An Evaluation of Electronic Nose for Space Program Applications

Rebecca C. Young, Bruce R. Linnell, and William J. Buttner
Applied Chemistry Laboratory, NASA Kennedy Space Center, Florida

Abstract
The ability to monitor air contaminants in the Shuttle and the International Space Station is important to ensure the health and safety of astronauts. Three specific space applications have been identified that would benefit from a chemical monitor: organic contaminants in crew cabins, propellant contaminants in the airlock, and pre-combustion fire detection. NASA has assessed several commercial and developing electronic noses (e-noses) for these applications. A preliminary series of tests identified those e-noses that exhibited sufficient sensitivity to the vapors of interest. These e-noses were further tested to assess their ability to identify vapors, and in-house software has been developed to enhance identification. This paper describes the tests, the classification ability of selected e-noses, and the software improvements made to meet the requirements for these space program applications.

Keywords
Electronic nose, air quality monitor, fire detection, hypergolic fuel, pattern recognition, classification, error estimation, feature selection, sample ratio.

Introduction
An electronic nose (e-nose) consists of an array of non-specific vapor sensors [1]. In general, a sensor array is designed such that each individual sensor responds to a broad range of chemicals, but with a unique sensitivity relative to the other sensors. Chemical identification is achieved by comparing the sensor response pattern of an unknown vapor to previously established patterns of known vapors using pattern recognition software. There are many different e-nose instruments which use various types of sensors, including metal oxide semiconductor (MOS), conducting polymer, surface acoustic wave (SAW), composite polymer (CP), and electrochemical. Some sensors are more sensitive to specific vapors, while others are more prone to drift due to changes in ambient conditions (temperature, relative humidity, air pressure). NASA at the Kennedy Space Center (KSC) has assessed the ability of several developing and commercially available e-noses to meet specific requirements for applications in the space program. Three applications have been identified:

- Monitoring air contaminants in a closed environment, such as the Shuttle and the International Space Station (ISS). Post-mission analyses of air samples from the Shuttle have confirmed the occasional presence of volatile organic contaminants [2]. The spacecraft maximum allowable concentration (SMAC) was established as a guideline to maintain the air quality in spacecraft [3,4], and air monitoring that provides both specific compound identification and concentration is required to assure SMAC compliance. Continuous air monitoring could also provide notification of adverse events such as spills or leaks.
- Monitoring hypergolic propellant contaminants in airlocks. During space walks, an astronaut's suit or other equipment may be exposed to hypergolic propellants [5] – hydrazine (Hz), monomethyl hydrazine (MMH), and nitrogen tetroxide (N_2O_4). Hydrazines are toxic and suspected human carcinogens, so it is important to verify that no residual vapor is present prior to reentry to the crew quarters through the airlock.
- Notification of an impending fire, which can be disastrous in a closed environment. Because overheated wires are a common precursor to electrical fires, an e-nose could detect the chemical vapors emitted from the heated insulation, and provide a warning even before combustion begins,

since vapor emissions occur prior to smoke generation or actual combustion. With sufficient sensitivity, it could also distinguish between electrical and other types of fires by the unique composition of the vapors.

The vapor levels required by these applications are low, and the deployment conditions can be quite challenging. This paper addresses the ability of the instruments to identify vapors at or near the conditions required for specific applications within the space program.

Experimental Procedure

A literature and market search for available e-noses was performed to identify instruments suitable for space program applications. A number of miniature commercial instruments were available in a moderate price range. In addition, several instruments that were not yet commercially available but were in an advanced stage of development were considered. The various e-noses used in this study are shown in Table 1. Not all instruments have been tested for each of the applications identified.

