Algorithm for Automatic Peak Detection and Quantification for GC-IMS Spectra

Jorge M. Fernandes[1,2(✉)], Valentina Vassilenko[1,2],
and Paulo H. Santos[1,2]

[1] Laboratory for Instrumentation, Biomedical Engineering and Radiation
Physics (LibPhys-UNL), NOVA School of Science and Technology, NOVA
University of Lisbon, Campus FCT-UNL, 2896-516 Caparica, Portugal
j.manuel@campus.fct.unl.pt
[2] NMT, S.A., Edifício Madan Parque, Rua dos Inventores,
2825-182 Caparica, Portugal

Abstract. Ion Mobility Spectrometry with a coupled Gas Chromatography (GC-IMS) pre-separation is an analytical technique suitable for detection of volatile organic compounds (VOCs) in complex sample matrices (indoor & outdoor air, breath samples, food, beverages, microbial cultures, etc.). Its outstanding sensitivity allows in-situ analysis of a very large range of organic compounds at low concentrations with detection limits typically in the low ppb or even ppt level. Automatic detection and quantification of VOCs through GC-IMS spectra is challenging and the lack of computational methodologies able to detect, quantify and deconvolute overlapped peaks are still scarce and diminished. In this work we present a preliminary algorithm and still in development for automatically identify and quantify VOC peaks directly from the spectra matrix with an established threshold, a noise filter, Reactive Ion Peak (RIP) measurements. Herein, proposed tools may be very useful for quick automatic detection and quantification of compounds in GC-IMS spectra.

Keywords: Ion Mobility Spectrometry (IMS) · Gas Chromatography (GC) · Volatile Organic Compounds (VOCs) · Peak detection · Quantification · Automation · Algorithm

1 Introduction

Ion mobility spectrometry (IMS) was developed in last the 50 years as a method for detecting and identifying trace levels (ppb_v and ppt_v ranges) of semi-volatile and volatile organic compounds (VOCs), mainly in security and military venues. IMS working principle is based on mobility determination in electric fields of gas phase ions in a sample from a large array of matrices [1].

Modern ion mobility analytical spectrometers were commercially available only in the late 1970s due to military and governmental control of the technology. Thus, the period afterwards saw a boom in intensive substance characterization by ion mobility spectrometry [2]. This analytical technique was initially known by other terms (*e.g.* plasma chromatography, gaseous electrophoresis or ion chromatography), however the

working principle of current and modern instrumentation remained [1–5]. Nonetheless, engineering and technological improvements have opened numerous applications and uses of IMS such as the development of portable spectrometers for field use [1, 2, 4]. General IMS operational principles are summarized in Fig. 1 and include:

- Transference of sample as vapor into an ion source (radioactive sources: ^{3}H, ^{63}Ni; Non-radioactive sources: corona discharges, electrospray or lasers);
- Production of ions from neutral sample molecules at atmospheric pressure
- Injection of an ion swarm into the drift region;
- Determination of drift velocities of ions under the influence of an electric field in the drift region and in a supporting atmosphere, the drift gas;
- Detection of ions and electrical signal storage or display, with or without automated analysis of the result.
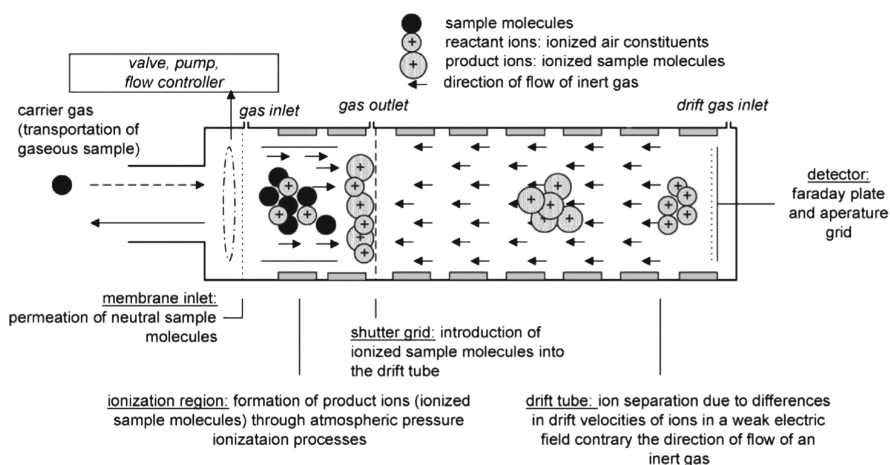

Fig. 1. Ion mobility spectrometry operation summary in a bidirectional flow system [1].

