Journal of Chromatography B 1092 (2018) 379–385

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/jchromb

Human odor and forensics: Towards Bayesian suspect identification using GC × GC–MS characterization of hand odor

Vincent Cuzuel[a,c], Roman Leconte[b], Guillaume Cognon[a], Didier Thiébaut[c], Jérôme Vial[c], Charles Sauleau[a], Isabelle Rivals[b,*]

[a] Institut de Recherche Criminelle de la Gendarmerie Nationale, Caserne Lange, 5 boulevard de l'Hautil, BP 20312 Pontoise, 95037 Cergy, Pontoise CEDEX, France
[b] Equipe de Statistique Appliquée, ESPCI Paris, PSL Research University, INSERM, UMRS 1158, Neurophysiologie Respiratoire Expérimentale et Clinique, 10 rue Vauquelin, Paris, France
[c] UMR CBI 8231, CNRS, Laboratoire Sciences Analytiques Bioanalytiques et Miniaturisation, ESPCI Paris, PSL Research University, 10 rue Vauquelin, Paris, France

ARTICLE INFO

Keywords:
Bayesian inference
Comprehensive two-dimensional gas chromatography
Forensics
Human hand odor
Mass spectrometry

ABSTRACT

A new method for identifying people by their odor is proposed. In this approach, subjects are characterized by a GC × GC–MS chromatogram of a sample of their hand odor. The method is based on the definition of a distance between odor chromatograms and the application of Bayesian hypothesis testing. Using a calibration panel of subjects for whom several odor chromatograms are available, the densities of the distance between chromatograms of the same person, and between chromatograms of different persons are estimated. Given the distance between a reference and a query chromatogram, the Bayesian framework provides an estimate of the probability that the corresponding two odor samples come from the same person. We tested the method on a panel that is fully independent from the calibration panel, with promising results for forensic applications.

1. Introduction

Forensic profiling of human hand odor using analytical devices and statistical tools is of particular interest to confirm the information provided by dogs in courts of law. Indeed, the identification dogs can perform remains challenged because of their inability to testify. For some years, several teams have tried to get a better understanding of the human odor and have aimed at characterizing it using analytical devices. The results and conclusions of these studies on odor sampling, sample analysis and data processing, three steps that are essential to a successful odor identification strategy, were recently reviewed [1].

The sampling procedure can either be performed directly, i.e. with contact, or indirectly, i.e. without contact: direct sampling necessitates putting an adsorbent phase on the skin of the subject in order to collect odor compounds [2], whereas indirect sampling is based on air suction around the object of interest [3]. In both cases, appropriate protocols can be implemented to reduce the contamination by exogenous substances. Most of the time, they consist in prewashing the sampling zone with tap water or perfume-free soap and subsequent "natural" drying. Subjects can be asked to use special soap and deodorant or no deodorant at all up to a week before the study [4,5]. Some authors asked their subjects not to eat specific foods (spices, garlic) or to do some exercise before sampling in order to increase the production of sweat [5,6]. To our knowledge, there is no validated, let alone standardized protocol yet.

The analytical separation is often performed by gas chromatography (GC) coupled with mass spectrometry (MS) [7,8]. Still, the information collected might not be sufficient, especially when hundreds of compounds need to be monitored within short-time analysis. Thus, the use of two-dimensional gas chromatography (GC × GC) is a relevant alternative and was already shown to provide more information than classic GC [9].

Finally, the statistical processing is essential to extract the information relevant for identification. To our knowledge, only few studies were carried out on a large panel, one of the largest being that collected by Colon-Crespo et al. involving 105 subjects, 54 women and 51 men belonging to 3 different ethnic groups (Caucasian, Hispanic and east Asian), with ages in the range 18–77 [2]. In their study, Curran et al. [7] worked with a panel of 60 people, 30 men and 30 women, but since only one sample (hands) per person was collected, the comparison of samples from the same individual was not possible. Brown et al. [8] worked with a more reduced panel (20 people), but sampled each subject 3 times from different body locations. However, though samples were taken in triplicate, average profiles were used, so that the degree of similarity between samples from the same person was again not evaluated. Gallagher et al. [5] sampled the backs and arms of 25 people

* Corresponding author.
E-mail address: isabelle.rivals@espci.fr (I. Rivals).

https://doi.org/10.1016/j.jchromb.2018.06.018
Received 12 April 2018; Received in revised form 6 June 2018; Accepted 8 June 2018
Available online 15 June 2018

twice to study the significance of the effects of three factors, sampling area (back/arm), sex (male/female) and age (young/old): significant differences were found for certain compounds across age and site of collection, but not between men and women. The study conducted by Penn et al. [4] is more thorough, their panel including 197 individuals who were sampled five times each, once every fortnight, over a ten-week period. The results of this study suggest that using pattern recognition on the entire profile pattern, rather than on a small set of compounds, could be essential to perform proper identification.

The aim of the present paper is to propose an attempt, not at classifying people into categories by gender and/or ethnicity, not at attributing chromatograms to a finite set of persons for which numerous odor samples are available, but at identification. For this purpose, we gathered an important and diversified panel of subjects sampled several times, and adopted a Bayesian approach, which does not suffer from several drawbacks of the frequentist approach. The first main drawback is the asymmetry between the null and alternative hypotheses, since evidence can essentially be collected against the null hypothesis and not in favor of it, and the second one is that very small *p*-values can be obtained despite weak evidence against the null hypothesis [10,11]. For these reasons, Barcaru and Vivò-Truyols used Bayesian hypothesis testing to spot differences between the GC × GC–MS chromatograms of pairs of diesel samples [12]. Bayesian approaches are successfully used in many fields of forensics, especially to assess the probative value of DNA [13,14]. Though there are large differences between DNA and odor profiles, the ultimate question is essentially the same: given a reference profile (the one collected on the crime scene), does the query profile (that of the suspect) correspond to the same person?

