Unlocking the full potential of voltammetric data analysis : a novel peak recognition approach for (bio)analytical applications

Reference:
Van Echelpoel Robin, De Jong Mats, Daems Devin, van Espen Pierre, De Wael Karolien.- Unlocking the full potential of voltammetric data analysis : a novel peak recognition approach for (bio)analytical applications
Talanta : the international journal of pure and applied analytical chemistry - ISSN 0039-9140 - 233(2021), 122605
Full text (Publisher's DOI): https://doi.org/10.1016/J.TALANTA.2021.122605
To cite this reference: https://hdl.handle.net/10067/1794170151162165141

Unlocking the full potential of voltammetric data analysis: a novel peak recognition approach for (bio)analytical applications

Authors: Robin Van Echelpoel, Mats de Jong, Devin Daems, Piet Van Espen, Karolien De Wael*

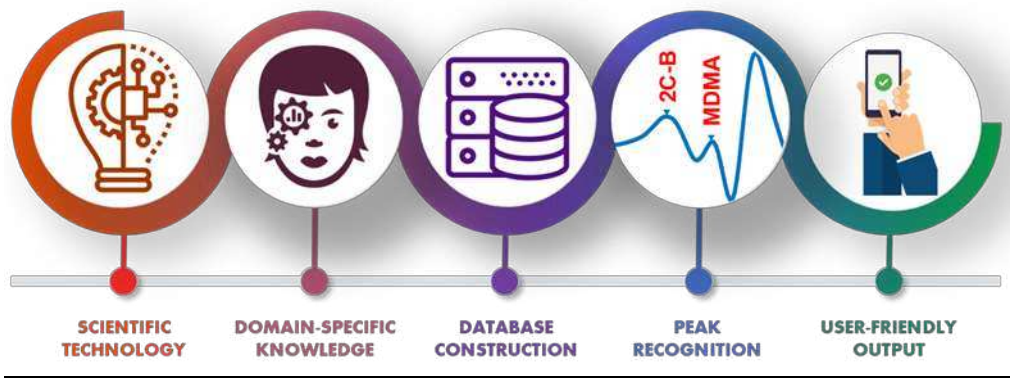
[a] Robin Van Echelpoel, Mats de Jong, Devin Daems, Piet Van Espen, Karolien De Wael
AXES Research Group, University of Antwerp Groenenborgerlaan 171, 2020 Antwerp (Belgium)
E-mail: karolien.dewael@uantwerpen.be

[b] Robin Van Echelpoel, Mats de Jong, Devin Daems, Karolien De Wael
NANOlab Centre of Excellence, University of Antwerp Groenenborgerlaan 171, 2020 Antwerp (Belgium)
E-mail: karolien.dewael@uantwerpen.be

Abstract

Bridging the gap between complex signal data output and clear interpretation by non-expert end-users is a major challenge many scientists face when converting their scientific technology into a real-life application. Currently, pattern recognition algorithms are the most frequently encountered signal data interpretation algorithms to close this gap, not in the least because of their straight-forward implementation via convenient software packages. Paradoxically, just because their implementation is so straight-forward, it becomes cumbersome to integrate the expert's domain-specific knowledge. In this work, a novel signal data interpretation approach is presented that uses this domain-specific knowledge as its fundament, thereby fully exploiting the unique expertise of the scientist. The new approach applies data preprocessing in an innovative way that transcends its usual purpose and is easy to translate into a software application. Multiple case studies illustrate the straight-forward application of the novel approach. Ultimately, the approach is highly suited for integration in various (bio)analytical applications that require interpretation of signal data.

Graphical abstract



Highlights

- Peak recognition approach using domain-specific knowledge
- Approach optimally suited for integration in (bio)analytical applications
- Data preprocessing tandem highlights hidden features in signal data
- Approach demonstrated using electrochemical sensors

Keywords

Peak recognition – signal data – voltammetry – data preprocessing – data analysis software - chemometrics

Introduction

Unraveling the valuable information hidden within complex data, and making that information understandable to non-experts, is a difficult task many scientists in (bio)analysis are confronted with when turning a scientific technology into a (bio)analytical application. Commonly, an expert in the research field is the only one capable of understanding and interpreting the complex data generated by a scientific device.[1],[2] However, if the complex output of the scientific device can be converted into a read-out comprehendible by non-experts, a major step is made towards the fulfillment of the point-of-need, paving the path for a successful application widely used.
Software is the ideal solution to bridge this gap between complex data output and non-expert, grace to its cost-, time- and labor efficiency.[3–6] Furthermore, a software program can be integrated in the large majority of scientific devices. As such, a workflow can be created where an expert develops a software program and integrates it in the scientific device, after which any end-user can use the device without needing any scientific background. As a result, the group of potential users is drastically enlarged, and in extent the valuable time of the expert can be used elsewhere.
A suiting illustration of the aforementioned is given by the research field of electrochemical (bio)sensors, more specifically voltammetric sensors. Voltammetric techniques such as linear sweep voltammetry (LSV), cyclic voltammetry (CV) and square wave voltammetry (SWV) gain both qualitative and quantitative information of an analyte by applying a varying potential to the analyte and subsequently measuring the resulting current. The voltammetric readout, usually in the form of voltammograms (current vs potential plots), can thus be considered as being an extension to a classic amperometric readout which solely considers the current resulting from a single, fixed potential. The latter has found commercial (bio)analytical applications (e.g. glucose sensor), while for voltammetric techniques this has proven more difficult due to the more complex readout.[7–9] It demands years of research to acquire the expertise to extract valuable information (qualitative and/or quantitative) from these voltammograms. The development of an effective voltammetric detection method for the illegal drug cocaine is an excellent example of the aforementioned. It required many years of extensive research to come to a highly accurate (> 98%), portable voltammetric cocaine sensor that outperforms the existing on-site identification tools.[10] The suppressing and shifting nature of certain cutting agents (e.g. levamisole, benzocaine), in particular, made it challenging to develop an optimal strategy.[11],[12] The last major obstacle this voltammetric sensor has to overcome to become the ultimate tool law enforcement needs, is a translation of all the domain-specific knowledge that has been gathered over the years into a clear-cut interpretation thereof.

