TECHNICAL NOTE

C. Sten L. Jonson,[1] *M.S. and Lars Strömberg,*[1] *Ph.D.*

Computer Aided Retrieval of Common-Batch Members in Leuckart Amphetamine Profiling

REFERENCE: Jonson, C. S. L. and Strömberg, L., **"Computer Aided Retrieval of Common-Batch Members in Leuckart Amphetamine Profiling,"** *Journal of Forensic Sciences*, JFSCA, Vol. 38, No. 6, November 1993, pp. 1472–1477.

ABSTRACT: Comparison of profiles is a well established way to find links between confiscated drugs. It is a laborious and time consuming task to manually compare large numbers of profiles to find common-batch links. To facilitate the comparison a computerized method has been developed. It is described and applied to a set of amphetamine impurity profiles. From each profile, areas of selected peaks are fed to the computer. By using quotients of corresponding peaks, the computer finds pairs of closely related profiles. With a sufficient number of peaks, the method is tolerant to variations in intensity between profiles, random peak area variations and a few strongly deviating peak areas. The program was written in Q-basic from Microsoft and may be run on any IBM-compatible personal computer. The method may also be used for analyzing data from other forensic objects, when the descriptors chosen are affected by errors like those described in the text.

KEYWORDS: forensic science, drug profiling, batch links, profile screening, computer aid

Common-batch links between confiscated Leuckart amphetamine items can be established by gas chromatographic profiling. The evidential value of such links is well recognized [*1*]. Those links are also of interest from the point of view of drugs intelligence, because they show how parts of a batch are spread over a country or different countries.

For court purposes, only a few profiles within a case, and already thought to be in some relation to each other, are compared but seldom profiles from different cases. With the introduction of modern technique, the stability of the analytical system has reached a level permitting the retrospective comparison of profiles. It is a waste of information not to look for all possible links, but then, a large number of profiles will have to be pairwise compared. To do that manually would take a lot of time but it is easily done with the help of a computer. Selected peak areas from the profiles are then fed into the computer and compared by some algorithm. An overview of such algorithms is given by Bratchell [*2*]. Janzen et al. used Euclidean distances in the comparison of cocaine profiles [*3*]. Most of the algorithms, however, have the drawback of being sensitive to

Received for publication 7 Dec. 1992; revised manuscript received 15 March 1993; accepted for publication 3 May 1993.

[1]Research Associate and Associate Professor, respectively, SKL National Laboratory of Forensic Science, Linköping, Sweden.

'errors' encountered in drug profiles from contaminated or missing peaks. With the present method, we have tried to overcome those problems.

Method Description

The profiles are approximated with selected peak areas that are then compared. If two samples come from the same batch, the profiles are similar and all the selected peak areas will show the same ratio. The computer excludes pairs of digitized profiles having different peak area ratios but reports those that may originate from the same batch. The original profiles may then be examined more closely. By this operation we don't have to examine a lot of obviously different pairs and tedious manual work is then avoided.

In such a system, profile variation between common-batch members may cause false exclusions. The variations may be of different kinds as described in the following.

When a batch of Leuckhart amphetamine is distributed and parts of it are confiscated at different times and places, the corresponding profiles undergo different kinds of changes. The profile intensity may change because of dilution or adulteration. Profiles of diluted or adulterated amphetamine may give peaks interfering with the selected peaks. Contamination, from plasticizers in polyethylene bags, for example, may also interfere with the selected peaks. Under extreme storage conditions, some peaks may change due to evaporation or chemical instability.

Profiles are also affected by the errors of the analytical procedure. Firstly, all peak areas are impaired by random errors and varying peak shapes may cause integration errors. Secondly, detector response varies over time, which affects profile intensity like dilution. Another problem arises when small peaks vary around the peak detection level. Then, while a peak just beyond the level will give the proper value, a peak just below it will be set to zero.

The variations can be summarized in three factors:

1. All peak areas in one profile differ from those of another profile by the same factor.
2. Random variation in peak areas.
3. Some peak areas deviate strongly.

These factors are commonly encountered and must be taken into consideration in order to not lose too many links.

From two digitized profiles, X and Y, the n corresponding peak areas are called $x_1, x_2, \ldots, x_i, \ldots, x_n$ and $y_1, y_2, \ldots, y_1, \ldots, y_n$ respectively. The quotients $q_i = x_i/y_i$ are calculated. Then, for every quotient q_i, the $n - 1$ distances, r_{ik}, to the other quotients are computed according to

$$r_{ik} = \text{abs}(q_i - q_k)/(q_i + q_k) \text{ for } k = 1, 2, \ldots n\ ;\ k \neq i$$

For each quotient the number of quotients with $r_{ik} < r_{max}$ is calculated where r_{max} is a preset value. The maximum of these numbers is denoted N and the comparison between the two profiles is said to be N quotients within r_{max}. If N is less than a preset value N_{min} the two profiles are regarded not to be a match. By choosing r_{max} and N_{min} properly, possibly matching pairs are reported.

