

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/12575173>

Neural Network Recognition of Chemical Class Information in Mobility Spectra Obtained at High Temperatures

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 2000

Impact Factor: 5.64 · DOI: 10.1021/ac991019r · Source: PubMed

CITATIONS

19

READS

15

5 AUTHORS, INCLUDING:



Suzanne Ehart Bell

West Virginia University

43 PUBLICATIONS 315 CITATIONS

SEE PROFILE



Erkinjon G. Nazarov

Draper Laboratory

93 PUBLICATIONS 2,255 CITATIONS

SEE PROFILE

Neural Network Recognition of Chemical Class Information in Mobility Spectra Obtained at High Temperatures

Suzanne Bell,^{*,†} E. Nazarov,[‡] Y. F. Wang,[‡] J. E. Rodriguez,[‡] and G. A. Eiceman[‡]

Department of Chemistry and Biochemistry, Eastern Washington University, 526 5th Street, MS 74, Cheney, Washington 99004-2431, and Department of Chemistry and Biochemistry, New Mexico State University, Las Cruces, New Mexico 88003

A minimal neural network was applied to a large library of high-temperature mobility spectra drawn from 16 chemical classes including 154 substances with 2000 spectra at various concentrations. A genetic algorithm was used to create a representative subset of points from the mobility spectrum as input to a cascade-type back-propagation network. This network demonstrated that significant information specific to chemical class was located in the spectral region near the reactant ions. This network failed to generalize the solution to unfamiliar compounds necessitating the use of complete spectra in network processing. An extended back-propagation network classified unfamiliar chemicals by functional group with a mean for average values of 0.83 without sulfides and 0.79 with sulfides. Further experiments confirmed that chemical class information was resident in the spectral region near the reactant ions. Deconvolution of spectra demonstrated the presence of ions, merged with the reactant ion peaks that originated from introduced samples. The ability of the neural network to generalize the solution to unfamiliar compounds suggests that these ions are distinct and class specific.

Ion mobility spectrometry (IMS) is an instrumental technique for detecting substances, principally organic chemicals, through the characterization of gas-phase ions in weak electric fields.¹ In IMS, response arises from ion formation and ion characterization in a drift tube maintained usually at atmospheric pressure in air. Ions are formed through atmospheric pressure chemical ionization (APCI) reactions between sample and a reservoir of charge known as the reactant ions. Then these ions, known as product ions, are characterized by drift time through a fixed distance and weak electric field. Ions strike a detector at the end of the drift tube creating a mobility spectrum, a plot of detector current (or signal intensity) versus ion drift time. Since IMS drift tubes are operated at ambient pressure, IMS instrumentation can be made small and rugged and is suited for on-site chemical measurements. Thus, IMS has been fielded in large numbers as the Chemical Agent

Monitor (CAM) for detecting chemical weapons² and the ION-SCAN for detecting explosives.³ In addition to gaseous analytes, nonvolatile substances have been examined by using laser ionization⁴ or electrospray ion sources⁵ and the utility of gas-phase mobility constants for chemical characterization is expected to grow.⁶

Despite the proven and promising uses of IMS, tools for processing and identifying spectra are still being developed and neural networks have shown promise for these tasks. The criteria necessary for acceptable network performance have been clarified recently for a large library of mobility spectra encompassing numerous chemical classes and compounds.⁷ Mobility spectra were successfully classified only when trained by using large numbers of spectra from different concentrations for a given chemical or substance. In addition, the spectra had to be obtained under uniform instrumental conditions without saturation of the ion source, i.e., where residual reactant ion peak intensity was maintained. Under such conditions, mobility spectra were identified by chemical class from spectra obtained at near-ambient temperatures;⁷ the successful classification by chemical class was unprecedented in IMS and the chemical or spectral basis for such classification was unknown.

Neural networks have been used in chemistry and analytical chemistry since the 1980s⁸ and networks have been used in IMS for probing the structural content of mobilities,⁹ for evaluating network architectures,¹⁰ and for quantitatively characterizing samples for volatile organic compounds.^{11–13} Among the advan-

* Corresponding author. Phone: 509-359-7932. E-mail: Suzanne.Bell@mail.ewu.edu.

[†] Eastern Washington University.

[‡] New Mexico State University.

(1) Eiceman, G. A.; Karpas, Z. *Ion Mobility Spectrometry*; CRC Press: Boca Raton, FL, 1994; p 145.

(2) Eiceman, G. A.; Karpas, Z. *Ion Mobility Spectrometry*; CRC Press: Boca Raton, FL, 1994; p 112.

(3) Danylewych-May, L. L.; Cummings, C. In: *Advances in Analysis and Detection of Explosives*; Yinon, J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1993; pp 385–402.

(4) Eiceman, G. A.; Anderson, G. K.; Danen, W. C.; Ferris, M. J.; Tiee, J. J. *Anal. Lett.* **1988**, 21, 539–552.

(5) Wu, C.; Siems, W. F.; Asbury, G. R.; Hill, H. H. *Anal. Chem.* **1998**, 70, 4929–4938.

(6) Baumbach, J. I.; Eiceman, G. A. *Appl. Spectrosc.* **1999**, 53, 338A–355A.

(7) Bell, S.; Nazarov, E. J.; Wang, W.-F.; Eiceman, G. A. *Anal. Chim. Acta* **1999**, 394, 121–133.

(8) Gasteiger, J.; Zupan, J. *Neural Networks for Chemists*; VCH: Weinheim, Federal Republic of Germany, and New York, 1993.

(9) Jurs, P. C.; Wessel, M. D. *Anal. Chem.* **1994**, 2480–2487.

