Investigation of various factors influencing Raman spectra interpretation with the use of likelihood ratio approach

Aleksandra Michalska[a,*], Agnieszka Martyna[b], Grzegorz Zadora[a,b]

[a] Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland
[b] Department of Analytical Chemistry, Institute of Chemistry, The University of Silesia, Szkolna 9, 40-006 Katowice, Poland

ABSTRACT

The main aim of this study was to verify whether selected analytical parameters may affect solving the comparison problem of Raman spectra with the use of the likelihood ratio (LR) approach. Firstly the LR methodologies developed for Raman spectra of blue automotive paints obtained with the use of 785 nm laser source (results published by the authors previously) were implemented for good quality spectra recorded for these paints with the use of 514.5 nm laser source. For LR models construction two types of variables were used i.e. areas under selected pigments bands and coefficients derived from discrete wavelet transform procedure (DWT).

Few experiments were designed for 785 nm and 514.5 nm Raman spectra databases after constructing well performing LR models (low rates of false positive and false negative answers and acceptable results of empirical cross entropy approach).

In order to verify whether objective magnification described by its numerical aperture affects spectra interpretation, three objective magnifications —20× (N.A. = 0.4.), 50× (N.A. = 0.75) and 100× (N.A. = 0.85) within each of the applied laser sources (514.5 nm and 785 nm) were tested for a group of blue solid and metallic automotive paints having the same sets of pigments depending on the applied laser source. The findings obtained by two types of LR models indicate the importance of this parameter for solving the comparison problem of both solid and metallic automotive paints regardless of the laser source used for measuring Raman signal. Hence, the same objective magnification, preferably 50× (established based on the analysis of within- and between-samples variability and F-factor value), should be used when focusing the laser on samples during Raman measurements.

Then the influence of parameters (laser power and time of irradiation) of one of the recommended fluorescence suppression techniques, namely photobleaching, was under investigation. Analysis performed on a group of solid automotive paint samples showed that time of irradiation upon established laser power does not affect solving the comparison problem with the use of LR test. Likewise upon established time of irradiation 5% or 10% laser power could be used interchangeably without changing conclusions within this problem. However, upon the established time of irradiation changes in laser power between control and recovered sample from 5% or 10% to 50% may cause erroneous conclusions. Additionally it was also proved that prolonged irradiation of paint does not quantitatively affect pigments bands areas revealed after such a pre-treatment.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

One of the most important tasks forensic experts face is a comparison problem based on confirmation or exclusion that two samples, one found on a scene of crime (so-called recovered sample) and one delivered as a control material, stem from the same object [1]. In case of comparative analysis of automotive paints various analytical techniques are implemented including microscopy [2,3], microspectrophotometry [4,5], elemental analysis with the use of scanning electron microscopy-energy dispersive X-ray analysis or X-ray fluorescence [2] as well as infrared spectroscopy [4,6], Raman spectroscopy [7,8] and pyrolysis gas chromatography mass spectrometry [9]. From the above mentioned techniques Raman spectroscopy is usually used for

* Corresponding author.
  E-mail addresses: amichalska@ies.krakow.pl (A. Michalska),
agnieszka.martyna@chemia.us.edu.pl (A. Martyna), gzadora@ies.krakow.pl
(G. Zadora).

determining pigments composition responsible for colour and shade of the compared paints.

Typically comparison problem of paints based on their Raman spectra is solved in the visual way [6,8,10,11] by overlying spectra obtained for the recovered and control samples and detecting the similarities and discrepancies between them. The method seems to be sufficient only when compared spectra are totally different which means they exhibit different sets of pigments or diverge at least in the presence of one intensive band. Then making a decision that both paints have different origins is straightforward. The problem arises when compared spectra differ only in a single non intensive band. Then visual examination does not allow to objectively decide whether small changes observed in the compared spectra are the effect of samples inhomogeneity, possible contamination or degradation of one of the samples or arise from the methodology which was applied for paints recovery [1,12] or point out different origins of samples. What is more even in case of full conformity of spectra there is lack of reliable estimation whether observed similarities result from common origin of the samples or may be observed by chance. It should be born in mind that automotive paints are mass product and their pigment composition is not their unique feature. As a consequence, information about a rarity of the analysed attribute (e.g. pigment composition determining specific colour of automotive paint) should always be taken into account. The method which allows to include information about the rarity as well as objectively decide whether compared samples share common origin is likelihood ratio (LR) approach.