Ions movement speed, or the drift velocity (v_d) dissipated by collisions with neutral molecules of the supporting gas atmosphere, is proportional to the strength of the electric field (E) with the constant of proportionality being the ions mobility (K) [2, 5]. Thermalized ions typically travel with a speed of approximately 2 m s^{-1} and traverses the drift region with lengths of 5 to 15 cm in a few milliseconds (2 to 15 ms). Ion drift time correlates with ions' mass, charge and collision cross section, which includes structural parameters (physical size and shape) and the electronic factors describing the ion-neutral interaction forces. Therefore, different drift velocities are attained for ions with different structure (shape) and mass, establishing the basis for ions separation in IMS [5].

Engineering advances provide to IMS-based methods a major advantage in ana-lytical application due to the analyser's low size, weight and power consumption,

making this instrumentation perfectly suitable for on-site or in-field monitoring, contrary to almost all analytical tools [1, 6, 7]. IMS analysers exhibit fast response and reliable performance (high sensitivity, recording of ion mobility spectra) and can be used in ambient pressure, with nitrogen, helium and air as drift gas.

Several IMS devices have been employed in airports worldwide for chemical-weapons monitoring and explosive detection in hand-held or bench analyser formats [8]. Applications in civilian fields are more diverse and include investigations with complex, humid gas-phase biological samples [6, 9], health and medical diagnostics [10], food quality and safety [11], as well as in the industrial process control [12], petrochemical, environmental analysis [13, 14] and air quality assessment [15–17]. However, in complex matrix analysis, a single IMS device has limitations, such as clustering forming in the ionisation region, thus making identification of the ions difficult or even impossible. Therefore, to solve this limitation and increase the selectivity, ion mobility spectrometry is usually coupled to a pre-separation method: Gas-Chromatographic column (GC), Multi-Capillary Column (MCC) or, not so frequently, Liquid-Chromatography (LC) [18].

2 Contributions to Life Improvement

Recent successful CG-IMS technology applications to environmental analysis, medical diagnostics, process control, air quality control, food quality control [19], biomolecules characterization and detection of biomarkers in bacteria [20] show a clear need for tools that allow quick and precise spectra processing.

Experimental research data derived from GC-IMS is represented by 3D graphs, also called heatmaps or spectra, where each analyte is given by retention and drift time for qualitative analysis, and intensity for quantitative analysis. Currently, software availability for automatically detect and process analyte peaks from 3D GC-IMS spectra is scarce, limited or functional for a single instrumentation type. Thus, a generalized automatic peak detection, identification and quantification algorithm will improve, accelerate and enrich IMS instrumentation when employed in the numerous life science fields previously mentioned.

3 Materials and Methods

3.1 Input Data Format

Ion mobility spectrometers produce data in 2D graphs format in which the x and y axis are respectively, drift time (t_d) and intensity (Fig. 2). Drift time is in milliseconds (ms) and it's usually expressed in relation to the Reactive Ion Peak (RIP) drift time (RIPrel). RIP refers to ionized ions of the drift gas and corresponds to the quantity of ions available to ionize analytes. RIP drift time varies with conditions such as drift gas type and humidity, and analytes' RIP relative drift times are employed to standardize drift times allowing their identification and peak comparison of analytes between measurements.

When IMS is coupled with pre-separation techniques, as CG or LC, experimental data obtained changes from 2D graphs to 3D plots, heatmaps, which often are called spectrum (singular) and spectra (plural), as it can be seen in the Fig. 2. Spectra in 3D format contains data with three variable: (a) retention time (t_r) of the gas or liquid chromatographic column, (b) drift time (t_d) for the separation of analytes in the drift tube and (c) intensity (I) detected in a faraday plate at the end of the drift tube. Retention time is expressed in seconds, drift time in relation to the RIP and intensity in volts. A more detailed description can be found elsewhere [21].

