

# Spectroscopic methods for authentication – an overview

Peter Belton

Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA(United Kingdom). E-mail : Peter.Belton@bbsrc.ac.uk

Received 24 July 2000, accepted 8 September 2000.

An analysis by the probability shows the criterion that an analysis method has to satisfy to allow the authentication of a product. There are two distinct ways of achieving this criterion. It defines two groups for the classification of authentication methods: those in which information is concentrated into few data points and those in which information is spread across many data points. The reason for requiring authenticity testing allows to choose between these two groups. The nature of the sample will then specify the appropriate method.

**Keywords.** Authentication, choice of an authentication method, classification of authentication methods, spectroscopy, probability.

**Méthodes spectroscopiques d'authentification – une synthèse.** Une analyse par la probabilité montre quel critère une méthode d'analyse doit satisfaire pour permettre l'authentification d'un produit. Il existe deux voies possibles pour remplir ce critère. Cela induit donc une classification des méthodes d'analyse en deux catégories : celles dans lesquelles l'information est concentrée sur un petit nombre de points et celles dans lesquelles l'information est étalée sur un grand nombre de points. Lors du choix d'une méthode d'authentification, la raison du contrôle permet de choisir la catégorie la plus appropriée. La nature de l'échantillon permettra ensuite de préciser son choix dans la méthode à appliquer.

**Mots-clés.** Authentification, choix d'une méthode d'authentification, classification des méthodes d'authentification, spectroscopie, probabilité.

## 1. INTRODUCTION

There are many requirements for authentication involving a large number of different types of samples and analytical needs. At present it is very difficult to begin to make a choice as to what type of spectroscopic method might be most appropriate. The choice is usually driven by the proximity of equipment, ease of use and the purpose for which the testing is required. Only after exhaustive and comparative experimentation can a decision as to the best method be made. Whilst this paper does not pretend to offer a solution to this problem it does attempt to offer a classification of methods based not on the details of the spectroscopy but on consideration of the type of information that is offered related to general analysis of the requirements of an authentication method.

## 2. THE GENERAL PROBLEM OF AUTHENTICATION

A spectrum consists of a series of paired values: an intensity  $I$ , associated with a value  $n$  which is usually a measure of wavelength, frequency or energy. The following discussion addresses what are conventionally referred to as one dimensional spectra, but it may be

generalized to any number of dimensions and also applied any suitable data set. The problem of authentication is to obtain spectra, under identical conditions, of an authentic sample,  $a$ , and an inauthentic sample,  $i$ . The requirement is that the probability of the intensity,  $I_i$ , at some point  $n$ , given the sample is inauthentic, having the same value as in the authentic sample,  $I_a$ , is small i.e. that:

$$P(I_i = I_a) \approx 0 \quad (1)$$

In general there are  $n$  values of  $I$  in a one dimensional spectrum. For an  $M$  dimensional spectrum there will be  $nM$  values. The total probability of all data points needs to be considered. Thus it is the product of all  $n$  data points that must meet the condition as follows:

$$\prod_{j=1}^n [P(I_j = I_{a_j})] \approx 0 \quad (2)$$

where

$$\prod_{j=1}^n [P(I_j = I_{a_j})] = P(I_1 = I_{a_1}) P(I_2 = I_{a_2}) P(I_3 = I_{a_3}) \dots$$

Equation (2) is essentially the criterion that the spectra be distinguishable. There are two distinct ways of achieving condition (2). The first is to have small  $n$  and small  $P$ , thus the product is small. The other route is to have relatively large  $P$  (of course  $P$  has a maximum value of 1) and large value of  $n$ . The value of  $P$  as defined in equation (2) will depend on the spectral variability. This will depend on the signal to noise ratio of the spectrometer, the variability of sampling and loading the spectrometer and the natural variability from sample to sample. The larger the variability, the greater will be the uncertainty that the two values of  $I$  are different and thus, in the limit that the noise is of the same order as  $I_n$  and  $I_a$ , they become indistinguishable. The left hand side of equation (2) may be considered as a measure of the degree of overlap of the volume occupied by the sets of authentic and inauthentic spectra in a multi-dimensional space. In the limit that the volumes do not overlap at all

$$\prod_{j=1}^n \left[ P(I_{i_j} = I_{a_j}) \right] = 0$$

and in the limit that there is complete coincidence

$$\prod_{j=1}^n \left[ P(I_{i_j} = I_{a_j}) \right] = 1.$$

The second factor that needs to be considered is the probability that deliberate adulteration can be used to mimic an authentic spectrum. This is a much more difficult parameter to estimate but since fraud is usually the result of a desire to make money, it is reasonable to assume that the fraud will only take place where it is profitable. There are therefore two criteria that may be applied. These are:

1. the fraud is technically possible
2. the fraud is financially worthwhile.