(10) Bell, S. C.; Mead, W. C.; Jones, R. D.; Eiceman, G. A.; Ewing, R. G. *J. Chem. Inf. Comput. Sci.* **1993**, 33, 609–615.

(11) Boger, Z.; Karpas, Z. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 576–580.

(12) Zheng, P.; Harrington, P. B.; Davis, D. *Chemom. Intell. Lab. Syst.* **1996**, 33, 121–132.

tages of networks is a facility to manage large numbers of spectra, flexible designs, and ability to model complex interactions. Most applications in analytical chemistry employ back-propagation networks consisting of interconnected layers of mathematical processing elements containing weighting factors. Spectra are presented to the network as individual vectors consisting of millivolt responses at each drift time channel along with a desired output (here, class identification). Training proceeds via iterative back-propagation of error until the overall error on a training and test set falls below a user-defined threshold. The ability of the network to generalize the solution is measured by independent validation spectra. Additional details about construction and operation of neural networks are available in the references cited above.⁷⁻¹³

The previous studies with networks, applied to large spectral libraries, employed near ambient temperature drift tubes and represented a successful proof-of-concept with delineation of the criteria necessary for networks to classify mobility spectra.⁷ However, instrumental parameters now known to affect mobility spectra were not controlled in those studies. Furthermore, for analytical purposes, it is desirable to use high temperature drift tubes to minimize memory effects and condensation. While a large body of work exists with high temperature mobility spectra, neither concentration of sample nor moisture in the drift gas were known or uniformly controlled,¹⁴⁻²⁰ and they are expected to add considerable complexity to ion-molecule processes. Indeed, the complexity of both temperature and moisture effects has been previously documented,²¹ in which it was shown that to interpret these complex interactions, it is essential to control one or both. Thus, the existing spectra from high-temperature mobility spectrometry are unsatisfactory to allow for clear delineation of phenomena that might be occurring across many compound classes. Consequently, the information content of whole spectra cannot be explored by using existing spectra. To address this problem, a standardized spectral library was created for the present studies.

In the present work, high-temperature mobility spectra were obtained for 154 chemicals representing 16 chemical classes under conditions of low moisture (0.1 ppm water in air), controlled concentration of vapors in the ion source, and drift tube temperatures from 175 to 250 °C. While such conditions might not be achieved for a field-deployable instrument, insights into the fundamental ion/molecule processes will be valuable and ultimately relevant to field applications. Neural networks were used to probe the regions of chemical information in mobility spectra from this spectral library. Results described below demonstrate that class specific information is encoded in mobility spectra in

the region of drift time in and around the reactant ion peaks. This region was not previously regarded as informative and useful for spectral classification.

EXPERIMENTAL SECTION

Instrumentation. For mobility studies, a Hewlett-Packard (Palo Alto, CA) 5880 gas chromatograph (GC) was interfaced with the mobility spectrometer which was built in-house at New Mexico State University. The IMS drift tube was operated at ambient pressure at four temperatures (175, 200, 225, and 250 °C) with constant low moisture (0.1 ppm) in purified air as the drift gas. All spectra were obtained in the positive polarity. Drift gas flow rates and other instrumental parameters were set after extensive optimization studies. Signals were processed with a Graseby Analytical Wasp board/software processor (Windows Advanced Signal Processor software, Graseby Ionics, Watford, UK) designed specifically for use in an IBM type computer. Low moisture in the drift gas was attained with a series of scrubbers. The moisture level was measured by using Moisture Image Series 2 m (Panametrics, Inc. Waltham, MA) and ranged from 30 to 100 ppb (0.1 ppm), which was considered nominal. Date, time, and moisture were logged continuously via an RS-232 interface. Stability of the reactant ion peak (RIP) was measured over a 24-h period resulting in a RSD of 0.5–1.0% for RIP drift time and 2–5% for RIP intensity.

Procedures and Analytical Standards. A GC was used as a pre-separation inlet to the mobility spectrometer. To maximize data collection efficiency, analytes were tested in mixtures prepared in GC-grade methylene chloride of confirmed purity. Considerable effort was directed toward ensuring that the purity of analytes and to determine the optimal chromatographic conditions necessary to ensure baseline separation between analytes contained a given mixture. Highly volatile chemicals were assayed for purity by GC-FID using 10 μ L of headspace vapors to avoid chromatographic interference from solvent. Chemicals of low volatility were assayed by using solutions of each chemical in methylene chloride. Standard safety precautions were followed in handling the standards and solvent. All chromatograms exhibited only minor peaks, less than 1% of the total area for the analyte peak. Chemical identity of the analytes was confirmed by GC/MS. Given that the different chemical classes have different response characteristics and sensitivities in IMS, it was not feasible to introduce the same concentration of each for this work. Rather, a detailed study of concentration and response was undertaken to determine the optimal amount of each analyte to use in a given mixture. Spectra were collected only when RIP depletion was between 20 and 80%. Response across the elution gradient of a chromatographic peak was considered representative of concentration-dependent spectral features.

Daily solvent blanks were used prior to the start of data collection and injection of a solution of 2,4-lutidine was used to confirm proper operation of the mobility spectrometer. If the RIP intensity exceeded the established average by more than $\pm 5\%$, data collection was halted. Between injections of the standards, the GC was ramped to elevated temperature (180 °C) and held 20 min, and overnight the drift tube was heated to 250 °C. These quality control procedures assured the reliability of the spectral library.

(13) Boger, Z.; Karpas, Z. *Anal. Chim. Acta* **1994**, *292*, 243–251.

(14) Poziomek, E. J.; Eiceman, G. A. *Environ. Sci. Technol.* **1992**, *26*, 1313–1318.

(15) Karasek, F. W.; Maican, A.; Tatone, O. S. *J. Chromatogr.* **1975**, *110*, 295–300.