IMS devices typically have their proprietary software for signal processing, saving measurements files and processing them. The spectra selected for the development of the algorithm to automatically detect and quantify peaks derived from a GC-IMS device, commercially available from G.A.S. Dortmund (*Gesellschaft für analytische Sensorsysteme*) sold as BreahSpec®. The software Laboratory Analyser Viewer (LAV) was provided along the device and can load the output files of the GC-IMS that come in a *.mea* file format. The software represents the *mea* file data in a heatmap, allows the definition of peak areas for quantification and extraction of drift and retentions times, management of measurements projects (reading several *mea* files in simultaneous) among many other features. *Mea* files can be extracted into a CSV format file, containing three degrees of information.
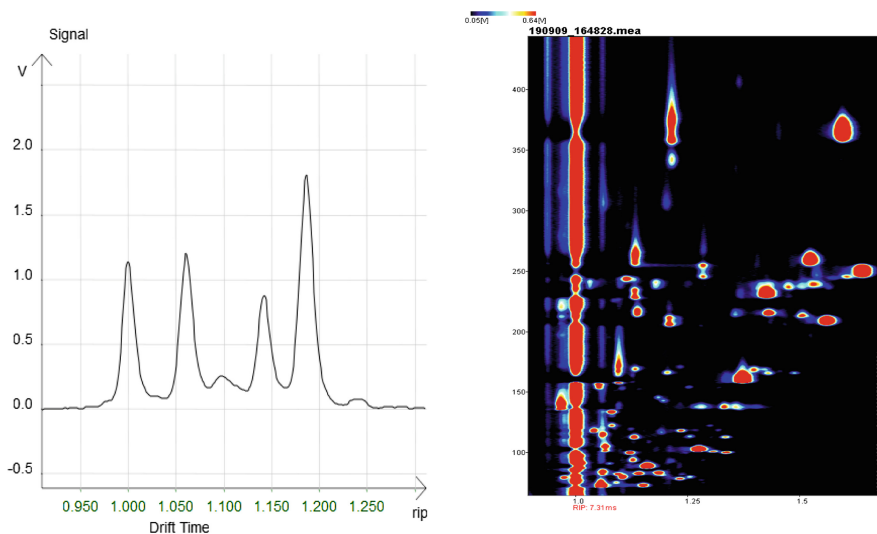

Fig. 2. Single IMS or 2D spectrum (right) and 3D spectrum or heatmap examples (left) (spectra processed and obtained from LAV of BreathSpec (GC-IMS) measurements).

Output data of a single IMS measurement, 2D spectrum, is a vector $S = (z_0, z_1, \ldots, z_n)$ of signal intensities z_i measured in equidistant time point dt_i, $i \in \{1 \ldots N\}$. If a pre-separation technique is coupled to an IMS (GC or LC), an additional dimension is provided in the 3D spectrum, the retention time. Therefore GC-IMS data become a series of R one dimensional IMS spectra recorded at equidistant retention time point rt_k, $k \in \{1 \ldots R\}$. Such data is represented in the mathematical matrix by exporting a measurement file (*mea*) with LAV, which adds additional information in text ($i_{textual}$) into a CSV file as simplified below.

$$i_{textual} = \begin{bmatrix} machine & \ldots & gas_{drift} & \ldots & \ldots & Timestamp \\ units & \ldots & \ldots & \text{kHz} & °\text{C} & \\ data & \ldots & Air & \ldots & serial & 2019\ldots \\ \ldots & & & blank\,line & & \ldots \\ \#specnum & RetTime[\text{s}]/Drift\,time[\text{ms}] & & \ldots & & \end{bmatrix} \quad (1)$$

$$M_{ims} = \begin{bmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & \ddots & \vdots \\ Z_{1R} & \cdots & Z_{nR} \end{bmatrix} \quad (2)$$

Exported CSV files, employed in the developed algorithm, include all available information from *mea* files and are characterized by a 5-line header containing device's mechanical and analytical textual information followed by a mathematical matrix.

3.2   Coding Language and Library

The algorithm was developed in Python, an open-source coding language, version 3.7 and, additionally, the libraries and/or functions were imported and include: *scikit-image* algorithms collection for image processing, *scipy.ndimage*, multi-dimensional image processing, *pandas* 0.25.3, Python Data Analysis Library, *mathplotlib* 3.1.1, Python 2D plotting library, *NumPy*, fundamental package for scientific computing and *operator* a standard operators as functions.

4   Results and Discussion

Developed algorithm is divided into four phases, (i) reading textual data, (ii) IMS matrix (spectra) processing, (iii) automatic peak detection and (iv) peak filtering and quantification. The algorithm uses *pandas* to read the csv file where all the textual and numerical information is contained.

(i) *Reading textual data*: this phase comprehends reading and printing of data relevant contained in the 5-line header. Useful information extracted in this phase includes, name and date of the file, machine type used, its serial number and GC column information. Textual information is printed after adding a line for the file format and origin. An example is presented below:

```
FILE ORIGIN:                              LAV conversion
FILE FORMAT:                              CSV DATA
FILE NAME:                                190410_133432_1.csv
DATE:                                     2019 04 10
TIME:                                     13 34 32
SAMPLE:
Machine type :                            BreathSpec®
Machine serial :                          3G1-00074
GC Column :                               MXT-200. 30m X 0.53mm ID
```

Supplementary information can be included in this phase, *e.g.* retention flow variation chart showing carrier gas flow changes the analysed measurement. Additional textual information can easily be tailored to each user's preference. Herein is shown only the most important information as an example.

