This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
Projective Mapping for interpreting wine aroma differences as perceived by naïve and experienced assessors

Luisa Torri, Caterina Dinnella, Annamaria Recchia, Tormod Naes, Hely Tuorila, Erminio Monteleone

Abstract

The perceptions of differences in the aroma of high quality Italian red wines were compared in experts and consumers by Projective Mapping. Quality and typicality assessments from experts, and liking ratings from consumers, were collected on the same wine set. The sensory profiles of the wines were described by a panel of trained subjects. The results suggest that product separation by experts was mainly based on the perceived overall quality rather than on specific sensory differences. Product differentiation by consumers was poor and worse than that of experts and trained subjects. Consumers' internal preference map showed a good sample separation based on liking data and allowed the identification of the aroma attributes that drove their preferences. Results from consumer tests indicated that differences among samples based on liking data were more evident than those from Projective Mapping. An increased differentiation ability was observed for those consumers able to match the duplicate samples in the Projective Mapping test. In this group, sample differentiation based mainly on liking was observed. The socio-cognitive traits of these subjects highlighted their high level of wine knowledge.

In general, the results indicate that Projective Mapping can be a valuable method for investigating the perceived similarities/dissimilarities among samples with subtle sensory differences when assessors share a high level of knowledge and experience about the product.

1. Introduction

The strategic role of consumer input for product development, advertisement, marketing positioning and communication led to the development of a number of methods to gather information about consumers' perceptions of the sensory properties of food products. Projective Mapping (PM) is a comparative sensory technique which allows consumers to evaluate products in an overall and simple way by expressing perceptual similarities/dissimilarities by a two-dimensional projection. Subjects are asked to use their own criteria to position objects according to the rule that the more similar two objects are perceived, the closer they are placed on the map and the product coordinates on the two-dimensional space quantify their separation (Risvik, McEwan, Colwill, Rogers, & Lyon, 1994; Risvik, McEwan, & Rodbotten, 1997). PM is supposed to be a simpler and faster way to obtain product inter-distances than similarity scaling (Risvik et al., 1997). This method may provide more graded information than sorting, because it is based on the individuation of similarities and differences using a graphic representation and not a nominal categorization (Nestrud & Lawless, 2008; Pagès, 2005). Perceptual maps are generated from PM data by means of multidimensional analysis methods (Multidimensional Scaling – MDS, Generalized Procrustes Analysis – GPA and Principal Component Analysis – PCA) (King, Cliff, & Hall, 1998; Risvik et al., 1994, 1997). Multiple factor analysis (MFA) has been proposed for PM data for the more specific “nappe” or mapping method (Morand & Pagès, 2005; Pagès, 2003, 2005).

The main technical advantages of PM are that training is not required, high numbers of samples (10–12) can be evaluated in each session and it is a user-friendly procedure (Pagès, 2005). Because of these characteristics this technique has become of interest in food sensory science and in wine research in particular. However, PM procedure shows some weakness. As reported by Nestrud and Lawless (2008) one important issue include the reliability of the results from this method.

With PM it is possible to get a representation of the products, which integrates the relative importance for the subjects of the
characteristics of the products; however, this does not characterize the product itself (Pagès, 2005). Sensory attributes have been shown to be a measure of consumers' perceptions of food sample similarities/dissimilarities by the use of a PM technique combined with descriptive sensory data from both conventional profiling (Kennedy & Heymann, 2009; Perrin et al., 2008; Risvik et al., 1997), and other descriptive methods such as free choice profile (Perrin et al., 2008) and flash profile (Albert, Varela, Salvador, Hough, & Fiszman, 2011; Moussaoui & Varela, 2010; Perrin et al., 2008; Veinand, Godefroy, Adam, & Delarue, 2011).

Risvik et al. (1997) compared consensus mapping dimensions from consumers to those from the profile data and noted a good agreement on the obvious aspects of the product. This tendency was confirmed in further studies (Barcenas, Pérez Elortondo, & Albisu, 2004).

Associating PM data collection with a verbalization task further highlighted the importance of sensory attributes in sample differentiation by consumers (Nestrud & Lawless, 2010; Albert et al., 2011; Veinand et al., 2011). Furthermore, the analysis of terms used by consumers to describe sample groups led to the identification of hedonic dimensions as relevant to product differentiation (Ares, Deliza, Barreiro, Giménez, & Gámbaro, 2010). The most liked samples tend to be positioned close together on perceptual maps (Ares, Varela, Rado, & Giménez, 2011; Risvik et al., 1997), however, a strong relationship between consumer preferences and perceptual space from PM has not been demonstrated.

Different configurations have been observed comparing results from PM carried out with naïve consumers, experts and trained subjects (Barcenas et al., 2004; Risvik et al., 1997; Pagès, 2005; Perrin et al., 2008; Nestrud & Lawless, 2010). The way subjects gain experience of a product (sensory experience) and their level of product knowledge (particularly for experts and professionals) significantly influence product differentiation (Mairet, Symoneaux, Jourjon, & Mehinagic, 2010). Specifically trained subjects tend to use non-hedonic criteria for sample discrimination irrespective of the product under evaluation (Delarue & Sieffermann, 2004).