(16) Karasek, F. W.; Kane, D. M. *J. Chrom. Sci.* **1972**, *10*, 673–677.

(17) Preston, J. M.; Karasek, F. W.; Kim, S. H. *Anal. Chem.* **1977**, *49*, 1746–1750.

(18) Karpas, Z.; Bell, S. E.; Wang, Y.-F.; Walsh, M.; Eiceman, G. A. *Struct. Chem.* **1994**, *5*, 135–140.

(19) Karpas, Z. *Anal. Chem.* **1989**, *61*, 684–689.

(20) Shumate, C.; St. Louis, R. H.; Hill, H. H., Jr. *J. Chromatogr.* **1986**, *373* (2), 141–173.

(21) Bell, S. E. Ion Mobility Spectrometry of Selected Organic Compound Classes, Ph.D. Thesis, New Mexico State University, 1990.

Once spectra were collected and stored in ASCII format, they were transferred to Excel spreadsheets for further analysis. Neural networks were implemented by using NeuralSim and NeuralWare Professional II/Plus (Aspen Technologies-NeuralWare, Pittsburgh, PA). Deconvolution was carried out using Peakfit (Jandel Scientific Software, Chicago, IL).

Computational Procedures. Raw spectra were collected as ASCII files of drift time and millivolt intensities. These spectra were collated into chemical classes. Previous work⁷ demonstrated the importance of prescreening spectra for use in the library and these procedures were followed here. Since spectral patterns rather than drift times and reduced mobilities were of interest, the drift times were converted to channels numbering 1–237 for all four temperatures. Care was taken to ensure that the 237 contiguous channels covered the regions of interest and did not exclude product ions with long drift times. Truncation to 237 channels was necessitated by the column limits in Excel; columns had to be reserved to insert desired outputs into the worksheets. These values were encoded by using a binary scheme and one output node per class. Thus, for any alcohol, the desired output would be coded as 1000000000, for any aldehyde 0100000000, etc. The node with the value of one assigned to it for a given class is referred to as the target output node.

The most common method of evaluating the performance of neural networks used for classification problems is classification rate, but this metric was not applied here for the following reasons. First, a threshold for correct classification must be established. By nature, this selection is somewhat arbitrary and can produce misleading results. Depending on the value selected, results could appear much worse (by selecting an inappropriately high threshold) or much better (by selecting an inappropriately low threshold) than they were. Second, since the goal here centered on interpretation and evaluation, a direct measure of numerical output was appropriate. Accordingly, network performance was reported by using an average value⁷ of the network outputs by class. For example, the performance for all alcohols in a given data set would be measured by averaging the values in the target output node for alcohols. It should be noted that AV does not provide information that can explain poor results for a given compound or class; however, this information is available by examining individual results across all of the output nodes.

An optimization study established default settings to be used with NeuralSim for construction of the initial networks, all back-propagation type. Several measures were taken to ensure acceptable performance and to reduce or eliminate overtraining. For these initial studies, a cascading algorithm was used to minimize the number of nodes in the hidden layer. In this approach, a maximum size of the hidden layer is set from the outset, 30 here. During training, candidate hidden nodes are added one at a time. If network performance improved, the node is retained; if not, the node is reinitialized with a new set of initial random weights, is tried, and is tested and likewise a third time if needed. If none of the three candidate nodes improves the network's performance, no further hidden nodes are added.

Regardless of the size of the input vectors (ranging from 15 to 237), the typical size of the hidden layer was ~10 nodes. By use of iterative trials of multiple variable and transforms by a genetic algorithm, the optimal number of input variables and their

preprocessing transforms were selected. Briefly, the genetic algorithm starts with random combinations of input variables, combines two parents (subsets of variables and preprocessing transforms), and determines the fitness of the combination (the offspring). Based on fitness, the new combination is either retained and recombined or discarded until maximum fitness is obtained. The maximum number of generations was set to 80, but selection ceased, if after seven successive generations, performance did not improve. Multiple linear regression was used to judge performance of the preprocessed variables. The most common preprocessing transforms selected were the hyperbolic tangent, inverse ($1/x$) and linear, all scaling the input between -1.0 and 1.0 . This combination of a genetic algorithm for variable selection and cascading addition of hidden nodes constituted the minimalist approach to constructing the networks.

Prior to training the networks, the spectral library was divided into working and validation sets. For the initial networks, an internal validation set was regenerated automatically by the software for each network. While this provides an excellent measure of network performance, changing composition of this validation set hampers comparisons between networks. Thus, two additional validation sets were created for each temperature. The first contained spectra selected at random from the working data prior to any network training or testing. Unlike the internal validation set, this contained exactly the same spectra for a given temperature, thus simplifying comparisons. The second external validation set consisted of representative compounds removed from the working data before network training. For example, all heptanol spectra were withheld from the network, including the entire elution profile, and such compounds are thus completely unfamiliar to the network. Performance of a network when analyzing unfamiliar compounds reflects the network's ability to generalize a solution and thus measures how applicable the network's solution is to novel situations. Compounds used for this validation set were consistent across all temperatures and networks, but differed from the other validation sets in that the entire elution profile (5–15) spectra from any one compound were included. Conversely, the other validation sets contained randomly selected spectra, not entire elution profiles.

