ELSEVIER

Journal of Chromatography A, 845 (1999) 3–11

JOURNAL OF CHROMATOGRAPHY A

Influence of capillary column efficiency on the reliability of the identification of complex mixtures by gas chromatography

G. Stoev

Central Laboratory of Veterinary Control and Ecology, Iskarsko Shosse Blvd. 5, 1528 Sofia, Bulgaria

Abstract

When identifying complex mixtures by gas chromatography the chromatograms of the samples are represented as maps with n and m strips along the abscissa (retention time) and the ordinate (signal intensity), respectively. The probability P of two samples with a number of characteristic peaks, k, being identical, when in fact they are not, can be determined by the equation:

$$P_{m,n}^{k} = \frac{(n-k)!k!}{mn!}$$

The probability P is a measure of certainty of the identification of the objects by their characteristic compounds and gives an opportunity for quantitative assessment of the selectivity of the chromatographic method. Increasing the efficiency of the capillary column renders a more reliable identification. Doubling of the number of characteristic peaks as a result of higher efficiency increases the reliability by several orders.

Keywords: Efficiency; Capillary columns; Milk; Oils; Gasoline; Peak overlap

1. Introduction

A great deal of pattern recognition algorithms exist for distinguishing investigated subjects by groups, types, clusters etc. However, in some cases applying this approach is not sufficient. For example, considering environmental analysis, it is not enough only to determine the type of oil contaminating the sea; but rather it would be more valuable if the analysis indicated the source, i.e., the contaminating tanker, that had polluted the sea. In forensic science it is not so important to say that an accident had been committed by a car painted with an alkyd- or nitrocellulose paint; it is necessary to specify which car among many others with the same paint had caused the accident. Because some ''invisible'' compounds determine taste, flavor, origin, etc., customers are interested in obtaining the essential oil, fragrance and beverage, with the same concentrations of these compounds. In these cases it is important to determine the probability that two samples would seem identical when in reality they are not. We could be confident that above this threshold the samples are different.

For three decades polymers were identified by pyrolysis–gas chromatography (GC). The pyrograms were compared as fingerprints by the relative concentration and relative retention time of their characteristic peaks: monomers, dimers or some others [1]. Later the application of capillary columns facilitated the distinguishing of polymers and natural products and an attempt for a qualitative assessment of the reliability of the identification [2].

Over the last decade several statistical theories

0021-9673/99/$ – see front matter
PII: S0021-9673(99)00342-8

have been developed to describe overlap in both simple and complex one-dimensional separations [3–10]. The statistical model of overlap [4,8,9] is among the simplest of these and it gives an opportunity for more precisely quantitative assessment of the identification of complex mixtures and determination of the selectivity of the analytical methods.

When identifying complex mixtures we use only the characteristic peaks (specific pollutants, specific aroma compounds, etc.) from the chromatograms of multi-compound samples. The other compounds of the analyzed objects are not important at the first step of identification. For example, when we want to know if the water is contaminated by gasoline, diesel oil or crude oil, we can use *n*-hydrocarbons as characteristic peaks. If a higher level of identification is required, e.g., for the purpose of distinguishing crude oils, then pristine, phytane, sulfur compounds, etc., will be used as characteristic peaks. In all cases the characteristic peaks must be separated from many others in the complex mixtures. Because the number of separated peaks strongly depends on the efficiency of the column this paper deals with the influence of the efficiency of the capillary column on the reliability of the identification of complex mixtures and their quantitative assessment.

2. Theory

The characteristic peaks used as a pattern of the analyzed sample must be fully separated from the other compounds in the sample. The peak capacity, n_c expresses the efficiency of the separating system [3] because it shows how many components can be separated in theory within a certain *k*-range as peaks of resolution $R=1.0$:

$$n_c = 1 + \frac{\sqrt{N}}{4}\ln(1 + k_{max}) \quad (1)$$

Here N is the number of theoretical plates of the separating system and k_{max} is the maximal acceptable value of the retention factor. Using Eq. (1) we deal with the peaks that are statistically distributed over the accessible time range. For example, if a separating system has a dead time of 1 min and the efficiency of the capillary column is 75 000 theoretical plates, the number of peaks of separation $R=1$ that can be separated over 30 min period will be:

$$n_c = 1 + \frac{\sqrt{75\,000}}{4}\ln\left(1 + \frac{30-1}{1}\right) = 234$$

Peak capacity of packed columns is small because their efficiency is low. The value of n_c is 39 at 3000 theoretical plates and $k_{max}=15$.