Moreover, the improved short term memory ability reported for trained panelists (Avancini De Almeida, Cubero, & O'Mahony, 1999) might account for the higher discrimination ability of experts when compared to naïve consumers, in a categorization task requiring the comparison of several samples (Chollet, Lelièvre, Abdi, & Valentin, 2011). Comparing trained with untrained subjects for repeatability and ability to match the duplicate sample in food categorization task, showed a similar performance of both subject groups (Chollet et al., 2011). However, the consensus level of their perceptual maps seemed to be affected by both the level and the kind of expertise. Formally trained subjects were more consensual than untrained subjects.

Comparing PM results from consumers and chefs, Nestrud and Lawless (2008) found a relatively low consensus level for chefs, consistently with the notion of a higher level of idiosyncratic criteria. Authors hypothesized that for panelists who have experience with tasting or a specific product set, the PM may be a useful tool to uncover criteria that are difficult elucidate with traditional consensus-derived attribute lists.

The potentiality for uncovering aspects of food perception related to psychophysical and cognitive models of individuals and subject groups, which are difficult to access by scaling sensory data collection methods, represents an original feature of the PM technique. From an applicative perspective, the perceptual maps obtained from this technique associated to descriptive data and hedonic responses could represent a useful tool to explain the consumer food like/dislike dimension thus helping for effective product development and marketing strategies.

Primary aroma is considered one of the most important distinctive traits of mono-varietal wines. In the current study, the perception of aroma similarities/dissimilarities among mono-varietal red wines by experts and consumers was assessed using the PM technique. Quality assessment from experts and liking ratings from consumers were also collected on the same wine set. Furthermore, sensory profiles of the wines were described by a separate panel of trained subjects. Perceptual maps were compared with the aim of gaining further insights into differentiation criteria used by assessors with different levels of expertise and to investigate the role of sensory properties and hedonic responses as drivers for wine differentiation by experts and consumers. Finally, the relationship between consumers’ ability to match duplicate samples in PM test and their background variables were explored; its effect on map consensus levels and on wine differentiation criteria was investigated.

2. Materials and methods

2.1. Wine samples

Eleven wines served as stimuli (Table 1). Six Tuscan PDO Sangiovese wines represented the whole sensory variability of PDO Sangiovese wines, based on a previous study aimed at describing sensory similarities and differences among 24 PDO Sangiovese wines (Recchia, Picchi, Fia, Bertucciolli, & Monteilemme, 2009). A further five Italian mono-varietal wines were selected by wine experts of the National “Enoteca” of Siena to represent high quality standard Italian regional wines, belonging to the same segment of Sangiovese wine from Tuscany in terms of price ($20–30 euros) and availability (mainly in wine shops rather than supermarkets).

2.2. Subjects

2.2.1. Trained panel

The trained panel was composed of nine subjects (4 males, 5 females, 22–28 years, mean age 25). They were selected from the wine-trained panel operating at the Agricultural Biotechnology Department of Florence University. They had participated in previous tests aimed at describing the aroma of red wines in general and Sangiovese wines in particular. Before evaluating the samples they participated to 10 one-hour training sessions.

2.2.2. Wine expert panel

The panel of experts was composed of thirteen Tuscan professionals (oenologists and wine producers; 8 males, 5 female, mean age of 40). They had an average of 10 years experience in the wine industry.

2.2.3. Consumers

Eighty-one wine consumers from Florence area (50 males, 31 females, 22–59 years, mean age 34) participated in the study. They had seen or received an invitation and volunteered based on their interest and availability. Subjects were informed that the test

<table>
<thead>
<tr>
<th>Wine code</th>
<th>Wine name</th>
<th>Grape cultivar</th>
<th>Origin region</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG1</td>
<td>Nobile di Montepulciano</td>
<td>Sangiovese</td>
<td>Tuscany</td>
</tr>
<tr>
<td>SG2</td>
<td>Chiante</td>
<td>Sangiovese</td>
<td>Tuscany</td>
</tr>
<tr>
<td>SG3</td>
<td>Brunello di Montalcino</td>
<td>Sangiovese</td>
<td>Tuscany</td>
</tr>
<tr>
<td>SG4</td>
<td>Nobile di Montepulciano</td>
<td>Sangiovese</td>
<td>Tuscany</td>
</tr>
<tr>
<td>SG5a–SG5b</td>
<td>Chiante</td>
<td>Sangiovese</td>
<td>Tuscany</td>
</tr>
<tr>
<td>SG6</td>
<td>Brunello di Montalcino</td>
<td>Sangiovese</td>
<td>Tuscany</td>
</tr>
<tr>
<td>PrM</td>
<td>Primitivo di Manduria</td>
<td>Sangiovese</td>
<td>Tuscany</td>
</tr>
<tr>
<td>BR</td>
<td>Barolo</td>
<td>Nebbiolo</td>
<td>Piedmont</td>
</tr>
<tr>
<td>AV</td>
<td>Aglianico del Vulture</td>
<td>Aglianico</td>
<td>Basilicata</td>
</tr>
<tr>
<td>NA</td>
<td>Nero d’Avola</td>
<td>Nero d’Avola</td>
<td>Sicily</td>
</tr>
<tr>
<td>CS</td>
<td>Cabernet Sauvignon</td>
<td>Cabernet Sauvignon</td>
<td>Veneto</td>
</tr>
</tbody>
</table>
would be run in a restaurant located on the university campus close to Florence and that there would be dinner for free after they completed the test. Six groups each consisting of 10–15 subjects were formed and were asked to come to the restaurant at half hour time intervals from 5 to 7.30 pm. All tests were conducted individually, social interaction was not permitted.