The last preliminary step involved the determination of reproducibility. Within a network, many factors contribute to uncertainties. Each new network starts with a new set of random weights and parents for the genetic algorithm used for variable selection. Thus, each network generates a unique solution even if all initial settings are identical. Likewise, differing compositions of training, test, and validation spectra will generate variation in results. To obtain a reasonable measure of these combined uncertainties, five networks were applied to spectra obtained at all four drift tube temperatures encompassing 11 compound classes. For each temperature, five different independent validation sets (familiar compounds) were randomly selected from the spectral library. Thus, each network started with different random weights, different training/testing/validation spectra, and different variables as selected by the genetic algorithm. Results are presented in Table 1. For the AVs, confidence intervals (95%) were generally 0.10 or less. Exceptions were most numerous at the lowest IMS operating temperature (175 °C). In discussions to follow, an uncertainty of ± 0.10 AV was assumed except for alkenes

Table 1. Reproducibility and Uncertainties^a

T, °C	all	1	2	3	4	5	6	7	8	9	10	11
175												
mean		0.89	0.65	0.82	0.71	0.64	0.88	0.64	0.64	0.81	0.86	0.72
95% CI		0.02	0.08	0.15	0.09	0.12	0.07	0.31	0.31	0.06	0.09	0.16
200												
mean		0.78	0.88	0.82	0.89	0.92	0.71	0.82	0.71	0.68	0.71	0.82
95% CI		0.05	0.11	0.04	0.05	0.05	0.07	0.08	0.33	0.12	0.14	0.08
225												
mean		0.95	0.88	0.93	0.77	0.69	0.94	0.85	0.59	0.93	0.60	0.99
95% CI		0.02	0.07	0.06	0.08	0.06	0.02	0.06	0.37	0.06	0.15	0
250												
mean		0.82	0.82	0.81	0.81	0.94	0.70	0.87	0.68	0.87	0.83	0.91
95% CI		0.08	0.06	0.13	0.07	0.07	0.08	0.12	0.45	0.05	0.11	0.08

^a Compound class identification; 1, alcohols; 2, aldehydes; 3, alkenes; 4, benzenes and substituted benzenes; 5, esters; 6, ketones; 7, nitro compounds; 8, polyaromatic hydrocarbons; 9, phenols; 10, sulfides; 11, RIP spectra. Results are in AV values.

and RIPs at 175 °C (± 0.20), nitro-compounds at 175° (± 0.30), and sulfides at 225 °C (± 0.20). Given the large uncertainty associated with polyaromatic hydrocarbons, extreme care was taken in data interpretation regarding this class. The reason for this was not known but could reflect an inadequate characterization of the class through the spectra and compounds selected.

To evaluate the relative contribution of different spectral regions, initial network studies involved six neural networks using different regions of the spectra. Network type one was considered the baseline and used the whole spectrum (channels 1–237), type 2 used the product ion peak region only, type 3 the “pre-RIP” section, type 4 the pre-RIP + RIP; type 5 the RIP only; and type 6 the whole spectrum minus the pre-RIP. These regions were determined by overlay plots of 10% of the library taken at random and visually selecting the regions separated by the baseline of the RIP.

Once results were collated and compared, deconvolution was performed on selected spectra to search for hidden features and patterns. Again, an extensive optimization study was performed to determine baseline parameters. Within the Peakfit program, options selected were baseline subtraction and fitting to a Voigt amplitude function. Peak width was varied between 12 and 14 based on full width half-maximum, a filter setting of 50–70 with variable peak widths enabled. For individual fits, these parameters were varied such that the correlation coefficient was 0.99 or better. Deconvolution was used to investigate underlying spectral features but was not used to preprocess spectra for the neural networks.

RESULTS AND DISCUSSION

High-Temperature Spectra Library and Network Results for Minimalist Approach. In the initial effort to examine mobility spectra by neural networks, the minimalist approach resulted in the selection of ~10–35 data points per whole spectrum (or a smaller number of only portions of spectra) unique for each network type and temperature. The results from 24 experiments are shown in Tables 2 and 3. A network application in which whole spectra were used (as opposed to only selected spectral regions) demonstrated that increases in temperature generally resulted in improved network performance overall and within individual classes. As shown in Table 2 (see Network Type 1), values from testing of familiar compounds showed overall AVs of 0.78–0.93.

Table 2. External Validation Results (AV values), Initial Network Studies, Familiar Compounds^a

T, °C	all	1	2	3	4	5	6	7	8	9	10	11
Network Type 1: Whole Spectrum												
175	0.78	0.93	0.69	0.79	0.67	0.64	0.88	0.90	0.96	0.67	0.68	0.81
200	0.74	0.78	0.87	0.70	0.84	0.92	0.52	0.84	0.76	0.63	0.31	0.95
225	0.83	0.98	0.83	0.87	0.87	0.68	0.84	0.72	0.61	0.85	0.88	10.00
250	0.93	0.98	0.93	0.89	0.91	0.98	0.92	0.95	0.96	0.96	0.93	0.96
Network Type 2: Products Ions												
175	0.59	0.77	0.36	0.60	0.61	0.59	0.81	0.38	0.98	0.45	0.42	0.52
200	0.57	0.38	0.38	0.65	0.74	0.71	0.34	0.73	0.52	0.68	0.56	0.58
225	0.51	0.69	0.56	0.38	0.38	0.38	0.46	0.45	0.47	0.63	0.39	0.85
250	0.65	0.48	0.38	0.52	0.73	0.97	0.42	0.84	0.44	0.54	0.87	0.92
Network Type 3: Pre-RIP												
175	0.48	0.73	0.34	0.64	0.36	0.19	0.46	0.55	0.55	0.57	0.42	0.48
200	0.39	0.64	0.37	0.42	0.37	0.59	0.29	0.23	0.47	0.22	0.19	0.54
225	0.61	0.92	0.80	0.74	0.31	0.29	0.72	0.27	0.89	0.62	0.19	10.00
250	0.53	0.39	0.48	0.47	0.45	0.55	0.28	0.71	0.76	0.45	0.48	0.85
Network Type 4: Pre-RIP and RIP												
175	0.66	0.81	0.67	0.72	0.42	0.60	0.81	0.74	0.69	0.78	0.38	0.72
200	0.79	0.79	0.96	0.72	0.79	0.96	0.65	0.89	0.96	0.86	0.34	0.80
225	0.82	0.99	0.97	0.83	0.57	0.64	0.90	0.82	0.99	0.77	0.52	10.00
250	0.73	0.68	0.95	0.52	0.88	0.82	0.34	0.82	0.76	0.72	0.58	0.95
Network Type 5: RIP												
175	0.59	0.46	0.67	0.64	0.77	0.46	0.68	0.66	0.68	0.69	0.34	0.66
200	0.64	0.46	0.69	0.65	0.84	0.56	0.50	0.72	0.43	0.86	0.58	0.73
225	0.70	0.96	0.86	0.71	0.65	0.41	0.69	0.64	0.52	0.75	0.62	0.93
250	0.51	0.68	0.74	0.32	0.50	0.59	0.54	0.20	0.20	0.45	0.52	0.86
Network Type 6: Whole Spectrum minus pre-RIP												
175	0.80	0.66	0.58	0.96	0.82	0.76	0.87	0.88	10.00	0.79	0.66	0.79
200	0.70	0.65	0.80	0.76	0.72	0.72	0.57	0.68	0.96	0.57	0.40	0.83
225	0.75	0.98	0.93	0.77	0.66	0.58	0.90	0.66	0.37	0.91	0.56	0.99
250	0.77	0.84	0.69	0.57	0.64	0.91	0.89	0.79	0.68	0.61	0.87	0.97