Eq. (1) determines the number of separated compounds at isothermal separation where the peak width increases roughly in proportion with the retention time. If the chromatogram of a system with temperature programming is homogeneous over the range from the dead time, t_R^0, to the maximum practical retention time, t_R^{max}, the separation number (SN) of the column represents the number of fully separated peaks and gives another possibility of expression of the peak capacity, n_c.

$$SN = \frac{t_{R(Cn+1)} - t_{R(Cn)}}{w_{0.5(Cn+1)} - w_{0.5(Cn)}} - 1 \quad (2)$$

Here $t_{R(Cn)}$ and $t_{R(Cn+1)}$ are the retention times of two homologues of hydrocarbons or fatty methyl esters with n and $n+1$ number of carbon atoms and $w_{0.5(Cn)}$ and $w_{0.5(Cn+1)}$ are their widths at the half the peak height. The sum of the separation numbers (ΣSN) represents the number of the separated compounds, n_c. It depends on the working temperature range and the selectivity of the liquid phase and the efficiency of the column.

The probability P^q that an analyte of the sample mixture with a number of compounds q will be eluted as a single compound peak (SCP) and will not overlap with other compounds of the mixture [4] is:

$$P^q \approx e^{-2q/n_c} \quad (3)$$

For example, when a system with a peak capacity $n_c=234$ separates a 10-component sample, $q=10$, the probability of sufficient resolution of the analyte from the neighboring peaks is 0.92. The probability that this peak consists of two, three or more components is 0.08 (1.00−0.92) [4]. The probability that the peak does not represent the analyte is 0.08, (8%). This value is above the accepted threshold value of the uncertainty of 5% at 95% coincidence level.

For a 10-component mixture Eq. (5) determines

that a 95% probability will be achieved at a peak capacity $n_c = 390$. It corresponds to a retention factor of $2.26 \cdot 10^3$ with a capillary column of $N = 7.5 \cdot 10^4$ or a retention factor of $3.6 \cdot 10^9$ with a packed column of $N = 5000$. If we carry out analysis at $k_{max} = 10$ then the column must possesses an efficiency $N = 420 \cdot 10^3$ ($k_{max} = 10$). These estimations are valid when the density of SCPs is constant throughout the chromatogram. The statistical model of overlap [4] was modified by Davis [8,9]. The significance of these works is the overcoming of the restriction that SCPs must be distributed with constant density throughout the separation [4], allowing their number to be determined more exactly. The improved method is more precise but the procedure of calculation is difficult and time consuming.

In real samples, where the peaks are of unequal size, the situation is worse [5]. The situation is more complicated when the sample possesses several analytes (characteristic peaks). The probability that all components of the sample will be separated, P^{all}, is markedly low [6]

$$P^{all} = \left(1 - \frac{q-1}{n_c - 1}\right)^{q-2} \quad (4)$$

especially when the number of compounds is above 20. For example, P^{all} is only 0.73 at $n_c = 234$ and $q = 10$. These tremendous values of efficiency and capacity factor are difficult to achieve in practice and another approach for distinguishing of the sample should be used.

The limited informing power of one-dimensional (1D) chromatographic systems and the desire to resolve increasingly complex mixtures have provided an incentive for the development of two-dimensional (2D) and multi-dimensional systems for higher total peak capacity or greater separation space. The theoretical peak capacity, n_{2D}, of the 2D separation [7,11] is a product of the peak capacities obtained on each of the two dimensions:

$$n_{2D} = n_{c,1} \times n_{c,2} \quad (5)$$

where $n_{c,1}$ and $n_{c,2}$ are the peak capacity obtained in the first and second dimension, respectively. However, in reality the actual peak capacity is somewhat less than the theoretical capacity because a truly orthogonal separation is seldom obtained [11–14].

For comparison of the subjects by their characteristic compounds let us represent the characteristic peaks of one sample in a 2D map (Fig. 1). The characteristic peaks are strictly arranged in the map by their retention times and areas. We refer to this specific arrangement as a ''fingerprint'' of the analyzed sample (identified object). The overlapping of the pattern of a ''standard'' object with the pattern of the analyzed object means that this subject is identical with the standard sample with respect to the characteristic compounds. The number of fully separated peaks determines the number of strips, *n*, along the abscissa (retention time). The number of strips, *m*, along the ordinate (signal) is determined by the sensitivity of the detector.