A commonly used approach to perform this translation, is the use of pattern recognition algorithms.[13] These algorithms have elevated research in many domains such as image processing and computer vision to new heights and are rightfully praised.[14–16] Voltammetry is no exception to this, and pattern recognition algorithms such as linear discriminant analysis (LDA), principal component analysis (PCA), soft independent modelling of class analogy (SIMCA) and more recently machine learning (ML), are frequently encountered.[17–20] One reason pattern recognition methods have become such a success is their easy implementation in software. Programming languages, such as Python or R, have highly convenient packages that allow a user to quickly run e.g. a PCA or construct and train a ML algorithm with limited prior knowledge. This is both a blessing and a curse, with the right expertise a user can quickly test different data analysis algorithms. However, the major risk

occurs that the user no longer fully understands what happens between input and output, making the correct interpretation of the output cumbersome (so-called black box).[21–23] The cohesion between the scientific technique that generates the data, and the data analysis method that interprets that data, is in danger of being lost. Especially, if it is also taken into account that the solutions these convenient software packages offer, are very general, precisely because they have to be so widely applicable. Incorporating domain-specific knowledge into them can therefore be anything but straight-forward, while it is precisely this knowledge that makes a scientific technology so valuable and powerful. Furthermore, the performance of pattern recognition algorithms, ML in particular, is strongly intertwined with the amount of available data. A lack thereof can result in poorly trained algorithms that are prone to overfitting and have bad generalizability, which is especially dangerous in combination with a black-box approach.[24]

In this work, we propose a novel approach in which the domain-specific knowledge is the protagonist, rather than the algorithm itself. The expert's unique, subject-specific knowledge and insight in the data, is the starting point and the fundament on which the novel approach is build. The approach is developed for interpretation of voltammetric data, however it is envisioned that the scope of the approach can be extended to interpretation of signal data in other research fields. In voltammetry, the expert has generally excellent control over the different signals, i.e. the expert can commonly authenticate the origin and presence of each signal. Therefore, instead of trying to unravel patterns in the data, the individual peaks themselves will be used to extract information from the data. As such, it is assured that all the domain-specific knowledge of the expert is fully exploited, and in extent the risk of a black-box approach becomes non-existent. The bottleneck of this approach is thus the domain-specific knowledge, as opposed to e.g. the amount of available data to train an algorithm. This is far more desirable since a lack of domain-specific knowledge is a bottleneck for the scientific technology itself, and the data analysis as such does not create an additional bottleneck.

Experimental

Reagents & Solutions. d,l-methyleendioxymethamphetamine.HCl (d,l-MDMA.HCl) standard was purchased from Lipomed (Arlesheim, Switzerland). A 2,4,6-trinitrotoluene (TNT) sample was provided by the Dutch Forensic Institute (NFI, The Hague, The Netherlands). 2,5-dimethoxy-4-bromophenethylamine (2-CB) and the cocaine and MDMA street samples were obtained from the National Institute for Criminalistics and Criminology (Brussels, Belgium). Analytical grade salts of potassium chloride and potassium phosphate, as well as potassium hydroxide, were purchased from Sigma-Aldrich (Overijse, Belgium). All solutions were prepared in 18.2 MΩ cm-1 doubly deionized water (Milli-Q water systems, Merck Millipore). The pH was measured using a CyberScan 510 pH-meter from Eutech Instruments (Landsmeer, The Netherlands) connected to a HI1131 glass bodied pH electrode from Hanna Instruments (Bedfordshire, United Kingdom). Adjustment of the pH was performed using a 100 mM KOH solution.

Voltammetric measurements. Electrochemical measurements, more specifically square wave voltammetric analyses, were carried out using a PalmSens4 potentiostat with PSTrace 5.7 software (Utrecht, The Netherlands). Disposable carbon ItalSens IS-C Screen Printed Electrodes (SPE) were purchased from PalmSens (Utrecht, The Netherlands) and were used during all electrochemical measurements. The SPE's contain an internal silver pseudo reference electrode and a carbon counter electrode. All experiments were performed by applying 50 µL of solution onto the SPE. All SWV measurements were carried out with a step potential of 5 mV, amplitude of 25 mV and frequency of 10 Hz. The following buffers were used during the electrochemical measurements: 0.1M phosphate

buffer pH5, 0.1M acetate buffer pH5, 0.1M phosphate buffer pH 7 and 0.1M phosphate buffer pH 12. The measurement of the MDMA/2-CB mixture (Figure 2 – example 1) was preceded by a pretreatment (-0.8V/300s). The measurement of TNT (Figure 2 – example 2) was preceded by a 10 minute argon purge of the SPE cell to remove the influence of oxygen.
The bio-matrix samples using saliva, were prepared as follows: a stock solution of the drug in milliQ water was diluted in saliva to obtain a concentration of 500µM (pure MDMA sample) or 0.3 mg/mL (street samples). Subsequently the stock solution containing the drug sample in saliva was diluted in the buffer to obtain the final concentrations (100µM and 0.03 mg/mL respectively) for the pure sample and the street samples. A waiting time of five minutes was incorporated before starting the measurement.

Data processing. The moving average baseline correction was applied using the function integrated in the PSTrace 5.7 software. All other processing of the data was performed with Matlab R2018a (MathWorks, Natick, MA, USA) software, including its Signal Processing Toolbox™.

Moving average baseline correction. The first step in the software framework is the removal of the baseline using a moving average baseline correction. The raw voltammogram is used as an input. For every two data points, the average current value is calculated, thereby reducing the amount of data points by a factor two. Every resulting data point A_i is subsequently compared to the average A'_i of the neighboring points A_{i-1} and A_{i+1}. Then it is checked if A_i is larger than A'_i (for oxidation peaks). If this is the case, A_i is replaced by A'_i. This is repeated until no more replacements take place or until the iteration threshold of 1000 is reached. Eventually, the number of data points is extrapolated to the original number. The resulting data points represent the increasing background and are thus subtracted from the raw voltammogram. The resulting voltammogram contains more visible oxidation peaks and is thus more easily interpretable.

