The identification of key components relevant to sensory perception of quality from commercial chocolate samples was accomplished after chemometric processing of GC×GC-MS (Comprehensive Two-dimensional Gas Chromatography with Mass Spectrometric Detection) profiles corresponding to HS-SPME ( Headspace Solid Phase Microextraction) extracts of the samples. Descriptive sensory evaluation of samples was carried out using Optimized Descriptive Profile (ODP) procedures, where sensory attributes of 24 commercial chocolate samples were used to classify them in two classes (low and high chocolate flavor). 2D Fisher Ratio analysis was applied to four-way chromatographic data tensors (1st dimension retention time $t_1$ × 2nd dimension retention time $t_2$ × m/z × sample), to identify the crucial areas on the chromatograms that resulted on ODP class separation on Principal Component Analysis (PCA) scores plot. Comparing the relevant sections of the chromatograms to the analysis of the corresponding mass spectra, it was possible to assess that most of the information regarding the sample main sensory attributes can be related to only 14 compounds (2,5-dimethylpyrazine, 2,6-dimethyl-4-heptanol, 1-octen-3-ol, trimethylpyrazine, β-pinene, o-cimene, 2-ethyl-3,5-dimethylpyrazine, tetramethylpyrazine, benzaldehyde, 1,3,5-trimethylbenzene, 6-methyl-5-hepten-2-one, limonene, ...
benzeneethanol and 1,1-dimethylbutylbenzene) among the complex blend of volatiles found on these extremely complex samples.

Keywords: chocolate flavor; sensorial analysis; GC×GC-MS, Fisher ratio; principal component analysis

INTRODUCTION

Chocolate is a complex matrix, exhibiting a volatile fraction with more than 500 different chemical components [1]. Several analytical tools have been used for identification of chocolate chemical constituents, mainly based on gas chromatography with mass spectrometry identification [2–4]. However, due to the high complexity of this matrix, in last years the use of comprehensive two-dimensional gas chromatography (GC×GC) [5] is becoming popular: compounds that could co-eluate in the first dimension chromatographic column could be separated in the second dimension column [6]. Described in 1991 [7], GC×GC technique is being constantly improved: e.g. with the introduction of new modulation devices [8,9] and data processing strategies (due to the complexity and large size of the chromatographic data sets) [10–12]. Consequently, the number of successful applications of GC×GC has increased in the literature, specially for complex samples such as petroleum [13,14], food [15–17], biological fluidics [18], among others.

Many components from the volatile fraction of chocolates are relatable to sensorial quality parameters assessable by procedures such as Quantitative Descriptive Analysis (QDA) [19,20]. QDA is the most complete and popular tool in food science, but requires long-term training of the panelists, being relatively slow and expensive [22]. In terms of time and cost, there are some more appropriate alternatives for QDA [20] such as Optimized Descriptive Profile (ODP) analysis [21]. ODP is based on semi-trained panelists, who evaluate the products using an unstructured anchored scale corresponding to the extremes (weak and strong) of the range of the sensorial property being assessed. Samples representing the extremes of the attribute scale (weak and strong standard) and samples being tested should be offered to panelists in the same tasting session; consequently, people with limited training will be able to consistently evaluate the samples. In this approach, on each tasting session just a single sensorial attribute should be evaluated, since the samples and the references are presented simultaneously [21,23,24].