Technical possibilities for fraud will be reduced if the intensities measured are from species that are impossible to obtain or that the cost of obtaining the species is prohibitive, some isotopic methods come into this latter class. If many intensities are measured and some that occur in the authentic samples are lower than those in inauthentic samples reduction of the levels of some species is required for fraud. This is likely to be technically very challenging. Similarly if the addition of large numbers of chemicals is required this will be costly. As technical difficulty often translates into high cost, the greater the technical challenge the more likely it is that fraud will become too expensive. The estimate of the probability of fraud requires expert knowledge which can only be properly factored into an estimate of probability by using Bayes' theorem (Malakoff, 1999).

### 3. SPECTROSCOPY, PROBABILITY AND INFORMATION

There are two ways of achieving

$$\prod_{j=1}^n \left[ P(I_{i_j} = I_{a_j}) \right] = 0$$

Type 1 approach: a small value of  $n$  but a very small chance that the respective intensities of authentic and inauthentic samples be identical whatever the value of  $j$  (from 1 to  $n$ ). i.e.  $P(I_{i_j} = I_{a_j}) \ll 1$ .

Type 2 approach:  $n$  may be large and  $P(I_{i_j} = I_{a_j})$  may be close to one for some values of  $j$  (from 1 to  $n$ ).

An example of a type 1 approach is SNIF NMR (Martin, Martin, 1995). The type 2 approach requires that there will be a large number of spectral elements. Example techniques for the type 2 approach are high resolution NMR and near and mid infrared spectroscopy.

It is interesting to compare the probability requirement that  $P(I_{i_j} = I_{a_j}) \rightarrow 0$  with the information content of the spectrum defined by

$$S = - \sum_{j=1}^n I_j \ln(I_j) \quad (3)$$

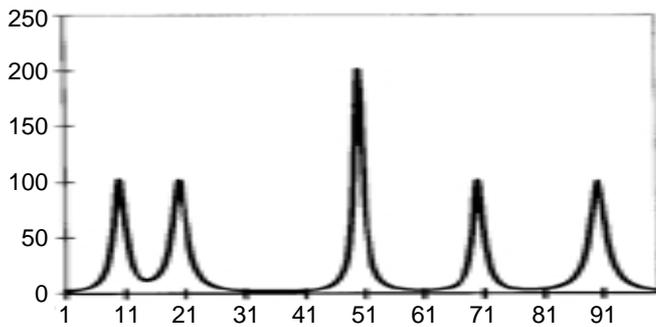
in this case  $I_j$  is the normalized intensity given by

$$I_j = \frac{I_j}{\sum_{j=1}^n I_j}$$

In some cases logarithms to other bases may be used, but this only makes a numerical difference to the value and, provided consistent definitions are used, this does not affect inter-comparison of spectra. In the case of phase encoded spectral data such as 2D FTIR or all of NMR,  $I_j$  may take negative values. This gives rise to problems with the use of equation (3) (Wright, Belton, 1986).

$S$  is the information content or entropy of the spectrum (Wright, 1996), generally small values of  $S$  arise from a few large peaks on a flat baseline and is a requirement for type 1 applications, where a few intensities and high signal to noise ratios are required. **Figure 1** is an example of a low  $S$  spectrum.

Type 2 applications require a large number of spectral elements whose intensities differ between authentic and inauthentic samples. Thus the value of  $S$  tends to be large as the spectra are typically broad and



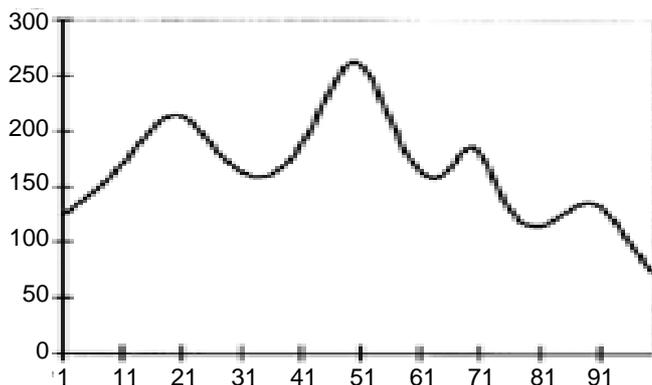
**Figure 1.** A low entropy spectrum, note that the information in the spectrum is concentrated into the relatively few peaks — *Spectre à basse entropie : l'information est concentrée seulement dans les quelques pics.*

poorly defined. For type 2 spectra sharp easily discriminated peaks are not required and thus poorly defined spectra such as are obtained from near infra red spectroscopy are valuable. **Figure 2** shows a typical high entropy spectrum.