Top hat filter. The top-hat filter is a so-called zero-area filter that has a central window with an odd number of channels w and two side windows each v channels wide. The value of the filter coefficients (k and h_k) follows from the zero-area constraint:

$$h_k = \begin{cases} -\frac{1}{2v}, -v - \frac{w}{2} \leq k < -\frac{w}{2} \\ \frac{1}{w}, -\frac{w}{2} \leq k \leq +\frac{w}{2} \\ -\frac{1}{2v}, +\frac{w}{2} < k \leq \frac{w}{2} + v \end{cases}.$$

The filtered (i.e. transformed) electrochemical response y_i^* is then obtained by the convolution of the electrochemical response with the filter:

$$y_i^* = \sum_{k=-v-w/2}^{v+w/2} h_k y_{i+k}.$$

Peak identification. Two parameters are defined to identify the peaks: the minimum peak height and the minimum peak prominence. The first one speaks for itself, whereas the second one might require some more explanation. A marker is placed on the top of a potential peak. Subsequently, a horizontal line is drawn through this marker until (i) it crosses the signal because it encounters a higher peak or (ii) it reaches the left or right end of the signal. Then, the minimum of the signal in each of the two intervals defined in the previous step is searched. This point is either a valley or one of the signal

endpoints. The higher of the two interval minima specifies the reference level. The height of the peak above this level is its prominence. Each peak that has a value higher than the defined minimum peak height and minimum peak prominence is identified as a peak that will be processed further throughout the approach.

Results

Approach overview. An innovative (voltammetric) signal data interpretation approach is developed which uses the locations of the individual peaks in a voltammogram to identify which compounds are present in a sample. **Figure 1** illustrates the different steps of the developed approach: initially a raw voltammogram is preprocessed to enrich the electrochemical fingerprint (i.e. the unique electrochemical signal or pattern specific for a certain analyte).[25] After this preprocessing, the different peaks in the voltammogram are identified, followed by assignment of compounds to these peaks using an internal database. An exception module can be introduced in front of the peak assignment to include additional rules and requirements.

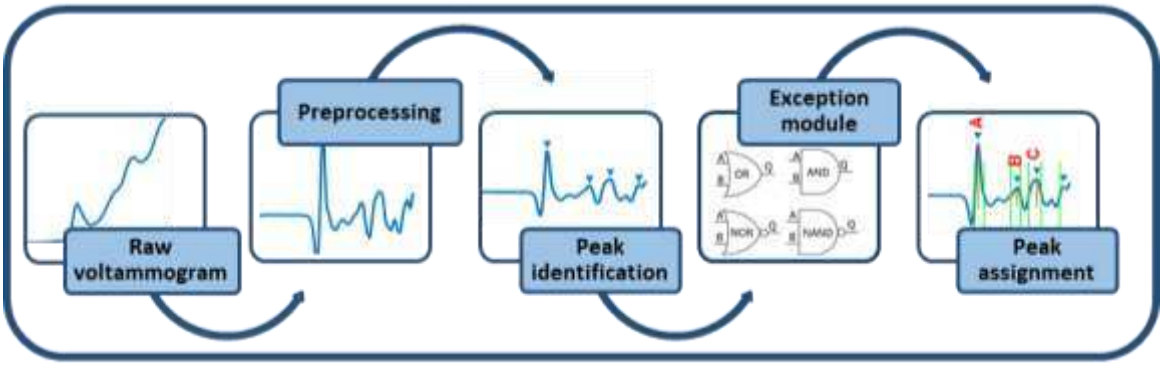


Figure 1*|Illustration of the approach: In the developed* approach*, a raw voltammogram is first modified with a baseline correction and a digital top hat filter to enrich the fingerprint and as such improve sensitivity. Then the relevant peaks are identified, and assigned to a compound using an internal database. An exception module can be introduced to incorporate additional rules prior to the peak assignment step.*

The novel approach utilizes individual peaks to assign the compounds present in the sample. It is therefore important to preprocess the raw voltammogram so that a voltammogram with distinct peaks is obtained. Preprocessing a raw signal to obtain a more easily interpretable signal is a common practice in signal processing.[26] The novel approach will however demand a more innovative use of the preprocessing that transcends its usual purpose. Solely obtaining a more easily interpretable signal is not enough, all meaningful peaks should be brought to light, even those that might initially be hidden by e.g. a shoulder. In the following steps, these meaningful peaks will be selected for further processing based on peak height and prominence. It is thus crucial that all meaningful peaks are resolved and sufficiently separated, and a good preprocessing of the raw signal therefore plays a vital role in the success of the approach.
Initially, the baseline of the raw signal is corrected by using a moving average baseline correction, increasing peak distinction in the resulting voltammogram due to the removal of the background. However, overlapping peaks are not fully resolved by a baseline correction, limiting clear interpretation (**Supporting S1**). A more sophisticated preprocessing tool, a digital filter, is thus subsequently applied to further improve peak demarcation. The general purpose of a digital filter is to smooth the data through a convolution, thereby improving the precision of the data without

distorting the signal tendency.[27] Many different filters have been developed over time, often with great success, claiming a vital position in a wide variety of real-life applications.[28],[29] For this approach, a zero-area filter, more specifically a top hat filter, was selected.[30] Besides their successful smoothing of the data, which is strived for here, zero-area filters have also proven to be beneficial for enhancing peak resolution.[31] This improved resolution is particularly useful here, as the approach is based on the identification of individual peaks. In this work, the zero-area filter of choice is the top hat filter, grace to its low computing time and promising performance in preliminary testing.