If we have only one characteristic peak in a multi-component sample the possibility of this characteristic peak with its retention time and intensity of the signal to be in any quadrant of the map is:

$$P = \frac{1}{nm} \quad (6)$$

The following combinations of *n* elements from *k*th class exist towards the abscissa for a *k* number of characteristic peaks, with progressively increasing retention times (Fig. 1):

$$C_n^k = \frac{n!}{k!(n-k)!} \quad (7)$$

Towards the ordinate the combination will consist of *m* elements from the 1st class:

$$C_m^1 = \frac{m!}{1!(m-1)!} = m \quad (8)$$

The number of all arrangements at *k* number characteristic peaks and *n* number of separated strips will be:

$$C_n^k C_m^1 = \frac{n!}{k!(n-k!)} \cdot m \quad (9)$$

The probability to result in one and the same arrangement in the map of these *k* points will be:

$$P = \frac{1}{C_n^k C_m^1} = \frac{(n-k)!k!}{mn!} \quad (10)$$

This probability is accepted as a threshold value

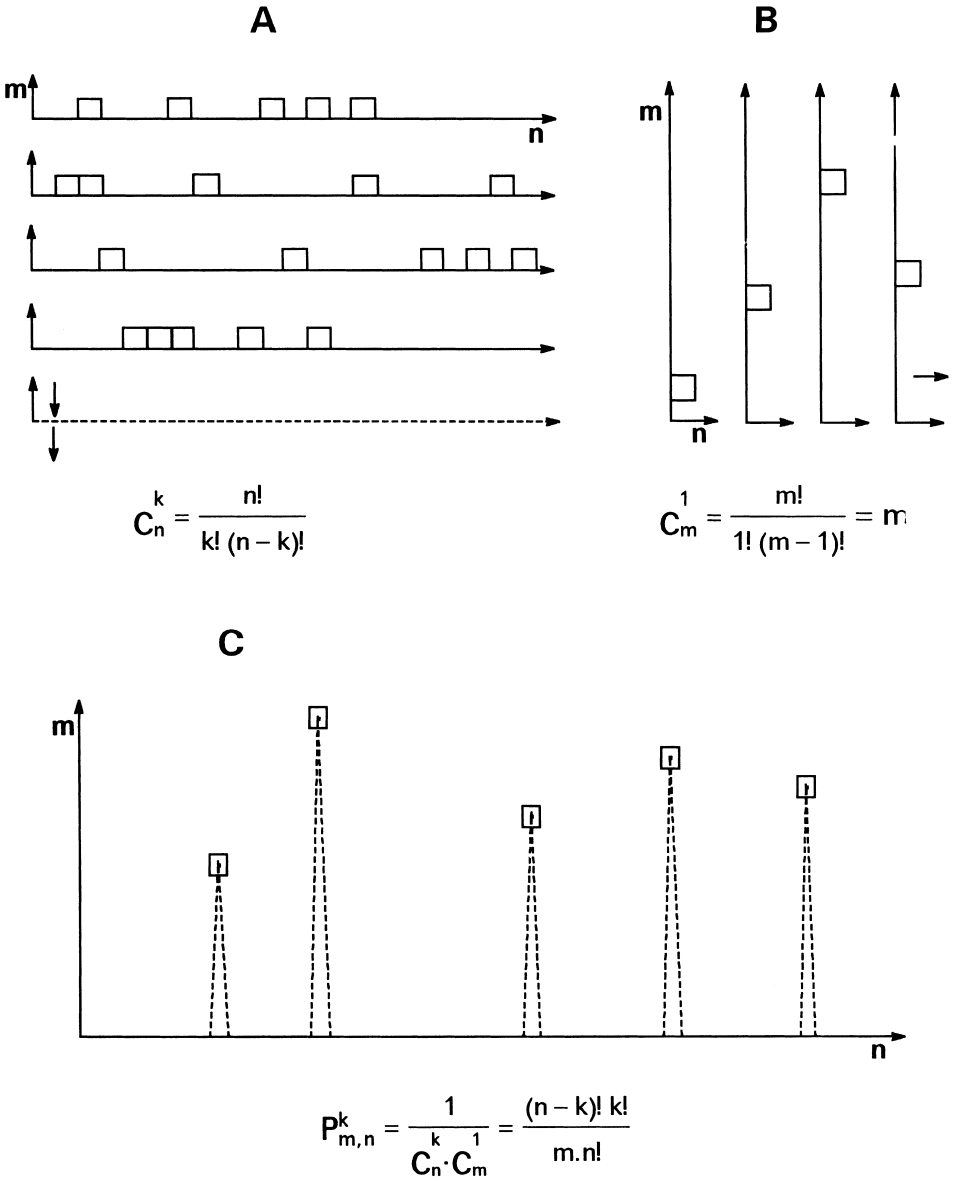

Fig. 1. Combinations of n elements from kth class toward the abscissa (A), combinations of m elements from 1st class toward the ordinate (B) and all arrangements at k number of separated strips (C).

for distinguishing of the subjects. Below this value the objects will be not distinguished, nevertheless some of them are different.