The proposed procedure can be outlined as follows. First, the available panel, in which subjects are described by several GC × GC–MS chromatograms of their hand odor reduced to the intensities of a large set of compounds, is to be split into a calibration set for training and validation, and an independent test set for performance estimation. Second, the calibration set is used to estimate the distributions of a distance between two chromatograms when they correspond to the same subject, i.e. under the null hypothesis, and when they belong to two different subjects, i.e. under the alternative hypothesis. In the absence of prior knowledge, the null and alternative hypotheses are considered equally plausible and the prior probability of the null hypotheses is taken equal to 0.5. The posterior probability of the null hypothesis given the distance between two chromatograms is computed using Bayes' formula, and the performance of the corresponding classifier is estimated using the independent test set, in terms of area under the receiver operating curve, sensitivity and specificity. The issues of the choice of a distance and of the selection of the most appropriate compounds to enhance the performance will be discussed.

2. Material and methods

2.1. Subjects

A panel of 119 subjects was set up. The subjects gave their consent for the analyses of the samples and filled an information sheet. The data were anonymized before the analysis. This panel gathers 61 men and 58 women, with 39 subjects aged 10 to 23, 39 subjects aged 24 to 36, and 41 subjects aged 37 to 81. These subjects were sampled four times to try to make sure that at least three repetitions per subject were available, should analytical problems occur.

2.2. Analytical separation and detection

2.2.1. Sorbent phase for compounds trapping

Sorb-Stars® were purchased from Action Europe (Sausheim, France) for direct sampling. This sorbent is a patented silicon-based polymeric phase and is subject to specific conditioning processes to avoid contaminations as much as possible. The Sorb-Star® is a 2 cm long cylinder with a cylindrical section of 2 mm in diameter, has a density of 1.12 g/cm^3 and is compliant with FDA 177.2600. It is physiologically safe, and even suitable for applications in the food industry.

2.2.2. Sample collection

Care was taken that the sampling conditions were the same for all subjects. To this end, identical sample collection kits of Sorb-Stars®, special soap (Topialyse, SVR laboratory), nitrile gloves and information sheets were dispatched to different samplers in France, who were given precise instructions for the sampling protocol. The subjects were asked to wash their hands for 30 s with the provided soap, to rinse them carefully with clear water for 1 min, to dry them with a paper towel and to rub them for 2 min. 5 min later, 4 Sorb-Stars® were placed into the hands of the subjects, who rubbed them for 15 min. In the meantime, a blank was obtained by placing an open vial containing a sorbstar® in front of the sampled person.

2.2.3. Analytical devices

The Sorbstars® were thermodesorbed prior to GC × GC–MS analysis. The development of the analytical method and its optimization were the topic of two previous studies [15,16].

The purge and trap system Versatile Sample Preparator (VSP4000) was purchased from Innovative Messtechnik GmbH (Vohenstrauß, Germany). Volatile substances were purged from the Sorbstar® by the carrier gas of the GC. This concentration step is done by adsorption on a Tenax TA® in the system trap by freezing out at −30 °C. After incubation of the sample at 190 °C and completion of the purging process (20 mL/min during 20 min), the concentrated substances were transferred by fast thermal desorption from the trap onto a transfer line, heated at 280 °C, and then separated by GC.

The thermodesorption device was coupled with a GC × GC–MS Q2010Plus purchased from Shimadzu (Kyoto, Japan). A ZB-1MS column (30 m × 0.25 mm, 0.25 µm) (Phenomenex, Torrance, USA) coupled with a ZB-1701 column (1.5 m × 0.1 mm, 0.1 µm) (Phenomenex, Torrance, USA) were used to conduct the chromatographic separation. The modulation was performed with a N_2 cooled Zoex ZX1 thermal modulator (Zoex, Houston, USA), and the modulation time was set at 8 s. The initial temperature was set to 40 °C for 1 min, then raised to 250 °C at 2.5 °C/min, and held for 1 min at 250 °C. The mass spectrometer was used with the electronic ionization source (70 eV) heated at 200 °C, the acquisition being performed in scan mode. The scan range was 29–250 *m/z*, and the sampling frequency 50 Hz. Fig. 1 displays a typical chromatogram collected on a 25-year-old man's hand.

2.3. Chromatogram processing

Data were acquired, converted to .mzXML file with GC Real Time Analysis 4.20 (Shimadzu software), and then processed with MatlabTM (Natick, MA, USA) version 9.3.0.713579 (R2017b), its Statistics and Machine Learning Toolbox version 11.2 and its Bioinformatics Toolbox version 4.9.

A preliminary manual processing of 25 chromatograms obtained on a subset of subjects of both genders sampled several times at different time instants enabled to draw up a list of several hundreds of peaks. A library was built to store their retention times, their linear retention index, their mass spectrum, and the name of the corresponding compound when it could be identified using the NIST library. Indeed, a compound does not need to be formally identified for the comparison of chromatograms, whereas the availability of its mass spectrum is compulsory. We also checked whether compounds described in the literature as constituents of the human hand odor [1] were present in this library, otherwise they were included. This preliminary work led us to a customized library of 600 compounds, identified or not, which were looked for in each chromatogram using a “home-made” Matlab script [17].