After enrichment of the raw voltammogram with a baseline correction and digital filter, the resulting peaks are evaluated on their relevance. In a further stage, compounds will be assigned to these peaks, and it is thus important to solely consider peaks that hold valuable information. Two parameters are introduced to define which peaks are interesting for further processing: (i) the minimum peak height and (ii) the minimum peak prominence. The prominence of a (signal) peak is a measure for how much the peak stands out relative to other peaks due to its intrinsic peak height and its peak location; it is akin to the concept of prominence in topography.[32] A low isolated peak can be more prominent than one that is higher but is an otherwise unremarkable member of a tall range. In the context of interpreting a voltammetric response, less prominent peaks are those which largely overlap with other peaks and/or which do not stick out considerably from the background signal.
Once the relevant peaks are identified, a compound is assigned to each of these peaks by exploiting the expert's domain-specific knowledge on voltammetric analysis or fingerprint. The voltammetric fingerprint of a compound represents the unique relationship between that compound and its specific voltammetric response, thus containing extremely valuable information.[25] Indeed, this relationship creates the perfect opportunity to link a (set of) voltammetric peak(s) to the presence of a specific compound. The approach takes full advantage of this opportunity by collecting all the voltammetric fingerprints into a database. The identified peaks are then one-by-one compared with the database, and a compound is assigned to a peak if a match is encountered.
The expert is the person par excellence to construct this database, since the latter is the sole person who has the required domain-specific knowledge. The database is one of the greatest assets of the novel approach, as it offers a general approach to incorporate subject-specific knowledge.
Depending on the application, a further processing of the identified compounds can be included. In a detection sensor for example, an alarm or warning message could be associated with the detection of a specific compound to warn the end-user about its presence.

Illustration of the approach with three case studies. **Figure 2** depicts three different applications of the developed software. In example 1, the analysis of an illicit drug mixture of 50.0% methyleendioxymethamphetamine (MDMA) and 50.0% 2,5-dimethoxy-4-bromophenethylamine (2-CB) (both commonly found in the drug ecstasy) in phosphate buffer pH 5 is shown. The baseline correction successfully removed the background, after which the top hat filter revealed two peaks that were not visible at first. This example truly demonstrates the power of the preprocessing steps, unravelling the presence of two peaks that were invisible in the raw voltammogram. The two revealed peaks are eventually correctly identified using the internal database, and assigned the appropriate tags, i.e. MDMA and 2-CB.

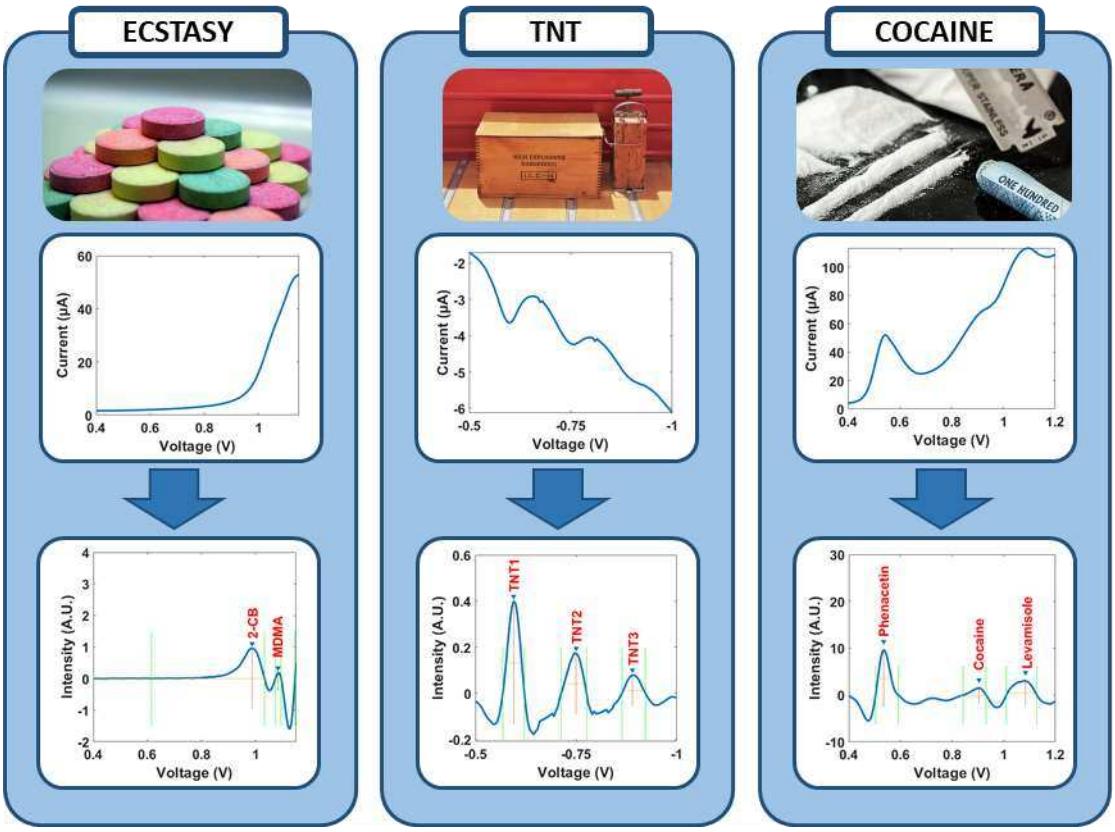


***Figure 2*|Case studies of the approach*:** *Three case studies of the approach are shown, from left to right: (i) an illicit drug mixture of 50.0% MDMA and 50.0% 2-CB in phosphate buffer pH 5, (ii) a 100.0% solution of the explosive 2,4,6-trinitrotoluene (TNT) in phosphate buffer pH 7 and (iii) a drug street sample containing 31.0% cocaine, 2.8% phenacetin, 5.7% levamisole and an unknown percentage of mannitol analyzed in phosphate buffer pH 12 buffer.*

In the second example, a pure solution of the explosive 2,4,6-trinitrotoluene (TNT) in phosphate buffer pH 7 was analyzed. Baseline removal and filtering led to a modified voltammogram with distinct peaks. The three characteristic reduction peaks of TNT were subsequently correctly identified and assigned the right tag after comparison with the corresponding database.[33] Note that the intensities were inverted since the software searches for peak maxima instead of peak minima.
The third example, a street sample of cocaine, illustrates that more complex samples are still correctly handled by the software. After analysis in phosphate buffer pH 12, the baseline correction and top hat filter converted yet again with success the raw voltammogram into a processed voltammogram without background and clear peak distinction. Cocaine was subsequently correctly assigned to the peak at 0.89 V, and in addition the cutting agents phenacetin and levamisole were correctly assigned as well.[11] Note that the feature between the peak of phenacetin and the peak of cocaine did not pass the prominence threshold, and therefore was not selected for further processing. In **Supporting S2**, all the different steps executed by the software are shown for example 3.