In order to test if two numbers (retention times or detector signals) are different considering their individual standard deviations, a two-sided t-test can be performed. Usually a 95% confidence level is used and the uncertainty, α, is $(1-95)/100=0.05$. In applying the two-sided t-test for the comparison of two means, $\bar{X}_A$ and $\bar{X}_B$, for five runs and looking up the t-test table, the value of the difference, u, is then evaluated from:

$$u = 1.03\sqrt{\sigma_A^2 + \sigma_B^2} \tag{11}$$

Here σ_A^2 and σ_B^2 are the variance in each of the values of compound A and compound B being compared, respectively. If the absolute value of the difference in the means, $|\bar{X}_A - \bar{X}_B|$, is greater than u, then the values are considered different at 95% confidential level and if $|\bar{X}_A - \bar{X}_B|$ is less than u, then $\bar{X}_A$ and $\bar{X}_B$ are not distinguishable.

The relative peak heights on modern GC instruments are reproducible to within 1 or 2%. Using Eq. (11) at $\sigma = 0.01$ the m value of 69 is determined. However, the ''history'' of the sample – origin of the sample, conservation and sample preparation, strongly influence the reproducibility of the signal amplitude. Deviations in the concentration of the characteristic compounds, especially at the biological samples must be taken into account. The use of the relative ratios of the detector signals determined by a characteristic compound, usually with the highest concentration, decrease the influence of these factors.

Since the relative retention times, determined again by one of the characteristic compounds, are very precise (~0.03–0.2%) many of the compounds can be distinguished simply by relative retention time. If we determine the number of strips along the abscissa by this reproducibility and Eq. (11), we will obtain a very high value of n. However, we can use this value for identification by Eq. (10) because we do not know if these strips represent single compound peaks.

3. Experimental

Glass capillary columns (30 m × 0.25 mm, 0.2 μm film thickness) with the stationary phases: OV-101, DEGS, Silar 10C and permethylated β-cyclodextrin (PM-β-CD) have been prepared. Packed columns (2.1 m × 2 mm with 3% OV-101, DEGS or Silar 10C on Chromosorb W 100–120 mesh) were prepared for comparison. Nitrogen was used as a carrier gas with a linear velocity of approximately 30 cm/s.

The liquid phases OV-101, DEGS and Silar 10 C were purchased from Applied Science, Netherlands, and PM-β-CD was synthesized [15].

The columns were tested with a Grob test mixture and the SN values were determined by fatty methyl esters (FAMEs) E11/E12. Columns having SN > 25, were used for the experiments. The ΣSN value was determined with a suitable mixture of *n*-hydrocarbons or FAMEs depending on the retention times of the characteristic compounds of the analyzed subjects.

Grob test mixture, *n*-hydrocarbons, FAMEs and standards α-, β-pinene, limonene, borneol, etc., were purchased from Fluka and Sigma.

Varian Star GC 3400 and Sichromat-2 gas chromatographs were used in the experiments.

Real samples from our practice – hydrocarbon air pollutants, fatty methyl esters, essential oils, etc., were used for this investigation.

4. Results and discussion

Air samples polluted with gasoline (A) and an air sample spiked with 50 μg/l gasoline from a suspected gasoline tank (cistern) are shown in Fig. 2. These samples have been analyzed with a capillary column having a ΣSN value of 186, determined by *n*-C_6–*n*-C_{14} mixture and a packed column with a ΣSN value of 36. The number of peaks with a signal/noise ratio ($S/N > 6$) was 32. The probability that all compounds of sample will be separated statistically with the capillary or packed columns was determined by Eq. (4) as $4 \cdot 10^{-3}$ and $5 \cdot 10^{-29}$. Obviously, it is practically impossible for all compounds to be separated even with a capillary column with high efficiency. The probabilities that every one of the characteristic compounds of gasoline: benzene, toluene, ethylbenzene and xylene, would be separated by a capillary or a packed column [Eq. (3)], were 0.71 and 0.24, respectively. The uncertainty of distinguishing every characteristic peak on the capillary column (0.29) is much less than the uncertainty for the packed column (0.76).