^a The compound classes are identified in the footnote to Table 1.

When spectral regions were selected for interrogation or use in the network training, results suggested substantial differences in information content in various regions of the spectra. For example, the spectral region in and immediately around the reactant ion peaks (hydrated cluster of NH_4^+ , NO^+ , and H^+) contained significant chemical information sufficient to associate substances by chemical class (Table 2, network type 4). This is seen in AVs of 0.66–0.82 with some high scoring chemical classes such as alcohols (column 1), aldehydes (column 2), polyaromatic hydrocarbons (column 8), and nitro compounds (column 7). Other chemical classes such as sulfides (column 10) and benzenes/alkylated benzenes (column 4) showed poor performance, while phenols (column 9) and alkenes (column 3) showed moderate success ($\text{AV} > 0.6$). None of these results in network type 4 included the region of a mobility spectrum containing the product ions, which have been traditionally considered the information rich region of mobility spectra. This means that significant chemical information is encoded or contained within the spectral region thought commonly to hold only details of reactant ions, i.e., only the reservoir of charge for APCI reactions. Results for this spectral region did not correlate with temperature, though the meaning and origins of this are unclear.

The region of the reactant ions is comprised of three peaks located in these studies between 2 and 4 ms at 250 °C and 3–5 ms at 175 °C, and the peaks were resolved at baseline. When the peaks for the two minor reactant ions with short drift times (the ammonium and nitric oxide peaks) were removed from the network, the AVs dropped 0.1–0.2 and when the pre-RIP region

Table 3. External Validation Results, Unfamiliar Compounds^a

T (°C)	all	1	2	3	4	5	6	7	8	9	10	11
Network Type 1: Whole Spectrum												
175	0.54	0.63	0.59	0.88	0.60	0.19	0.79	0.06	0.09	0.99	0.31	0.80
200	0.64	0.78	0.84	0.59	0.73	0.98	0.81	0.15	0.26	0.97	0.08	0.88
225	0.60	0.98	0.68	0.98	0.56	0.47	0.36	0.05	0.49	0.99	0.06	10.00
250	0.63	0.98	0.19	0.77	0.72	10.00	0.75	0.09	0.29	0.97	0.21	0.99
Network Type 2: Products Ions Only												
175	0.41	0.63	0.08	0.76	0.30	0.20	0.48	0.00	0.62	0.77	0.08	0.58
200	0.35	0.35	0.37	0.57	0.39	0.35	0.11	0.11	0.07	0.77	0.05	0.71
225	0.34	0.38	0.39	0.59	0.23	0.16	0.07	0.02	0.24	0.73	0.06	0.91
250	0.35	0.11	0.18	0.05	0.19	0.84	0.43	0.00	0.37	0.69	0.07	0.87
Network Type 3: Pre-RIP Only												
175	0.50	0.76	0.24	0.77	0.49	0.40	0.59	0.35	0.50	0.56	0.62	0.22
200	0.44	0.72	0.53	0.44	0.38	0.74	0.31	0.40	0.20	0.24	0.40	0.52
225	0.67	0.98	0.87	0.74	0.38	0.40	0.69	0.49	0.80	0.62	0.36	10.00
250	0.46	0.38	0.43	0.40	0.54	0.76	0.22	0.11	0.46	0.45	0.38	0.89
Network Type 4: Pre-RIP and RIP												
175	0.66	0.72	0.46	0.95	0.68	0.62	0.73	0.66	0.68	0.66	0.83	0.29
200	0.75	0.88	0.96	0.84	0.92	0.88	0.76	0.33	0.62	0.74	0.50	0.81
225	0.76	10.00	0.97	0.86	0.82	0.49	0.99	0.22	0.78	0.79	0.47	10.00
250	0.63	0.83	0.69	0.36	0.78	0.98	0.62	0.04	0.26	0.69	0.70	0.96
Network Type 5: RIP Only												
175	0.61	0.63	0.33	0.84	0.78	0.49	0.67	0.58	0.80	0.61	0.72	0.27
200	0.58	0.68	0.78	0.42	0.84	0.30	0.56	0.17	0.61	0.76	0.42	0.82
225	0.65	0.97	0.85	0.70	0.79	0.22	0.66	0.19	0.77	0.74	0.31	0.94
250	0.48	0.61	0.41	0.02	0.31	0.99	0.40	0.27	0.28	0.56	0.57	0.91
Network Type 6: Whole Spectrum minus pre-RIP												
175	0.73	0.56	0.69	0.90	0.51	0.47	0.91	0.40	0.99	0.97	0.79	0.80
200	0.51	0.59	0.83	0.38	0.27	0.45	0.32	0.14	0.79	0.92	0.07	0.83
225	0.63	0.96	0.73	0.90	0.33	0.51	0.85	0.00	0.67	0.96	0.02	0.98
250	0.53	0.59	0.49	0.06	0.32	0.99	0.83	0.00	0.78	0.73	0.04	0.95