Written informed consent was obtained from each subject after the experiment had been described to them.

2.3. Experimental procedure

2.3.1. Conventional profiling

The trained panel described the aroma (odor by nose) characteristics of the wines. Subjects were trained to recognize and rate 18 attributes widely used to describe red wine aroma (Table 2). All attributes were evaluated on a 9-point category scale (1 = extremely weak; 9 = extremely strong). Subjects were presented with two standards corresponding to “weak/moderate” and “strong” intensities for each attribute (Carlucci & Monteleone, 2008; Recchia et al., 2009). Subjects participated in a total of three training sessions.

Samples (30 ml) were presented at room temperature in 100 ml amber glasses, covered with a plastic cover and marked with a three digit code. In each session 11 samples were evaluated in two subsets of 5 and 6 samples each. The presentation order of samples was balanced for first order and carry over effects. Subjects rested for 30 s after each evaluation; a further break of 45 min was held between sample sets. Subjects participated in a total of four sessions, each sample was replicated four times. All evaluations were performed in individual booths under red lights in order to eliminates visual clues. Scores were collected by FIZZ software (version 2.47B, Biosystems, Courtenon, France).

2.3.2. Wine expert test

Sessions were held in the testing room of a private wine analysis laboratory close to Siena (Tuscany) and consisted of three parts.

2.3.2.1. Smell test. The olfactory ability of experts was tested by a “Sniffin’ test” (Burghart Medical Technology, Wedel, Germany). Twelve odor sticks, each associated with a list of four descriptors, were presented to each subject. Subjects sniffed each stick and selected the appropriate descriptor from the corresponding list (Gudziol, Hummel, Negosas, Ishimaru, & Hummel, 2007). More than four mistakes in this test indicated a possible odor perception disorder in the subject.

2.3.2.2. Mapping test. Subjects were presented with 12 three-digit coded samples (11 wines reported in Table 1 plus a replicate sample – Chianti wine SG5 a and b). Experts were provided with a white rectangular paper (60 × 40 cm) and instructed according to Pagès (2005):

Principle: You have to be being asked to evaluate the similarities (or dissimilarities) between several wines. You have to do this according to your own criteria, those that are significant for you. You do not have to indicate your criteria. There are no good or bad answers.

Procedure: You have to position the wines on the tablecloth in such a way that two wines are very near if they seem identical to you and that two wines are distant from one another if they seem different to you. This must be done according to your own criteria. Do not hesitate to express strongly the differences you perceive by using the whole of the sheet. If you have any questions or are unclear what to do, please don’t hesitate to ask.

2.3.2.3. Evaluation of quality and typicality. After completing the mapping test and after a 45 min break, subjects were presented with the 12 wine samples. They smelled wine samples and rated their perceived quality on a 9-point category scale anchored on the left end with “very poor” and on the right end with “excellent”. After a 20 min break, a new set of the same 12 samples and a new score card were presented. Experts answered the following question: “If you had to explain to a friend the typical odour of a Tuscan Sangiovese wine, what kind of example is this wine?”. Responses were collected using nine categories from “bad example of Sangiovese” (1) to “very good example of Sangiovese” (9). The order of collecting quality and typicality responses was counterbalanced across subjects. Sample presentation was the same as described in the conventional profiling section.

Different three digit codes were used to identify samples in the three different evaluation sessions.

2.3.3. Consumer test

Consumer test took place from 5 to 8 pm. Subjects participated in a session organized in four parts. The first part consisted of the “Smell test” carried out with the same procedure described above for experts. In the second part, respondents filled in a background questionnaire in which they recorded their socio-demographic background (age, gender and education); their familiarity/previous use of 20 wines including the 11 selected wines; their frequency of wine consumption and finally their socio-cognitive background (innovativeness and involvement). Innovativeness is a type of personality trait. Personality traits are thought to be relatively enduring patterns of behavior or cognition that differentiate people. As reported in Goldsmith and Foxall (2003), innovativeness describes reactions to the new and different, openness to experience, motivation towards learning. Responses to wine attributes may be affected by different levels of involvement with the product. In particular, for the product category of wine product involvement has been conceptualized as the interest, enthusiasm and excitement that consumers exhibit towards wine (Goldsmith, d’Hauteville, & Flynn, 1998). Because of their characteristics, both innovativeness and involvement were assumed to be potentially relevant to explain consumers’ responses in the present study. Details about the questionnaire and related scales are reported in Table 3. All the scales associated with questionnaire variables were presented in Italian. After filling in the questionnaire, subjects participated in the mapping test using the same procedure as described above for experts. After completing the test there was a 45 min break that subjects spent in the leisure room of the restaurant with magazines and equipped with TV and board games. Then subjects were presented with 12 wines samples (11 wines reported in Table 1 plus a replicate sample – Chianti wine SG5 a and b), smelled and rated their liking on a Labeled Affecive Magnitude Scale from 0 to 100, anchored at the ends with ‘greatest imaginable dislike’ and ‘greatest imaginable like’ (Cardello & Schutz, 2004).

