexity increases greatly.

Future work in this area would include the investigation of performance in larger feature space, and the design of an electro/optical correlator to perform classification on-line. Also, the performance of CMF compared to the POF needs to be investigated.

7. ACKNOWLEDGEMENT

K Gudmundsson was supported by a grant from Wright State University College of Engineering, while the work was done. AASA was with the Lawrence Livermore Laboratories, and R. Gutierrez-Osuna was with the Department of Computer Science, Texas A&M University.

8. REFERENCES:

[1] H. T. Nagle, S. S. Schiffman and R. Gutierrez-Osuna , "The how and why of electronic noses" IEEE Spectrum, September 1998, pp. 22-34.

[2] A. V. Oppenheim and J. S. Lim, ”The importance of phase in signals” Proc. IEEE 69, 529-541 (1984)

[3] M. A. Karim and A. A.S. Awwal, “Optical Computing: An Introduction” New York,John Wiley & sons, Inc, 1992.