The combination of volatile profile characterization by GC×GC and sensory analysis can be a powerful tool to study chocolate. If sensory analysis results could be objectively correlated to GC×GC data, it would be possible to pinpoint volatile chemical components responsible for specific quality features of the samples. The modelling of quantitative sensory data correlation to GC×GC chromatograms should be made using multivariate chemometric tools [10,25,26] such as Principal Component Analysis (PCA), Partial Least Squares regression (PLS) [27] as well as neural network modelling [28] have been successfully applied in these studies. Among them, PLS [29] is by far the most useful when chromatographic data has to be correlated to some macroscopic parameter of the samples; however, it has some limitations for large and complex data structures [30] such as GC×GC chromatograms. In these situations, preliminary selection of sections of the chromatograms potentially related to the sensory property being assessed before multivariate modelling can be crucial. Fisher Ratio analysis (FR) has been successfully applied as pre-processing strategy to filter non-relevant section from large GC×GC data sets and improve the quality of subsequent multivariate analysis [31–33]. Essentially, FR analysis locates regions of the chromatograms were extra-class variations (pertinent to the distinction of samples of different sample classes) are significantly larger than intra-class variations (that does not contain useful information regarding sample categorization). Only the relevant regions of the chromatographic data sets are feed to multivariate data analysis algorithms, improving the quality of the chemometric processing and demanding less computational power and resources [34].

For each signal point in the \( \mathbf{t}_R \times \mathbf{t}_R \times \mathbf{m}/\mathbf{z} \) chromatographic space, Fisher Ratio is defined as a scalar value calculated as:
\[
FR = \frac{\sigma_{ci}^2}{\sigma_{err}^2}
\]

where
\[
\sigma_{ci}^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 \cdot n_i}{k - 1}
\]

and
\[
\sigma_{err}^2 = \frac{\sum (\bar{x}_{ij} - \bar{x}_i)^2 - (\sum (\bar{x}_i - \bar{x})^2 \cdot n_i)}{(N - k)}
\]

In these equations, \(n_i\) is the number of measurements in the \(i\)th class, \(x_i\) is the mean of the \(i\)th class, \(x\) is the overall mean, \(k\) is the number of classes, \(x_{ij}\) is the \(i\)th measurement of the \(j\)th class and \(N\) is the total number of sample profiles. Sections of the chromatogram with values for \(FR\) significantly larger than 1 are relevant to the distinction of sample classes; points where \(FR \sim 1\) correspond to sections of the chromatograms where intra- and extra-class variations of the signal are similar and are not relevant to sample qualification. To select regions of the chromatographic space for multivariate analysis, data points with \(FR\) smaller than a user-defined threshold \(FR_{cut}\) are excluded without loss of information.

In this work, the volatile fraction of commercial chocolate samples was isolated by headspace Solid Phase Microextraction (HS-SPME) an analyzed by GC×GC-MS; the so-called “chocolate flavor” sensorial attribute of these samples was previously determined by ODP, which were split in two classes according to the corresponding chocolate flavor values (low and high). After preliminary selection of potentially relevant areas of the chromatograms using 2D Fisher Ratio analysis, the data tensors were subject to exploratory inspection by Principal Component Analysis. The loadings from PCA were compared to the chromatograms, and the mass spectra and retention indexes of the peaks appearing on the relevant regions allowed identification of the key chemical compounds responsible for class separation.

MATERIALS AND METHODS

Samples and Sample Preparation

Twenty-four chocolate samples of different suppliers and with nominal cocoa content from 45 to 90% were obtained at the local market. The volatile fraction of the samples was isolated by Headspace Solid Phase Microextraction (HS-SPME) using a previously described procedure [35]. Aliquots of (1.000 ± 0.005) g of sample were exposed to N\(_2\)-frozen and then grounded manually and then the grounded samples were weighed in 15 mL septum-sealed extraction vials. After a 5 min period for sample/headspace equilibration at 60 °C a 50/30 μm Divinylbenzene / Carboxen / Polydimethylsiloxane (DVB/CAR/PDMS) SPME fiber (Supelco, Bellefont – PA, USA) was exposed for 50 min to the sample headspace and then immediately inserted into the GC×GC system injector for 5 min. Before the extractions, the fiber was conditioned for 2 h at 250 °C according to the supplier instructions.