^a The compound classes are identified in the footnote to Table 1.

was considered alone, performance was the poorest of any network with the exception of the 225 °C results. By contrast, removal of the product ion peak (Table 2, network 4) showed smaller decreases in performance and even improvement at 200 °C. These results suggest that the information encoded in this region was relatively minor.

In sharp contrast to results obtained by networks for the spectral region around the reactant ion peaks, results from isolating and testing the spectral region associated with product ions yielded uniformly poor results across chemical classes (Table 2, network type 2) with AV of 0.51–0.65. This low performance implies that drift times for product ions alone do not contain information sufficient to assign a spectrum to a particular chemical class. Since product ions in IMS are commonly protonated monomers of the kind $M \cdot H^+ \cdot (H_2O)_n$, there is no reason to expect class specific information to be contained in and around such peaks. This finding, as with the reactant ion peak experiments noted above, was not apparently affected by temperature.

In summary, networks were able to classify spectra from familiar chemicals according to chemical class by using principally the spectral region near the reactant ion peak. A secondary contribution was seen from the remaining portion of spectra, i.e., that region where product ions appear. These results were accomplished by using spectra drawn from substances known to the network, i.e., familiar compounds. An improved test of the network ability to generalize a solution may be obtained by using totally unknown or unfamiliar substances (i.e., network had seen an example only for the chemical class but not any spectra from the substance).

The use of minimal networks with the high-temperature spectral library showed reduced performance when tested with unfamiliar chemicals. In some instances the effects were dramatic as with 2-octanone (Table 3, network type 1, column 6 at 225 °C). When spectra for 2-octanone were presented to a trained network, the network could not unambiguously recognize these spectra as a ketone. Results for AVs ranged from 0.02 to 0.86 across concentrations (not shown) with a mean of 0.36. In other instances including alcohols (column 1), alkenes (column 3), ketones (column 9), and phenols (column 6), the AVs suggested fairly strong performance with some isolated, inexplicable but strong exceptions. Indeed, the effects of temperature as seen in Tables 1, 2, and 3 are significant yet inconsistent. For example, in Table 3, there is a large general decrease in performance between 225 and 250 °C. The reasons for the observed temperature effects are unknown and currently under investigation.

The decline in performance of the minimal networks using unfamiliar compounds suggested that networks were not able to generalize in a satisfactory manner. Since the minimal network's select variables based upon all the spectra including examples of the validation spectra, removal of all spectra from a given chemical (i.e., those deemed unfamiliar) in this experiment would have affected the choice of variables and hence the generalizing capabilities. This was strong evidence that minimal networks were limited for detailed exploration of spectral content. Accordingly, further investigations were made by using full spectra.

A network was constructed by using spectra obtained at 250 °C. This network, rather than using selected channels, incorporated all points of the spectrum as input and used a cascading algorithm to add hidden units in a stepwise manner as described previously, resulting in a hidden layer consisting of 11 units. Overall AV for familiar compounds was 0.93 and ranged from 0.83 for the phenols to 1.00 for aldehydes and nitro compounds. With unfamiliar compounds, the overall AV dropped to 0.67, but excellent performance was seen for alcohols (0.97), aldehydes (0.91), esters (1.00), PAHs (0.93), and phenols (1.00). While performance is still not optimal across all compound classes, important information could still be gleaned from this network.

Full Spectral Network with an Expanded Back-Propagation Network. The use of an extended back-propagation network had been shown to provide satisfactory performance when networks were trained by using whole spectra and numerous examples from a range of concentrations.⁷ Given that the networks discussed in the previous section still showed limitations in performance, the next step was to apply such an extended back-propagation network to the spectral library to determine if performance could improve while still providing acceptable performance as judged by the use of unfamiliar compounds. This network, larger and covering more compounds and classes (16 versus 11), used the whole spectrum as input, a log function for preprocessing, and a hidden layer of 100 nodes. In recognition of the dangers inherent in such large networks (i.e., overtraining and an inability to generalize), five different validation sets were used, each comprising different combinations of familiar spectra as well as a separate collection of unfamiliar spectra. Uncertainties for a similar network have been previously established,⁷ with the 95% confidence intervals ranging from 0.03 to 0.12. Results from network experiments with various combinations of familiar