Table 1: E-nose Instruments

Instrument	Manufacturer	Array
KAMINA	Karlsruhe Research Center	38 MOS
i-Pen2, i-Pen3	Airsense	10 MOS
Sam Detect	Daimler-Benz Aerospac	5 MOS
Cyranose	Cyrano Sciences	32 CP
JPL	N/A	32 CP
Early Fire Detection System	Marconi Applied Technologies	4 SAW
VaporLab	SawTek	4 SAW

Calibrated Vapor Generation

Test vapors were generated using calibrated permeation devices (PD) and ovens (Kintek Model 360). Table 2 summarizes the vapors generated for this research. Vapors from the vapor generators were either used directly or blended with air from a temperature, humidity and flow rate controller (Miller-Nelson Model HCS-40), providing dilution factors up to 25. See Figure 1.
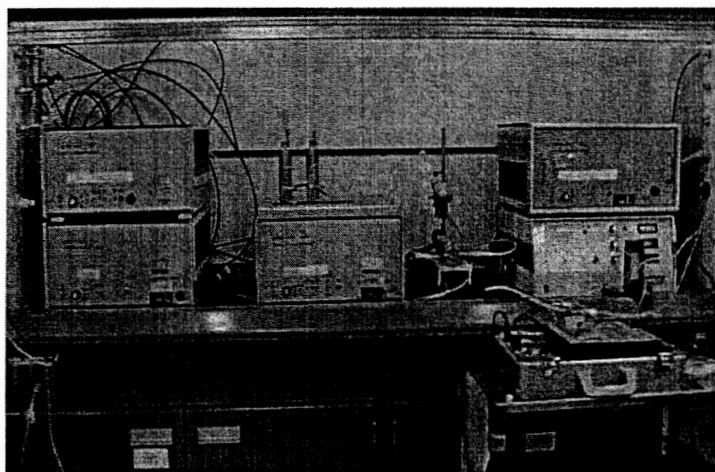

Figure 1 – Vapor Generation System

Organic Vapor Tests

The first series of tests were a quick screening to determine which instruments were more suitable for detecting the organic vapors at the SMAC levels in Table 2. From these tests, the Sam Detect, KAMINA, and i-Pen2 were selected for further testing. These instruments were then subjected to more extensive testing, including exposures at various concentrations and relative humidities. These datasets were used to evaluate the ability of the e-noses to identify various gasses, as discussed in the "Vapor Identification" section below.

Hypergolic Fuel Tests

The current Threshold Limit Value for exposures to hydrazines set by the American Conference of Governmental Industrial Hygienists is 10 ppb, which KSC has adopted as the standard for its operations. Moreover, vapor monitoring must be done at the operating pressure of the airlock, which ranges from about 150 to 750 torr. While the KAMINA, Cyranose, and i-Pen2 all showed reasonable sensitivity to ppm levels of hydrazine or MMH, only the KAMINA was able to respond to 10 ppb levels with a signal to noise ratio greater than three.

Table 2 : Test Vapors

Vapor	Abrv.	Concentration (ppm)	7-day SMAC (ppm)
Hydrazine	Hz	0.07, 0.20, 1.25, 17.7, 35.0	0.04
Monomethyl Hydrazine	MMH	0.15, 1.06, 5.32, 16.0	0.002
Ammonia	NH3	9	10
Nitrogen Dioxide	NO2	12, 46	Not defined
Toluene	Tol	59	16
Acetone	Ace	115	22
Trichloroethylene	TCE	90	9
Methylethyl Ketone	MEK	48	10
Xylene	Xyl	40	50
Isopropyl alcohol	IPA	55	60

Pre-combustion Fire Detection Tests

For the Shuttle and other space program applications, wire insulation is typically Teflon®, Kapton® or silicone-based materials. To mimic a pending electrical fire, a section of electrical wire was mounted in an enclosed wire-heating apparatus, and voltage was gradually increased to heat the wire. Wires with PVC, Teflon, and Kapton were tested. The KAMINA, Sam Detect, and the Cyranose were used in this test, and all three responded to the PVC vapors. However, only the KAMINA and Sam Detect were able to detect vapors from Teflon or Kapton wires, which were at much lower concentrations. Only limited testing has been performed to date on this application, and it will not be discussed further in this paper.

For more details on the experimental setups and testing procedures, see [6].