GC×GC-MS analysis

GC×GC-MS analysis was performed using a prototype based on a Shimadzu (Tokyo, Japan) QP2010+ GC-MS [36]. The column set consisted on a 30 m × 0.25 mm × 0.25 μm HP-5 column (Agilent Technologies, Wilmington, DE) connected to a 0.80 m × 0.1 mm × 0.1 μm Solgel Wax column (SGE Analytical Science, Ringwood - Victoria, Australia). The injector was operated on splitless mode at 260 °C. The column oven temperature was programmed as: 40 °C to 110 °C at 3 °C min\(^{-1}\) and 110 °C to 240 °C at 10 °C min\(^{-1}\); the modulation period was set to 6.0 s and a cryogenic modulator was used. The interface temperature was 260 °C. MS detector photomultiplier high voltage was programmed from 0.8 kV up to 10 min and then increased to 0.9 kV until final of the chromatographic run. Data was collected at an acquisition rate of 25 scan s\(^{-1}\) with scanned mass range set from \(m/z = 40\) D to 340 D. The determination 1\(^{st}\) dimension linear temperature programming retention indexes (LTPRI) of relevant peaks was performed after analysis of selected samples spiked with n-alkane mixture (C\(_6\) to C\(_{20}\)). Detection of chromatographic peaks on the chromatograms was performed using GCImage software (Zoex, Houston – TX, USA) and analyte identification achieved by matching their mass spectra and LTPRI with literature data and spectra on NIST 2010 library.
**Optimized Descriptive Profile (ODP)**

ODP of samples was performed according to the methodology described by Silva et al. [21,23]. Triangular tests were applied in 20 judges using two samples from the same supplier but with different cocoa content (50% and 75% cocoa). The triangular test was performed in triplicate and the judges that have at least two right answers were selected for the ODP tests. In this case, 75% of the judges were approved in the triangular test.

**2D Fisher ratio for four-way data**

In order to simplify further data processing, Fisher Ratio Analysis was used to select areas on the $t_R \times t_R \times m/z$ chromatographic space generated by GC×GC-MS that really contained pertinent information related to the sample differentiation. The main algorithm used to calculate the variance inside one class, between two classes and the Fisher Ratios is described elsewhere [37]; this procedure was later extended to be applied directly in four-way data [34] such as the case here. Fisher Ratio analysis applied in four-way chromatograms allows both pixel-based (direct application to the raw Signal = f($t_{R1}$, $t_{R2}$, m/z) tensors without peak detection and integration) and peak-level (where input data are peak area tables obtained after conventional detection and integration) [31,38].

Figure 1 shows the general scheme of 2D Fisher ratio calculation in the four-way space. Samples with highest and lowest values for the chocolate flavor attribute obtained in ODP analyses were chosen to define the two classes and (Sample × $t_{R1} \times t_{R2} \times m/z$) data cubes for these samples were generated for each mass channel (Figure 1-A). These cubes were unfolded by concatenating the second and third dimensions (retentions times), resulting in matrices where each line corresponds to a sample and retention times are in columns for each mass channel (Figure 1-B). Fisher Ratios for all matrices were calculated, resulting in a vector with dimensions equal of the number of retention times for each mass channel (Figure 1-C). Finally, a new matrix was formed with Fisher Ratios for all mass channels (Figure 1-D); this matrix was folded back, resulting in a data cube with (Fisher Ratios × $t_{R1} \times t_{R2}$) (Figure 1-E). Each signal point in this cube is obtained as the sum of all individual m/z signals (Figure 1-F) [31,34]. All operations here were performed in MATLAB R2016b (MathWorks, USA) using self-made routines for 2D Fisher ratio calculation.

![Diagram of 2D Fisher ratio calculation in GC×GC-MS](image)

**Figure 1.** General scheme of 2D Fisher ratio calculation in GC×GC-MS. N is the number of samples, M is the number of mass (m/z) channels, $t_{R1}$ is the retention time in the first column and $t_{R2}$ is the retention time in the second column.
Identification of volatile components essential to predict chocolate flavor sensory attribute