(ii) *IMS matrix (spectra) processing*: conducted after reading the csv file, and includes RIP automated detection, without showing RIP intensity or losing any peaks' information. Such processing is performed by detecting maximum intensity values in the mathematical matrix first line. Since recording is always started before any analyte is injected into the drift tube, the first line contains solely information about intensity from the drift and carrier gas without any sample. However, RIP is contained in a drift time interval and, to detect this interval or window, a simple idea is implemented in the algorithm. RIP is defined and identified as the number of columns in the first line that contain intensity above, 0.280 V and 0.100 V prior (left) and posteriorly (right) of the RIP maximum, respectively. Values before and after the RIP are not equal, due to the drift and carrier gas humidity influence. All matrix portions previously fragmented were correctly reconstructed by *matplotlib* functions (as done by LAV).

(iii) *Automatic peak detection* is achieved by applying a module from *skimage* know as measure, *skimage.measure.find_contours (array, level)*. By finding iso-valued level of the IMS matrix (or above a threshold), clusters concerning intensity peaks were found, contoured and marked. However, since IMS spectra regularly show low intensity regions with the same intensity value, "noise", *skimage* module outputs a high number of regions which are not to be considered as peaks (Fig. 3). Hence, the algorithm was able to detect and mark matrix regions with a threshold-value above 0.150 V, set for *skimage level*, and account the total number of contours, however this value did not have a direct correspondence with the effective total number of peaks.
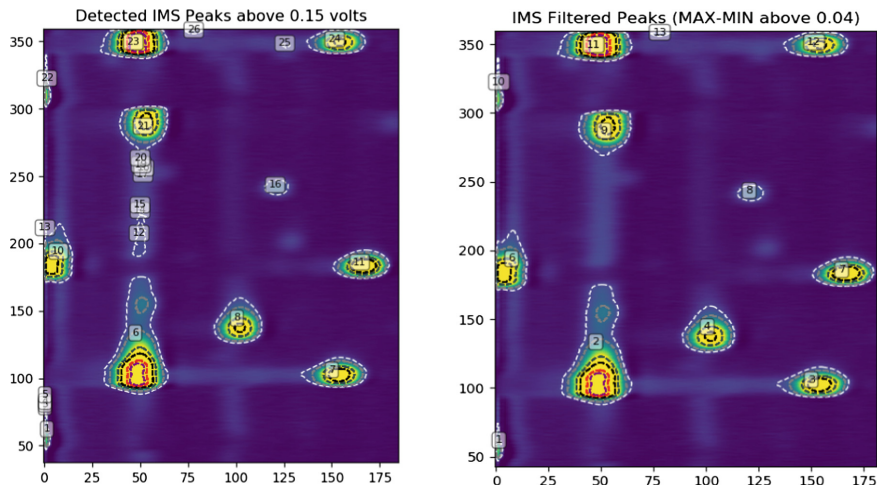
Fig. 3. Algorithm detected peaks before (left) and after using a filtering method (right) marked by dashed lines and labelled with number in a grey square. *Y*-axis represents the retention time in seconds and the *x*-axis the drift time with the RIP position as zero.

(iv) *Peak filtering and quantification*: a simple tactic was employed to filter "noise" regions (or ineffective peaks) from the peaks found by the *skimage.measure* module. Matrix maximum and minimum values were obtained for each region with the module *skimage.measure*, and if the difference between those values was lower or equal to, a defined threshold of 0.04 such region would not be classified as a peak. This threshold was defined based on observed values for general noise regions and peaks, nonetheless this value can be adjusted by the user in accordance to its study targets.

Once the noise filter was applied to the detection algorithm, results enabled the recognition of the effective (real) peaks of the spectra and allowed to estimate peak intensity. Such estimation was executed by summing all the matrix values inside the peak areas delimited previously. Furthermore, the matrix index for the maximum value of each detected peak was obtained with the intention of using it for peak identification based on a database of drift and retention times.

5   Conclusions and Future Work

The developed algorithm was able to read a csv file directly exported from the LAV software; a type of software used in ion mobility spectrometers. From csv files, the algorithm interpreted and separated text information and spectral matrix. A graphical representation was correctly performed from the reconstruction of the matrix values and peak detection was achieved by applying a *skimage* module. To reduce spectral

noise, a filter was applied resulting in the detection and isolation of relevant peaks. Furthermore, specific maximum and total intensity of peaks were found and calculated respectively.