Vapor Identification

While the KAMINA, i-Pen, and Sam Detect e-noses provide PC-based graphical user interface programs to control the collection of vapors and to display the results of classification, these programs are very limited in terms of the kinds of features and classifiers that can be used, and do not provide an estimate of the future classification success rate of the model. Thus programs were written for The Mathworks' Matlab® (a data analysis package) to calculate the classification success rate, and to explore ways of improving the classification by using different features and different classifiers. Using Matlab also allowed us to experiment with many different kinds of pre- and post-processing, such as noise filtering, various normalizations, outlier removal, etc.

Feature Extraction

In pattern classification, a "feature" is any direct or derived measurement of the system that helps differentiate between classes. Many different possible features can be extracted from e-nose data, but it has been shown that either the final value or maximum initial slope (see Figure 2) are among the best single features for discriminating between classes [12]. However, the value of the maximum initial slope is very sensitive to noise, which can be significant at low vapor concentrations. While the final value is robust and simple to calculate, most sensors require a long time to stabilize (one to 20 minutes depending on the application and the sensor response time). Instead, we used the sensor values at specific times, which were considerably earlier than the time needed for the sensor to reach steady-state. While sensor values turn out to be a very good feature, they are usually the only feature option offered by commercial e-nose software. Using Matlab allowed us to explore other options such as maximum initial slope, areas under the curve, signal modeling constants, etc.

Sample Ratio

The term "sample ratio" (SR) is defined here as the number of samples per class divided by the number of features. A commonly held rule of thumb in pattern recognition is that the SR should be at least five to ten [8]. This can be difficult to achieve, due to long sample times (up to 30 to 60 minutes per example) and multiple sensors. For example, the KAMINA e-nose has 38 sensors, and so even if only one feature is extracted per sensor (such as the final value), there are a total of 38 features per sample. This means that almost 200 samples per class should be gathered for an SR of 5, which could take more than 100 hours. The problem is compounded if two or more features are extracted per sensor. Therefore, small SR's are quite common in e-nose research. SR's near to or less than one make all aspects of the pattern classification process much more difficult [8,9,10].

Classification

There are many methods available to classify an unknown sample, given a model. Statistical methods include the nearest Euclidean mean classifier, the Linear classifier, the nearest Mahalanobis distance classifier, and the Quadratic classifier [11,12], all of which have the advantage of being very fast. However, the Euclidean classifier ignores the shapes of the distributions of the classes, the Linear classifier assumes the class shapes are identical, and the Mahalanobis and Quadratic classifiers require relatively large amounts of data (SR ≥ 3) in order to model the classes well. The Kth nearest neighbor classifier is also very popular [11,12], but it tends to overfit the data for K=1, and does not generalize well at small SR's [12].
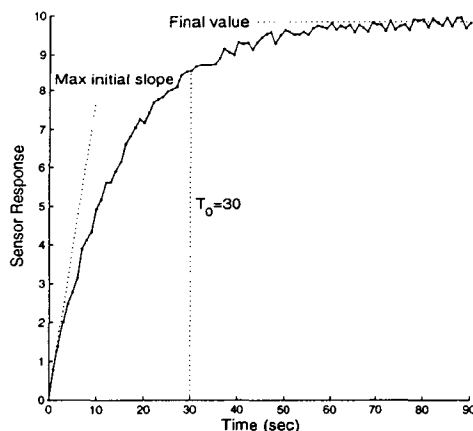

Figure 2 – Typical Time Response of a Single Sensor

Iterative methods, including the relaxation, Widrow-Huff, and Ho-Kashyap linear classifiers [11], as well as neural networks and support vector machines, are generally very slow and have many user-selectable parameters which can significantly affect the classifier's performance. We used the Linear classifier for our

application because of its simplicity, and it has been shown to perform better than the Quadratic classifier at small SR's [13]. This is often the only classifier offered by commercial e-nose software, but if enough data is available, we now have the ability to see if other classifiers can provide better classification.

Error Estimation

In order to determine how well the classes can be differentiated, some estimate of the classifier's future performance is required [14]. None of the vendor-supplied software could automatically calculate this value. It should be noted that the separation between classes in a Linear Discriminant Analysis (LDA) plot has no correlation to the classification success rate of the original data [18].