Table 4. Unfamiliar Compound Results, Expanded Back Propagation Network

class	compound	no. of spectra	AV	AV class	class	compound	no. of spectra	AV	AV class
aldehydes	hexanal	13	0.95	0.95	benzenes	ethyl benzene	8	1.00	0.80
	benzaldehyde	15	0.82			styrene	8	0.11	
	decanal	13	0.99			<i>p</i> -cymene	9	0.97	
	undecanal	15	0.99			<i>p</i> -xylene	7	0.96	
	pentanal	11	1.00			<i>p</i> -diisopropylbenzene	11	0.94	
alkenes	2,5-dimethyl-2,4-hexadiene	12	0.77	0.81	esters	2-ethyl butanoic acid	7	0.98	0.95
	<i>cis</i> -cyclooctene	12	0.99			methyl hexanoate	18	0.93	
	1-undecene	17	0.70			ethyl octanoate	11	0.98	
	isoprene	12	0.58			2-ethylbutanoic acid methyl ester	8	0.96	
	1-octene	12	1.00			propyl butanoate	10	0.92	
cycloalkanes	<i>cis</i> -1,3-dimethylcyclohexane	14	1.00	0.98	ketones	3-octanone	10	0.99	0.62
	1,4-dimethylcyclohexane	13	1.00			4-methylcyclohexanone	10	0.40	
	cyclooctane	15	1.00			2-decanone	10	1.00	
	isopropylcyclohexane	8	1.00			4-methylcyclohexanone	10	0.42	
	propylcyclohexane	11	0.90			2-hexanone	10	0.28	
ethers	<i>tert</i> -amyl methyl ether	10	0.96	0.93	nitro compounds	2-nitropropane	14	0.84	0.43
	pentyl ether	12	1.00			2-nitrotoluene	16	0.00	
	hexyl ether	14	0.88			nitrobenzene	15	0.87	
	hexyl ether	14	0.85			nitromethane	6	0.00	
	<i>tert</i> -butyl methyl ether	16	0.96			<i>m</i> -cresol	21	1.00	
mercaptans	1-pentanethiol	10	1.00	0.91	phenols	hydroquinone	16	0.41	0.87
	nonylmercaptan	16	0.99			catechol	21	0.99	
	1,8-octanedithiol	17	0.55			<i>o</i> -cresol	16	1.00	
	3-methyl-1-butanethiol	14	1.00			3,4-dimethylphenol	18	0.96	
	2-methyl-1-butanethiol	11	0.99			valeric acid	14	0.96	
organophosphates	diethyl methyl phosphonate	9	0.98	0.99	carboxylic acids	2-methylbutyric acid	9	0.66	0.59
	diethyl ethyl phosphonate	9	1.00			heptanoic acid	15	0.31	
	triisobutyl phosphate	7	1.00			4-pentanoic acid	17	0.94	
	triisoprophosphate	6	0.98			butyric acid	15	0.09	
	triisobutyl phosphate	7	1.00			ethyl sulfide	12	0.10	
PAHs	fluorene	7	0.30	0.60	alkyl sulfides	propyl sulfide	12	0.34	0.15
	phenanthrene	8	0.42			butyl sulfide	10	0.24	
	phenanthrene	8	0.97			ethyl methyl sulfide	12	0.01	
	acenaphthylene	6	0.99			<i>tert</i> -butyl sulfide	10	0.07	
	naphthalene	7	0.30						
alcohols	3-methyl-3-pentanol	9	0.92	0.96	summary	av		0.79	
	amyl alcohol	19	0.93			standard deviation		0.32	
	1-nonanol	21	1.00			95% CI		0.07	
	octanol	15	1.00						
	1-undecanol	11	0.93			minus alkyl sulfides		0.83	
alkanes	decane	17	0.97	0.99				0.28	
	<i>n</i> -heptane	10	1.00					0.06	
	<i>n</i> -undecane	15	1.00						
	tridecane	15	0.99						
	2,2,4-trimethylpentane	8	1.00						

compounds at 250 °C gave performance with AVs ranging from 0.78 to 1.00, with the vast majority of the familiar compounds yielding AVs of 0.95 and above.

Results with unfamiliar compounds are presented in Table 4. The use of whole spectra (with all data points) including spectral regions for the reactant ion peak and product ion peak yielded superb results for generalization of spectra by chemical class. For example, pentanal (see Table 4 under aldehydes) was properly assigned to the aldehyde class even though the network had not seen spectra for pentanal and had seen only spectra for other aldehydes and all other chemical classes. This was done without error across a range of concentrations for pentanal (i.e., all values were perfect so the AV = 1.00). As shown in Table 4, perfect results or very favorable results (AVs > 0.95) were obtained for 42 of the 75 tests (~56%). When the number of good results (AVs > 0.90) were included, the results were 50 of 75 or 67%. The number of poor results (AVs < 0.75) were 15 of 75 or 20%. Those chemical classes that yielded the worst results were alkyl sulfides,

ketones, carboxylic acids, and nitrogen-containing chemicals and in these several tests were very poor in each instance.

In some cases (hexyl ether and phenanthrene), two different sets of spectra were included in the different validation sets as replicates. For phenanthrene, a dramatic drop in performance is noted and this was attributed to a small population of phenanthrene spectra. Also, there are instances where similar compounds such as 2-nitropropane and nitromethane show dramatically different results (AVs of 0.84 and 0.00, respectively). The reasons for this are unknown and require further investigation.

Clearly, these spectra contained chemical information that permitted categorization by chemical class and has been historically considered impossible or unrealistic for IMS since spectra were thought to be relatively simple and to contain only protonated cluster ions. However, findings here show that structural information common to a chemical group or moiety is encoded in the spectra. These results demonstrated that when a mobility spectrum is obtained at high temperature and low moisture, informa-