The three case studies highlight the true power of the two data preprocessing steps. The preprocessing of voltammetry is a hot topic in voltammetry, involving the development and application of baseline removal algorithms and digital filters.[34],[35] The task of preprocessing the data is often unfairly limited to 'cleaning' the data (reducing noise and improving signal-to-noise ratio). Here, we have proven that preprocessing the data can handle more ambitious tasks. The moving average baseline correction in combination with the top hat filter are the ideal tandem to reveal

hidden features, improve peak demarcation and as such improve sensitivity. The resulting enriched voltammogram after preprocessing can thus be an objective in itself. An expert might come to new insights as new features that were masked in the initial voltammogram are revealed in the enriched voltammogram.

Demonstration of the approach in bio-matrices. In the previous section, it was shown that the approach is highly suited to correctly identify a complex drug street sample. This result encouraged us to further explore the limits of the approach' applicability, this time by applying the latter to samples in bio-matrix. Since a database was already constructed for illicit drugs, it was decided to remain in this research area and investigate drugs samples in saliva. This bio-matrix is highly suited for illicit drug detection, especially when considering that many drugs can be found in saliva after administration.[36] The decision was made to focus on a highly relevant target molecule, i.e. MDMA, as this illicit drug is almost exclusively administered orally, and research has proven its presence in saliva after a single dose.[37]

In **Figure 3**, the approach is demonstrated on three MDMA samples in saliva. The first example depicts a pure MDMA sample (100 µM) spiked in saliva. Due to the higher complexity of the saliva, the electrochemical fingerprint becomes more complex. However, the MDMA peak around 1.04V is still visible and nicely highlighted by the preprocessing steps. The correct identification is subsequently executed by the approach. In the second example, the difficulty is raised by analyzing a MDMA street sample spiked in saliva (white crystal, 96.77%, 0.03mg/mL). The approach correctly identifies MDMA in this second example as well. The third example contains a MDMA street sample of low purity (orange tablet in form of Donald Trump, 39.46%, 0.03mg/mL). Once again, the approach correctly identifies MDMA.

The bio-matrix, i.e. saliva, didn't cause any additional issues for the approach. It is interesting to note that some features at e.g. 0.36V, 1.21V and 1.39V are recurring, probably originating from analytes in the saliva.

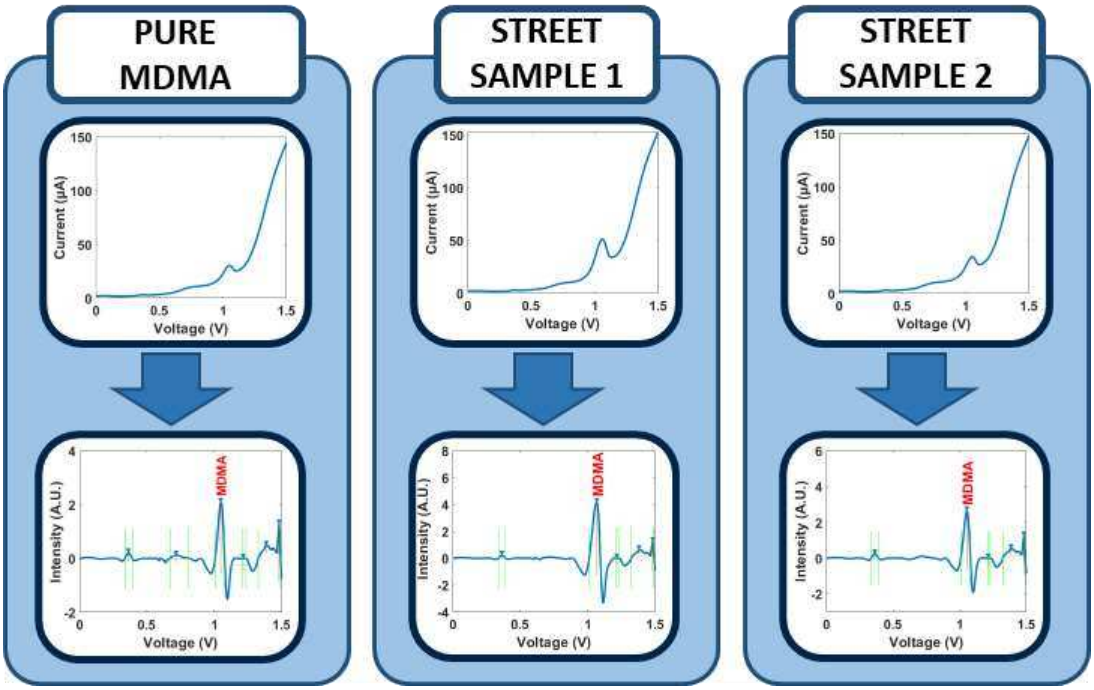


***Figure 3*|Demonstration on samples in saliva.** *Three MDMA samples (one pure, two street samples) were spiked in saliva, after which the respective SWVs were measured in an acetate pH5 buffer. Subsequent application of the approach clearly indicates for all three samples that the assets of the*

approach translate to a bio-matrix. The fingerprint itself is, as expected, more complex due to higher complexity of the bio-matrix.

Special cases. In each of the examples presented in **Figures 2 and 3**, each peak could easily be assigned to a single compound. This 1-on-1 assignment will be sufficient in most of the cases where this software is applied. However, some applications require additional rules. The approach facilitates the incorporation of such additional rules through an exception module that is placed right in front of the compound assignment (**Figure 4**). This exception module is the perfect tool to bring domain-specific knowledge into the approach in a straight-forward manner. A parallel with the concept of logic gates is made to illustrate what an additional rule looks like.[38–40] However, the additional rules that can be incorporated are not limited to this concept (**Supporting S3**).
Figure 4a illustrates how the presence of two (or more) characteristic peaks of a compound can be required when assigning a compound. An example is the voltammogram of the illicit drug ketamine in phosphate buffer pH 12.[41] This compound showcases two characteristic peaks, located at 0.90 V and 1.25 V respectively, and it can thus be required that both peaks need to be identified before ketamine is assigned to the voltammogram. The identification of only one of the peaks is thus not sufficient for assignment. Considering the TNT example in **Figure 2**, the aforementioned thus allows to incorporate an exception module that requires the presence of all three TNT peaks prior to assignment.