While the ideal method to calculate the classification success rate would use part of the data for building the model and the other part for testing the model (called the Holdout estimator), this is not possible when the SR is very small. Holdout requires a large number of examples to build an accurate model, and even more examples to see how well the model performs. Thus Holdout is applicable only in high-SR situations.

On the other hand, if all the data is used to build the model, and is also used to estimate the success rate (known as Resubstitution), the estimate will be too optimistic. This problem is usually solved by using techniques such as "N-Fold Cross Validation", which sets aside part of the data, builds a model with the remaining data, and uses the first part to estimate the performance. N different portions are set aside, and the N estimates are then averaged. If N equals the number of samples, the result is the "Leave-One-Out" estimator, which is often called simply "cross validation". As the SR gets smaller, Leave-One-Out becomes increasingly pessimistic (that is, underestimates the classification success rate), while Resubstitution becomes increasingly optimistic [9]. However, the average of these two estimators has been shown to be an excellent estimate of future classification performance at small SR's [10]. This average will be called the "RU" estimator. Another method, known as "Bootstrap" [15], has also been shown to be an excellent estimator at small SR's [10]. Bootstrap creates a model of the same size as the original dataset by randomly selecting examples from the dataset (allowing each example to be used more than once), and then uses the entire dataset to test the model. This process, called a bootstrap trial, is repeated multiple times to create an estimate.

Results

The conditions of each test are summarized in Tables 3 and 4, and the classification results are shown in Table 5. In all cases, the features were the sensor values at the given time, severe outliers were removed first, and the Linear classifier was used. Each estimate is the average of the RU and Bootstrap estimators, while the margin of error term is half the difference between them. For example, if RU returned a value of 90% and Bootstrap returned 92%, the table entry would read 91% ± 1%. Twenty bootstrap trials were used for the Bootstrap estimator.

It should be noted that for the KAMINA data, a different method of calculating the baseline sensor resistance (R_0) was implemented than that used by their software. In addition, we used all 38 sensor values for classification, whereas the KAMINA software used a much smaller number of LDA components, which can only loose information. In a comparison test using a Linear classifier on identical datasets with the Holdout estimator, these changes resulted in an improvement of 13 percentage points in the classification as compared to the KAMINA software (the Holdout estimate for the KAMINA was calculated manually).

Table 3   Test Parameters

Test Name	E-nose	# Sensors	Gasses Tested	Avg SR
i-Pen2/organic	i-Pen2	10	Ace IPA MEK TCE Tol Xyl NO2 NH3	0.90
SAM/organic	Sam Detect	5	Ace IPA MEK TCE Tol Xyl NO2	3.4
Chip 20/organic	KAMINA	38	Ace IPA MEK Tol Xyl	0.37
Chip 20/fuel	KAMINA	38	Hz MMH	0.93
Chip 20/both	KAMINA	38	Ace IPA MEK Tol Xyl Hz MMH	0.52
Chip 9/organic	KAMINA	38	Ace IPA MEK Xyl	3.5
Chip 15/organic	KAMINA	38	Ace IPA MEK Xyl	3.5
i-Pen3/organic	i-Pen3	10	Ace IPA MEK TCE Tol Xyl NO2 NH3	12.3
Chip 15/fuel	KAMINA	38	Hz MMH	4.0

Table 4   Environmental Parameters and Vapor Concentrations

Test name	Temp	RH	Air Pressure	Concentrations
i-Pen2/organic	21° C	8%, 33%	Ambient (≈760 torr)	2-23 ppm
SAM/organic	21° C	8%	Ambient	11-116 ppm
Chip 20/organic	21° C	8%, 33%	Ambient	11-116 ppm
Chip 20/fuel	21° C	8%, 33%, 60%	150, 200, 450, 710 torr	10-220 ppb
Chip 9/organic	21° C	8%, 33%, 48%	Ambient	11-44 ppm
Chip 15/organic	21° C	8%, 33%, 48%	Ambient	11-44 ppm
i-Pen3/organic	21° C	8%, 33%, 50%	Ambient	1-33 ppm
Chip 15/fuel	21° C	8%, 33%, 50%	Ambient	10-100 ppb