This method is based on interpretation of obtained data (here: Raman spectra) in the context of two contrasting hypotheses, i.e.: H_1 mostly associated with prosecutor, who states that compared samples stem from the same object, and alternative hypothesis, H_2, which describes defence statement about different origins of the samples. In practice this methodology is based upon establishing the conditional probabilities—$Pr(E|H_1)$ and $Pr(E|H_2)$, and evaluation of the likelihood ratio value expressed by equation:

$$LR = \frac{Pr(E|H_1)}{Pr(E|H_2)}.$$

The likelihood ratio approach allows not only for objective evaluation which hypothesis is more credible based on the evidence (here E similarity of the compared spectra) but, in contrary to visual examination, gives also the information about the strength of support for particular hypothesis, which is crucial from forensic perspective. The only inconvenience is obtaining a value equal to the threshold value which is given at one and provides indecisive conclusion (both of tested hypotheses are equally supported). In other cases LR results are interpreted as follows: LR value greater than unity indicates that evidence (E) supports prosecutor's hypothesis (H_1), while value below unity supports defence's hypothesis (H_2). Generally, the higher the value of the likelihood ratio, the stronger the support for the prosecution proposition is (H_1). Likewise, the lower the value of the likelihood ratio, the stronger the support for the defence proposition is (H_2). Apart from objective estimation of the similarity of the compared samples and information about the rarity of the determined physicochemical characteristics in the relevant population incontestable advantage of LR approach is taking into consideration possible sources of uncertainty like within-object variability related to e.g. inhomogeneity of the analysed sample as well as between-object variability, which takes into account the information about variation of measurements of characteristics between various samples in the relevant population. All mentioned factors pose significant elements which should always be considered during the evaluation process.

However, most of them is partially ignored by other statistical tests like t-test or chemometric approaches like Principal Component Analysis. All these advantages make the likelihood ratio approach recommended and well accepted measure of evidence value in the forensic sphere [1,8,13–27].

Despite its merits, LR methodology was not applied for solving the comparison problem of samples based on their Raman spectra before 2015 [23]. The reason for that relates to the fact that LR models are easily developed and constructed for data comprising of limited number of variables in comparison to the number of samples they describe. Handling datasets, such as spectral databases, containing only limited number of samples described by thousands of variables may pose serious difficulties in reliable estimation of LR models parameters like means, variances and co-variances. So for such multidimensional data like Raman spectra it is hard to imagine that any analyst will spend plenty of time analysing thousands of samples in order to reliably estimate models parameters. As a consequence the only reasonable solution seems to be smart depletion of the number of variables without losing significant information about the samples and then using these new variables in LR calculations.

Till now, authors have proposed three various approaches for reducing the dimensionality of Raman spectra recorded for blue automotive paints. In the first one each spectrum was described by new set of variables being areas under the selected pigment bands [23]. In the remaining two chemometric tools including discrete wavelet transform [20] or distance representation coupled with linear discriminant analysis [21] were applied. New sets of variables selected upon each procedure were further used for LR models construction. It was proved that either of the data dimensionality reduction methodologies LR approach enabled for solving the comparison problem of paints based on their Raman spectra. Apart from making the evaluation process more objective, it was also stated, that the levels of false positive answers (when results indicate that compared samples stem from the same object while in fact they originate from different sources) delivered by these models were definitely lower or at least comparable with those yielded during visual spectra comparison performed on the same group of samples [23]. In contrary to visual examination, in LR models the information about inhomogeneity of paint samples expressed by the level of false negative rates (when results indicate that compared samples stem from different objects while in fact they originate from the same source) could be included.

Till now the comparison problem of blue automotive paints based on Raman data and evaluated with the use of LR approach pertained only to good quality spectra obtained with the use of near infrared semiconductor laser at 785 nm. When the same paints were analysed using argon ion laser (514.5 nm) as an excitation source most of them exhibited only fluorescence in their spectra. To deal with such an inconvenience one of the recommended fluorescence suppression techniques [28,29] namely photobleaching pre-treatment was applied. The procedure was based on constant sample irradiation for a long period of time before acquiring the spectrum. The source of light was simply the Raman spectrometer's laser with the same line as that used for sample excitation. The employed methodology allowed to quench fluorescence background in 18 out of 20 paint samples revealing previously masked weaker Raman signal. Moreover, the results indicated that the process of quenching fluorescence is repeatable and durable and, what was crucial, the position and shape of pigments bands revealed after such a pre-treatment was not visually affected [30]. Therefore the question arises: could Raman spectra obtained after photobleaching process be used in solving the comparison problem with the use of LR approach? And