A ''fingerprint'' comparison of the samples was done by calculating the relative area of the characteristic peaks of benzene, toluene (reference peak), ethylbenzene and xylene ($k = 4$), using Eq. (10). The reproducibility of flame ionization detection towards these compounds was 0.4%, but the reproducibility of the recovery of the hydrocarbons at the preconcentration was 13%, and Eq. (11) gave an m value of 5 ($u = 0.19$; $m = 100\%/19\% \approx 5$).

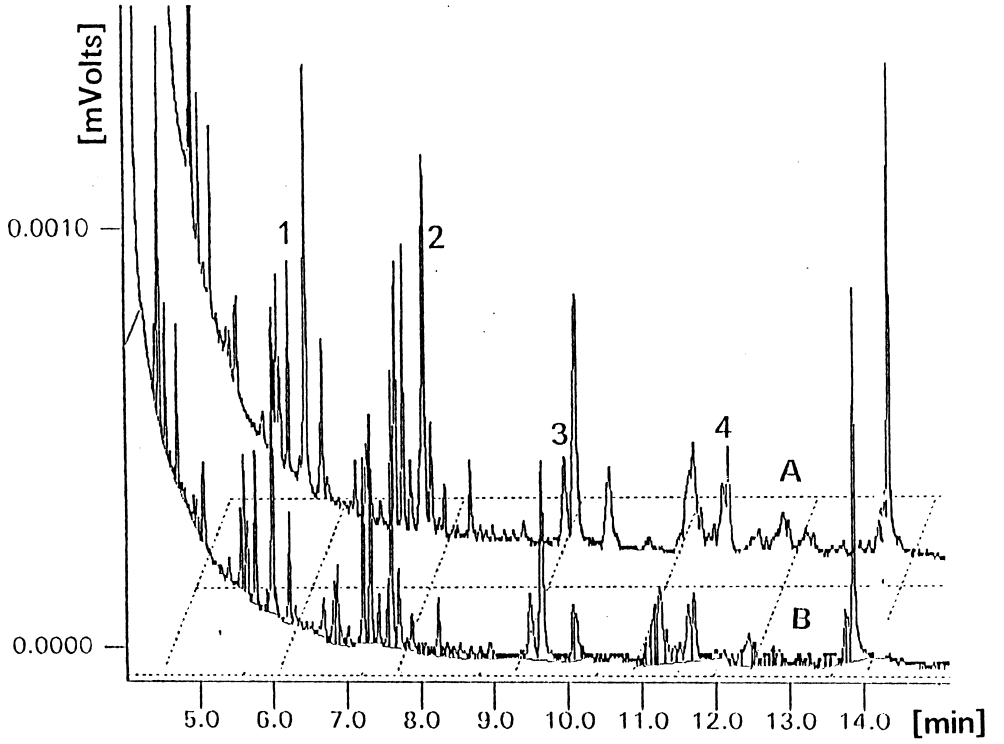

Fig. 2. Chromatograms of 1-l air samples polluted with gasoline (A) and chromatogram of gasoline from a suspected cistern (B). Peaks: 1=benzene, 2=toluene, 3=ethylbenzene and 4=xylene.

$$P^4_{5,186} = \frac{(186-4)!4!}{5\times 186} = 4.0\cdot 10^{-9} \text{ (capillary column)}$$

$$P^4_{5,36} = \frac{(36-4)!4!}{5\times 36!} = 4.4\cdot 10^{-6} \text{ (packed column)}$$

Taking into account the uncertainties of separating the characteristic peaks, the probability of incidental overlaying of the compared chromatograms is $4.0\cdot 10^{-9}\cdot 0.29 = 1.2\cdot 10^{-9}$ for the capillary column and $4.4\cdot 10^{-6}\cdot 0.76 = 3.3\cdot 10^{-6}$ for the packed column. A comparison of these values shows that the identification of samples by their characteristic compounds using a capillary column is approximately $1.1\cdot 10^3$ more reliable compared to the packed column ($4.4\cdot 10^{-6}/4.0\cdot 10^{-9} = 1.1\cdot 10^3$).