Table 5   Classification Success Results (RU and Boot)

Test Name	30 sec, All Features	30 sec, Best Features	90 sec, All Features	90 sec, Best Features
i-Pen2/organic	90% ± 1%	90% ± 0.9% (5/10 used)	89% ± 0.4%	88% ± 0.7% (4/10 used)
SAM/organic	95% ± 0.5%	96% ± 0.3% (4/5 used)	100% ± 0%	100% ± 0.2% (3/5 used)
Chip 20/organic	87% ± 4%	98% ± 2% (17/38 used)	87% ± 4%	97% ± 0.9% (18/38 used)
Chip 20/fuel	83% ± 0.7%	94% ± 3% (26/38 used)	84% ± 2%	97% ± 3% (20/38 used)
Chip 20/both	79% ± 2%	84% ± 1% (17/38 used)	81% ± 1%	82% ± 0.5% (19/38 used)
Chip 9/organic	94% ± 0.4%	94% ± 0.3% (23/38 used)	100% ± 0.01%	100% ± 0% (5/38 used)
Chip 15/organic	91% ± 0.03%	91% ± 0.4% (23/38 used)	97% ± 0.1%	97% ± 0.1% (9/38 used)
i-Pen3/organic	95% ± 0.02%	95% ± 0.1% (6/10 used)	99% ± 0.08%	99% ± 0.07% (4/10 used)
Chip 15/fuel	83% ± 1%	91% ± 1% (15/38 used)	89% ± 1%	98% ± 0.2% (3/38 used)

Table 6 Classification Success Results (Holdout)

Test Name	30 sec, All	30 sec, Best	90 sec, All	90 sec, Best
SAM/organic	94% ± 7%	95% ± 7%	100% ± 0%	99% ± 2%
Chip 9/organic	93% ± 5%	93% ± 4%	100% ± 0.4%	100% ± 0.4%
Chip 15/organic	88% ± 5%	90% ± 6%	97% ± 1%	98% ± 2%
i-Pen3/organic	95% ± 3%	94% ± 3%	99% ± 0.8%	98% ± 3%
Chip 15/fuel	80% ± 8%	88% ± 7%	87% ± 7%	94% ± 6%

The best features (i.e., sensors) were found by using standard feature selection techniques [10,16], and generally increases the classification success rate by removing redundant, irrelevant, or faulty sensors (such as in some KAMINA sensor chips). This also has the advantage of increasing the SR. In some cases, the classification success rate increased by up to 10-13 percentage points, while in other cases, the same success rate was achieved with far fewer sensors. Intelligent sensor selection is another function that is not available with any of the e-nose software packages we evaluated.

The close agreement between the RU and Bootstrap estimators (usually within 2 percentage points, except for Chip 20) indicates the usefulness of the RU estimator in small-sample situations. RU was up to 8 times faster than Bootstrap in these tests, which used only the minimum recommended number of bootstrap samples. In addition, Bootstrap returns slightly different estimates each time it is run, which may cause problems with feature selection [17].

The estimates for the higher-SR datasets could be verified using the Holdout method described earlier. Since the test data was not used to build the model, this provides an independent measure of the quality of the model. Table 6 shows the results using twenty random model/test partitions of the data, which agree very well with the estimates in Table 5. The margin-of-error term in Table 6 is half the difference between the maximum and minimum holdout estimates.

Finally, the long-term nature of the Chip 9 and Chip 15 data (44 days) showed that sensor drift was not a significant problem for those KAMINA sensors over that time frame.

Summary and Conclusion
The viability of e-nose technology for the target applications was confirmed. Several instruments were identified that could detect the organic vapors at the SMAC levels (KAMINA, i-Pen, and Sam Detect) and the pre-combustion vapors (KAMINA and Sam Detect). Only the KAMINA was able to detect hypergolic fuels at 10 ppb.