The identification of the set of key volatile compounds that need to be assessed to model the evaluated sensory attribute (chocolate flavor) from GC×GC-MS chromatograms was performed through an iterative process. After calculation of the tensor of Fisher Ratio values, selection of potentially relevant sections of the chromatograms for differentiation between sample classes (high and low chocolate flavor) was performed using an initial tentative value for FR_{cut}. Data sets containing only the selected areas of the GC×GC-MS were unfolded, and the mass spectra channel was summed obtaining the TIC chromatograms and after that the data were submitted to Principal Component Analysis, using MatLab PLS Toolbox v. 7.3.1 (Eigenvector, USA) with pareto scaling as pre-processing method and the resulting scores plots inspected to check for distinction between sample classes. This process was repeated, at each time using larger FR_{cut} values (and therefore smaller input data sets on PCA) until clear separation of sample classes (low and high chocolate flavor) appears in the scores plots.

After definition of the maximum Fisher Ratio threshold for PCA class separation – and therefore isolating only the absolutely essential parts of the chromatograms that contain information regarding class differentiation, the correspondent loadings matrixes were matched to GC×GC chromatograms to point the location on the $t_1 \times t_2$ plane of compounds relevant to characterization of these categories.

RESULTS AND DISCUSSION

Figure 2 shows the average values of panelist responses for chocolate flavor attribute for the studied samples studied. Samples with intermediary sensory parameter values between the classes were marked as “undefined”; Samples with chocolate flavor attribute defined as “low” and “high” are indicated in the figure.

![Figure 2. Chocolate flavor values of all commercial chocolates studied. Samples indicated with a circle were considering as ‘low chocolate flavor’ class and samples with an arrow were considered as ‘high chocolate flavor’ class.](image)

Figure 3 shows the chromatogram parts selected as relevant to sensory classification of samples using the final threshold value of FR_{cut} = 2×10^7 overlapped with a typical chromatogram. The spots selected by Fisher Ratio analysis were fed as input data for PCA modelling.
Figure 3. Spots selected by 2D Fisher ratio (in red) overlapped in a typical chromatogram (black color).

Figure 4 compares PCA scores plots obtained after modeling using the raw, whole chromatographic data sets without variable selection and scores plots of PCA using partial chromatographic data selected after FR preliminary analysis. In the former, it is quite clear that samples group in two distinct regions of the plane - confirming that this value of threshold was adequate for sample classification; as for the PCA scores plot resulting after using the whole, unfiltered data no class separation is observable.

Figure 4. PCA scores plot. (a) all variables included in the model, (b) only variables with Fisher ratio above $2 \times 10^7$. ▼ samples of “low chocolate flavor” class and ● samples of “high chocolate flavor” class.

By using the loadings of the first PC of the data set formed by the selected variables, it is possible to point the areas in the GC×GC chromatograms where compounds responsible for class separation elute. Those compounds were tentatively identified from their retention indexes and mass spectra. Figure 5 shows the positive (Figure 5a) and negative loadings (Figure 5b), where is possible to note a reduced number of spots with significant loading values. These results indicated that for the class separation (low and high chocolate flavor) a reduced number of variables were selected by the 2D Fisher ratio (only 5% of
the total of variables). Positive loadings were related to samples with high chocolate flavor and negative loadings with samples with low values for this attribute.

Figure 5. Plot of the loadings of the first PC. (a) positive loadings. (b) negative loadings.

As a reduced number of spots in GC×GC chromatograms were selected, the number of identified components was also low. The identified compounds are listed in Table I. The compounds with positive loadings were mostly pyrazines (tetramethylpyrazine, 2-ethyl-3,5-dimethylpyrazine, trimethylpyrazine and 2,5-dimethylpyrazine) which are known to be related to the chocolate flavor [39,40]. Also, Counet et al. [41] described in their work that the tetramethylpyrazine was the most abundant pyrazine found in dark chocolates evaluated and that this compound could have a reasonable influence in the final odor of chocolates. Besides the recognized influence of pyrazines in the chocolate odor 1-octen-3-ol was also observed as important compound by Nightingale et al. [2] when evaluating the changes in the headspace of dark chocolates during different storage times.