The selectivity of the liquid phase is important at a higher level of identification. Isomers or enantiomers are used as identifying signs in this case and Silar 10C, OV-275, or chiral stationary phases are useful for this purpose. For the differentiation of milk samples by their FAMEs we initially used columns with DEGS as stationary phase. Reproducibility of the recovery of the peaks of FAMEs at sample preparation (extraction, derivatization, etc.) was 8.3% and by Eq. (11) an m value of 8 was determined. Seven FAMEs were used as characteristic compounds: $C_{12:0}$, $C_{14:0}$, $C_{16:0}$(reference peak), $C_{16:1}$ $C_{18:1}$, $C_{20:0}$ and $C_{20:1}$. Twenty-three peaks, having $S/N>6$, were detected by the capillary column possessing a ΣSN value of 208. The probability that each characteristic peak will be separated is: $P^{23} = e^{-46/208} = 0.80$. The DEGS packed column that was used with a ΣSN value of 39 distinguished each of the characteristic peaks with a probability $P^{23} = e^{-46/39} = 0.31$ and an uncertainty of 0.69.

Incidental overlaying of the chromatograms obtained with a capillary column is

$$P^{7}_{5,208} = \frac{(208-7)!7!}{8\times 208!} = 4.1\cdot 10^{-14}\cdot 0.20 = 8.3\cdot 10^{-15},$$

while this probability for the packed column is

$$P^{7}_{5,39} = \frac{(39-7)!7!}{8\times 39!} = 8.8\cdot 10^{-8}\cdot 0.69 = 6.1\cdot 10^{-9}$$

and as a result the distinguishing of the sample by the capillary column is $7.3\cdot 10^{6}$-times more reliable.

For fine distinguishing of milk samples by FAMEs an OV-275 capillary column was used, which separated the *cis*/*trans* $C_{18:1}$ isomers (Fig. 3). *cis*/*trans* $C_{18:1}$ isomers were eluted between $C_{8:0}$ and $C_{20:0}$ with an SN value of 33. Since the number of peaks between $C_{18:0}$ and $C_{20:0}$ is 7 they will be separated with a certainty $P^{7}=e^{-14/33}=0.65$. The probability for distinguishing of the samples only by *cis*/*trans* $C_{18:1}$ is:

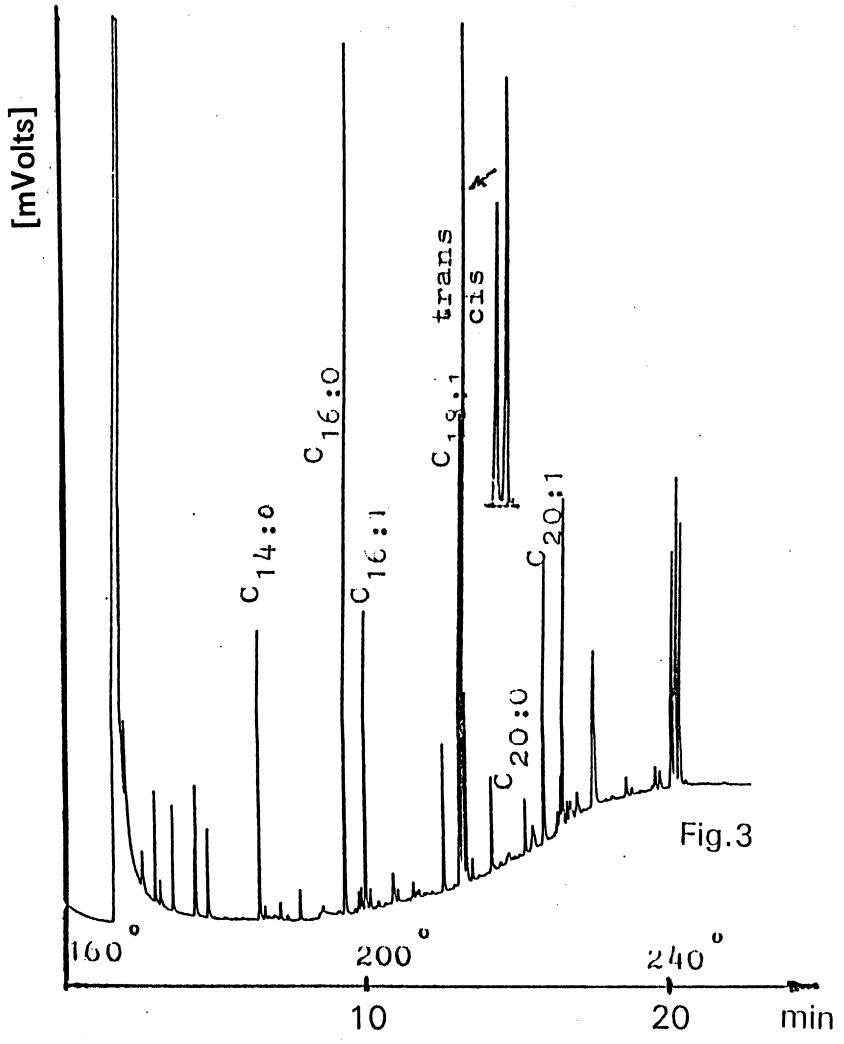

Fig. 3. *cis*/*trans* $C_{18:1}$ isomers used as a characteristic peaks for distinguishing milk samples.