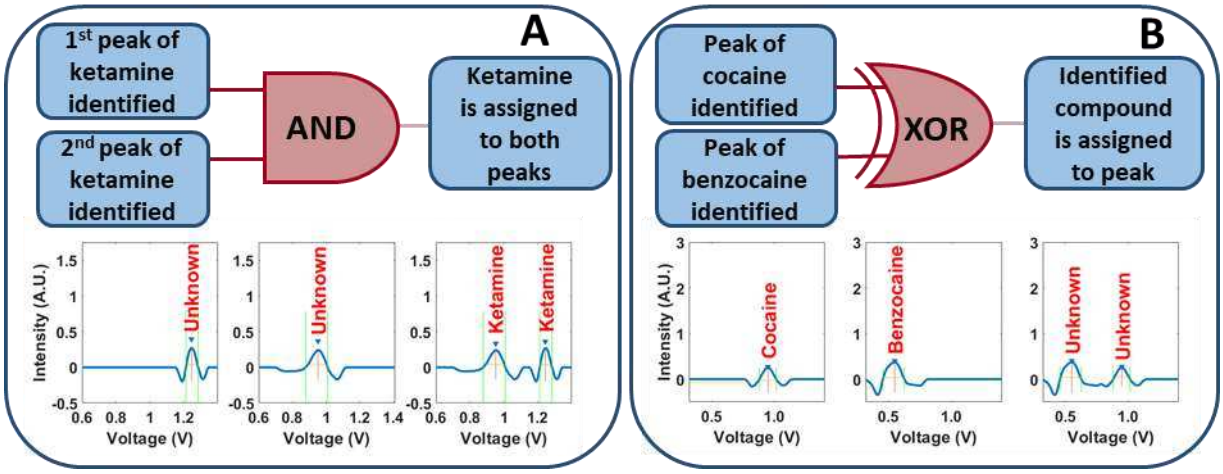


Figure 4*|Incorporating additional rules via exception module: Additional rules, to extent the scope, can be integrated in the software by means of an exception module that is placed right before the peak assignment. The first example (**A**) shows how an additional rule can be incorporated that requires the presence of both peaks of ketamine ahead of assignment. Another possibility would be a rule where two compounds can never be assigned together (**B**).*

Another possible exception is an additional rule that prevents two compounds from being assigned together (**Figure 4b**). Such an additional rule can be of great use when one of two compounds is known to cause a shift of the other one's signal. In that case it is a good safety measure to avoid their mutual assignment to prevent assignment errors. The implementation of this additional rule can be compared to an exclusive OR-gate, i.e. a true output is only given in case of an odd number of true inputs. In the example shown in **Figure 4b**, cocaine has one signal (0.82 V), as has the cutting agent benzocaine (0.41 V). Benzocaine is known to cause a shift of the signal of cocaine, and an additional rule is thus incorporated.[12] If both aforementioned signals (0.41 V and 0.82 V) are encountered together, cocaine and benzocaine will not be assigned to these peaks.

Discussion

In this work we have introduced a novel approach for the interpretation of voltammetric signal data. Instead of relying on the extraction and interpretation of complex patterns in the data, the interpretation is executed directly by comparing the individual peaks with an internal database. This database is constructed using the expert's domain-specific knowledge, thereby fully utilizing this unique expertise.

Before comparison with the database, two preprocessing steps, a baseline correction and a digital filter, are applied to the signal to remove the background and improve peak demarcation. This preprocessing tandem greatly enriches the voltammetric signal, revealing peaks that were visually hidden in the raw signal. These peaks are later on used for compound assignment and the preprocessing tandem consequently claims a key role in the approach. The increased sensitivity thus obtained is therefore of great importance in this approach, but can equally well be of exceptional value for any researcher seeking to extract hidden information from signal data.

Two parameters, minimum peak height and minimum peak prominence, are introduced to select the relevant peaks for further processing. A good parameter value choice is necessary to ensure that only those peaks are selected that hold valuable information. Values that are too lenient will result in a large pool of peaks that will further be processed, some of which might solely originate from noise.

Eventually, the selected peaks are compared with the internal database and assigned a compound if a match is encountered. The complexity of the database depends on the application and the amount of different compounds that need to be identified. The database of a detection sensor which targets a single compound could consist of that sole compound itself, whereas the database of an identification tool for e.g. waste water analysis might consist of a large list of compounds. Correct identification of samples in saliva demonstrates the approach' ability to transcend its assets into more complex bio-matrices. Additional rules, to e.g. build in restrictions or exceptions, can be included in the approach via an exception module. The database and exception module enable a very user-friendly integration of subject-specific knowledge into the approach, thereby providing quick access to tailor-made solutions. The approach allows integration of novel modules to facilitate further processing of the data. In voltammetry, amongst others research fields, the intensity of a signal can often be associated with the concentration of the corresponding analyte through means of a calibration curve. As such, a quantification module is an example of a potential module that could be introduced in the approach, in this case specifically dedicated to perform a quantitative analysis of the sample.

Furthermore, the approach is designed in such a way that a single expert can optimize the approach (choose peak identification parameter values, build database, …), translate it into a suitable software program and integrate the latter in a scientific device. Ideally, a non-expert can then use the device without prior knowledge of voltammetry or even science. The expert is only involved in the setup and integration, which frees up time to perform more analyses.