Aiming to apply the present algorithm in all kind of IMS spectra, additional functions or tools are planned for future iterations. For instance, the deconvolution of potential overlapping peaks, a major issue in IMS spectra, is intended to be solved with the application of adjusted Gaussian functions. With this, the algorithm will be able to automatically detect and quantify all peaks which could later be cross-checked with IMS drift time libraries for compound identification.

References

1. Borsdorf, H., Eiceman, G.A.: Ion mobility spectrometry: principles and applications. Appl. Spect. Rev. **41**(4), 323–375 (2006)
2. Cumeras, R., Figueras, E., Davis, C.E., Baumbach, J.I.: Review on ion mobility spectrometry. Part 1: current instrumentation. Analyst **140**(5), 1376–1390 (2015)
3. Kanu, A.B., Hill Jr., H.H.: Ion mobility spectrometry detection for gas chromatography. J. Chromatogr. A **1177**(1), 12–27 (2008)
4. St. Louis, R.H., Hill Jr., H.H., Eiceman, G.A.: Ion mobility spectrometry in analytical chemistry. Crit. Rev. Anal. Chem. **21**(5), 321–355 (1990)
5. Gabelica, V., Marklund, E.: Fundamentals of ion mobility spectrometry. Curr. Opin. Chem. Biol. **42**, 51–59 (2018)
6. Borsdorf, H., Mayer, T., Zarejoushe, M., Eiceman, G.A.: Recent developments in ion mobility spectrometry. Appl. Spect. Rev. **46**(6), 472–521 (2011)
7. Hopfgartner, G.: Current developments in ion mobility spectrometry. Anal. Bioanal. Chem. **411**(24), 6227 (2019). https://doi.org/10.1007/s00216-019-02028-1
8. Ewing, R.G., Atkinson, D.A., Eiceman, G.A., Ewing, G.J.: A critical review of ion mobility spectrometry for the detection of explosives and explosive related compounds. Talanta **54** (3), 515–529 (2001)
9. Kirk, A.T., Allers, M., Cochems, P., Langejuergen, J., Zimmermann, S.: A compact high resolution ion mobility spectrometer for fast trace gas analysis. Analyst **138**, 5200–5207 (2013)
10. Chouinard, C.D., Wei, M.S., Beekman, C.R., Kemperman, R.H., Yost, R.A.: Ion mobility in clinical analysis: current progress and future perspectives. Clin. Chem. **62**(1), 124–133 (2016)
11. Karpas, Z.: Applications of ion mobility spectrometry (IMS) in the field of foodomics. Food Res. Int. **54**(1), 1146–1151 (2013)
12. Baumbach, J.I.: Process analysis using ion mobility spectrometry. Anal. Bioanal. Chem. **384** (5), 1059–1070 (2006). https://doi.org/10.1007/s00216-005-3397-8
13. Salthammer, T.: Organic Indoor Air Pollutants: Occurence, Measurement, Evaluation, 1st edn. Wiley-VCH, Weinheim (1999)
14. Gallart-Mateu, D., Armenta, S., de la Guardia, M.: Indoor and outdoor determination of pesticides in air by ion mobility spectrometry. Talanta **161**, 632–639 (2016)

15. Śmiełowska, M., Marć, M., Zabiegala, B.: Indoor air quality in public utility environments - a review. Environ. Sci. Pollut. Res. **24**, 11166–11176 (2017). https://doi.org/10.1007/s11356-017-8567-7
16. Vautz, W., Ruszany, V., Sielemann, S., Baumbach, J.: Sensitive ion mobility spectrometry of humid ambient air using 10.6 eV UV-IMS. Int. J. Ion Mob. Spectrom. 3–8 (2004)
17. Fetter, V., Vassilenko, V., Fernandes, J., Moukhamedieva, L., Orlov, O.: Validation of analytical instrumentation for continuous online monitoring of large spectra of VOCs in closed habitat during simulation of space fligh. In: Proceedings of 69th International Astronautical Congress, Bremen, Germany (2018)
18. Vautz, W., Franzke, J., Zampolli, S., Elmi, I., Liedtke, S.: On the potential of ion mobility spectrometry coupled to GC pre-separation - a tutorial. Anal. Chim. Acta **1024**, 52–64 (2018)
19. Espalha, C., Fernandes, J., Diniz, M., Vassilenko, V.: Fast and direct detection of biogenic amines in fish by GC-IMS technology. In: 2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG), Lisbon (2019)
20. Gonçalves, M., Fernandes, J., Fetter, V., Diniz, M., Vassilenko, V.: Novel methodology for quick detection of bacterial metabolites. In: 2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG), Lisbon (2019)