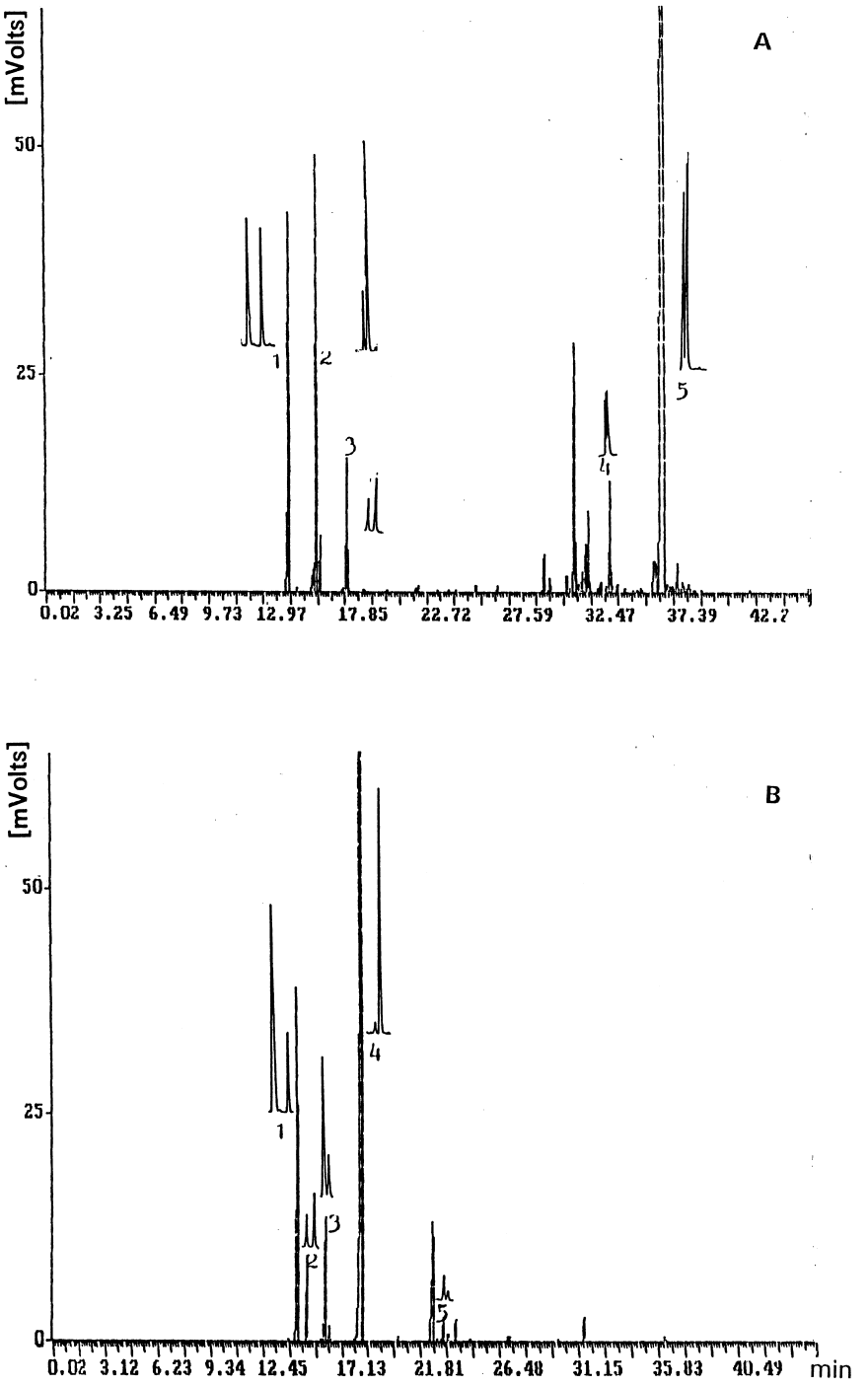

Fig. 4. Enantiomers of caroten oil (A) and rosemary oil (B). Peaks (A): 1=α-pinene, 2=β-pinene, 3=limonen, 4=cineol, 5=rotol; (B) 1=α-pinene, 2=camphene 3=β-pinene, 4=cineol, 5=borneol.

$$P^{2}_{5,33} = \frac{(33-2)!2!}{8\times 33!} = 2.4\cdot 10^{-5}\cdot 0.65$$
$$= 1.5\cdot 10^{-5}\ (n=33, m=8, k=2).$$

Separation of *cis*/*trans* $C_{18:1}$ was impossible by packed column and for this reason distinguishing of the samples was impossible.