2.4. Data analysis

2.4.1. Descriptive data

Intensity data from the trained panel were analyzed by multiblock PCA (Tucker-1) and by p*MSE plot (Panel Check software,
ver 1.4.0, Nofima, Norway) to assess panel calibration and assessor performance, respectively (Naes, Brockhoff, & Tomic, 2010). In particular, the occurrence of assessor–attribute combinations deviating from the rest of the panel was checked, for each attribute, by means of the analysis of a correlation plot generated by the multiblock PCA. For an assessor interpreting an attribute differently from the rest of the panel, the corresponding correlation loadings are located closer to the center than others, making it possible to detect assessor–attribute combinations with weak relation to the general underlying data structure. No cases of disagreement among panellists were found for all the attributes. Furthermore differences among assessors in product differentiation ability and in consistency over replicates were analyzed by p*MSE plots. These plots are derived from the computation of a one way ANOVA for each assessor and each attribute combination. For each attribute, individual ability of discriminating among samples, expressed as p-value, is plotted along the vertical axis, while the consistency over replicates, expressed as ratio of sums of squares of the differences between average values of products and the variance of replicates (MSE), is plotted along the horizontal axis. The lower are both p and MSE values the better is the assessor performance and thus the presence of weak performers can be detected (see Naes et al., 2010 for details). On the basis of the p*MSE plots, all assessors were considered reliable.

Table 3
Questionnaire variables (familiarity/previous use of wines, frequency of wine consumption and socio-cognitive variables) and relevant scales.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Familiarity with 20 wines including the 11 wine samples</td>
<td>5-Point scale (Bäckström, Pirttilä-Backman, &amp; Tuorila, 2004):&lt;br&gt;1 = I do not recognize the product&lt;br&gt;2 = I recognize the product, but I have not tasted it&lt;br&gt;3 = I have tasted, but I do not use the product&lt;br&gt;4 = I occasionally eat the product&lt;br&gt;5 = I regularly eat the product</td>
</tr>
<tr>
<td>2. Consumption frequency of red and white wine</td>
<td>7-Point scale:&lt;br&gt;1 = once a month or less&lt;br&gt;2 = 2–3 times a month&lt;br&gt;3 = once a week&lt;br&gt;4 = 2–3 times a week&lt;br&gt;5 = 4–5 times a week&lt;br&gt;6 = once a day&lt;br&gt;7 = more than once a day</td>
</tr>
<tr>
<td>3. Consumption frequency of red and white wine in a wine bar</td>
<td>6-Point scale:&lt;br&gt;1 = 1–3 times a year or less&lt;br&gt;2 = once a month&lt;br&gt;3 = 2–3 times a month&lt;br&gt;4 = once a week&lt;br&gt;5 = 2–3 times a week&lt;br&gt;6 = more than three times a week</td>
</tr>
<tr>
<td>4. Buying frequency of red and white PDO wines</td>
<td>7-Point scale:&lt;br&gt;1 = 1–3 times a year or less&lt;br&gt;2 = once a month&lt;br&gt;3 = 2–3 times a month&lt;br&gt;4 = once a week&lt;br&gt;5 = 2–3 times a week&lt;br&gt;6 = 4–5 times a week&lt;br&gt;7 = once a day</td>
</tr>
<tr>
<td>5. Average price of the most frequently bought wine</td>
<td>7-Point scale:&lt;br&gt;1 = 2.5 € or less&lt;br&gt;2 = 3–5 €&lt;br&gt;3 = 6–10 €&lt;br&gt;4 = 11–15 €&lt;br&gt;5 = 16–20 €&lt;br&gt;6 = 21–25 €&lt;br&gt;7 = more than 25 €</td>
</tr>
<tr>
<td>7. Innovativeness</td>
<td>Domain Specific Innovativeness scale (from 1 = “strongly disagree” to 7 = “strongly agree”) (Goldsmith &amp; Hofacker, 1991):&lt;br&gt;1. In general, I am among the last in my circle of friends to purchase a new wine&lt;br&gt;2. If I heard that a new wine was available through a local store, I would be interested enough to buy it*&lt;br&gt;3. Compared to my friends, I do little shopping for new wine&lt;br&gt;4. I would consider buying a new wine, even if I hadn’t heard of it yet*&lt;br&gt;5. In general, I am the last in my circle of friends to know the names of the latest wines and wine trends&lt;br&gt;6. I know more about new wines than other people do*</td>
</tr>
</tbody>
</table>

Ratings of the marked items* have been reversed for the mean scores.
Product differences for each attribute were assessed by a three-way ANOVA mixed model using assessor and replicate as random factors, while sample was the fixed factor. A Fisher LSD post hoc test was used to test the significance (p < 0.05) of relative mean differences for the main factors (sample and replicate).

Differences among samples from descriptive analysis were studied by means of Principal Component Analysis (PCA) (Unscrambler version 10.1. Camo). PCA models were computed on panel averages of each significant attribute (p < 0.05) arising from the ANOVA models. Samples were included as dummy variables (down-weighted in the data matrix) to improve the visual interpretation (Martens & Martens, 2001). The full cross validation was computed to validate the interpretation of the first two components.