Table I. Compounds identified as significant for classes separation

<table>
<thead>
<tr>
<th>Positive loadings</th>
<th>Negative loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,5-dimethylpyrazine</td>
<td>benzaldehyde</td>
</tr>
<tr>
<td>2,6-dimethyl-4-heptanol</td>
<td>1,3,5-trimethylbenzene</td>
</tr>
<tr>
<td>1-octen-3-ol</td>
<td>6-methyl-5-hepten-2-one</td>
</tr>
<tr>
<td>trimethylpyrazine</td>
<td>limonene</td>
</tr>
<tr>
<td>β-pinene</td>
<td>benzeneethanol</td>
</tr>
<tr>
<td>o-cimene</td>
<td>1,1-dimethylbutylbenzene</td>
</tr>
<tr>
<td>2-ethyl-3,5-dimethylpyrazine</td>
<td></td>
</tr>
<tr>
<td>tetramethylpyrazine</td>
<td></td>
</tr>
</tbody>
</table>

The opposite can be understood for the negative loadings. Samples with low chocolate flavor values show this behavior due to the compounds indicated in the negative loadings. Among them, the compounds were benzaldehyde, 1,3,5-trimethylbenzene, 6-methyl-5-hepten-2-one and limonene.

Among the identified compounds limonene was already described for being a important compound in chocolates related with the crystallization process, physical properties and the aroma [42–44]. Furthermore,
benzaldehyde was previously related with the bitter sensorial attribute [45] and leads to confirm the obtained results in this work, since chocolates described as low chocolate flavor present more bitterness flavor (benzaldehyde). Here is important to know that Brazilian chocolate consumers are more used to consume milk chocolate (that present more sugar than dark chocolates used in this study) [46] and for this reason describe chocolates with less chocolate flavor as chocolates with more bitterness.

Finally, the threshold value stipulated by the Fisher ratio was applied for all 24 samples studied. A new data set was formed with all samples, and the PCA was calculated. The plot of the scores of the first three PCs for this new data is presented in Figure 6. We can verify that the samples were again separated into high and low chocolate flavor values (red and black samples), and between them, samples with intermediate values of this attribute (in blue). This is an additional indication of the correct variable selection by Fisher ratio, showing that samples with intermediate chocolate flavor attribute have also intermediate content of the chemical constituents.

![Figure 6. Plot of the scores of the PCA model using all samples studied. ▼ low values of chocolate flavor attribute ♦ intermediate values of chocolate flavor attribute and ● high values of the chocolate flavor attribute.](image)

CONCLUSIONS

In this work we demonstrated a strategy for identification of the essential set of volatile components related to the sensory parameter “chocolate flavor” from GC×GC-MS data in tandem with sensory analysis PCA combined to variable selection using Fisher Ratio analysis. The 2D Fisher ratio was able to point the sections of chromatograms where key compounds that separate the predefined classes (low and high chocolate flavor) in the PCA scores plot were located. It is important to emphasize that the separation was accomplished using a reduced number of variables and consequently only 14 chemical components were identified. This result does not mean that only these compounds are related to the differentiation of the chocolate flavor – several other minor components can contribute with this attribute, but they certainly are associated with this feature; also, this approach does not account for synergistic effects that could appear on compounds which are on non-selected areas of the chromatographic space. The compounds identified for positive loadings, and with high chocolate flavor attribute, are mostly pyrazines that are recognized for the characterization of chocolate flavor.
Conflicts of interest
No conflicts of interest – financial or otherwise – existed on the composition of this manuscript.

Acknowledgements
This work was supported by grants provided by Fapesp - São Paulo Research Foundation [Grants 2012/14695-8 and 2014/50867-3], CNPq - National Council for Scientific and Technological Development [Grant 465389/2014-7] and CAPES - Brazilian Coordination for the Improvement of Higher Education Personnel [Finance Code 001].

This paper is the last paper resulting from the last and fruitful cooperation between Prof. Ronei Poppi and our research group. He left this world prematurely in April 2020; saying that he is going by to be missed as scientist, associate and friend is a bitter understatement. It is in honor of his memory we dedicate this paper.

REFERENCES