The method is designed to deal with all three different kinds of variation described above. The use of quotients between corresponding peak areas will eliminate the variation of profile intensity and by choosing the proper values of r_{max} and N_{min} the random variation in peak areas and the problem with some strongly deviating peak areas will be overcome. Another important quality of this algorithm is that $r_{ik} = r_{ki}$, that is, it doesn't matter if we compare the X profile with the Y profile or in the reverse order.

Two other problems encountered are peak areas below the peak detection level and

TABLE 1—*Number of pairs retrieved for different r_{max} and N (not N_{min}) when comparing 31 profiles from one amphetamine sample analyzed over a period of 6 months. Very low N-values give no reports because of the similarity of the profiles. The total number of reports, however, will depend on the preset N_{min}. Thus, 59 + 208 + 132 pairs are reported at r_{max} = 12% and N_{min} = 12. To retrieve all 465 pairs, different settings are possible e.g. r_{max} = 8% and N_{min} = 6 or r_{max} = 20% and N_{min} = 12.*

Number of quotients N	r_{max} = 8%	r_{max} = 12%	r_{max} = 16%	r_{max} = 20%
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	2	0	0	0
7	10	0	0	0
8	25	5	0	0
9	32	1	0	0
10	51	29	1	0
11	59	31	5	0
12	85	59	28	2
13	133	208	176	146
14	68	132	255	317

missing peak area values. In neither case we know the true value but we must decide whether the variable involved should contribute to the number of true quotients or not.

Peak areas below the peak detection level will be reported as zeros by the integrator. In the case with two zeroes the true peak area values are both very small and the program is made to always accept this as a quotient contributing to the number *N*. Quotients involving one zero may be set to be either accepted or rejected as contributing.

In some cases, when the integrator has obviously failed in reporting a correct peak area, for example, due to contamination, "−99" is put into the computer instead of the

TABLE 2—*Peak areas from the 4 profiles in Figs. 1a-b and 2a-b. Quotients are calculated for the 2 matching pairs 1a/1b and 2a/2b and for the pair 1a/2a. The quotients that don't fit in the comparison are underlined. In the comparison between 1a and 2a only 3 quotients fit within 12%.*

Peak	Peak areas 1a	Peak areas 1b	Quotients 1a/1b	Peak areas 2a	Peak areas 2b	Quotients 2a/2b	Quotients 1a/2a
1	4466.1	4602.7	0.97	1286.2	2252.6	0.57	3.47
2	1091.3	1249.0	0.87	116.2	223.5	0.52	9.39
3	254.4	234.0	1.09	437.3	445.3	0.98	0.58
4	−99	24649.7	−99	64796.0	66012.1	0.98	−99
5	58.7	51.5	1.14	239.4	231.6	1.03	0.25
6	421.8	481.7	0.88	53.2	45.2	1.18	7.92
7	77.1	55.2	1.40	93.1	86.9	1.07	0.83
8	67.1	64.6	1.04	161.0	165.2	0.97	0.42
9	26.1	26.2	1.00	76.2	77.1	0.99	0.34
10	62.5	58.2	1.07	85.0	84.8	1.00	0.74
11	122.6	113.2	1.08	724.6	724.7	1.00	0.17
12	29.6	28.9	1.02	86.7	85.9	1.01	0.34
13	1319.0	1363.9	0.97	1118.7	1312.4	0.85	1.18
14	752.2	752.1	1.00	685.8	786.7	0.87	1.10