2.4.2. Mapping data

The mapping data from consumers and experts were analyzed separately and organized into 12 rows representing the wines, and 2 x n (where n is the number of subjects) columns representing the X and Y coordinates of the samples for each subject. Matrices were analyzed by Generalized Procrustes Analysis (GPA) (Gower, 1975) using Senstools 3.3.2. (OP&P & Talcott, Utrecht, The Netherlands). In order to estimate the significance of the GPA results, a Permutation Test was performed.

The relationship between the product coordinates from the two first dimensions of consensus maps and the mean intensity scores from descriptive analysis were studied by linear least square regressions. Similarly, the relationship between the product coordinates from the two first dimensions of consensus maps and the product coordinates from the two first dimensions of the map from descriptive analysis were studied by linear regressions.

The relations between the product coordinates along the first dimension of consensus maps for the consumers and consumer mean liking ratings were analyzed by linear regression analysis. Similarly the relations between the product coordinates along the first dimension of consensus map for the experts and mean quality scores of wines were investigated by linear regression analysis.

For each individual map generated by consumers, the ratio of the distance between the two replicated samples (dr) and the distance between two most distant samples (ds) on the map was computed and expressed as percentage ratio between ds and dr values (D_r%).

2.4.3. Quality, typicality and liking data

Quality and typicality data from experts and liking data from consumers were independently submitted to a two-way ANOVA model, assuming sample and subject as main effects, with a Fisher LSD post hoc test considered significant for p < 0.05.

Individual differences in consumer liking for wines and their relationship with sensory descriptive data were analyzed by means of Principal Component Regression (PCR) using Unscrambler version 10.1 (Camo). For this purpose liking data were used as the X matrix and mean sensory descriptive data as the Y matrix (Internal Preference Map). Samples were included as dummy variables (down-weighted in the X data matrix) to improve the visual interpretation of the results.

2.4.4. Consumer background variables

Individual involvement and innovativeness values were computed as the sum of ratings given to the statements, after the ratings of negative items had been reversed (Goldsmith & Hofacker, 1991; Kähkönen & Tuorila, 1999).

Based on characteristic values of a percentile distribution (first and third quartiles) three subject groups were defined according to three levels of variation (low, medium, and high) of D_r% values (Dinnella, Recchia, Vincenzi, Tuorila, & Monteleone, 2010). In particular the “low” group (G1) was composed with 20 subjects with a D_r% value lower than the first quartile. The “high” (G3) group was composed with 20 subjects with a D_r% value higher than the third quartile, while the “medium” group (G2) was composed with subjects with the 20 D_r% values closest to the median. The associations of groups with age, involvement, and innovativeness (continuous variables) were estimated using one-way ANOVA tests, with a Fisher-LS post hoc test considered significant for p < 0.05, and those with gender, familiarity, education level, consumption and buying frequencies and price range (categorical variables) by a homogeneity chi-square test. Liking data for 12 wine samples expressed by all the three groups were treated with a two-way ANOVA model using group (3 levels) and product (12 levels) as main effects, with a Fisher LSD post hoc test considered significant for p < 0.05.

3. Results

3.1. Perceptual space based on descriptive analysis

Results from the ANOVA model showed a significant sample effect for 8 of the 18 attributes namely, “Blackberry”, “Cherry”, “Prune”, “Wood”, “Vanilla”, “Grass”, “Pears”, and “Asparagus”. No significant effects of replicate and replicate x sample were found, thus panel performance was validated. Non-significant attributes were not included in further data analyses.

Mean intensity data of significant attributes were submitted to Principal Component Analysis (PCA) (Fig. 1). The first two components accounted for 92% of the variation (PC1:64% and PC2:28%). PC1 was positively associated with “herbaceous”/“vegetative” aroma attributes (asparagus, peas and grass) while a negative correlation was found for “fruity” (blackberry, cherry, prune) and “wood” (vanilla and wood) aroma attributes. CS and SG3 samples were separated from the rest of the samples along PC1 and show a positive correlation with herbaceous aroma descriptors. PC2 showed a positive correlation with wood aroma descriptors (wood, vanilla) while a negative correlation was observed with prune aroma. SG5 and SG2 samples were positioned in the upper side of the bi-plot and were associated with wood aroma. The prune descriptor characterized the aromas of PrM and NA samples. Overall the results from PCA indicated that aroma descriptors did not clearly separate Sangiovese wines, nor wines produced in Tuscany, from the rest of the samples.

3.2. Experts evaluations

On average, the number of correct answers (correct association between odor and descriptor) from the smell test for experts was 11 (range 9–12), thus indicating the good odor discrimination ability of these subjects. The mapping data from experts were submitted to a GPA. The Permutation Test indicated a probability of less than 0.1% that the consensus could have arisen by chance.