Conclusions

We propose a novel peak recognition approach, designed to interpret signal data. The approach combines a preprocessing tandem with a tailor-made database, thereby ensuring that the operating scientist can optimally integrate their expertise in the approach. In addition to the tailor-made database, an exception module grants the approach the versatility to readily be applied to different research problems. This adaptability makes the approach highly suited for integration in various scientific applications, e.g. (bio)sensors, that require interpretation of signal data. After integration by

the expert, the approach opens up the scientific application to non-expert end-users, allowing society to fully enjoy the benefits of the scientific technology.
Even though the approach was developed for interpretation of voltammetric signal data, it is envisioned that the scope can transcend the field of voltammetry and be applied in other fields of science that handle signal data such as optical (bio)sensors and applications in MedTech and life sciences.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No. 833787, BorderSens. The authors acknowledge financial support from IOF-SBO and IOF-POC from University of Antwerp, Antwerp, Belgium; and VLAIO IM [HBC.2019.2181], Brussels, Belgium.

Author contributions

K.D.W. conceived the project, obtained funding and supervised the research.
All authors contributed to the conceptualization of the novel methodology.
R.V.E., M.D.J. and P.V.E. wrote the peak recognition software with input from D.D. for the conceptualization of the exception module.
R.V.E. wrote the manuscript with input from D.D., M.D.J. and K.D.W.
All authors discussed the results and commented on the manuscript.

ORCID

Robin Van Echelpoel: 0000-0001-5513-1321
Mats de Jong: 0000-0001-8930-7962
Devin Daems: 0000-0002-8401-9141
Karolien De Wael: 0000-0003-4495-0748

References

[1] A. V Maltese, J.A. Harsh, D. Svetina, Data Visualization Literacy: Investigating Data Interpretation Along the Novice-Expert Continuum, J. Coll. Sci. Teach. 45 (2015) 84–90. http://www.jstor.org/stable/43631889.

[2] N. Zamboni, E. Fischer, U. Sauer, FiatFlux – a software for metabolic flux analysis from 13C-glucose experiments, BMC Bioinformatics. 6 (2005) 209. https://doi.org/10.1186/1471-2105-6-209.

[3] S. Buckingham, Bioinformatics: Programmed for success, Nature. 425 (2003) 209–211. https://doi.org/10.1038/425209a.

[4] A.F. Albuquerque, F. Alet, P. Corboz, P. Dayal, A. Feiguin, S. Fuchs, L. Gamper, E. Gull, S. Gürtler, A. Honecker, R. Igarashi, M. Körner, A. Kozhevnikov, A. Läuchli, S.R. Manmana, M. Matsumoto, I.P. McCulloch, F. Michel, R.M. Noack, G. Pawłowski, L. Pollet, T. Pruschke, U. Schollwöck, S. Todo, S. Trebst, M. Troyer, P. Werner, S. Wessel, The ALPS project release 1.3: Open-source software for strongly correlated systems, J. Magn. Magn. Mater. 310 (2007) 1187–1193. https://doi.org/https://doi.org/10.1016/j.jmmm.2006.10.304.

[5] W. St John, P. Johnson, The Pros and Cons of Data Analysis Software for Qualitative Research,

J. Nurs. Scholarsh. 32 (2000) 393–397. https://doi.org/10.1111/j.1547-5069.2000.00393.x.

[6] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data, Bioinformatics. 28 (2012) 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

[7] N. Elgrishi, K.J. Rountree, B.D. McCarthy, E.S. Rountree, T.T. Eisenhart, J.L. Dempsey, A Practical Beginner's Guide to Cyclic Voltammetry, J. Chem. Educ. 95 (2018) 197–206. https://doi.org/10.1021/acs.jchemed.7b00361.

[8] V. Mirceski, S. Skrzypek, L. Stojanov, Square-wave voltammetry, ChemTexts. 4 (2018) 17. https://doi.org/10.1007/s40828-018-0073-0.

[9] A.P. Periasamy, Y.-J. Chang, S.-M. Chen, Amperometric glucose sensor based on glucose oxidase immobilized on gelatin-multiwalled carbon nanotube modified glassy carbon electrode, Bioelectrochemistry. 80 (2011) 114–120. https://doi.org/https://doi.org/10.1016/j.bioelechem.2010.06.009.

[10] M. de Jong, N. Sleegers, J. Kim, F. Van Durme, N. Samyn, J. Wang, K. De Wael, Electrochemical fingerprint of street samples for fast on-site screening of cocaine in seized drug powders, Chem. Sci. 7 (2016) 2364–2370. https://doi.org/10.1039/C5SC04309C.

[11] M. de Jong, A. Florea, A.-M. de Vries, A.L.N. van Nuijs, A. Covaci, F. Van Durme, J.C. Martins, N. Samyn, K. De Wael, Levamisole: a Common Adulterant in Cocaine Street Samples Hindering Electrochemical Detection of Cocaine, Anal. Chem. 90 (2018) 5290–5297. https://doi.org/10.1021/acs.analchem.8b00204.

[12] M. de Jong, N. Sleegers, J. Schram, D. Daems, A. Florea, K. De Wael, A Benzocaine-Induced Local Near-Surface pH Effect: Influence on the Accuracy of Voltammetric Cocaine Detection, Anal. Sens. n/a (2020). https://doi.org/10.1002/anse.202000012.

[13] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 4–37. https://doi.org/10.1109/34.824819.

[14] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015) 436–444. https://doi.org/10.1038/nature14539.

[15] S. Vyas, L. Kumaranayake, Constructing socio-economic status indices: how to use principal components analysis, Health Policy Plan. 21 (2006) 459–468. https://doi.org/10.1093/heapol/czl029.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, Int. J. Comput. Vis. 115 (2015) 211–252. https://doi.org/10.1007/s11263-015-0816-y.

[17] P. Nikolaou, E. Deskoulidis, E. Topoglidis, A.T. Kakoulidou, F. Tsopelas, Application of chemometrics for detection and modeling of adulteration of fresh cow milk with reconstituted skim milk powder using voltammetric fingerpriting on a graphite/ SiO2 hybrid electrode, Talanta. 206 (2020) 120223. https://doi.org/https://doi.org/10.1016/j.talanta.2019.120223.