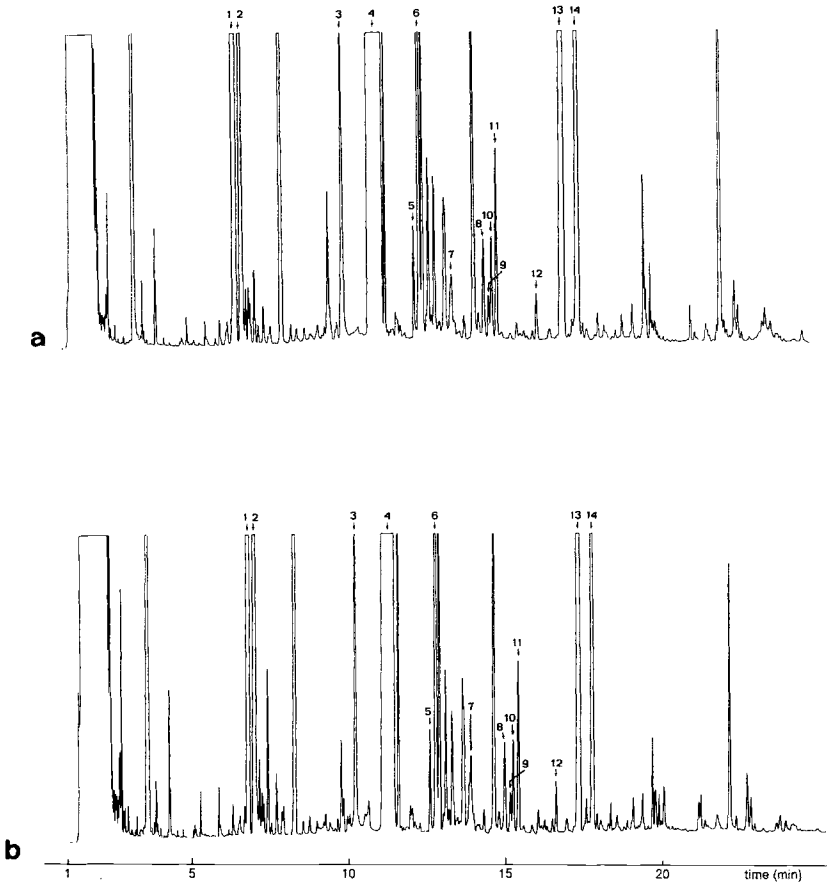

FIG. 1*a-b—The first pair of matching profiles. The peaks used for profile digitizing are indicated by arrows and numbered. The long arrow in 1b indicates the contaminated peak 7, giving rise to the non-matching quotient of 1.40 (underlined in the table).*

peak area. Quotients involving −99 can be chosen to be accepted or rejected independent of how zeroes are handled as described above. Profiles containing too many −99:s to be of forensic interest can be excluded from the comparison by pre-setting a maximum number.

The computer program was written in Q-basic from Microsoft and may be run on any IBM-compatible personal computer. The digitized profiles are stored on the computer disk in order of case numbers and when running the program it is possible to compare any sequence of profiles with themselves or with those of another sequence. Several such runs can be chained to perform extensive searches. Each comparison takes about 0.02 seconds (calculated on the run described below which was performed on a Copam model 486B/33, the disk having an average time of access of 19 ms).

Practical Example

The run that will be described here was done on a set of 345 authentical profiles. Fourteen peaks were used, all originating from the Leuckart synthesis. The selection of