The consensus map is reported in Fig. 2. The first two dimensions accounted for 36% and 11% of the consensus variance, respectively. The close positioning on the consensus map of replicate samples (SG5a and SG5b) indicated that experts tended to consider these samples to be very similar. In fact D_r% value was lower than 10 for 9 experts. For further 4 experts the D_r% value ranged from 12 to 20. Chianti wines (SG5a, SG5b and SG2) and two wines from southern Italian regions (PrM and NA) were positioned on the opposite direction along PC1, thus indicating that experts perceived strong aroma dissimilarities between these two wine groups. The remaining wines were not clearly separated. In order to understand whether the GPA consensus from experts was related to specific sensory descriptors from conventional profile, a
regression of sensory attributes on GPA dimensions was calculated. The first dimension showed a strong negative relation to “wood” aroma descriptors (wood: \( r = -0.86; p < 0.05 \); vanilla: \( r = -0.79; p < 0.05 \)) while a positive relation to the prune descriptor on sensory differences analytically evaluated. Based on a sensory concept (e.g. quality or typicality) rather than evaluated aroma similarities/dissimilarities among wines on the basis of these results it is possible to hypothesize that experts differentiated along the second dimension of the descriptive sensory plot (Brunello di Montalcino, Nobile di Montepulciano).

The relationship between product scores from the consensus map and quality ratings was investigated. A significant negative relation was found between quality ratings and the product score on the first dimension of the map (\( r = -0.95; p < 0.05 \)). No significant relation was found between product scores on the second dimension and quality ratings. This evidence suggests that product differentiation by experts was mainly based on the perceived overall quality and supports the validity of the hypothesis above. Wood descriptors can be considered as the main sensory attributes that determine the appreciation of Sangiovese wine aroma quality by experts. Typicality scores for Sangiovese wines tended to be higher than those for the other mono-varietal wines with the only exception of SG1 wine. It seems that experts have developed a concept of Sangiovese wine, although the boundary of this concept is not precisely defined. However, the correlation between product scores along the first dimension of consensus map and mean typicality scores was low (\( r = 0.27; p = 0.08 \)).

### 3.3. Consumers’ evaluations

The mean number of correct answers (correct associations between odors and descriptors) from the smell test with consumers was 10.3 (range 6–12). Fourteen subjects correctly matched all odors and descriptors and the number of correct answers was lower than nine in only seven consumers.

Mapping data from consumers were submitted to a GPA (Fig. 3). The Permutation Test indicated a probability of less than 0.1% that the consensus could have arisen by chance. The first two dimensions accounted for 15.2% and 6.5% of the consensus variance, respectively. Along the first dimension, samples NA and PrM were separated from SG4 and SG2, while the rest of samples were positioned close to the center of the model. CS was the only sample differentiated along the second dimension.

Visual inspection of the maps indicated that sample distribution along the first GPA dimension is similar to their positioning on the second component of the descriptive sensory plot (\( r = 0.8; p < 0.01 \)) and the sample distribution along the second GPA dimension is similar to their positioning on the first PCA component (\( r = 0.63; p < 0.05 \)).

**Table 4:** Liking, quality and typicality scores: mean values and standard deviation.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Subjects</th>
<th>Consumer (n = 81)</th>
<th>Experts (n = 13)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Liking</td>
<td>Quality</td>
<td>Typicality</td>
</tr>
<tr>
<td>SG5a</td>
<td>66.2 (14.4)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.8 (1.8)&lt;sup&gt;bcd&lt;/sup&gt;</td>
<td>5.5 (1.9)&lt;sup&gt;ab&lt;/sup&gt;</td>
</tr>
<tr>
<td>SG5b</td>
<td>63.0 (14.1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>7.0 (1.5)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.9 (2.1)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>SG2</td>
<td>61.5 (17.1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6.9 (1.1)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.2 (1.2)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>SG1</td>
<td>59.4 (15.9)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.5 (1.1)&lt;sup&gt;bde&lt;/sup&gt;</td>
<td>4.8 (2.4)&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>SG4</td>
<td>58.9 (20.7)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.5 (2.3)&lt;sup&gt;bde&lt;/sup&gt;</td>
<td>5.1 (2.2)&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>SG6</td>
<td>56.2 (16.0)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.3 (1.4)&lt;sup&gt;de&lt;/sup&gt;</td>
<td>5.4 (2.0)&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>AV</td>
<td>56.1 (13.0)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>4.5 (2.2)&lt;sup&gt;ef&lt;/sup&gt;</td>
<td>4.1 (2.5)&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>PrM</td>
<td>54.5 (18.2)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.3 (1.8)&lt;sup&gt;g&lt;/sup&gt;</td>
<td>4.0 (2.0)&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>BR</td>
<td>52.7 (19.7)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.2 (2.0)&lt;sup&gt;de&lt;/sup&gt;</td>
<td>5.3 (1.8)&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>NA</td>
<td>52.3 (21.0)&lt;sup&gt;cd&lt;/sup&gt;</td>
<td>4.1 (2.4)&lt;sup&gt;ef&lt;/sup&gt;</td>
<td>4.1 (2.2)&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>CS</td>
<td>50.4 (19.1)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.2 (2.4)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.6 (2.4)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>SG1</td>
<td>47.5 (18.0)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.2 (2.6)&lt;sup&gt;de&lt;/sup&gt;</td>
<td>3.9 (1.8)&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Different letters indicate significantly different values (\( p < 0.05 \)).

Ratings scales:
- Liking: “greatest imaginable dislike” (0) – “greatest imaginable like” (100).
- Quality: “very poor” (1) – “excellent” (9).
- Typicality: “bad example of Sangiovese” (1) – “very good example of Sangiovese” (9).
The first dimension of the consensus map was negatively correlated with the attributes wood ($r = -0.81; p < 0.001$) and vanilla ($r = -0.81; p < 0.001$) while the second dimension showed a negative relation to prune aroma descriptor ($r = -0.59; p = 0.04$) and a strong positive relation to grass, peas and asparagus aroma descriptors (grass: $r = 0.65; p = 0.02$; peas: $r = 0.75; p < 0.001$; asparagus: $r = 0.62; p = 0.03$).