The enantiomers of α-pinene, camphene, β-pinene (reference peak), cineol and borneol ($k=5\cdot 2=10$) in adulterated rosemary oil (A) and caroten oil (B) with synthetic compounds were separated by multi-dimensional capillary GC with a PEG 20M column (first) and PM-β-CD, respectively, (Fig. 4A and B). In this case the practical peak capacity is near the theoretical peak capacity [Eq. (5)] because the separation mechanisms are different. Separation of DEGS is based on dipole–dipole and Van der Vaals interactions while the enantioseparation of β-cyclodextrin is due of the inclusion phenomena, (orthogonality). The peak capacities of the DEGS and permethylated β-CD columns were 208 and 166, respectively and therefore the total peak capacity was accepted as a product of these values $n_{ct}=208\cdot 166=34\ 528$. 142 compounds ($S/N>6$) were detected in the chromatogram of rosemary oil and the probability of every one of them to be separated was: $P^{142}=e^{-284/34528}=$ 0.99. This means practically every compound of rosemary oil is fully separated and the reliability of identification is very high. In our case, where only five pairs of enantiomers were used, (reproducibility of the peak area of 6.9% and $m=10$), the probability of incidental overlaying of the chromatograms of two samples with different origins was:

$$P^{10}_{10,35\ 528} = \frac{(34\ 528-10)!10!}{10\times 34\ 528!}\rightarrow 0,$$

which was extremely small. By using a 2D separation system we were able to distinguish the geographic origin of the rosemary and the rose oils and to prevent them from adulteration. This example shows that the multi-dimensional separation system is able to distinguish the characteristic analytes as single compound peaks without applying of the improved but more complicated method of statistical model of overlap [8,9].

Many of the chemometric methods concerning the optimization of the separation gave opportunities to improve the distinguishing of the characteristic compounds. For example, Beens has reported a prediction of the comprehensive 2D GC separation of analytes with different concentration in multi-compound mixtures [16]. Separation of overlapped peaks among many others is achieved easy and quickly by varying of the temperature programming of the column [17].

5. Conclusion

An algorithm for the quantitative estimation of the reliability of distinguishing multi-component samples is proposed. The reliability of the identification depends on the number of the compounds in the identified samples, the purpose of the identification, the number of the suitable characteristic compounds, and especially the efficiency of the separation system. Multi-dimensional GC is a powerful technique for identification of very complex subjects.

References

[1] V.R. Alishoev, V.G. Berezkin, A.A. Korolev, I.A. Tugorskii, Zh. Anal. Chim. 22 (1967) 151.
[2] G. Stoev, Ph.D.Thesis, University of Chemical Technology, Sofia, 1979
[3] E. Grushka, Anal. Chem. 42 (1970) 1142.
[4] J.M. Davis, J.C. Giddings, Anal. Chem. 55 (1983) 418.
[5] M.Z. Fallon, M. Martin, J. Chromatogr. 557 (1991) 23.
[6] M. Martin, D.P. Herman, G. Guiochon, Anal. Chem. 58 (1986) 2200.
[7] J.C. Giddings, Anal. Chem. 56 (1984) 1258A.
[8] J.M. Davis, Anal. Chem. 66 (1994) 735.
[9] J.M. Davis, Chromatographia 44 (1997) 81.
[10] A. Felinger, L. Pasti, F. Dondi, Anal. Chem. 63 (1991) 2627.
[11] Z. Liu, D.G. Patterson Jr., Anal. Chem. 67 (1995) 3840.
[12] J.M. Davis, Anal. Chem. 63 (1991) 2141.
[13] P.J. Slonacker, X. Li, Th.H. Ridway, J.G. Dorcey, Anal. Chem. 68 (1996) 682.
[14] F. Cuesta Sanchez, J. Toft, B. van den Bogaert, D.L. Massart, Anal. Chem. 68 (1996) 79.
[15] V. Schurig, H.P. Novotny, J. Chromatogr. 441 (1988) 155.
[16] J. Beens, Ph.D.Thesis, Faculty of Chemistry, University of Amsterdam, 1998, p. 119
[17] DryLab software, LC Resources, Orinda, CA, 1998