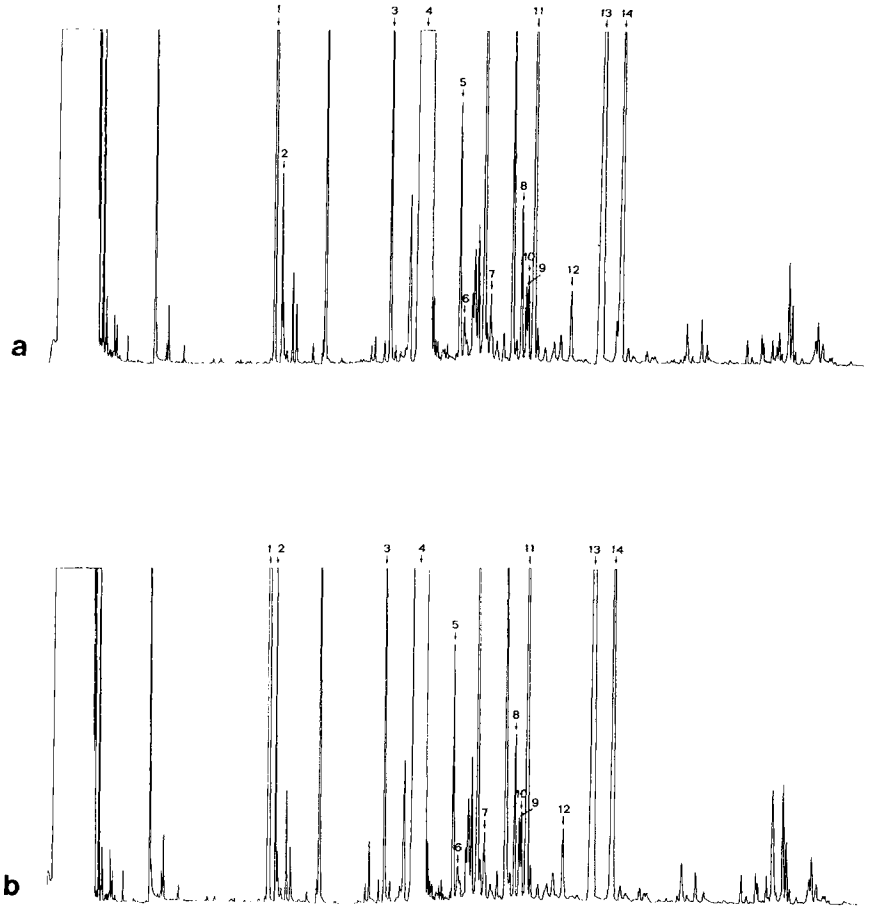

FIG. 2*a-b*—*The second pair of matching profiles. The deviating quotients for peaks 1 and 2 is probably caused by evaporation losses in object 2a.*

those was grounded on experience from case work and controlled syntheses and regarding factors such as abundance and chromatographic performance.

To get an idea of the settings of the parameters r_{max} and N_{min}, 31 profiles from one amphetamine sample analyzed over a period of 6 months were compared. Four different levels of r_{max}, 8%, 12%, 16% and 20% were chosen, accepting zeroes but rejecting −99:s as explained above. The maximum number of −99:s was set to 1. The numbers of reported pairs under different conditions are listed in Table 1.

In the search for similar pairs among the 345 profiles the r_{max} value was set to 12% whereas N_{min} was set to 12. Zeroes and −99:s were treated as above. Using these settings for the repetition series, 399 pairs would have been retrieved which corresponds to 86%.

The number of outputs within 12% was 140 which is 0.24% of the total number of compared pairs (59 340). Of these, 113 were "trivial," that is, matching pairs from case-to-case comparisons. In these cases, the average number of quotients within 12% was 12.7. The remaining 27 outputs revealed new links of great interest for drugs intelligation. Here, the corresponding number of quotients was 12.2. This difference is expected, because non-trivial common-batch members have travelled along different paths through the illegal distribution system. Moreover, long-term changes of the GC-system perfor-

mance may also render the matches worse as the analyses are carried out at different times.

Peak areas and quotients from two of the non-trivial pairs are listed in Table 2.

The profiles in each pair chosen have similar total intensity which facilitates the comparison for the reader. This gives quotient values around unity. Quotients that don't fit in the comparison are underlined. The profiles are shown in Figs. 1*a* and *b* and 2*a* and *b*. In the first pair, one quotient differs from the others depending on the contaminated peak in profile 1*b*. Together with the missing value for peak 4 replaced by −99 there will be two quotients beyond the limit. In the second pair, 2 quotients differ from the others probably depending on evaporation in object 2a. Despite the deviating quotients, both the profile matches are very good. An example of a non-matching pair is 1*a* and 2*a* where the quotients differ from 0.17 to 9.39. Only 3 quotients can be found within the preset limit of 12%.

Final Remarks

The method may also be used for analyzing data from other forensic objects, when the descriptors chosen are affected by errors like those described. The following example of another application is given. Amphetamine synthesized according to the Alles route (electrolytic reduction of phenyl nitropropene) will give by-products typical for that route. By properly selecting among those an analogous method will easily be set up for the retrieval of common-batch links among them.

As far as the choice of r_{max} and N_{min} in a new application is concerned, this depends on what the computer output is used for. With "hard" settings as in the example above, few pairs are reported but the links will be strong, because of the high quality of the matches. However, the price for this may be that some true links are lost. With "softer" settings, the number of pairs reported increases rapidly, but the links are then weaker although they may be useful for drugs intelligence purposes. In the example given above, 99.8% of all pairs were excluded but 27 new links were found.

As a consequence of the drug trafficking, a batch of production is often spread over more than one country. Therefore, international common-batch links are of interest for drugs intelligation. By using the method described such links may be established from a combined data base, built up by exchanging peak data from different laboratories. A prerequisite for this is that the profiling methods used are harmonized.

References

[*1*] Sanger, D. G., Humphreys, I. J., Patel, A. C., Japp, M., and Osborne, R. G. L., "The Significance of Gas Chromatographic Impurity Patterns Obtained from Illicitly Produced Amphetamine," *Forensic Science International*, Vol. 28, 1985, pp. 7–17.

[*2*] Bratchell, N., "Cluster Analysis," *Chemometrics and Intelligent Laboratory Systems*, Vol. 6, 1989, pp. 105–125.

[*3*] Janzen, K. E., Walter, L., and Fernando, A. R., "Comparison Analysis of Illicit Cocaine Samples," *Journal of Forensic Sciences*, Vol. 37, No. 2, March 1992, pp. 436–445.

Address requests for reprints or additional information to
Sten Jonson
SKL National Laboratory of Forensic Science
S-581 94 Linköping
Sweden