Choosing the method of authentication for a particular application requires a consideration of the reasons for requiring authenticity testing. Some possible reasons for testing are geographic origin, processed as claimed, adulterants, intake screening of regulatory requirements

Depending on the requirements different approaches may be needed. In order to examine these it is useful to classify the methods of authentication available. This is done in **table 1**.

Typically large n,P methods are chemical survey methods whilst small n,P methods are elemental,



**Figure 2.** A high entropy spectrum, note that the relative width of peaks is greater than in the low entropy spectrum and the information is more widely spread across the spectrum — *Spectre à haute entropie : la largeur des pics est plus grande que dans un spectre à basse entropie et l'information est plus dispersée sur la largeur du spectre.*

**Table 1.** Classification of authentication methods — *Classification des méthodes d'authentification.*

Low n, small P (type 1)	Large n, large P (type 2)
SNIF	NIR
Mass spectroscopic isotope analysis	FTIR
Elemental analysis	NMR
immunoassay (e.g. Elisa)	Chromatography mass spectroscopy
	HPLC
	Some forms of DNA testing

isotopic or biospecific methods. DNA testing may be considered in both categories depending on how one chooses to define I. If the unit of intensity is the single base or the three letter code a significant fragment of DNA will fall into the type 2 category. On the other hand if the whole gene is considered as the entity, a type 1 classification is appropriate.

The choice of method will be affected by the nature of the sample. Four factors are important:

1. chemical nature of sample
2. prior knowledge
3. timescale required for results
4. physical state of sample.

Typically spectroscopic techniques are fast but there are limitations due to other factors. For methods in the type 2 category:

1. FTIR works well with spreadable solids;
2. NMR requires liquid or liquid containing samples;
3. NIR has the advantage that it is insensitive to the nature of the sample and thus has wide range of applicability;
4. chromatographic methods require extraction processes and are slower than spectroscopy.

A problem with type 1 methods is that a work up procedure is generally required.

For intake screening, it is clear that a rapid method will be needed. Thus type 2 spectroscopic methods will be the most likely to be used. Cost considerations will certainly come into play here. It may be that the optimal solution will involve a cheap, fast method to identify suspect intake and more expensive method, located at some central point may be used to check the findings. For regulatory requirements a method defensible and comprehensible in a court of law is required. This points to a type 1 method where differences can be easily seen and understood. Generally for legal requirements the measurement step is not rate determining.

In order to guide the choice of methods **table 2** offers a résumé of spectroscopic methods available together with some comments and classification.

**Table 2.** Spectroscopic methods available: classification and comments — *Méthodes spectroscopiques disponibles : classification et commentaires.*

Spectral region	Type of information	Comments	Classification
X-ray, far UV, gamma-ray	elemental, nuclear analysis		Type 1
Near UV, visible	valence electron transitions	absorption mode limited in applications	Type 2
Near UV, visible, NIR	valence electron virtual transitions	Raman spectra for chemical survey	Type 2
NIR	vibrational combinations, overtones	powerful chemical survey method, limited bonds containing H atoms	Type 1
Mid IR	vibrational modes, fundamentals	powerful, unlimited chemically, complemented by Raman data	Type 2
Far IR, microwave	low frequency vibrations, rotational modes	limited chemical information	Type 2
Microwave, radio frequency radio frequency	coherence magnetic methods	NMR powerful survey method, isotope analysis, ESR limited	Type 1, type 2

#### 4. CONCLUSIONS

Authentication methods can be divided into those in which information is concentrated into a relatively few data points and those in which information is spread across many data points. In the former case the probability that the intensity of a data point containing information does not occur in the authentic and inauthentic sample must be high. In the latter case this probability may be lower because there are many more data points to be compared and thus the product of the probabilities may still remain low.

#### Bibliography

- Malakoff D. (1999). Bayes offers a new way to make sense of numbers. *Science* **286**, p. 1460–1464.
- Martin GJ., Martin ML. (1995). Stable isotope analysis of food and beverages by nuclear magnetic resonance. *Annu. Rep. NMR Spectrosc.* **31**, p. 81–103.
- Wright KM., Belton PS. (1986). A new definition of the information content of NMR spectra suitable for use in maximum entropy signal processing. *Mol. Phys.* **58**, p. 485–495.
- Wright KM. (1996). Maximum entropy methods in NMR data processing. In Rutledge D. (ed.). *Signal and signal analysis in NMR*. Amsterdam: Elsevier.

(4 ref.)