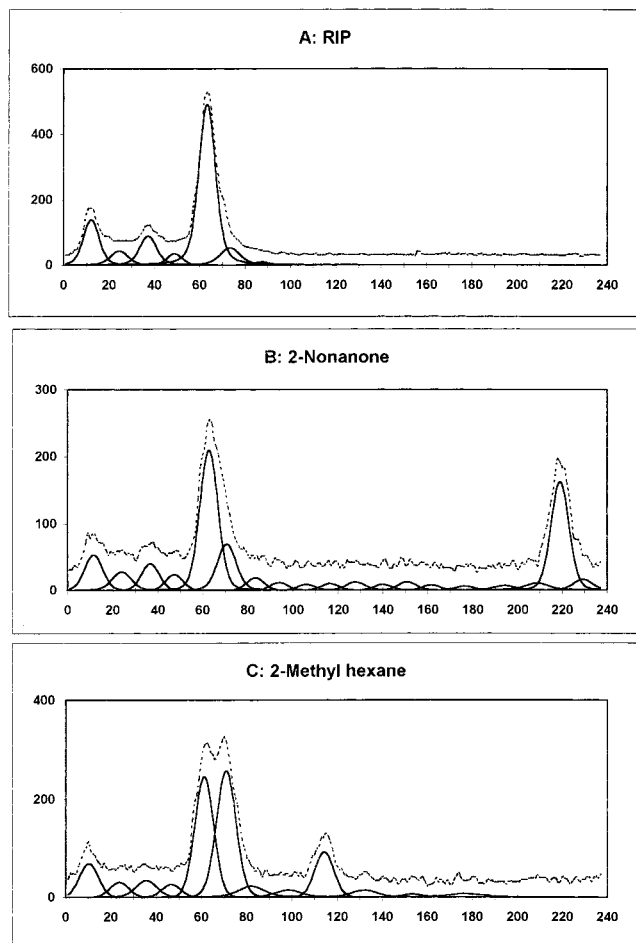


Figure 1. Results of deconvolution. The x-axis is the channel number and the y-axis is the intensity in millivolts. The original spectrum is offset and shown above the deconvoluted spectrum.

tion can be found in the region associated with the RIP, that is, a portion of the drift time separate from the commonly observed product ions of the MH^+ or M_2H^+ variety. That is, ions with fast drift times and high mobilities contain chemical class specific information.

Deconvolution of Spectra. Once neural networks had provided evidence that class specific information was contained in mobility spectra, examples of spectra from various classes were examined by deconvolution methods. Results from the reactant peak are shown in Figure 1 (trace A) and the peaks from left to right are $NH_4^+(H_2O)_n$, $NO^+(H_2O)_n$, and $H^+(H_2O)_n$ (at 250 °C, values for $n = 1-2$). Deconvolution of this spectrum showed the presence of a small peak near channel 75. The origins and identity of this peak are unknown and are believed to be due to off-gassing of drift tube materials. Ketones were a chemical class for which classification by functional group was fair (AVs of ~ 0.6) and a spectrum for 2-nonanone is shown in trace B. In the spectrum, the reactant ion peaks are evident along with a pronounced product ions peak at channel 220 and this should be a proton bound dimer, $MH^+(H_2O)_n$, based upon prior IMS/MS findings.²¹ However, the region for the RIP for the deconvoluted spectrum of 2-nonanone differs from that for the deconvoluted RIP spectrum

(trace A) in the same region near the hydrated proton. Studies with spectra from various concentrations for this chemical (and all others) demonstrated that the change around the RIP with a chemical was not governed by concentration. The location of these peaks, due to presence of 2-nonanone, is suggestive of small ions or ions with low mass such as might be seen from fragment ions. The shoulder peak at channel 75 is likely not associated with the contamination mentioned above given the slight shift in peak location and the larger intensity. Other ketones showed similar profiles as this including a strong product ion at a drift time corresponding to a protonated monomer and a second or third peak in or near the hydrated proton of the RIP.

Ions with drift times near the RIP for some chemicals were evident without deconvolution as shown in trace C. The mobility spectrum for 2-methylhexane shows a product ion at channel ~ 115 that has a drift time anticipated for M^{+} .²² However, an intense peak at channel 75 is merged with the RIP and can be understood as fragment ion and is readily apparent in the deconvoluted spectrum. This pattern of a small ion near the RIP was seen with other alkanes and mass-identification from APCI-MS of alkanes supports a conclusion that this is a fragment ion.²² The presence of ions near the RIP and in some instances between the RIP and expected drift times for intact protonated molecules was also evident with ethers, phenols, alcohols, alkenes, mercaptans, and aldehydes. The peaks characteristic of these small ions suggested that their size was comparable to that for the ketone already described. However, for neural networks to categorize unfamiliar spectra by functional group, the location of these ions is presumably distinct and recognizable by class. The process of generalization is indicative of class-specific fragmentation that is reflected in the spectral region near the RIP.

CONCLUSION

Neural networks were demonstrated as tools for probing the information content of mobility spectra and revealed that significant amounts of class-specific information, presumably fragmentation, was located in the region of a mobility spectrum occupied by the reactant ion peaks. This region has rarely been considered significant for mobility spectrometry and was probed by using tests of relative variable contribution and deconvolution of original spectra. The results strongly support a conclusion that mobility spectra obtained at elevated temperatures contain small ions that presumably arise through fragmentation reactions. Such reactions, based upon the ability for networks to generalize solutions for unfamiliar chemicals, may be expected to be common within a chemical class and distinct from other chemical classes. Further investigations into the nature of the fragments using IMS-MS are currently under way as are detailed studies of deconvoluted spectra and their relationships to the mass-identified ions. Temperature effects as a function of drift tube length are also ongoing.

ACKNOWLEDGMENT

The financial support from NASA (grant no. NAGY-4558) and US Army Research Office (grant no. DAAH04-95-1-0541) is gratefully acknowledged. Additional support from the Oak Ridge Institute for Science and Education are gratefully acknowledged.

Received for review September 3, 1999. Accepted January 5, 2000.

AC991019R

(22) Bell, S. B.; Ewing, R. G.; Eiceman, G. A.; Karpas, Z. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 177-185.