Liking scores were submitted to a two-way ANOVA model. A significant ($p < 0.001$) sample effect was found. Mean scores and their relative significant differences are reported in Table 4. No significant correlation between product scores along the first consensus dimension and mean liking scores was found ($r = 0.15$). The most liked wines were samples SG5a, SG5b and SG2 which were the most intense in “wood” and “vanilla” attributes, whereas wines with the lowest liking scores were not clearly related to specific aroma descriptors.

In order to further explore the relationship between liking and sensory profile, a principal component regression (PCR) was computed. This is an Internal Preference Map. The correlation loading plot from the PCR of the eight significant sensory attributes for the 11 wine samples is presented in Fig. 4. The first dimension indicates that consumer liking was oriented towards samples SG4, SG2 and SG5, on the left side of the map in opposition to samples located on the right side of the plot (PrM and NA). Most consumers were located on the left of the first component and their preference was mainly driven by “wood” and “vanilla” attributes. Consumers were widely spread along PC2 in which sample CS was separated from the rest in opposition to sample SG5. Consumers in the bottom left of the plot showed a clear preference toward sample SG5, thus both “woody” and “fruity” notes positively drove the liking of these subjects. General results from the PCR plot showed a good separation among samples as a function of liking data and indicated that sensory attributes drove consumers’ preferences for the wines.

The visual inspection of the preference map and the consumer consensus configuration, showed that the distribution of the wines in both spaces were comparable. The first dimension of the preference map was strongly correlated to the first PM map dimension ($r = 0.86; p < 0.001$). Similarly the second components of the two maps were significantly correlated each-other ($r = 0.56; p = 0.049$). It is evident that the preference map and the consensus map from consumers gave the same information in terms of similarities/dissimilarities among products suggesting that liking was the main criterion used by consumers to evaluate similarities among products. However, it is also evident that differences among samples based only on liking data were easier to interpret than the information from the mapping test.

The consumers’ GPA plot (Fig. 3) and the experts’ map (Fig. 2) show some similarities in terms of the relative positions of the samples. In both cases samples PrM and NA were differentiated along the first dimension, opposing samples SG2 and SG5, whereas samples SG1, SG6, AV and BR were positioned in the central part of the plot. Some differences were also evident. Sample CS was clearly differentiated only along the second dimension in the consumer map whereas the first dimension in the experts’ map contributed its separation. Among the samples located close to the origin of the consumer map was sample SG3 which was, in contrast, clearly distinguished by experts along the second dimension. A significant correlation ($r = 0.8; p < 0.05$) between mean liking and quality scores was found. This evidence may contribute to explain the similarities between the GPA configurations from consumers and experts.

To determine the reasons for the low differentiation level of the mapping test conducted with consumers, the individual maps were visually inspected. There was a great variability among consumers in the positioning of the two replicated wines (SG5a and SG5b) on the map. In order to quantify this variability, for each individual map the $D_{ci}$ value was computed. This value ranged from 3 to 100 with a mean value of 38.6 thus indicating very large individual differences in matching the two replicates. Characteristic values of a percentile distribution were used in order to define three groups of 20 subjects each. Group 1 (G1: $D_{ci} < 14.8$) was composed of subjects who coherently positioned the replicate samples in the map. Group 3 (G3: $D_{ci} > 59.10$), was composed of subjects who did not coherently position the replicate samples in the map and group 2 (G2: $25.8 < D_{ci} < 45.8$) was an intermediate group and included subjects with the twenty $D_{ci}$ values closest to the median.

In order to evaluate the influence of the subject discriminative ability on the consensus map configuration, GPA was independently applied to data provided by consumers belonging to G1, G2 and G3. Results of a GPA model from G1 are shown in Fig. 5.
with a consensus map based on the first two dimensions. The first dimension of the map explained 21%, while the second dimension explained 7.4% of the consensus variance, respectively. Along the first dimension, sample PrM was clearly separated from SG4, SG2, and SG5. The remaining wines from Tuscany (SG1, SG3, SG6) were not differentiated from the other Italian wines. Consumers belonging to G1 exhibited a greater ability to differentiate along the first dimension than all other consumers. In fact, these consumers were more able than other consumers to group together the four Sangiovese wines (SG2, SG4, SG5a, SG5b). The ability to differentiate samples along the first component of G2 subjects was lower than G1 (17.8%). No clear separation of Sangiovese wines was observed for these subjects. For G3 subject the Permutation Test indicated a probability higher than 6% that the consensus maps from the mapping test, quite different sample grouping when PCA plots from the DA data were compared with the consensus maps from the mapping test, quite different sample grouping were observed.

The comparison between perceptual map from the trained panel and from wine experts showed that the second PCA component is closely related to the first GPA dimension. On the other hand, samples distributed along the second GPA dimension differ from those along the first PCA component. In particular, on the PCA plots, samples CS and SG3 are in the same group, whereas in the PM map CS is close to SG2 and SG5, in the opposite position to SG3, along the second dimension.