Data sets were collected and modeled using vendor supplied algorithms and protocols. The vendor software did not provide information on the probability of future identification, and this made it difficult to do a complete assessment of the models. To alleviate this shortfall, in-house modeling was performed on the data sets. This not only provided an independent means of assessing the information content of the raw sensor data, but also allowed greater flexibility in developing models than that provided by vendors.

It was found that selecting a subset of sensors could either significantly increase the classification success rate or reduce the number of required sensors. The RU estimator was found to be in close agreement to the Bootstrap estimator for small-sample problems, and considerably faster.

The final result is that the expected accuracy of identification can exceed 95% for organic and hypergolic vapors.

References

[1] H. Nagle, R. Gutierrez-Osuna, S. Schiffman, "The How and Why of Electronic Noses", IEEE Spectrum, September 1998, pp. 22-34.

[2] J. James, T. Linero, H. Leano, J. Boyd, P.Covington, "Volatile Organic Contaminants Found in the Habitable Environment of the Space Shuttle: STS-26 to STS-55", Aviation, Space, and Environmental Medicine, September 1994, pp. 851-857.

[3] "Spacecraft Maximum Allowable Concentrations for Airborne Contaminants", JSC-20584, NASA Johnson Space Center, May 1999.

[4] "Safety Requirements Document for International Space Station Program", SSP50021, NASA Johnson Space Center, December 12, 1995.

[5] "Adequacy of the Intravehicular Detector and Removal of Hypergolic Propellant Contamination Following Extravehicular Activities", JSC2000, JSC0024, March 31, 2001.

[6] R. Young, W. Buttner, B. Linnell, R. Ramesham, "Electronic Nose for Space Program Applications", presented at the Ninth International Conference on Chemical Sensors, Boston, MA, July 2002, and submitted for publication in the Proceedings.

[7] S. Roussel, G. Forsberg, V. Steinmetz, P. Grenier, V. Bellon-Maurel, "Optimization of Electronic Nose Measurements, Part I : Methodology of Output Feature Selection", Journal of Food Engineering, Vol. 37, pp. 207-222, 1998.

[8] A. Jain and B. Chandrasekaran, "Dimensionality and Sample Size Considerations in Pattern Recognition Practice", in P.R. Krishnaiah and L.N. Kanal (ed.), Handbook of Statistics, Vol. 2, North-Holland Publishing Company, New York, 1982, pp. 835-855.

[9] S. Raudys and A. Jain, "Small Sample Size Effects in Statistical Pattern Recognition : Recommendations for Practioners", IEEE PAMI, Vol. 13, pp. 252-264, 1991.

[10] B. Linnell, "The Effects of Small Samples on Statistical Pattern Recognition Techniques", Ph.D. dissertation, Electrical and Computer Engineering Dept., North Carolina State University, 2001.

[11] R. Duda and P. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York, 1973.

[12] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 2nd ed., 1990.

[13] S. Marks and O.J. Dunn, "Discriminant Functions When Covariance Matrices are Unequal", Journal of the American Statistical Association, Vol. 69, No. 346, pp. 555-559, June 1974.
[14] G. Toussaint, "Bibliography on Estimation of Misclassification", IEEE Transactions on Information Theory, Vol. 20, pp. 472-479, 1974.

[15] A.K. Jain, R.C. Dubes, and C.C. Chen, "Bootstrap Techniques for Error Estimation", IEEE PAMI, Vol. 9, No. 5, pp. 629-633, Sept. 1987.

[16] P. Pudil, J. Novovicova, J. Kittler, "Floating Search Methods in Feature Selection", Pattern Recognition Letters, Vol. 15, pp. 1119-1125, 1994.

[17] R. Kohavi, "Feature Subset Selection as Search with Probabilistic Estimates", American Association for Artificial Intelligence Fall Symposium on Relevance, pp. 109-113, 1994.

[18] "Electronic Nose Vapor Identification for Space Program Application" presented at the Ninth International Symposium on Olfaction and Electronic Nose, Rome, Italy, September 2002, and submitted for publication in the Proceedings.

Acknowledgement