[18] F. Tsopelas, D. Konstantopoulos, A.T. Kakoulidou, Voltammetric fingerprinting of oils and its combination with chemometrics for the detection of extra virgin olive oil adulteration, Anal. Chim. Acta. 1015 (2018) 8–19. https://doi.org/10.1016/j.aca.2018.02.042.

[19] X. Cetó, A.M. O' Mahony, J. Wang, M. del Valle, Simultaneous identification and quantification of nitro-containing explosives by advanced chemometric data treatment of cyclic voltammetry at screen-printed electrodes, Talanta. 107 (2013) 270–276. https://doi.org/https://doi.org/10.1016/j.talanta.2012.12.042.

[20] J. Massah, K. Asefpour Vakilian, An intelligent portable biosensor for fast and accurate nitrate determination using cyclic voltammetry, Biosyst. Eng. 177 (2019) 49–58. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2018.09.007.

[21] M.T. Ribeiro, S. Singh, C. Guestrin, Why Should I Trust You?: Explaining the Predictions of Any Classifier, in: 2020: pp. 1135–1144. https://doi.org/10.1145/2939672.2939778.

[22] S. Falsaperla, H. Langer, C. Hammer, Advantages and Pitfalls of Pattern Recognition, 2020. https://doi.org/10.1016/C2016-0-00254-1.

[23] P. Riley, Three pitfalls to avoid in machine learning., Nature. 572 (2019) 27–29. https://doi.org/10.1038/d41586-019-02307-y.

[24] M.A. Pitt, I.J. Myung, When a good fit can be bad, Trends Cogn. Sci. 6 (2002) 421–425. https://doi.org/https://doi.org/10.1016/S1364-6613(02)01964-2.

[25] G. Moro, H. Barich, K. Driesen, N. Felipe Montiel, L. Neven, C. Domingues Mendonça, S. Thiruvottriyur Shanmugam, E. Daems, K. De Wael, Unlocking the full power of electrochemical fingerprinting for on-site sensing applications, Anal. Bioanal. Chem. 412 (2020) 5955–5968. https://doi.org/10.1007/s00216-020-02584-x.

[26] G. Vivó-Truyols, J.R. Torres-Lapasió, A.M. van Nederkassel, Y. Vander Heyden, D.L. Massart, Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: Part I: Peak detection, J. Chromatogr. A. 1096 (2005) 133–145. https://doi.org/https://doi.org/10.1016/j.chroma.2005.03.092.

[27] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures., Anal. Chem. 36 (1964) 1627–1639. https://doi.org/10.1021/ac60214a047.

[28] D. Acharya, A. Rani, S. Agarwal, V. Singh, Application of adaptive Savitzky–Golay filter for EEG signal processing, Perspect. Sci. 8 (2016) 677–679. https://doi.org/https://doi.org/10.1016/j.pisc.2016.06.056.

[29] Y. Li, B. Yong, H. Wu, R. An, H. Xu, An Improved Top-Hat Filter with Sloped Brim for Extracting Ground Points from Airborne Lidar Point Clouds, Remote Sens. 6 (2014) 12885–12908. https://doi.org/10.3390/rs61212885.

[30] J. Gilfrich, Handbook of X-Ray Spectrometry: Methods and Techniques Edited by: Rene E. Van Grieken and Andrzej A. Markowicz Published by Marcel Dekker, Inc., New York, 1993; xiv + 704 pp., $195, ISBN 0-8247-8483-9, X-Ray Spectrom. 23 (1994) 45–46. https://doi.org/10.1002/xrs.1300230110.

[31] F. Janssens, J.P. Francois, Evaluation of three zero-area digital filters for peak recognition and interference detection in automated spectral data analysis, Anal. Chem. 63 (1991) 320–331. https://doi.org/10.1021/ac00004a005.

[32] J. O'Driscoll, Landscape prominence: Examining the topographical position of Irish hillforts using a cumulative viewshed approach, J. Archaeol. Sci. Reports. 16 (2017) 73–89. https://doi.org/https://doi.org/10.1016/j.jasrep.2017.09.033.

[33] C.K. Chua, M. Pumera, L. Rulíšek, Reduction Pathways of 2,4,6-Trinitrotoluene: An Electrochemical and Theoretical Study, J. Phys. Chem. C. 116 (2012) 4243–4251.

https://doi.org/10.1021/jp209631x.

[34] P. Puthongkham, B.J. Venton, Recent advances in fast-scan cyclic voltammetry, Analyst. 145 (2020) 1087–1102. https://doi.org/10.1039/C9AN01925A.

[35] M. Jakubowska, Signal Processing in Electrochemistry, Electroanalysis. 23 (2011) 553–572. https://doi.org/10.1002/elan.201000465.

[36] W. Schramm, R.H. Smith, P.A. Craig, D.A. Kidwell, Drugs of Abuse in Saliva: A Review, J. Anal. Toxicol. 16 (1992) 1–9. https://doi.org/10.1093/jat/16.1.1.

[37] A.J. Barnes, K.B. Scheidweiler, E.A. Kolbrich-Spargo, D.A. Gorelick, R.S. Goodwin, M.A. Huestis, MDMA and metabolite disposition in expectorated oral fluid after controlled oral MDMA administration, Ther. Drug Monit. 33 (2011) 602–608. https://doi.org/10.1097/FTD.0b013e3182281975.

[38] J. Daintith, E. Wright, A Dictionary of Computing, Oxford University Press, 2008. https://doi.org/10.1093/acref/9780199234004.001.0001.

[39] A. Singh, Designing protein logic gates, Nat. Methods. 17 (2020) 565. https://doi.org/10.1038/s41592-020-0865-1.

[40] J. Macdonald, D. Stefanovic, M. Stojanovic, DNA Computers for Work and Play, Sci. Am. 299 (2008) 84–91. https://doi.org/10.1038/scientificamerican1108-84.

[41] J. Schram, M. Parrilla, N. Sleegers, N. Samyn, S.M. Bijvoets, M.W.J. Heerschop, A.L.N. van Nuijs, K. De Wael, Identifying electrochemical fingerprints of ketamine with voltammetry and LC-MS for its detection in seized samples, Anal. Chem. (2020). https://doi.org/10.1021/acs.analchem.0c02810.