These differences between the trained and expert panels can be explained by the different criteria used by the two groups to differentiate samples. Despite the significant relationship of the first consensus map dimension with specific wine aroma attributes (wood, vanilla and prune), general results led to the conclusion that experts tended to separate samples mainly on their overall quality, rather than on their specific sensory properties. Wines with different sensory profiles and having similar quality level are closely positioned on the consensus map (SG5, SG2 and CS). In the same way, SG3 and CS samples are both characterized by herbaceous aroma notes but, since their perceived quality levels are different, their positioning on the GPA plot is quite distant. Ballester, Abdi, Langlois, Peyron, and Valentin (2009), in a study conducted in Burgundy (France) on single-grape variety wines, showed that experts are able to categorize wines according to grape variety. According to the authors, experts are supposed to differentiate wines on the basis of their mental representation of odors of wines and their capacity to recognize them. They speculated that expert perception could be enhanced by top down processes in which knowledge on different wine styles and varieties affects similarity assessment. In the present study a product differentiation based on a mental representation of aroma Sangiovese grape volatile precursors contribute to distinctive aroma in wines from different grape cultivars. However, differences in aroma sensory descriptors only partially explain the assessment of aroma similarities/dissimilarities by experts and consumers. In fact, when PCA plots from the DA data were compared with the consensus maps from the mapping test, quite different sample grouping were observed.

The correlation between product scores along the first consensus dimension and mean liking scores from G1 were significant ($r = 0.85; p < 0.001$), showing that consumers who coherently positioned the replicate samples on the map, differentiated the products primarily as a function of their liking.
do not support this hypothesis. In fact the correlation between product scores along the first dimension of consensus map and mean typicality scores was low. Our results seem to suggest that the way experts grouped wines in relation to their similarity and differences is affected by the degree of consensus about what is good quality and what is not. In fact a strong relationship between product scores along the first dimension of the consensus map and mean quality scores was found. Experts seem to mainly refer to a common memorized wine prototype which represents the synthesis of high quality red wines previous tasting experiences.

When experts were asked to perform the mapping test and were left free to use their own criteria in differentiating among samples, it is likely to obtain information similar to that coming from a simple and less time consuming quality assessment. However, mapping performed with experts could be used to explicitly differentiate wines on the basis of their quality. The bi-dimensional feature of PM could provide more informative data regarding quality perception with respect to the quality rating thus possibly allowing for a better exploration of the relationship between the differentiation task result and the sensory profile. As already suggested it would be worthwhile performing a PM test with experienced subjects by providing instruction specifically suited for the aim of the test (Pagès, 2005).

Product differentiation by consumers was worse than that of experts and trained subjects.

When the consumer GPA plot was compared with the trained group’s descriptive sensory plot, it was noted that they agreed well with wines PrM and NA but not with sample SG3. In fact, sample SG3 was grouped with sample CS only and separated from all others wines along the first component in the descriptive sensory plot. However, in the consumers’ GPA plot, this sample was close to the SG origin, thus indicating that consumers did not find clear differences between sample SG3 and the others wines. However, the similarity of samples positioning on the first GPA dimension and on the second PCA component indicates wood and prune aroma descriptors as sensory attributes considered by consumers for wine differentiation.

Consumers’ internal preference map showed a good sample differentiation based on liking data and allowed the identification of the aroma attributes that drove their preferences. Thus, sample differentiation based on liking data appeared to be more informative and easier to interpret than from the map obtained with the PM test. Moreover, the similar sample positioning on the preference map and on the consumer consensus configuration indicates that consumers grouped samples mainly as a function of liking.

Despite the lower differentiation ability observed for consumers, the consumers’ GPA plot and the experts’ map show some similarities in terms of the relative positions of the samples. These similarities could be related to the relationship between the main criteria used by the two subject groups to differentiate wine samples (liking for consumers and quality for experts).

Consumers exhibited large individual differences in their ability to match the two replicate wines on their perceptual maps. Consequently, the poor discriminative ability appears as one of the major drawbacks affecting sample separation. The computation of $D_{\Delta}$ values seems an effective tool for selecting subjects able to differentiate the samples. In fact, when GPA was applied exclusively to the data provided by such consumers (G1) an increased differentiation will be weak if idiosyncratic criteria prevail (Nestrud & Lawless, 2008). The experiential background of subjects was the main factor affecting differentiation, at least in this study, where the sensory differences among samples are subtle. In this case it seems that assessors should share a common high knowledge level of the sample set to share a common criterion for sample differentiation, thus giving significant and interpretable consensus maps.

5. Conclusions

Criteria driving differentiation of high quality red wine on the basis of their aroma similarities/dissimilarities were different in wine experts and naive consumers. Experts seem to mainly refer to a common memorized wine prototype which represents the synthesis of high quality red wines previous tasting experiences. The low consensus level of consumer perceptual space makes it difficult to identify criteria which lead to sample positioning. However, liking can be considered as the main criterion for aroma similarities/dissimilarities evaluation by experienced consumers. In conclusion, it seems that PM is informative about perceived similarities/dissimilarities among samples with subtle sensory differences only when assessors have a high level of experience and knowledge of the products and thus can refer to a common object/category when performing the differentiation task.

References


